Elect. Comm. in Probab. **14** (2009), 334–342

ELECTRONIC COMMUNICATIONS in PROBABILITY

CONCENTRATION OF THE SPECTRAL MEASURE OF LARGE WISHART MATRICES WITH DEPENDENT ENTRIES

ADITYANAND GUNTUBOYINA
Department of Statistics, Yale University, P.O. Box 208290, New Haven, CT 06520, USA
email: `adityanand.guntuboyina@yale.edu`

HANNES LEEB
Department of Statistics, Yale University, P.O. Box 208290, New Haven, CT 06520, USA
email: `hannes.leeb@yale.edu`

Submitted October 15, 2008, *accepted in final form* June 30, 2009

AMS 2000 Subject classification: AMS 2000 Subject Classification: Primary 15A52 Secondary 60F10, 15A18
Keywords: Wishart matrices, concentration inequalities, spectral measure

Abstract

We derive concentration inequalities for the spectral measure of large random matrices, allowing for certain forms of dependence. Our main focus is on empirical covariance (Wishart) matrices, but general symmetric random matrices are also considered.

1 Introduction

In this short paper, we study concentration of the spectral measure of large random matrices whose elements need not be independent. In particular, we derive a concentration inequality for Wishart matrices of the form $X'X/m$ in the important setting where the rows of the $m \times n$ matrix X are independent but the elements within each row may depend on each other; see Theorem 1. We also obtain similar results for other random matrices with dependent entries; see Theorem 5, Theorem 6, and the attending examples, which include a random graph with dependent edges, and vector time series.

Large random matrices have been the focus of intense research in recent years; see Bai [3] and Guionnet [8] for surveys. While most of this literature deals with the case where the underlying matrix has independent entries, comparatively little is known for dependent cases. Götze and Tikhomirov [7] showed that the expected spectral distribution of an empirical covariance matrix $X'X/m$ converges to the Marčenko-Pastur law under conditions that allow for some form of dependence among the entries of X. Bai and Zhou [2] analyzed the limiting spectral distribution of $X'X/m$ when the row-vectors of X are independent (allowing for certain forms of dependence within the row-vectors of X). Mendelson and Pajor [16] considered $X'X/m$ in the case where the row-vectors of X are independent and identically distributed (i.i.d.); under some additional assumptions, they derive a concentration result for the operator norm of $X'X/m - E(X'X/m)$.

Boutet de Monvel and Khorunzhy [5] studied the limiting behavior of the spectral distribution and of the operator norm of symmetric Gaussian matrices with dependent entries.
For large random matrices similar to those considered here, concentration of the spectral measure was also studied by Guionnet and Zeitouni [9], who considered Wishart matrices $X'X/m$ where the entries $X_{i,j}$ of X are independent, as well as Hermitian matrices with independent entries on and above the diagonal, and by Houdre and Xu [11], who obtained concentration results for random matrices with stable entries, thus allowing for certain forms of dependence. For matrices with dependent entries, we find that concentration of the spectral measure can be less pronounced than in the independent case. Technically, our results rely on Talagrand's inequality [17] and on the Azuma/Hoeffding/McDiarmid bounded difference inequality [1, 10, 15].

2 Results

Throughout, the eigenvalues of a symmetric $n \times n$ matrix M are denoted by $\lambda_1(M) \leq \cdots \leq \lambda_n(M)$, and we write $F_M(\lambda)$ for the cumulative distribution function (c.d.f.) of the spectral distribution of M, i.e., $F_M(\lambda) = n^{-1}\sum_{i=1}^n 1\{\lambda_i(M) \leq \lambda\}$, $\lambda \in \mathbb{R}$. The integral of a function $f(\cdot)$ with respect to the measure induced by F_M is denoted by $F_M(f)$, i.e.,

$$F_M(f) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i(M)).$$

For certain classes of random matrices M and certain classes of functions f, we will show that $F_M(f)$ is concentrated around its expectation $\mathbb{E}F_M(f)$ or around any median med $F_M(f)$. For a Lipschitz function g, we write $||g||_L$ for its Lipschitz constant. Moreover, we also consider functions $f : (a,b) \to \mathbb{R}$ that are of bounded variation on (a,b) (where $-\infty \leq a < b \leq \infty$), in the sense that

