Electron. Commun. Probab. **19** (2014), no. 33, 1–9.
DOI: 10.1214/ECP.v19-3121
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Local semicircle law with imprimitive variance matrix

Oskari Ajanki* László Erdős† Torben Krüger‡

Abstract

We extend the proof of the local semicircle law for generalized Wigner matrices given in [4] to the case when the matrix of variances has an eigenvalue -1. In particular, this result provides a short proof of the optimal local Marchenko-Pastur law at the hard edge (i.e. around zero) for sample covariance matrices $\mathbf{X}^*\mathbf{X}$, where the variances of the entries of $\mathbf{X}$ may vary.

Keywords: Generalised Wigner matrices; generalised random sample covariance matrices; hard edge; local semicircle law.
AMS MSC 2010: 15B52 ; 60B20.
Submitted to ECP on November 8, 2013, final version accepted on May 22, 2014.
Supersedes arXiv:1311.2016.

1 Model and results

The local semicircle law on the local distribution of eigenvalues of large Wigner matrices has been the basic technical input in the recent works on the Wigner-Dyson-Gaudin-Mehta universality (see [6] and references therein). The analysis was extended to generalized Wigner matrices [7, 4] but always practically assuming that the matrix of variances is primitive [1], in particular -1 is not in its spectrum. This assumption naturally holds for random band matrices that were the main motivation to generalize Wigner matrices in [7]. However, some important matrices with a certain block structure do not satisfy this condition. Most notable example is the $2N \times 2N$ matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{0} & \mathbf{X}^* \\ \mathbf{X} & \mathbf{0} \end{bmatrix} \tag{1.1}$$

where the $N \times N$ matrix $\mathbf{X}$ has independent entries. The matrix $\mathbf{H}$ is the linearization of the of the sample covariance matrix $\mathbf{X}^*\mathbf{X}$. In this paper we show how to remove the primitivity assumption in [4].

We consider generalized $N \times N$ hermitian or symmetric Wigner matrix $\mathbf{H} = (h_{ij})_{i,j=1}^N$ with independent entries (up to the symmetry constraint $\mathbf{H} = \mathbf{H}^*$) such that

$$\mathbb{E}\, h_{ij} \,=\, 0\,, \qquad \text{and} \qquad s_{ij} := \mathbb{E}\,|h_{ij}|^2 < \infty\,. \tag{1.2}$$

*IST Austria, Am Campus 1, A-3400 Klosterneuburg, Austria. Partially supported by ERC Advanced Grant "RANMAT", No. 338804. E-mail: oskari.ajanki@ist.ac.at
†IST Austria, Am Campus 1, A-3400 Klosterneuburg, Austria. Partially supported by ERC Advanced Grant "RANMAT", No. 338804. E-mail: lerdos@ist.ac.at
‡IST Austria, Am Campus 1, A-3400 Klosterneuburg, Austria. Partially supported by SFB-TR 12 Grant of the German Research Council. E-mail: torben.krueger@ist.ac.at

[1] A non-negative $d \times d$-matrix $\mathbf{M}$ is said to be *primitive* (cf. Definition 4.1 of [1]) if there exists an integer k such that every element of $\mathbf{M}^k$ is strictly positive.

We assume that all moments are bounded in the sense that,

$$\mathbb{E}\left|\frac{h_{ij}}{s_{ij}^{1/2}}\right|^p < C_p\,, \qquad \forall\, p<\infty\,, \tag{1.3}$$

with constants C_p independent of N. In order to avoid unnecessary clutter we have suppressed the N-dependence in the notations, e.g., we use $\mathbf{H}$ and $\mathbf{S}$ to refer to the sequences of matrices $\mathbf{H}^{(N)} = (h_{ij}^{(N)})_{i,j=1}^N$ and $\mathbf{S}^{(N)} = (s_{ij}^{(N)})_{i,j=1}^N$, respectively.

Besides the natural constraints, $\mathbf{S}^{\mathrm{T}} = \mathbf{S}$, $s_{ij} \geq 0$, we make the following additional assumptions on the variance matrix:

(A1) **Boundedness:** There exists a sequence $N^\delta \leq M = M_N \leq N$, with $\delta > 0$, such that

$$0 \,\leq\, s_{ij} \,\leq\, M^{-1}\,; \tag{1.4}$$

(A2) **Constant row sums:** $\mathbf{S}$ is (double) stochastic:

$$\sum_{j=1}^N s_{ij} \,=\, 1\,, \qquad \forall\, i = 1,\dots,N\,; \tag{1.5}$$

(A3) **Isolated extremum eigenvalues:** There exists (N-independent) constant $0 < \rho < 1$, such that

$$\mathrm{Spec}(\mathbf{S}) \subset \{-1\} \cup [-\rho,\rho] \cup \{+1\}\,. \tag{1.6}$$

This setup is similar to that in [4], except that here we explicitly allow -1 in the spectrum of $\mathbf{S}$. This allows us to consider $\mathbf{S}$ which contain imprimitive irreducible components. The results in [4] practically excluded this case since the estimates became unstable, see Section 7 of [4]. The main observation of this paper is that this instability is not present.

