Journal of Multivariate Analysis 101 (2010) 434–446

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Singular value decomposition of large random matrices (for two-way classification of microarrays)[☆]

Marianna Bolla[a,*], Katalin Friedl[b], András Krámli[c]

[a] Institute of Mathematics, Budapest University of Technology and Economics, Hungary

[b] Department of Computer Science, Budapest University of Technology and Economics, Hungary

[c] Bolyai Institute, University of Szeged, Hungary

ARTICLE INFO

Article history:
Available online 18 September 2009

AMS 1991 subject classifications:
15A42
15A52
62H30

Keywords:
Blown up matrix
Noise matrix
Random perturbation
Two-way classification
Microarray
Correspondence matrix

ABSTRACT

Asymptotic behavior of the singular value decomposition (SVD) of blown up matrices and normalized blown up contingency tables exposed to random noise is investigated. It is proved that such an $m \times n$ random matrix almost surely has a constant number of large singular values (of order $\sqrt{mn}$), while the rest of the singular values are of order $\sqrt{m+n}$ as $m, n \to \infty$. We prove almost sure properties for the corresponding isotropic subspaces and for noisy correspondence matrices. An algorithm, applicable to two-way classification of microarrays, is also given that finds the underlying block structure.

1. Introduction

A common problem in modern multivariate statistics is that of finding linear structures in large real-world data sets like internet or microarray measurements. In Bolla [1], large symmetric blown up matrices burdened with a so-called symmetric Wigner noise were investigated. It was proved that such an $n \times n$ matrix has some outstanding eigenvalues of order n, while the majority of the eigenvalues are at most of order $\sqrt{n}$, with probability tending to 1 as $n \to \infty$. Our goal is to generalize these results for the stability of SVD of large rectangular random matrices and to apply them to the contingency table matrix of categorical variables in order to perform two-way clustering of these variables.

Throughout this paper so-called blown up structures burdened with a general kind of noise are investigated. An $m \times n$ real matrix is a blown up matrix if after rearranging its rows and columns, it consists of blocks of the same entries. Such schemes are sought for in microarray analysis, and they are called chessboard patterns; cf. Kluger et al. [2]. The problem is that, in practical applications, with m and n large, we need sophisticated algorithms to find the convenient permutation of the rows and columns producing the chessboard, not to mention the noise added to our observations. There may be measurement errors that are very typical in microarray data. Therefore, we suppose that there is noise added to the entries independently.

The noise matrix has independent entries of zero expectation whose distributions are usually not identical. The entries are uniformly bounded, though this condition can be relaxed. In the proofs we use the well-known fact that the spectral norm of

[☆] Research supported in part by the Hungarian National Research Fund and by the National Office for Research and Technology (Grant No. OTKA 76481, OTKA-NKTH 67651 and 77778).

* Corresponding author.
E-mail addresses: marib@math.bme.hu (M. Bolla), friedl@cs.bme.hu (K. Friedl), kramli@informatika.ilab.sztaki.hu (A. Krámli).

0047-259X/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2009.09.006

an $m \times n$ noise matrix is of order $\sqrt{m+n}$ with high probability if m and n are large. If the entries of all matrices are defined on thesame probability space (this is often the case when we take larger and larger samples from the same population), a large deviation theorem of Alon et al. [3] — formulated for the eigenvalues of symmetric matrices — is applicable to get an almost sure statement for the largest singular value of the noise. A stronger result is proved by Litvak et al. [4] for the whole singular spectrum of rectangular matrices with special sub-Gaussian entries under the condition that $m \geq n$ with $n \to \infty$ in such a way that m is proportional to n.

Our almost sure statements are formulated in terms of $m, n \to \infty$ independently of each other under some growth conditions imposed on the cluster sizes, while the number of clusters remains constant under the blow up. First Füredi and Komlós [5] viewed a non-centered symmetric Wigner-type matrix not merely a noise by introducing a shift that puts off the edge of the spectrum. We generalize this setting to rectangular matrices with different shifts, and prove that the noisy matrix will almost surely have outstanding singular values whose number is equal to the rank k of the blown up matrix. In addition we prove that the sum of the inner variances of the clusters, formed by means of the corresponding singular vector pairs, is of order $\frac{m+n}{mn}$, which tends to 0 almost surely as $m, n \to \infty$ under the growth conditions. This fact can be interpreted as the good classification property of such a structure, irrespective of the noise. Thus, the coordinates of the singular vector pairs belonging to the leading k singular values will be used to form k-dimensional representatives of the rows and columns, and on this basis, we can find the row- and column-clusters, as well as the row- and column-memberships by using the k-means algorithm. On the other hand, in the presence of some outstanding singular values, an explicit construction is also given for finding a blown up structure in a noisy matrix.

This is the underlying linear structure only if the error matrix is comparable with the noise matrix in spectral norm. Otherwise, there may be a nonlinear structure or chaos behind our data. These questions have little significance in the classical multivariate analysis, as with m and n small, the whole singular spectrum, together with the singular vectors, is at our disposal and statistical hypotheses can be tested. In contrast, in modern data mining we deal with enormous data sets. On the basis of some leading singular values and corresponding singular vector pairs, merely approximate inferences can be made due to the computational limitations and random errors.

In most applications the underlying matrix is a contingency table with non-negative integer entries that contain, for example, counts for two categorical variables with m and n different categories. As the categories may be measured in different units, a normalization is necessary. This normalization is achieved by dividing the entries by the square roots of the corresponding row- and column-sums. In Kluger et al. [2], the authors use a similar transformation and find the clusters by means of just one eigenvector or singular vector pair belonging usually to an outstanding singular value. We show that the above transformation is identical to that of the correspondence analysis (cf. Greenacre [6]), and prove that there is a remarkable gap between the k largest and the other singular values of the noisy correspondence matrix. This implies good two-way classification properties of the row- and column-categories.

In microarray measurements the rows correspond to different genes, the columns correspond to different conditions or samples, and the entries are the expression levels of a specific gene under a specific condition or in a specific sample. Multivariate analysis and spectral decomposition of some similarity matrix derived from the microarray are frequently used in the steadily developing literature on microarrays. But even if the SVD is applied, usually only one singular vector pair is used for classification purposes, e.g., Higham et al. [7] and Liu et al. [8]. In Omberg et al. [9], the authors use a more sophisticated algorithm for three-dimensional fixed (usually small) size arrays that result in traditional microarrays in their special two-dimensional marginals. They introduce a higher order SVD but neither using noise and preprocessing of the data nor investigating asymptotics is included.

The organization of the paper is as follows. In Section 2, we introduce the precise notions of a noise matrix, and a blown up matrix; further the frequently used convergence facts and conditions are formulated together with a summary of the relevant literature. In Section 3, it is proved that the $m \times n$ noisy matrix almost surely has k outstanding singular values of order $\sqrt{mn}$. In Section 4, the distances of the corresponding isotropic subspaces are estimated, and this gives rise to a two-way classification of the rows and columns of the noisy matrix with sum of inner variances $\mathcal{O}(\frac{m+n}{mn})$, almost surely. In Section 5, noisy correspondence matrices are investigated, while in Section 6 an explicit construction is given showing how a blown up structure behind a real-life matrix with a few outstanding singular values and "well classifiable" corresponding singular vector pairs can be found.

2. Preliminaries

First we introduce some notation and facts needed in the sequel.

Definition 1. The $m \times n$ real matrix $\mathbf{B}$ is a blown up matrix if there is an $a \times b$ so-called *pattern matrix* $\mathbf{P}$ with entries $0 \leq p_{ij} \leq 1$, and there are positive integers $m_1, \ldots, m_a$ with $\sum_{i=1}^{a} m_i = m$ and $n_1, \ldots, n_b$ with $\sum_{i=1}^{b} n_i = n$, such that the matrix $\mathbf{B}$, after rearranging its rows and columns, can be divided into $a \times b$ blocks, where block (i, j) is an $m_i \times n_j$ matrix with entries equal to p_{ij} $(1 \leq i \leq a, 1 \leq j \leq b)$.

Definition 2. Let w_{ij} $(1 \leq i \leq m, 1 \leq j \leq n)$ be independent random variables defined on the same probability space. $\mathbb{E}(w_{ij}) = 0$ $(\forall i, j)$ and the w_{ij}'s are uniformly bounded (i.e., there is a constant $K > 0$, independently of m and n, such that $|w_{ij}| \leq K, \forall i, j$). The $m \times n$ real matrix $\mathbf{W} = (w_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ is called a noise matrix.

Sometimes we use the notation $\mathbf{W}_{m\times n}$ to emphasize that the sequence of noise matrices is expanding. In fact, our results are valid under relaxed conditions with a noise matrix of independent, sub-Gaussian entries and anypattern matrix of fixed sizes and non-negative entries. The restriction that the entries of $\mathbf{P}$ take on values in the [0,1] interval is used merely in a construction of Section 4, where p_{ij}'s are regarded as probabilities.

Our model is the following. Let us fix the matrix $\mathbf{P}$, blow it up to obtain matrix $\mathbf{B}$, and let $\mathbf{A} = \mathbf{B} + \mathbf{W}$, where $\mathbf{W}$ is a noise matrix of appropriate size. We are interested in the properties of $\mathbf{A}$ when $m_1, \dots, m_a \to \infty$ and $n_1, \dots, n_b \to \infty$, roughly speaking, at the same rate. More precisely, we impose two different constraints on the growth of the sizes m and n, and the growth rate of their components. The first one is required for all our reasonings, while the second one will be used in the case of noisy correspondence matrices, only.

