

FULL CONDITIONAL DISTRIBUTIONS

The joint posterior distribution of interest can be written as $p(\mathbf{g}, \mathbf{c}, \boldsymbol{\theta} \mid \mathbf{X}, \boldsymbol{\psi})$. Here, \mathbf{g} is the vector of binary gene labels, \mathbf{c} the vector of binary condition labels, $\boldsymbol{\theta}$ the vector of model parameters, $\boldsymbol{\psi}$ the total set of hyperparameters ($\kappa, \nu, s, \varphi, \xi_{g0}, \xi_{g1}, \xi_{c0}, \xi_{c1}$), \mathbf{X} the expression data matrix (with rows referring to genes, and columns to conditions), $B(\xi_{g1}, \xi_{g0})$ the Beta prior distribution on the probability that a gene belongs to the bicluster, $B(\xi_{c1}, \xi_{c0})$ the Beta prior distribution on the probability that a condition belongs to the bicluster,

Derivation of full conditional Bernoulli distribution for the gene labels

1. Assume that the prior probability that a gene (condition) label equals one is given by a Beta distribution:

$$B(\xi_{g1}, \xi_{g0}) \text{ and } B(\xi_{c1}, \xi_{c0})$$

2. The probability distribution of a collection \mathbf{g} of n binary gene labels is then found by integrating out over this (conjugate) Beta prior distribution:

$$\begin{aligned} p(\mathbf{g} \mid \xi_{g1}, \xi_{g0}) &= \int p(\mathbf{g} \mid x) B(x, \xi_{g1}, \xi_{g0}) dx \\ &= \int \prod_{i=1}^n \text{Bern}(g_i, x) B(x, \xi_{g1}, \xi_{g0}) dx \\ &= \int x^{\|\mathbf{g}\|_1} (1-x)^{n-\|\mathbf{g}\|_1} B(x, \xi_{g1}, \xi_{g0}) dx \\ &= \frac{\Gamma(\xi_{g0} + n - \|\mathbf{g}\|_1) \Gamma(\xi_{g1} + \|\mathbf{g}\|_1)}{\Gamma(\xi_{g0} + \xi_{g1} + n)} \end{aligned}$$

n refers to the total number of genes and $\|\mathbf{g}\|_1$ to the one norm of the current (binary) gene label vector (number of genes currently in the bicluster).

3. The derivation for the length m vector of binary condition labels \mathbf{c} is analogous, and leads to:

$$p(\mathbf{c} \mid \xi_{c1}, \xi_{c0}) = \frac{\Gamma(\xi_{c0} + m - \|\mathbf{c}\|_1) \Gamma(\xi_{c1} + \|\mathbf{c}\|_1)}{\Gamma(\xi_{c0} + \xi_{c1} + m)}$$

4. Now consider the full conditional distributions of the gene labels. The conditional for gene label i is modeled by a Bernoulli distribution with parameter α_i , that gives the probability that the label equals one:

$$\begin{aligned}
\alpha_i &= p(g_i = 1 | \mathbf{X}, \mathbf{g}_{\neq i}, \mathbf{c}, \boldsymbol{\theta}, \boldsymbol{\psi}) = p(g_i = 1 | \mathbf{X}, \mathbf{g}_{\neq i}, \mathbf{c}, \boldsymbol{\theta}, \xi_{g_1}, \xi_{g_0}) \\
&= \frac{p(\mathbf{X}, g_i = 1, \mathbf{g}_{\neq i} | \mathbf{c}, \boldsymbol{\theta}, \xi_{g_1}, \xi_{g_0})}{p(\mathbf{X}, \mathbf{g}_{\neq i} | \mathbf{c}, \boldsymbol{\theta}, \xi_{g_1}, \xi_{g_0})} = \frac{p(\mathbf{X}, g_i = 1, \mathbf{g}_{\neq i} | \mathbf{c}, \boldsymbol{\theta}, \xi_{g_1}, \xi_{g_0})}{N} \\
&= \frac{1}{N} p(g_i = 1, \mathbf{g}_{\neq i} | \xi_{g_1}, \xi_{g_0}) p(\mathbf{X}_i | g_i = 1, \mathbf{c}, \boldsymbol{\theta}) \prod_{\substack{k=1 \\ k \neq i}}^n p(\mathbf{X}_k | g_k, \mathbf{c}, \boldsymbol{\theta}) \\
&= \frac{1}{N} p(g_i = 1, \mathbf{g}_{\neq i} | \xi_{g_1}, \xi_{g_0}) \prod_{j|c_j=1} p^{\text{bcl}}(x_{ij}) \prod_{j|c_j=0} p^{\text{bgd},0}(x_{ij}) \prod_{\substack{k=1 \\ k \neq i}}^n p(\mathbf{X}_k | g_k, \mathbf{c}, \boldsymbol{\theta})
\end{aligned}$$

In the above, \mathbf{X}_k indicates the k -th row in the expression matrix \mathbf{X} and $\mathbf{g}_{\neq i}$ refers to the set of all gene labels except for label i .

