DOI: 10.1214/ECP.v14-1504

Elect. Comm. in Probab. **14** (2009), 495–500

ELECTRONIC COMMUNICATIONS in PROBABILITY

AN OBSERVATION ABOUT SUBMATRICES

SOURAV CHATTERJEE[1]
367 Evans Hall #3860, Department of Statistics, University of California at Berkeley, CA 94720-3860, USA.
email: sourav@stat.berkeley.edu

MICHEL LEDOUX[2]
Institut de Mathématiques, Université de Toulouse, 31062 Toulouse Cedex 9, France
email: ledoux@math.univ-toulouse.fr

Submitted August 19, 2008, *accepted in final form* October 7, 2009

AMS 2000 Subject classification: 60E15, 15A52
Keywords: Random matrix, concentration of measure, empirical distribution, eigenvalue

Abstract
Let M be an arbitrary Hermitian matrix of order n, and k be a positive integer $\leq n$. We show that if k is large, the distribution of eigenvalues on the real line is almost the same for almost all principal submatrices of M of order k. The proof uses results about random walks on symmetric groups and concentration of measure. In a similar way, we also show that almost all $k \times n$ submatrices of M have almost the same distribution of singular values.

1 Introduction

Let M be a square matrix of order n. For any two sets of integers $i_1, \ldots, i_k$ and $j_1, \ldots, j_l$ between 1 and n, $M(i_1, \ldots, i_k; j_1, \ldots, j_l)$ denotes the submatrix of M formed by deleting all rows except rows $i_1, \ldots, i_k$, and all columns except columns $j_1, \ldots, j_l$. A submatrix like $M(i_1, \ldots, i_k; i_1, \ldots, i_k)$ is called a principal submatrix.

For a Hermitian matrix M of order n with eigenvalues $\lambda_1, \ldots, \lambda_n$ (repeated by multiplicities), let F_M denote the empirical spectral distribution function of M, that is,

$$F_M(x) := \frac{\#\{i : \lambda_i \leq x\}}{n}.$$

The following result shows that given $1 \ll k \leq n$ and any Hermitian matrix M of order n, the empirical spectral distribution is *almost the same* for *almost every* principal submatrix of M of order k.

[1]RESEARCH PARTIALLY SUPPORTED BY NSF GRANT DMS-0707054 AND A SLOAN RESEARCH FELLOWSHIP
[2]RESEARCH PARTIALLY SUPPORTED BY THE ANR GRANDES MATRICES ALÉATOIRES

Theorem 1. *Take any $1 \le k \le n$ and a Hermitian matrix M of order n. Let A be a principal submatrix of M chosen uniformly at random from the set of all $k \times k$ principal submatrices of M. Let F be the expected spectral distribution function of A, that is, $F(x) = \mathbb{E}F_A(x)$. Then for each $r \ge 0$,*

$$\mathbb{P}(\|F_A - F\|_\infty \ge k^{-1/2} + r) \le 12\sqrt{k}e^{-r\sqrt{k/8}}.$$

Consequently, we have

$$\mathbb{E}\|F_A - F\|_\infty \le \frac{13 + \sqrt{8}\log k}{\sqrt{k}}.$$

Exactly the same results hold if A is a $k \times n$ submatrix of M chosen uniformly at random, and F_A is the empirical distribution function of the singular values of A. Moreover, in this case M need not be Hermitian.

