

УДК 519.2

ЛОКАЛЬНЫЙ ЗАКОН МАРЧЕНКО–ПАСТУРА ДЛЯ ПРОРЕЖЕННЫХ ПРЯМОУГОЛЬНЫХ СЛУЧАЙНЫХ МАТРИЦ

© 2021 г. Ф. Гётце^{1,*}, Д. А. Тимушев^{2,**}, А.Н. Тихомиров^{2,***}

Представлено академиком РАН И.А. Ибрагимовым 11.08.2021 г.

Поступило 09.09.2021 г.

После доработки 09.09.2021 г.

Принято к публикации 27.10.2021 г.

Рассматриваются прореженные выборочные ковариационные матрицы с вероятностью прореживания $p_n \geq C_0 \log^{\frac{2}{\kappa}} n/n$, для некоторого $\kappa > 0$. В предположении, что распределение элементов матриц имеет конечный абсолютный момент порядка $4 + \delta$, $\delta > 0$, показано, что расстояние между преобразованиями Стилтеса эмпирической спектральной функции распределения и закона Марченко–Пастура имеет порядок $\log n(1/(nv) + 1/(np_n))$, где v – расстояние до действительной оси в комплексной плоскости.

Ключевые слова: распределение Марченко–Пастура, локальный режим, прореженные случайные матрицы, спектр случайной матрицы, преобразование Стилтеса

DOI: 10.31857/S2686954321060060

Рассмотрим серию случайных матриц $\mathbf{X}_{n,m} = \{X_{jk}^{(n,m)}\}$, $1 \leq j \leq n$, $1 \leq k \leq m$, элементы которых – независимые одинаково распределенные вещественные случайные величины, определенные на одном и том же вероятностном пространстве. Будем полагать, что $\mathbb{E}X_{jk}^{(n,m)} = 0$ и $\mathbb{E}|X_{jk}^{(n,m)}|^2 = 1$. Назовем выборочной ковариационной матрицей матрицу вида

$$\mathbf{Z} = \frac{1}{m} \mathbf{X}_{n,m} \mathbf{X}_{n,m}^*.$$

Впервые выборочные ковариационные матрицы были рассмотрены в работе Вишарта [1]. В контексте задач многомерного статистического анализа строки матрицы $\mathbf{X}_{n,m}$ представляют собой выборку объема n , состоящую из m -мерных векторов наблюдений. При росте размеров матрицы $\mathbf{X}_{n,m}$ к бесконечности таким образом, что отношение n/m сходится к некоторой константе, эмпирическая спектральная функция распределения матрицы $\mathbf{Z}$ (функция равномерного распределе-

ния на множестве сингулярных чисел) сходится к некоторой неслучайной функции распределения, плотность которой была явно найдена в работе Марченко и Пастура [2]. Факт такой сходимости принято называть законом Марченко–Пастура, а само предельное распределение – распределением Марченко–Пастура.

Важным частным случаем выборочных ковариационных матриц являются так называемые прореженные выборочные ковариационные матрицы (sparse sample covariance matrices). Пусть $\xi_{jk}^{(n,m)}$ – семейство независимых бернуллиевских случайных величин, независимых от $X_{jk}^{(n,m)}$, с $\mathbb{E}\xi_{ij}^{(n,m)} = p_n$. Везде далее мы будем опускать верхний индекс $^{(n,m)}$. Для прореженной матрицы $\mathbf{W} = \{X_{jk}\xi_{jk}\}$, $1 \leq j \leq n$, $1 \leq k \leq m$, рассмотрим прореженную выборочную ковариационную матрицу

$$\mathbf{Z} = \frac{1}{mp_n} \mathbf{W} \mathbf{W}^*.$$

В дальнейшем мы будем рассматривать только параметр n , а про параметр m будем предполагать, что он меняется так, что

$$\lim_{n,m \rightarrow \infty} \frac{n}{m} =: y \in (0,1).$$

Особый интерес в моделях с прореженными выборочными ковариационными матрицами представляет случай, когда вероятность прореживания

¹ Билефельдский университет, Билефельд, Германия

² Коми научный центр Уральского отделения Российской академии наук, Сыктывкар, Россия

*E-mail: goetze@math.uni-bielefeld.de

**E-mail: timushev@ipm.komisc.ru

***E-mail: tikhomirov@ipm.komisc.ru

p_n стремится к нулю, с ростом n . В частности, если p_n убывает так, что $np_n \rightarrow \infty$, то для прореженной выборочной ковариационной матрицы остается справедливым упомянутый выше закон Марченко–Пастура.

