ANOTHER OBSERVATION
ABOUT OPERATOR COMPRESSIONS

ELIZABETH S. MECKES AND MARK W. MECKES

(Communicated by Marius Junge)

ABSTRACT. Let T be a self-adjoint operator on a finite dimensional Hilbert space. It is shown that the distribution of the eigenvalues of a compression of T to a subspace of a given dimension is almost the same for almost all subspaces. This is a coordinate-free analogue of a recent result of Chatterjee and Ledoux on principal submatrices. The proof is based on measure concentration and entropy techniques, and the result improves on some aspects of the result of Chatterjee and Ledoux.

1. INTRODUCTION

Let T be an operator on a (real or complex) n-dimensional Hilbert space $\mathcal{H}$, and let $E \subseteq \mathcal{H}$ be a subspace. The *compression* of T to E is the operator $T_E = \pi_E T|_E = \pi_E T \pi_E^*$ on E, where $\pi_E : \mathcal{H} \to E$ is the orthogonal projection. The spectral distribution of a self-adjoint operator T is the probability measure

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

on $\mathbb{R}$, where $\lambda_1(T) \geq \cdots \geq \lambda_n(T)$ are the eigenvalues of T, counted with multiplicity.

The following result shows that for $1 \leq k \leq n$ and a self-adjoint operator T on an n-dimensional Hilbert space $\mathcal{H}$, the empirical spectral distribution of the compression T_E is almost the same for almost every k-dimensional subspace $E \subseteq \mathcal{H}$. The notations σ_k and ρ are explained after the statement of the theorem; d_1 denotes the Kantorovich-Rubinstein metric on probability measures, also defined below.

Theorem 1. *Let $\mathcal{H}$ be an n-dimensional Hilbert space, T a self-adjoint operator on $\mathcal{H}$, and $1 \leq k \leq n$. Let E be a k-dimensional subspace of $\mathcal{H}$ chosen at random with respect to the rotationally invariant probability measure on the Grassmann manifold. Let μ_E be the empirical spectral distribution of the compression of T to E, and let $\mu = \mathbb{E}\mu_E$. Then*

$$(1) \qquad \mathbb{E}d_1(\mu_E, \mu) \leq c_1 \frac{\sigma_k(T)^{4/7} \rho(T)^{3/7}}{(kn)^{2/7}}$$

and

$$(2) \qquad \mathbb{P}\left[d_1(\mu_E, \mu) \geq c_1 \frac{\sigma_k(T)^{4/7}\rho(T)^{3/7}}{(kn)^{2/7}} + t\right] \leq c_2 \exp\left[-c_3 \frac{kn}{\sigma_k(T)^2} t^2\right]$$

for every $t > 0$, where $c_1, c_2, c_3 > 0$ are absolute (computable) constants.

Here $\rho(T) = \frac{1}{2}(\lambda_1 - \lambda_n)$ denotes one half the spectral diameter of T (which is different in general from the classical spectral radius); it is easy to check that $\rho(T)$ is the distance of T from the space of real scalar operators with respect to the operator norm. For $1 \leq k \leq n$,

$$\sigma_k(T) = \inf_{\lambda \in \mathbb{R}} \sqrt{\sum_{i=1}^{k} s_i(T - \lambda I)^2},$$

where $s_1 \geq \cdots \geq s_n \geq 0$ denote singular values. That is, $\sigma_k(T)$ is the distance of T from the space of real scalar operators with respect to the norm $\|T\|_{(k),2} = \sqrt{\sum_{i=1}^{k} s_i(T)^2}$.

The space of probability measures (with finite first moment) on $\mathbb{R}$ is equipped with the Kantorovich-Rubinstein or L^1-Wasserstein distance d_1, which may be equivalently defined in the following three ways:

(3)
$$d_1(\mu, \nu) = \inf_{\pi} \int_{\mathbb{R} \times \mathbb{R}} |x - y| \ d\pi(x, y) = \sup_{f} \left(\int f \ d\mu - \int f \ d\nu\right) = \|F_\mu - F_\nu\|_{L^1(\mathbb{R})}.$$

Here π varies over all probability measures on $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν, f varies over all Lipschitz continuous functions $\mathbb{R} \to \mathbb{R}$ with Lipschitz constant at most 1, and F_μ, F_ν are the cumulative distribution functions of μ, ν. All three characterizations will be used in this note; for the equalities see [10, Chapter 1].

