Electron. Commun. Probab. **22** (2017), no. 63, 1–13.
DOI: 10.1214/17-ECP97
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Singularities of the density of states of random Gram matrices

Johannes Alt*

Abstract

For large random matrices X with independent, centered entries but not necessarily identical variances, the eigenvalue density of XX^* is well-approximated by a deterministic measure on $\mathbb{R}$. We show that the density of this measure has only square and cubic-root singularities away from zero. We also extend the bulk local law in [5] to the vicinity of these singularities.

Keywords: local law; Dyson equation; square-root edge; cubic cusp; general variance profile.
AMS MSC 2010: Primary 60B20, Secondary 15B52.
Submitted to ECP on August 28, 2017, final version accepted on November 2, 2017.
Supersedes arXiv:1708.08442.

1 Introduction

The empirical eigenvalue density or *density of states* of many large random matrices is well-approximated by a deterministic probability measure, the *self-consistent density of states*. If X is a $p \times n$ random matrix with independent, centered entries of identical variances then the limit of the eigenvalue density of the *sample covariance matrix* XX^* for large p and n with p/n converging to a constant has been identified by Marchenko and Pastur in [9]. However, some applications in wireless communication require understanding the spectrum of XX^* without the assumption of identical variances of the entries of $X = (x_{kq})_{k,q}$ [6, 8, 10]. In this case, the matrix XX^* is a *random Gram matrix*.

For constant variances, the self-consistent density of states is obtained by solving a scalar equation for its Stieltjes transform, the *scalar Dyson equation*. In case the variances $s_{kq} := \mathbb{E}|x_{kq}|^2$ depend nontrivially on k and q, the self-consistent density of states is obtained from the solution $m(\zeta) = (m_1(\zeta), \ldots, m_p(\zeta)) \in \mathbb{H}^p$ of the *vector Dyson equation* [7]

$$-\frac{1}{m_k(\zeta)} = \zeta - \sum_{q=1}^{n} s_{kq}\Big(1 + \sum_{l=1}^{p} s_{lq} m_l(\zeta)\Big)^{-1} \qquad \text{for all } k \in [p], \tag{1.1}$$

for all $\zeta \in \mathbb{H}$. Here, we introduced $\mathbb{H} := \{\zeta \in \mathbb{C} \colon \operatorname{Im} \zeta > 0\}$ and $[p] := \{1, \ldots, p\}$. Indeed, the average $\langle m(\zeta)\rangle_1 := p^{-1} \sum_{k=1}^{p} m_k(\zeta)$ is the Stieltjes transform of the self-consistent

*Institute of Science and Technology Austria. Partially funded by ERC Advanced Grant RANMAT No. 338804.
E-mail: johannes.alt@ist.ac.at

density of states denoted by $\langle \nu \rangle_1$. If the limit of $\langle \nu \rangle_1$ as $p, n \to \infty$ exists then it can be studied via an infinite-dimensional version of (1.1) (see (2.3) below).

For Wigner-type matrices, i.e., Hermitian random matrices with independent (up to the Hermiticity constraint), centered entries, the analogue of (1.1) is a quadratic vector equation (QVE) in the language of [1, 3]. In these papers, finite and infinite-dimensional versions of the QVE have been extensively studied to analyze the self-consistent density of states whose Stieltjes transform is the average of the solution to the QVE. The authors show that the self-consistent density of states has a $1/3$-Hölder continuous density. Except for finitely many square-root and cubic-root singularities this density is real-analytic. The square-root behaviour emerges solely at the edges of the connected components of the support of the self-consistent density of states, whereas the cubic-root singularities lie inside these components. The detailed stability analyis in [1] is then used in [2] to obtain the local law for Wigner-type matrices. A *local law* typically refers to a statement about the convergence of the eigenvalue density to a deterministic measure on a scale slightly above the typical local eigenvalue spacing.

For the Dyson equation for random Gram matrices, we obtain away from $\zeta = 0$ the same results as mentioned above in the QVE setup. Furthermore, we extend our local law for random Gram matrices in [5] to the vicinity of the singularities of the self-consistent density of states. This can be seen as another instance of the universality phenomenon in random matrix theory. Despite the different structure of Gram and Wigner-type matrices, the densities of states of these Hermitian random matrices have the same types of singularities. We refer to [5] and the references therein for related results about random Gram matrices.

There is a close connection between Gram and Wigner-type matrices. The Dyson equation, (1.1), can be transformed into a QVE in the sense of [1] and the spectrum of XX^* is closely related to that of a Wigner-type matrix in the sense of [2]. This is easiest explained on the random matrix level through a special case of the linearization tricks: If X has independent and centered entries then the random matrix

$$\boldsymbol{H} = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix} \tag{1.2}$$

is a Wigner-type matrix and the spectra of $\boldsymbol{H}^2$ and XX^* agree away from zero. Therefore, instead of trying to analyze (1.1) and XX^* directly, it is more efficient to study the corresponding QVE and Wigner-type matrix as in [5]. However, owing to the large zero blocks in $\boldsymbol{H}$, its variance matrix is not uniformly primitive (see **A3** in [1]), a key assumption for the analysis in [1]. Indeed, the stability operator of the QVE possesses an additional unstable direction $\boldsymbol{f}_-$, which has to be treated separately. In [5], this study has been conducted in the bulk spectrum and away from the support of $\langle \nu \rangle_1$, where $\boldsymbol{f}_-$ did not play an important role at least away from zero.

In this note, we present a new argument needed in the analysis of the cubic equation (see (3.19) below) describing the stability of the QVE close to its singularities in order to incorporate the additional unstable direction. In fact, the analysis of the cubic equation in [1] heavily relies on the uniform primitivity of the variance matrix. Adapting this argument to the current setup cannot exclude that the coefficients of the cubic and the quadratic term in the cubic equation vanish at the same time due to the presence of $\boldsymbol{f}_-$. A nonvanishing cubic or quadratic coefficient is however absolutely crucial for the cubic stability analysis in [1]. Otherwise not only square-root or cubic-root but also higher order singularities would emerge. Our main novel ingredient, a very detailed analysis of these coefficients, actually excludes this scenario. With this essential new input, the regularity and the singularity structure of (1.1) as well as the local law for XX^* follow by correctly combining the arguments in [1, 2, 5].

Acknowledgments. The author is very grateful to László Erdős for many fruitful discussions and many valuable suggestions. The author would also like to thank Torben Krüger for several helpful conversations.

2 Main results

2.1 Structure of the solution to the Dyson equation

Let $(\mathfrak{X}_1, \mathcal{S}_1, \pi_1)$ and $(\mathfrak{X}_2, \mathcal{S}_2, \pi_2)$ be two finite measure spaces such that $\pi_1(\mathfrak{X}_1)$ and $\pi_2(\mathfrak{X}_2)$ are strictly positive. Moreover, we denote the spaces of bounded and measurable functions on $\mathfrak{X}_1$ and $\mathfrak{X}_2$ by

$$\mathscr{B}_i := \left\{ u\colon \mathfrak{X}_i \to \mathbb{C} : \|u\|_\infty := \sup_{x\in\mathfrak{X}_i} |u(x)| < \infty \right\}$$

for $i = 1, 2$. We consider $\mathscr{B}_1$ and $\mathscr{B}_2$ equipped with the supremum norm $\|\cdot\|_\infty$. We denote the induced operator norms by $\|\cdot\|_{\mathscr{B}_1\to\mathscr{B}_2}$ and $\|\cdot\|_{\mathscr{B}_2\to\mathscr{B}_1}$. For $u \in \mathscr{B}_1$, we write $u_k = u(k)$ for $k \in \mathfrak{X}_1$. We use the same notation for $v \in \mathscr{B}_2$.