5. Similarly, an expression for $1-\alpha_i$ can be derived:

$$\begin{aligned}
1-\alpha_i &= p(g_i = 0 | \mathbf{X}, \mathbf{g}_{\neq i}, \mathbf{c}, \boldsymbol{\theta}, \boldsymbol{\psi}) \\
&= \frac{1}{N} p(g_i = 0, \mathbf{g}_{\neq i} | \xi_{g_1}, \xi_{g_0}) \prod_{j|c_j=1} p^{\text{bgd},1}(x_{ij}) \prod_{j|c_j=0} p^{\text{bgd},0}(x_{ij}) \prod_{\substack{k=1 \\ k \neq i}}^n p(\mathbf{X}_k | g_k, \mathbf{c}, \boldsymbol{\theta})
\end{aligned}$$

6. Now, the ratio of those two expressions has a simple form and an elegant interpretation, as discussed in some detail in the main text:

$$\begin{aligned}
\frac{\alpha_i}{1-\alpha_i} &= \frac{p(g_i = 1, \mathbf{g}_{\neq i} | \xi_{g_1}, \xi_{g_0}) \prod_{j|c_j=1} p^{\text{bcl}}(x_{ij})}{p(g_i = 0, \mathbf{g}_{\neq i} | \xi_{g_1}, \xi_{g_0}) \prod_{j|c_j=1} p^{\text{bgd},1}(x_{ij})} \\
&= \frac{\xi_{g_1} + \|\mathbf{g}_{\neq i}\|_1}{\xi_{g_0} + n - \|\mathbf{g}_{\neq i}\|_1} \prod_{j|c_j=1} \frac{p^{\text{bcl}}(x_{ij})}{p^{\text{bgd},1}(x_{ij})}
\end{aligned}$$

The last equation is a direct consequence of the result described under 2.

Missing values are dealt with naturally in this expression, by excluding them from the likelihood ratio calculations (or, equivalently, by assuming the ratio for the corresponding condition equals one). If the expression values for a certain gene are missing in all bicluster conditions, the score of the corresponding gene is completely determined by the prior factor.

Derivation of full conditional Bernoulli distribution for the condition labels

The odds for the Bernoulli parameters of the condition label conditionals can be obtained similarly:

$$\begin{aligned}
\beta_j &= p(c_j = 1 | \mathbf{X}, \mathbf{c}_{\neq j}, \mathbf{g}, \boldsymbol{\theta}, \boldsymbol{\psi}) = p(c_j = 1 | \mathbf{X}, \mathbf{c}_{\neq j}, \mathbf{g}, \boldsymbol{\theta}, \xi_{c1}, \xi_{c0}) \\
&= \frac{p(\mathbf{X}, c_j = 1, \mathbf{c}_{\neq j} | \mathbf{g}, \boldsymbol{\theta}, \xi_{c1}, \xi_{c0})}{p(\mathbf{X}, \mathbf{c}_{\neq j} | \mathbf{g}, \boldsymbol{\theta}, \xi_{c1}, \xi_{c0})} = \frac{p(\mathbf{X}, c_j = 1, \mathbf{c}_{\neq j} | \mathbf{g}, \boldsymbol{\theta}, \xi_{c1}, \xi_{c0})}{M} \\
&= \frac{1}{M} p(c_j = 1, \mathbf{c}_{\neq j} | \xi_{c1}, \xi_{c0}) p(\mathbf{X}_j | c_j = 1, \mathbf{g}, \boldsymbol{\theta}) \prod_{\substack{k=1 \\ k \neq j}}^m p(\mathbf{X}_k | c_k, \mathbf{g}, \boldsymbol{\theta}) \\
&= \frac{1}{M} p(c_j = 1, \mathbf{c}_{\neq j} | \xi_{c1}, \xi_{c0}) \prod_{i|g_i=1} p^{\text{bcl}}(x_{ij}) \prod_{i|g_i=0} p^{\text{bgd},1}(x_{ij}) \prod_{\substack{k=1 \\ k \neq j}}^m p(\mathbf{X}_k | c_k, \mathbf{g}, \boldsymbol{\theta})
\end{aligned}$$

$$\begin{aligned}
1 - \beta_j &= p(c_j = 0 | \mathbf{X}, \mathbf{c}_{\neq j}, \mathbf{g}, \boldsymbol{\theta}, \boldsymbol{\psi}, \xi_{g1}, \xi_{g0}, \xi_{c1}, \xi_{c0}) \\
&= \frac{1}{M} p(c_j = 0, \mathbf{c}_{\neq j} | \xi_{c1}, \xi_{c0}) \prod_{i|g_i=1} p^{\text{bgd},0}(x_{ij}) \prod_{i|g_i=0} p^{\text{bgd},0}(x_{ij}) \prod_{\substack{k=1 \\ k \neq j}}^m p(\mathbf{X}_k | c_k, \mathbf{g}, \boldsymbol{\theta})
\end{aligned}$$

