SINGULAR VALUE DECOMPOSITION OF LARGE RANDOM MATRICES

Marianna Bolla[1], Katalin Friedl[2], András Krámli[3]

[1]Institute of Mathematics, Budapest University of Technology and Economics
[2]Department of Computer Science, Budapest University of Technology and Economics
[3]Bolyai Institute, University of Szeged

Abstract
Asymptotic behavior of the singular value decomposition (SVD) of blown up matrices and normalized blown up contingency tables exposed to Wigner-noise is investigated. It is proved that such an $m \times n$ 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 $m, n \to \infty$. Concentration results of Alon at al. for the eigenvalues of large symmetric random matrices are adapted to the rectangular case, and on this basis, almost sure results for the singular values as well as for the corresponding isotropic subspaces are proved. An algorithm, applicable to two-way classification of microarrays, is also given that finds the underlying block structure.

AMS classification: 15A42, 15A52, 60E15

Keywords: Concentration of singular values; Two-way classification of microarrays; Perturbation of correspondence matrices; Almost sure convergence by large deviations

– – – – – – – – – –

[1]Corresponding author: P.O.Box 91, Bldg. H5/7, 1521 Budapest, Hungary

E-mail addresses: [1]marib@math.bme.hu, [2]friedl@cs.bme.hu, [3]kramli@sztaki.hu

1. Introduction

The purpose of this paper is to fill the gap between the theory of random matrices and the challenge of finding linear structure in large real-world data sets like internet or microarray measurements.

In [5], 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 protruding eigenvalues (of order n), while the majority of the eigenvalues is at most of order $\sqrt{n}$ with probability tending to 1, as $n \to \infty$. These provide a useful tool to recognize linear structure in large symmetric real matrices, such as weight matrices of random graphs on a large number of vertices produced by communication, social, or cellular networks. 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 formed by categorical variables in order to perform two-way clustering of these variables.

First we introduce some notation.

Definition 1.1. The $m \times n$ real matrix $\mathbf{W}$ is a *Wigner-noise* if its entries w_{ij} ($1 \le i \le m$, $1 \le j \le n$) are independent random variables, $\mathbb{E}(w_{ij}) = 0$, $\mathrm{Var}\,(w_{ij}) \le \sigma^2$ with some $0 < \sigma < \infty$ that does not depend on n and m, and the w_{ij}'s are uniformly bounded (i.e., there is a constant $K > 0$ such that $|w_{ij}| \le K$).

Though, the main results of this paper can be extended to w_{ij}'s with any light-tail distribution (especially for the case of Gaussian distributed w_{ij}'s), our almost sure results will be based on the assumptions of Definition 1.1.

According to a generalization of a theorem of Füredi and Komlós [8] to rectangular matrices, the following result is valid for $\mathbf{W}$ (see [1]).

Lemma 1.2. The maximum singular value of the Wigner-noise $\mathbf{W}$ is at most of order $\sqrt{m+n}$ with probability tending to 1, as $n, m \to \infty$.

Definition 1.3. 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 $0 \le p_{ij} \le 1$, further there are positive integers $m_1, \dots, m_a$ with $\sum_{i=1}^{a} m_i = m$ and $n_1, \dots, n_b$ with $\sum_{i=1}^{b} n_i = n$, respectively, such that the matrix $\mathbf{B}$ can be divided into $a \times b$ blocks, the block (i, j) being an $m_i \times n_j$ matrix with entries all equal to p_{ij} ($1 \le i \le a$, $1 \le j \le b$).

Such schemes are sought for in microarray analysis and they are called chess-board patterns, cf. [10]. 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 Wigner-noise 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, both at the same rate. More precisely, we make two different constraints on the growth of the sizes m, n, and the growth rate of their components. The first one is needed for all our reasonings, while the second one will be used in the case of noisy correspondence matrices, only.

Definition 1.4.

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 $C \geq 1$, $D \geq 1$, and $C_0 > 0$, $D_0 > 0$ such that $m \leq C_0 \cdot n^C$ and $n \leq D_0 \cdot m^D$ hold for sufficiently large m and n.

Remark 1.5.

1. *GC1* implies that
$$c \leq \frac{m_k}{m_i} \leq \frac{1}{c} \quad \text{and} \quad d \leq \frac{n_l}{n_j} \leq \frac{1}{d} \tag{1.1}$$
hold for any pair of indices $k, i \in \{1, \dots, a\}$ and $l, j \in \{1, \dots, b\}$.
2. *GC2* implies that
$$(\frac{1}{D_0})^{1/D} \cdot n^{1/D} \leq m \leq C_0 \cdot n^C \quad \text{and} \quad (\frac{1}{C_0})^{1/C} \cdot m^{1/C} \leq n \leq D_0 \cdot m^D$$
hold for sufficiently large m and n.

Now, let $\mathbf{B}$ be a blown up matrix (Definition 1.3) and $\mathbf{W}$ be a Wigner-noise of the corresponding size (Definition 1.1). We want to establish some property $\mathcal{P}_{m,n}$ that holds for the $m \times n$ random matrix $\mathbf{A} = \mathbf{B} + \mathbf{W}$ (briefly, $\mathbf{A}_{m\times n}$) with m and n large enough. In this paper $\mathcal{P}_{m,n}$ is mostly related to the SVD of $\mathbf{A}_{m\times n}$. We will consider two types of convergences.

Definition 1.6.

