

Gene expression

Shifting and scaling patterns from gene expression data

Jesús S. Aguilar-Ruiz

BIGS Bioinformatics Group Seville, University of Seville, Pablo de Olavide University, Spain

Received on May 24, 2005; revised on August 7, 2005; accepted on August 22, 2005

Advance Access publication September 6, 2005

ABSTRACT

Motivation: During the last years, the discovering of biclusters in data is becoming more and more popular. Biclustering aims at extracting a set of clusters, each of which might use a different subset of attributes. Therefore, it is clear that the usefulness of biclustering techniques is beyond the traditional clustering techniques, especially when datasets present high or very high dimensionality. Also, biclustering considers overlapping, which is an interesting aspect, algorithmically and from the point of view of the result interpretation. Since the Cheng and Church's works, the mean squared residue has turned into one of the most popular measures to search for biclusters, which ideally should discover shifting and scaling patterns.

Results: In this work, we identify both types of patterns (shifting and scaling) and demonstrate that the mean squared residue is very useful to search for shifting patterns, but it is not appropriate to find scaling patterns because even when we find a perfect scaling pattern the mean squared residue is not zero. In addition, we provide an interesting result: the mean squared residue is highly dependent on the variance of the scaling factor, which makes possible that any algorithm based on this measure might not find these patterns in data when the variance of gene values is high. The main contribution of this paper is to prove that the mean squared residue is not precise enough from the mathematical point of view in order to discover shifting and scaling patterns at the same time.

Contact: aguilar@lsi.us.es

1 INTRODUCTION

Clustering of microarray expression data can lead to molecular classification of disease states, identification of co-fluctuation of functionally related genes, functional groupings of genes and logical descriptions of gene regulation, among others. Understanding the direct correlation patterns of genes with experimental conditions is a starting point for understanding the large-scale network that they comprise. The qualitative or quantitative correlation of genes with experimental conditions is a key to understand the biological interactions among genes in a better way. Clustering has been applied to gene expression data (Ben-Dor *et al.*, 1999), which usually refers to conditions or patients, although genes can also be grouped in order to search for functional similarities. However, relevant genes are not necessarily related to every condition, or in other words, there are genes that might be relevant for a subset of conditions (Wang *et al.*, 2002).

Clustering methods can be applied to either the rows or the columns of the data matrix, separately. Biclustering methods, on the other hand, perform simultaneous clustering of rows and columns. Each bicluster is selected using only a subset of the attributes

(genes) and each attribute in a bicluster is selected using only a subset of the examples (conditions) (Cheng and Church, 2000). The goal of biclustering techniques is thus to identify subgroups of examples and subgroups of attributes, by performing simultaneous clustering of both rows and columns of the data matrix, instead of clustering rows and columns separately. Biologically, biclustering is a very interesting technique, as it is possible to discriminate groups of conditions by using different groups of genes, i.e. to identify groups of genes that show a 'similar' level of expression (or more generally, similar behavior) under a specific subset of experimental conditions.

The mean squared residue was the first measure used to find biclusters in gene expression data (Cheng and Church, 2000), and it is still very popular. It is based on the variability of the neighborhood with regard to the arithmetic mean. Therefore, when all the elements in a submatrix are similar, the mean squared residue is low. However, biclusters should contain subset of genes showing similar behavior, not similar values. Thus, other types of patterns, such as shifting and scaling patterns, should be found in biclusters that were obtained by using the mean squared residue. Interestingly, this is not so, because the mean squared residue is not a good measure to evaluate scaling patterns. Therefore, there exists a range of values for the mean squared residue, for which those biclusters will contain poor shifting patterns or good scaling patterns, and we might not be able to discriminate them.

In this work, we present the analytical result of observing the general behavior of the shifting and the scaling pattern with regard to the mean squared residue. We demonstrate that shifting patterns can be discovered by the mean squared residue, but scaling patterns cannot be discovered. This is due to the effect of the scaling variance on the measure.

The paper is organized as follows: Section 2 presents several approaches to discover biclusters in data; the basic concepts are defined in Section 3; the analysis of the mean squared residue is given in Section 4, demonstrating its strength and weakness with regard to the shifting and scaling patterns, respectively; in Section 5 we discuss the consequences of the theorems presented in this work, providing examples of the effect of shifting and scaling factors on the mean squared residue; finally, the main conclusions are summarized in Section 6.