$$\begin{aligned}
\frac{\beta_j}{1 - \beta_j} &= \frac{p(c_j = 1, \mathbf{c}_{\neq j} | \xi_{c1}, \xi_{c0}) \prod_{i|g_i=1} p^{\text{bcl}}(x_{ij}) \prod_{i|g_i=0} p^{\text{bgd},1}(x_{ij})}{p(c_j = 0, \mathbf{c}_{\neq j} | \xi_{c1}, \xi_{c0}) \prod_{i=1}^n p^{\text{bgd},0}(x_{ij})} \\
&= \frac{\xi_{c1} + \|\mathbf{c}_{\neq j}\|_1}{\xi_{c0} + m - \|\mathbf{c}_{\neq j}\|_1} \prod_{i|g_i=1} \frac{p^{\text{bcl}}(x_{ij})}{p^{\text{bgd},0}(x_{ij})} \prod_{i|g_i=0} \frac{p^{\text{bgd},1}(x_{ij})}{p^{\text{bgd},0}(x_{ij})}
\end{aligned}$$

Again, a Beta prior allows introducing prior knowledge on the bicluster size. Compared to the gene label conditional distribution, an additional second likelihood term appears due to the fact that the likelihood ratio for the genes in column j is in general different from one, even if a gene is part of the background. If condition label j is zero, the entire j -th column is captured by one background model (bgd,0). On the other hand, if the j -th column label equals one, the background model that describes the expression of those background genes should only be constructed with the background genes (bgd,1). So, although we keep the model parameters fixed here, strictly speaking there are two background models rather than one single model. In most cases, however, this distinction is primarily a technical issue and the difference between bgd,0 and bgd,1 only matters when very large biclusters occur. In other words, the last factor is approximately equal to one. It is a result of the asymmetrical nature of the patterns we are looking for ('striped' biclusters rather than constant-value biclusters).

Missing values are dealt with in a similar way as for the gene labels.

Derivation of full conditional Bernoulli distribution for model parameters

Due to conjugacy, the full conditional distributions for the model are in the same form as the priors.

1. One can easily derive that for the mean, the product of the normal likelihood with the normal prior (conditioned on σ) leads to a posterior conditional normal distribution for the mean (conditioned on σ).
2. Similarly, combined with a normal likelihood, the scaled inverse χ^2 prior on σ^2 gives rise to a scaled inverse χ^2 posterior on σ^2 . We briefly sketch the derivation:

$$p(\sigma_j^2 | \mathbf{g}, \mathbf{c}, \boldsymbol{\theta}_{\neq \sigma_j^2}, \mathbf{X}, \boldsymbol{\Psi}) = p(\sigma_j^2 | \mathbf{X}_j, \mu_j, \nu, s_j^2, \mathbf{g}) = \frac{p(\mathbf{X}_j | \mu_j, \sigma_j^2, \mathbf{g}) p(\sigma_j^2 | \nu, s_j^2)}{p(\mathbf{X}_j | \mu_j, \nu, s_j^2, \mathbf{g})}$$

Plugging in the expression for the Normal distribution of $p(\mathbf{X}_j | \mu_j, \sigma_j^2)$ and the scaled inverse χ^2 distribution of $p(\sigma_j^2 | \nu, s_j^2)$ yields:

$$p(\mathbf{X}_j | \sigma_j^2, \mu_j, \mathbf{g}) p(\sigma_j^2 | \nu, s_j^2) = Z \sigma_j^{-2\left(\frac{\nu + \|\mathbf{m}\|_1}{2} + 1\right)} \exp\left\{-\frac{\sum_{i|m_i=1} (x_{ij} - \mu_j)^2 + \nu s_j^2}{2\sigma_j^2}\right\}$$

$$p(\mathbf{X}_j | \mu_j, \nu, s_j^2, \mathbf{g}) = Z \int \left[\sigma_j^{-2\left(\frac{\nu + \|\mathbf{m}\|_1}{2} + 1\right)} \exp\left\{-\frac{\sum_{i|m_i=1} (x_{ij} - \mu_j)^2 + \nu s_j^2}{2\sigma_j^2}\right\} \right] d\sigma_j^2$$

$$\propto \left(\frac{\sum_{i|m_i=1} (x_{ij} - \mu_j)^2 + \nu s_j^2}{2} \right)^{\frac{-\nu - \|\mathbf{m}\|_1}{2}} \Gamma\left(\frac{\nu + \|\mathbf{m}\|_1}{2}\right)$$

To simplify notation, we did not add the superscript ‘bcl’ or ‘bgd’. Depending on those superscripts, \mathbf{m} is either the set of genes in the background (1- \mathbf{g}) or in the bicluster (\mathbf{g}) and $\|\mathbf{m}\|_1$ refers to the corresponding number of genes.