$$V_f(a,b) \quad = \quad \sup_{n\geq 1} \sup_{a<x_0\leq x_1\leq\cdots\leq x_n<b} \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|$$

is finite; cf. Section X.1 in [12]. [A function f is of bounded variation on (a,b) if and only if it can be written as the difference of two bounded monotone functions on (a,b). Note that the indicator function $g : x \mapsto 1\{x \leq \lambda\}$ is of bounded variation on $\mathbb{R}$ with $V_g(\mathbb{R}) = 1$ for each $\lambda \in \mathbb{R}$.]
The following result establishes concentration of $F_S(f)$ for Wishart matrices S of the form $S = X'X/m$ where we only require that the rows of X are independent (while allowing for dependence *within* each row of X). See also Example 9 and Example 10, which follow, for scenarios that also allow for some dependence among the rows of X.

Theorem 1. *Let X be an $m \times n$ matrix whose row-vectors are independent, set $S = X'X/m$, and fix $f : \mathbb{R} \to \mathbb{R}$.*

(i) Suppose that f is such that the mapping $x \mapsto f(x^2)$ is convex and Lipschitz, and suppose that $|X_{i,j}| \leq 1$ for each i and j. For each $\epsilon > 0$, we then have

$$\mathbb{P}\left(\left|F_S(f) - med\ F_S(f)\right| \geq \epsilon\right) \quad \leq \quad 4\exp\left[-\frac{nm}{n+m}\frac{\epsilon^2}{8||f(\cdot^2)||_L^2}\right]. \tag{1}$$

(ii) Suppose that f is of bounded variation on $\mathbb{R}$. For each $\epsilon > 0$, we then have

$$\mathbb{P}\left(\left|F_S(f) - \mathbb{E}F_S(f)\right| \geq \epsilon\right) \quad \leq \quad 2\exp\left[-\frac{n^2}{m}\frac{2\epsilon^2}{V_f^2(\mathbb{R})}\right]. \tag{2}$$

In particular, for each $\lambda \in \mathbb{R}$ and each $\epsilon > 0$, the probability $\mathbb{P}(|F_S(\lambda) - \mathbb{E}F_S(\lambda)| \geq \epsilon)$ is bounded by the right-hand side of (2) *with $V_f(\mathbb{R})$ replaced by* 1.

Remark 2. *From the upper bound* (1) *one can also obtain a similar bound for $\mathbb{P}(|F_S(f) - \mathbb{E}F_S(f)| \geq \epsilon)$ using standard methods.*

The upper bounds in Theorem 1 are of the form

$$\mathbb{P}(|F_S(f) - A| \geq \epsilon) \quad \leq \quad B \exp\left[-nC\right], \tag{3}$$

where A, B, and C equal $\operatorname{med} F_S(f)$, 4, and $m\epsilon^2/\left((n+m)8||f(\cdot^2)||_L^2\right)$ in part (i) and $\mathbb{E}F_S(f)$, 2, and $n2\epsilon^2/(mV_f^2)$ in part (ii), respectively. For the interesting case where n and m both go to infinity at the same rate, the next example shows that these bounds can not be improved qualitatively without imposing additional assumptions.

Example 3. Let $n = m = 2^k$, and let X be the $n \times n$ matrix whose i-th row is $R_i v_i'$, where $R_1, \ldots, R_n$ are i.i.d. with $\mathbb{P}(R_1 = 0) = \mathbb{P}(R_1 = 1) = 1/2$, and where $v_1, \ldots, v_n$ are orthogonal n-vectors with $v_i \in \{-1, 1\}^n$ for each i. [The v_i's can be obtained, say, from the first n binary Walsh functions; cf. [18].] Note that the eigenvalues of $S = X'X/m$ are $R_1^2, \ldots, R_n^2$. Set $f(x) = x$ for $x \in \{0, 1\}$. Then $nF_S(f)$ is binomial distributed with parameters n and $1/2$, i.e., $nF_S(f) \sim B(n, 1/2)$. By Chernoff's method (cf. Theorem 1 of [6]), we hence obtain that

$$\mathbb{P}(F_S(f) - \mathbb{E}F_S(f) \geq \epsilon) \quad = \quad \exp\left[-n(C(\epsilon) + o(1))\right], \tag{4}$$

