Electron. Commun. Probab. **18** (2012), no. 25, 1–8.
DOI: 10.1214/ECP.v18-2340
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

On Euclidean random matrices in high dimension

Charles Bordenave*

Abstract

In this note, we study the $n \times n$ random Euclidean matrix whose entry (i,j) is equal to $f(\|X_i - X_j\|)$ for some function f and the X_i's are i.i.d. isotropic vectors in $\mathbb{R}^p$. In the regime where n and p both grow to infinity and are proportional, we give some sufficient conditions for the empirical distribution of the eigenvalues to converge weakly. We illustrate our result on log-concave random vectors.

Keywords: Euclidean random matrices ; Marcenko-Pastur distribution ; Log-concave distribution.
AMS MSC 2010: 60B20 ; 15A18.
Submitted to ECP on September 28, 2012, final version accepted on March 28, 2013.
Supersedes arXiv:1209.5888.

1 Introduction

Let Y be an *isotropic* random vector in $\mathbb{R}^p$, i.e. $\mathbb{E}Y = 0$, $\mathbb{E}[YY^T] = I/p$, where I is the identity matrix. Let $(X_1, \cdots, X_n)$ be independent copies of Y. We define the $n \times n$ matrix A by, for all $1 \leqslant i, j \leqslant n$,

$$A_{ij} = f(\|X_i - X_j\|^2),$$

where $f : [0, \infty) \to \mathbb{R}$ is a measurable function and $\| \cdot \|$ denotes the Euclidean norm. The matrix A is a random Euclidean matrix. It has already attracted some attention see e.g. Mézard, Parisi and Zhee [16], Vershik [18] or Bordenave [7] and references therein.

If B is a symmetric matrix of size n, then its eigenvalues, say $\lambda_1(B), \cdots, \lambda_n(B)$ are real. The empirical spectral distribution (ESD) of B is classically defined as

$$\mu_B = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(B)},$$

where δ_x is the Dirac delta function at x. In this note, we are interested in the asymptotic convergence of μ_A as p and n converge to $+\infty$. This regime has notably been previously considered in El Karoui [10] and Do and Vu [9]. More precisely, we fix a sequence $p(n)$ such that

$$\lim_{n \to \infty} \frac{p(n)}{n} = y \in (0, \infty). \tag{1.1}$$

*CNRS & Université de Toulouse, Institut de Mathématiques de Toulouse, France.
E-mail: charles.bordenave@math.univ-toulouse.fr

Throughout this note, we consider, on a common probability space, an array of random variables $(X_k(n))_{1\leqslant k\leqslant n}$ such that $(X_1(n),\cdots,X_n(n))$ are independent copies of $Y(n)$, an isotropic vector in $\mathbb{R}^{p(n)}$. For each n, we define the Euclidean matrix $A(n)$ associated. For ease of notation, we will often remove the explicit dependence in n: we write p, Y, X_k or A in place of $p(n)$, $Y(n)$, $X_k(n)$ or $A(n)$.

The Marcenko-Pastur probability distribution with parameter $1/y$ is given by

$$\nu_{MP}(dx) = (1-y)^+\delta_0(dx) + \frac{y}{2\pi x}\sqrt{(y_+ - x)(x-y_-)}\mathbf{1}_{[y_-,y_+]}(x)dx,$$

where $x^+ = (x \vee 0)$, $y_\pm = (1 \pm \frac{1}{\sqrt{y}})^2$ and dx denotes the Lebesgue measure. Since the celebrated paper of Marcenko and Pastur [15], this distribution is known to be closely related to empirical covariance matrices in high-dimension.

We say that Y has a *log-concave distribution*, if Y has a density on $\mathbb{R}^p$ which is log-concave. Log-concave random vectors have an increasing importance in convex geometry, probability and statistics (see e.g. Barthe [5]). For example, uniform measures on convex sets are log-concave. We will prove the following result.