C1 (Convergence in probability). We say that the property $\mathcal{P}_{m,n}$ holds for $\mathbf{A}_{m\times n}$ in probability (with probability tending to 1) if
$$\lim_{m,n\to\infty} \mathbb{P}\left(\mathbf{A}_{m\times n} \text{ has } \mathcal{P}_{m,n}\right) = 1.$$

C2 (Almost sure convergence). We say that the property $\mathcal{P}_{m,n}$ holds for $\mathbf{A}_{m\times n}$ almost surely (with probability 1) if
$$\mathbb{P}\left(\exists\ m_0, n_0 \in \mathbb{N} \text{ such that for } m \geq m_0 \text{ and } n \geq n_0\ \mathbf{A}_{m\times n} \text{ has } \mathcal{P}_{m,n}\right) = 1.$$

Here we may assume *GC1* or *GC2* for the growth of m and n.

Remark 1.7.

1. *C2* always implies *C1*.
2. Conversely, if in addition to *C1*, $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} p_{mn} < \infty$ 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, $\mathbf{A}_{m\times n}$ has $\mathcal{P}_{m,n}$ almost surely and so, *C2* is true.

In combinatorics literature *C1* is frequently called almost sure convergence (this was also the case in [5]). However, from probabilistic point of view, type *C2* convergence is much stronger than *C1*, and it makes a difference in practice: *C2* guarantees that no matter how $\mathbf{A}_{m\times n}$ is selected, it must have property $\mathcal{P}_{m,n}$ if m and n are large enough.

For example, Lemma 1.2 states that the spectral norm of a Wigner-noise $\mathbf{W}_{m\times n}$ is $\mathcal{O}(\sqrt{m+n})$ in probability (type *C1* convergence). To prove almost sure (type

C2) convergence, a recent sharp concentration theorem of N. Alon at al. plays a crucial role. For the completeness we formulate this result (cf. [2]).

Lemma 1.8. Let $\widetilde{\mathbf{W}}$ be a $q \times q$ real symmetric matrix whose entries in and above the main diagonal are independent random variables with absolute value at most 1. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$ be the eigenvalues of $\widetilde{\mathbf{W}}$. The following estimate holds for the deviation of the ith largest eigenvalue from its expectation with any positive real number t:

$$\mathbb{P}\left(|\lambda_i - \mathbb{E}(\lambda_i)| > t\right) \leq \exp\left(-\frac{(1-o(1))t^2}{32i^2}\right) \quad \text{when} \quad i \leq \frac{q}{2},$$

and the same estimate holds for the probability $\mathbb{P}\left(|\lambda_{q-i+1} - \mathbb{E}(\lambda_{q-i+1})| > t\right)$.

Now let $\mathbf{W}$ be a Wigner-noise with entries uniformly bounded by K. The $(m+n) \times (m+n)$ symmetric matrix

$$\widetilde{\mathbf{W}} = \frac{1}{K} \cdot \begin{pmatrix} \mathbf{0} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{0} \end{pmatrix}$$

satisfies the conditions of Lemma 1.8, its largest and smallest eigenvalues being

$$\lambda_i(\widetilde{\mathbf{W}}) = -\lambda_{n+m-i+1}(\widetilde{\mathbf{W}}) = \frac{1}{K} \cdot s_i(\mathbf{W}), \qquad i = 1, \dots, \min\{m, n\},$$

the others are zeros, where $\lambda_i(.)$ and $s_i(.)$ denote the ith largest eigenvalue and singular value of the matrix in the argument, respectively (cf. [3]). Therefore

$$\mathbb{P}\left(|s_1(\mathbf{W}) - \mathbb{E}(s_1(\mathbf{W}))| > t\right) \leq \exp\left(-\frac{(1-o(1))t^2}{32K^2}\right). \tag{1.2}$$

Lemma 1.2 asserts that $\mathbf{W}$'s spectral norm $\|\mathbf{W}\| = s_1(\mathbf{W}) = \mathcal{O}(\sqrt{m+n})$ in probability. This fact together with inequality (1.2) ensures 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 holds.

Lemma 1.9. There exist two positive constants C_{K1} and C_{K2}, depending on the common bound for the entries of $\mathbf{W}$, such that

$$\mathbb{P}\left(\|\mathbf{W}\| > C_{K1} \cdot \sqrt{m+n}\right) \leq \exp[-C_{K2} \cdot (m+n)]. \tag{1.3}$$

The exponential decay of the right hand side of (1.3) together with the second part of Remark 1.7 implies that the spectral norm of a Wigner-noise $\mathbf{W}_{m\times n}$ is $\mathcal{O}(\sqrt{m+n})$, almost surely. This observation will provide the base of *C2* (almost sure) results of Sections 2 and 3.

In case of a Wigner noise with Gaussian distributed entries relying upon the Tracy-Widom distribution [13] of the maximal eigenvalue of the above type matrices, only *C1* (convergence in probability) results can be proved. Also, for other type of distribution of $\mathbf{W}$, the methods of Mehta [11] and Olkin [12] may be used to find the joint distribution of its singular values (for the distribution of the ordered singular values we do not know similar results).

In Section 2 we shall prove that the $m \times n$ noisy matrix $\mathbf{A} = \mathbf{B} + \mathbf{W}$ almost surely has $r = \text{rank}\,(\mathbf{P})$ protruding singular values of order $\sqrt{mn}$. In Section 3 the distances of the corresponding isotropic subspaces are estimated and this gives rise to a two-way classification of the row and column items of $\mathbf{A}$ with sum of inner variances $\mathcal{O}(\frac{m+n}{mn})$, almost surely.

In Definition 1.3 we required that the entries of the pattern matrix $\mathbf{P}$ be in the [0,1] interval. We made this restriction only for the sake of the generalized Erdős–Rényi hypergraph model to be introduced with the entries of $\mathbf{P}$ as probabilities. In fact, our results are valid for any pattern matrix with fixed sizes and with nonnegative entries. For example, in microarray measurements the rows correspond to different genes, the columns correspond to different conditions, and the entries are the expression levels of a specific gene under a specific condition.