Let $s\colon \mathfrak{X}_1 \times \mathfrak{X}_2 \to \mathbb{R}_0^+, s(k, q) = s_{kq}$ be a measurable nonnegative function such that

$$\sup_{k\in\mathfrak{X}_1} \int_{\mathfrak{X}_2} s_{kq}\pi_2(\mathrm{d}q) < \infty, \qquad \sup_{q\in\mathfrak{X}_2} \int_{\mathfrak{X}_1} s_{kq}\pi_1(\mathrm{d}k) < \infty. \tag{2.1}$$

We define the bounded linear operators $S\colon \mathscr{B}_2 \to \mathscr{B}_1$ and $S^t\colon \mathscr{B}_1 \to \mathscr{B}_2$ through

$$(Sv)_k = \int_{\mathfrak{X}_2} s_{kr}v_r\pi_2(\mathrm{d}r), \;\; k \in \mathfrak{X}_1,\; v \in \mathscr{B}_2, \quad (S^tu)_q = \int_{\mathfrak{X}_1} s_{lq}u_l\pi_1(\mathrm{d}l), \;\; q \in \mathfrak{X}_2,\; u \in \mathscr{B}_1. \tag{2.2}$$

We are interested in the solution $m\colon \mathbb{H} \to \mathscr{B}_1$ of the *Dyson equation*

$$-\frac{1}{m(\zeta)} = \zeta - S\frac{1}{1 + S^tm(\zeta)}, \tag{2.3}$$

for $\zeta \in \mathbb{H}$, which satisfies $\operatorname{Im} m(\zeta) > 0$ for all $\zeta \in \mathbb{H}$.

Proposition 2.1 (Existence and Uniqueness)**.** *If* (2.1) *holds true then there is a unique function $m\colon \mathbb{H} \to \mathscr{B}_1$ satisfying* (2.3) *and* $\operatorname{Im} m(\zeta) > 0$ *for all* $\zeta \in \mathbb{H}$*. Moreover, $m\colon \mathbb{H} \to \mathscr{B}_1$ is analytic. For each $k \in \mathfrak{X}_1$, there is a unique probability measure ν_k on $\mathbb{R}$ such that m_k is the Stieltjes transform of ν_k, i.e.,*

$$m_k(\zeta) = \int_0^\infty \frac{1}{E - \zeta}\nu_k(\mathrm{d}E) \tag{2.4}$$

for all $\zeta \in \mathbb{H}$. The support of ν_k is contained in $[0, \Sigma]$ for each $k \in \mathfrak{X}_1$, where

$$\Sigma := 4 \max\left\{\|S\|_{\mathscr{B}_2\to\mathscr{B}_1}, \|S^t\|_{\mathscr{B}_1\to\mathscr{B}_2}\right\}. \tag{2.5}$$

Further assumptions on π_1, π_2 and S will yield a more detailed understanding of the measures ν_k. To formulate these assumptions, we introduce the averages of $u \in \mathscr{B}_1$ and $v \in \mathscr{B}_2$ through

$$\langle u\rangle_1 = \frac{1}{\pi_1(\mathfrak{X}_1)} \int_{\mathfrak{X}_1} u_k\pi_1(\mathrm{d}k), \qquad \langle v\rangle_2 = \frac{1}{\pi_2(\mathfrak{X}_2)} \int_{\mathfrak{X}_2} v_q\pi_2(\mathrm{d}q).$$

Additionally, we set $\|u\|_t := \langle|u|^t\rangle_1^{1/t}$ and $\|v\|_t := \langle|v|^t\rangle_2^{1/t}$ for $u \in \mathscr{B}_1$, $v \in \mathscr{B}_2$ and $t \geq 1$. Moreover, for $k \in \mathfrak{X}_1$ and $q \in \mathfrak{X}_2$, we define the functions $S_k\colon \mathfrak{X}_2 \to \mathbb{R}_0^+,\ S_k(r) = s_{kr}$

and $(S^t)_q\colon \mathfrak{X}_1 \to \mathbb{R}_0^+,\ (S^t)_q(l) = s_{lq}$. We call S_k and $(S^t)_q$ the *rows* and *columns* of S, respectively.

Assumptions 2.2. **(A1)** The total measures $\pi_1(\mathfrak{X}_1)$ and $\pi_2(\mathfrak{X}_2)$ are comparable, i.e., there are constants $0 < \pi_* < \pi^*$ such that

$$\pi_* \leq \frac{\pi_1(\mathfrak{X}_1)}{\pi_2(\mathfrak{X}_2)} \leq \pi^*.$$

(A2) The operators S and S^t are irreducible in the sense that there are $L_1, L_2 \in \mathbb{N}$ and $\kappa_1, \kappa_2 > 0$ such that

$$\big((SS^t)^{L_1} u\big)_k \geq \kappa_1 \langle u \rangle_1, \quad \big((S^tS)^{L_2} v\big)_q \geq \kappa_2 \langle v \rangle_2,$$

for all $u \in \mathscr{B}_1$, $v \in \mathscr{B}_2$ satisfying $u \geq 0$ and $v \geq 0$ and for all $k \in \mathfrak{X}_1$, $q \in \mathfrak{X}_2$.

(A3) The rows and columns of S are sufficiently close to each other in the sense that there is a continuous strictly monotonically decreasing function $\gamma\colon (0,1] \to \mathbb{R}_0^+$ such that $\lim_{\varepsilon \downarrow 0} \gamma(\varepsilon) = \infty$ and for all $\varepsilon \in (0,1]$, we have

$$\gamma(\varepsilon) \leq \min\left\{ \inf_{k \in \mathfrak{X}_1} \frac{1}{\pi_1(\mathfrak{X}_1)} \int_{\mathfrak{X}_1} \frac{\pi_1(\mathrm{d}l)}{\varepsilon + \|S_k - S_l\|_2^2}, \inf_{q \in \mathfrak{X}_2} \frac{1}{\pi_2(\mathfrak{X}_2)} \int_{\mathfrak{X}_2} \frac{\pi_2(\mathrm{d}r)}{\varepsilon + \|(S^t)_q - (S^t)_r\|_2^2} \right\}.$$

(A4) The operators S and S^t map square-integrable functions continuously to bounded functions, i.e., there are constants $\Psi_1, \Psi_2 > 0$ such that

$$\|S\|_{L^2(\pi_2/\pi_2(\mathfrak{X}_2)) \to \mathscr{B}_1} \leq \Psi_1, \qquad \|S^t\|_{L^2(\pi_1/\pi_1(\mathfrak{X}_1)) \to \mathscr{B}_2} \leq \Psi_2.$$

Our estimates will be uniform in all models that satisfy Assumptions 2.2 with the same constants. Therefore, the constants π_*, π^* from **(A1)**, L_1, L_2, κ_1, κ_2 from **(A2)**, the function γ from **(A3)** and Ψ_1, Ψ_2 from **(A4)** are called *model parameters*. We refer to Remark 2.4 below for an easily checkable sufficient condition for **(A3)**. We now state our main result about the regularity and the possible singularities of ν_k defined in (2.4).

Theorem 2.3. *If we assume **(A1)** – **(A4)** then the following statements hold true:*

(i) *(Regularity of ν) There are $\nu^0 \in \mathscr{B}_1$ and $\nu^d\colon \mathfrak{X}_1 \times (0,\infty) \to [0,\infty),\ (k,E) \mapsto \nu_k^d(E)$ such that*

$$\nu_k(\mathrm{d}E) = \nu_k^0 \delta_0(\mathrm{d}E) + \nu_k^d(E)\mathrm{d}E \tag{2.6}$$

for all $k \in \mathfrak{X}_1$. For all $\delta > 0$, the function ν^d is uniformly $1/3$-Hölder continuous on $[\delta,\infty)$, i.e.,

$$\sup_{k \in \mathfrak{X}_1} \sup_{E_1 \neq E_2,\, E_1,E_2 \geq \delta} \frac{|\nu_k^d(E_1) - \nu_k^d(E_2)|}{|E_1 - E_2|^{1/3}} < \infty.$$

For all $k \in \mathfrak{X}_1$, we have

$$\{E \in (0,\infty)\colon \langle \nu^d(E) \rangle > 0\} = \{E \in (0,\infty)\colon \nu_k^d(E) > 0\}.$$

We set $\mathfrak{P} := \{E \in (0,\infty)\colon \langle \nu^d(E) \rangle > 0\}$. For each $\delta > 0$, the set $\mathfrak{P} \cap (\delta,\infty)$ is a finite union of open intervals. The map $\nu^d\colon (0,\infty) \setminus \partial\mathfrak{P} \to \mathscr{B}_1$ is real-analytic. There is $\rho_* > 0$ depending only on the model parameters and δ such that the Lebesgue measure of each connected component of $\mathfrak{P} \cap (\delta,\infty)$ is at least $2\rho_*$.