Theorem 1.1. *If Y has a log-concave distribution and f is three times differentiable at 2, then, almost surely, as $n\to\infty$, μ_A converges weakly to μ, the law of $f(0)-f(2)+2f'(2)-2f'(2)S$, where S has distribution ν_{MP}.*

With the weaker assumption that f is differentiable at 2, Theorem 1.1 is conjectured in Do and Vu [9]. (For more background, we postpone to the end of the introduction). Their conjecture has motivated this note. It would follow from the thin-shell hypothesis which asserts that there exists $c>0$, such that for any isotropic log-concave vector Y in $\mathbb{R}^p$, $\mathbb{E}(\|Y\|-1)^2 \leqslant c/p$ (see Anttila, Ball and Perissinaki [3] and Bobkov and Koldobsky [6]). Klartag [14] has proved the thin-shell hypothesis for isotropic unconditional log-concave vectors.

The proof of Theorem 1.1 will rely on two recent results on log-concave vectors. Let $X = X(n)$ be the $n\times n$ matrix with columns given by $(X_1(n),\cdots,X_n(n))$. Pajor and Pastur have proved the following :

Theorem 1.2 ([17])**.** *If Y has a log-concave distribution, then, in probability, as $n\to\infty$, μ_{X^TX} converges weakly to ν_{MP}.*

We will also rely on a theorem due to Guédon and Milman.

Theorem 1.3 ([12])**.** *There exist positive constants c_0, c_1 such that if Y is an isotropic log-concave vector in $\mathbb{R}^p$, for any $t\geqslant 0$,*

$$\mathbb{P}(|\|Y\|-1|\geqslant t) \leqslant c_1 \exp\left(-c_0\sqrt{p}\big(t\wedge t^3\big)\right).$$

With Theorems 1.2 and 1.3 in hand, the heuristic behind Theorem 1.1 is simple. Theorem 1.3 implies that $\|X_i\|^2 \simeq 1$ with high probability. Hence, since $\|X_i - X_j\|^2 = \|X_i\|^2 + \|X_j\|^2 - 2X_i^TX_j$, a Taylor expansion of f around 2 gives

$$A_{ij} \simeq \begin{cases} f(2) - 2f'(2)X_i^TX_j & \text{if } i\neq j \\ f(0) & \text{if } i=j. \end{cases}$$

In other words, the matrix A is close to the matrix

$$M = (f(0)-f(2)+2f'(2))I + f(2)J - 2f'(2)X^TX, \tag{1.2}$$

where I is the identity matrix and J is the matrix with all entries equal to 1. From Theorem 1.2, μ_{X^TX} converges weakly to ν_{MP}. Moreover, since J has rank one, it is negligible for the weak convergence of ESD. It follows that μ_M is close to μ. The actual proof of Theorem 1.1 will be elementary and it will follow this heuristic. We shall use some standard perturbation inequalities for the eigenvalues. The idea to perform a Taylor expansion was already central in [10, 9].

Beyond Theorems 1.2-1.3, the proof of Theorem 1.1 is not related to log-concave vectors. In fact, it is nearly always possible to linearize f as soon as the norms of the vectors concentrate around their mean. More precisely, let us say that two sequences of probability measures (μ_n), (ν_n), are asymptotically weakly equal, if for any bounded continuous function f, $\int f d\mu_n - \int f d\nu_n$ converges to 0.

Theorem 1.4. *Assume that there exists an integer $\ell \geqslant 1$ such that $\mathbb{E}|\|Y\| - 1|^{2\ell} = O(p^{-1})$, and that for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{1\leqslant i,j\leqslant n} \left\{ \left|\|X_i - X_j\|^2 - 2\right| \vee \left|\|X_i\|^2 - 1\right| \right\} \leqslant \varepsilon \right) = 1. \tag{1.3}$$

Then, if f is ℓ times differentiable at 2, almost surely, μ_A is asymptotically weakly equal to the law of $f(0) - f(2) + 2f'(2) - 2f'(2)S$, where S has distribution $\mathbb{E}\mu_{X^TX}$.

The case $\ell = 1$ of Theorem 1.4 is contained in Do and Vu [9, Theorem 5]. Besides Theorem 1.2, some general conditions on the matrix X guarantee the convergence of μ_{X^TX}, see Yin and Krishnaiah [19], Götze and Tikhomirov [11] or Adamczak [1].