Прореженные выборочные ковариационные матрицы возникают во многих прикладных задачах, например, при исследовании свойств двудольного случайного графа — графа, вершины которого можно разделить на две группы, в каждой из которых вершины не соединены друг с другом, а наличие ребер — случайно, с некоторой вероятностью p_n .

Прореженные случайные матрицы изучались разными авторами (см. [3–8] и литературу к ним). Поведение сингулярных чисел и собственных векторов прореженных случайных матриц значительно зависит от вероятности прореживания, и это не позволяет простым образом распространить на них результаты, полученные ранее для непрореженного ($p_n = 1$) случая. В частности, непростой задачей является изучение поведения минимального сингулярного числа. Прогресс в решении этой проблемы для случая вигнеровских матриц был достигнут в работе [9].

Обозначим через $s_1^2 \geq s_2^2 \geq \dots \geq s_n^2$ собственные числа матрицы $\mathbf{Z}$ (соответственно, $s_1, \dots, s_n$ — сингулярные числа матрицы $\frac{1}{\sqrt{mp_n}} \mathbf{W}$).

Назовем “симметризацией” некоторой положительной случайной величины η^2 случайную величину $\varepsilon|\eta|$, где ε — независимая от η случайная величина, равновероятно принимающая значения 1 и -1 . Симметризованная эмпирическая спектральная функция распределения матрицы $\mathbf{Z}$ имеет вид

$$F_n(x) = \frac{1}{2n} \sum_{j=1}^n (\mathbb{I}\{s_j < x\} + \mathbb{I}\{-s_j < x\})$$

и является спектральной функцией распределения эрмитовой матрицы

$$\mathbf{V} = \begin{bmatrix} \mathbf{O} & \frac{1}{\sqrt{mp_n}} \mathbf{W} \\ \frac{1}{\sqrt{mp_n}} \mathbf{W}^* & \mathbf{O} \end{bmatrix},$$

где $\mathbf{O}$ — матрица с нулевыми элементами.

Для произвольного $\delta > 0$ определим величину $\kappa = \kappa(\delta) := \frac{\delta}{2(4+\delta)}$. Мы будем предполагать, что вероятность прореживания p_n и моменты элементов матрицы X_{ij} удовлетворяют ряду условий:

- условие (C0): для некоторого $c_0 > 0$ и всех $n \geq 1$,

$$np_n \geq c_0 \log^{\frac{2}{\kappa}} n;$$

- условие (C1): для некоторого $\delta > 0$,

$$\mu_{4+\delta} := \mathbb{E}|X_{11}|^{4+\delta} < \infty;$$

- условие (C2): существует такая константа $c_1 > 0$, что для всех $1 \leq j \leq n, 1 \leq k \leq m$ выполнено

$$|X_{jk}| \leq c_1 (np_n)^{\frac{1-\kappa}{2}}.$$

Важной задачей в изучении свойств случайной матрицы является оценивание ее наименьшего положительного и наибольшего сингулярных чисел. Подобные задачи решались в ряде работ, посвященных прореженным матрицам. Так, в [6], в предположении $np_n \geq n^\delta$, для сколь угодно малого $\delta > 0$, и $\mathbb{E}|X_{jk}|^q \leq (C_1 q)^{C_2 q}$, для всех $q \geq 3$, показано, что найдутся такие положительные константы $K = K(\delta, C_1, C_2)$, $Q = Q(\delta, C_1, C_2)$ и $C_3 = C_3(\delta, C_1, C_2)$, что выполнено неравенство $\Pr\{s_1 \geq K\} \leq C_3 n^{-Q}$. В работе [7], в предположении субгауссовости элементов матрицы и $np_n \geq C_1 \log n$, получена оценка $\Pr\{s_n \leq n^{-C_2}\} \leq C_3 n^{-Q}$, для некоторых положительных констант $Q = Q(C_1)$, $C_2 = C_2(C_1)$ и $C_3 = C_3(C_1)$. В нашей работе доказаны следующие результаты.