Remarks. (i) Note that the bounds do not depend at all on the entries of M, nor on the dimension n.
(ii) We think it is possible to improve the $\log k$ to $\sqrt{\log k}$ using Theorem 2.1 of Bobkov [2] instead of the spectral gap techniques that we use. (See also Bobkov and Tetali [3].) However, we do not attempt to make this small improvement because $\sqrt{\log k}$, too, is unlikely to be optimal. Taking M to be the matrix which has $n/2$ 1's on the diagonal and the rest of the elements are zero, it is easy to see that there is a lower bound of $const.k^{-1/2}$. We conjecture that the matching upper bound is also true, that is, there is a universal constant C such that $\mathbb{E}\|F_A - F\|_\infty \le Ck^{-1/2}$.
(iii) The function F is determined by M and k. If M is a diagonal matrix, then F is exactly equal to the spectral measure of M, irrespective of k. However it is not difficult to see that the spectral measure of M cannot, in general, be reconstructed from F.
(iv) The result about random $k \times n$ submatrices is related to the recent work of Rudelson and Vershynin [6]. Let us also refer to [6] for an extensive list of references to the substantial volume of literature on random submatrices in the computing community. However, most of this literature (and also [6]) is concerned with the largest eigenvalue and not the bulk spectrum. On the other hand, the existing techniques are usually applicable only when M has low rank or low 'effective rank' (meaning that most eigenvalues are negligible compared to the largest one).

A numerical illustration. The following simple example demonstrates that the effects of Theorem 1 can kick in even when k is quite small. We took M to be a $n \times n$ matrix for $n = 100$, with (i,j)th entry $= \min\{i,j\}$. This is the covariance matrix of a simple random walk up to time n. We chose $k = 20$, and picked two $k \times k$ principal submatrices A and B of M, uniformly and independently at random. Figure 1 plots to superimposed empirical distribution functions of A and B, after excluding the top 4 eigenvalues since they are too large. The classical Kolmogorov-Smirnov test from statistics gives a p-value of 0.9999 (and $\|F_A - F_B\|_\infty = 0.1$), indicating that the two distributions are *statistically indistinguishable*.

2 Proof

Markov chains. Let us now quote two results about Markov chains that we need to prove Theorem 1. Let $\mathscr{X}$ be a finite or countable set. Let $\Pi(x,y) \ge 0$ satisfy

$$\sum_{y \in \mathscr{X}} \Pi(x,y) = 1$$

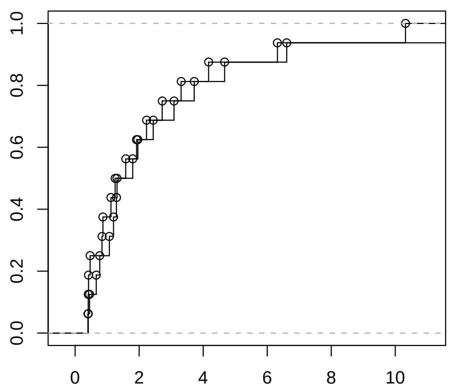

Figure 1: Superimposed empirical distribution functions of two submatrices of order 20 chosen at random from a deterministic matrix of order 100.

for every $x \in \mathscr{X}$. Assume furthermore that there is a symmetric invariant probability measure μ on $\mathscr{X}$, that is, $\Pi(x,y)\mu(\{x\})$ is symmetric in x and y, and $\sum_x \Pi(x,y)\mu(\{x\}) = \mu(\{y\})$ for every $y \in \mathscr{X}$. In other words, (Π,μ) is a reversible Markov chain. For every $f : \mathscr{X} \to \mathbb{R}$, define

$$\mathscr{E}(f,f) = \frac{1}{2}\sum_{x,y\in\mathscr{X}} (f(x)-f(y))^2 \Pi(x,y)\mu(\{x\}).$$

The spectral gap or the Poincaré constant of the chain (Π,μ) is the largest $\lambda_1 > 0$ such that for all f's,

$$\lambda_1 \mathrm{Var}_\mu(f) \le \mathscr{E}(f,f).$$

Set also

$$|||f|||_\infty^2 = \frac{1}{2}\sup_{x\in\mathscr{X}} \sum_{y\in\mathscr{X}} (f(x)-f(y))^2 \Pi(x,y). \tag{1}$$

The following concentration result is a copy of Theorem 3.3 in [5].