In settings where $\mathbb{E}|\|Y\| - 1|^2 = O(p^{-1})$, statements analogous to Theorem 1.4 were already known, notably in the case where the entries of Y are i.i.d., see El Karoui [10, Theorem 2.2] or Do and Vu [9, Corollary 3]. When the vector Y satisfies a concentration inequality for all Lipschitz functions, see El Karoui [10, Theorem 2.3]. (it applies notably to log-concave vectors which density in $\mathbb{R}^p$ of the form $e^{-V(x)}$ with $\mathrm{Hess}(V) \geqslant cI$ and $c > 0$).

2 Proofs

2.1 Perturbation inequalities

We first recall some basic perturbation inequalities of eigenvalues and introduce a good notion of distances for ESD. For μ, ν two real probability measures, the *Kolmogorov-Smirnov distance* can be defined as

$$d_{KS}(\mu,\nu) = \sup\left\{ \int f d\mu - \int f d\nu : \|f\|_{BV} \leqslant 1 \right\},$$

where, for $f : \mathbb{R} \to \mathbb{R}$, the bounded variation norm is $\|f\|_{BV} = \sup \sum_{k\in\mathbb{Z}} |f(x_{k+1}) - f(x_k)|$, and the supremum is over all real increasing sequence $(x_k)_{k\in\mathbb{Z}}$. The following inequality is a classical consequence of the interlacing of eigenvalues (see e.g. Bai and Silverstein [4, Theorem A.43]).

Lemma 2.1 (Rank inequality)**.** *If B, C are $n \times n$ Hermitian matrices, then,*

$$d_{KS}(\mu_B, \mu_C) \leqslant \frac{\mathrm{rank}(B - C)}{n}.$$

For $p \geqslant 1$, let μ, ν be two real probability measures such that $\int |x|^p d\mu$ and $\int |x|^p d\nu$ are finite. We define the *L^p-Wasserstein distance* as

$$W_p(\mu,\nu) = \left(\inf_\pi \int_{\mathbb{R}\times\mathbb{R}} |x - y|^p d\pi \right)^{\frac{1}{p}}$$

where the infimum is over all coupling π of μ and ν (i.e. π is probability measure on $\mathbb{R} \times \mathbb{R}$ whose first marginal is equal to μ and second marginal is equal to ν). Hölder inequality implies that for $1 \leqslant p \leqslant q$, $W_p \leqslant W_q$. Moreover, the Kantorovich-Rubinstein duality gives a variational expression for W_1:

$$W_1(\mu, \nu) = \sup \left\{ \int f d\mu - \int f d\nu : \|f\|_L \leqslant 1 \right\},$$

where $\|f\|_L = \sup_{x \neq y} |f(x) - f(y)|/|x - y|$ is the Lipschitz constant of f. The next classical inequality is particularly useful (see e.g. Anderson, Guionnet and Zeitouni [2, Lemma 2.1.19]).

Lemma 2.2 (Hoffman-Wielandt inequality). *If B, C are $n \times n$ Hermitian matrices, then*

$$W_2(\mu_B, \mu_C) \leqslant \sqrt{\frac{1}{n} \mathrm{tr}(B - C)^2}.$$

We finally introduce the distance

$$d(\mu, \nu) = \sup \left\{ \int f d\mu - \int f d\nu : \|f\|_L \leqslant 1 \text{ and } \|f\|_{BV} \leqslant 1 \right\}.$$

By Lemmas 2.1 and 2.2, we obtain that for any $n \times n$ Hermitian matrices B, C,

$$d(\mu_B, \mu_C) \leqslant \sqrt{\frac{1}{n} \mathrm{tr}(B - C)^2} \wedge \frac{\mathrm{rank}(B - C)}{n}. \tag{2.1}$$

Notice that $d(\mu_n, \mu) \to 0$ implies that μ_n converges weakly to μ.