Теорема 1. *Предположим, что условия (C0)–(C2) выполнены. Тогда для любого $Q \geq 1$ найдутся такие положительные константы*

$$K_1 = K_1(Q, \delta, \mu_{4+\delta}, c_0, c_1), K_2 = K_2(Q, \delta, \mu_{4+\delta}, c_0, c_1) \\ \text{и } C_0 = C_0(Q, \delta, \mu_{4+\delta}, c_0, c_1),$$

что

$$\Pr\{s_1 \geq K_1\} \leq C_0 n^{-Q}, \quad \Pr\{s_n \leq K_2\} \leq C_0 n^{-Q}.$$

Рассмотрим преобразование Стильтеса симметризованной эмпирической функции распределения $F_n(x)$:

$$s_n(z) = \frac{1}{n} \sum_{j=1}^n \frac{z}{z^2 - s_j^2}.$$

Преобразование Стильтеса симметризованного распределения Марченко–Пастура имеет вид

$$S_y(z) = \frac{-z + \frac{1-y}{z} + \sqrt{(z^2 - a^2)(z^2 - b^2)}}{2y},$$

где $a = 1 - \sqrt{y}$, $b = 1 + \sqrt{y}$.

Следуя работам Х.-Ц. Яу, мы будем называть локальным законом Марченко–Пастура равномерную по $z = u + iv$ сходимость

$$|s_n(z) - S_y(z)| \rightarrow 0$$

в области $v \geq n^{-1}\alpha_n$, $u_- \leq |u| \leq u_+$, где α_n “медленно” растёт к бесконечности с ростом n ($\alpha_n \sim \log^\beta n$ или $\alpha_n \ll n^\gamma$, с некоторыми $\beta, \gamma > 0$), а u_- и u_+ — некоторые положительные числа.

Введем обозначения $\Lambda_n(z) := s_n(z) - S_y(z)$, $b(z) = z - \frac{1-y}{z} + 2yS_y(z)$, $x \wedge y = \min\{x, y\}$.

Теорема 2. *Предположим, что условия (C0)–(C2) выполнены. Тогда для любого $Q \geq 1$ существуют такие константы $C_0 = C_0(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, $C_1 = C_1(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, $K = K(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, что для всех $z = u + iv$ из области $(a-v)_+ \leq |u| \leq b+v$ и $v \geq C_1 n^{-1} \log^4 n$ выполнено*

$$\Pr \left\{ |\Lambda_n| \geq K \left(\frac{\text{Im} b(z)}{nv|b(z)|} + \frac{1}{np_n|b(z)|} \right) \right\} \leq C_0 n^{-Q}.$$

Следующий результат относится к $\text{Im} \Lambda_n(z)$.

Теорема 3. *Предположим, что условия (C0)–(C2) выполнены. Тогда для любого $Q \geq 1$ существуют такие константы $C_0 = C_0(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, $C_1 = C_1(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, $C_2 = C_2(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, $K = K(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, что для всех $\alpha, u_0 > 0$ и $z = u + iv$ из области $0 < \alpha \leq |u| \leq u_0$, $v \geq C_1 n^{-1} \log^4 n$ и $|b(z)| \geq C_2/(nv)$ выполнено*

$$\Pr \left\{ |\text{Im} \Lambda_n| \geq K \left(\frac{\text{Im} b(z)}{nv|b(z)|} + \frac{1}{np_n|b(z)|} \right) \right\} \leq C_0 n^{-Q}.$$

Следствие 1. *Обозначим через $\mathbf{u}_j$ собственный вектор матрицы $\mathbf{Z}$, отвечающий собственному значению s_j^2 . Тогда, в предположениях (C0)–(C2), для любого $Q \geq 1$ найдутся такие константы $C_0 = C_0(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, $C_1 = C_1(Q, \delta, \mu_{4+\delta}, c_0, c_1)$, что*

$$\Pr \left\{ \max_{1 \leq j, k \leq n} |u_{jk}| \leq C_1 \log^2 n / \sqrt{n} \right\} \leq C_0 n^{-Q}.$$

Теорема 5. *Предположим, что выполнены только условия (C0) и (C1). Тогда для любого $0 < \gamma < \frac{b-a}{2}$ существуют такие константы $C_0 = C_0(Q, \delta, \mu_{4+\delta}, c_0, \gamma)$, $C_1 = C_1(Q, \delta, \mu_{4+\delta}, c_0, \gamma)$, $C_2 = C_2(Q, \delta, \mu_{4+\delta}, c_0, \gamma)$, что*

$$\Pr \left\{ |\Lambda_n| \leq C_1 \log n \left(\frac{1}{nv} + \frac{1}{np_n} \right) \right\} \geq 1 - \frac{C_0 \log n}{np_n},$$

для всех $z = u + iv$ из области $v \geq C_2 n^{-1} \log^4 n$, $a + \gamma \leq |u| \leq b - \gamma$.