Definition 3. ***GC1*** (Growth Condition 1). There exists a constant $0 < c < 1$ such that $m_i/m \ge c\ (i = 1, \dots, a)$ and there exists a constant $0 < d < 1$ such that $n_i/n \ge d\ (i = 1, \dots, b)$.
GC2 (Growth Condition 2). There exist constants $C \ge 1, D \ge 1$, and $C_0 > 0, D_0 > 0$ such that $m \le C_0 \cdot n^C$ and $n \le D_0 \cdot m^D$ hold for sufficiently large m and n.

Remark 4. *GC1* implies that

$$c \le \frac{m_k}{m_i} \le \frac{1}{c} \quad \text{and} \quad d \le \frac{n_\ell}{n_j} \le \frac{1}{d} \tag{1}$$

hold for any pair of indices $k, i \in \{1, \dots, a\}$ and $\ell, j \in \{1, \dots, b\}$.

We want to establish some property $\mathcal{P}_{m,n}$ that holds for the $m \times n$ random matrix $\mathbf{A} = \mathbf{B} + \mathbf{W}$ (for short, $\mathbf{A}_{m\times n}$) with m and n large enough. In this paper $\mathcal{P}_{m,n}$ is mostly related to the SVD of $\mathbf{A}_{m\times n}$.

Usually convergence in probability, that is

$$\lim_{m,n\to\infty} \mathbb{P}\left(\mathbf{A}_{m\times n} \text{ has } \mathcal{P}_{m,n}\right) = 1,$$

is considered. If in addition $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} p_{mn} < \infty$ also holds, where $p_{mn} = \mathbb{P}\left(\mathbf{A}_{m\times n} \text{ does not have } \mathcal{P}_{m,n}\right)$, then — by the Borel–Cantelli Lemma — the property $\mathcal{P}_{m,n}$ holds almost surely for $\mathbf{A}_{m\times n}$. This means that if the independent entries of the expanding sequence of noise and noisy matrices are defined on the same probability space, then

$$\mathbb{P}\left(\exists\, m_0, n_0 \in \mathbb{N} \text{ such that for } m \ge m_0 \text{ and } n \ge n_0 \mathbf{A}_{m\times n} \text{ has } \mathcal{P}_{m,n}\right) = 1.$$

Here we may assume *GC1* or *GC2* for the growth of m and n, while the uniform bound K, as defined in Definition 2, is kept fixed.

As an easy consequence of Füredi and Komlós [5] for rectangular matrices, the spectral norm (i.e., the largest singular value) of an $m \times n$ noise matrix is of order $\sqrt{m+n}$ in probability. Then a concentration theorem of Alon at al. [3], using Talagrand's technique [10], is applied for the $(m+n)\times(m+n)$ symmetric matrix

$$\tilde{\mathbf{W}} = \frac{1}{K}\begin{pmatrix} \mathbf{0} & W \\ W^{\mathrm{T}} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{W}$ is the $m \times n$ noise matrix and K is the uniform bound for the entries. Hence,

$$\mathbb{P}\left(|s_1(\mathbf{W}) - \mathbb{E}(s_1(\mathbf{W}))| > t\right) \le \exp\left(-\frac{(1-o(1))t^2}{32K^2}\right). \tag{2}$$

The fact that $s_1(\mathbf{W}) = \|\mathbf{W}\| = \mathcal{O}(\sqrt{m+n})$ in probability and inequality (2) together ensure that $\mathbb{E}(\|\mathbf{W}\|) = \mathcal{O}(\sqrt{m+n})$. Hence, no matter how $\mathbb{E}(\|\mathbf{W}\|)$ behaves when $m \to \infty$ and $n \to \infty$, the following rough estimate is valid: there exist positive constants c_1 and c_2 which depend on the uniform bound K for the entries of $\mathbf{W}$, such that

$$\mathbb{P}\left(\|\mathbf{W}\| > c_1\sqrt{m+n}\right) \le \exp(-c_2(m+n)). \tag{3}$$

Remark 5. The exponential decay of the right-hand side of (3) implies, by the Borel–Cantelli Lemma, that the spectral norm of a noise matrix $\mathbf{W}_{m\times n}$ is of order $\sqrt{m+n}$, almost surely. This observation suffices for us to get almost sure results in Sections 3 and 4.

We note that the above result remains valid if the noise matrix consists of symmetrically distributed entries with some special moment conditions, while $m, n \to \infty$ $(m \ge n)$ with m/n kept near constant; see Litvak et al. [4]. We remark that the latter growth condition is a special case of *GC2* with $C = D = 1$ and C_0D_0 near 1. The authors prove that under the above conditions not only the largest, but also the smallest singular value, and, hence, also the whole singular spectrum are of order $\sqrt{m}$ with high probability. Therefore, the random map embodied by the noise matrix becomes more and more an "isomorphism" onto its image, if its sizes grow proportionally. The above conditions also imply the sharper estimate $\|\mathbf{W}\| = \Theta(\sqrt{m+n})$ instead of $\|\mathbf{W}\| = \mathcal{O}(\sqrt{m+n})$. Analogous results for different norms of random rectangular matrices with complex, uniformly bounded entries were obtained by Meckes et al. [11] using Talagrand's technique. However, we included Remark 5 since it fits best into our framework.

For square symmetric matrices, since Wigner's famous semicircle law [12] is applicable to the mass spectrum, finer results for the individual eigenvalues have been obtained in a steadily increasing number in the last 30 years. Hence, the Tracy–Widom distribution of [13] came into existence for the distribution of the first few largest eigenvalues of Gaussian orthogonal ensembles, while Soshnikov [14] proved the universality of random Wigner matrices in the sense that the limiting distribution of their largest eigenvalues becomes independent of the entries' distribution as their size tends to infinity. Füredi and Komlós [5] perturbed the Wigner matrix with independent and uniformly bounded entries using a fixed constant matrix and obtained that the limit distribution of the largest eigenvalue is normal. Recently, Féral and Péché [15] determined the limiting behavior of the largest eigenvalue of a rank 1 deformation of a Wigner matrix with more general entries normalized in such a way that the perturbation is comparable to the perturbed matrix in spectral norm. In this framework, we use rectangular matrices and add more pronounced shifts that may differ from block to block. Other generalizations for sample covariance (i.e., Wishart-type) matrices (with dependent entries) can be found in the statistics literature, e.g., Johnstone [16], and Baik and Silverstein [17]. The only common thing that this approach shares with our setting is that with $\mathbf{W}$ having independent Gaussian or sub-Gaussian entries, $\mathbf{W}\mathbf{W}^{\mathrm{T}}$ is a Wishart-type matrix with eigenvalues being the squared singular values of $\mathbf{W}$. On this basis, inferences for the singular values of a rectangular noise matrix can be made; cf. Litvak et al. [4] and Olkin [18].

3. Singular values of a noisy matrix

We now prove the main asymptotic results for the singular values of a noisy matrix.

Proposition 6. *If GC1 holds, then all the non-zero singular values of the $m \times n$ blown up matrix $\mathbf{B}$ are of order $\sqrt{mn}$.*

Proof. As there are at most a and b linearly independent rows and linearly independent columns in $\mathbf{B}$, respectively, the rank r of the matrix $\mathbf{B}$ cannot exceed $\min\{a, b\}$. Let $s_1 \ge s_2 \ge \cdots \ge s_r > 0$ be the positive singular values of $\mathbf{B}$. Let $\mathbf{v}_k \in \mathbb{R}^m$, $\mathbf{u}_k \in \mathbb{R}^n$ be a singular vector pair corresponding to s_k, $k = 1, \dots, r$. Without loss of generality, $\mathbf{v}_1, \dots, \mathbf{v}_r$ and $\mathbf{u}_1, \dots, \mathbf{u}_r$ can be unit-norm, pairwise orthogonal vectors in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively.

For the subsequent calculations we drop the subscript k, and $\mathbf{v}$, $\mathbf{u}$ denotes a singular vector pair corresponding to the singular value $s > 0$ of the blown up matrix $\mathbf{B}$, $\|\mathbf{v}\| = \|\mathbf{u}\| = 1$. It is easy to see that they have piecewise constant structures: $\mathbf{v}$ has m_i coordinates equal to $v(i)$ $(i = 1, \dots, a)$ and $\mathbf{u}$ has n_j coordinates equal to $u(j)$ $(j = 1, \dots, b)$. Then, with these coordinates the singular value–singular vector equation

$$\mathbf{B}\mathbf{u} = s \cdot \mathbf{v} \tag{4}$$

has the form

$$\sum_{j=1}^{b} n_j p_{ij} u(j) = s \cdot v(i) \quad (i = 1, \dots, a). \tag{5}$$

With the notation

$$\tilde{\mathbf{u}} = (u(1), \dots, u(a))^{\mathrm{T}}, \qquad \tilde{\mathbf{v}} = (v(1), \dots, v(b))^{\mathrm{T}},$$
$$\mathbf{D}_m = \mathrm{diag}\,(m_1, \dots, m_a), \qquad \mathbf{D}_n = \mathrm{diag}\,(n_1, \dots, n_b)$$

the equations in (5) can be written as

$$\mathbf{P}\mathbf{D}_n\tilde{\mathbf{u}} = s \cdot \tilde{\mathbf{v}}.$$

Introducing the following transformations of $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$:

$$\mathbf{w} = \mathbf{D}_n^{1/2}\tilde{\mathbf{u}}, \qquad \mathbf{z} = \mathbf{D}_m^{1/2}\tilde{\mathbf{v}}, \tag{6}$$

the equation is equivalent to

$$\mathbf{D}_m^{1/2}\mathbf{P}\mathbf{D}_n^{1/2}\mathbf{w} = s \cdot \mathbf{z}. \tag{7}$$