2.2 Concentration inequality

For $x = (x_1, \cdots, x_n) \in \mathcal{M}_{p,n}(\mathbb{R})$, define $a(x)$ as the Euclidean matrix obtained from the columns of x : $a(x)_{ij} = f(\|x_i - x_j\|^2)$. In particular, we have $A = a(X)$. Let $i \in \{1, \cdots, n\}$, $x' = (x'_1, \cdots, x'_n) \in \mathcal{M}_{p,n}(\mathbb{R})$ and assume that $x'_j = x_j$ for all $j \neq i$. Then $a(x)$ and $a(x')$ have all entries equal but the entries on the i-th row or column. We get

$$\mathrm{rank}(a(x) - a(x')) \leqslant 2.$$

It thus follows from Lemma 2.1 that for any function f with $\|f\|_{BV} < \infty$,

$$\left| \int f d\mu_{a(x)} - \int f d\mu_{a(x')} \right| \leqslant \frac{2\|f\|_{BV}}{n}.$$

Using Azuma-Hoeffding's inequality, it is then straightforward to check that for any $t \geqslant 0$,

$$\mathbb{P}\left(\int f d\mu_A - \mathbb{E} \int f d\mu_A \geqslant t \right) \leqslant \exp\left(-\frac{nt^2}{8\|f\|_{BV}^2} \right). \tag{2.2}$$

(For a proof, see [8, proof of Lemma C.2] or Guntuboyina and Leeb [13]). Using the Borel-Cantelli Lemma, this shows that for any such function f, a.s.

$$\int f d\mu_A - \int f d\mathbb{E}\mu_A \to 0. \tag{2.3}$$

Now, recall that M was defined by (1.2). Note that the matrix J has rank one. We get from Theorem 1.2 and Lemma 2.1 that $\mathbb{E}\mu_M$ converges weakly to μ.

Proposition 2.3. *Under the assumptions of Theorem 1.1, we have*

$$\lim_{n \to \infty} d(\mathbb{E}\mu_A, \mathbb{E}\mu_M) = 0.$$

Theorem 1.1 is a corollary of Proposition 2.3. Indeed, it implies that $\mathbb{E}\mu_A$ is a tight sequence of probability measures. Hence, a.s. μ_A is also tight. Then, since the set of continuous functions on an interval endowed with the uniform norm is separable, from (2.3) we get that a.s. μ_A and $\mathbb{E}\mu_A$ are asymptotically weakly equal. Now, Theorem 1.1 follows from a new application of Proposition 2.3.

2.3 Proof of Proposition 2.3

The idea is to perform a multiple Taylor expansion which takes the best out of (2.1).

Step 1 : concentration of norms

By assumption, there exists an open interval $K = (2-\delta, 2+\delta)$ such that f is C^1 in K and, for any $x \in K$,

$$f(x) = f(2) + f'(2)(x-2) + \frac{f''(2)}{2}(x-2)^2 + \frac{f'''(2)}{6}(x-2)^3(1+o(1)).$$

For any $i \neq j$, $(X_i - X_j)/\sqrt{2}$ is an isotropic log-concave vector. Define the sequence $\varepsilon(n) = n^{-\kappa} \wedge (\delta/2)$ with $0 < \kappa < 1/6$. It follows from Theorem 1.3 and the union bound that the event

$$\mathcal{E} = \left\{ \max_{i,j} \left\{ \left| \|X_i - X_j\|^2 - 2 \right| \vee \left| \|X_i\|^2 - 1 \right| \right\} \leqslant \varepsilon(n) \right\}$$

has probability tending to 1 as n goes to infinity.

Step 2 : Taylor expansion around $\|X_i\|^2 + \|X_j\|^2$

We consider the matrix

$$B_{ij} = \begin{cases} f(\|X_i\|^2 + \|X_j\|^2) - 2f'(\|X_i\|^2 + \|X_j\|^2) X_i^T X_j & \text{if } i \neq j \\ f(0) & \text{if } i = j. \end{cases}$$