2 RELATED WORK

Biclustering is a NP-hard problem, considered first by Morgan and Sonquist (1963), and subsequently by Hartigan (1972) and by Mirkin (1996). In the context of microarray analysis it was considered by Cheng and Church (2000), who proposed the biclustering

of gene expression, introducing the residue of an element in the bicluster and the mean squared residue of a submatrix. In addition, they adjusted that measure to reject trivial biclusters by means of the row variance. In Getz *et al.* (2000) the authors presented the coupled two-way clustering. It uses hierarchical clustering applied separately to each dimension. Then they defined the process to combine both results. Obviously, the quality of biclusters depends on the clusters generated at each dimension, which in turn, allow us to experiment with different types of clustering algorithms. In Lazzeroni and Owen (2002) they used ‘plaid models’ in the same context, where the concept of ‘layers’ (bicluster) is used to compute the values in the data matrix, which is described as a linear function of layers. Basically, each element is seen as a superposition of layers. In Yang *et al.* (2002) it was presented as δ -clusters, and a year later, the same authors improved the Cheng and Church’s approach in FLOC (Yang *et al.* 2005), paying attention to missing values. FLOC follows the same technique as Cheng and Church’s algorithm, by adding/removing each row/column to a set of initial biclusters, improving its quality iteratively. Also in Tanay *et al.*, (2002) the authors identified biclusters by means of a bipartite graph-based model and using a greedy approach to add/remove vertices in order to find maximum weight subgraphs, which is related to its statistical significance. An approach based on projected clusters is presented in Yip *et al.* (2004), in which the algorithm HARP includes a measure, called relevance index, that uses local and global variances along dimensions. Cho *et al.* (2004) proposed an algorithm, named COCLUS, also based on the mean squared residue, to find non-overlapping biclusters. Although slightly different, the xMOTIF (conserved gene expression motif) is also a special case of bicluster. A probabilistic algorithm to find xMOTIFs is presented in Murali and Kasif (2003). Other heuristics, such as simulated annealing, have been also tested together with the mean squared residue (Bryan *et al.* 2005).

Many of these algorithms find exclusive biclusters and that is inappropriate in the biological context, because genes can contribute to different biological processes. The search method (random, exhaustive, heuristic, etc.) and the evaluation function (mean squared residue, relevance index, etc.) are critical in finding high-quality biclusters in gene expression data.

3 DEFINITIONS

Throughout the paper, we will refer to examples as conditions and attributes as genes. Next, we will give some necessary definitions.

3.1 Bicluster

DEFINITION 1 (microarray). A microarray is defined as a triplet $\mathcal{M} = (\mathcal{C}, \mathcal{G}, \ell)$, where \mathcal{C} and \mathcal{G} are two finite sets referred to as the set of conditions and the set of genes, respectively, and $\ell: \mathcal{C} \times \mathcal{G} \rightarrow \mathbb{R}$ is the level function.

A microarray $\mathcal{M} = (\mathcal{C}, \mathcal{G}, \ell)$, is therefore defined by $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$, $\mathcal{G} = \{g_1, g_2, \dots, g_m\}$ and the level function ℓ , where

$$(\ell(c_i, g_j)) = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \dots & \dots & \dots & \dots \\ v_{n1} & v_{n2} & \dots & v_{nm} \end{pmatrix}.$$

Figure 1 shows an example of microarray, where conditions are represented in the x-axis, and the values of gene expression are

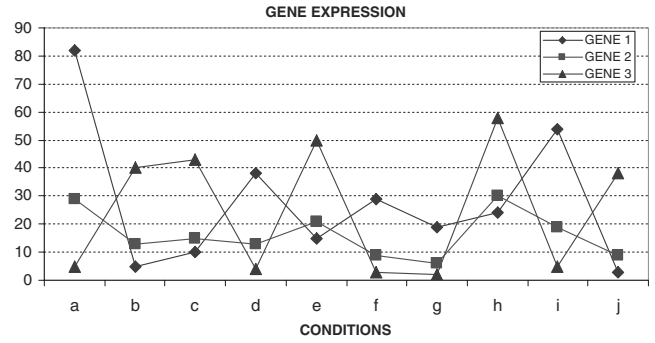


Fig. 1. Example of a dataset.

represented in the y-axis. In this example, there are three genes {gene 1, gene 2, gene 3} and ten conditions {a,b,c,d,e,f,g,h,i,j}. Note that the order in which the conditions are drawn in the x-axis is not important, as we are just interested in the shape of each gene.