The relaxation of the irreducibility condition is elementary algebra (cf. Lemma 2.1 below), and this extension was already mentioned in [4]. However, the inclusion of -1's in the spectrum of $\mathbf{S}$ requires a new algebraic identity that is stated as Lemma 2.2 below. We will show here how to incorporate this identity into the proof given in [4] with minor modifications.

The condition (A2) guarantees that the diagonal elements of the resolvent matrix,

$$\mathbf{G}(z) \,:=\, \frac{1}{\mathbf{H}-z}\,, \qquad z := E + \mathrm{i}\eta\,, \quad E\in\mathbb{R}\,,\; \eta>0\,, \tag{1.7}$$

converge towards the Stieltjes transform

$$m(z) \,=\, \frac{-z \,+\, \sqrt{z^2-4}}{2} \tag{1.8}$$

of the Wigner semicircle law, $\varrho(x) = (2\pi)^{-1}\sqrt{\max\{4-x^2,0\}}$, as N approaches infinity.

In order to state this main result, we recall the concept of stochastic domination (Definition 2.1 in [4]). We say that a (sequence) of random variables $X = X^{(N)}$ is stochastically dominated by another (sequence) of random variables $Y = Y^{(N)}$, in notation $X \prec Y$, if for any $\varepsilon, D > 0$ there exists $N_0 = N_0(\varepsilon, D) < \infty$ such that

$$\mathbb{P}\big\{X > N^\varepsilon Y\big\} \leq\, N^{-D} \qquad \forall\, N \geq N_0\,.$$

If X, Y depend on some other parameters (like z or labels like i,j), then the definition is always taken uniform in these parameters (i.e. N_0 depends only on $\varepsilon, D > 0$). The notation $X = \mathcal{O}_\prec(Y)$ means same as $|X| \prec Y$.

Theorem 1.1 (Main Result)**.** *Suppose* $\mathbf{S}$ *satisfies the assumptions (A1)–(A3), and denote*

$$\mathbb{D}(\gamma) \,=\, \mathbb{D}^{(N)}(\gamma) \,:=\, \big\{ z : \, |z| \le 10\,,\, \mathrm{Im}\, z \ge M_N^{-1+\gamma} \big\}\,, \qquad \gamma > 0\,. \tag{1.9}$$

Then for any fixed $\gamma > 0$*, the resolvent* $\mathbf{G}(z) = (G_{ij}(z))_{i,j=1}^N$ *of* $\mathbf{H}$ *satisfies the local estimates*

$$\max_{i,j=1}^{N} \big| G_{ij}(z) - m(z)\delta_{ij} \big| \;\prec\; \sqrt{\frac{\mathrm{Im}\, m(z)}{M\eta}} + \frac{1}{M\eta} \tag{1.10a}$$

$$\Big| \frac{1}{N} \mathrm{Tr}\, \mathbf{G}(z) - m(z) \Big| \;\prec\; \frac{1}{M\eta}\,, \tag{1.10b}$$

uniformly in $z = E + \mathrm{i}\eta \,\in\, \mathbb{D}(\gamma)$*.*

Moreover, outside the spectrum (1.10b) *can be strengthened by introducing the distance,* $\kappa := \max\{\,|E-2|, |E+2|\,\}$*, of* E *from the spectral edges:*

$$\Big| \frac{1}{N} \mathrm{Tr}\, \mathbf{G}(z) - m(z) \Big| \;\prec\; \frac{1}{M(\kappa+\eta)} \,+\, \frac{1}{(M\eta)^2\sqrt{\kappa+\eta}}\,. \tag{1.11}$$

This estimate is also uniform in $z = E + \mathrm{i}\eta \in \mathbb{D}(\gamma)$ *as long as the constraints* $|E| \ge 2$ *and* $\eta\sqrt{\kappa+\eta} \ge M^{-1+\gamma}$ *are satisfied.*

This theorem is a generalization of the following result.