Theorem 1 is a coordinate-free analogue of a recent result of Chatterjee and Ledoux [3], which considered the empirical spectral measure of a random $k \times k$ principal submatrix of a fixed $n \times n$ Hermitian matrix. The approach taken in [3] is rather different from the one taken here; the result of [3] is also given in terms of the Kolmogorov distance between measures rather than Wasserstein distance. See section 3 below for a more detailed comparison of the results.

2. Proof of Theorem 1

Throughout this section let $\mathcal{H}$ and T be fixed, and let μ_E and μ be as defined in the statement of the theorem. For brevity we write $\sigma_k = \sigma_k(T)$ and $\rho = \rho(T)$. The notation $\lesssim B$ means $\leq cB$, where $c > 0$ is some absolute constant.

Recall that the Grassmann manifold $G_k(\mathcal{H})$ of k-dimensional subspaces of $\mathcal{H}$ is equipped with the metric

$$d(E, F) = \inf \sqrt{\sum_{i=1}^{k} \|e_i - f_i\|^2},$$

where the infimum is over all orthonormal bases $\{e_1, \ldots, e_k\}$ and $\{f_1, \ldots, f_k\}$ of E and F respectively.

Lemma 2. *For any $E, F \in G_k(\mathcal{H})$, $d_1(\mu_E, \mu_F) \leq \frac{2\sigma_k}{\sqrt{k}} d(E, F)$.*

Proof. Define a coupling π of μ_E and μ_F by $\pi = \frac{1}{k} \sum_{i=1}^{k} \delta_{(\lambda_i(T_E), \lambda_i(T_F))}$. Then

$$d_1(\mu_E, \mu_F) \leq \frac{1}{k} \sum_{i=1}^{k} |\lambda_i(T_E) - \lambda_i(T_F)| \leq \sqrt{\frac{1}{k} \sum_{i=1}^{k} |\lambda_i(T_E) - \lambda_i(T_F)|^2}.$$

Now if $\{e_1, \ldots, e_k\}$ and $\{f_1, \ldots, f_k\}$ are orthonormal bases of E and F, then the matrices of T_E and T_F with respect to these bases are $\left[\langle T(e_j), e_i \rangle \right]_{i,j=1}^{k}$ and $\left[\langle T(f_j), f_i \rangle \right]_{i,j=1}^{k}$ respectively. As a consequence of Lidskii's theorem (see [1, §III.4]), for any $k \times k$ Hermitian matrices A and B,

$$\sqrt{\sum_{i=1}^{k} |\lambda_i(A) - \lambda_i(B)|^2} \leq \|A - B\|_{HS} = \sqrt{\sum_{i,j=1}^{k} |a_{ij} - b_{ij}|^2}.$$

Thus by the self-adjointness of T and the Cauchy-Schwarz inequality,

$$
\begin{aligned}
d_1(\mu_E, \mu_F) &\leq \sqrt{\frac{1}{k} \sum_{i,j=1}^{k} |\langle T(e_j), e_i \rangle - \langle T(f_j), f_i \rangle|^2} \\
&= \sqrt{\frac{1}{k} \sum_{i,j=1}^{k} |\langle T(e_j), e_i - f_i \rangle + \langle e_j - f_j, T(f_i) \rangle|^2} \\
&\leq \sqrt{\frac{1}{k} \sum_{i,j=1}^{k} \left[\|T(e_j)\| \, \|e_i - f_i\| + \|T(f_i)\| \, \|e_j - f_j\| \right]^2} \\
&\leq \frac{d(E,F)}{\sqrt{k}} \left(\sqrt{\sum_{j=1}^{k} \|T(e_j)\|^2} + \sqrt{\sum_{i=1}^{k} \|T(f_i)\|^2} \right) \leq \frac{2 d(E,F)}{\sqrt{k}} \|T\|_{(k),2}.
\end{aligned}
$$

Observing that $d_1(\mu_E, \mu_F)$ is invariant under the addition of a real scalar matrix to T, the lemma is proved. $\square$

The same proof as above can be carried out (and is slightly simpler) with the Kantorovich-Rubinstein distance replaced by the L^2-Wasserstein distance, although this observation will not be used here.

The following concentration inequality goes back to Gromov and Milman [5]; see also section 2.1 of [6], where it is pointed out explicitly that the same result applies in the complex case.