DEFINITION 2 (bicluster). A bicluster is a triplet $\mathcal{B} = (\mathcal{I}, \mathcal{J}, h)$, where $\mathcal{I} \subseteq \mathcal{C}$, $\mathcal{J} \subseteq \mathcal{G}$ and $h: \mathcal{I} \times \mathcal{J} \rightarrow \mathbb{R}$ is the level function of the bicluster, such as $h(c_i, g_j) = \ell(c_i, g_j)$, $\forall (c_i, g_j) \in \mathcal{I} \times \mathcal{J}$.

A bicluster is a subset of genes that exhibit similar behavior across a subset of conditions, and vice versa. The problem of discovering a set of biclusters \mathcal{B}_k from a microarray \mathcal{M} is determined by some specific measure of homogeneity, which assigns a figure of merit to biclusters that identify valid patterns in the data.

From now on, let us suppose that a bicluster $\mathcal{B} = (\mathcal{I}, \mathcal{J}, h)$ is defined by the function h as follows:

$$(h(c_i, g_j)) = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1J} \\ w_{21} & w_{22} & \dots & w_{2J} \\ \dots & \dots & \dots & \dots \\ w_{I1} & w_{I2} & \dots & w_{IJ} \end{pmatrix},$$

where w_{ij} is $\ell(c_i, g_j)$, for some $i' \in \{1, \dots, n\}$ and $j' \in \{1, \dots, m\}$, and $i \in \{1, \dots, I\}$ and $j \in \{1, \dots, J\}$. Henceforth, by simplicity, a bicluster will be defined by the pair $(\mathcal{I}, \mathcal{J})$, enumerating the conditions and genes that compose the submatrix.

3.2 Patterns

A perfect bicluster is a submatrix $(\mathcal{I}, \mathcal{J})$ where all the values are a constant π . A perfect bicluster with constant conditions is a submatrix $(\mathcal{I}, \mathcal{J})$ where all the values within the bicluster can be obtained using one of the following expressions:

$$w_{ij} = \pi + \alpha_i \quad (1)$$

$$w_{ij} = \pi \times \alpha_i, \quad (2)$$

where π is the typical value within the bicluster and α_i is the adjustment for condition $i \in I$. This adjustment can be obtained either in an additive (1) or multiplicative way (2).

Similarly, a perfect bicluster with constant genes is a submatrix $(\mathcal{I}, \mathcal{J})$, where all the values within the bicluster can be obtained using one of the following expressions:

$$w_{ij} = \pi + \beta_j \quad (3)$$

$$w_{ij} = \pi \times \beta_j, \quad (4)$$

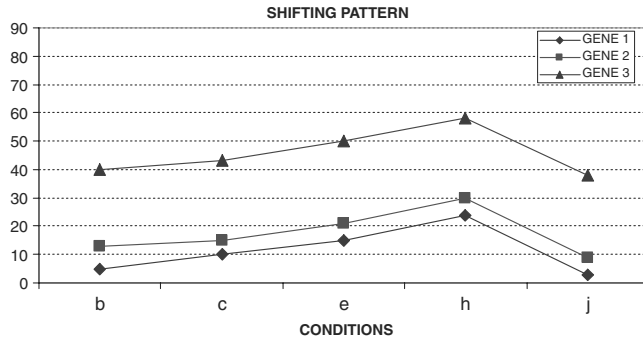


Fig. 2. Shifting pattern from the dataset in Figure 1.

where π is the typical value within the bicluster and β_j is the adjustment for gene $j \in J$. This adjustment can be obtained either in an additive (3) or multiplicative way (4).

A more general view of biclusters includes those where a subset of genes and a subset of conditions have coherent values on both rows and columns. A perfect bicluster with coherent values, $(\mathcal{I}, \mathcal{J})$, is defined as a subset of rows and a subset of columns, whose values w_{ij} are predicted using one of the following expressions:

$$w_{ij} = \pi + \alpha_i + \beta_j \quad (5)$$

$$w_{ij} = \pi \times \alpha_i \times \beta_j. \quad (6)$$

Biologically, the interest resides in the subsets of conditions that present similar behavior across a subset of genes. In this case, we usually do not find constant behaviors, but variable ones. Therefore, we should replace the constant π above with the value π_{ij} . Fortunately, the behavior that a gene expresses across a subset of conditions should be similar to that of another gene if both are related, which in turn simplifies the patterns regarding the π values. This means that we can typically find two types of patterns between two (or more) related genes: shifting and scaling patterns.