Combining the above expressions yields the full conditional posterior distribution for the variance:

$$p(\sigma_j^2 | \mathbf{g}, \mathbf{c}, \boldsymbol{\theta}_{\neq \sigma_j^2}, \mathbf{X}, \boldsymbol{\Psi}) = p(\sigma_j^2 | \mathbf{g}, \nu, s_j^2, \mathbf{X}) \propto \text{Inv-}\chi^2(\eta_j, \zeta_j^2)$$

$$\begin{cases} \eta_j = \nu + \|\mathbf{m}\|_1 \\ \zeta_j^2 = \frac{\nu s_j^2 + \|\mathbf{m}\|_1 \bar{\sigma}_j^2}{\nu + \|\mathbf{m}\|_1} \end{cases}$$

This interpretation of the full conditionals for the model parameters directly motivates the parameterization of the prior, as it allows to interpret both κ and ν as a number of prior observations (sometimes referred to as pseudocounts) associated with the prior. For a discussion on these and other conjugate priors, we refer the reader to standard textbooks on Bayesian statistics (Gelman A.B. et al., 2004).

SCORING MEASURES

Biclustering results were scored with *module recovery* and *bicluster relevance* scores as described in (Prelic et al., 2006). The *module recovery* score indicates how well the gene content of the ‘ideal’ modules is on average reflected in the (best matching bicluster of the) bicluster results:

$$S_{MR}(M_1, M_2) = \frac{1}{M_1} \sum_{(G_1, C_1) \in M_1} \max_{(G_2, C_2) \in M_2} \frac{|G_1 \cap G_2|}{|G_1 \cup G_2|}$$

In the above, M_1 indicates the set of biclusters or modules that are considered ‘ideal’ or ‘real’, M_2 indicates the set of biclusters that are really returned by the algorithm. Hence, module recovery runs over all ‘ideal’ biclusters and averages out the match scores (in terms of gene content G) of the best matching biclusters from the obtained biclustering results.

The *bicluster relevance* score S_{BR} is obtained by interchanging the roles of the ‘ideal’ and reported biclusters. It is related to the relevance of the set of modules in the output. In other words, it penalizes ‘incorrect’ modules in the output.

INITIALIZATION

A dummy gene with a profile equal to the mean profile of the seed genes is added to the data set. When the seed contains more than one gene, this yields an improved initialization for the resolution sweep (start with one dummy gene and all conditions, rather than no genes and no conditions). In all cases, the label of the dummy gene is initially set to 1 while all other gene labels equal 0. When reporting likelihoods, we do not include the likelihood of the dummy gene.

BACKGROUND MEAN AND VARIANCE PRIORS

We do not make the background prior parameters κ^{bgd} and ν^{bgd} available to the user, because any background of interest contains many genes and hence the likelihood (of the background data, including all genes except for those that are currently in the bicluster) and the prior (constructed with all genes except for the seed genes) usually agree. Therefore, in most cases the strength of these priors is rather unimportant. However, to avoid those cases in which very large modules push away the background distribution, we require the background statistics to remain close to the expected ones (equal in terms of means, similar in terms of variance) by setting the number of prior observations for the

background variance (v^{bgd}) equal to the total number of genes by default. The outcome of the algorithm is robust against the precise choice of these background parameters (and more robust if the module signal is stronger).

MODULE SELECTION VIA THE AIC CRITERION

We assume that interesting resolutions are those resolutions for which the corresponding modules score better than modules at neighboring (slightly smaller or larger) resolutions, according to some score function. The most obvious score function is the posterior probability. However, the Conditional Maximization approach only assures we remain in the local posterior mode without knowing, for instance, the height of the distribution in that node. The posterior probability is therefore unknown. A related quantity is the likelihood (of bicluster and background together) given the models at a certain resolution. However, likelihood scores favor models for which the bicluster contains more conditions:

- In a bicluster condition, the expression is modeled by a mixture of two Gaussian distributions (4 parameters), whereas expression in a background condition is modeled by only one Gaussian (2 parameters).
- Therefore, the model complexity increases with the number of conditions in the bicluster.
- A complexer model is expected to fit the data better (even if it is not really a ‘better’ model).

Akaike’s Information Criterion (Akaike, 1974) describes the tradeoff between model fit (likelihood) and model complexity:

AIC score = $2l - 2k$ with l the log likelihood and k the number of model parameters

As such, it represents a sensible choice for the score function. We use $k = 2m + 2\|\mathbf{c}\|_1$ with m the total number of conditions, reflecting 2 parameters for each background probability distribution and 2 additional parameters for each bicluster probability distribution. To limit the functional enrichment analysis, we only considered the most pronounced¹ local optima in the AIC score.

RESOLUTION SWEEP

Setup: For simplicity, assume there is only one gene in the query set and the priors on mean and variance are infinitely strong.

The Conditional Maximization (CM) resolution sweep approach then simply consists of an iteration of the following steps:

- 1) Start with zero variance around the seed profile

¹ If $(\text{aic})_i > (\text{aic})_k$ for $k \in \{i-2, i-1, i+1, i+2\}$, the optimum in i was considered more pronounced than if the condition only holds for $k \in \{i-1, i+1\}$.