Theorem 3. *Let $f : G_k(\mathcal{H}) \to \mathbb{R}$ be 1-Lipschitz with respect to the metric d on $G_k(\mathcal{H})$, and let $E \in G_k(\mathcal{H})$ be distributed according to the rotationally invariant probability measure on $G_k(\mathcal{H})$. Then*

$$\mathbb{P}\big[|f(E) - \mathbb{E}f(E)| \geq t \big] \lesssim \exp\big[-cnt^2 \big]$$

for $t > 0$, where $c > 0$ is an absolute constant.

Observe that (1), Lemma 2, and Theorem 3 together imply (2), so it suffices now to prove (1).

Let $E \in G_k(\mathcal{H})$ be distributed according to the rotationally invariant probability measure on $G_k(\mathcal{H})$. For a given function $f : \mathbb{R} \to \mathbb{R}$, define the random variable

$X_f = \int f d\mu_E - \int f d\mu$. By Lemma 2 and Theorem 3, for functions f, g,

$$(4) \qquad \mathbb{P}\big[\, |X_f - X_g| \geq t \,\big] = \mathbb{P}\big[\, |X_{f-g}| \geq t \,\big] \lesssim \exp\left[-c\frac{kn}{\sigma_k^2 \,|f-g|_L^2}\right]$$

for $t \geq 0$, where $|f|_L$ denotes the Lipschitz constant of f.

The inequality (4) shows that the random process X_f, indexed by some family $\mathcal{F}$ of Lipschitz continuous test functions (to be determined), satisfies a sub-Gaussian increment condition with respect to the norm $\|\cdot\|' = \frac{\sigma_k}{\sqrt{kn}} \|\cdot\|_{C^1}$ on $\mathcal{F}$ (here, $\|f\|_{C^1} := \max\{\|f\|_\infty, \|f'\|_\infty\}$, so that for $f \in C^1$, $|f|_L \leq \|f\|_{C^1}$). This raises the possibility to estimate its expected supremum by Dudley's entropy bound [4] (see also [8]):

$$(5) \qquad \mathbb{E}\sup_{f \in \mathcal{F}} X_f \lesssim \int_0^\infty \sqrt{\log N(\mathcal{F}, \|\cdot\|', \varepsilon)}\; d\varepsilon,$$

where $N(\mathcal{F}, \|\cdot\|', \varepsilon)$ is the minimum number of sets of diameter ε with respect to $\|\cdot\|'$ needed to cover $\mathcal{F}$. Since μ_E and μ are supported on $[\lambda_n, \lambda_1]$,

$$d_1(\mu_E, \mu) = \sup\{X_f : |f|_L \leq 1\} = \sup\{X_f : |f|_L \leq 1, \|f\|_\infty \leq 2\rho\}.$$

Thus to prove (1), it suffices to estimate $\mathbb{E}\sup_{f \in \mathcal{F}} X_f$ for $\mathcal{F} = \{f : \|f\|_{C^1} \leq 1 + 2\rho\}$. However, since C^1 is an infinite dimensional function space, for this choice of $\mathcal{F}$ the covering numbers $N(\mathcal{F}, \|\cdot\|', \varepsilon)$ in (5) will always be infinite for small ε.

Instead, define $\mathcal{F} = \{f : \|f\|_{C^2} \leq 1\}$, where $\|f\|_{C^2} := \max\{\|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty\}$. The covering numbers $N(\mathcal{F}, \|\cdot\|_{C^1}, \varepsilon)$ can be estimated using the methods of [9, §2.7]; see [7] for explicit estimates which, combined with (5) and a linear change of variables, yield

$$(6) \quad \mathbb{E}\sup\{X_f : \|f\|_{C^2} \leq 1\} \lesssim \frac{\sigma_k}{\sqrt{kn}} \int_0^1 \sqrt{1 + \log\frac{1}{\varepsilon} + \frac{1}{\varepsilon}(\rho+1)}\; d\varepsilon \lesssim \frac{\sigma_k\sqrt{\rho+1}}{\sqrt{kn}}.$$

The bound (1) is now derived from (6) via a smoothing and scaling argument. Fix $f : \mathbb{R} \to \mathbb{R}$ with $|f|_L \leq 1$ and $\|f\|_\infty \leq 2\rho$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a smooth probability density with finite first absolute moment and $\varphi' \in L^1(\mathbb{R})$. For $t > 0$ define $\varphi_t(x) = \frac{1}{t}\varphi(\frac{x}{t})$, and let $g_t = f * \varphi_t$. Then
$$(7)$$
$$\|g_t\|_\infty \leq \|f\|_\infty \|\varphi_t\|_1 \leq 2\rho, \quad \|g_t'\|_\infty \leq |f|_L \|\varphi_t\|_1 \leq 1, \quad \|g_t''\|_\infty \leq |f|_L \|\varphi_t'\|_1 \lesssim \frac{1}{t}.$$