DEFINITION 3 (shifting pattern). A bicluster shows a shifting pattern when it follows the expression

$$w_{ij} = \pi_j + \beta_i. \quad (7)$$

In Figure 2, a bicluster obtained from the dataset shown in Figure 1, which contains a shifting pattern, is illustrated.

DEFINITION 4 (scaling pattern). A bicluster shows a scaling pattern when it follows the expression

$$w_{ij} = \pi_j \times \alpha_i. \quad (8)$$

In Figure 3, a bicluster obtained from the dataset shown in Figure 1, which contains a scaling pattern, is illustrated.

As it can be observed in Figures 2 and 3, both shifting and scaling patterns are very interesting from the biological point of view. Sometimes, related genes express the same way, with the only difference that they started with different initial values. This produces that shapes are similar, although values are not equal. It is also likely to find genes that present the same behavior with regard to the regulation, but not with the same energy, i.e. changes are more abrupt for one gene than for the other. This scaling behavior is more difficult to detect, but it is more probable in nature. Regulatory pathways seem to show this type of behavior.

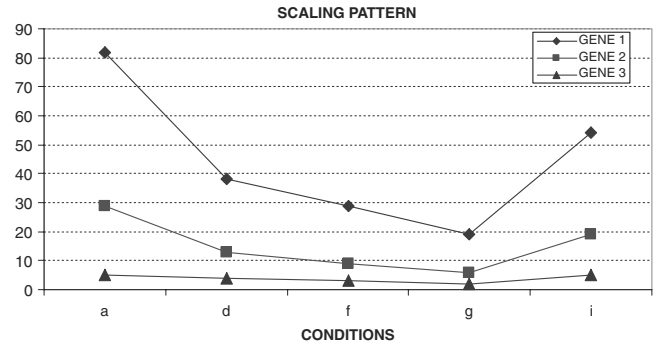


Fig. 3. Scaling pattern from the dataset in Figure 1.

3.3 Mean squared residue

The residue is an indicator of the degree of coherence of an element with respect to the remaining ones in the bicluster, given the tendency of the relevant gene and the relevant condition. The lower the residue, the stronger the coherence.

DEFINITION 5 (residue). The residue R of an entry w_{ij} of a bicluster $\mathcal{B} = (\mathcal{I}, \mathcal{J}, h)$ is

$$R(w_{ij}) = w_{ij} - w_{iI} - w_{Ij} + w_{IJ}, \quad (9)$$

where w_{iI} is the mean of the i -th row,

$$w_{iI} = \frac{1}{|J|} \sum_{j \in J} w_{ij},$$

w_{Ij} is the mean of the j -th column,

$$w_{Ij} = \frac{1}{|I|} \sum_{i \in I} w_{ij},$$

and w_{IJ} is the mean of all elements in the bicluster,

$$w_{IJ} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} w_{ij}.$$

The mean squared residue measures the variance of the set of all elements in the bicluster, plus the mean row variance and the mean column variance. The lower the mean squared residue, the stronger the coherence exhibited by the bicluster, and the better the quality of the bicluster.

DEFINITION 6 (mean squared residue). The mean squared residue H of a bicluster $\mathcal{B} = (\mathcal{I}, \mathcal{J}, h)$ is

$$H(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} R^2(w_{ij}). \quad (10)$$

If the bicluster has a mean squared residue lower than a given value δ , then we call the bicluster a δ -bicluster. The problem of finding the largest square δ -bicluster is NP-hard (Cheng and Church, 2000). In addition to the mean squared residue, we may prefer the row variance to be relatively large to reject the trivial bicluster. This can help to rank biclusters when they show similar mean squared residue. Several works adopt this approach to limit the number of biclusters (Cheng and Church, 2000; Yang *et al.*, 2002). The only problem is that the value of δ has to be estimated in advance, and it is different for every dataset (Tavazoie *et al.* 1999).

4 ANALYSIS

In order to show the behavior of the mean squared residue, we will consider a bicluster that includes both types of patterns: shifting and scaling. Next, we will analyze which result would be for a perfect bicluster including both the patterns.