for $0 < \epsilon < 1/2$ and as $n \to \infty$ with $k \to \infty$, where here $C(\epsilon)$ equals $\log(2) + (1/2 + \epsilon)\log(1/2 + \epsilon) + (1/2 - \epsilon)\log(1/2 - \epsilon)$; the same is true if $\mathbb{E}F_S(f) - F_S(f)$ replaces $F_S(f) - \mathbb{E}F_S(f)$ in (4). These statements continue to hold with $\operatorname{med} F_S(f)$ replacing $\mathbb{E}F_S(f)$, because the mean coincides with the median here. To apply Theorem 1(i), we extend f by setting $f(x) = \sqrt{|x|}$ for $x \in \mathbb{R}$; to apply Theorem 1(ii), extend f as $f(x) = 1\{x \leq 1/2\}$. Theorem 1(i) and Theorem 1(ii) give us that the left hand side of (4) is bounded by terms of the form $4\exp\left[-nC_1(\epsilon)\right]$ and $2\exp\left[-nC_2(\epsilon)\right]$, respectively, for some functions C_1 and C_2 of ϵ. It is easy to check that $C(\epsilon)/C_i(\epsilon)$ is increasing in ϵ for $i = 1, 2$, and that

$$\lim_{\epsilon \downarrow 0} \frac{C(\epsilon)}{C_1(\epsilon)} = 32 \qquad \text{and} \qquad \lim_{\epsilon \downarrow 0} \frac{C(\epsilon)}{C_2(\epsilon)} = 1.$$

In this example, both parts of Theorem 1 give upper bounds with the correct rate $(-n)$ in the exponent. The constants $C_i(\epsilon)$, $i = 1, 2$, both are sub-optimal, i.e., they are too small, but the constant $C_2(\epsilon)$, which is obtained from Theorem 1(ii), is close to the optimal constant for small ϵ.

Under additional assumptions on the law of X, $F_S(f)$ can concentrate faster than indicated by (3). In particular, in the setting of Theorem 1(i) and for the case where all the elements $X_{i,j}$ of X are independent, Guionnet and Zeitouni [9] obtained bounds of the same form as (3) but with n^2 replacing n in the exponent, for functions f such that $x \mapsto f(x^2)$ is convex and Lipschitz. (This should be compared with Example 10 below.) However, if f does not satisfy this requirement, but is of bounded variation on $\mathbb{R}$ so that Theorem 1(ii) applies, then the upper bound in (2) can not be improved qualitatively without additional assumptions, even in the case when all the elements $X_{i,j}$ of X are independent. This is demonstrated by the following example.

Example 4. Let X be the $n \times n$ diagonal matrix $\operatorname{diag}(R_1, \ldots, R_n)$, where $R_1, \ldots, R_n$ are as in Example 3. Set $f(x) = 1\{x \leq 0\}$. Clearly, Theorem 1(ii) applies here so that the left hand side (2) is bounded by $2\exp\left[-nC_2(\epsilon)\right]$ for $C_2(\epsilon)$ as in Example 3. Moreover, since for each i, $f(R_i^2/n) = 1 - R_i$, it follows that $nF_S(f) \sim B(n, 1/2)$, and then (4) holds again.

Theorem 1 can also be used to get concentration inequalities for the empirical distribution of the singular values of a non-symmetric $m \times n$ matrix X with independent rows. Indeed, the i-th singular value of $X/\sqrt{m}$ is just the square root of the i-th eigenvalue of $X'X/m$.
Both parts of Theorem 1 are in fact special cases of more general results that are presented next. The following two theorems, the first of which should be compared with Theorem 1.1(a) of [9], apply to a variety of random matrices besides those considered in Theorem 1; some examples are given later in this section. In the following theorem, we view symmetric $n \times n$ matrices as elements of $\mathbb{R}^{n(n+1)/2}$ by collecting the entries on and above the diagonal.