(ii) *(Singularities of ν^d) Fix $\delta > 0$. For any $E_0 \in (\partial\mathfrak{P}) \cap (\delta,\infty)$, there are two cases*

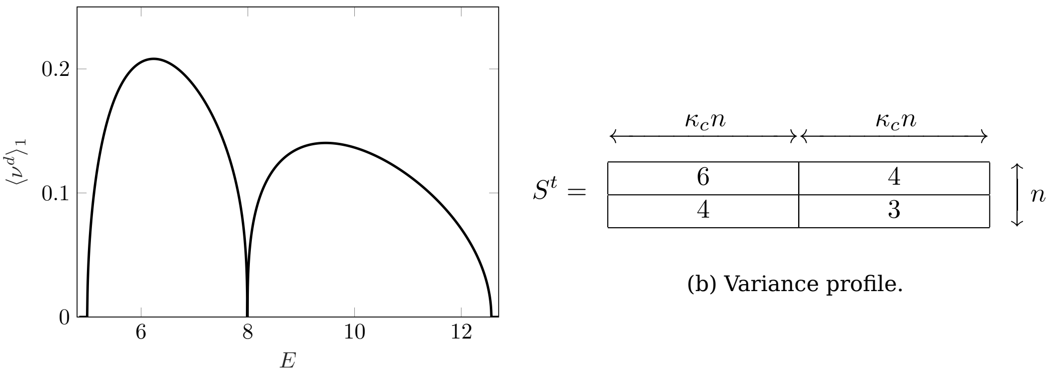
CUSP: *The point* E_0 *is the intersection of the closures of two connected components of* $\mathfrak{P} \cap (\delta, \infty)$ *and* ν^d *has a cubic root singularity at* E_0*, i.e., there is* $c \in \mathscr{B}_1$ *satisfying* $\inf_{k \in \mathfrak{X}_1} c_k > 0$ *such that*

$$\nu_k^d(E_0 + \lambda) = c_k |\lambda|^{1/3} + \mathcal{O}(|\lambda|^{2/3}), \qquad \lambda \to 0.$$

EDGE: *The point* E_0 *is the left or right endpoint of a connected component of* $\overline{\mathfrak{P}} \cap (\delta, \infty)$ *and* ν^d *has a square root singularity at* E_0*, i.e., there is* $c \in \mathscr{B}_1$ *satisfying* $\inf_{k \in \mathfrak{X}_1} c_k > 0$ *such that*

$$\nu_k^d(E_0 + \theta\lambda) = c_k \lambda^{1/2} + \mathcal{O}(\lambda), \qquad \lambda \downarrow 0,$$

where $\theta = +1$ *if* E_0 *is a left endpoint of* $\mathfrak{P}$ *and* $\theta = -1$ *if* E_0 *is a right endpoint.*



(a) Self-consistent density of states $\langle \nu^d \rangle_1$.

(b) Variance profile.

Figure 1: Example of a self-consistent density of states with variance profile S. It has square-root edges at the left and right endpoint of its support and a cubic cusp at $E \approx 8$.

Remark 2.4 (Piecewise Hölder-continuous rows and columns of S imply **(A3)**)**.** Let $\mathfrak{X}_1$ and $\mathfrak{X}_2$ be two nontrivial compact intervals in $\mathbb{R}$ and π_1 and π_2 the Lebesgue measures. In this case, **(A3)** holds true if the maps $k \mapsto S_k$ and $r \mapsto (S^t)_r$ are piecewise $1/2$-Hölder continuous in the sense that there are two finite partitions $(I_\alpha)_{\alpha \in A}$ and $(J_\beta)_{\beta \in B}$ of $\mathfrak{X}_1$ and $\mathfrak{X}_2$, respectively, such that, for all $\alpha \in A$ and $\beta \in B$, we have

$$\|S_k - S_l\|_2 \le C_\alpha |k - l|^{1/2} \ \text{ for } k, l \in I_\alpha, \qquad \|(S^t)_q - (S^t)_r\|_2 \le D_\beta |q - r|^{1/2} \ \text{ for } q, r \in J_\beta.$$

There is a similar condition for **(A3)** if $\mathfrak{X}_1 = [p]$ and $\mathfrak{X}_2 = [n]$ for some $p, n \in \mathbb{N}$ and the measures π_1 and π_2 are the (unnormalized) counting measures on $[p]$ and $[n]$, respectively.

2.2 Local law for random Gram matrices

In this subsection, we state our results on random Gram matrices. We now set $\mathfrak{X}_1 = [p]$, $\mathfrak{X}_2 = [n]$ as well as π_1 and π_2 the (unnormalized) counting measures on $[p]$ and $[n]$, respectively. In particular, $\pi_1(\mathfrak{X}_1) = p$ and $\pi_2(\mathfrak{X}_2) = n$.

Assumptions 2.5. Let $X = (x_{kq})_{k,q}$ be a $p \times n$ random matrix with independent, centered entries and variance matrix $S = (s_{kq})_{k,q}$, i.e., $\mathbb{E}\, x_{kq} = 0$ and $s_{kq} := \mathbb{E}|x_{kq}|^2$ for $k \in [p]$, $q \in [n]$. Moreover, we assume that **(A1)**, **(A2)** and **(A3)** in Assumptions 2.2 and the following conditions are satisfied.

(B1) The variances are bounded in the sense that there exists $s^* > 0$ such that

$$s_{kq} \leq \frac{s^*}{p+n} \qquad \text{for } k \in [p],\ q \in [n].$$

(B2) All entries of X have bounded moments in the sense that there are $\mu_m > 0$ for $m \geq 3$ such that

$$\mathbb{E}|x_{kq}|^m \leq \mu_m s_{kq}^{m/2} \qquad \text{for all } k \in [p],\ q \in [n].$$

The sequence $(\mu_m)_{m\geq 3}$ in **(B2)** is also considered a model parameter.

Since **(B1)** implies **(A4)**, we can apply Theorem 2.3. By its first part, for every $\delta > 0$, there are $\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_K \in [\delta, \infty)$ for some $K \in \mathbb{N}$ such that

$$\operatorname{supp} \left\langle \nu^d|_{[\delta,\infty)} \right\rangle_1 = \bigcup_{i=1}^K [\alpha_i, \beta_i], \qquad \alpha_j < \beta_j < \alpha_{j+1}$$

and $\rho_* > 0$ depending only on the model parameters and δ such that $\beta_i - \alpha_i \geq 2\rho_*$ for all $i \in [K]$. For $\rho \in [0, \rho_*)$, we introduce the *local gap size* Δ_ρ via

$$\Delta_\rho(E) := \begin{cases} \alpha_{i+1} - \beta_i, & \text{if } \beta_i - \rho \leq E \leq \alpha_{i+1} + \rho \text{ for some } i \in [K], \\ 1, & \text{if } E \leq \alpha_1 + \rho \text{ or } E \geq \beta_K - \rho, \\ 0, & \text{otherwise.} \end{cases} \tag{2.7}$$

For $\delta, \gamma > 0$, we define the spectral domain $\mathbb{D}_{\delta,\gamma} := \{\zeta \in \mathbb{H} \colon |\zeta| \geq \delta,\ \operatorname{Im}\zeta \geq p^{-1+\gamma}\}$. We introduce the resolvent $R(\zeta) := (XX^* - \zeta)^{-1}$ of XX^* at $\zeta \in \mathbb{H}$ and denote its entries by $R_{kl}(\zeta)$ for $k, l \in [p]$.