Now for any probability measure ν on $\mathbb{R}$,

$$\left|\int f\, d\nu - \int g_t\, d\nu\right| = \left|\int\int [f(x) - f(x-y)]\varphi(y)\, dy\, d\nu(x)\right| \lesssim t.$$

Thus

$$|X_f| \leq \left|\int f\, d\mu_E - \int g_t\, d\mu_E\right| + \left|\int g_t\, d\mu_E - \int g_t\, d\mu\right| + \left|\int g_t\, d\mu - \int f\, d\mu\right|$$
$$\lesssim t + \|g_t\|_{C^2} \sup\{X_h : \|h\|_{C^2} \leq 1\},$$

and so by (6) and (7),

$$\mathbb{E}d_1(\mu_E, \mu) \lesssim t + \left(1 + 2\rho + \frac{1}{t}\right)\frac{\sigma_k\sqrt{\rho+1}}{\sqrt{kn}}.$$

Picking t of the order $\frac{\sqrt{\sigma_k}(\rho+1)^{1/4}}{(kn)^{1/4}}$ yields

$$(8) \qquad \mathbb{E}d_1(\mu_E, \mu) \lesssim \frac{\sqrt{\sigma_k}(\rho+1)^{1/4}}{(kn)^{1/4}} + \frac{\sigma_k(\rho+1)^{3/2}}{\sqrt{kn}}.$$

Now apply (8) with the operator T replaced by sT for $s > 0$. It is easy to check that the Kantorovich-Rubinstein distance $d_1(\mu_E, \mu)$ is homogeneous with respect to this rescaling, as are σ_k and ρ. Thus one obtains

$$\mathbb{E}d_1(\mu_E, \mu) \lesssim \frac{1}{s}\left(\frac{\sqrt{s\sigma_k}(s\rho+1)^{1/4}}{(kn)^{1/4}} + \frac{s\sigma_k(s\rho+1)^{3/2}}{\sqrt{kn}}\right).$$

Picking s of the order $\frac{(kn)^{1/7}}{\sigma_k^{2/7}\rho^{5/7}}$ now yields (1).   □

3. Discussion

In [3], Chatterjee and Ledoux proved a version of Theorem 1 for principal submatrices. Namely, let $\mathcal{H} = \mathbb{C}^n$, and suppose that E is now uniformly distributed among k-dimensional *coordinate* subspaces of $\mathbb{C}^n$. Then [3] shows that

$$(9) \qquad \mathbb{P}\big[d_\infty(\mu_E, \mu) \geq k^{-1/2} + t\big] \leq 12\sqrt{k}e^{-t\sqrt{k/8}}$$

for $t > 0$, and consequently,

$$(10) \qquad \mathbb{E}d_\infty(\mu_E, \mu) \leq \frac{13 + \sqrt{8}\log k}{\sqrt{k}}.$$

Here $d_\infty(\mu, \nu) = \|F_\mu - F_\nu\|_\infty$ is the Kolmogorov distance between probability measures μ and ν on $\mathbb{R}$.

It is likely that the methods of this paper could be used to prove a result in the setting of [3], by replacing Theorem 3, which follows from concentration inequalities on the unitary or special orthogonal group, with an appropriate concentration inequality on the symmetric group S_n. Furthermore, it may be possible to prove a result in the setting of this paper using methods related to those of [3], such as adapting the approach of Chatterjee in [2]. Below some quantitative comparison will be offered between Theorem 1 and the result of [3], ignoring the fact that the random subspace E has a different distribution in each setting. In particular, the distribution of E is probably responsible for the difference between the subexponential tail decay in (9) and the sub-Gaussian tail decay in (2).

Before discussing more specific quantitative comparisons, we note that the clearest difference between the two results is that ours is coordinate-free. While there are settings in which coordinates have meaning and thus coordinate-oriented results are natural, there are many settings in which there would be no clearly preferred basis in which to view an operator. Take, for example, the Laplacian Δ on the sphere $\mathbb{S}^{n-1}$. It has eigenvalues (up to sign convention) $0 > -\lambda_1 > -\lambda_2 > \cdots \to -\infty$, and the corresponding eigenspaces are multidimensional. If $\mathcal{H}$ is the span of the first m eigenspaces, with $T = \Delta|_{\mathcal{H}}$, then there is no canonical choice of basis within each eigenspace, and so it would seem more natural to consider compressions of T to *all* subspaces of a given dimension, rather than only to the coordinate subspaces for some choice of basis.