- 2) Apply a posterior mode search algorithm (via iteration of conditional maximization steps)
- 3) After convergence, increase the prior variance slightly and repeat the mode search. Some conditions may be lost (or possibly gained) and some genes may be gained (or possibly lost).
- 4) Iterate the previous steps over a prespecified resolution range
- 5) Report ‘interesting’ modules by studying the evolution of the Akaike Information Criterion

Figure 1 illustrates how the general setup looks like in the probability landscape. Figure 2 shows an example plot on an artificial data set in scenario S2A (noise level 0.06), using gene 1 and 2 as query. The correct module is characterized by a very pronounced local optimum in the AIC score, as shown by the colored marks. Figure 3 shows a similar plot on an artificial data set in scenario S1B (overlap parameter 7), using genes. Again, the correct module corresponds to a local optimum in the likelihood.

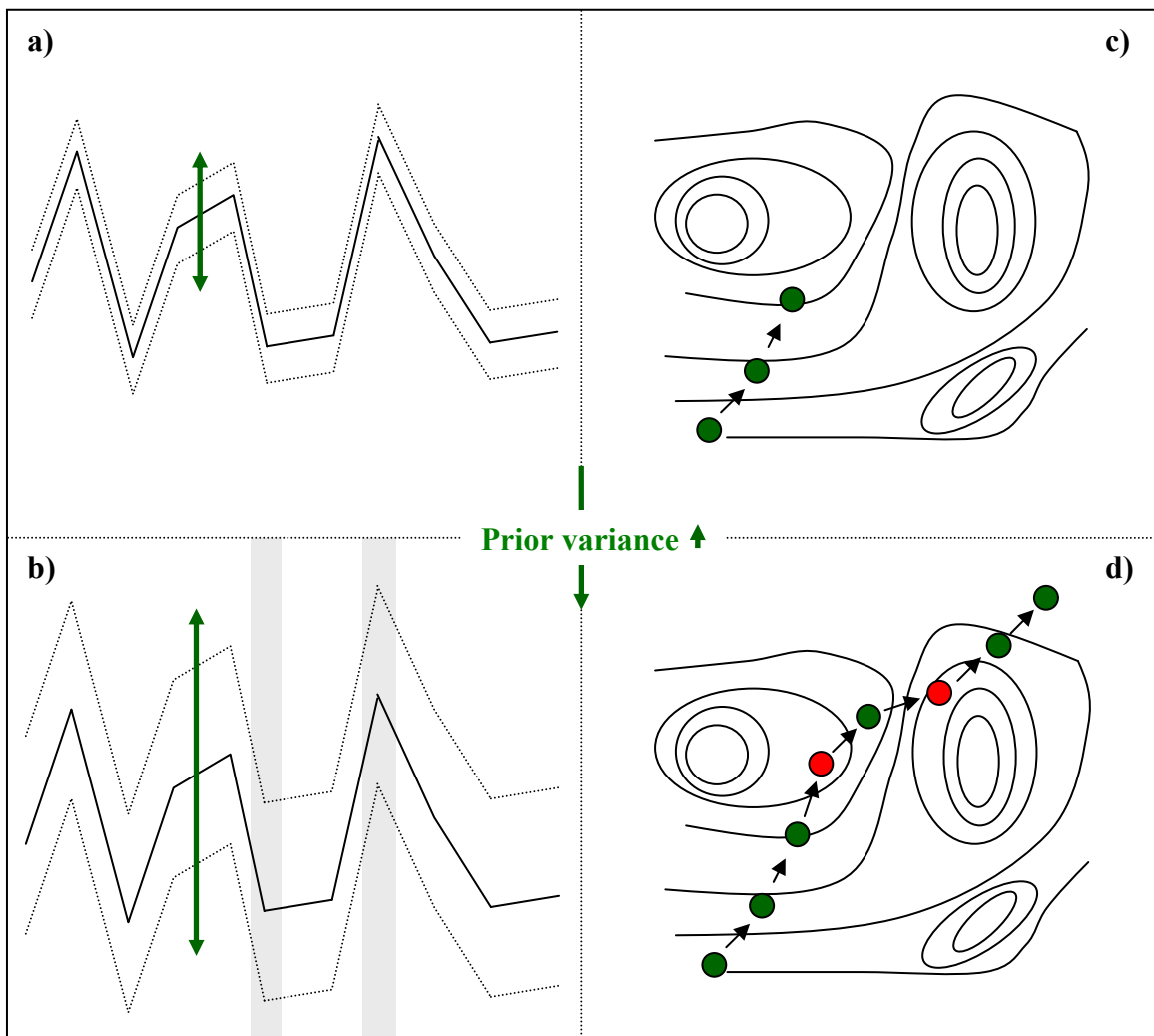


Figure 1: General resolution sweep setup. a) represents a situation with a small prior variance. From a) to b) the variance is gradually increased and some conditions (shaded gray) may no longer belong to the module. c) and d) show the (hypothetical 2D) corresponding AIC score function and the

evolution in the corresponding landscape upon variation of the prior variance. Every dot corresponds to a local optimum in the posterior for the corresponding prior setting. However, the posterior landscape itself remains unknown². Therefore, a reasonable way to select interesting modules on the resolution sweep path is to identify local optima (red dots) in the AIC score on the resolution sweep path.