Theorem 2.6 (Local law for Gram matrices)**.** *Let Assumptions 2.5 hold true. Fix $\delta > 0$ and $\gamma \in (0,1)$. Then there is $\rho \in (0, \rho_*)$ depending only on the model parameters and δ such that if we define $\kappa = \kappa^{(p)} \colon \mathbb{H} \to (0, \infty]$ through*

$$\kappa(\zeta) = \left(\Delta_\rho(\operatorname{Re}\zeta)^{1/3} + \langle \operatorname{Im} m(\zeta)\rangle\right)^{-1}$$

then, for each $\varepsilon > 0$ and $D > 0$, there is a constant $C_{\varepsilon,D} > 0$ such that

$$\mathbb{P}\left(\sup_{\substack{\zeta \in \mathbb{D}_{\delta,\gamma} \\ k,l \in [p]}} p^{-\varepsilon} \left| R_{kl}(\zeta) - m_k(\zeta)\delta_{kl} \right| \leq \sqrt{\frac{\langle \operatorname{Im} m(\zeta)\rangle}{p \operatorname{Im}\zeta}} + \min\left\{ \frac{1}{\sqrt{p\operatorname{Im}\zeta}}, \frac{\kappa(\zeta)}{p\operatorname{Im}\zeta} \right\} \right) \geq 1 - \frac{C_{\varepsilon,D}}{p^D}. \tag{2.8a}$$

Furthermore, for any $\varepsilon > 0$ and $D > 0$, there is a constant $C_{\varepsilon,D} > 0$ such that, for any deterministic vector $w \in \mathbb{C}^p$ satisfying $\max_{k\in[p]}|w_k| \leq 1$, we have

$$\mathbb{P}\left(\sup_{\zeta \in \mathbb{D}_{\delta,\gamma}} \left| \frac{1}{p} \sum_{k=1}^p w_k \Big(R_{kk}(\zeta) - m_k(\zeta) \Big) \right| \leq p^\varepsilon \min\left\{ \frac{1}{\sqrt{p\operatorname{Im}\zeta}}, \frac{\kappa(\zeta)}{p\operatorname{Im}\zeta} \right\} \right) \geq 1 - \frac{C_{\varepsilon,D}}{p^D}. \tag{2.8b}$$

The constant $C_{\varepsilon,D}$ in (2.8) *depends only on the model parameters as well as δ and γ in addition to ε and D.*

Remark 2.7. (i) (Corollaries of the local law) In the same way as in [2] and in [5], the standard corollaries of a local law – convergence of cumulative distribution function, rigidity of eigenvalues, anisotropic law and delocalization of eigenvectors – may be proven.

(ii) (Local law in the bulk and away from $\operatorname{supp}\nu$) In the bulk, Theorem 2.6 has already been proven in [5]. Away from $\operatorname{supp}\nu$, the convergence rate in (2.8a) and (2.8b) can be improved and thus the condition $\operatorname{Im}\zeta \geq p^{-1+\gamma}$ can be removed. See [5] for Gram matrices and [4] for Kronecker matrices.

(iii) (Local law close to zero) Strengthening the assumption **(A2)**, we have proven the local law close to zero in the cases, $n = p$ and $|p - n| \geq cn$, in [5].

3 Quadratic vector equation

In this section, we translate (2.3) into a quadratic vector equation of [1] (see (3.2) below) and show that Proposition 2.1 trivially follows from [1]. However, the singularity analysis in [1] has to be changed essentially due to the violation of the uniform primitivity condition, **A3** in [1], on $\boldsymbol{S}$ (cf. (3.1) below) in our setup.

Let $\mathfrak{X} := \mathfrak{X}_1 \sqcup \mathfrak{X}_2$ be the disjoint union of $\mathfrak{X}_1$ and $\mathfrak{X}_2$ and π the probability measure defined through

$$\pi(A \sqcup B) = (\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))^{-1}(\pi_1(A) + \pi_2(B)), \qquad \text{for } A \subset \mathfrak{X}_1,\ B \subset \mathfrak{X}_2.$$

Moreover, we denote the set of bounded measurable functions $\mathfrak{X} \to \mathbb{C}$ by $\mathscr{B} := \{\boldsymbol{w}\colon \mathfrak{X} \to \mathbb{C}\colon \|\boldsymbol{w}\|_\infty := \sup_{x\in\mathfrak{X}}|\boldsymbol{w}(x)| < \infty\}$ with the supremum norm $\|\cdot\|_\infty$. Finally, on $\mathscr{B} = \mathscr{B}_1 \oplus \mathscr{B}_2$, we define the linear operator $\boldsymbol{S}\colon \mathscr{B} \to \mathscr{B}$ through

$$\boldsymbol{S} := \begin{pmatrix} 0 & S \\ S^t & 0 \end{pmatrix}, \qquad \text{i.e.,} \quad \boldsymbol{S}\boldsymbol{w} = S(\boldsymbol{w}|_{\mathfrak{X}_2}) + S^t(\boldsymbol{w}|_{\mathfrak{X}_1}) \quad \text{for } \boldsymbol{w} \in \mathscr{B}. \tag{3.1}$$

Here, we consider $S(\boldsymbol{w}|_{\mathfrak{X}_2})$ and $S^t(\boldsymbol{w}|_{\mathfrak{X}_1})$ as functions $\mathfrak{X} \to \mathbb{C}$, extended by zero outside of $\mathfrak{X}_1$ and $\mathfrak{X}_2$, respectively. Instead of (2.3), we study the quadratic vector equation (QVE)

$$-\frac{1}{\boldsymbol{m}} = z + \boldsymbol{S}\boldsymbol{m} \tag{3.2}$$

for $z \in \mathbb{H}$. Here, we used the change of variables $z^2 = \zeta$. We now explain how $\boldsymbol{m}$ and m are related. If $\boldsymbol{m}$ is a solution of (3.2) then $m_1 := \boldsymbol{m}|_{\mathfrak{X}_1}$ and $m_2 := \boldsymbol{m}|_{\mathfrak{X}_2}$ satisfy $-m_1^{-1} = z + Sm_2$ and $-m_2^{-1} = z + S^t m_1$. Solving the second equation for m_2, plugging the result into the first relation and choosing $z = \sqrt{\zeta} \in \mathbb{H}$, we see that m defined through

$$m(\zeta) = \frac{m_1(\sqrt{\zeta})}{\sqrt{\zeta}} \tag{3.3}$$

for $\zeta \in \mathbb{H}$ is a solution of (2.3). If $\boldsymbol{m}$ has positive imaginary part then m as well.

For $\boldsymbol{u} \in \mathscr{B}$, we write $\boldsymbol{u}_x := \boldsymbol{u}(x)$ with $x \in \mathfrak{X}$. For $\boldsymbol{u}, \boldsymbol{w} \in \mathscr{B}$, we denote the scalar product of $\boldsymbol{u}$ and $\boldsymbol{w}$ and the average of $\boldsymbol{u}$ by

$$\langle \boldsymbol{u}, \boldsymbol{w} \rangle := \int_{\mathfrak{X}} \overline{\boldsymbol{u}_x}\, \boldsymbol{w}_x \pi(\mathrm{d}x), \quad \langle \boldsymbol{u} \rangle := \langle 1, \boldsymbol{u} \rangle = \int_{\mathfrak{X}} \boldsymbol{u}_x \pi(\mathrm{d}x). \tag{3.4}$$

We also introduce the Hilbert space $L^2(\pi) := \{\boldsymbol{u}\colon \mathfrak{X} \to \mathbb{C}\colon \langle \boldsymbol{u}, \boldsymbol{u} \rangle < \infty\}$. The operator $\boldsymbol{S}$ is symmetric on $\mathscr{B}$ with respect to $\langle \cdot\,,\cdot \rangle$ and positivity preserving, as $s_{kr} \geq 0$ for all $k \in \mathfrak{X}_1$ and $r \in \mathfrak{X}_2$. Therefore, by Theorem 2.1 in [1], there exists $\boldsymbol{m}\colon \mathbb{H} \to \mathscr{B}$ which satisfies (3.2) for all $z \in \mathbb{H}$. This function is unique if we require that the solution of (3.2) satisfies $\operatorname{Im}\boldsymbol{m}(z) > 0$ for $z \in \mathbb{H}$. Moreover, $\boldsymbol{m}\colon \mathbb{H} \to \mathscr{B}$ is analytic and, for all $z \in \mathbb{H}$, we have

$$\|\boldsymbol{m}(z)\|_2 \leq 2|z|^{-1}. \tag{3.5}$$

Furthermore, for all $x \in \mathfrak{X}$, there are symmetric probability measures $\boldsymbol{\rho}_x$ on $\mathbb{R}$ such that

$$\boldsymbol{m}_x(z) = \int_{\mathbb{R}} \frac{1}{\tau - z} \boldsymbol{\rho}_x(\mathrm{d}\tau) \tag{3.6}$$

for all $z \in \mathbb{H}$ [1]. That means that $\boldsymbol{m}_x$ is the Stieltjes transform of $\boldsymbol{\rho}_x$. By (2.7) in [1], the definition of Σ in (2.5) and $\|\boldsymbol{S}\| = \|\boldsymbol{S}\|_{\mathscr{B}\to\mathscr{B}} = \max\{\|S\|_{\mathscr{B}_2\to\mathscr{B}_1}, \|S^t\|_{\mathscr{B}_1\to\mathscr{B}_2}\}$, the support of $\boldsymbol{\rho}_x$ is contained in $[-\Sigma^{1/2}, \Sigma^{1/2}]$.