Sometimes the pattern matrix $\mathbf{P}$ is an $a \times b$ contingency table with entries that are nonnegative integers. Then the blown up matrix $\mathbf{B}$ can be regarded as a larger ($m \times n$) contingency table that contains e.g., counts for two categorical variables with m and n different categories, respectively. For finding maximally correlated factors with respect to the marginal distributions of these two discrete variables, the technique of correspondence analysis is widely used, see [6]. In case of a general pattern matrix $\mathbf{P}$ (with nonnegative real entries), the blown-up matrix $\mathbf{B}$ can also be regarded as a data matrix for two not independent categorical variables. As the categories may be measured in different units, a normalization is necessary. This normalization is made by dividing the entries of $\mathbf{B}$ by the square roots of the corresponding row and column sums (cf. [10]). This transformation is identical to that of the correspondence analysis, and the transformed matrix remains the same when we multiply the initial matrix by a positive constant. Thus, it does not matter whether we started with a contingency or frequency table or just with a matrix with nonnegative entries. The transformed matrix $\mathbf{B}_{corr}$, which belongs to $\mathbf{B}$, has entries in [0,1] and maximum singular value 1. It is proved that there is a remarkable gap between the $\text{rank}\,(\mathbf{B}) = \text{rank}\,(\mathbf{P})$ largest and the other singular values of $\mathbf{A}_{corr}$, the matrix obtained from the noisy matrix $\mathbf{A} = \mathbf{B} + \mathbf{W}$ by the correspondence transformation. This implies well two-way classification properties of the row and column categories (genes and expression levels) in Section 4.

In Section 5 a construction is given, how a blown up structure behind a real-life matrix with a few protruding singular values and "well classifiable" corresponding singular vector pairs can be found. To find SVD of large rectangular matrices randomized algorithms are favored. They exploit the randomness of our data and provide good approximations of the underlying clusters only if originally there was a linear structure in our matrix.

2. Singular values of a noisy matrix

Proposition 2.1. 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 $s_1 \geq s_2 \geq \cdots \geq s_r > 0$ 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 [9].

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

$$\mathbf{B}\mathbf{u} = s \cdot \mathbf{v} \tag{2.1}$$

has the form

$$\sum_{j=1}^{b} n_j p_{ij} u(j) = s \cdot v(i) \qquad (i = 1, \dots, a). \tag{2.2}$$

With the notations

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

(2.2) can be written as

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

Further, introducing the transformations

$$\mathbf{w} = \mathbf{D}_n^{1/2} \tilde{\mathbf{u}}, \qquad \mathbf{z} = \mathbf{D}_m^{1/2} \tilde{\mathbf{v}}, \tag{2.3}$$

the equivalent equation

$$\mathbf{D}_m^{1/2} \mathbf{P} \mathbf{D}_n^{1/2} \mathbf{w} = s \cdot \mathbf{z} \tag{2.4}$$

is obtained. It is very important that the transformation (2.3) results in unit-norm vectors, that is

$$\|\mathbf{w}\|^2 = \sum_{j=1}^{b} n_j u^2(j) = \|\mathbf{u}\|^2 = 1 \quad \text{and} \quad \|\mathbf{z}\|^2 = \sum_{i=1}^{a} m_i v^2(i) = \|\mathbf{v}\|^2 = 1.$$

Furthermore, applying the transformation (2.3) for the $\tilde{\mathbf{u}}_k, \tilde{\mathbf{v}}_k$ pairs obtained from the $\mathbf{u}_k, \mathbf{v}_k$ pairs $(k = 1, \dots, r)$, the orthogonality is also preserved, since

$$\mathbf{w}_k{}^T \cdot \mathbf{w}_l = \sum_{j=1}^{b} n_j u_k(j) u_l(j) = 0 \quad \text{and} \quad \mathbf{z}_k{}^T \cdot \mathbf{z}_l = \sum_{i=1}^{a} m_i v_k(i) v_l(i) = 0 \quad (k \neq l).$$

Consequently, $\mathbf{z}_k$, $\mathbf{w}_k$ is a singular vector pair corresponding to the 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

$$\widetilde{\mathbf{D}}_m = \frac{1}{m}\mathbf{D}_m, \quad \widetilde{\mathbf{D}}_n = \frac{1}{n}\mathbf{D}_n$$

(2.4) is also equivalent to

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

that is the $a \times b$ matrix $\widetilde{\mathbf{D}}_m^{1/2}\mathbf{P}\widetilde{\mathbf{D}}_n^{1/2}$ 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 $\widetilde{\mathbf{D}}_m^{1/2}\mathbf{P}\widetilde{\mathbf{D}}_n^{1/2}$.

Now we want to establish relations between the singular values of $\mathbf{P}$ and $\widetilde{\mathbf{D}}_m^{1/2}\mathbf{P}\widetilde{\mathbf{D}}_n^{1/2}$. Let $s_k(\mathbf{Q})$ denote the kth largest singular value of a matrix $\mathbf{Q}$. By the Courant–Fischer–Weyl minimax principle (cf. [3], 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 $r = \operatorname{rank} \mathbf{B} = \operatorname{rank} \widetilde{\mathbf{D}}_m^{1/2}\mathbf{P}\widetilde{\mathbf{D}}_n^{1/2}$, it is sufficient to consider vectors $\mathbf{x}$, for which $\widetilde{\mathbf{D}}_m^{1/2}\mathbf{P}\widetilde{\mathbf{D}}_n^{1/2}\mathbf{x} \neq \mathbf{0}$. Therefore with $k \in \{1, \dots, r\}$ and an arbitrary k-dimensional subspace $H \subset \mathbb{R}^b$ one can write

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

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

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

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

Theorem 2.2. *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 Wigner-noise of the same size. Then under GC1 the matrix* $\mathbf{A}$ *almost surely has* r *singular values* $z_1, \dots, z_r$ *with*

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

and for the other singular values

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

hold almost surely.

Proof. The statement follows from the analog of the Weyl's perturbation theorem for singular values of rectangular matrices (see [3], p.99) and from Lemma 1.9. If $s_i(\mathbf{A})$ and $s_i(\mathbf{B})$ denote the ith singular values of the matrix in the argument in decreasing order then for the difference of the corresponding pairs

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

By Lemma 1.9,

$$\begin{aligned} &\mathbb{P}\left(|s_i(\mathbf{A}) - s_i(\mathbf{B})| > C_{K1} \cdot \sqrt{m+n}\right) \\ \le &\mathbb{P}\left(\, \|\mathbf{W}\| > C_{K1} \cdot \sqrt{m+n}\,\right) \le \exp[-C_{K2} \cdot (m+n)]. \end{aligned}$$

The right hand side of the last inequality is the general term of a convergent series (defined as a double summation), thus the second part of Remark 1.7 implies the almost sure statement of the theorem. □

Remark 2.3. A more precise estimation of the individual differences can be obtained in the following way. With any positive constant C

$$\begin{aligned} &\mathbb{P}\left(|z_i - s_i| > C\sqrt{m+n}\right) \le \mathbb{P}\left(|z_i - \mathbb{E}(z_i)| + |\mathbb{E}(z_i) - s_i| > C\sqrt{m+n}\right) \\ &= \mathbb{P}\left(|z_i - \mathbb{E}(z_i)| > C\sqrt{m+n} - |\mathbb{E}(z_i) - s_i|\right) \le \mathbb{P}\left(|z_i - \mathbb{E}(z_i)| > C_1(\sqrt{m+n}\right) \\ &\le \exp\left[\frac{(1-o(1))C_1^2(m+n)}{32i^2(K+1)^2}\right] =: p_{mn}, \qquad i = 1, \dots, r \end{aligned}$$

with some constant $C_1 \le C$. We used Lemma 1.8 and the fact that the difference between the constants $\mathbb{E}(z_i)$ and s_i is $\mathcal{O}(\sqrt{m+n})$. As $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} p_{mn} < \infty$, the Borel–Cantelli Lemma implies that $|z_i - s_i| = \mathcal{O}(\sqrt{m+n})$ holds almost surely for $i = 1, \dots, r$.

Corollary 2.4. With notations

$$\varepsilon := \|\mathbf{W}\| = \mathcal{O}(\sqrt{m+n}) \quad \text{and} \quad \Delta := \min_{1\le i\le r} s_i(\mathbf{B}) = \min_{1\le i\le r} s_i = \Theta(\sqrt{mn}) \qquad (2.5)$$

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.

3. Classification via singular vector pairs

With the help of Theorem 2.2 we can 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 $\mathbf{B}$,

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

Let us also denote the unit-norm, pairwise orthogonal left- and right-hand side singular vectors corresponding to the r protruding singular values $z_1, \dots, z_r$ of $\mathbf{A}$ 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. For them

$$\mathbf{A}\mathbf{x}_i = z_i \cdot \mathbf{y}_i \qquad (i = 1, \dots, r)$$

holds true. Let

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

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

Proposition 3.1. With the above notation, under *GC1* the following estimate holds almost surely for the sum of the squared distances between $\mathbf{y}_1, \dots, \mathbf{y}_r$ and F:

$$\sum_{i=1}^{r} \operatorname{dist}^2(\mathbf{y}_i, F) \le r \frac{\varepsilon^2}{(\Delta - \varepsilon)^2} = \mathcal{O}\left(\frac{m+n}{mn}\right), \tag{3.1}$$

and analogously, for the sum of the squared distances between $\mathbf{x}_1, \dots, \mathbf{x}_r$ and G:

$$\sum_{i=1}^{r} \operatorname{dist}^2(\mathbf{x}_i, G) \le r \frac{\varepsilon^2}{(\Delta - \varepsilon)^2} = \mathcal{O}\left(\frac{m+n}{mn}\right). \tag{3.2}$$

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, similarly 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

$$\mathbf{A}\mathbf{x} = (\mathbf{B} + \mathbf{W})\mathbf{x} = \sum_{i=1}^{r} t_i s_i \mathbf{v}_i + \mathbf{W}\mathbf{x}, \tag{3.3}$$

and on the other hand,

$$\mathbf{A}\mathbf{x} = z\mathbf{y} = \sum_{i=1}^{m} z l_i \mathbf{v}_i. \tag{3.4}$$

Equating the right-hand sides of (3.3) and (3.4) we obtain

$$\sum_{i=1}^{r}(zl_i - t_i s_i)\mathbf{v}_i + \sum_{i=r+1}^{m} zl_i \mathbf{v}_i = \mathbf{Wx}.$$

Applying the Pythagorean Theorem

$$\sum_{i=1}^{r}(zl_i - t_i s_i)^2 + z^2 \sum_{i=r+1}^{m} l_i^2 = \|\mathbf{Wx}\|^2 \le \varepsilon^2, \tag{3.5}$$

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

As $z \ge \Delta - \varepsilon$ holds almost surely by Theorem 2.2,

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

The order of the above estimate follows from the order of ε and Δ of (2.5):

$$\operatorname{dist}^2(\mathbf{y}, F) = \mathcal{O}(\frac{m+n}{mn}) \tag{3.6}$$

almost surely.