Theorem 1.2 ([4])**.** *Assume that* $\mathbf{S}$ *satisfies (A1)–(A2), but (A3) is strengthened to* $\mathrm{Spec}(\mathbf{S}) \subset [-\rho, \rho] \cup \{+1\}$*. Then conclusions of Theorem 1.1 hold true.*

Theorem 1.2 is a special case of the more general Theorem 2.3 of [4] which also covers the cases $\mathrm{Spec}(\mathbf{S}) \subset [-\rho_-^{(N)}\,, \rho_+^{(N)}] \cup \{+1\}$ where the spectral gaps $1-\rho_\pm^{(N)}$ may close at certain rates as $N \to \infty$. However, the estimates near $E = 0$ in [4] deteriorated if the smallest eigenvalue approached -1, and in particular -1 was not allowed belong to the spectrum. The condition (A3) rules out closing of the upper gap, and hence the spectral domain $\widetilde{\mathbb{S}}(\gamma)$, defined by formulas (2.14) and (2.17) in [4], has been replaced here by the simpler set $\mathbb{D}(\gamma)$. It is straighforward to extend Theorem 1.1 to the entire set $\widetilde{\mathbb{S}}(\gamma)$ in the spirit of Theorem 2.3 in [4] but for brevity of this note we refrain from doing so.

Theorem 1.1 directly implies a rigidity result for the increasingly ordered eigenvalues $(\lambda_\alpha)_{\alpha=1}^N$ of $\mathbf{H}$ in terms of the N-th quantiles $(\gamma_\alpha)_{\alpha=1}^N$ of the semicircle density:

$$|\,\lambda_\alpha - \gamma_\alpha| \;\prec\; \frac{1}{M}\Big(\frac{N}{\widehat{\alpha}}\Big)^{1/3}, \quad \text{when} \quad \widehat{\alpha} := \min\{\alpha, N+1-\alpha\} \;\ge\; NM^{-1+\varepsilon}\,, \tag{1.12}$$

with $\varepsilon > 0$ arbitrary. See Theorem 7.6 in [4] for a proof in a more general setup and for the estimates on the extreme eigenvalues.

We remark that there have been many results on local semicircle laws prior to [4], in fact most methods used in [4] stem from [7, 8, 9]. See [4] for a complete account of the history and for the most concise general proof.

Finally, we mention a simple application. The eigenvalues of $\mathbf{H}$ in (1.1) generically come in pairs, $\pm\lambda$, (see (2.7) below) and their squares λ^2 are the eigenvalues of the sample covariance matrices $\mathbf{X}\mathbf{X}^*$ and $\mathbf{X}^*\mathbf{X}$. We assume that the elements of the square matrix $\mathbf{X}$ are independent, centred and their variances are chosen such that $\mathbf{S}$ and $\mathbf{H}$ satisfy (1.3)–(1.6) (note that -1 is an eigenvalue of $\mathbf{S}$). Under these conditions, Theorem 1.1 can be directly used to estimate the resolvent matrix elements and the trace of $\mathbf{X}^*\mathbf{X}$. Indeed, by applying the Schur formula to the $N \times N$-block decomposition

$$\mathbf{G}(z) \,=\, \frac{1}{\mathbf{H}-z} \,=\, \begin{bmatrix} -z & \mathbf{X}^* \\ \mathbf{X} & -z \end{bmatrix}^{-1} =: \begin{bmatrix} \mathbf{G}_{11}(z) & \mathbf{G}_{12}(z) \\ \mathbf{G}_{21}(z) & \mathbf{G}_{22}(z) \end{bmatrix}\,,$$

we see that the blocks on the diagonal equal:

$$\mathbf{G}_{11}(z) \;=\; \frac{z}{\mathbf{X}^*\mathbf{X} - z^2}\,, \qquad \mathbf{G}_{22}(z) \;=\; \frac{z}{\mathbf{X}\mathbf{X}^* - z^2}\,.$$

Thus Theorem 1.1 implies that the local Marchenko-Pastur law holds in the critical "hard-edge" case, when the limiting density

$$\varrho_{\mathrm{MP}}(x) \;:=\; \frac{1}{2\pi}\sqrt{\frac{\max\{4-x,0\}}{x}}\,, \tag{1.13}$$

is singular at the origin. In fact $\varrho_{\mathrm{MP}}(x) = x^{-1/2}\varrho(x^{1/2})$, $x > 0$, where ϱ is the Wigner semicircle density. By denoting the Stieltjes transform of the Marchenko-Pastur law (1.13) by m_{MP} and and writing $w := z^2$, an elementary calculation from Theorem 1.1 yields the following result.

