ERRATUM

Erratum to: Outliers in the spectrum of iid matrices with bounded rank perturbations

Terence Tao

Erratum to: Probab. Theory Relat. Fields (2013) 155:231–263 DOI 10.1007/s00440-011-0397-9

The demonstration of a key lemma in this is only valid in the sense of convergence in probability instead of in the almost sure sense. However, a refinement of the argument, which we give here, can be used to give convergence in the almost sure sense.

In this erratum we report an issue in the originally published article that was pointed out to us by David Renfrew and Sean O'Rourke. Specifically, Lemma 2.3 of that paper asserts a bound

$$\left\langle \left(\frac{1}{\sqrt{n}} X_n \right)^m u_n, v_n \right\rangle = o(1)$$

to hold in the almost sure sense, but the argument given, which is based on establishing the second moment bound

$$\mathbf{E} \left| \left\langle \left(\frac{1}{\sqrt{n}} X_n \right)^m u_n, v_n \right\rangle \right|^2 = O\left(\frac{1}{n} \right)$$

after truncating the atom distribution to be bounded, only establishes convergence in probability instead of in the almost sure sense (note that $1/n$ is not absolutely summable and so the Borel–Cantelli lemma does not apply here). Thus, the main theorems in the

---

T. Tao (✉)
Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA
e-mail: tao@math.ucla.edu

originally published article, are only proven in that paper in probability rather than in the almost sure sense.

However, the almost sure version of Lemma 2.3 can be recovered using the methods of proof of that lemma by obtaining[1] the fourth moment bound

$$\mathbf{E}\left|\left\langle\left(\frac{1}{\sqrt{n}}X_n\right)^m u_n, v_n\right\rangle\right|^4 = O(n^{-3/2}) \tag{1}$$

assuming bounded atom distribution. By Markov's inequality, this shows that $\langle(\frac{1}{\sqrt{n}}X_n)^m u_n, v_n\rangle = O(n^{-1/16})$ with probability $1 - O(n^{-5/4})$, which by the Borel–Cantelli lemma gives Lemma 2.3 in the almost sure sense for bounded atom distribution, and then the case of general atom distributions then follows by the usual truncation argument.

The bound $O(n^{-3/2})$ in (1) is not optimal; $O(n^{-2})$ should be the truth (cf. Proposition 4.1), but we do not attempt to optimise exponents here.

It remains to establish (1). Following the argument for Lemma 2.3 given in the originally published article, we first consider the case when the real and imaginary parts of the atom distribution x are independent gaussians. Then, as in the originally published article, we can reduce to the case when u_n and v_n are real unit vectors, in which case the distribution of $\langle(\frac{1}{\sqrt{n}}X_n)^m u_n, v_n\rangle$ is in fact independent of u_n and v_n. In particular, the left-hand side of (1) is then equal to

$$\mathbf{E}\left|\left\langle\left(\frac{1}{\sqrt{n}}X_n\right)^m u_n, w\right\rangle\right|^4$$

where w is a unit real vector drawn uniformly at random on S^{n-1} independently of X_n and u_n. However, a short computation in cylindrical coordinates (or Levy's concentration of measure theorem) shows that for any deterministic vector x, one has $\mathbf{E}|\langle x, w\rangle|^4 \ll \|x\|^4/n^2$ if w is a unit vector drawn uniformly from the sphere. Thus we may bound (1) by

$$O\left(\frac{1}{n^2}\mathbf{E}\left\|\left(\frac{1}{\sqrt{n}}X_n\right)^m u_n\right\|^4\right).$$

By Theorem 1.4, the inner expectation is $O(1)$, and the claim (1) follows in the Gaussian case.

To handle the non-Gaussian case, we argue as in the originally published article and establish the bound

$$\mathbf{E}\left|\left\langle\left(\frac{1}{\sqrt{n}}X_n\right)^m u_n, v_n\right\rangle\right|^4 - \mathbf{E}\left|\left\langle\left(\frac{1}{\sqrt{n}}G_n\right)^m u_n, v_n\right\rangle\right|^4 = O(n^{-3/2}).$$

---

[1] In the case when the atom distribution obeys a log-Sobolev inequality, one could also use concentration of measure techniques to obtain the desired result, and by matching moments with a distribution with a log-Sobolev inequality one could largely remove the log-Sobolev hypothesis, although a few distributions (and in particular Bernoulli distributions) have too rigid of a moment sequence in order to apply this strategy.

The left-hand side expands as

$$n^{-2m} \sum_{a_0,\ldots,a_m,b_0,\ldots,b_m,c_0,\ldots,c_m,d_0,\ldots,d_m \in \{1,\ldots,n\}} u_{n,a_0} \overline{v_{n,a_m}} u_{n,b_0} v_{n,b_m} u_{n,c_0} \overline{v_{n,c_m}} v_{n,d_0} v_{n,d_m}$$

$$\left(\mathbf{E} \prod_{j=1}^{m} x_{a_j,a_{j-1}} \overline{x_{b_j,b_{j-1}}} x_{c_j,c_{j-1}} \overline{x_{d_j,d_{j-1}}} \right.$$

$$\left. - \mathbf{E} \prod_{j=1}^{m} g_{a_j,a_{j-1}} \overline{g_{b_j,b_{j-1}}} g_{c_j,c_{j-1}} \overline{g_{d_j,d_{j-1}}} \right)$$

(note that the analogous formula in the originally published article has a typo in that only one product $\prod_{j=1}^{m}$ factor is present instead of two). Arguing as in the originally published article, we can bound this as

$$O\left(n^{-2m} \sum_{*} |u_{n,a_0}| |v_{n,a_m}| |u_{n,b_0}| |v_{n,b_m}| |u_{n,c_0}| |v_{n,c_m}| |u_{n,d_0}| |v_{n,d_m}| \right)$$

where $\sum_{*}$ denotes the sum over all tuples $a_0, \ldots, a_m, b_0, \ldots, b_m, c_0, \ldots, c_m, d_0, \ldots, d_m$ such that the ordered pairs

$$(a_{j-1}, a_j), (b_{j-1}, b_j), (c_{j-1}, c_j), (d_{j-1}, d_j) \tag{2}$$

for $j = 1, \ldots, m$ are such that each pair occurs at least twice, and at least one pair occurs three or more times; in particular there are at most $2m - 1$ distinct pairs. Our task is now to show that

$$\sum_{*} |u_{n,a_0}| |v_{n,a_m}| |u_{n,b_0}| |v_{n,b_m}| |u_{n,c_0}| |v_{n,c_m}| |u_{n,d_0}| |v_{n,d_m}| \ll n^{2m-3/2} \tag{3}$$

Suppose that $a_0, \ldots, a_m, b_0, \ldots, b_m, c_0, \ldots, c_m, d_0, \ldots, d_m$ are such that (2) is of the stated form. Let G be the unordered looped graph G with edges being the unordered pairs associated to (2) and with vertices being the elements of these pairs, and let r be the number of connected components of G. Then $1 \leq r \leq 4$, and G has at most $2m - 1$ edges and thus at most $2m + r - 1$ vertices.

We consider first the contribution $\sum_{**}$ of those tuples for which G has at most $2m + r - 2$ vertices; this is for instance the case if G contains a cycle or a looped edge, or has strictly fewer than $2m - 1$ edges. Then if one fixes a_0, b_0, c_0, d_0, then one has fixed at least one vertex in each component of G, leaving at most $2m - 2$ remaining vertices. Thus there are $O(2^{2m-2})$ choices for the remaining data $a_1, \ldots, a_m, b_1, \ldots, b_m, c_1, \ldots, c_m, d_1, \ldots, d_m$; summing over a_0, b_0, c_0, d_0 using the fact that u_n is a unit vector yields that

$$\sum_{**} |u_{n,a_0}|^2 |u_{n,b_0}|^2 |u_{n,c_0}|^2 |u_{n,d_0}|^2 \ll n^{2m-2}$$

and similarly

$$\sum_{**} |v_{n,a_m}|^2 |v_{n,b_m}|^2 |v_{n,c_m}|^2 |v_{n,d_m}|^2 \ll n^{2m-2}$$

and so by Cauchy–Schwarz the contribution of the $\sum_{**}$ tuples to (3) is acceptable.

Now consider the contribution $\sum_{***}$ of those tuples for which G has exactly $2m + r - 1$ vertices, but such that two of the a_0, b_0, c_0, d_0 are distinct elements of a common component of G. Then as in the $\sum_{**}$ case, fixing a_0, b_0, c_0, d_0 leaves only $O(2^{2m-2})$ choices for the remaining data, so that

$$\sum_{***} |u_{n,a_0}|^2 |u_{n,b_0}|^2 |u_{n,c_0}|^2 |u_{n,d_0}|^2 \ll n^{2m-2}$$

while we also have the cruder bound

$$\sum_{***} |v_{n,a_m}|^2 |v_{n,b_m}|^2 |v_{n,c_m}|^2 |v_{n,d_m}|^2 \ll n^{2m-1}$$

so by Cauchy–Schwarz this contribution is also acceptable. Similarly if G has $2m+r-1$ vertices but two of the a_m, b_m, c_m, d_m are distinct elements of a common component of G.

The only remaining contribution $\sum_{****}$ comes from the case when G has exactly $2m+r-1$ vertices, the a_0, b_0, c_0, d_0 agree whenever they lie in a common component of G, and a_m, b_m, c_m, d_m agree whenever they lie in a common component of G. As discussed previously, the requirement that G has exactly $2m + r - 1$ vertices forces G to be a forest (a union of r disjoint trees, with no cycles or looped edges) and to contain exactly $2m - 1$ edges. Among other things, this implies that the tuples (2) do not contain a loop (a, a), nor do these tuples contain a pair (a, b) together with its reversal (b, a). If $a_0, \ldots, a_m$ and $b_0, \ldots, b_m$ (for instance) lie in the same component of G, then we must then $a_0 = b_0$ and $a_m = b_m$, and then $a_i = b_i$ for all $0 \le i \le m$ (otherwise there would be a cycle, looped edge, or a pair (a, b) and its reversal). Thus each component has m edges, which is only consistent with the total edge count of $2m - 1$ if $r = 1$ and $m = 1$. But in this case the left-hand side of (3) simplifies to

$$\sum_{i \ne j} |u_{n,i}|^4 |v_{n,j}|^4$$

which is easily seen to be $O(1)$, so that (3) easily follows in this case.