Comparisons of the results are made somewhat difficult as the Kantorovich-Rubinstein distance d_1 and the Kolmogorov distance d_∞ are not comparable in

general. However, since the measures here are all supported in the interval $[\lambda_n(T),$ $\lambda_1(T)]$, from the third representation of d_1 in (3) one obtains the estimate

$$(11) \qquad d_1(\mu_E, \mu) \leq 2\rho(T)d_\infty(\mu_E, \mu)$$

in the present context. This estimate is related to a qualitative difference between d_1 and d_∞: whereas d_1 is homogeneous with respect to a rescaling of the supports of measures (a fact which was exploited in the proof of Theorem 1), d_∞ is invariant under rescaling. Which behavior is more convenient may vary by the context.

Inequality (11) makes some quantitative comparisons between the results of [3] and Theorem 1 possible. Observe that (9) and (10) only yield nontrivial information if $k \gg 1$ (which of course requires $n \gg 1$), whereas under appropriate scaling, Theorem 1 is nontrivial for $n \gg 1$ even if k is small. In particular, (9) and (11) imply that the fluctuations of $d_1(\mu_E, \mu)$ above its mean are of order (ignoring logarithmic factors) at most $k^{-1/2}\rho(T)$, whereas (2) together with the general estimate $\sigma_k(T) \leq \sqrt{k}\rho(T)$ yields fluctuations of order at most $(kn)^{-1/2}\sigma_k(T) \leq n^{-1/2}\rho(T)$.

The issue of the expected distance is more complicated. The general estimate $\rho(T) \leq \sigma_k(T)$ and inequalities (10) and (11) imply that

$$(12) \qquad \mathbb{E}d_1(\mu_E, \mu) \leq c\frac{\rho(T)\log k}{\sqrt{k}} \leq c\frac{\sigma_k(T)^{4/7}\rho(T)^{3/7}\log k}{\sqrt{k}},$$

which is slightly weaker than (1) for k large (in which case the lossy estimates used to arrive at (12) mean that the comparison should probably not be taken too seriously) and significantly weaker for k small. Since the different distributions of E are being ignored here there is little point in making the comparison very precise.

Finally, the comparison of fluctuations highlights that the methods of this paper are more sensitive to the proximity of T to the space of scalar operators. If T is a (real) scalar operator then μ_E is a constant point mass, so it is natural to expect that if T is nearly scalar in some sense then μ_E will be more tightly concentrated than in general. The results of [3] do not directly reflect this at all, although the estimate (11) allows one to insert this effect by hand when changing metrics. However, $\sigma_k(T)$ provides a sharper measure than $\rho(T)$ of how close T is to scalar, and in some cases the bound $(kn)^{-1/2}\sigma_k(T)$ on the order of the fluctuations may be even much smaller than $n^{-1/2}\rho(T)$. This is the case, for example, if T has a large number of tightly clustered eigenvalues with a small number of outliers.

References

[1] R. Bhatia. *Matrix Analysis*, volume 169 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1997. MR1477662 (98i:15003)

[2] S. Chatterjee. Concentration of Haar measures, with an application to random matrices. *J. Funct. Anal.*, 245(2):379–389, 2007. MR2309833 (2008a:60011)

[3] S. Chatterjee and M. Ledoux. An observation about submatrices. *Elect. Comm. in Probab.*, 14:495–500, 2009. MR2559099

[4] R.M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, 1:290–330, 1967. MR0220340 (36:3405)

[5] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105(4):843–854, 1983. MR708367 (84k:28012)

[6] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001. MR1849347 (2003k:28019)

[7] E. Meckes. Approximation of projections of random vectors. To appear in *J. Theoret. Probab.*; DOI 10.1007/s10959-010-0299-2.

[8] M. Talagrand. Majorizing measures: the generic chaining. *Ann. Probab.*, 24(3):1049–1103, 1996. MR1411488 (97k:60097)

[9] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics. MR1385671 (97g:60035)

[10] C. Villani. *Topics in Optimal Transportation*, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003. MR1964483 (2004e:90003)

Department of Mathematics, Case Western Reserve University, Cleveland, Ohio 44106

E-mail address: `elizabeth.meckes@case.edu`

Department of Mathematics, Case Western Reserve University, Cleveland, Ohio 44106

E-mail address: `mark.meckes@case.edu`