Замечание 1. В работе [6], в предположении, что моменты элементов матрицы удовлетво-

ряют условию $\mathbb{E}|X_{jk}|^q \leq (C_1 q)^{C_2 q}$, для всех $q \geq 3$, в области $\{z = u + iv: a^2/2 \leq u \leq b^2 + 1, 0 < v < 3\}$ показано, что для несимметризованных преобразований Стильеса, для сколь угодно малого $\varepsilon > 0$ и произвольного $D > 0$ найдется такая положительная константа $C_3 = C_3(\varepsilon, D, C_1, C_2)$, что для достаточно больших n выполнено неравенство

$$\Pr \left\{ |\Lambda_n| > n^\varepsilon \left(\frac{1}{nv} + \frac{1}{np_n} \right) \right\} \leq C_3 n^{-D}.$$

Следствие 2. *В условиях теоремы 4 найдутся такие константы $C_0 = C_0(Q, \delta, \mu_{4+\delta}, c_0, \gamma)$, $C_1 = C_1(Q, \delta, \mu_{4+\delta}, c_0, \gamma)$, что*

$$\Pr \left\{ \sup_x |F_n(x) - G_y(x)| \leq \frac{C_1 \log n}{np_n} \right\} \geq 1 - \frac{C_0 \log n}{np_n}.$$

Следствие 3. *Предположим, что выполнены условия (C1), (C2) и $np_n \geq c'_0 n^{2/3}$, $c'_0 > 0$. Пусть γ_j , $j = 1, \dots, n$, — квантили порядка $(1 - j/n)$ распределения $G_y(x)$. Тогда для некоторого малого $\alpha > 0$ и $b - \alpha \leq \gamma_j \leq b$, для любого $Q \geq 1$ найдутся такие константы $C_0 = C_0(Q, \delta, \mu_{4+\delta}, c'_0, c_1, \alpha)$, $C_1 = C_1(Q, \delta, \mu_{4+\delta}, c'_0, c_1, \alpha)$, что*

$$\Pr \left\{ |s_j - \gamma_j| \geq C_1 \log n \left(\frac{1}{(j \wedge (n-j))^{1/3} n^{2/3}} + \frac{1}{np_n} \right) \right\} \leq C_0 n^{-Q}.$$

Замечание 2. В работе [6], в предположении, что моменты элементов матрицы удовлетворяют условию $\mathbb{E}|X_{jk}|^q \leq (C_1 q)^{C_2 q}$, для всех $q \geq 3$, при $np_n \geq C_3 n^{2/3}$ показано, что для сколь угодно малого $\varepsilon > 0$ и произвольного $D > 0$ найдется такая положительная константа $C_4 = C_4(\varepsilon, D, C_1, C_2, C_3)$, что для достаточно больших n , при $b - \alpha \leq \gamma_j \leq b$, выполнено неравенство

$$\Pr \left\{ |s_j - \gamma_j| > n^\varepsilon (j^{-1/3} n^{-2/3} + (np_n)^{-1}) \right\} \leq C_4 n^{-D}.$$

Схема доказательства. Для оценки спектральной нормы матрицы $\frac{1}{\sqrt{mp_n}} \mathbf{W}$ используется результат Сегинера [10].

В оценке наименьшего сингулярного числа мы следуем методу Рудельсона и Вершинина (см. [11]), разработанному для прямоугольных матриц, дополняя его рекуррентной процедурой (см. [12]), которая позволяет получить необходимую оценку.

Для доказательства неравенства делокализации используется представление

$$\operatorname{Im} R_{jj} = \frac{1}{n} \sum_{k=1}^n \frac{v}{|z - s_j|^2} |u_{jk}|^2 = \int \operatorname{Im} \frac{1}{z - x} dF_{nj}(x),$$

где $F_{nj}(x) = \sum_{k=1}^n |u_{jk}|^2 \mathbb{I}\{s_j \leq x\}$. Отсюда, для произвольного $v > 0$, имеем

$$\max_k \{|u_{jk}|^2\} \leq 2v \sup_u \operatorname{Im} R_{jj}(u + iv).$$

Для исследования преобразования Стилтеса прореженной матрицы используется стандартное представление диагональных элементов резольвентной матрицы $\mathbf{R}(z) = (\mathbf{Z} - z\mathbf{I})^{-1}$:

$$R_{jj}(z) = S_y(z)(1 - \varepsilon_j R_{jj}(z) + y\Lambda_n R_{jj}(z)),$$

$$R_{l+n,l+n}(z) = \left(yS_y(z) - \frac{1-y}{z} \right) (1 - \varepsilon_{l+n} R_{l+n,l+n}(z) + y\Lambda_n(z) R_{l+n,l+n}(z)),$$