Corollary 1.3 (Local Marchenko-Pastur law at the hard edge)**.** *Under the conditions on* $\mathbf{X}$ *above, we have for any fixed* $\gamma > 0$*,*

$$\max_{i,j=1}^{N}\left|\left(\frac{1}{\mathbf{X}^*\mathbf{X} - w}\right)_{ij} - m_{\mathrm{MP}}(w)\delta_{ij}\right| \;\prec\; \sqrt{\frac{\operatorname{Im} m_{\mathrm{MP}}(w)}{M\operatorname{Im} w}} + \frac{1}{M\operatorname{Im} w} \tag{1.14a}$$

$$\left|\frac{1}{N}\operatorname{Tr}\frac{1}{\mathbf{X}^*\mathbf{X} - w} - m_{\mathrm{MP}}(w)\right| \;\prec\; \frac{1}{M\operatorname{Im} w}\,, \tag{1.14b}$$

uniformly in $w \in \mathbb{C}$ *satisfying* $|w| \le 100$ *and* $\operatorname{Im} w \ge \sqrt{|\operatorname{Re} w|}\, M^{-1+\gamma}$*.*

The estimate outside of the spectrum (1.11) and the rigidity bound (1.12) can also be directly translated to the similar statements for the sample covariance matrices.

We remark that local Marchenko-Pastur law on the smallest local scale was first proven in [5] away from the critical case. The hard-edge case was independently considered in [3] and in [2], the latter providing an optimal error bound. Both works dealt with the case when the variances $\mathbb{E}\,|x_{ij}|^2$ are constant, the above corollary extends the result to the case of non-constant variances.

Acknowledgments. The authors are grateful to Ofer Zeitouni who pointed out the importance of removing the primitivity condition on $\mathbf{S}$ for various applications. An alternative albeit somewhat weaker extension of [4] to treat the case $-1 \in \operatorname{Spec}(\mathbf{S})$ was given in [10].

2 Two algebraic lemmas

Let us define for arbitrary square matrices $\mathbf{M}_1, \dots, \mathbf{M}_k$, the diagonal block matrix by:

$$\mathbf{D}(\mathbf{M}_1,\dots,\mathbf{M}_k) \;:=\; \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_k \end{bmatrix}. \tag{2.1}$$

General algebraic results for non-negative matrices yield the following decomposition when applied to $\mathbf{S}$ satisfying the assumptions of Theorem 1.1.

Lemma 2.1. *Suppose that* $\mathbf{S}$ *is a symmetric (double) stochastic matrix that satisfies conditions (A1)–(A3). Then, after an appropriate permutation* $\mathbf{P}$ *of the indices,* $\mathbf{S}$ *has a block structure*

$$\mathbf{S} \;=\; \mathbf{P}\,\mathbf{D}\big(\mathbf{S}_1,\dots,\mathbf{S}_p,\widetilde{\mathbf{S}}_1,\dots,\widetilde{\mathbf{S}}_q\big)\,\mathbf{P}^{-1}\,, \tag{2.2}$$

where $\mathbf{S}_\alpha$*,* $1 \le \alpha \le p$*, and* $\widetilde{\mathbf{S}}_\beta$*,* $1 \le \beta \le q$*, are irreducible doubly stochastic matrices with some* p, q*. The spectrums of the blocks*

$$\widetilde{\mathbf{S}}_\beta = (\tilde{s}_{\beta;ij})_{i,j=1}^{\tilde{d}_\beta}\,, \quad 1 \le \beta \le q\,,$$

satisfy $\mathrm{Spec}(\widetilde{\mathbf{S}}_\beta) \subset [-\rho, \rho] \cup \{+1\}$*. The blocks* $\mathbf{S}_\alpha$ *have both* $+1$ *and* -1 *as simple eigenvalues, and they have the structure*

$$\mathbf{S}_\alpha = \begin{bmatrix} \mathbf{0} & \mathbf{A}_\alpha^{\mathrm{T}} \\ \mathbf{A}_\alpha & \mathbf{0} \end{bmatrix}\,, \qquad \mathbf{A}_\alpha = (a_{\alpha;ij})_{i,j=1}^{d_\alpha}\,, \quad 1 \le \alpha \le p\,, \tag{2.3}$$

where both the rows and the columns of the (generally) non-symmetric matrices $\mathbf{A}_\beta$ *sum to one, i.e.,* $\sum_i a_{\beta;ij} = \sum_j a_{\beta;ij} = 1$*. The matrix elements are bounded,*

$$\tilde{s}_{\beta;ij}\,,\; a_{\beta;ij} \;\le\; \frac{1}{M}\,,$$

while the dimensions satisfy $d_\alpha,\, \tilde{d}_\beta \ge M$*, and*

$$2d_1 \,+\, \ldots \,+\, 2d_p \,+\, \tilde{d}_1 \,+\, \ldots \,+\, \tilde{d}_q \;=\; N\,. \tag{2.4}$$

Proof. Irreducible components of $\mathbf{S}$ can be permuted into diagonal square blocks (2.2) and the properties (A1)–(A3) are preserved under relabelling. In particular, the constant row and column sums for $\mathbf{S}$, as well as the bound $0 \le s_{ij} \le M^{-1}$, translate directly to analogous bounds for $\widetilde{S}_\beta$'s and $\mathbf{A}_\alpha$'s, and this in turn implies $d_\alpha, \tilde{d}_\beta \ge M$.