Applying (3.6) for the left-hand side singular vectors $\mathbf{y}_1, \dots, \mathbf{y}_r$, by the definition of *C2*

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

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

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

consequently,

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

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

$$\mathbf{A}^T\mathbf{y} = z \cdot \mathbf{x}$$

and using the fact that $\mathbf{A}^T$ has the same singular values as $\mathbf{A}$. □

Proposition 3.1 implies that 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 Wigner-noise $\mathbf{W}$ 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 3.1 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. Let the r-dimensional representatives of the genes be 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 be 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 [4], of the genes' representatives in the clustering $A_1, \dots, A_a$:

$$S_a^2(\mathbf{Y}) = \sum_{i=1}^{a} \sum_{j \in A_i} \|\mathbf{y}^j - \bar{\mathbf{y}}^i\|^2, \quad \text{where} \quad \bar{\mathbf{y}}^i = \frac{1}{m_i} \sum_{j \in A_i} \mathbf{y}^j, \tag{3.7}$$

while $S_b^2(\mathbf{X})$ denotes the b-variance of the conditions' representatives in the clustering $B_1, \dots, B_b$:

$$S_b^2(\mathbf{X}) = \sum_{i=1}^{b} \sum_{j \in B_i} \|\mathbf{x}^j - \bar{\mathbf{x}}^i\|^2, \quad \text{where} \quad \bar{\mathbf{x}}^i = \frac{1}{n_i} \sum_{j \in B_i} \mathbf{x}^j. \tag{3.8}$$

Theorem 3.2. *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 [4] it can be easily seen that $S_a^2(\mathbf{Y})$ and $S_b^2(\mathbf{X})$ is equal to the left-hand side of (3.1) and (3.2), respectively, therefore they are also of order $\mathcal{O}(\frac{m+n}{mn})$. □

Hence, the addition of any kind of a Wigner-noise to a rectangular matrix that has a blown up structure $\mathbf{B}$ will not change the order of the protruding 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 Wigner-noise, 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

$$w_{lk} := \begin{cases} 1 - p_{ij} & \text{with probability} \quad p_{ij} \\ -p_{ij} & \text{with probability} \quad 1 - p_{ij} \end{cases} \tag{3.9}$$

be independent random variables. This $\mathbf{W}$ satisfies the conditions of Definition 1.1 with entries uniformly bounded by 1, zero expectation and variance

$$\sigma^2 = \max_{1 \le i \le a;\, 1 \le j \le b} p_{ij}(1 - p_{ij}) \le \frac{1}{4}.$$

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 (3.9) of $\mathbf{W}$, the vertices of A_i are contained in edges of B_j with probability p_{ij} (set i of genes equally influences set j of conditions, like the chess-board pattern of [10]). It is a generalization of the classical Erdős–Rényi model for random hypergraphs and for several blocks, see [7]. The question, how such a chess-board pattern behind a random (especially 0-1) matrix can be found under specific conditions, is discussed in Section 5.

4. Perturbation results for correspondence matrices

Now the pattern matrix $\mathbf{P}$ contains arbitrary non-negative entries, 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 notations:

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

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

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

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

Further, set

$$\mathbf{B}_{corr} := \mathbf{D}_{Brow}^{-1/2} \mathbf{B} \mathbf{D}_{Bcol}^{-1/2} \quad \text{and} \quad \mathbf{A}_{corr} := \mathbf{D}_{Arow}^{-1/2} \mathbf{A} \mathbf{D}_{Acol}^{-1/2} \tag{4.1}$$

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

$$\mathbf{v}_{corr\,i} := \mathbf{D}_{Brow}^{-1/2} \mathbf{v}_i \quad \text{and} \quad \mathbf{u}_{corr\,i} := \mathbf{D}_{Bcol}^{-1/2} \mathbf{u}_i \tag{4.2}$$

the so-called correspondence vector pairs are obtained. If the coordinates $u_{corr\,i}(j)$, $v_{corr\,i}(j)$ 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 previous random variables $\beta_1, \dots, \beta_{i-1}$ and $\alpha_1, \dots, \alpha_{i-1}$, respectively ($i > 1$).

If $s_1 = 1$ is a simple 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 $k \le \operatorname{rank} \mathbf{B}_{corr} = \operatorname{rank} \mathbf{B} = \operatorname{rank} \mathbf{P}$ 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 underlying two discrete variables is obtained:

$$\mathbf{v}^j_{corr} := (v_{corr\,2}(j), \dots, v_{corr\,k}(j)) \quad \text{and} \quad \mathbf{u}^l_{corr} := (u_{corr\,2}(l), \dots, u_{corr\,k}(l)) \,.$$

This representation has the following optimality 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 4.1. 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 $r = \operatorname{rank} \mathbf{B} = \operatorname{rank} \mathbf{P}$.

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

$$\tilde{p}_{ij} = \frac{p_{ij}}{\sqrt{(\sum_{l=1}^b p_{il} n_l)(\sum_{k=1}^a p_{kj} m_k)}} .$$

Following the considerations of the proof of Proposition 2.1, the blown up matrix $\mathbf{B}_{corr}$ has exactly $r = \operatorname{rank} \mathbf{P} = \operatorname{rank} \tilde{\mathbf{P}}$ 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

$$p'_{ij} = \frac{p_{ij}\sqrt{m_i}\sqrt{n_j}}{\sqrt{(\sum_{l=1}^b p_{il} n_l)(\sum_{k=1}^a p_{kj} m_k)}} = \frac{p_{ij}}{\sqrt{(\sum_{l=1}^b p_{il} \frac{n_l}{n_j})(\sum_{k=1}^a p_{kj} \frac{m_k}{m_i})}} .$$

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 (1.1). (Here the compactness is understood in the topology induced by the spectral norm.) The range of the non-zero singular values depends continuously on the matrix that does not depend on m and n. Therefore, the minimum non-zero singular value does not depend on m or n. The largest singular value being 1, this finishes the proof. □

Theorem 4.2. *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

$$\mathbf{A}_{corr} = \mathbf{D}_{Arow}^{-1/2}\mathbf{A}\mathbf{D}_{Acol}^{-1/2} = \mathbf{D}_{Arow}^{-1/2}\mathbf{B}\mathbf{D}_{Acol}^{-1/2} + \mathbf{D}_{Arow}^{-1/2}\mathbf{W}\mathbf{D}_{Acol}^{-1/2}. \tag{4.3}$$

Observe, that 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. [5], Lemma 4.2):

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

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

$$\begin{aligned}
&\mathbb{P}\left(|d_{Arow\, i} - d_{Brow\, i}| > n^{1-\tau} \text{ for all } \quad i \in \{1, \dots, m\}\right) \\
&\le m \cdot \exp\left\{-\frac{n^{1-2\tau}}{2(\sigma^2 + Kn^{-\tau}/3)}\right\} \\
&\le C_0 \cdot n^C \cdot \exp\left\{-\frac{n^{1-2\tau}}{2(\sigma^2 + Kn^{-\tau}/3)}\right\} \\
&= \exp\left\{\ln C_0 + C\ln n - \frac{n^{1-2\tau}}{2(\sigma^2 + Kn^{-\tau}/3)}\right\}.
\end{aligned} \tag{4.4}$$

The estimation of probability

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

can be treated analogously (with $D_0 > 0$ and $D \ge 1$ of *GC2*). Now the second part of Remark 1.7 can be applied since the right-hand side of (4.4) forms a convergent series. Therefore

$$\begin{aligned}
\min_{i \in \{1,\dots,m\}} |d_{Arow\, i}| &= \Theta(n), \\
\min_{j \in \{1,\dots,n\}} |d_{Acol\, j}| &= \Theta(m)
\end{aligned} \tag{4.5}$$

hold almost surely.

Now it is straightforward to bound the norm of the second term of (4.3) by

$$\|\mathbf{D}_{Arow}^{-1/2}\| \cdot \|\mathbf{W}\| \cdot \|\mathbf{D}_{Acol}^{-1/2}\|. \tag{4.6}$$

As by Lemma 1.9 $\|\mathbf{W}\| = \mathcal{O}(\sqrt{m+n})$ holds almost surely, the quantity (4.6) is at most of order $\sqrt{\frac{m+n}{mn}}$ almost surely. Hence, it is almost surely less than $\max\{n^{-\tau}, m^{-\tau}\}$.

In order to estimate the norm of the first term of (4.3) let us write it in the form

$$\begin{aligned} \mathbf{D}_{Arow}^{-1/2}\mathbf{B}\mathbf{D}_{Acol}^{-1/2} &= \mathbf{D}_{Brow}^{-1/2}\mathbf{B}\mathbf{D}_{Bcol}^{-1/2}+ \\ &+ \left[\mathbf{D}_{Arow}^{-1/2} - \mathbf{D}_{Brow}^{-1/2}\right]\mathbf{B}\mathbf{D}_{Bcol}^{-1/2}+ \\ &+ \mathbf{D}_{Arow}^{-1/2}\mathbf{B}\left[\mathbf{D}_{Acol}^{-1/2} - \mathbf{D}_{Bcol}^{-1/2}\right]. \end{aligned} \tag{4.7}$$

The first term is just $\mathbf{B}_{corr}$, so – due to Proposition 4.1 – 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

$$\|\mathbf{D}_{Arow}^{-1/2} - \mathbf{D}_{Brow}^{-1/2}\| \cdot \|\mathbf{B}\| \cdot \|\mathbf{D}_{Bcol}^{-1/2}\|. \tag{4.8}$$

The estimation of the first factor in (4.8) is as follows:

$$\begin{aligned} \|\mathbf{D}_{Arow}^{-1/2} - \mathbf{D}_{Brow}^{-1/2}\| &= \max_{i\in\{1,\dots,m\}} \left(\frac{1}{\sqrt{d_{Arow\,i}}} - \frac{1}{\sqrt{d_{Brow\,i}}}\right) \\ &= \max_{i\in\{1,\dots,m\}} \frac{|d_{Arow\,i} - d_{Brow\,i}|}{\sqrt{d_{Arow\,i}\cdot d_{Brow\,i}}(\sqrt{d_{Arow\,i}} + \sqrt{d_{Brow\,i}})} \\ &\le \max_{i\in\{1,\dots,m\}} \frac{|d_{Arow\,i} - d_{Brow\,i}|}{\sqrt{d_{Arow\,i}\cdot d_{Brow\,i}}} \cdot \max_{i\in\{1,\dots,m\}} \frac{1}{(\sqrt{d_{Arow\,i}} + \sqrt{d_{Brow\,i}})}. \end{aligned} \tag{4.9}$$

By relations (4.5), $\sqrt{d_{Arow\,i}\cdot d_{Browi}} = \Theta(n)$ for any $i = 1,\dots,m$ almost surely, and hence,

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

holds 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 (4.9) can be estimated by $n^{-\tau-1/2}$ from above almost surely. For the further factors in (4.8) we obtain $\|\mathbf{B}\| = \Theta(\sqrt{mn})$ (see Proposition 2.1), 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 (4.3), and by he Weyl's perturbation theorem the proof, too. □

Remark 4.3. In the Gaussian case the large deviation principle can be replaced by the simple estimation of the Gaussian probabilities with any $\kappa > 0$:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} w_{ij}\right| > \kappa\right) < \min\left(1, \frac{4\sigma}{\kappa\sqrt{2\pi n}}\exp\left\{-\frac{n}{2\sigma^2}\kappa^2\right\}\right). \tag{4.10}$$

Setting $\kappa = n^{-\tau}$ we get an estimate, analogous to (4.4).

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

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

The correspondence vector pairs obtained by the transformations

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

contain coordinates of the discrete random variables β_i and α_i, respectively. Apart from the first (trivial) pair – with all 1 coordinates – they are of zero expectation and unit variance with respect to the marginal distributions

$$d_{Brow\,1}, \dots, d_{Brow\,m} \quad \text{and} \quad d_{Bcol\,1}, \dots, d_{Bcol\,n},$$

respectively. Further, the different α_i's and β_i's are uncorrelated with respect to the joint discrete distribution embodied by the entries of $\mathbf{B}$.

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 4.2 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 4.4. With the above notation, under *GC1* and *GC2* the following estimate holds almost surely for the distance between $\mathbf{y}_i$ and F:

$$\text{dist}(\mathbf{y}_i, F) \leq \frac{\epsilon}{(\delta - \epsilon)} = \frac{1}{(\frac{\delta}{\epsilon} - 1)} \qquad (i = 1, \dots, r) \tag{4.11}$$

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

$$\text{dist}(\mathbf{x}_i, G) \leq \frac{\epsilon}{(\delta - \epsilon)} = \frac{1}{(\frac{\delta}{\epsilon} - 1)} \qquad (i = 1, \dots, r). \tag{4.12}$$

Proof. Follow the method of proving Proposition 3.1 – under *GC1* – with δ instead of Δ and ϵ instead of ε! Here *GC2* is necessary only for $\mathbf{A}_{corr}$ to have r protruding singular values. □

Remark 4.5. The left-hand sides of (4.11) and (4.12) are almost surely of order $\max\{n^{-\tau}, m^{-\tau}\}$ that tend to zero, as $m, n \to \infty$ under *GC1* and *GC2*.

Proposition 4.4 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 $\mathbf{Y}_{corr}$ be the $m \times r$ matrix containing the left-hand side vectors $\mathbf{y}_{corr\,1}, \dots, \mathbf{y}_{corr\,r}$ in its columns. Similarly, let $\mathbf{X}_{corr}$ be the $n \times r$ matrix containing the right-hand side vectors $\mathbf{x}_{corr\,1}, \dots, \mathbf{x}_{corr\,r}$ in its columns. Let the r-dimensional representatives of α be the row vectors of $\mathbf{Y}_{corr}$: $\mathbf{y}^1_{corr}, \dots, \mathbf{y}^m_{corr} \in \mathbb{R}^r$, while the r-dimensional representatives of β be the row vectors of $\mathbf{X}_{corr}$: $\mathbf{x}^1_{corr}, \dots, \mathbf{x}^n_{corr} \in \mathbb{R}^r$. With respect to the marginal distributions, let the a- and b-variances of these representatives be defined by

$$S_a^2(\mathbf{Y}_{corr}) = \sum_{i=1}^{a} \sum_{j \in A_i} d_{Arow\,j} \|\mathbf{y}^j_{corr} - \bar{\mathbf{y}}^i_{corr}\|^2, \quad \text{where} \quad \bar{\mathbf{y}}^i_{corr} = \sum_{j \in A_i} d_{Arow\,j} \mathbf{y}^j_{corr},$$

while

$$S_b^2(\mathbf{X}_{corr}) = \sum_{i=1}^{b} \sum_{j \in B_i} d_{Acol\,j} \|\mathbf{x}^{(j)}_{corr} - \bar{\mathbf{x}}^i_{corr}\|^2, \quad \text{where} \quad \bar{\mathbf{x}}^i_{corr} = \sum_{j \in B_i} d_{Acol\,j} \mathbf{x}^j_{corr}.$$

Theorem 4.6. *With the above notation, under GC1 and GC2*

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

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

Proof. An easy calculation shows that

$$S_a^2(\mathbf{Y}_{corr}) = \sum_{i=1}^{r} \text{dist}^2(\mathbf{y}_i, F) \quad \text{and} \quad S_b^2(\mathbf{X}_{corr}) = \sum_{i=1}^{r} \text{dist}^2(\mathbf{x}_i, G),$$

hence the result of Proposition 4.4 can be used. □

Under *GC1* and *GC2* with m, n large enough, Theorem 4.6 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 $\mathbf{A}$.

5. 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 $a\sqrt{mn}$, in 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 Wigner-noise.

So 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 3.2 (cf. the introduction to this theorem). Under these conditions we can reconstruct a blown up structure behind our matrix.

Theorem 5.1. *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 $a \geq k$ and $b \geq k$ such that the a- and b-variances of the row- and column-representatives are $\mathcal{O}(\frac{m+n}{mn})$, then there is 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. The proof gives an explicit construction for $\mathbf{B}_{m\times n}$. 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, respectively in $\mathbb{R}^k$ with sum of inner variances $\mathcal{O}(\frac{m+n}{mn})$. Reorder the rows and columns of $\mathbf{A}$ according to the clusters. 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 be row vectors of the $m \times k$ matrix $\widetilde{\mathbf{Y}}$ such that the first m_1 rows of $\widetilde{\mathbf{Y}}$ be 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 $\widetilde{\mathbf{Y}}$ be equal to $\bar{\mathbf{y}}^a$; similarly, the conditions' representatives be row vectors of the $n \times k$ matrix $\widetilde{\mathbf{X}}$ such that the first n_1 rows of $\widetilde{\mathbf{X}}$ be equal to $\bar{\mathbf{x}}^1$, and so on, the last n_b rows of $\widetilde{\mathbf{X}}$ be equal to $\bar{\mathbf{x}}^b$.

By the considerations of Theorem 3.2 and the assumption for the clusters

$$\sum_{i=1}^{k} \operatorname{dist}^2(\mathbf{y}_i, F) = S_a^2(\mathbf{Y}) = \mathcal{O}(\frac{m+n}{mn}) \tag{5.1}$$

and

$$\sum_{i=1}^{k} \operatorname{dist}^2(\mathbf{x}_i, G) = S_b^2(\mathbf{X}) = \mathcal{O}(\frac{m+n}{mn}) \tag{5.2}$$

hold respectively, where the k-dimensional subspace $F \subset \mathbb{R}^m$ is spanned by the column vectors of $\widetilde{\mathbf{Y}}$, while the k-dimensional subspace $G \subset \mathbb{R}^n$ is spanned by the column vectors of $\widetilde{\mathbf{X}}$. We follow the construction given in [4] (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

$$\sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{v}_i\|^2 = \min_{\substack{\mathbf{v}'_1,\dots,\mathbf{v}'_k \\ \mathbf{v}'^T_i \mathbf{v}'_j = \delta_{ij}}} \sum_{i=1}^{k} \|\mathbf{y}_i - \mathbf{v}'_i\|^2 \le 2 \sum_{i=1}^{k} \operatorname{dist}^2(\mathbf{y}_i, F) \tag{5.3}$$

and

$$\sum_{i=1}^{k} \|\mathbf{x}_i - \mathbf{u}_i\|^2 = \min_{\substack{\mathbf{u}'_1,\dots,\mathbf{u}'_k \\ \mathbf{u}'^T_i \mathbf{u}'_j = \delta_{ij}}} \sum_{i=1}^{k} \|\mathbf{x}_i - \mathbf{u}'_i\|^2 \le 2 \sum_{i=1}^{k} \operatorname{dist}^2(\mathbf{x}_i, G) \tag{5.4}$$

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

$$\mathbf{Y}^T \mathbf{V}' = \mathbf{Q}\mathbf{S}\mathbf{Z}^T$$

be SVD, where the matrix $\mathbf{S}$ contains the singular values of the $k \times k$ matrix $\mathbf{Y}^T\mathbf{V}'$ 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 $\mathbf{R} = \mathbf{Z}\mathbf{Q}^T$ 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:

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

In view of (5.1) – (5.4)

$$\sum_{i=1}^{k} \|\mathbf{r}_i\|^2 = \mathcal{O}(\frac{m+n}{mn}) \quad \text{and} \quad \sum_{i=1}^{k} \|\mathbf{q}_i\|^2 = \mathcal{O}(\frac{m+n}{mn}) \tag{5.5}$$

hold.

Consider the following decomposition:

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

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

$$\sum_{i=1}^{k} z_i \mathbf{y}_i \mathbf{x}_i^T = \sum_{i=1}^{k} z_i (\mathbf{v}_i + \mathbf{r}_i)(\mathbf{u}_i^T + \mathbf{q}_i^T) = \\ = \sum_{i=1}^{k} z_i \mathbf{v}_i \mathbf{u}_i^T + \sum_{i=1}^{k} z_i \mathbf{v}_i \mathbf{q}_i^T + \\ + \sum_{i=1}^{k} z_i \mathbf{r}_i \mathbf{u}_i^T + \sum_{i=1}^{k} z_i \mathbf{r}_i \mathbf{q}_i^T . \tag{5.6}$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_k$ and $\mathbf{u}_1, \dots, \mathbf{u}_k$ are unit vectors, the last three terms in (5.6) can be estimated by means of the relations

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

Taking into account that z_i cannot exceed $\Theta(\sqrt{mn})$ and k is fixed, due to (5.5) we get that the spectral norms of the last three terms in (5.6) – 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.,

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

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

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

that finishes the proof. □

Then, provided the conditions of Theorem 5.1 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 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 after we use the above construction. If we decide to perform correspondence analysis on $\mathbf{A}$ then by (4.3) and (4.7), $\mathbf{B}_{corr}$ 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 obtain SVD of large matrices, randomized algorithms are at our disposal, e.g., [1]. There is nothing to loose when applying these algorithms because they give the required results only if our matrix had a primary linear structure.

Acknowledgements

The authors are indebted to Gábor Tusnády for suggesting the microarray problem and supplying computer simulations, and also to Péter Major for inspiring discussions on Chernoff-type inequalities.

References

[1] Achlioptas, D., McSherry, F., Fast computation of low rank matrix approximations, in: *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, 611-618, ACM, New York (2001).

[2] Alon, N., Krivelevich, M., Vu, V. H., On the concentration of eigenvalues of random symmetric matrices, *Israel Journal of Mathematics* 131 (2002) 259-267.

[3] Bhatia, R., Matrix Analysis, Graduate Texts in Mathematics 169, Springer, New York (1996).

[4] Bolla M., Distribution of the eigenvalues of random block-matrices, *Linear Algebra and its Applications* 377 (2004) 219-240.

[5] Bolla, M., Recognizing linear structure in noisy matrices, *Linear Algebra and its Applications* 402 (2005) 228-244.

[6] Bolla, M., Krámli, A., Statisztikai következtetések elmélete (Theory of Statistical Inference), Graduate Textbook in Hungarian, Typotex, Budapest (2005).

[7] Bollobás, B., Random Graphs, second ed., Cambridge University Press, Cambridge (2001).

[8] Füredi, Z., Komlós, J., The eigenvalues of random symmetric matrices, *Combinatorica* 1 (3) (1981) 233-241.

[9] Golub, G., Van Loan, C., Matrix Computations, Johns Hopkins University Press (1989).

[10] Kluger, Y., Basri, R., Chang, J. T., Gerstein, M., Spectral biclustering of microarray data: coclustering genes and conditions, *Genome Rerearch* 13 (2003) 703-716.

[11] Mehta, M. L., Random Matrices, Academic Press, London (1991).

[12] Olkin, I., The 70th anniversary of the distribution of random matrices: A survey, *Linear Algebra and its Applications* 354 (2002) 231-243.

[13] Tracy, C. A., Widom, H., Level spacing distributions and the Airy kernel, *Comm. Math. Phys.* 159 (1994) 151-174.