с некоторыми (см. [13]) поправочными членами ε_j , $j = 1, \dots, n$, и ε_{l+n} , $l = 1, \dots, m$. На первом этапе получаем оценку $\mathbb{E}|R_{jk}|^q \leq C^q$, для некоторого $C > 0$, при $q \sim \log n$. Для этого используется так называемый “метод мультипликативного спуска”, основанный на неравенстве

$$|R_{jj}(u + iv)| \leq s |R_{jj}(u + isv)|, \quad s \geq 1.$$

Для оценки величины Λ_n используется представление $\Lambda_n(z) = T_n(z)/b_n(z)$, где

$$T_n(z) := \frac{1}{n} \sum_{j=1}^n \varepsilon_j R_{jj},$$

$$b_n(z) := z - \frac{1-y}{z} + 2yS_y(z) + y\Lambda_n(z).$$

СПИСОК ЛИТЕРАТУРЫ

1. *Wishart J.* The generalised product moment distribution in samples from a normal multivariate population // *Biometrika*. 1928. V. 20A. № 1–2. P. 32–52.
2. *Марченко В.А., Пастур Л.А.* Распределение собственных значений в некоторых ансамблях случайных матриц // *Матем. сб.* 1967. Т. 72. № 4. С. 507–536.
3. *Erdős L., Knowles A., Yau H.-T., Yin J.* Spectral statistics of Erdős–Rényi graphs I: Local semicircle law // *Ann. Probab.* 2013. V. 41. № 3B. P. 2279–2375.
4. *Erdős L., Knowles A., Yau H.-T., Yin J.* Spectral statistics of Erdős–Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues // *Comm. Math. Phys.* 2012. V. 314. № 3. P. 587–640.
5. *Lee J.O., Schnelli K.* Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population // *Ann. Appl. Probab.* 2016. V. 26. № 6. P. 3786–3839.
6. *Hwang J.Y., Lee J.O., Schnelli K.* Local law and Tracy–Widom limit for sparse sample covariance matrices // *Ann. Appl. Probab.* 2019. V. 29. № 5. P. 3006–3036.
7. *Hwang J.Y., Lee J.O., Yang W.* Local law and Tracy–Widom limit for sparse stochastic block models // *Bernoulli*. 2020. V. 26. № 3. P. 2400–2435.
8. *Lee J.O., Schnelli K.* Local law and Tracy–Widom limit for sparse random matrices // *Probab. Theory Relat. Fields*. 2018. V. 171. P. 543–616.
9. *Tikhomirov K., Youssef P.* Outliers in spectrum of sparse Wigner matrices // *Random Structures and Algorithms*. 2020. V. 58. № 3. P. 517–605.
10. *Seginer Y.* The expected norm of random matrices // *Combinatorics, Probability and Computing*. 2000. V. 9. № 2. P. 149–166.
11. *Rudelson M., Vershynin R.* Smallest singular value of a random rectangular matrix // *Communications on Pure and Applied Mathematics*. 2009. V. 62. № 12. P. 1707–1739.
12. *Götze F., Tikhomirov A.N.* The circular law for random matrices // *Ann. Probab.* 2010. V. 38. № 4. P. 1444–1491.
13. *Götze F., Tikhomirov A.N.* Rate of Convergence of the Expected Spectral Distribution Function to the Marchenko–Pastur Law // Preprint. 2014. <https://arxiv.org/abs/1412.6284>.

LOCAL MARCHENKO–PASTUR LAW FOR SPARSE RECTANGULAR RANDOM MATRICES

F. Götze^a, A. N. Tikhomirov^b, and D. A. Timushev^b

^a *Bielefeld University, Bielefeld, Germany*

^b *Komi Science Center of Ural Division of Russian Academy of Sciences, Syktyvkar, Russia*

We consider sparse sample covariance matrices with sparse probability $p_n \geq c_0 \log^{\frac{2}{\kappa}} n/n$, with $\kappa > 0$. Assuming that the distribution of matrix elements has a finite absolute moment of order $4 + \delta$, $\delta > 0$, it is shown that the distance between the Stieltjes transforms of the empirical spectral distribution function and the Marchenko–Pastur law is of order $\log n(1/(nv) + 1/(np_n))$, where v is the distance to the real axis in the complex plane.

Keywords: random matrices, sparse matrices, local Marchenko–Pastur law, Stieltjes transform