Theorem 5. *Let M be a random symmetric $n \times n$ matrix that is a function of m independent $[-1,1]^p$-valued random vectors $Y_1, \ldots, Y_m$ i.e., $M = M(Y_1, \ldots, Y_m)$. Assume that $M(\cdot)$ is linear and Lipschitz with Lipschitz constant C_M when considered as a function from $[-1,1]^{mp}$ with the Euclidean norm to the set of all symmetric $n \times n$ matrices with the Euclidean norm on $\mathbb{R}^{n(n+1)/2}$. Finally, assume that $f : \mathbb{R} \to \mathbb{R}$ is convex and Lipschitz with Lipschitz constant $||f||_L$. For $S = M/\sqrt{m}$, we then have*

$$\mathbb{P}\left(\left|F_S(f) - med\, F_S(f)\right| \geq \epsilon\right) \quad \leq \quad 4\exp\left[-\frac{nm}{p}\frac{\epsilon^2}{32C_M^2||f||_L^2}\right] \tag{5}$$

for each $\epsilon > 0$.

Theorem 6. *Let M be a random symmetric $n \times n$ matrix that is a function of m independent random quantities $Y_1, \ldots, Y_m$, i.e., $M = M(Y_1, \ldots, Y_m)$. Write $M_{(i)}$ for the matrix obtained from M after replacing Y_i by an independent copy, i.e., $M_{(i)} = M(Y_1, \ldots, Y_{i-1}, Y_i^*, Y_{i+1}, \ldots, Y_m)$ where Y_i^* is distributed as Y_i and independent of $Y_1, \ldots, Y_m$ ($i = 1, \ldots, m$). For $S = M/\sqrt{m}$ and $S_{(i)} = M_{(i)}/\sqrt{m}$, assume that*

$$||F_S - F_{S_{(i)}}||_\infty \quad \leq \quad r/n \tag{6}$$

holds (almost surely) for each $i = 1, \ldots, m$ and for some (fixed) integer r. Finally, assume that $f : \mathbb{R} \to \mathbb{R}$ is of bounded variation on $\mathbb{R}$. For each $\epsilon > 0$, we then have

$$\mathbb{P}\left(\left|F_S(f) - \mathbb{E}F_S(f)\right| \geq \epsilon\right) \quad \leq \quad 2\exp\left[-\frac{n^2}{m}\frac{2\epsilon^2}{r^2 V_f^2(\mathbb{R})}\right]. \tag{7}$$

Also, if a and b, $-\infty \leq a < b \leq \infty$, are such that $\mathbb{P}\left(a < \lambda_1(S) \text{ and } \lambda_n(S) < b\right) = 1$, then (7) holds for each function $f : (a,b) \to \mathbb{R}$ of bounded variation on (a,b), where now $V_f(a,b)$ replaces $V_f(\mathbb{R})$ on the right hand side of (7).

To apply Theorem 6, one needs to establish the inequality in (6) for each $i = 1, \ldots, m$. This can often be accomplished by using the following lemma, which is taken from Bai [3], Lemma 2.2 and 2.6, and which is a simple consequence of the interlacing theorem. [Consider a symmetric $n \times n$ matrix A and denote its $(n-1) \times (n-1)$ major submatrix by B. The interlacing theorem, a direct consequence of the Courant-Fisher formula, states that $\lambda_i(A) \leq \lambda_i(B) \leq \lambda_{i+1}(A)$ for $i = 1, \ldots, n-1$.]

Lemma 7. *Let A and B be symmetric $n \times n$ matrices and let X and Y be $m \times n$ matrices. Then the following inequalities hold:*

$$||F_A - F_B||_\infty \quad \leq \quad \frac{rank(A-B)}{n},$$

and

$$||F_{X'X} - F_{Y'Y}||_\infty \quad \leq \quad \frac{rank(X-Y)}{n}.$$

We now give some examples where Theorem 5 or Theorem 6 can be applied, the latter with the help of Lemma 7.

Example 8. Consider a network of, say, social connections or relations between a group of n entities that enter the group sequentially and that establish connections to group members that entered before as follows: For the i-th entity that enters the group, connections to the existing group members, labeled $1, \ldots, i-1$, are chosen according to some probability distribution, independently of the choices made by all the other entities. Denote the $n \times n$ adjacency matrix of the resulting random graph by M, and write Y_i for the n-vector $(M_{i,1}, M_{i,2}, \ldots, M_{i,i}, 0, \ldots, 0)'$ for $i = 1, \ldots, n$. By construction, $Y_1, \ldots, Y_n$ are independent and M (when considered as a function of $Y_1, \ldots, Y_n$ as in Theorem 5) is linear and Lipschitz with Lipschitz constant 1. Hence Theorem 5 is applicable with $m = p = n$ and $C_M = 1$.
Theorem 6 can also be applied here. To check condition (6), write $M_{(i)}$ for the matrix obtained from M by replacing Y_i by an independent copy denoted by Y_i^* as in Theorem 6. Clearly, the i-th row of the matrix $M - M_{(i)}$ equals $\delta_i = (Y_{i,1} - Y_{i,1}^*, \ldots Y_{i,i} - Y_{i,i}^*, 0, \ldots, 0)$, the i-th column of $M - M_{(i)}$ equals δ_i', and the remaining elements of $M - M_{(i)}$ all equal zero. Therefore, the rank of $M - M_{(i)}$ is at most two. Using Lemma 7, we see that Theorem 6 is applicable here with $r = 2$ and $m = n$.

The following two examples deal with the sample covariance matrix of vector moving average (MA) processes. For the sake of simplicity, we only consider MA processes of order 1. Our arguments can be extended to also handle MA processes of any fixed and finite order. In Example 9, we consider an MA(1) process with independent innovations, allowing for arbitrary dependence within each innovation, and obtain concentration inequalities of the form (3). In Example 10, we consider the case where each innovation has independent components (up to a linear function) and obtain a concentration inequality of the form (3) but with n^2 replacing n in the exponent.

Example 9. Consider an $m \times n$ matrix X whose row-vectors follow a vector MA process of order 1 i.e., $(X_{i,\cdot})' = Y_{i+1} + BY_i$ for $i = 1 \ldots m$, where $Y_1, \ldots Y_{m+1}$ are $m+1$ independent n-vectors and B is some fixed $n \times n$ matrix. Set $S = X'X/m$.

(i) Suppose that f is such that the mapping $x \mapsto f(x^2)$ is convex and Lipschitz, and suppose that $Y_i \in [-1, 1]^n$ for each $i = 1, \ldots, m+1$. For each $\epsilon > 0$, we have

$$\mathbb{P}\left(\left|F_S(f) - \text{med } F_S(f)\right| \geq \epsilon\right) \quad \leq \quad 4 \exp\left[-\frac{nm}{n+m} \frac{\epsilon^2}{8C_B^2 \|f(\cdot^2)\|_L^2}\right]. \tag{8}$$

Here C_B equals $1 + \|B\|$, where $\|B\|$ is the operator norm of the matrix B.

(ii) Suppose that f is of bounded variation on $\mathbb{R}$. For each $\epsilon > 0$, we then have

$$\mathbb{P}\left(\left|F_S(f) - \mathbb{E}F_S(f)\right| \geq \epsilon\right) \quad \leq \quad 2 \exp\left[-\frac{n^2}{m+1} \frac{\epsilon^2}{2V_f^2(\mathbb{R})}\right]. \tag{9}$$

The proofs of (8) and (9) follow essentially the same argument as used in the proof of Theorem 1 using the particular structure of the matrix X as considered here.

Example 10. As in Example 9, consider an $m \times n$ matrix X whose row-vectors follow a vector MA(1) process $(X_{i,\cdot})' = Y_{i+1} + BY_i$ for some fixed $n \times n$ matrix B, $i = 1, \ldots, m$. For the innovations Y_i, we now assume that $Y_i = UZ_i$, where U is a fixed $n \times n$ matrix, and where the $Z_{i,j}$, $i = 1, \ldots, m+1$,

$j = 1, \ldots, n$, are independent and satisfy $|Z_{i,j}| \leq 1$. Set $S = X'X/m$. For a function f such that the mapping $x \mapsto f(x^2)$ is convex and Lipschitz, we then obtain that

$$\mathbb{P}\left(\left|F_S(f) - \text{med } F_S(f)\right| \geq \epsilon\right) \quad \leq \quad 4\exp\left[-\frac{n^2 m}{n+m}\frac{\epsilon^2}{8C^2||f(\cdot^2)||_L^2}\right] \tag{10}$$

for each $\epsilon > 0$, where C is shorthand for $C = (1 + ||B||)||U||$ with $||B||$ and $||U||$ denoting the operator norms of the indicated matrices. The relation (10) is derived by essentially repeating the proof of Theorem 1(i) and by employing the particular structure of the matrix X as considered here.
We note that the statement in the previous paragraph reduces to Corollary 1.8(a) in [9] if one sets B to the zero matrix and U to the identity matrix. Moreover, we note that Theorem 6 can also be applied here (similarly to Example 9(ii)), but the resulting upper bound does not improve upon (9).