On the event $\mathcal{E}$, $\|X_i\|^2 + \|X_j\|^2 \in K$. Since f is C^1 in K, we may perform a Taylor expansion of $f(\|X_i - X_j\|^2)$ around $\|X_i\|^2 + \|X_j\|^2$. It follows that for $i \neq j$,

$$|A_{ij} - B_{ij}| = o\big(\|X_i - X_j\|^2 - \|X_i\|^2 - \|X_j\|^2\big) \leqslant \delta(n) \big| X_i^T X_j \big|,$$

where $\delta(n)$ is a sequence going to 0. From (2.1) and Jensen's inequality, we get

$$\begin{aligned} d(\mathbb{E}\mu_A, \mathbb{E}\mu_B) \leqslant \mathbb{E} d(\mu_A, \mu_B) &\leqslant \mathbb{P}(\mathcal{E}^c) + \left(\frac{1}{n} \sum_{i \neq j} \mathbb{E}|A_{ij} - B_{ij}|^2 \mathbf{1}_{\mathcal{E}} \right)^{1/2} \\ &\leqslant \mathbb{P}(\mathcal{E}^c) + \delta(n) \Big(n \mathbb{E} \big| X_1^T X_2 \big|^2 \Big)^{1/2}. \end{aligned}$$

Now, from the assumption that X_1 and X_2 are independent and isotropic, we find

$$\mathbb{E}\big|X_1^T X_2\big|^2 = \mathbb{E}\left(\sum_{k=1}^p X_{k1} X_{k2} \right)^2 = \sum_{k=1}^p \left(\mathbb{E} X_{k1}^2 \right)^2 = \frac{1}{p}.$$

By assumption (1.1), we deduce that

$$\lim_{n \to \infty} d(\mathbb{E}\mu_A, \mathbb{E}\mu_B) = 0.$$

It thus remains to compare $\mathbb{E}\mu_B$ and $\mathbb{E}\mu_M$.

Step 3 : Taylor expansion around 2

We define the matrix

$$C_{ij} = \begin{cases} f(\|X_i\|^2 + \|X_j\|^2) - 2f'(2)X_i^T X_j & \text{if } i \neq j \\ f(0) & \text{if } i = j. \end{cases}$$

We now use the fact that f' is locally Lipschitz at 2. It follows that if $\mathcal{E}$ holds, for $i \neq j$,

$$|B_{ij} - C_{ij}| = O\big(X_i^T X_j(\|X_i\|^2 + \|X_j\|^2 - 2)\big) \leqslant c\,\varepsilon(n)\big|X_i^T X_j\big|.$$

The argument of step 2 implies that

$$\lim_{n\to\infty} d(\mathbb{E}\mu_B, \mathbb{E}\mu_C) = 0.$$

It thus remains to compare $\mathbb{E}\mu_C$ and $\mathbb{E}\mu_M$.

Step 4 : Taylor expansion around 2 **again**

We now consider the matrix

$$D_{ij} = \begin{cases} f(2) + f'(2)(\|X_i\|^2 + \|X_j\|^2 - 2) + \frac{f''(2)}{2}(\|X_i\|^2 + \|X_j\|^2 - 2)^2 & \\ \qquad + \frac{f'''(2)}{6}(\|X_i\|^2 + \|X_j\|^2 - 2)^3 - 2f'(2)X_i^T X_j & \text{if } i \neq j \\ f(0) & \text{if } i = j. \end{cases}$$

We are going to prove that

$$\lim_{n\to\infty} d(\mathbb{E}\mu_C, \mathbb{E}\mu_D) = 0. \tag{2.4}$$

We perform a Taylor expansion of order 3 of $f(\|X_i\|^2 + \|X_j\|^2)$ around 2. It follows that if $\mathcal{E}$ holds, for $i \neq j$,

$$|C_{ij} - D_{ij}| = o\big(\|X_i\|^2 + \|X_j\|^2 - 2\big)^3 \leqslant \delta(n)\big|\|X_i\|^2 + \|X_j\|^2 - 2\big|^3,$$

where $\delta(n)$ is a sequence going to 0. Using (2.1) and arguing as in step 2, in order to prove (2.4), it thus suffices to show that

$$\frac{1}{n}\sum_{i\neq j} \mathbb{E}|\|X_i\|^2 + \|X_j\|^2 - 2|^6 \mathbf{1}_{\mathcal{E}} = O(1).$$