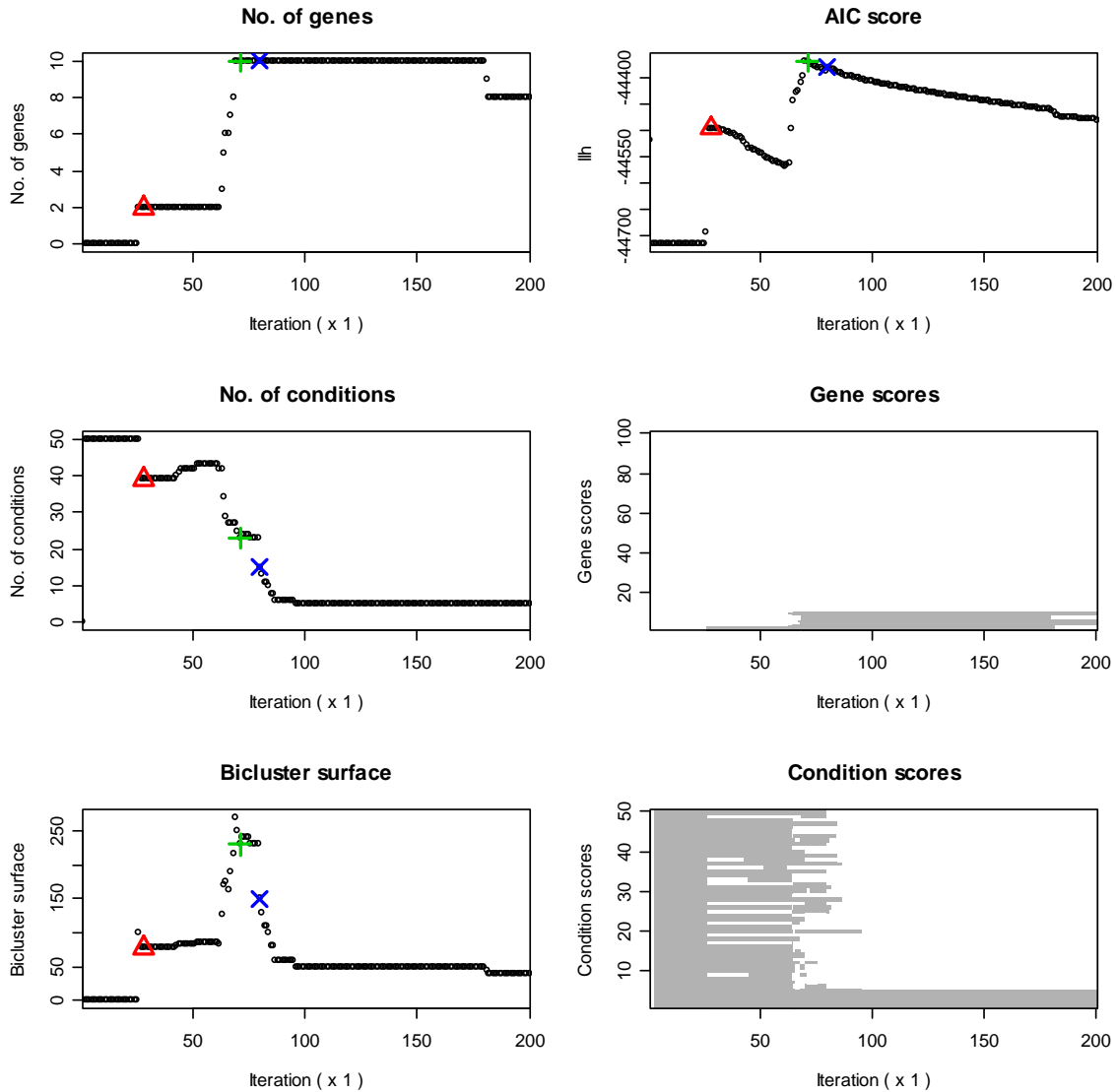


Figure 2: Example plot on an artificial data set in scenario S2A (noise level 0.06), using gene 1 and 2 as query (200 iteration). The correct module is characterized by a very pronounced local optimum in the AIC score, as shown by the colored marks.

² we only assure we remain in the local posterior mode without knowing, for instance, the height of the distribution in that node

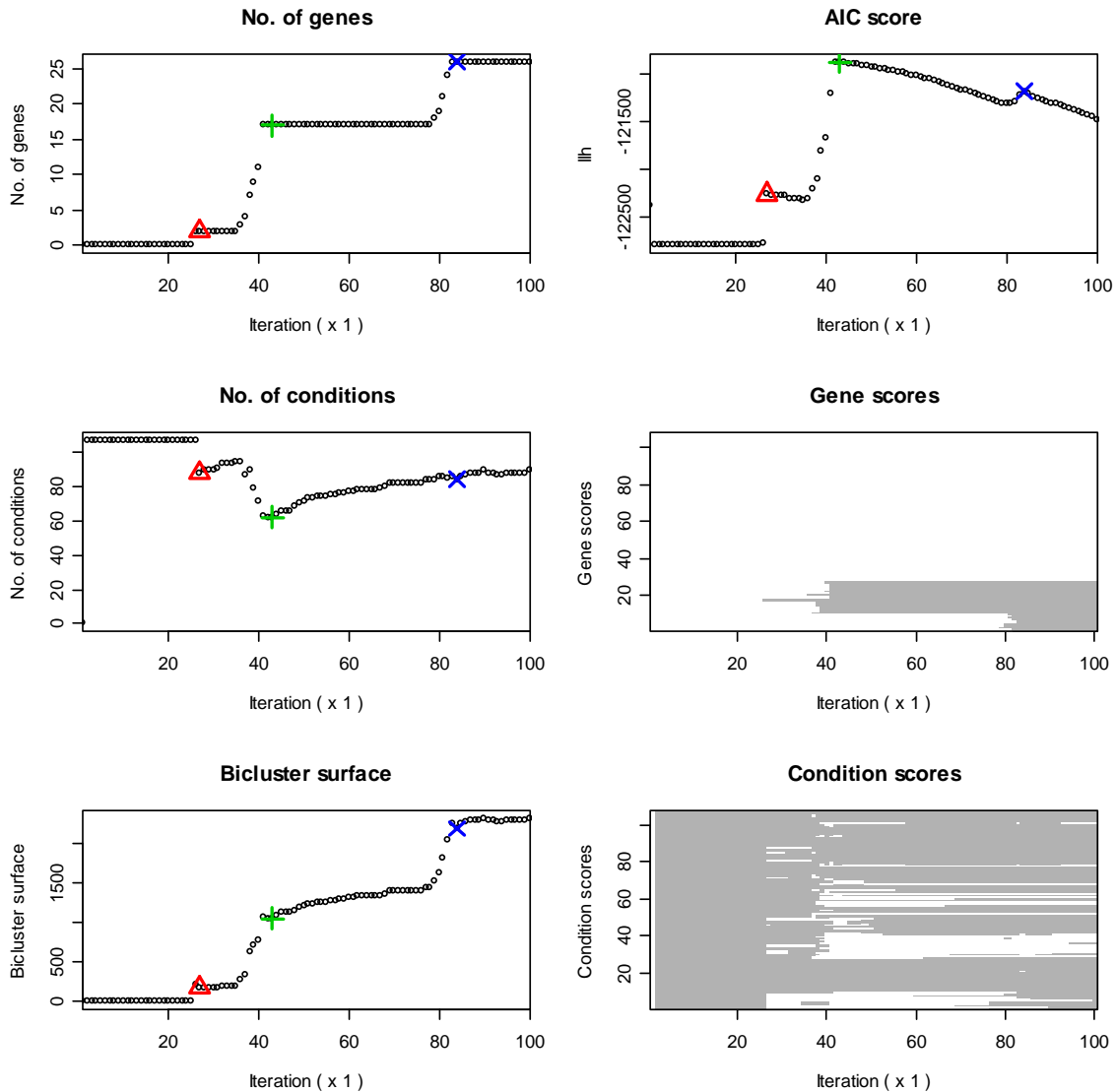


Figure 3: Example plot on an artificial data set in scenario S1B (overlap parameter 7), using gene 17 and 18 as query (100 iterations). The correct module is characterized by a very pronounced local optimum in the AIC score, as shown by the colored marks. As the prior variance increases, the bicluster evolves from a module of 17 genes to the union of two modules (27 genes = 17 genes of module 1 + 17 genes of module 2 – 7 genes in overlap) around iteration 80. This new module also corresponds to a local optimum in the AIC score, but the absolute value of its score is smaller.

AN INTUITIVE COMMENT ON THE NOTION OF RESOLUTION

In order to understand the effect of the resolution sweep, it is important to realize that we do not define resolution in terms of the number of genes or conditions in the bicluster directly. Instead, ‘resolution’ is related to the notion of bicluster homogeneity or coherence, reflected by the variances of the Gaussian distributions for the bicluster expression values in the selected conditions (see general introduction and introduction to section 3 for a definition of ‘resolution’). Of course, if the homogeneity constraint is

made more stringent, the corresponding valid patterns also tend to contain less entries of the data matrix.