Theorem 2 ([5], Theorem 3.3). *Let (Π,μ) be a reversible Markov chain on a finite or countable space $\mathscr{X}$ with a spectral gap $\lambda_1 > 0$. Then, whenever $f : \mathscr{X} \to \mathbb{R}$ is a function such that $|||f|||_\infty \le 1$, we have that f is integrable with respect to μ and for every $r \ge 0$,*

$$\mu(\{f \ge \textstyle\int f\,d\mu + r\}) \le 3e^{-r\sqrt{\lambda_1}/2}.$$

Let us now specialize to $\mathscr{X} = S_n$, the group of all permutations of n elements. The following transition kernel Π generates the 'random transpositions walk'.

$$\Pi(\pi,\pi') = \begin{cases} 1/n & \text{if } \pi' = \pi, \\ 2/n^2 & \text{if } \pi' = \pi\tau \text{ for some transposition } \tau, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It is not difficult to verify that the uniform distribution μ on S_n is the unique invariant measure for this kernel, and the pair (Π, μ) defines a reversible Markov chain.

Theorem 3 (Diaconis & Shahshahani [4], Corollary 4). *The spectral gap of the random transpositions walk on S_n is $2/n$.*

We are now ready to prove Theorem 1.

Proof of Theorem 1. Let π be a uniform random permutation of $\{1,\ldots,n\}$. Let $A = A(\pi) = M(\pi_1,\ldots,\pi_k;\pi_1,\ldots,\pi_k)$. Fix a point $x \in \mathbb{R}$. Let

$$f(\pi) := F_A(x).$$

Let Π be the transition kernel for the random transpositions walk defined in (2), and let $|||\cdot|||_\infty$ be defined as in (1).
Now, by Lemma 2.2 in Bai [1], we know that for any two Hermitian matrices A and B of order k,

$$\|F_A - F_B\|_\infty \le \frac{\operatorname{rank}(A-B)}{k}. \tag{3}$$

Let $\tau = (I,J)$ be a random transposition, where I,J are chosen independently and uniformly from $\{1,\ldots,n\}$. Multiplication by τ results in taking a step in the chain defined by Π. Now, for any $\sigma \in S_n$, the $k\times k$ Hermitian matrices $A(\sigma)$ and $A(\sigma\tau)$ differ at most in one column and one row, and hence $\operatorname{rank}(A(\sigma)-A(\sigma\tau)) \le 2$. Thus,

$$|f(\sigma) - f(\sigma\tau)| \le \frac{2}{k}. \tag{4}$$

Again, if I and J both fall outside $\{1,\ldots,k\}$, then $A(\sigma) = A(\sigma\tau)$. Combining this with (3) and (4), we get

$$|||f|||_\infty^2 = \frac{1}{2}\max_{\sigma\in S_n}\mathbb{E}(f(\sigma)-f(\sigma\tau))^2 \le \frac{1}{2}\left(\frac{2}{k}\right)^2\frac{2k}{n} \le \frac{4}{kn}.$$

Therefore, from Theorems 2 and 3, it follows that for any $r \ge 0$,

$$\mathbb{P}(|F_A(x) - F(x)| \ge r) \le 6\exp\left(-\frac{r\sqrt{2/n}}{2\sqrt{4/kn}}\right) = 6\exp\left(-\frac{r\sqrt{k}}{\sqrt{8}}\right). \tag{5}$$

The above result is true for any x. Now, if $F_A(x-) := \lim_{y\uparrow x} F_A(y)$, then by the bounded convergence theorem we have $\mathbb{E}F_A(x-) = \lim_{y\uparrow x} F(y) = F(x-)$. It follows that for every r,

$$\begin{aligned}\mathbb{P}(|F_A(x-) - \mathbb{E}F_A(x-)| > r) &\le \liminf_{y\uparrow x}\mathbb{P}(|F_A(y) - F(y)| > r)\\ &\le 6\exp\left(-\frac{r\sqrt{k}}{\sqrt{8}}\right).\end{aligned}$$