Since, for $\ell \geqslant 1$, $|x+y|^\ell \leqslant 2^{\ell-1}(|x|^\ell + |y|^\ell)$, it is sufficient to show that

$$n\mathbb{E}\big(\|X_1\|^2 - 1\big)^6 \mathbf{1}_{\mathcal{E}} = O(1).$$

To this end, for integer $\ell \geqslant 1$, we write

$$\mathbb{E}\big|\|X_1\|^2 - 1\big|^\ell \mathbf{1}_{\mathcal{E}} \quad = \quad \mathbb{E}|\|X_1\| - 1|^\ell |\|X_1\| + 1|^\ell \mathbf{1}_{\mathcal{E}} \leqslant 3^\ell \mathbb{E}|\|X_1\| - 1|^\ell.$$

Then, Theorem 1.3 implies that there exists c_ℓ such that

$$\mathbb{E}|\|X_1\| - 1|^\ell \leqslant c_\ell\, p^{-\ell/6}.$$

It follows that

$$\mathbb{E}\big|\|X_1\|^2 - 1\big|^\ell \mathbf{1}_{\mathcal{E}} = O\Big(p^{-\ell/6}\Big). \tag{2.5}$$

This proves (2.4). It finally remains to compare $\mathbb{E}\mu_D$ and $\mathbb{E}\mu_M$.

Step 5 : End of proof

We set
$$z_i = (\|X_i\|^2 - 1).$$
We note that for $i \neq j$,
$$D_{ij} = M_{ij} + \sum_{1 \leqslant k+\ell \leqslant 3} c_{k\ell} z_i^k z_j^\ell,$$
for some coefficients $c_{k\ell}$ depending on $f'(2), f''(2), f'''(2)$. Note that $c_{10} = c_{01} = f'(2)$. Similarly,
$$D_{ii} = M_{ii} + 2f'(2) z_i = M_{ii} + c_{10} z_i + c_{01} z_i.$$
Define the matrix E, for all $1 \leqslant i, j \leqslant n$,
$$E_{ij} = M_{ij} + \sum_{1 \leqslant k+\ell \leqslant 3} c_{k\ell} z_i^k z_j^\ell.$$
If $\mathcal{E}$ holds, then $\max_i |z_i| \leqslant \varepsilon(n)$ and we find
$$|E_{ij} - D_{ij}| = \mathbf{1}(i = j) \left| \sum_{2 \leqslant k+\ell \leqslant 3} c_{k\ell} z_i^k z_i^\ell \right| \leqslant c\mathbf{1}(i = j)\varepsilon(n)^2.$$
It follows from (2.1) that
$$\begin{aligned} d(\mathbb{E}\mu_D, \mathbb{E}\mu_E) \leqslant \mathbb{E} d(\mu_D, \mu_E) \quad &\leqslant \quad \mathbb{P}(\mathcal{E}^c) + \left(\frac{1}{n} \sum_{i,j} \mathbb{E} |E_{ij} - D_{ij}|^2 \mathbf{1}_{\mathcal{E}} \right)^{1/2} \\ &\leqslant \quad \mathbb{P}(\mathcal{E}^c) + c\varepsilon(n)^2. \end{aligned}$$
We deduce that
$$\lim_{n \to \infty} d(\mathbb{E}\mu_D, \mathbb{E}\mu_E) = 0.$$
We notice finally that the matrix $E - M$ is equal to
$$\sum_{1 \leqslant k+\ell \leqslant 3} c_{k\ell} Z_k {Z_\ell}^T,$$
where Z_k is the vector with coordinates $(z_i^k)_{1 \leqslant i \leqslant n}$. It implies in particular that $\mathrm{rank}(E - M) \leqslant 9$, indeed the rank is subadditive and $\mathrm{rank}(Z_k {Z_\ell}^T) \leqslant 1$. In particular, it follows from (2.1) that
$$d(\mathbb{E}\mu_E, \mathbb{E}\mu_M) \leqslant \mathbb{E} d(\mu_E, \mu_M) \leqslant \frac{9}{n}.$$
This concludes the proof of Proposition 2.3 and of Theorem 1.1.