The sweep is initialized with small variance (high resolution, good homogeneity) and the corresponding bicluster contains few genes (one gene at initialization) exhibiting tight coexpression over many conditions. When the variance is increased, less homogeneous patterns are allowed. This is what we refer to as ‘lower resolution’. Intuitively, it is clear that these patterns can contain more genes because more profiles may fit well in a wider band around the mean seed profile. At the same time, for a number of conditions the corresponding profiles may not fit well enough (as quantified by the likelihood ratio of the bicluster distribution fit to the background distribution fit) and hence the number of conditions decreases in most cases, especially when there is a crisp increase in the number of genes.

In addition, it is important to realize that the algorithm always compares the model fit of the bicluster and the background distributions. An increase in bicluster variance (decrease in resolution) makes the bicluster distributions wider but also less peaked (lower). This change affects the bicluster content via the likelihood ratio (between bicluster and background) in the full conditional distributions for the labels. Therefore, it is not necessarily so that an increase in resolution corresponds to an increase in bicluster size. However, as pointed out above, there is often a connection between resolution and the number of data matrix entries in the bicluster patterns.

In summary, we relate resolution to coherence (of expression values in the included conditions) rather than the number of genes or conditions in the bicluster. The bicluster size at various resolutions depends on the local optimum in the posterior that we are tracking.

TIME COMPLEXITY

The time complexity per iteration is linear in the number of conditions and the number of genes. Indeed, all algorithmic steps scale linearly with the number of entries in the data matrix. This is also illustrated in Figure 4.

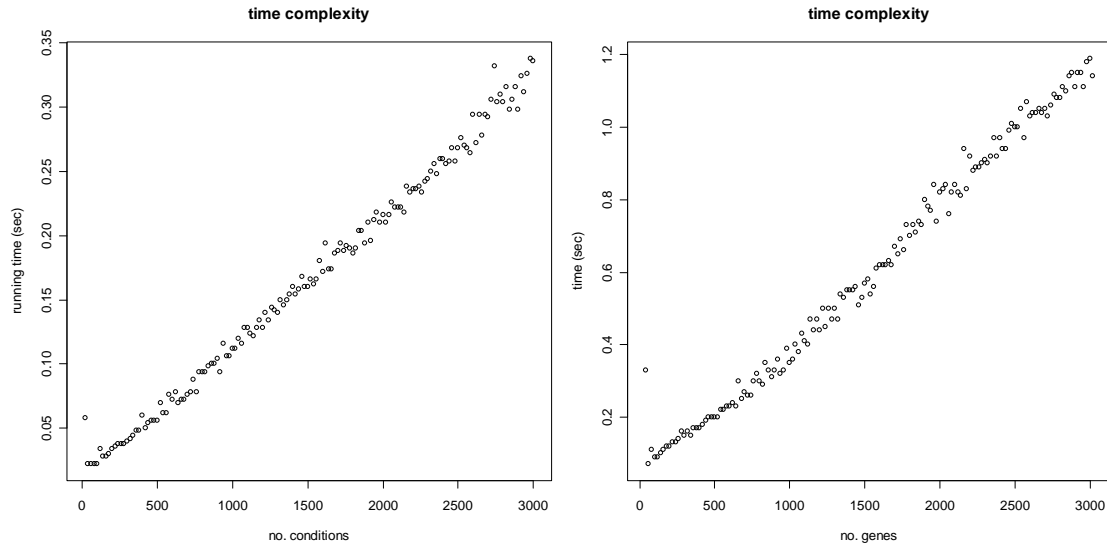


Figure 4: Computational complexity increases linearly with number of conditions for fixed number of genes (left) and with number of genes for fixed number of conditions (right).

ITERATIVE SIGNATURE ALGORITHM

The Iterative Signature Algorithm (Bergmann *et al.*, 2003) intuitively defines a bicluster as a set of genes and conditions with significant average overexpression or underexpression of the bicluster genes in the bicluster conditions. The exact definition relies on fixed points of an alternating iterative procedure in which genes are scored with respect to the condition set and conditions with respect to the gene set. In every step, the scores are thresholded using two user defined parameters (one for the gene scores and one for the condition scores) to eliminate noise and enforce sparseness. The initial gene set determines to which fixed point the algorithm converges. One can either use a clever initialization (query-based), or many random initializations (global biclustering).

GENE RECOMMENDER

The Gene Recommender algorithm (Owen *et al.*, 2003) consists of two steps. In a first step, experimental conditions are scored with a preference for experiments with extreme expression levels and experiments with tight clustering of expression levels. A small grid of condition score thresholds is explored. In a second step, genes are scored based on the statistical significance of the correlation of their expression profiles with the query in the selected conditions. The algorithm is able to work with missing values and returns a ranked list of genes.

REFERENCES

- Bergmann,S. et al. (2003) Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.*, **67**, 031902.
- Gelman A.B. et al. (2004) *Bayesian data analysis*. Chapman & Hall/CRC.
- Owen,A.B. et al. (2003) A gene recommender algorithm to identify coexpressed genes in *C. elegans*. *Genome Res.*, **13**, 1828-1837.
- Prelic,A. et al. (2006) A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics.*, **22**, 1122-1129.
- Akaike,H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control.*, **19**(6), 716–723.