Since this holds for all r, the $>$ can be replaced by $\ge$. Similarly it is easy to show that F is a legitimate cumulative distribution function. Now fix an integer $l \ge 2$, and for $1 \le i < l$ let

$$t_i := \inf\{x : F(x) \ge i/l\}.$$

Let $t_0 = -\infty$ and $t_l = \infty$. Note that for each i, $F(t_{i+1}-) - F(t_i) \le 1/l$. Let

$$\Delta = (\max_{1 \le i < l} |F_A(t_i) - F(t_i)|) \vee (\max_{1 \le i < l} |F_A(t_i-) - F(t_i-)|).$$

Now take any $x \in \mathbb{R}$. Let i be an index such that $t_i \le x < t_{i+1}$. Then

$$F_A(x) \le F_A(t_{i+1}-) \le F(t_{i+1}-) + \Delta \le F(x) + 1/l + \Delta.$$

Similarly,

$$F_A(x) \ge F_A(t_i) \ge F(t_i) - \Delta \ge F(x) - 1/l - \Delta.$$

Combining, we see that

$$\|F_A - F\|_\infty \le 1/l + \Delta.$$

Thus, for any $r \ge 0$,

$$\mathbb{P}(\|F_A - F\|_\infty \ge 1/l + r) \le 12(l-1)e^{-r\sqrt{k/8}}.$$

Taking $l = [k^{1/2}] + 1$, we get for any $r \ge 0$,

$$\mathbb{P}(\|F_A - F\|_\infty \ge 1/\sqrt{k} + r) \le 12\sqrt{k}e^{-r\sqrt{k/8}}.$$

This proves the first claim of Theorem 1. To prove the second, using the above inequality, we get

$$\begin{aligned}\mathbb{E}\|F_A - F\|_\infty &\le \frac{1 + \sqrt{8}\log k}{\sqrt{k}} + \mathbb{P}\left(\|F_A - F\|_\infty \ge \frac{1 + \sqrt{8}\log k}{\sqrt{k}}\right)\\ &\le \frac{13 + \sqrt{8}\log k}{\sqrt{k}}.\end{aligned}$$

For the case of singular values, we proceed as follows. As before, we let π be a random permutation of $\{1, \ldots, n\}$; but here we define $A(\pi) = M(\pi_1, \ldots, \pi_k; 1, \ldots, n)$. Since singular values of A are just square roots of eigenvalues of AA^*, therefore

$$\|F_A - \mathbb{E}(F_A)\|_\infty = \|F_{AA^*} - \mathbb{E}(F_{AA^*})\|_\infty,$$

and so it suffices to prove a concentration inequality for F_{AA^*}. As before, we fix x and define

$$f(\pi) = F_{AA^*}(x).$$

The crucial observation is that by Lemma 2.6 of Bai [1], we have that for any two $k \times n$ matrices A and B,

$$\|F_{AA^*} - F_{BB^*}\|_\infty \le \frac{\operatorname{rank}(A-B)}{k}.$$

The rest of the proof proceeds exactly as before. $\square$

Acknowledgment. We thank the referees for helpful comments.

References

[1] BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** no. 3, 611–677. MR1711663

[2] BOBKOV, S. G. (2004). Concentration of normalized sums and a central limit theorem for noncorrelated random variables. *Ann. Probab.* **32** no. 4, 2884–2907. MR2094433

[3] BOBKOV, S. G. and TETALI, P. (2006). Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.* **19** no. 2, 289–336. MR2283379

[4] DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** no. 2, 159–179. MR0626813

[5] LEDOUX, M. (2001). *The concentration of measure phenomenon.* Amer. Math. Soc., Providence, RI. MR1849347

[6] RUDELSON, M. and VERSHYNIN, R. (2007). Sampling from large matrices: an approach through geometric functional analysis. *J. ACM* **54** no. 4, Art. 21, 19 pp. MR2351844