Proof of Proposition 2.1. The existence of m directly follows from the transform in (3.3) and the existence of $\boldsymbol{m}$. The uniqueness of m and the existence of ν_k, $k \in \mathfrak{X}_1$, are obtained as in the proof of Theorem 2.1 in [5]. □

The special structure of $\boldsymbol{S}$ (cf. (3.1)) implies an important symmetry of the solution $\boldsymbol{m}$. We multiply (3.2) by $\boldsymbol{m}$ and take the scalar product of the result with $\boldsymbol{e}_- \in \mathscr{B}$ defined through $\boldsymbol{e}_-(k) = 1$ if $k \in \mathfrak{X}_1$ and $\boldsymbol{e}_-(q) = -1$ if $q \in \mathfrak{X}_2$. As $\langle \boldsymbol{e}_-, \boldsymbol{m}(\boldsymbol{S}\boldsymbol{m})\rangle = 0$, we have

$$z\langle \boldsymbol{e}_-, \boldsymbol{m}\rangle = -\langle \boldsymbol{e}_-\rangle = -\frac{\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)}{\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2)}, \qquad \text{for all } z \in \mathbb{H}. \tag{3.7}$$

Assumptions 3.1. In the remainder of this section, we assume that **(A1)**, **(A2)**, **(A4)** and the following condition hold true:

(C1) There are $\tilde{\delta} > 0$ and $\Phi > 0$ such that for all $z \in \mathbb{H}$ satisfying $|z| \geq \tilde{\delta}$, we have

$$\|\boldsymbol{m}(z)\|_\infty \leq \Phi.$$

Remark 3.2 (Relation between **(A3)** and **(C1)**)**.** By slightly adapting the proofs of Theorem 6.1 (ii) and Proposition 6.6 in [1], we see that, by **(A3)**, for each $\tilde{\delta} > 0$, there is $\Phi_{\tilde{\delta}} > 0$ such that **(C1)** is satisfied with a constant $\Phi \equiv \Phi_{\tilde{\delta}}$.

Since our estimates in this section will be uniform in all models that satisfy **(A1)**, **(A2)**, **(A4)** and **(C1)** with the same constants, we introduce the following notion.

Convention 3.3 (Comparison relation)**.** *For nonnegative scalars or vectors f and g, we will use the notation $f \lesssim g$ if there is a constant $c > 0$, depending only on π_*, π^* in **(A1)**, $L_1, L_2, \kappa_1, \kappa_2$ in **(A2)**, Ψ_1, Ψ_2 in **(A4)** as well as $\tilde{\delta}$ and Φ in **(C1)**, such that $f \leq cg$. Moreover, we write $f \sim g$ if both, $f \lesssim g$ and $f \gtrsim g$, hold true.*

3.1 Hölder continuity and analyticity

We recall Σ from (2.5) and introduce the set $\mathbb{H}_{\tilde{\delta}}^{\Sigma} := \{z \in \mathbb{H} \colon 2\tilde{\delta} \leq |z| \leq 10\Sigma^{1/2}\}$ and its closure $\overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$.

Proposition 3.4 (Regularity of $\boldsymbol{m}$)**.** *Assume **(A1)**, **(A2)**, **(A4)** and **(C1)**.*

(i) The restriction $\boldsymbol{m} \colon \mathbb{H}_{\tilde{\delta}}^{\Sigma} \to \mathscr{B}$ is uniformly $1/3$-Hölder continuous, i.e.,

$$\|\boldsymbol{m}(z) - \boldsymbol{m}(z')\|_\infty \lesssim |z - z'|^{1/3} \tag{3.8}$$

for all $z, z' \in \mathbb{H}_{\tilde{\delta}}^{\Sigma}$. In particular, $\boldsymbol{m}$ can be uniquely extended to a uniformly $1/3$-Hölder continuous function $\overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma} \to \mathscr{B}$, which we also denote by $\boldsymbol{m}$.

(ii) The measure $\boldsymbol{\rho}$ from (3.6) is absolutely continuous, i.e., there is a function $\boldsymbol{\rho}^d \colon \mathfrak{X} \times \mathbb{R} \setminus (-2\tilde{\delta}, 2\tilde{\delta}) \to [0, \infty)$, $(x, \tau) \mapsto \boldsymbol{\rho}_x^d(\tau)$ such that

$$\left(\boldsymbol{\rho}_x|_{\mathbb{R}\setminus(-2\tilde{\delta},2\tilde{\delta})}\right)(\mathrm{d}\tau) = \boldsymbol{\rho}_x^d(\tau)\mathrm{d}\tau, \qquad \textit{for all } x \in \mathfrak{X}. \tag{3.9}$$

The components $\boldsymbol{\rho}_x^d$ are comparable with each other, i.e., $\boldsymbol{\rho}_x^d(\tau) \sim \boldsymbol{\rho}_y^d(\tau)$ for all $x, y \in \mathfrak{X}$ and $\tau \in \mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}]$. Moreover, the function $\boldsymbol{\rho}^d \colon \mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}] \to \mathscr{B}$ is uniformly $1/3$-Hölder continuous, symmetric in τ, $\boldsymbol{\rho}^d(\tau) = \boldsymbol{\rho}^d(-\tau)$, and real-analytic around any $\tau \in \mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}]$ apart from points $\tau \in \operatorname{supp}\langle \boldsymbol{\rho}^d \rangle$, where $\boldsymbol{\rho}^d(\tau) = 0$.

A similar result has been obtained in Theorem 2.4 in [1] essentially relying on the uniform primitivity assumption **A3** in [1]. For discrete $\mathfrak{X}_1$ and $\mathfrak{X}_2$ without assuming **(C1)**, Lemma 3.8 in [5] shows Hölder continuity of $\langle \boldsymbol{m} \rangle$ instead of $\boldsymbol{m}$ with a smaller exponent than $1/3$. Both conditions, **A3** in [1] and the discreteness of $\mathfrak{X}_1$ and $\mathfrak{X}_2$, are violated in our setup. However, based on the proof of Theorem 2.4 in [1], we now explain how to extend the arguments of [1] and [5] to show Proposition 3.4.

Lemma 3.5. *Uniformly for all $z \in \mathbb{H}_{\tilde{\delta}}^{\Sigma}$, we have*

$$|\boldsymbol{m}(z)| \sim 1, \tag{3.10}$$

$$\operatorname{Im} \boldsymbol{m}(z) \sim \langle \operatorname{Im} \boldsymbol{m}(z) \rangle. \tag{3.11}$$

Using the arguments in the proof of Lemma 5.4 in [1], Lemma 3.5 follows immediately from **(A2)**, **(C1)** and (3.2). Here, as in the proof of Lemma 3.1 in [5], the uniform primitivity assumption **A3** of [1] has to be replaced by (B') in [5], which is a direct consequence of **(A2)**.

The Hölder continuity and the analyticity of $\boldsymbol{m}$ and hence $\boldsymbol{\rho}^d$ will be consequences of analyzing the perturbed QVE

$$-\frac{1}{\boldsymbol{g}} = z + \boldsymbol{S}\boldsymbol{g} + \boldsymbol{d} \tag{3.12}$$

for $z \in \mathbb{H}$ and $\boldsymbol{d} = z - z'$ as well as the stability operator $\boldsymbol{B}$ defined through

$$\boldsymbol{B}(z)\boldsymbol{u} = \frac{|\boldsymbol{m}(z)|^2}{\boldsymbol{m}(z)^2}\boldsymbol{u} - \boldsymbol{F}(z)\boldsymbol{u}, \tag{3.13}$$

where $\boldsymbol{F}(z) \colon \mathscr{B} \to \mathscr{B}$ is defined through $\boldsymbol{F}(z)\boldsymbol{u} = |\boldsymbol{m}(z)|\boldsymbol{S}\left(|\boldsymbol{m}(z)|\boldsymbol{u}\right)$ for any $\boldsymbol{u} \in \mathscr{B}$ (cf. [1, 5]). Correspondingly, we introduce $F(z) \colon \mathscr{B}_2 \to \mathscr{B}_1$ via $F(z)w = |m_1(z)|S(|m_2(z)|w)$ for $w \in \mathscr{B}_2$ and $F^t(z) \colon \mathscr{B}_1 \to \mathscr{B}_2$ via $F^t(z)u = |m_2(z)|S^t(|m_1(z)|u)$ for $u \in \mathscr{B}_1$.