Let $\mathcal{B} = (\mathcal{I}, \mathcal{J}, h)$ be the bicluster defined by level function h , where

$$(h(c_i, g_j)) = \begin{pmatrix} \pi_1 \alpha_1 + \beta_1 & \pi_2 \alpha_1 + \beta_1 & \cdots & \pi_J \alpha_1 + \beta_1 \\ \pi_1 \alpha_2 + \beta_2 & \pi_2 \alpha_2 + \beta_2 & \cdots & \pi_J \alpha_2 + \beta_2 \\ \cdots & \cdots & \cdots & \cdots \\ \pi_1 \alpha_I + \beta_I & \pi_2 \alpha_I + \beta_I & \cdots & \pi_J \alpha_I + \beta_I \end{pmatrix}, \quad (11)$$

where α_i is a scaling factor, β_i is a shifting factor for each condition i and π_j are the base values for each gene j . Note that each column from this matrix is represented as a piece-wise linear function in Figures 1–3.

Let

$$\mu_\pi = \frac{1}{|J|} \sum_{j=1}^{|J|} \pi_j$$

be the mean of the values π_j ,

$$\mu_\alpha = \frac{1}{|I|} \sum_{i=1}^{|I|} \alpha_i$$

be the mean of the values α_i and

$$\mu_\beta = \frac{1}{|I|} \sum_{i=1}^{|I|} \beta_i$$

be the mean of the values β_i .

The mean of a row (condition) i is $\mu_{r_i} = \alpha_i \mu_\pi + \beta_i$, and the mean of a column (gene) j is given by $\mu_{c_j} = \pi_j \mu_\alpha + \mu_\beta$.

THEOREM 1. *A shifting pattern has no effect on the residue.*

PROOF. In the bicluster illustrated in Equation (11), $w_{ij} = \pi_j \alpha_i + \beta_i$, $w_{iI} = \mu_{r_i} = \alpha_i \mu_\pi + \beta_i$, $w_{IJ} = \mu_{c_j} = \pi_j \mu_\alpha + \mu_\beta$ and $w_{IJ} = \mu_\pi \mu_\alpha + \mu_\beta$. Substituting these expressions in Equation (9), we have

$$R(w_{ij}) = (\pi_j - \mu_\pi)(\alpha_i - \mu_\alpha). \quad (12)$$

We can observe that the factor β is not present. Therefore there is no influence of the shifting pattern on the value of the residue. In other words, the residue cannot differentiate quantitatively two shifting patterns from each other.

The result presented in Theorem 1 is interesting, since based on it we would not be able to rank shifting patterns in order to measure their quality.

COROLLARY 1. *The perfect bicluster with a shifting pattern has mean squared residue equal to zero.*

PROOF. If there is a shifting pattern in the data (and no scaling patterns), the value of the residue is zero, since all the α_i are constant and the value of μ_α would also be the same constant. As the residue of any element in the bicluster is zero, then the mean squared residue is also zero.

THEOREM 2. *A scaling pattern has significant effect on the mean squared residue.*

PROOF. Substituting Equation (12) in Equation (10), we have the following expression:

$$H(I, J) = \sigma_\alpha^2 \sigma_\pi^2, \quad (13)$$

where σ_α^2 is the variance of the scaling factor α and σ_π^2 is the variance of the values π in the submatrix.

The result of Theorem 2 is extremely important, as the quality of a bicluster based on the mean squared residue depends on the product of two variances: the scaling factor and the values of π , and it does not depend at all on the shifting factor.

5 DISCUSSION

Both theorems extend the results of Cheng and Church (2000) and have important consequences on the search for biclusters.

Theorem 1 states that the shifting pattern has no effect on the residue. In Cheng and Church (2000) authors stated (without proof) that a shifting (addition by a constant) to the matrix would not affect the mean squared residue. They referred to a global shifting of the matrix, i.e. when the values follow the expression $w_{ij} = \pi_j + \alpha$. In this paper we have proved that not only a global shifting but also a local shifting, i.e. when the values follow the expression $w_{ij} = \pi_j + \alpha_i$, is not affected by the score. This means that every experimental condition might suffer a different shifting without any impact on the final score.

EXAMPLE 1. Figure 4 illustrates an example of a perfect bicluster, including shifting and scaling patterns. The values of π , α and β , and the values of the submatrix are shown. Figure 5 shows the effect of a global shifting in which all the β values are 50. The mean squared residue of this submatrix is exactly the same as the one shown in Figure 4. It is obvious that not only a global shifting but also local shiftings have no effect on the score.

Cheng and Church (2000) authors stated (without proof) that a scaling (multiplication by a constant) might affect the score. They referred to a global scaling of the matrix, i.e. when the values follow the expression $w_{ij} = \pi_j \alpha$. We have proved here, in contrast with the above statement, that a global scaling has no effect, as if the variance of α is zero, then the mean squared residue is also zero (see Theorem 2). However, in the case of local scaling, i.e. when the values follow the expression $w_{ij} = \pi_j \alpha_i$, then the effect of the variance of α is very important.

EXAMPLE 2. Figure 6 shows the effect of a global scaling. All the α values remain constant ($\alpha = 10$). Therefore, the mean squared residue is zero. In this case, the gene values are shifted, representing a perfect shifting pattern. However, the case of local scaling is significant, as depicted in Figures 4 and 5.

Theorem 2 goes further, since it demonstrates that when both shifting and scaling patterns are present in the data the mean squared residue only takes into account the variance of the scaling factor and the variance of the base values π . The product in Equation (13) is zero when one (or both) factors are zero. Therefore, if there is no scaling, then H is zero (Fig. 6). However, when the variance of the values of π is zero, i.e. when all of them are equal, then H is also zero, independent of the scaling factors (Fig. 7).

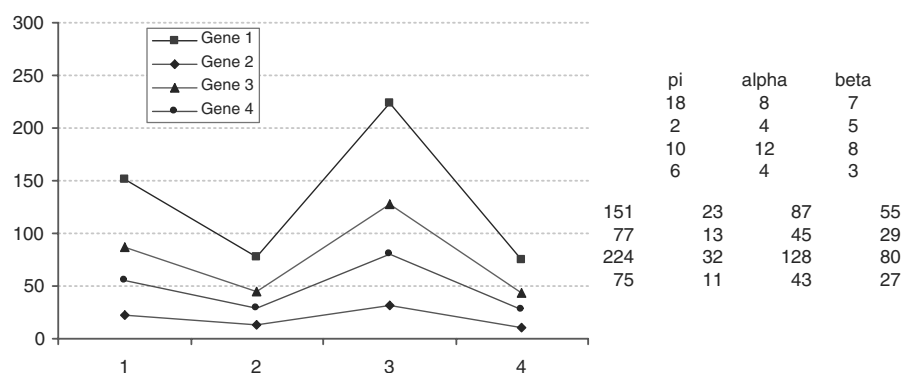


Fig. 4. A perfect bicluster with shifting and scaling patterns. The variance $\sigma_{\pi}^2=35$, and the variance $\sigma_{\alpha}^2=11$ ($H = 385$).

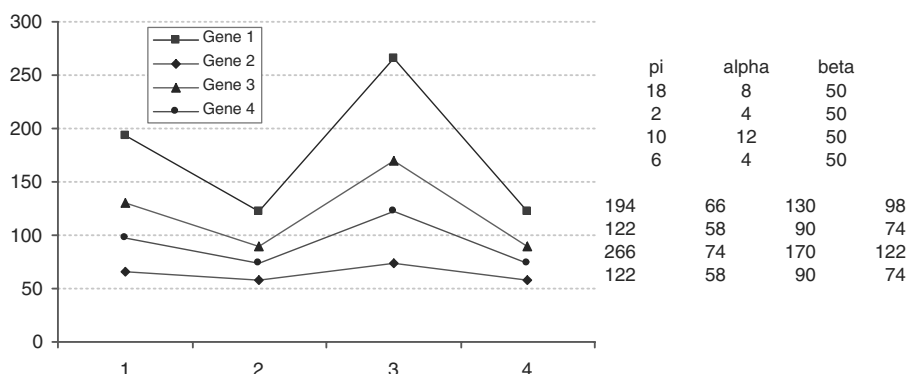


Fig. 5. A perfect bicluster, obtained from the submatrix in Figure 4, maintaining the shifting constant, $\beta = 50$, for all the genes ($H = 385$).