The structure of the block decomposition of $\mathbf{S}_\alpha$ (2.3) follows from the general theory of non-negative irreducible matrices $\mathbf{M} = (m_{ij})_{i,j=1}^{d}$, $m_{ij} \ge 0$, e.g. Theorem 2.20 of [1]: If $\mathbf{M}$ has k-eigenvalues on its spectral circle, $\{z \in \mathbb{C} : |z| = r\}$, then those eigenvalues are precisely the k complex roots of r^2, i.e., they equal $r\mathrm{e}^{\mathrm{i}2\pi j/k}$, $1 \le j \le k$. Moreover, the matrix $\mathbf{M}$ has the block representation:

$$\mathbf{M} \;=\; \begin{bmatrix} \mathbf{0} & \mathbf{D}\big(\mathbf{M}_1, \ldots, \mathbf{M}_{k-1}\big) \\ \mathbf{M}_k & \mathbf{0} \end{bmatrix} \tag{2.5}$$

for some matrices $\mathbf{M}_1, \ldots, \mathbf{M}_k$. The dimensions of $\mathbf{M}_j$'s are such that the zero blocks along the diagonal (not visible in (2.5)) are square, so that the rows of $\mathbf{M}_j$ and the columns of $\mathbf{M}_{j+1}$ have the same dimensions for each $j = 1, 2, \ldots, k$ if one identifies $\mathbf{M}_{k+1} := \mathbf{M}_1$.

Applying the decomposition (2.5) to $\mathbf{M} := \mathbf{S}_\alpha$ yields the representation (2.3) since the symmetry $\mathbf{S}_\alpha^{\mathrm{T}} = \mathbf{S}_\alpha$ implies $1 \le k \le 2$, while $-1 \in \mathrm{Spec}(\mathbf{S}_\alpha)$ excludes the case $k \neq 1$. □

The random matrix $\mathbf{H}$ inherits the structure (2.2) of $\mathbf{S}$ through (1.2):

$$\mathbf{H} \,=\, \mathbf{P}\,\mathbf{D}\big(\mathbf{H}_1, \ldots, \mathbf{H}_p, \widetilde{\mathbf{H}}_1, \ldots, \widetilde{\mathbf{H}}_q\big)\,\mathbf{P}^{-1}\,.$$

Here $\mathbf{H}_\alpha$ and $\widetilde{\mathbf{H}}_\beta$ are independent generalised Wigner matrices satisfying $\mathbb{E}\,|h_{\alpha;ij}|^2 = s_{\alpha;ij}$ and $\mathbb{E}\,|\tilde{h}_{\beta;ij}|^2 = \tilde{s}_{\beta;ij}$, respectively. This decomposition means that it suffices to prove Theorem 1.1 for the irreducible components separately. The components $\widetilde{\mathbf{H}}_\beta$ are already covered by Theorem 1.2. Hence, dropping the indices $\alpha \ge 1$, we are left to prove Theorem 1.1 in the case

$$\mathbf{H} \,=\, \begin{bmatrix} \mathbf{0} & \mathbf{X}^* \\ \mathbf{X} & \mathbf{0} \end{bmatrix} \qquad \text{and} \qquad \mathbf{S} \,=\, \begin{bmatrix} \mathbf{0} & \mathbf{A}^{\mathrm{T}} \\ \mathbf{A} & \mathbf{0} \end{bmatrix}\,, \tag{2.6}$$

where $\mathbf{S}$ is irreducible, and the entries x_{ij} of the square matrices $\mathbf{X}$ are independent, and satisfy $\mathbb{E}\,|x_{ij}|^2 = a_{ij}$. For the sake of convenience, we also redefine N to be equal to the dimension of $\mathbf{A}$, so that $\mathbf{H}$ and $\mathbf{S}$ are $2N \times 2N$ matrices.