To formulate the key properties of $\boldsymbol{F}$ and $\boldsymbol{B}$, we now introduce some notation. The operator norms for operators on $\mathscr{B}$ and $L^2(\pi)$ are denoted by $\|\cdot\|_\infty$ and $\|\cdot\|_2$, respectively. If $T \colon L^2 \to L^2$ is a compact self-adjoint operator then the *spectral gap* $\operatorname{Gap}(T)$ is the difference between the two largest eigenvalues of $|T|$. We remark that S and hence FF^t are compact operators due to **(A4)**.

Lemma 3.6 (Properties of $\boldsymbol{F}$)**.** *The eigenspace of $\boldsymbol{F}$ associated to $\|\boldsymbol{F}\|_2$ is one-dimensional and spanned by a unique $L^2(\pi)$-normalized positive $\boldsymbol{f}_+ \in \mathscr{B}$. The eigenspace associated to $-\|\boldsymbol{F}\|_2$ is one-dimensional and spanned by $\boldsymbol{f}_- := \boldsymbol{f}_+ \boldsymbol{e}_- \in \mathscr{B}$. We have*

$$\boldsymbol{f}_+ \sim 1 \tag{3.14}$$

uniformly for $z \in \mathbb{H}_{\tilde{\delta}}^{\Sigma}$. There is $\varepsilon \sim 1$ such that

$$\|\boldsymbol{F}\boldsymbol{u}\|_2 \le (\|\boldsymbol{F}\|_2 - \varepsilon)\|\boldsymbol{u}\|_2 \tag{3.15}$$

uniformly for $z \in \mathbb{H}_{\tilde{\delta}}^{\Sigma}$ and for all $\boldsymbol{u} \in \mathscr{B}$ satisfying $\langle \boldsymbol{f}_+, \boldsymbol{u} \rangle = 0$ and $\langle \boldsymbol{f}_-, \boldsymbol{u} \rangle = 0$. Furthermore, we have $\|\boldsymbol{F}\|_2 \le 1$, $\operatorname{Gap}(F(z)F^t(z)) \sim 1$ uniformly for $z \in \mathbb{H}_{\tilde{\delta}}^{\Sigma}$.

Lemma 3.6 is a consequence of the proof of Lemma 3.3 in [5] with $r = |\boldsymbol{m}|$ and (3.10).

Lemma 3.7. *Uniformly for $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}}$, we have*

$$\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim \frac{1}{\langle \operatorname{Im} \boldsymbol{m}(z)\rangle^2}. \tag{3.16}$$

Proof. We describe the modifications in the proof of Lemma 3.5 in [5] necessary to obtain (3.16). We remark that (3.10) in [5] holds true due to **(A4)**.

Let $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}}$. Taking the real part in (3.2), using (3.10) and Lemma 3.6, we obtain the bound $\left\|\operatorname{Re} \boldsymbol{m}|\boldsymbol{m}|^{-1}\right\|_2 \geq |\operatorname{Re} z|\|\boldsymbol{m}\|_2/2 \gtrsim |\operatorname{Re} z|$. Therefore, using $\langle(\operatorname{Im}\boldsymbol{m})^2\rangle \geq \langle \operatorname{Im}\boldsymbol{m}\rangle^2$ by Jensen's inequality, we obtain (3.27) in [5] with $\kappa = 2$. Employing $\operatorname{Gap}(F(z)F^t(z)) \sim 1$, we get $\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim (\operatorname{Re} z)^{-2}\langle \operatorname{Im}\boldsymbol{m}(z)\rangle^{-2}$. As $\|\boldsymbol{B}^{-1}(z)\|_2 \leq (1-\|\boldsymbol{F}(z)\|_2)^{-1} \lesssim (\operatorname{Im} z)^{-1}$ by (3.21) in [5] we conclude from $\operatorname{Im}\boldsymbol{m} \lesssim \min\{1, (\operatorname{Im} z)^{-1}\}$ that $\|\boldsymbol{B}^{-1}(z)\|_\infty \lesssim |z|^{-2}\langle \operatorname{Im}\boldsymbol{m}(z)\rangle^{-2}$. This concludes the proof of (3.16) since $|z| \geq 2\tilde{\delta}$. ☐

Note that if $\boldsymbol{\rho}$ has a density $\boldsymbol{\rho}^d$ around a point τ_0 then, uniformly for τ in a neighbourhood of τ_0, we have

$$\boldsymbol{\rho}^d(\tau) = \pi^{-1} \lim_{\eta \downarrow 0} \operatorname{Im} \boldsymbol{m}(\tau + \mathrm{i}\eta). \tag{3.17}$$

Proof of Proposition 3.4. Following the proof of Proposition 7.1 in [1] yields the uniform $1/3$-Hölder continuity of $\boldsymbol{m}$ and $\boldsymbol{\rho}^d$. In this proof, the estimate (5.40b) has to be replaced by (3.16). Furthermore, (3.11) substitutes Proposition 5.3 (ii) in [1], in particular, $\boldsymbol{\rho}^d_x(\tau) \sim \boldsymbol{\rho}^d_y(\tau)$. We remark that now the same proofs extend Lemma 3.5, Lemma 3.6 and Lemma 3.7 to all $z \in \overline{\mathbb{H}}^{\Sigma}_{\tilde{\delta}}$. Hence, the proof of Corollary 7.6 in [1] yields the analyticity using (3.17) for $\tau \in \mathbb{R} \cap \overline{\mathbb{H}}^{\Sigma}_{\tilde{\delta}}$. ☐

3.2 Singularities of $\boldsymbol{\rho}^d$ and the cubic equation

We now study the behaviour of $\boldsymbol{\rho}^d$ near points $\tau \in \mathbb{R}$, where $\boldsymbol{\rho}^d$ is not analytic. Theorem 2.6 in [1] describes the density near the edges and the cusps as well as the transition between the bulk and the singularity regimes in a quantitative manner. The same results hold for $\boldsymbol{\rho}^d$ as well:

Proposition 3.8. *We assume **(A1)**, **(A2)**, **(A4)** and **(C1)**. Then all statements of Theorem 2.6 in [1] hold true on $\mathbb{R} \setminus [-2\tilde{\delta}, 2\tilde{\delta}]$.*

For the proof of Proposition 3.8 we follow Chapter 8 and 9 in [1] which contain the proof of the analogue of Proposition 3.8, Theorem 2.6 in [1], and describe the necessary changes as well as the main philosophy.

The shape of the singularities of $\boldsymbol{m}$ as well as the stability of the QVE (cf. Chapter 10 in [1]) will be a consequence of the stability of a cubic equation. We note that similar as in Lemma 8.1 of [1], the following properties of the stability operator $\boldsymbol{B} = \boldsymbol{B}(z)$ defined in (3.13) can be proven. There is $\varepsilon_* \sim 1$ such that for $z \in \overline{\mathbb{H}}^{\Sigma}_{\tilde{\delta}}$ satisfying $\langle \operatorname{Im}\boldsymbol{m}(z)\rangle \leq \varepsilon_*$, $\boldsymbol{B}$ has a unique eigenvalue $\beta = \beta(z)$ of smallest modulus and $|\beta'| - |\beta| \gtrsim 1$ for all $\beta' \in \operatorname{Spec}(\boldsymbol{B}) \setminus \{\beta\}$. The eigenspace associated to β is one-dimensional and there is a unique vector $\boldsymbol{b} = \boldsymbol{b}(z) \in \mathscr{B}$ in this eigenspace such that $\langle \boldsymbol{b}(z), \boldsymbol{f}_+\rangle = 1$.