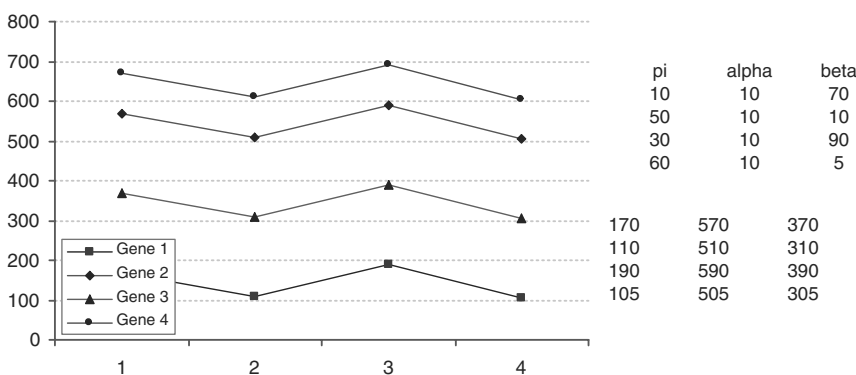


Fig. 6. A perfect bicluster with global scaling, i.e. all the α values are equal ($\alpha = 10$), so $H = 0$. This case is equivalent to a shifting pattern.

EXAMPLE 3. Figure 7 depicts the effect of null variance of the π values. Obviously, the mean squared residue is zero, because all the π values are equal, in spite of the non-null variance of α .

It can be inferred from Figures 6 and 7 that the mean squared residue is not useful to detect scaling patterns, since $\sigma_{\alpha}^2=0$ represents a perfect shifting pattern, and $\sigma_{\pi}^2=0$ represents a special case of shifting pattern (a trivial case in which all the genes have the same expression value).

The result presented in Theorem 2 has enormous implications on the search for patterns. As it was mentioned in Subsection 3.3,

typical algorithms look for δ -biclusters, i.e. biclusters whose mean squared residue is lower than δ . Obviously, this requires some prior knowledge about the correct choice of δ , which might vary depending on the dataset [the reader can find a brief discussion on this topic in Tavazoie *et al.* (1999)]. However, if the values of σ_{α}^2 or σ_{π}^2 are high, then we will miss those biclusters that provide a mean squared residue greater than δ .

It is very interesting to note that if we change $\alpha_3 = 10$ in Figure 4 (instead of $\alpha_3 = 12$), then the mean squared residue decreases from 385 to 236.25. This means that the Cheng and Church's algorithm would be able to find this new bicluster (using $\delta = 300$ as for the

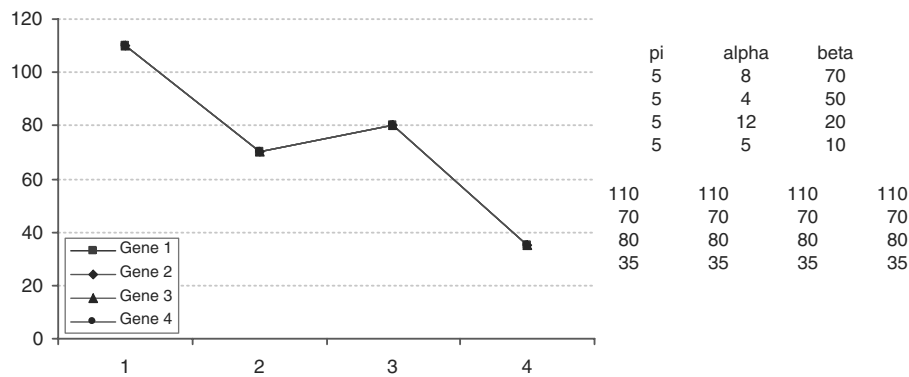


Fig. 7. A perfect bicluster without scaling or shifting (all the π values are equal, so $\sigma_{\pi}^2=0$).

yeast dataset), but it would not find the bicluster shown in Figure 4. This fact justifies the behavior of the Cheng and Church's algorithm, which includes a large number of genes per bicluster, since this way the variance σ_{α}^2 is decreased, and therefore, the value of the mean squared residue H is decreased as well.

The most immediate consequence of Theorem 2 is that if we use the mean squared residue to search for biclusters, and we set the value of δ as a threshold (upper bound), then we might miss many high-quality biclusters. Therefore, we can state that the mean squared residue is not an accurate measure to search for biclusters and therefore to elucidate functional similarities of a group of genes.

6 CONCLUSIONS

Cheng and Church (2000) authors stated (without proof) that a shifting (addition by a constant) to the matrix would not affect the mean squared residue, and a scaling (multiplication by a constant) might affect the score. These states have been proven in this paper, although not only for the global types mentioned above, but also for the local shiftings and scalings. However, the implications of the fact that scaling patterns affect the score are very important in order to discover useful patterns from gene expression data.

This paper also shows that the mean squared residue depends on the scaling variance, which we do not know a priori, and not on the shifting variance. This fact is critical, since a high scaling variance would lead us to miss those biclusters that provide a mean squared residue greater than δ . Specifically, interesting biclusters comprising few genes are likely to show high scaling variance, so they will go unnoticed.

These findings will help to develop new approaches, specially in order to discover scaling patterns, which awakes the interest of biologists on gene regulation and pathways.

From our point of view, the characterization of an objective function H that allows to detect simultaneously both shifting and scaling patterns would be very beneficial.

ACKNOWLEDGEMENTS

The author thanks Dr Dan Simovici (UMass, Boston) for very interesting discussions about biclustering, and reviewers for useful suggestions. The research was supported by the Spanish Research

Agency under the grant TIN2004—00159 and Junta de Andalucia (III Research Program).

Conflict of Interest: none declared.

REFERENCES

- Ben-Dor, A. *et al.* (1999) Clustering gene expression patterns. *J. Comput. Biol.*, **6**, 281–297.
- Bryan, K., Cunningham, P. and Bolshakova, N. (2005) Biclustering of expression data using simulated annealing. In *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems*, (CBMS 2005) Dublin, Ireland, 23–24 June, 2005, pp. 383–388.
- Cheng, Y. and Church, G. M. (2000) Biclustering of expression data. In *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*, (ISMB 2000) La Jolla, CA, 20–23 August, 2000, pp. 93–103.
- Cho, H., Dhillon, I. S., Guan, Y. and Sra, S. (2004) Minimum Sum-Squared Residue Co-clustering of Gene Expression Data. In *Proceedings of the 4th SIAM International Conference on Data Mining*, (SIAM 2004) Lake Buena Vista, Florida, TX, 22–24 April, 2004, ISBN 0-8971-568-7.
- Getz, G. *et al.* (2000) Coupled two-way clustering analysis of gene microarray data. *Proc. Natl Acad. Sci. USA*, pp. 12079–12084.
- Hartigan, J. A. (1972) Direct clustering of a data matrix. *J. Am. Stat. Assoc.*, **67**, 123–129.
- Lazzeroni, L. and Owen, A. (2002) Plaid models for gene expression data. *Statistica Sinica*, **12**, 61–86.
- Mirkin, B. (1996) *Mathematical Classification and Clustering*. Kluwer.
- Morgan, J. and Sonquist, J. (1963) Problems in the analysis of survey data, and a proposal. *JASA*, **58**, 415–434.
- Murali, T. M. and Kasif, S. (2003) Extracting conserved gene expression motifs from gene expression data. *Pac. Symp. Biocomput.*, 77–88.
- Tanay, A. *et al.* (2002) Discovering statistically significant biclusters in gene expression data. *Bioinformatics*, **19** (Suppl 2), 196–205.
- Tavazoie, S. *et al.* (1999) Systematic determination of genetic network architecture. *Proc. Natl Acad. Sci. USA*, **22**, 281–285.
- Wang, H., Wang, W., Yang, J. and Yu, P. S. (2002) Clustering by pattern similarity in large data sets. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Madison, WI, ACM Press, ISBN 1-58113-497-5, pp. 394–405.
- Yang, J., Wang, W., Wang, H. and Yu, P. S. (2002) δ -clusters: capturing subspace correlation in a large data set. In *Proceedings of the 18th IEEE Conference on Data Engineering (ICDE 2002)* San Jose, CA, IEEE Computer Society Press, ISBN 0-7695-1531-2, pp. 517–528.
- Yang, J., Wang, W., Wang, H. and Yu, P. S. (2003) Enhanced biclustering on expression data. In *Proceedings of the 3rd IEEE Conference on Bioinformatics and Bioengineering (BIBE 2003)*, pp. 321–327.
- Yip, K. Y., Cheung, D. W., Ng, M. K. and Cheung, K.-H. (2004) Identifying projected clusters from gene expression profiles. In *Proceedings of the 3rd IEEE Conference on Bioinformatics and Bioengineering (BIBE 2003)*, pp. 259–266.