Using the special structure (2.6) it follows that if $\lambda \in \mathrm{Spec}(\mathbf{H})$ then also $-\lambda \in \mathrm{Spec}(\mathbf{H})$, and the corresponding eigenvectors are related in the following simple way:

$$\mathbf{H}\begin{bmatrix}\mathbf{u}\\ \pm\mathbf{w}\end{bmatrix} = \pm\,\lambda\begin{bmatrix}\mathbf{u}\\ \pm\mathbf{w}\end{bmatrix}. \tag{2.7}$$

In particular, the same reasoning can be applied to the ± 1 eigenvalues of $\mathbf{S}$. Moreover, since $\mathbf{S}$ is double stochastic and irreducible, the eigenvectors of $\mathbf{S}$ belonging to the non-degenerate eigenvalues ± 1, equal

$$\begin{aligned}\mathbf{e} &:= \tfrac{1}{\sqrt{2N}}(1,\dots,1,1,\dots,1)^{\mathrm{T}}\\ \mathbf{f} &:= \tfrac{1}{\sqrt{2N}}(1,\dots,1,-1,-1,\dots,-1)^{\mathrm{T}},\end{aligned} \tag{2.8}$$

so that $\mathbf{S}\mathbf{e} = \mathbf{e}$ and $\mathbf{S}\mathbf{f} = -\mathbf{f}$.

Let us denote the complex inner product between vectors $\mathbf{a},\mathbf{b} \in \mathbb{C}^{2N}$ by $(\mathbf{a},\mathbf{b}) = \sum_i \overline{a_i}\,b_i$. Combining the symmetries (2.7) and (2.8) of $\mathbf{H}$ and $\mathbf{S}$ yields the following algebraic identity.

Lemma 2.2. *Let $\mathbf{H}$ be a $2N$-dimensional self-adjoint matrix that has the block structure* (2.6) *but is otherwise arbitrary, e.g.,* (1.2) *is not assumed. Then the diagonal elements of the Green function $\mathbf{G} := (\mathbf{H}-z)^{-1}$, $\mathrm{Im}\,z \geq 0$, constitute a vector orthogonal to the -1 eigenspace of $\mathbf{S}$,*

$$\big(\mathbf{f}, \mathrm{diag}(\mathbf{G})\big) \,=\, 0\,, \tag{2.9}$$

where $\mathrm{diag}(\mathbf{G}) = (G_{11},\dots,G_{2N,2N})$ and $\mathbf{f}$ is defined in (2.8).

Proof. Let us additionally assume that $0 \notin \mathrm{Spec}(\mathbf{H})$, and, that besides the pairs (2.7), there are no further degeneracies. Suppose

$$\mathbf{v} := \begin{bmatrix}\mathbf{u}\\ \mathbf{w}\end{bmatrix} \qquad\text{and}\qquad \widetilde{\mathbf{v}} := \begin{bmatrix}\mathbf{u}\\ -\mathbf{w}\end{bmatrix}, \qquad\text{with}\qquad \mathbf{u},\mathbf{w} \in \mathbb{C}^N,$$

are the eigenvectors corresponding to the eigenvalues λ and $-\lambda$ in (2.7), respectively. Since $\lambda \neq 0$ (so that $-\lambda \neq \lambda$) these eigenvectors are orthogonal

$$0 \,=\, (\mathbf{v},\widetilde{\mathbf{v}}) \,=\, \sum_{i=1}^N |u_i|^2 - \sum_{k=1}^N |w_k|^2\,,$$

i.e., the first and the second blocks are balanced: $\|\mathbf{u}\|_2 = \|\mathbf{w}\|_2$.

Now, let $\mathbf{v}^{(\alpha)}$, $\alpha = 1,\dots,2N$, be the $2N$ eigenvectors of H, and let $\mathbf{u}^{(\alpha)}$ and $\mathbf{w}^{(\alpha)}$ contain the first and the last N-components of $\mathbf{v}^{(\alpha)}$. Then combining the spectral theorem,

$$G_{kk} \,=\, \sum_{\alpha=1}^{2N} \frac{|v_k^{(\alpha)}|^2}{\lambda_\alpha - z}\,,$$

with the balancing conditions, $\|\mathbf{u}^{(\alpha)}\|_2 = \|\mathbf{w}^{(\alpha)}\|_2$, yields

$$\sum_{k=1}^N G_{kk} \,=\, \sum_{\alpha=1}^{2N} \frac{\|\mathbf{u}^{(\alpha)}\|_2^2}{\lambda_\alpha - z} \,=\, \sum_{\alpha=1}^{2N} \frac{\|\mathbf{w}^{(\alpha)}\|_2^2}{\lambda_\alpha - z} \,=\, \sum_{k=N+1}^{2N} G_{kk}\,, \tag{2.10}$$

which is equivalent to (2.9). Finally, by using a basic continuity argument, one sees that (2.10) must apply also without the extra assumptions concerning the degeneracies and the exclusion of 0 from the spectrum of $\mathbf{H}$. □

3 Translating proof of Theorem 1.2 to cover case (2.6)

With the identity (2.9) at hand we may translate the proof of Theorem 1.2 from [4] to the setting (2.6) without significant changes. In order to see this, we recall that the -1 eigenvalue of $\mathbf{S}$ enters the proofs in [4] only when one needs to bound the inverse of the operator $1 - m(z)^2\mathbf{S}$. However, using the identity (2.9) one can show that in all such cases it suffices to restrict the analysis to the orthogonal complement of the eigendirection $\mathbf{f}$ corresponding to the eigenvalue -1. The lower spectral gap assumption (1.6) above -1 then guarantees that $(1-m(z)^2\mathbf{S})^{-1}$ stays uniformly bounded even when $m(z)^2$ becomes close to -1 (equivalently, $z \approx 0$), i.e.,

$$(1 - m(z)^2\mathbf{S})^{-1} \;\approx\; (1 + \mathbf{S})^{-1}\,, \qquad \text{for} \quad z \approx 0\,.$$

Since the 'bad direction' $\mathbf{f}$ will not play any role in the analysis, the only necessary modification of [4] in the end, is to replace the operator norm $\widetilde{\Gamma}(z)$ (cf. equation (2.11) in [4]) of $(1 - m^2\mathbf{S})^{-1}$ in $\mathbf{e}^\perp$, by the analogous norm in the orthogonal complement of both $\mathbf{e}$ and $\mathbf{f}$:

$$\widehat{\Gamma}(z) \;:=\; \big\|(1 - m(z)^2\mathbf{S})^{-1}|_{\mathrm{span}\{\mathbf{e},\mathbf{f}\}^\perp}\big\|_{\ell^\infty\to\ell^\infty}\,. \tag{3.1}$$

The estimate (A.3) of [4],

$$\widehat{\Gamma}(z) \;\le\; C(\rho)\log N$$

remains valid since the operator norm of $(1 - m(z)^2\mathbf{S})^{-1}$ from ℓ^2 to itself is bounded by $1/(1-\rho)$ in the complement of $\mathrm{span}\{\mathbf{e},\mathbf{f}\}$. Here the logarithm comes from the fact that the ℓ^∞-norm is bigger by this factor over the ℓ^2-norm (cf. p. 46 of [4]).

In order to simplify the notations let us drop out the explicit z-dependence and write $m = m(z)$, $\mathbf{G} = \mathbf{G}(z)$, etc., as in [4]. It remains to demonstrate why the inversion of $1 - m^2\mathbf{S}$ can be always be restricted to the orthogonal complement of $\mathbf{f}$. This inversion was used to bound the random fluctuations of the diagonal resolvent elements,

$$v_i \;:= G_{ii} - m \tag{3.2}$$

in terms of the small random error terms $\Upsilon_i = \mathcal{O}_\prec(N^{-c})$ appearing the self-consistent vector equation (cf. (5.9) in [4]):

$$-\sum_k s_{ik}v_k \;+\; \Upsilon_i \;=\; \frac{1}{m+v_i} \;-\; \frac{1}{m}\,. \tag{3.3}$$

Under the assumption, $|v_i| \prec \Lambda \prec N^{-c}$, with some control parameter Λ, and using $|m| \sim 1$, (3.3), takes the form

$$(1 - m^2\mathbf{S})\mathbf{v} \;=\; \mathcal{O}_\prec(\|\boldsymbol{\Upsilon}\|_\infty + \Lambda^2)\,. \tag{3.4}$$

Writing (3.2) as

$$\mathbf{v} = \mathrm{diag}(\mathbf{G}) - \sqrt{2N}m\,\mathbf{e}\,,$$

recalling $(\mathbf{f},\mathbf{e}) = 0$, and then applying Lemma 2.2 yields:

$$(\mathbf{f},\mathbf{v}) \;=\; (\mathbf{f},\mathrm{diag}(\mathbf{G})) \;=\; 0\,. \tag{3.5}$$

The identity (3.5) shows that inversion of $1 - m^2\mathbf{S}$ can be indeed restricted to the complement of $\mathbf{f}$ in the case of (3.4).

The inverse of $1 - m^2\mathbf{S}$ becomes unbounded also in the direction $\mathbf{e}$ when $m^2 \approx 1$. However, unlike with direction $\mathbf{f}$, the inversions of $1-m^2\mathbf{S}$ can not be straightforwardly restricted to the complement of $\mathbf{e}$, since the average of $\mathbf{v}$,

$$[\mathbf{v}] \,:=\, \frac{1}{2N}\sum_{i=1}^{2N} v_i \,=\, \frac{(\mathbf{e},\mathbf{v})}{\sqrt{2N}}\,, \tag{3.6}$$

is not small. For this reason the critical part $[\mathbf{v}]$ was treated separately from the remainder, $\mathbf{v} - (\mathbf{e},\mathbf{v})\,\mathbf{e} \in \mathrm{span}\{\mathbf{e},\mathbf{f}\}^\perp$ in a more precise second order scalar equation in [4]. The remainder part satisfies a linearised vector equation for which one needs to again invert $1 - m^2\mathbf{S}$. We will now demonstrate that also in this case the component $\mathbf{f}$ is not present due to Lemma 2.2. Indeed, in order to get from (6.19) to (6.20) in [4] one applies the fluctuation averaging estimate (4.14) (with the choice $t_{ij} = s_{ij}$) to bound the remainder $\mathbf{v} - (\mathbf{e},\mathbf{v})\,\mathbf{e}$. The crucial steps appear in the proof of (4.14) located at the end of the proof of Theorem 4.7 on p. 54 of [4], where a bound for

$$w_a := \sum_i t_{ai}(v_i - [v])\,, \tag{3.7}$$

is derived from a linearised self-consistent equation

$$\sum_i t_{ai}(v_i - [\mathbf{v}]) \,=\, m^2 \sum_{b,j} s_{ab} t_{bj}(v_j - [\mathbf{v}]) \,+\, \mathcal{O}_\prec(\Psi^2)\,,\,, \tag{3.8}$$

in terms of the small control parameter $\Psi \le N^{-c}$. Writing $\mathbf{w} = \mathbf{T}(\mathbf{v} - (\mathbf{e},\mathbf{v})\,\mathbf{e})$, and recalling $[\mathbf{T},\mathbf{S}] = \mathbf{0}$ (actually we need only the case $\mathbf{T} = \mathbf{S}$ here) and $(\mathbf{f},\mathbf{v}) = 0$ by (3.5), we obtain:

$$(\mathbf{f},\mathbf{w}) \,=\, 0\,. \tag{3.9}$$

Thus by expressing (3.8) in the vector form,

$$(1 - m^2\mathbf{S})\mathbf{w} \,=\, \mathcal{O}_\prec(\Psi^2)\,, \tag{3.10}$$

we see that $1 - m^2\mathbf{S}$ can be also inverted in the subspace orthogonal to both the $+1$ and -1 eigendirections. Hence (3.10) yields

$$\mathbf{w} \,=\, \mathcal{O}_\prec\big(\widehat{\Gamma}\,\Psi^2\big)\,,$$

which is exactly the fluctuation averaging bound (4.14) of [4] with $\widetilde{\Gamma}$ updated to $\widehat{\Gamma}$.

Besides these observations and the replacement of $\widetilde{\Gamma}$ by $\widehat{\Gamma}$ the proof from [4] can be carried out without further modifications.

References

[1] Abraham Berman and Robert J Plemmons, *Nonnegative matrices in the mathematical sciences*, Classics in Applied Mathematics, vol. 9, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. MR-1298430

[2] Paul Bourgade, Horng-Tzer Yau, and Jun Yin, *Local circular law for random matrices*, Probability Theory and Related Fields (2013), 1–51.

[3] Claudio Cacciapuoti, Anna Maltsev, and Benjamin Schlein, *Local marchenko-pastur law at the hard edge of sample covariance matrices*, J. Math. Phys **54** (2013), 043302. MR-3088802

[4] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin, *The local semicircle law for a general class of random matrices*, Electronic Journal of Probability **18** (2013), no. 0, 1–58. MR-3068390

[5] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin, *The local relaxation flow approach to universality of the local statistics of random matrices*, Ann. Inst. Henri Poincaré (B) **48** (2012), 1–46. MR-2919197

[6] László Erdős and H.-T. Yau, *Universality of local spectral statistics of random matrices*, Bull. Amer. Math. Soc **49** (2012), 377–414. MR-2917064

[7] László Erdős, Horng-Tzer Yau, and Jun Yin, *Bulk universality for generalized Wigner matrices*, Probability Theory and Related Fields **154** (2011), no. 1-2, 341–407. MR-2981427

[8] ______, *Universality for generalized Wigner matrices with Bernoulli distribution*, Journal of Combinatorics **2** (2011), no. 1, 15–82. MR-2847916

[9] ______, *Rigidity of eigenvalues of generalized Wigner matrices*, Advances in Mathematics **229** (2012), no. 3, 1435–1515. MR-2871147

[10] Mark Rudelson and Ofer Zeitouni, *In preparation*, (private communication) (2013).