Applying the transformation (6) for the $\tilde{\mathbf{u}}_k$, $\tilde{\mathbf{v}}_k$ pairs obtained from the $\mathbf{u}_k$, $\mathbf{v}_k$ pairs $(k = 1, \dots, r)$, orthonormal systems in $\mathbb{R}^a$ and $\mathbb{R}^b$ are obtained:

$$\mathbf{w}_k{}^{\mathrm{T}} \cdot \mathbf{w}_\ell = \sum_{j=1}^{b} n_j u_k(j) u_\ell(j) = \delta_{k\ell} \quad \text{and} \quad \mathbf{z}_k{}^{\mathrm{T}} \cdot \mathbf{z}_\ell = \sum_{i=1}^{a} m_i v_k(i) v_\ell(i) = \delta_{k\ell}.$$

Consequently, $\mathbf{z}_k$, $\mathbf{w}_k$ is a singular vector pair corresponding to singular value s_k of the $a \times b$ matrix $\mathbf{D}_m^{1/2}\mathbf{P}\mathbf{D}_n^{1/2}$ $(k = 1, \dots, r)$. With the shrinking

$$\tilde{\mathbf{D}}_m = \frac{1}{m}\mathbf{D}_m, \qquad \tilde{\mathbf{D}}_n = \frac{1}{n}\mathbf{D}_n$$

an equivalent form of (7) is

$$\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{w} = \frac{s}{\sqrt{mn}} \cdot \mathbf{z},$$

that is the $a \times b$ matrix $\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}$ has non-zero singular values $\frac{s_k}{\sqrt{mn}}$ with the same singular vector pairs $\mathbf{z}_k, \mathbf{w}_k$ $(k = 1, \dots, r)$. If the s_k's are not distinct numbers, the singular vector pairs corresponding to a multiple singular value are not unique, but still they can be obtained from the SVD of the shrunken matrix $\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}$.

Now we want to establish relations between the singular values of $\mathbf{P}$ and $\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}$. Let $s_k(\mathbf{Q})$ denote the kth-largest singular value of a matrix $\mathbf{Q}$. By the Courant–Fischer–Weyl minimax principle (cf. [19, p. 75])

$$s_k(\mathbf{Q}) = \max_{\dim H = k} \min_{\mathbf{x} \in H} \frac{\|\mathbf{Q}\mathbf{x}\|}{\|\mathbf{x}\|}.$$

Since we are interested only in the first r singular values, where $r = \text{rank}\,(\mathbf{B}) = \text{rank}\,(\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2})$, it is sufficient to consider vectors $\mathbf{x}$ for which $\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{x} \neq \mathbf{0}$. Therefore with $k \in \{1, \dots, r\}$ and an arbitrary k-dimensional subspace $H \subset \mathbb{R}^b$ one can write

$$\begin{aligned}\min_{\mathbf{x} \in H} \frac{\|\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|}{\|\mathbf{x}\|} &= \min_{\mathbf{x} \in H} \frac{\|\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|}{\|\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|} \cdot \frac{\|\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|}{\|\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|} \cdot \frac{\|\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|}{\|\mathbf{x}\|} \\ &\geq s_a(\tilde{\mathbf{D}}_m^{1/2}) \cdot \min_{\mathbf{x} \in H} \frac{\|\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|}{\|\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|} \cdot s_b(\tilde{\mathbf{D}}_n^{1/2}) \geq \sqrt{cd} \cdot \min_{\mathbf{x} \in H} \frac{\|\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|}{\|\tilde{\mathbf{D}}_n^{1/2}\mathbf{x}\|},\end{aligned}$$

with c, d of *GC1*. Now taking the maximum for all possible k-dimensional subspaces H we obtain that $s_k(\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}) \geq \sqrt{cd} \cdot s_k(\mathbf{P}) > 0$. On the other hand,

$$s_k(\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}) \leq \|\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}\| \leq \|\tilde{\mathbf{D}}_m^{1/2}\| \cdot \|\mathbf{P}\| \cdot \|\tilde{\mathbf{D}}_n^{1/2}\| \leq \|\mathbf{P}\| \leq \sqrt{ab}.$$

These inequalities imply that $s_k(\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2})$ is a non-zero constant, and because of $s_k(\tilde{\mathbf{D}}_m^{1/2}\mathbf{P}\tilde{\mathbf{D}}_n^{1/2}) = \frac{s_k}{\sqrt{mn}}$ we obtain that $s_1, \dots, s_r = \Theta(\sqrt{mn})$. □

Theorem 7. *Let $\mathbf{A} = \mathbf{B} + \mathbf{W}$ be an $m \times n$ random matrix, where $\mathbf{B}$ is a blown up matrix with positive singular values $s_1, \dots, s_r$ and $\mathbf{W}$ is a noise matrix. Then, under GC1, the matrix $\mathbf{A}$ almost surely has r singular values $z_1, \dots, z_r$, such that*

$$|z_i - s_i| = \mathcal{O}(\sqrt{m+n}), \quad i = 1, \dots, r$$

and for the other singular values almost surely

$$z_j = \mathcal{O}(\sqrt{m+n}), \quad j = r+1, \dots, \min\{m, n\}.$$

Proof. The statement follows from the analog of Weyl's perturbation theorem for singular values of rectangular matrices (see [19, p. 99]) and from Remark 5. If $s_i(\mathbf{A})$ and $s_i(\mathbf{B})$ denote the ith-largest singular values of the matrix in the argument then for the difference of the corresponding pairs

$$|s_i(\mathbf{A}) - s_i(\mathbf{B})| \leq \max_i s_i(\mathbf{W}) = \|\mathbf{W}\|, \quad i = 1, \dots, \min\{m, n\}.$$

By Remark 5, $\|\mathbf{W}\|$ is of order $\sqrt{m+n}$ almost surely; that finishes the proof. □

Corollary 8. *With the notation*

$$\varepsilon := \|\mathbf{W}\| = \mathcal{O}(\sqrt{m+n}) \quad \textit{and} \quad \Delta := \min_{1 \leq i \leq r} s_i(\mathbf{B}) = \min_{1 \leq i \leq r} s_i = \Theta(\sqrt{mn}) \tag{8}$$

there is a spectral gap of size $\Delta - 2\varepsilon$ between the r largest and the other singular values of the perturbed matrix $\mathbf{A}$, and this gap is significantly larger than ε.

Under the growth conditions of Litvak et al. [4], $\varepsilon = \Theta(\sqrt{m+n})$ is valid, indicating that the error cannot be decreased.

4. Classification via singular vector pairs

In this section perturbation results for the singular vector pairs corresponding to the leading singular values are established. To this end, with the help of Theorem 7, we estimate the distances between the corresponding right- and left-hand side eigenspaces (isotropic subspaces) of the matrices $\mathbf{B}$ and $\mathbf{A} = \mathbf{B} + \mathbf{W}$. Let $\mathbf{v}_1, \dots, \mathbf{v}_m \in \mathbb{R}^m$ and $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{R}^n$ be orthonormal left- and right-hand side singular vectors of **B**,

$$\mathbf{B}\mathbf{u}_i = s_i \cdot \mathbf{v}_i \quad (i = 1, \dots, r) \quad \text{and} \quad \mathbf{B}\mathbf{u}_j = 0 \quad (j = r+1, \dots, n).$$

Let us also denote the unit-norm, pairwise orthogonal left- and right-hand side singular vectors corresponding to the r outstanding singular values $z_1, \dots, z_r$ of **A** by $\mathbf{y}_1, \dots, \mathbf{y}_r \in \mathbb{R}^m$ and $\mathbf{x}_1, \dots, \mathbf{x}_r \in \mathbb{R}^n$, respectively. Then $\mathbf{A}\mathbf{x}_i = z_i \cdot \mathbf{y}_i$ ($i = 1, \dots, r$). Let

$$F := \text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_r\} \quad \text{and} \quad G := \text{Span}\,\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$$

denote the spanned linear subspaces in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively; further, let $\text{dist}(\mathbf{y}, F)$ denote the Euclidean distance between the vector $\mathbf{y}$ and the subspace F.

Proposition 9. *With the above notation, under GC1, the following estimate holds almost surely:*

$$\sum_{i=1}^{r} \text{dist}^2(\mathbf{y}_i, F) \le r\frac{\varepsilon^2}{(\Delta - \varepsilon)^2} = \mathcal{O}\left(\frac{m+n}{mn}\right) \tag{9}$$

and analogously,

$$\sum_{i=1}^{r} \text{dist}^2(\mathbf{x}_i, G) \le r\frac{\varepsilon^2}{(\Delta - \varepsilon)^2} = \mathcal{O}\left(\frac{m+n}{mn}\right). \tag{10}$$

Proof. Let us choose one of the right-hand side singular vectors $\mathbf{x}_1, \dots, \mathbf{x}_r$ of $\mathbf{A} = \mathbf{B} + \mathbf{W}$ and denote it simply by $\mathbf{x}$ with corresponding singular value z. We shall estimate the distance between $\mathbf{x}$ and G, and similarly that between $\mathbf{y} = \mathbf{A}\mathbf{x}/z$ and F. For this purpose we expand $\mathbf{x}$ and $\mathbf{y}$ in the orthonormal bases $\mathbf{u}_1, \dots, \mathbf{u}_n$ and $\mathbf{v}_1, \dots, \mathbf{v}_m$, respectively:

$$\mathbf{x} = \sum_{i=1}^{n} t_i \mathbf{u}_i \quad \text{and} \quad \mathbf{y} = \sum_{i=1}^{m} l_i \mathbf{v}_i.$$

Then

$$\mathbf{A}\mathbf{x} = (\mathbf{B} + \mathbf{W})\mathbf{x} = \sum_{i=1}^{r} t_i s_i \mathbf{v}_i + \mathbf{W}\mathbf{x}, \tag{11}$$

and, on the other hand,

$$\mathbf{A}\mathbf{x} = z\mathbf{y} = \sum_{i=1}^{m} z l_i \mathbf{v}_i. \tag{12}$$

Equating the right-hand sides of (11) and (12) we obtain

$$\sum_{i=1}^{r} (z l_i - t_i s_i)\mathbf{v}_i + \sum_{i=r+1}^{m} z l_i \mathbf{v}_i = \mathbf{W}\mathbf{x}.$$

Applying the Pythagorean Theorem,

$$\sum_{i=1}^{r} (z l_i - t_i s_i)^2 + z^2 \sum_{i=r+1}^{m} l_i^2 = \|\mathbf{W}\mathbf{x}\|^2 \le \varepsilon^2, \tag{13}$$

because $\|\mathbf{x}\| = 1$ and $\|\mathbf{W}\| = \varepsilon$.

As $z \ge \Delta - \varepsilon$ holds almost surely by Theorem 7,

$$\text{dist}^2(\mathbf{y}, F) = \sum_{i=r+1}^{m} l_i^2 \le \frac{\varepsilon^2}{z^2} \le \frac{\varepsilon^2}{(\Delta - \varepsilon)^2}.$$

The order of the above estimate follows from the order of ε and Δ of (8):

$$\text{dist}^2(\mathbf{y}, F) = \mathcal{O}\left(\frac{m+n}{mn}\right) \tag{14}$$

almost surely. Applying (14) for the left-hand side singular vectors $\mathbf{y}_1, \dots, \mathbf{y}_r$, by the definition of an almost sure property,

$$\mathbb{P}\left\{\exists m_{0i}, n_{0i} \in \mathbb{N} \text{ such that for } m \ge m_{0i} \text{ and } n \ge n_{0i}\text{: } \text{dist}^2(\mathbf{y}_i, F) \le \varepsilon^2/(\Delta - \varepsilon)^2\right\} = 1$$

for $i = 1, \dots, r$. Hence,

$$\mathbb{P}\left\{\exists m_0, n_0 \in \mathbb{N} \text{ such that for } m \ge m_0 \text{ and } n \ge n_0\text{: } \text{dist}^2(\mathbf{y}_i, F) \le \varepsilon^2/(\Delta - \varepsilon)^2, i = 1, \dots, r\right\} = 1,$$

and consequently,

$$\mathbb{P}\left\{\exists m_0, n_0 \in \mathbb{N} \text{ such that for } m \geq m_0 \text{ and } n \geq n_0\text{: } \sum_{i=1}^{r} \operatorname{dist}^2(\mathbf{y}_i, F) \leq r\varepsilon^2/(\Delta-\varepsilon)^2\right\} = 1$$

also holds, and this finishes the proof of the first statement.

The estimate for the squared distance between G and a right-hand side singular vector $\mathbf{x}$ of $\mathbf{A}$ follows in the same way starting with $\mathbf{A}^{\mathrm{T}}\mathbf{y} = z \cdot \mathbf{x}$ and using the fact that $\mathbf{A}^{\mathrm{T}}$ has the same singular values as $\mathbf{A}$. □

By Proposition 9, the individual distances between the original and the perturbed subspaces and also the sum of these distances tend to zero almost surely as $m, n \to \infty$.

Now let $\mathbf{A}$ be a microarray on m genes and n conditions, with a_{ij} denoting the expression level of gene i under condition j. We suppose that $\mathbf{A}$ is a noisy random matrix obtained by adding a noise matrix $\mathbf{W}$ to the blown up matrix $\mathbf{B}$. Let us denote by $A_1, \dots, A_a$ the partition of the genes and by $B_1, \dots, B_b$ the partition of the conditions with respect to the blow up (they can also be thought of as clusters of genes and conditions).

Proposition 9 also implies the well-clustering property of the representatives of the genes and conditions in the following representation. Let $\mathbf{Y}$ be the $m \times r$ matrix containing the left-hand side singular vectors $\mathbf{y}_1, \dots, \mathbf{y}_r$ of $\mathbf{A}$ in its columns. Similarly, let $\mathbf{X}$ be the $n \times r$ matrix containing the right-hand side singular vectors $\mathbf{x}_1, \dots, \mathbf{x}_r$ of $\mathbf{A}$ in its columns. The r-dimensional representatives of the genes are the row vectors of $\mathbf{Y}$: $\mathbf{y}^1, \dots, \mathbf{y}^m \in \mathbb{R}^r$, while the r-dimensional representatives of the conditions are the row vectors of $\mathbf{X}$: $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^r$. Let $S_a^2(\mathbf{Y})$ denote the a-variance, introduced in [20], of the genes' representatives

$$S_a^2(\mathbf{Y}) = \min_{\{A'_1,\dots,A'_a\}} \sum_{i=1}^{a} \sum_{j\in A'_i} \|\mathbf{y}^j - \bar{\mathbf{y}}^i\|^2, \quad \text{where } \bar{\mathbf{y}}^i = \frac{1}{m_i}\sum_{j\in A'_i} \mathbf{y}^j,$$

while $S_b^2(\mathbf{X})$ denotes the b-variance of the conditions' representatives

$$S_b^2(\mathbf{X}) = \min_{\{B'_1,\dots,B'_b\}} \sum_{i=1}^{b} \sum_{j\in B'_i} \|\mathbf{x}^j - \bar{\mathbf{x}}^i\|^2, \quad \text{where } \bar{\mathbf{x}}^i = \frac{1}{n_i}\sum_{j\in B'_i} \mathbf{x}^j,$$

the partitions $\{A'_1, \dots, A'_a\}$ and $\{B'_1, \dots, B'_b\}$ varying over all a- and b-partitions of the genes and conditions, respectively.

Theorem 10. *With the above notation, under GC1, for the a- and b-variances of the representation of the microarray* $\mathbf{A}$ *the relations*

$$S_a^2(\mathbf{Y}) = \mathcal{O}\left(\frac{m+n}{mn}\right) \quad \text{and} \quad S_b^2(\mathbf{X}) = \mathcal{O}\left(\frac{m+n}{mn}\right)$$

hold almost surely.

Proof. By the proof of Theorem 3 of [20] it can be easily seen that $S_a^2(\mathbf{Y}) \leq \sum_{i=1}^{a}\sum_{j\in A_i} \|\mathbf{y}^j - \bar{\mathbf{y}}^i\|^2$ and $S_b^2(\mathbf{X}) \leq \sum_{i=1}^{b}\sum_{j\in B_i} \|\mathbf{x}^j - \bar{\mathbf{x}}^i\|^2$, the right-hand sides being equal to the left-hand sides of (9) and (10), respectively; therefore they are also of order $\frac{m+n}{mn}$. □

Hence, the addition of any kind of a noise matrix to a rectangular matrix that has a blown up structure $\mathbf{B}$ will not change the order of the outstanding singular values, and the block structure of $\mathbf{B}$ can be reconstructed from the representatives of the row and column items of the noisy matrix $\mathbf{A}$.

With an appropriate noise matrix, we can achieve that the matrix $\mathbf{B}+\mathbf{W}$ in its (i, j)-th block contains 1's with probability p_{ij}, and 0's otherwise. That is, for $i = 1, \dots, a$, $j = 1, \dots, b$, $l \in A_i$, $k \in B_j$, let

$$w_{lk} := \begin{cases} 1 - p_{ij}, & \text{with probability } p_{ij} \\ -p_{ij} & \text{with probability } 1 - p_{ij} \end{cases} \tag{15}$$

be independent random variables. This $\mathbf{W}$ satisfies the conditions of Definition 2 with entries uniformly bounded by 1 in absolute value and of zero expectation. The noisy matrix $\mathbf{A}$ becomes a 0–1 matrix that can be regarded as the incidence matrix of a hypergraph on m vertices and n edges. (Vertices correspond to the genes and edges correspond to the conditions. The incidence relation depends on whether a specific gene is expressed or not under a specific condition.)

By the choice (15) of $\mathbf{W}$, vertices of the vertex set A_i appear in edges of the edge set B_j with probability p_{ij} (set i of genes equally influences set j of conditions, like the chessboard pattern of [2]). It is a generalization of the classical Erdős–Rényi model [21] for random hypergraphs and for several blocks. The question of how such a chessboard pattern behind a random (especially 0–1) matrix can be found under specific conditions is discussed in Section 6.

5. Perturbation results for correspondence matrices

Now the pattern matrix $\mathbf{P}$ contains arbitrary non-negative entries, and so does the blown up matrix $\mathbf{B}$. Let us suppose that there are no identically zero rows or columns. We perform the correspondence transformation described below on $\mathbf{B}$.

We are interested in the order of singular values of matrix $\mathbf{A} = \mathbf{B} + \mathbf{W}$ when the same correspondence transformation is applied to it. To this end, we introduce the following notation:

$$\mathbf{D}_{Brow} = \mathrm{diag}\,(d_{Brow\,1}, \ldots, d_{Brow\,m}) := \mathrm{diag}\left(\sum_{j=1}^{n} b_{1j}, \ldots, \sum_{j=1}^{n} b_{mj}\right)$$

$$\mathbf{D}_{Bcol} = \mathrm{diag}\,(d_{Bcol\,1}, \ldots, d_{Bcol\,n}) := \mathrm{diag}\left(\sum_{i=1}^{m} b_{i1}, \ldots, \sum_{i=1}^{m} b_{in}\right)$$

$$\mathbf{D}_{Arow} = \mathrm{diag}\,(d_{Arow\,1}, \ldots, d_{Arow\,m}) := \mathrm{diag}\left(\sum_{j=1}^{n} a_{1j}, \ldots, \sum_{j=1}^{n} a_{mj}\right)$$

$$\mathbf{D}_{Acol} = \mathrm{diag}\,(d_{Acol\,1}, \ldots, d_{Acol\,n}) := \mathrm{diag}\left(\sum_{i=1}^{m} a_{i1}, \ldots, \sum_{i=1}^{m} a_{in}\right).$$

Further, set

$$\mathbf{B}_{corr} := \mathbf{D}_{Brow}^{-1/2}\mathbf{B}\mathbf{D}_{Bcol}^{-1/2} \quad \text{and} \quad \mathbf{A}_{corr} := \mathbf{D}_{Arow}^{-1/2}\mathbf{A}\mathbf{D}_{Acol}^{-1/2}$$

for the transformed matrices obtained from $\mathbf{B}$ and $\mathbf{A}$ while carrying out correspondence analysis on $\mathbf{B}$ and the same correspondence transformation on $\mathbf{A}$. It is well known [6] that the leading singular value of $\mathbf{B}_{corr}$ is equal to 1 and the multiplicity of 1 as a singular value coincides with the number of irreducible blocks in $\mathbf{B}$. Let s_i denote a non-zero singular value of $\mathbf{B}_{corr}$ with unit-norm singular vector pair $\mathbf{v}_i$, $\mathbf{u}_i$. With the transformations

$$\mathbf{v}_{corr\,i} := \mathbf{D}_{Brow}^{-1/2}\mathbf{v}_i \quad \text{and} \quad \mathbf{u}_{corr\,i} := \mathbf{D}_{Bcol}^{-1/2}\mathbf{u}_i$$

the so-called correspondence vector pairs are obtained. If the coordinates $v_{corr\,i}(j)$, $u_{corr\,i}(\ell)$ of such a pair are regarded as possible values of two discrete random variables β_i and α_i (often called the ith correspondence factor pair) with the prescribed marginals, then, as in canonical analysis, their correlation is s_i, and this is the largest possible correlation under the condition that they are uncorrelated with the first $i-1$ correspondence factors within their own sets ($i = 2, \ldots, \min\{m, n\}$).

If $s_1 = 1$ is a single singular value, then $\mathbf{v}_{corr\,1}$ and $\mathbf{u}_{corr\,1}$ are the all 1 vectors and the corresponding β_1, α_1 pair is regarded as a trivial correspondence factor pair. This corresponds to the general case. Keeping $k \le \mathrm{rank}\,(\mathbf{B}_{corr}) = \mathrm{rank}\,(\mathbf{B}) = \mathrm{rank}\,(\mathbf{P})$ singular values with the coordinates of the corresponding $k-1$ non-trivial correspondence factor pairs, the following $(k-1)$-dimensional representation of the jth and ℓth categories of the two underlying discrete variables is obtained:

$$\mathbf{v}_{corr}^{j} := (v_{corr\,2}(j), \ldots, v_{corr\,k}(j)) \quad \text{and} \quad \mathbf{u}_{corr}^{\ell} := (u_{corr\,2}(\ell), \ldots, u_{corr\,k}(\ell)).$$

This representation has the following optimum properties: the closeness of categories of the same variable reflects the similarity between them, while the closeness of categories of different variables reflects their frequent simultaneous occurrence. For example, $\mathbf{B}$ being a microarray, the representatives of similar function genes as well as representatives of similar conditions are close to each other; also, representatives of genes that are responsible for a given condition are close to the representatives of those conditions. Now we prove the following.

Proposition 11. *Given the blown up matrix* $\mathbf{B}$*, under GC1 there exists a constant* $\delta \in (0, 1)$*, independent of* m *and* n*, such that all the* r *non-zero singular values of* $\mathbf{B}_{corr}$ *are in the interval* $[\delta, 1]$*, where* $r = \mathrm{rank}\,(\mathbf{B}) = \mathrm{rank}\,(\mathbf{P})$.

Proof. It is easy to see that $\mathbf{B}_{corr}$ is the blown up matrix of the $a \times b$ pattern matrix $\tilde{\mathbf{P}}$ with entries

$$\tilde{p}_{ij} = \frac{p_{ij}}{\sqrt{\left(\sum_{\ell=1}^{b} p_{i\ell} n_\ell\right)\left(\sum_{k=1}^{a} p_{kj} m_k\right)}}.$$

Following the considerations of the proof of Proposition 6, the blown up matrix $\mathbf{B}_{corr}$ has exactly $r = \mathrm{rank}\,(\mathbf{P}) = \mathrm{rank}\,(\tilde{\mathbf{P}})$ non-zero singular values that are the singular values of the $a \times b$ matrix $\mathbf{P}' = \mathbf{D}_m^{1/2}\tilde{\mathbf{P}}\mathbf{D}_n^{1/2}$ with entries

$$p'_{ij} = \frac{p_{ij}\sqrt{m_i}\sqrt{n_j}}{\sqrt{\left(\sum_{\ell=1}^{b} p_{i\ell} n_\ell\right)\left(\sum_{k=1}^{a} p_{kj} m_k\right)}} = \frac{p_{ij}}{\sqrt{\left(\sum_{\ell=1}^{b} p_{i\ell} \frac{n_\ell}{n_j}\right)\left(\sum_{k=1}^{a} p_{kj} \frac{m_k}{m_i}\right)}}.$$

Since the matrix $\mathbf{P}$ contains no identically zero rows or columns, the matrix $\mathbf{P}'$ varies on a compact set of $a \times b$ matrices determined by the inequalities (1). The range of the non-zero singular values depends continuously on the matrix, which does not depend on m and n. Therefore, the minimum non-zero singular value does not depend on m or n. Because the largest singular value is 1, this finishes the proof. □

Theorem 12. *Under GC1 and GC2, there exists a positive number δ (independent of m and n) such that for every $0 < \tau < 1/2$ the following statement holds almost surely: the r largest singular values of $\mathbf{A}_{corr}$ are in the interval $[\delta - \max\{n^{-\tau}, m^{-\tau}\}, 1 + \max\{n^{-\tau}, m^{-\tau}\}]$, while all the others are at most $\max\{n^{-\tau}, m^{-\tau}\}$.*

Proof. First notice that

$$\mathbf{A}_{corr} = \mathbf{D}_{Arow}^{-1/2}\mathbf{A}\mathbf{D}_{Acol}^{-1/2} = \mathbf{D}_{Arow}^{-1/2}\mathbf{B}\mathbf{D}_{Acol}^{-1/2} + \mathbf{D}_{Arow}^{-1/2}\mathbf{W}\mathbf{D}_{Acol}^{-1/2}. \tag{16}$$

The entries of $\mathbf{D}_{Brow}$ and those of $\mathbf{D}_{Bcol}$ are of order $\Theta(n)$ and $\Theta(m)$, respectively. Now we prove that for every $i = 1, \dots, m$ and $j = 1, \dots, n$, $|d_{Arow\,i} - d_{Brow\,i}| < n \cdot n^{-\tau}$ and $|d_{Acol\,j} - d_{Bcol\,j}| < m \cdot m^{-\tau}$ hold almost surely. To this end, we use Chernoff's inequality for large deviations (cf. [1, Lemma 4.2]):

$$\begin{aligned}\mathbb{P}\left(|d_{Arow\,i} - d_{Brow\,i}| > n \cdot n^{-\tau}\right) &= \mathbb{P}\left(\left|\sum_{j=1}^{n} w_{ij}\right| > n^{1-\tau}\right)\\ &< \exp\left\{-\frac{n^{2-2\tau}}{2\left(\mathrm{Var}\left(\sum_{j=1}^{n} w_{ij}\right) + Kn^{1-\tau}/3\right)}\right\} \le \exp\left\{-\frac{n^{2-2\tau}}{2(n\sigma^2 + Kn^{1-\tau}/3)}\right\}\\ &= \exp\left\{-\frac{n^{1-2\tau}}{2(\sigma^2 + Kn^{-\tau}/3)}\right\} \quad (i = 1, \dots, m),\end{aligned}$$

where the constant K is the uniform bound for $|w_{ij}|$'s and σ^2 is the bound for their variances. By virtue of *GC2* the following estimate holds with some $C_0 > 0$ and $C \ge 1$ (constants of *GC2*) and large enough n:

$$\begin{aligned}\mathbb{P}\left(|d_{Arow\,i} - d_{Brow\,i}| > n^{1-\tau} \text{ for all} i \in \{1, \dots, m\}\right) &\le m \cdot \exp\left\{-\frac{n^{1-2\tau}}{2(\sigma^2 + Kn^{-\tau}/3)}\right\}\\ &\le C_0 \cdot n^C \cdot \exp\left\{-\frac{n^{1-2\tau}}{2(\sigma^2 + Kn^{-\tau}/3)}\right\}\\ &= \exp\left\{\ln C_0 + C\ln n - \frac{n^{1-2\tau}}{2(\sigma^2 + Kn^{-\tau}/3)}\right\}.\end{aligned} \tag{17}$$

The estimation of probability

$$\mathbb{P}\left(|d_{Acol\,j} - d_{Bcol\,j}| > m^{1-\tau} \quad \text{for all } j \in \{1, \dots, n\}\right)$$

can be treated analogously (with $D_0 > 0$ and $D \ge 1$ of *GC2*). The right-hand side of (17) forms a convergent series; therefore

$$\min_{i\in\{1,\dots,m\}} |d_{Arow\,i}| = \Theta(n), \qquad \min_{j\in\{1,\dots,n\}} |d_{Acol\,j}| = \Theta(m) \tag{18}$$

hold almost surely.

Now it is straightforward to bound the norm of the second term of (16) by

$$\|\mathbf{D}_{Arow}^{-1/2}\| \cdot \|\mathbf{W}\| \cdot \|\mathbf{D}_{Acol}^{-1/2}\|. \tag{19}$$

As by Remark 5, $\|\mathbf{W}\| = \mathcal{O}(\sqrt{m+n})$ holds almost surely, the quantity (19) is at most of order $\sqrt{\frac{m+n}{mn}}$ almost surely. Hence, it is almost surely less than $\max\{n^{-\tau}, m^{-\tau}\}$.

To estimate the norm of the first term of (16) let us write it in the form

$$\mathbf{D}_{Arow}^{-1/2}\mathbf{B}\mathbf{D}_{Acol}^{-1/2} = \mathbf{D}_{Brow}^{-1/2}\mathbf{B}\mathbf{D}_{Bcol}^{-1/2} + \left[\mathbf{D}_{Arow}^{-1/2} - \mathbf{D}_{Brow}^{-1/2}\right]\mathbf{B}\mathbf{D}_{Bcol}^{-1/2} + \mathbf{D}_{Arow}^{-1/2}\mathbf{B}\left[\mathbf{D}_{Acol}^{-1/2} - \mathbf{D}_{Bcol}^{-1/2}\right]. \tag{20}$$

The first term is just $\mathbf{B}_{corr}$, so due to Proposition 11, we should prove only that the norms of both remainder terms are almost surely less than $\max\{n^{-\tau}, m^{-\tau}\}$. These two terms have a similar appearance; therefore it is enough to estimate one of them. For example, the second term can be bounded by

$$\|\mathbf{D}_{Arow}^{-1/2} - \mathbf{D}_{Brow}^{-1/2}\| \cdot \|\mathbf{B}\| \cdot \|\mathbf{D}_{Bcol}^{-1/2}\|. \tag{21}$$

The estimation of the first factor in (21) is as follows:

$$\|\mathbf{D}_{Arow}^{-1/2} - \mathbf{D}_{Brow}^{-1/2}\| = \max_{i\in\{1,\dots,m\}}\left(\frac{1}{\sqrt{d_{Arow\,i}}} - \frac{1}{\sqrt{d_{Brow\,i}}}\right)$$

$$= \max_{i\in\{1,\dots,m\}} \frac{|d_{Arow\,i} - d_{Brow\,i}|}{\sqrt{d_{Arow\,i}\cdot d_{Brow\,i}}(\sqrt{d_{Arow\,i}}+\sqrt{d_{Brow\,i}})}$$
$$\le \max_{i\in\{1,\dots,m\}} \frac{|d_{Arow\,i} - d_{Brow\,i}|}{\sqrt{d_{Arow\,i}\cdot d_{Brow\,i}}} \cdot \max_{i\in\{1,\dots,m\}} \frac{1}{(\sqrt{d_{Arow\,i}}+\sqrt{d_{Brow\,i}})}. \tag{22}$$

By relations (18), $\sqrt{d_{Arow\,i}\cdot d_{Browi}} = \Theta(n)$ for any $i = 1,\dots,m$, and hence,

$$\frac{|d_{Arow\,i} - d_{Brow\,i}|}{\sqrt{d_{Arow\,i}\cdot d_{Brow\,i}}} \le n^{-\tau}$$

almost surely; further $\max_{i\in\{1,\dots,m\}} \frac{1}{\sqrt{d_{Arowi}}+\sqrt{d_{Browi}}} = \Theta(\frac{1}{\sqrt{n}})$ almost surely.

Therefore the left-hand side of (22) can be estimated by $n^{-\tau-1/2}$ from above almost surely. For the further factors in (21) we obtain $\|\mathbf{B}\| = \Theta(\sqrt{mn})$ (see Proposition 6), while $\|\mathbf{D}_{Bcol}^{-1/2}\| = \Theta(\frac{1}{\sqrt{m}})$ almost surely. These together imply that

$$n^{-\tau-1/2}\cdot n^{1/2}m^{1/2}\cdot m^{-1/2} \le n^{-\tau} \le \max\{n^{-\tau}, m^{-\tau}\}.$$

This finishes the estimation of the first term in (16), and by Weyl's perturbation theorem the proof, too. □

Remark 13. In the Gaussian case the large deviation principle can be replaced by the simple estimation of the Gaussian probabilities with any $\kappa > 0$:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} w_{ij}\right| > \kappa\right) < \min\left(1, \frac{4\sigma}{\kappa\sqrt{2\pi n}}\exp\left\{-\frac{n}{2\sigma^2}\kappa^2\right\}\right).$$

Setting $\kappa = n^{-\tau}$ we get an estimate analogous to (17).

Suppose that the blown up matrix $\mathbf{B}$ is irreducible and its non-negative entries sum up to 1. This restriction does not affect the result of the correspondence analysis, that is the SVD of the matrix $\mathbf{B}_{corr}$. Remember that the non-zero singular values of $\mathbf{B}_{corr}$ are the numbers $1 = s_1 > s_2 \ge \dots \ge s_r > 0$ with unit-norm singular vector pairs $\mathbf{v}_i, \mathbf{u}_i$ having piecewise constant structure ($i = 1,\dots,r$). Set

$$F := \text{Span}\{\mathbf{v}_1,\dots,\mathbf{v}_r\} \quad \text{and} \quad G := \text{Span}\{\mathbf{u}_1,\dots,\mathbf{u}_r\}.$$

Let $0 < \tau < 1/2$ be arbitrary and $\epsilon := \max\{n^{-\tau}, m^{-\tau}\}$. Let us also denote the unit-norm, pairwise orthogonal left- and right-hand side singular vectors corresponding to the r singular values $z_1,\dots,z_r \in [\delta-\epsilon, 1+\epsilon]$ of $\mathbf{A}_{corr}$ — guaranteed by Theorem 12 under *GC2* – by $\mathbf{y}_1,\dots,\mathbf{y}_r \in \mathbb{R}^m$ and $\mathbf{x}_1,\dots,\mathbf{x}_r \in \mathbb{R}^n$, respectively.

Proposition 14. *With the above notation, under GC1 and GC2 the following estimate holds almost surely for the distance between* $\mathbf{y}_i$ *and F:*

$$\text{dist}(\mathbf{y}_i, F) \le \frac{\epsilon}{(\delta-\epsilon)} = \frac{1}{(\frac{\delta}{\epsilon}-1)} \quad (i = 1,\dots,r) \tag{23}$$

and analogously, for the distance between $\mathbf{x}_i$ *and G:*

$$\text{dist}(\mathbf{x}_i, G) \le \frac{\epsilon}{(\delta-\epsilon)} = \frac{1}{(\frac{\delta}{\epsilon}-1)} \quad (i = 1,\dots,r). \tag{24}$$

Proof. Follow the method of proving Proposition 9 — under *GC1* — with δ instead of Δ and ϵ instead of ε. Here *GC2* is necessary only for $\mathbf{A}_{corr}$ to have r outstanding singular values. □

Remark 15. The left-hand sides of (23) and (24) are almost surely of order $\max\{n^{-\tau}, m^{-\tau}\}$, tending to zero as $m, n \to \infty$ under *GC1* and *GC2*.

Proposition 14 implies the well-clustering property of the representatives of the two discrete variables by means of the noisy correspondence vector pairs

$$\mathbf{y}_{corr\,i} := \mathbf{D}_{Arow}^{-1/2}\mathbf{y}_i, \qquad \mathbf{x}_{corr\,i} := \mathbf{D}_{Acol}^{-1/2}\mathbf{x}_i \quad (i = 1,\dots,r).$$

Let $\mathbf{Y}_{corr}$ denote the $m \times r$ matrix that contains the left-hand side vectors $\mathbf{y}_{corr\,1},\dots,\mathbf{y}_{corr\,r}$ in its columns. Similarly, let $\mathbf{X}_{corr}$ denote the $n \times r$ matrix that contains the right-hand side vectors $\mathbf{x}_{corr\,1},\dots,\mathbf{x}_{corr\,r}$ in its columns. The r-dimensional representatives of α are the row vectors of $\mathbf{Y}_{corr}$ denoted by $\mathbf{y}^1_{corr},\dots,\mathbf{y}^m_{corr} \in \mathbb{R}^r$, while the r-dimensional representatives of β are the row vectors of $\mathbf{X}_{corr}$ denoted by $\mathbf{x}^1_{corr}, \dots, \mathbf{x}^n_{corr} \in \mathbb{R}^r$. With respect to the marginal distributions, let the a- and b-variances of these representatives be defined by

$$S_a^2(\mathbf{Y}_{corr}) = \min_{\{A'_1,\dots,A'_a\}} \sum_{i=1}^{a} \sum_{j\in A'_i} d_{Arow\,j} \|\mathbf{y}^j_{corr} - \bar{\mathbf{y}}^i_{corr}\|^2 ,$$

$$S_b^2(\mathbf{X}_{corr}) = \min_{\{B'_1,\dots,B'_b\}} \sum_{i=1}^{b} \sum_{j\in B'_i} d_{Acol\,j} \|\mathbf{x}^{(j)}_{corr} - \bar{\mathbf{x}}^i_{corr}\|^2,$$

where $\{A'_1, \dots, A'_a\}$ and $\{B'_1, \dots, B'_b\}$ are a- and b-partitions of the genes and conditions, respectively,

$$\bar{\mathbf{y}}^i_{corr} = \sum_{j\in A'_i} d_{Arow\,j}\mathbf{y}^j_{corr} \quad \text{and} \quad \bar{\mathbf{x}}^i_{corr} = \sum_{j\in B'_i} d_{Acol\,j}\mathbf{x}^j_{corr}.$$

Theorem 16. *With the above notation, under GC1 and GC2,*

$$S_a^2(\mathbf{Y}_{corr}) \le \frac{r}{(\frac{\delta}{\epsilon}-1)^2} \quad \textit{and} \quad S_b^2(\mathbf{X}_{corr}) \le \frac{r}{(\frac{\delta}{\epsilon}-1)^2}$$

hold almost surely, where $\epsilon = \max\{n^{-\tau}, m^{-\tau}\}$ *with every* $0 < \tau < 1/2$.

Proof. An easy calculation shows that

$$S_a^2(\mathbf{Y}_{corr}) \le \sum_{i=1}^{a} \sum_{j\in A_i} d_{Arow\,j} \|\mathbf{y}^j_{corr} - \bar{\mathbf{y}}^i_{corr}\|^2 = \sum_{i=1}^{r} \operatorname{dist}^2(\mathbf{y}_i, F),$$

$$S_b^2(\mathbf{X}_{corr}) \le \sum_{i=1}^{b} \sum_{j\in B_i} d_{Acol\,j} \|\mathbf{x}^{(j)}_{corr} - \bar{\mathbf{x}}^i_{corr}\|^2 = \sum_{i=1}^{r} \operatorname{dist}^2(\mathbf{x}_i, G),$$

and hence the result of Proposition 14 can be used. □

Under *GC1* and *GC2* with m, n large enough, Theorem 16 implies that after performing correspondence analysis on the noisy matrix $\mathbf{A}$, the representation through the correspondence vectors belonging to $\mathbf{A}_{corr}$ will also reveal the block structure behind $\mathbf{A}$.

6. Recognizing the structure

One might wonder where the singular values of an $m \times n$ matrix $\mathbf{A} = (a_{ij})$ are located if $a := \max_{i,j} |a_{ij}|$ is independent of m and n. On one hand, the maximum singular value cannot exceed $\mathcal{O}(\sqrt{mn})$, as it is at most $\sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$. On the other hand, let $\mathbf{Q}$ be an $m \times n$ random matrix with entries a or $-a$ (independently of each other). Consider the spectral norm of all such matrices and take the minimum of them: $\min_{\mathbf{Q}\in\{-a,+a\}^{m\times n}} \|\mathbf{Q}\|$. This quantity measures the minimum linear structure that a matrix of the same size and magnitude as $\mathbf{A}$ can possess. As the Frobenius norm of $\mathbf{Q}$ is $a\sqrt{mn}$, by virtue of inequalities between spectral and Frobenius norms, the above minimum is at least $\frac{a}{\sqrt{2}}\sqrt{m+n}$, which is exactly the order of the spectral norm of a noise matrix.

In summary, an $m \times n$ random matrix – whose entries are independent and uniformly bounded – under very general conditions has at least one singular value of order greater than $\sqrt{m+n}$. Suppose there are k such singular values and the representatives by means of the corresponding singular vector pairs can be well classified in the sense of Theorem 10 (cf. the introduction to that theorem). Under these conditions we can reconstruct a blown up structure behind our matrix.

Theorem 17. *Let* $\mathbf{A}_{m\times n}$ *be a sequence of* $m \times n$ *matrices, where m and n tend to infinity. Assume that* $\mathbf{A}_{m\times n}$ *has exactly k singular values of order greater than* $\sqrt{m+n}$ *(k is fixed). If there are integers* $a \ge k$ *and* $b \ge k$ *such that the a- and b-variances of the row- and column-representatives are* $\mathcal{O}(\frac{m+n}{mn})$*, then there is an explicit construction for a blown up matrix* $\mathbf{B}_{m\times n}$ *such that* $\mathbf{A}_{m\times n} = \mathbf{B}_{m\times n} + \mathbf{E}_{m\times n}$*, with* $\|\mathbf{E}_{m\times n}\| = \mathcal{O}(\sqrt{m+n})$.

Proof. In the sequel the subscripts m and n will be dropped. We shall speak in terms of microarrays (genes and conditions).

Let $\mathbf{y}_1, \dots, \mathbf{y}_k \in \mathbb{R}^m$ and $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ denote the left- and right-hand side unit-norm singular vectors corresponding to $z_1, \dots, z_k$, the singular values of $\mathbf{A}$ of order larger than $\sqrt{m+n}$. The k-dimensional representatives of the genes and conditions — that are row vectors of the $m \times k$ matrix $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_k)$ and those of the $n \times k$ matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$, respectively — by the condition of the theorem form a and b clusters in $\mathbb{R}^k$, respectively, with sums of inner variances $\mathcal{O}(\frac{m+n}{mn})$. Reorder the rows and columns of $\mathbf{A}$ according to their respective cluster memberships. Denote by $\mathbf{y}^1, \dots, \mathbf{y}^m \in \mathbb{R}^k$ and $\mathbf{x}^1, \dots, \mathbf{x}^n \in \mathbb{R}^k$ the Euclidean representatives of the genes and conditions (the rows of the reordered $\mathbf{Y}$ and $\mathbf{X}$), and let $\bar{\mathbf{y}}^1, \dots, \bar{\mathbf{y}}^a \in \mathbb{R}^k$ and $\bar{\mathbf{x}}^1, \dots, \bar{\mathbf{x}}^b \in \mathbb{R}^k$ denote the cluster centers, respectively. Now let us choose the following new representation of the genes and conditions. The genes' representatives are row vectors of the $m \times k$ matrix $\tilde{\mathbf{Y}}$ such that the first m_1 rows of $\tilde{\mathbf{Y}}$ are equal to $\bar{\mathbf{y}}^1$, the next m_2 rows to $\bar{\mathbf{y}}^2$, and so on; the last m_a rows of $\tilde{\mathbf{Y}}$ are equal to $\bar{\mathbf{y}}^a$. And similarly, the conditions' representatives are row vectors of the $n \times k$ matrix $\tilde{\mathbf{X}}$ such that the first n_1 rows of $\tilde{\mathbf{X}}$ are equal to $\bar{\mathbf{x}}^1$, and so on; the last n_b rows of $\tilde{\mathbf{X}}$ are equal to $\bar{\mathbf{x}}^b$.

By the considerations of Theorem 10 and the assumption for the clusters,

$$\sum_{i=1}^{k} \mathrm{dist}^2(\mathbf{y}_i, F) = S_a^2(\mathbf{Y}) = \mathcal{O}\left(\frac{m+n}{mn}\right) \tag{25}$$

and

$$\sum_{i=1}^{k} \mathrm{dist}^2(\mathbf{x}_i, G) = S_b^2(\mathbf{X}) = \mathcal{O}\left(\frac{m+n}{mn}\right) \tag{26}$$

hold respectively, where the k-dimensional subspace $F \subset \mathbb{R}^m$ is spanned by the column vectors of $\tilde{\mathbf{Y}}$, while the k-dimensional subspace $G \subset \mathbb{R}^n$ is spanned by the column vectors of $\tilde{\mathbf{X}}$. We follow the construction given in [20] (see Proposition 2) of a set $\mathbf{v}_1, \dots, \mathbf{v}_k$ of orthonormal vectors within F and another set $\mathbf{u}_1, \dots, \mathbf{u}_k$ of orthonormal vectors within G such that

$$\sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{v}_i\|^2 = \min_{\mathbf{v}'_1,\dots,\mathbf{v}'_k} \sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{v}'_i\|^2 \le 2\sum_{i=1}^{k} \mathrm{dist}^2(\mathbf{y}_i, F) \tag{27}$$

and

$$\sum_{i=1}^{k} \|\mathbf{x}_i - \mathbf{u}_i\|^2 = \min_{\mathbf{u}'_1,\dots,\mathbf{u}'_k} \sum_{i=1}^{k} \|\mathbf{x}_i - \mathbf{u}'_i\|^2 \le 2\sum_{i=1}^{k} \mathrm{dist}^2(\mathbf{x}_i, G) \tag{28}$$

hold, where the minimum is taken over orthonormal sets of vectors $\mathbf{v}'_1, \dots, \mathbf{v}'_k \in F$ and $\mathbf{u}'_1, \dots, \mathbf{u}'_k \in G$, respectively. The construction of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ is as follows ($\mathbf{u}_1, \dots, \mathbf{u}_k$ can be constructed in the same way). Let $\mathbf{v}'_1, \dots, \mathbf{v}'_k \in F$ be an arbitrary orthonormal system (obtained, e.g., by the Schmidt orthogonalization method). Let $\mathbf{V}' = (\mathbf{v}'_1, \dots, \mathbf{v}'_k)$ be an $m \times k$ matrix and

$$\mathbf{Y}^{\mathrm{T}}\mathbf{V}' = \mathbf{QSZ}^{\mathrm{T}}$$

be a SVD, where the matrix $\mathbf{S}$ contains the singular values of the $k \times k$ matrix $\mathbf{Y}^{\mathrm{T}}\mathbf{V}'$ in its main diagonal and zeros otherwise, while $\mathbf{Q}$ and $\mathbf{Z}$ are $k \times k$ orthogonal matrices (containing the corresponding unit-norm singular vector pairs in their columns). The orthogonal matrix $\mathbf{R} = \mathbf{ZQ}^{\mathrm{T}}$ will give the convenient orthogonal rotation of the vectors $\mathbf{v}'_1, \dots, \mathbf{v}'_k$. That is, the column vectors of the matrix $\mathbf{V} = \mathbf{V}'\mathbf{R}$ form also an orthonormal set that is the desired set $\mathbf{v}_1, \dots, \mathbf{v}_k$.

Define the error terms $\mathbf{r}_i$ and $\mathbf{q}_i$, respectively:

$$\mathbf{r}_i = \mathbf{y}_i - \mathbf{v}_i \quad \text{and} \quad \mathbf{q}_i = \mathbf{x}_i - \mathbf{u}_i \quad (i = 1, \dots, k).$$

In view of (25)–(28),

$$\sum_{i=1}^{k} \|\mathbf{r}_i\|^2 = \mathcal{O}\left(\frac{m+n}{mn}\right) \quad \text{and} \quad \sum_{i=1}^{k} \|\mathbf{q}_i\|^2 = \mathcal{O}\left(\frac{m+n}{mn}\right). \tag{29}$$

Consider the following decomposition:

$$\mathbf{A} = \sum_{i=1}^{k} z_i \mathbf{y}_i \mathbf{x}_i^{\mathrm{T}} + \sum_{i=k+1}^{\min\{m,n\}} z_i \mathbf{y}_i \mathbf{x}_i^{\mathrm{T}}.$$

The spectral norm of the second term is at most of order $\sqrt{m+n}$. Now consider the first term,

$$\sum_{i=1}^{k} z_i \mathbf{y}_i \mathbf{x}_i^{\mathrm{T}} = \sum_{i=1}^{k} z_i (\mathbf{v}_i + \mathbf{r}_i)(\mathbf{u}_i^{\mathrm{T}} + \mathbf{q}_i^{\mathrm{T}}) = \sum_{i=1}^{k} z_i \mathbf{v}_i \mathbf{u}_i^{\mathrm{T}} + \sum_{i=1}^{k} z_i \mathbf{v}_i \mathbf{q}_i^{\mathrm{T}} + \sum_{i=1}^{k} z_i \mathbf{r}_i \mathbf{u}_i^{\mathrm{T}} + \sum_{i=1}^{k} z_i \mathbf{r}_i \mathbf{q}_i^{\mathrm{T}}. \tag{30}$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_k$ and $\mathbf{u}_1, \dots, \mathbf{u}_k$ are unit vectors, the last three terms in (30) can be estimated by means of the relations

$$\|\mathbf{v}_i \mathbf{u}_i^{\mathrm{T}}\| = \sqrt{\|\mathbf{v}_i \mathbf{u}_i^{\mathrm{T}} \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}\|} = 1 \quad (i = 1, \dots, k),$$

$$\|\mathbf{v}_i \mathbf{q}_i^{\mathrm{T}}\| = \sqrt{\|\mathbf{q}_i \mathbf{v}_i^{\mathrm{T}} \mathbf{v}_i \mathbf{q}_i^{\mathrm{T}}\|} = \|\mathbf{q}_i\| \quad (i = 1, \dots, k),$$

$$\|\mathbf{r}_i \mathbf{u}_i^{\mathrm{T}}\| = \sqrt{\|\mathbf{r}_i \mathbf{u}_i^{\mathrm{T}} \mathbf{u}_i \mathbf{r}_i^{\mathrm{T}}\|} = \|\mathbf{r}_i\| \quad (i = 1, \dots, k),$$

$$\|\mathbf{r}_i \mathbf{q}_i^{\mathrm{T}}\| = \sqrt{\|\mathbf{r}_i \mathbf{q}_i^{\mathrm{T}} \mathbf{q}_i \mathbf{r}_i^{\mathrm{T}}\|} = \|\mathbf{q}_i\| \cdot \|\mathbf{r}_i\| \quad (i = 1, \dots, k).$$

Taking into account that z_i cannot exceed $\Theta(\sqrt{mn})$ and k is fixed, due to (29) we get that the spectral norms of the last three terms in (30) – for their finitely many subterms the triangle inequality is applicable – are at most of order $\sqrt{m+n}$. Let $\mathbf{B}$ be the first term, i.e.,

$$\mathbf{B} = \sum_{i=1}^{k} z_i \mathbf{v}_i \mathbf{u}_i^{\mathrm{T}}.$$

Then $\|\mathbf{A} - \mathbf{B}\| = \mathcal{O}(\sqrt{m+n})$.

By definition, the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ and the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ are in the subspaces F and G, respectively. Both spaces consist of piecewise constant vectors; thus the matrix $\mathbf{B}$ is a blown up matrix containing $a \times b$ blocks. The noise matrix is

$$\mathbf{E} = \sum_{i=1}^{k} z_i \mathbf{v}_i \mathbf{q}_i^{\mathrm{T}} + \sum_{i=1}^{k} z_i \mathbf{r}_i \mathbf{u}_i^{\mathrm{T}} + \sum_{i=1}^{k} z_i \mathbf{r}_i \mathbf{q}_j^{\mathrm{T}} + \sum_{i=k+1}^{\min\{m,n\}} z_i \mathbf{y}_i \mathbf{x}_i^{\mathrm{T}}$$

which finishes the proof. □

Then, provided the conditions of Theorem 17 hold, by the construction given in the proof above, an algorithm can be written that uses several SVD's and produces the blown up matrix $\mathbf{B}$. This $\mathbf{B}$ can be regarded as the best blown up approximation of the microarray $\mathbf{A}$. At the same time clusters of the genes and conditions are also obtained. More precisely, first we conclude the clusters from the SVD of $\mathbf{A}$, rearrange the rows and columns of $\mathbf{A}$ accordingly, and afterwards we use the above construction. If we decide to perform correspondence analysis on $\mathbf{A}$ then by (16) and (20), $\mathbf{B}_{corr}$ will give a good approximation to $\mathbf{A}_{corr}$ and similarly, the correspondence vectors obtained by the SVD of $\mathbf{B}_{corr}$ will give representatives of the genes and conditions.

To find the SVD for large rectangular matrices randomized algorithms are favored, e.g., [22]. In the case of random matrices with an underlying linear structure (outstanding singular values), the random noise of the algorithm is just added to the noise in our data, but their sum is also a noise matrix, so it does not change the effect of our algorithm in finding the clusters. Under the conditions of Theorem 17, the separated error matrix is comparable with the noise matrix, and this fact guarantees that the underlying block structure can be extracted.

Acknowledgments

The authors are indebted to Gábor Tusnády for suggesting the microarray problem and supplying computer simulations, to Péter Major for inspiring discussions on concentration inequalities, and also to the anonymous referees of the paper for their constructive remarks.

References

[1] M. Bolla, Recognizing linear structure in noisy matrices, Linear Algebra and its Applications 402 (2005) 228–244.
[2] Y. Kluger, R. Basri, J.T. Chang, M. Gerstein, Spectral biclustering of microarray data: Coclustering genes and conditions, Genome Research 13 (2003) 703–716.
[3] N. Alon, M. Krivelevich, V.H. Vu, On the concentration of eigenvalues of random symmetric matrices, Israel Journal of Mathematics 131 (2002) 259–267.
[4] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, Advances in Mathematics 195 (2005) 491–523.
[5] Z. Füredi, J. Komlós, The eigenvalues of random symmetric matrices, Combinatorica 1 (3) (1981) 233–241.
[6] M.J. Greenacre, Correspondence Analysis, Academic Press, New York, 1984.
[7] D.J. Higham, G. Kalna, J.K. Vass, Analysis of the singular value decomposition as a tool for processing microarray expression data, in: Proc. of ALGORITMY, 2005, pp. 250–259.
[8] L. Liu, D.M. Hawkins, S. Ghosh, S.S. Young, Robust singular value decomposition analysis of microarray data, Proceedings of the National Academy of Sciences of the US 100 (23) (2003) 13167–13172.
[9] L. Omberg, G.H. Golub, O. Alter, A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies, Proceedings of the National Academy of Sciences of the US 104 (47) (2007) 18371–18376.
[10] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Publications Mathématiques. Institute de Hautes Études Scientifiques 81 (1995) 73–203.
[11] M. Meckes, Concentration of norms and eigenvalues of random matrices, Journal of Functional Analysis 211 (2) (2004) 508–524.
[12] E. Wigner, On the distribution of the roots of certain symmetric matrices, Annals of Mathematics 67 (1958) 325–328.
[13] C.A. Tracy, H. Widom, Level-spacing distributions and the Airy kernel, Communications in Mathematical Physics 159 (1994) 151–174.
[14] A. Soshnikov, Universality at the edge of the spectrum of Wigner random matrices, Communications in Mathematical Physics 207 (1999) 697–733.
[15] D. Féral, S. Péché, The largest eigenvalue of rank one deformation of large Wigner matrices, Communications in Mathematical Physics 272 (1) (2007) 185–228.
[16] I. Johnstone, On the distribution of the largest eigenvalue in principal component analysis, Annals of Statistics 29 (2001) 295–327.
[17] J. Baik, J. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, Journal of Multivariate Analysis 97 (2006) 1382–1408.
[18] I. Olkin, The 70th anniversary of the distribution of random matrices: A survey, Linear Algebra and its Applications 354 (2002) 231–243.
[19] R. Bhatia, Matrix Analysis, Graduate Texts in Mathematics, vol. 169, Springer, New York, 1996.
[20] M. Bolla, Distribution of the eigenvalues of random block-matrices, Linear Algebra and its Applications 377 (2004) 219–240.
[21] B. Bollobás, Random Graphs, second ed., Cambridge University Press, Cambridge, 2001.
[22] D. Achlioptas, F. McSherry, Fast computation of low rank matrix approximations, in: Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing, ACM, New York, 2001, pp. 611–618.