Let $z \in \overline{\mathbb{H}}^{\Sigma}_{\tilde{\delta}}$ such that $\langle \operatorname{Im}\boldsymbol{m}(z)\rangle \leq \varepsilon_*$ and $\boldsymbol{g} \in \mathscr{B}$ satisfy the perturbed QVE, (3.12), at z. We define

$$\Theta(z) := \left\langle \frac{\bar{\boldsymbol{b}}(z)}{\langle \boldsymbol{b}(z)^2\rangle}, \frac{\boldsymbol{g} - \boldsymbol{m}(z)}{|\boldsymbol{m}(z)|}\right\rangle. \tag{3.18}$$

By possibly shrinking $\varepsilon_* \sim 1$, we obtain that if $\|\boldsymbol{g} - \boldsymbol{m}(z)\|_\infty \leq \varepsilon_*$ then it can be shown as in Proposition 8.2 in [1] that Θ satisfies

$$\mu_3\Theta^3 + \mu_2\Theta^2 + \mu_1\Theta + \langle |\boldsymbol{m}|\bar{\boldsymbol{b}}, \boldsymbol{d}\rangle = \kappa\left((\boldsymbol{g}-\boldsymbol{m})/|\boldsymbol{m}|, \boldsymbol{d}\right), \tag{3.19}$$

where μ_1, μ_2 and μ_3, which depend only on S and z, as well as κ are given in [1].

The main ingredient that needs to be changed in our setup is the estimate in (8.13) of [1]. It gives a lower bound on the nonnegative quadratic form

$$\mathcal{D}(\boldsymbol{w}) := \Big\langle \boldsymbol{Q}_+\boldsymbol{w}\ ,\ (\|\boldsymbol{F}\|_2 + \boldsymbol{F})\,(1-\boldsymbol{F})^{-1}\,\boldsymbol{Q}_+\boldsymbol{w}\Big\rangle \tag{3.20}$$

for $\boldsymbol{w} \in \mathscr{B}$, where the projection $\boldsymbol{Q}_+$ is defined through $\boldsymbol{Q}_+\boldsymbol{w} := \boldsymbol{w} - \langle \boldsymbol{f}_+ , \boldsymbol{w}\rangle \boldsymbol{f}_+$. For some $c(z) > 0$ and all $\boldsymbol{w} \in \mathscr{B}$, this lower bounds reads as follows

$$\mathcal{D}(\boldsymbol{w}) \geq c(z)\|\boldsymbol{Q}_+\boldsymbol{w}\|_2^2. \tag{3.21}$$

However, in our setup, owing to the second unstable direction $\boldsymbol{f}_- \perp \boldsymbol{f}_+$, $\boldsymbol{F}\boldsymbol{f}_- = -\|\boldsymbol{F}\|_2\boldsymbol{f}_-$, we have $\mathcal{D}(\boldsymbol{f}_-) = 0$ which contradicts (3.21). In [1], the estimate (3.21) is only used to obtain

$$|\mu_3(z)| + |\mu_2(z)| \gtrsim 1 \tag{3.22}$$

(cf. (8.34) in [1]) for all $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ satisfying $\langle \operatorname{Im}\boldsymbol{m}(z)\rangle \leq \varepsilon_*$ and $\|\boldsymbol{g} - \boldsymbol{m}(z)\|_\infty \leq \varepsilon_*$ for $\varepsilon_* \sim 1$ small enough. In fact, it is shown above (8.50) in [1] that

$$|\mu_3| \gtrsim \psi + \mathcal{O}(\alpha) \qquad |\mu_2| \gtrsim |\sigma| + \mathcal{O}(\alpha). \tag{3.23}$$

Here, we introduced the notations $\psi := \mathcal{D}(\boldsymbol{p}\boldsymbol{f}_+^2)$ with $\boldsymbol{p} := \operatorname{sign}(\operatorname{Re}\boldsymbol{m})$ as well as $\alpha := \langle \boldsymbol{f}_+ \operatorname{Im}\boldsymbol{m}/|\boldsymbol{m}|\rangle$ and $\sigma := \langle \boldsymbol{f}_+ , \boldsymbol{p}\boldsymbol{f}_+^2\rangle$. The proof used in [1] to show (3.23) works in our setup as well. Since $\alpha = \langle \boldsymbol{f}_+ \operatorname{Im}\boldsymbol{m}/|\boldsymbol{m}|\rangle \sim \langle \operatorname{Im}\boldsymbol{m}\rangle \leq \varepsilon_*$ by (3.10) and (3.14), we conclude that $|\mu_3| + |\mu_2| \gtrsim \psi + |\sigma|$ for $\varepsilon_* \sim 1$ small enough. Hence, (3.22) is a consequence of

Lemma 3.9 (Stability of the cubic equation)**.** *There exists $\varepsilon_* \sim 1$ such that*

$$\psi(z) + \sigma^2(z) \sim 1 \tag{3.24}$$

uniformly for all $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ satisfying $\langle \operatorname{Im}\boldsymbol{m}(z)\rangle \leq \varepsilon_*$.

Proof. We first remark that due to (3.10), (3.11) and possibly shrinking $\varepsilon_* \sim 1$ we can assume

$$|\operatorname{Re}\boldsymbol{m}(z)| \sim 1 \tag{3.25}$$

for $z \in \overline{\mathbb{H}}_{\tilde{\delta}}^{\Sigma}$ satisfying $\langle \operatorname{Im}\boldsymbol{m}(z)\rangle \leq \varepsilon_*$. Second, owing to (3.15), for all $\boldsymbol{w} \in \mathscr{B}$, we have the following analogue of (3.21)

$$\mathcal{D}(\boldsymbol{w}) \gtrsim \|\boldsymbol{Q}_\pm\boldsymbol{w}\|_2^2, \tag{3.26}$$

where $\boldsymbol{Q}_\pm$ is the projection onto the orthogonal complement of $\boldsymbol{f}_+$ and $\boldsymbol{f}_-$, i.e, $\boldsymbol{Q}_\pm\boldsymbol{w} = \boldsymbol{w} - \langle \boldsymbol{f}_+ , \boldsymbol{w}\rangle\boldsymbol{f}_+ - \langle \boldsymbol{f}_- , \boldsymbol{w}\rangle\boldsymbol{f}_-$. Note that (3.15) also yields the upper bound $\mathcal{D}(\boldsymbol{w}) \lesssim \|\boldsymbol{Q}_+\boldsymbol{w}\|_2^2$ and hence the upper bound in (3.24) by (3.14). Therefore, it suffices to prove the lower bound in (3.24). A straightforward computation starting from (3.26) and using $\boldsymbol{f}_- = \boldsymbol{e}_-\boldsymbol{f}_+$ yields

$$\psi + \sigma^2 = \mathcal{D}(\boldsymbol{p}\boldsymbol{f}_+^2) + \langle \boldsymbol{p}\boldsymbol{f}_+^3\rangle^2 \gtrsim \|\boldsymbol{p}\boldsymbol{f}_+^2 - \langle \boldsymbol{f}_- , \boldsymbol{p}\boldsymbol{f}_+^2\rangle\boldsymbol{f}_-\|_2^2 = \Big\langle \boldsymbol{f}_+^2\left(\boldsymbol{p}\boldsymbol{f}_+ - \langle \boldsymbol{p}\boldsymbol{e}_-\boldsymbol{f}_+^3\rangle\boldsymbol{e}_-\right)^2\Big\rangle. \tag{3.27}$$

Using (3.14), (3.25) and $|\operatorname{Re}\boldsymbol{m}| = \boldsymbol{p}\operatorname{Re}\boldsymbol{m}$, we conclude

$$\begin{aligned}\psi + \sigma^2 &\gtrsim \Big\langle (\operatorname{Re}\boldsymbol{m})^2\left(\boldsymbol{p}\boldsymbol{f}_+ - \langle \boldsymbol{p}\boldsymbol{e}_-\boldsymbol{f}_+^3\rangle\boldsymbol{e}_-\right)^2\Big\rangle \\ &\geq \langle \boldsymbol{f}_+|\operatorname{Re}\boldsymbol{m}|\rangle\left(\langle \boldsymbol{f}_+|\operatorname{Re}\boldsymbol{m}|\rangle + 2\langle \boldsymbol{p}\boldsymbol{e}_-\boldsymbol{f}_+^3\rangle\langle \boldsymbol{e}_-\rangle\operatorname{Re}\frac{1}{z}\right)\end{aligned} \tag{3.28}$$