2.4 Proof of Theorem 1.4

The isotropy implies that
$$\int x^2 \mathbb{E}\mu_{X^T X}(dx) = \frac{1}{n} \mathbb{E}\mathrm{tr}(X^T X) = 1.$$
It follows that $\mathbb{E}\mu_{X^T X}$ and $\mathbb{E}\mu_M$ are tight sequences of probability measures. Note also that the concentration inequality (2.2) holds. It is thus sufficient to prove the analog of Proposition 2.3. If $\ell \geqslant 2$, the proof is essentially unchanged. In step 1, the assumption (1.3) implies the existence of a sequence $\varepsilon = \varepsilon(n)$ going to 0 such that $\mathbb{P}(\mathcal{E}) \to 1$. Then, in step 4, it suffices to extend the Taylor expansion up to ℓ.

For the case $\ell = 1$: in step 2, we perform directly the Taylor expansion around 2, for $i \neq j$ we write $f(\|X_i - X_j\|^2) = f(2) - 2f'(2) X_i^T X_j (1 + o(1))$. We then move directly to step 5. (As already pointed, this case is treated in [9]).

References

[1] Radosław Adamczak, *On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries*, Electron. J. Probab. **16** (2011), no. 37, 1068–1095. MR-2820070

[2] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR-2760897

[3] Milla Anttila, Keith Ball, and Irini Perissinaki, *The central limit problem for convex bodies*, Trans. Amer. Math. Soc. **355** (2003), no. 12, 4723–4735 (electronic). MR-1997580

[4] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010. MR-2567175

[5] Franck Barthe, *Un théorème de la limite centrale pour les ensembles convexes (d'après Klartag et Fleury-Guédon-Paouris)*, Astérisque (2010), no. 332, Exp. No. 1007, ix, 287–304, Séminaire Bourbaki. Volume 2008/2009. Exposés 997–1011. MR-2648682

[6] S. G. Bobkov and A. Koldobsky, *On the central limit property of convex bodies*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 1807, Springer, Berlin, 2003, pp. 44–52. MR-2083387

[7] Charles Bordenave, *Eigenvalues of Euclidean random matrices*, Random Structures Algorithms **33** (2008), no. 4, 515–532. MR-2462254

[8] Charles Bordenave, Pietro Caputo, and Djalil Chafaï, *Spectrum of non-Hermitian heavy tailed random matrices*, Comm. Math. Phys. **307** (2011), no. 2, 513–560. MR-2837123

[9] Yen Do and Van Vu, *The spectrum of random kernel matrices*, preprint, arXiv:1206.3763

[10] Noureddine El Karoui, *The spectrum of kernel random matrices*, Ann. Statist. **38** (2010), no. 1, 1–50. MR-2589315

[11] F. Götze and A. Tikhomirov, *Limit theorems for spectra of positive random matrices under dependence*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **311** (2004), no. Veroyatn. i Stat. 7, 92–123, 299. MR-2092202

[12] Olivier Guédon and Emanuel Milman, *Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures*, Geom. Funct. Anal. **21** (2011), no. 5, 1043–1068. MR-2846382

[13] Adityanand Guntuboyina and Hannes Leeb, *Concentration of the spectral measure of large Wishart matrices with dependent entries*, Electron. Commun. Probab. **14** (2009), 334–342. MR-2535081

[14] Bo'az Klartag, *A Berry-Esseen type inequality for convex bodies with an unconditional basis*, Probab. Theory Related Fields **145** (2009), no. 1-2, 1–33. MR-2520120

[15] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Mat. Sb. (N.S.) **72 (114)** (1967), 507–536. MR-0208649

[16] M. Mézard, G. Parisi, and A. Zee, *Spectra of Euclidean random matrices*, Nuclear Phys. B **559** (1999), no. 3, 689–701. MR-1724455

[17] A. Pajor and L. Pastur, *On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution*, Studia Math. **195** (2009), no. 1, 11–29. MR-2539559

[18] A. M. Vershik, *Random metric spaces and universality*, Uspekhi Mat. Nauk **59** (2004), no. 2(356), 65–104. MR-2086637

[19] Y. Q. Yin and P. R. Krishnaiah, *Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic*, Teor. Veroyatnost. i Primenen. **30** (1985), no. 4, 810–816. MR-0816299