A Proofs

We first prove Theorem 5 and Theorem 6 and then use these results to deduce Theorem 1. The proof of Theorem 5 is modeled after the proof of Theorem 1.1(a) in Guionnet and Zeitouni [9]. It rests on a version of Talagrand's inequality (see Talagrand [17] and Maurey [14]) that is given as Theorem 11 below, and also on Lemma 1.2 from Guionnet and Zeitouni [9] that is restated as Lemma 12, which follows.

Theorem 11. *Fix $m \geq 1$ and $p \geq 1$. Consider a function $T : [-1,1]^{mp} \to \mathbb{R}$ that is quasi-convex[1] and Lipschitz with Lipschitz constant σ. Let $Y_1, \ldots, Y_m$ be independent p-vectors, each taking values in $[-1,1]^p$ and consider the random variable $T = T(Y_1, \ldots, Y_m)$. For each $\epsilon > 0$, we then have*

$$\mathbb{P}(|T - \mathit{med}\ T| \geq \epsilon) \leq 4\exp\left(-\frac{1}{p\sigma^2}\frac{\epsilon^2}{16}\right). \tag{11}$$

The result is derived from Theorem 6.1 of [17] by arguing just like in the proof of Theorem 6.6 of [17], but now using $[-1,1]^p$ instead of $[-1,1]$. (When $p = 1$, Theorem 11 reduces to Theorem 6.6 of [17].) Alternatively, it also follows from Theorem 3 of [14] using standard arguments.

Lemma 12. *Let $\mathscr{A}^n$ denote the set of all real symmetric $n \times n$ matrices and let $u : \mathbb{R} \to \mathbb{R}$ be a fixed function. Let us denote by Λ_u^n the functional $A \mapsto F_A(u)$ on $\mathscr{A}^n$. Then*

(i) If u is convex, then so is Λ_u^n.

(ii) If u is Lipschitz, then so is Λ_u^n (when considering $\mathscr{A}^n$ with the Euclidean norm on $\mathbb{R}^{n(n+1)/2}$ by collecting the entries on and above the diagonal). Moreover, the Lipschitz constant of Λ_u^n satisfies

$$||\Lambda_u^n||_L \leq \frac{\sqrt{2}}{\sqrt{n}}||u||_L.$$

Remark 13. *For a proof of this lemma, see Guionnet and Zeitouni [9, Proof of Lemma 1.2]. A simpler proof (along with other similar results) of Lemma 12(i) can be found in Lieb and Pedersen [13]. See also [9] and [13] for earlier references for Lemma 12(i).*

[1] A real valued function T is said to be quasi-convex if all the level sets $\{T \leq a\}$, $a \in \mathbb{R}$, are convex.

Proof of Theorem 5. Set $T = F_S(f)$ and let $\mathscr{A}^n$ be as in Lemma 12. In view of Theorem 11, it suffices to show that $T = T(Y_1, \ldots, Y_m)$ is such that the function $T(\cdot)$ is quasi-convex and Lipschitz with Lipschitz constant $\leq (2/(nm))^{1/2} C_M ||f||_L$. To this end, we write T as the composition $T_2 \circ T_1$, where $T_1 : ([-1,1]^p)^m \to \mathscr{A}^n$ and $T_2 : \mathscr{A}^n \to \mathbb{R}$ denote the mappings $(y_1, \ldots, y_m) \mapsto M(y_1, \ldots, y_m)/\sqrt{m}$ and $A \mapsto F_A(f)$, respectively. By assumption, T_1 is linear and Lipschitz with $||T_1||_L = C_M/\sqrt{m}$. Also, since f is assumed to be convex and Lipschitz, Lemma 12 entails that T_2 is convex and Lipschitz with $||T_2||_L \leq (2/n)^{1/2} ||f||_L$. It follows that T is convex (and hence quasi-convex) and Lipschitz with $||T||_L \leq (2/(nm))^{1/2} C_M ||f||_L$. The proof is complete. □

To prove Theorem 6, we use the Azuma/Hoeffding/McDiarmid bounded difference inequality. The following version of this inequality taken from Proposition 12 in [4]:

Proposition 14. *Consider independent random quantities $Y_1, \ldots, Y_m$, and let $Z = f(Y_1, \ldots, Y_m)$ where f is a Borel measurable function. For each $i = 1, \ldots, m$, define $Z_{(i)}$ like Z, but with Y_i replaced by an independent copy; that is, $Z_{(i)} = f(Y_1, \ldots, Y_{i-1}, Y_i^*, Y_{i+1}, \ldots, Y_m)$, where Y_i^* is distributed as Y_i and independent of $Y_1, \ldots, Y_m$. If*

$$\left| Z - Z_{(i)} \right| \leq c_i$$

holds (almost surely) for each $i = 1, \ldots, m$, then, for each $\epsilon > 0$, both $\mathbb{P}(Z - \mathbb{E}Z \geq \epsilon)$ and $\mathbb{P}(Z - \mathbb{E}Z \leq -\epsilon)$ are bounded by $\exp\left[-2\epsilon^2 / \sum_{i=1}^m c_i^2\right]$.

Proof of Theorem 6. It suffices to prove the second claim. Hence assume that a and b, $-\infty \leq a < b \leq \infty$ are such that $\mathbb{P}\left(a < \lambda_1(S) \text{ and } \lambda_n(S) < b\right) = 1$ and that $f : (a,b) \to \mathbb{R}$ is of bounded variation on (a,b). We shall now show that

$$|F_S(f) - F_{S_{(i)}}(f)| \leq r V_f(a,b)/n \text{ for each } i = 1, \ldots, m. \tag{12}$$

With this, we can use the bounded difference inequality, i.e., Proposition 14, with Z, $Z_{(i)}$, and c_i $(1 \leq i \leq m)$ replaced by $F_S(f)$, $F_{S_{(i)}}(f)$, and $rV_f(a,b)/n$, respectively, to obtain (7), completing the proof.
To obtain (12), set $G(\lambda) = F_S(\lambda) - F_{S_{(i)}}(\lambda)$ and choose α and β satisfying $a < \alpha < \min\{\lambda_1(S), \lambda_1(S_{(i)})\}$ and $b > \beta > \max\{\lambda_n(S), \lambda_n(S_{(i)})\}$. With these choices, we can write $F_S(f) - F_{S_{(i)}}(f)$ as the Riemann-Stieltjes integral $\int_\alpha^\beta f\, dG$. In particular, we have

$$\left| F_S(f) - F_{S_{(i)}}(f) \right| = \left| \int_\alpha^\beta f\, dG \right| = \left| \int_\alpha^\beta G\, df \right| \leq ||G||_\infty V_f(a,b),$$

where the second equality is obtained through integration by parts upon noting that $G(\alpha) = G(\beta) = 0$. By assumption, $||G||_\infty = ||F_S - F_{S_{(i)}}||_\infty \leq r/n$, and (12) follows. □

Proof of Theorem 1. Our reasoning is similar to that used in the proof of Corollary 1.8 of Guionnet and Zeitouni [9]. Set $\tilde{n} = m + n$ and write $\tilde{M}$ as shorthand for $\tilde{n} \times \tilde{n}$ matrix

$$\tilde{M} = \begin{pmatrix} 0_{n\times n} & X'_{n\times m} \\ X_{m\times n} & 0_{m\times m} \end{pmatrix}.$$

Moreover, set $\tilde{S} = \tilde{M}/\sqrt{m}$, and write Y_i for the i-th row of X, $1 \leq i \leq m$, i.e., $Y_i = (X_{i,\cdot})'$. We view $\tilde{M}$ as a function of $Y_1, \ldots, Y_m$. Also let $\tilde{f}(x) = f\left(x^2\right)$. Clearly

$$\tilde{S}^2 = \begin{pmatrix} X'X/m & 0 \\ 0 & XX'/m \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & XX'/m \end{pmatrix}.$$

This, along with the fact that the matrices $S = X'X/m$ and XX'/m have the same nonzero eigenvalues, allows us to deduce that

$$F_{\tilde{S}}(\tilde{f}) = \frac{2n}{\tilde{n}} F_S(f) + \frac{m-n}{\tilde{n}} f(0),$$

and hence

$$\mathbb{P}\left(|F_S(f) - \mu| > \epsilon\right) \quad = \quad \mathbb{P}\left(|F_{\tilde{S}}(\tilde{f}) - \tilde{\mu}| > \frac{2n}{\tilde{n}}\epsilon\right),$$

where μ ($\tilde{\mu}$) can be either $\mathbb{E}F_S(f)$ ($\mathbb{E}F_{\tilde{S}}(\tilde{f})$) or med $F_S(f)$ (med $F_{\tilde{S}}(\tilde{f})$).
To prove (i), it suffices to note that Theorem 5 applies with $\tilde{M}$, $\tilde{S}$, $\tilde{n}$, n, $\tilde{f}$, and 1 replacing M, S, n, p, f, and C_M, respectively. Using Theorem 5 with these replacements and with $\frac{2n}{\tilde{n}}\epsilon$ replacing ϵ, we see that the left hand side of (1) is bounded as claimed.
To prove (ii), we first note that $||F_{\tilde{S}} - F_{\tilde{S}^{(i)}}||_\infty \leq 2/\tilde{n}$ in view of Lemma 7 (where $\tilde{S}^{(i)}$ is defined as $\tilde{S}$ but with the i-th row of X replaced by an independent copy). Also, note that $\tilde{f}$ is of bounded variation on $\mathbb{R}$ with $V_{\tilde{f}}(\mathbb{R}) \leq V_f(\mathbb{R})$. Hence, Theorem 6 applies with $\tilde{M}, \tilde{S}, \tilde{n}, X_{i,\cdot}, 2$, and $\tilde{f}$ replacing M, S, n, Y_i, r, and f respectively and (2) follows after elementary simplifications. □

Acknowledgment

We sincerely thank Professor Ofer Zeitouni and an anonymous referee for valuable comments that greatly helped us to improve earlier versions of the paper.

References

[1] Azuma, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Math. J. (2)* **19**, 357–367. MR0221571

[2] Bai, Z. D. and Zhou, W. (2008). Large sample covariance matrices without independence sturctures in columns. *Statistica Sinica* **18**, 425–442. MR2411613

[3] Bai, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9**, 3, 611–677. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author. MR1711663

[4] Boucheron, S., Lugosi, G., and Massart, P. (2003). Concentration inequalities using the entropy method. *Ann. Probab.* **31**, 3, 1583–1614. MR1989444

[5] Boutet de Monvel, A. and Khorunzhy, A. (1998). Limit theorems for random matrices. *Markov Process. Related Fields* **4**, 2, 175–197. MR1641617

[6] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23**, 493–507. MR0057518

[7] Götze, F. and Tikhomirov, A. (2004). Limit theorems for spectra of positive random matrices under dependence. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **311**, Veroyatn. i Stat. 7, 92–123, 299. MR2092202

[8] Guionnet, A. (2004). Large deviations and stochastic calculus for large random matrices. *Probab. Surv. 1*, 72–172 (electronic). MR2095566

[9] Guionnet, A. and Zeitouni, O. (2000). Concentration of the spectral measure for large matrices. *Electron. Comm. Probab. 5*, 119–136 (electronic). MR1781846

[10] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30. MR0144363

[11] Houdre, C. and Xu, H. (2008). Concentration of the spectral measure for large random matrices with stable entries. *Electron. J. Probab.* **5**, 107–134. MR2375601

[12] Lang, S. (1993). *Real and functional analysis*, Third ed. Graduate Texts in Mathematics, Vol. **142**. Springer-Verlag, New York. MR1216137

[13] Lieb, E. and Pedersen, G. (2002). Convex multivariable trace functions. *Rev. Math. Phys.* **14**, 631–648. MR1932659

[14] Maurey, B. (1991). Some deviation inequalities. *Geom. Funct. Anal.* **1**, 188–197. MR1097258

[15] McDiarmid, C. (1989). On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*. London Math. Soc. Lecture Note Ser., Vol. **141**. Cambridge Univ. Press, Cambridge, 148–188. MR1036755

[16] Mendelson, S. and Pajor, A. (2006). On singular values of matrices with independent rows. *Bernoulli* **12**, 5, 761–773. MR2265341

[17] Talagrand, M. (1996). A new look at independence. *Ann. Probab.* **24**, 1, 1–34. MR1387624

[18] Walsh, J. L. (1923). A closed set of normal orthogonal functions. *Amer. J. Math.* **45**, 5–24. MR1506485