Here, we employed Jensen's inequality and (3.7) in the second step. Since $z \in \overline{\mathbb{H}}^{\Sigma}_{\tilde{\delta}}$ and $\langle \boldsymbol{e}_- \rangle = 0$ for $\pi_1(\mathfrak{X}_1) = \pi_2(\mathfrak{X}_2)$, there exists $\iota_* \sim 1$ such that the last factor on the right-hand side of (3.28) is bounded from below by $\langle \boldsymbol{f}_+ |\mathrm{Re}\, \boldsymbol{m}|\rangle/2$ for all $z \in \overline{\mathbb{H}}^{\Sigma}_{\tilde{\delta}}$ and $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| \leq \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$. Since $\langle \boldsymbol{f}_+ |\mathrm{Re}\, \boldsymbol{m}|\rangle^2 \gtrsim 1$ by (3.14) and (3.25), this finishes the proof of (3.24) for $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| \leq \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$. For the proof of (3.24) in the remaining regime, $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| > \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$, we introduce $\boldsymbol{y} := \boldsymbol{e}_- \boldsymbol{p} \boldsymbol{f}_+$ and use $\boldsymbol{y}^2 = \boldsymbol{f}_+^2 \sim 1$ and $\big(\boldsymbol{y} + \langle \boldsymbol{y}^3 \rangle\big)^2 \lesssim 1$ by (3.14) to obtain from (3.27) the bound

$$\psi + \sigma^2 \gtrsim \left\langle \big(\boldsymbol{y} - \langle \boldsymbol{y}^3 \rangle\big)^2 \big(\boldsymbol{y} + \langle \boldsymbol{y}^3 \rangle\big)^2 \right\rangle = \left\langle \big((\boldsymbol{y}^2 - 1) + (1 - \langle \boldsymbol{y}^3 \rangle^2)\big)^2 \right\rangle \geq \left\langle \big(\boldsymbol{y}^2 - 1\big)^2 \right\rangle. \quad (3.29)$$

Here, we used $\langle \boldsymbol{y}^2 \rangle = \langle \boldsymbol{f}_+^2 \rangle = 1$ and $\big(1 - \langle \boldsymbol{y}^3 \rangle^2\big)^2 \geq 0$. Since $0 = \langle \boldsymbol{f}_-, \boldsymbol{f}_+ \rangle = \langle \boldsymbol{e}_- \boldsymbol{y}^2 \rangle$, using (3.29), we conclude

$$\langle \boldsymbol{e}_- \rangle^2 = \langle \boldsymbol{e}_- (1 - \boldsymbol{y}^2) \rangle^2 \leq \langle (1 - \boldsymbol{y}^2)^2 \rangle \lesssim \psi + \sigma^2. \quad (3.30)$$

This implies (3.24) for $|\pi_1(\mathfrak{X}_1) - \pi_2(\mathfrak{X}_2)| > \iota_*(\pi_1(\mathfrak{X}_1) + \pi_2(\mathfrak{X}_2))$ as $\langle \boldsymbol{e}_- \rangle^2 \geq \iota_*^2 \sim 1$. This completes the proof of Lemma 3.9. □

Following the remaining arguments of chapter 8 and 9 in [1] yields Proposition 3.8.

4 Proofs of Theorem 2.3 and Theorem 2.6

Proof of Theorem 2.3. By Remark 3.2, we can apply Proposition 3.4 for each $\tilde{\delta} > 0$. Hence, there are $\boldsymbol{\rho}^0 \in \mathscr{B}$ and $\boldsymbol{\rho}^d \colon \mathfrak{X} \times \mathbb{R} \setminus \{0\} \to [0, \infty)$ such that

$$\boldsymbol{\rho}_x(\mathrm{d}\tau) = \boldsymbol{\rho}_x^0 \delta_0(\mathrm{d}\tau) + \boldsymbol{\rho}_x^d(\tau)\mathrm{d}\tau$$

for all $x \in \mathfrak{X}$. For $k \in \mathfrak{X}_1$, we set $\nu_k^0 := \boldsymbol{\rho}_k^0$ and

$$\nu_k^d(E) := E^{-1/2} \boldsymbol{\rho}_k^d(E^{1/2}) \chi(E > 0) \quad (4.1)$$

with $E \in \mathbb{R}$. Therefore, using (3.3), we obtain (2.6) (cf. the proof of Theorem 2.1 in [5]). The $1/3$-Hölder continuity of $\boldsymbol{\rho}^d$ implies the $1/3$-Hölder continuity of ν^d. Similarly, the analyticity of ν^d is obtained from the analyticity of $\boldsymbol{\rho}^d$. From Proposition 3.8 with $\tilde{\delta} = \sqrt{\delta}/2$, we conclude that $\mathfrak{P} \cap (\delta, \infty)$ is a finite union of open intervals and its connected components have a Lebesgue measure of at least $2\rho_*$ for some ρ_* depending only on the model parameters and δ. This completes the proof (i).

For the proof of (ii), we follow the proof of Theorem 2.6 in [3]. We replace the estimates (4.1), (4.2), (5.3) and (6.7) as well as their proofs in [3] by (3.10), (3.11), (3.16) and (3.24) as well as their proofs in this note, respectively. This proves a result corresponding to Theorem 2.6 in [3] for $\boldsymbol{\rho}^d$ and $\tau_0 \in (\partial \mathfrak{P}) \cap (0, \infty)$ in our setup. Using the transform (4.1) completes the proof of Theorem 2.3. □

Proof of Theorem 2.6. Note that **(B1)** implies **(A4)**. By Remark 3.2, **(A3)** implies **(C1)**. Using (3.22) to replace (8.34) in [1], we obtain an analogue of Proposition 10.1 in [1] in our setup on $\overline{\mathbb{H}}^{\Sigma}_{\tilde{\delta}}$. Therefore, we have proven in our setup analogues of all the ingredients provided in [1] and used in [2] to prove a local law for Wigner-type random matrices with a uniform primitive variance matrix. Thus, following the arguments in [2], we obtain a local law for the resolvent of $\boldsymbol{H}$ defined in (1.2) and spectral parameters $z \in \mathbb{H}^{\Sigma}_{\tilde{\delta}} \cap \{w \in \mathbb{H} \colon \mathrm{Im}\, w \geq (p+n)^{-1+\gamma}\}$, where $\tilde{\delta} = \sqrt{\delta}/2$ and $\gamma \in (0,1)$. Proceeding as in the proof of Theorem 2.2 in [5] yields Theorem 2.6. □

References

[1] O. Ajanki, L. Erdős, and T. Krüger, *Quadratic vector equations on complex upper half-plane*, arXiv:1506.05095v4, 2015.

[2] O. Ajanki, L. Erdős, and T. Krüger, *Universality for general Wigner-type matrices*, Prob. Theor. Rel. Fields **169** (2017), no. 3-4, 667–727. MR-3719056

[3] O. Ajanki, L. Erdős, and T. Krüger, *Singularities of solutions to quadratic vector equations on the complex upper half-plane*, Comm. Pure Appl. Math. **70** (2017), no. 9, 1672–1705. MR-3684307

[4] J. Alt, L. Erdős, T. Krüger, and Yu. Nemish, *Location of the spectrum of Kronecker random matrices*, arXiv:1706.08343, 2017.

[5] J. Alt, L. Erdős, and T. Krüger, *Local law for random Gram matrices*, Electron. J. Probab. **22** (2017), no. 25, 41 pp. MR-3622895

[6] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, Cambridge University Press, 2011. MR-2884783

[7] V.L. Girko, *Theory of stochastic canonical equations: Volumes i and ii*, Mathematics and Its Applications, Springer Netherlands, 2012. MR-1887675

[8] W. Hachem, P. Loubaton, and J. Najim, *Deterministic equivalents for certain functionals of large random matrices*, Ann. Appl. Probab. **17** (2007), no. 3, 875–930. MR-2326235

[9] V. A. Marchenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Mat. Sbornik **1** (1967), no. 4, 457–483.

[10] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*, Found. Trends Commun. Inf. Theory **1** (2004), no. 1, 1–182.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm