

Eigenvalue Exclusion for Sample Covariance Matrices with Dependent Entries

Alex Hassett

Faculty Mentor: Dr. Simona Diaconu

October 2, 2026

Abstract

Let

$$B_n = \frac{1}{n} T_n^{1/2} X_n X_n^* T_n^{1/2}, \quad p_n \asymp n,$$

where $X_n = (X_{ij}) \in \mathbb{R}^{p_n \times n}$ is a random matrix and $T_n \in \mathbb{R}^{p_n \times p_n}$ is deterministic and positive semidefinite, with $\sup_n \|T_n\|_{\text{op}} < \infty$. We assume that the sample columns are independent, while the coordinates within each column form a conditionally standardized martingale-difference sequence. Under a uniform 2626th coordinate-moment bound, we prove a conditional quadratic-form estimate at the covariance-weighted Hilbert–Schmidt scale. On compact intervals satisfying the stated deterministic support and stability conditions, we obtain, on an event of high probability, a 164th-moment bound for the difference between the random and deterministic Stieltjes transforms at imaginary parts down to $n^{-1/68}$. For each such interval, the probability that B_n has an eigenvalue in that interval is bounded by $Cn^{-379/272}$, yielding eventual almost-sure exclusion. We also treat the polynomial moment bound $\max_{i,j} \mathbb{E}|X_{ij}|^{8000} \leq Cn^5$, for which the corresponding probability is bounded by $Cn^{-387/272}$.

Keywords. Sample covariance matrices, dependent coordinates, martingale differences, quadratic forms, Stieltjes transforms, spectral exclusion.

Mathematics Subject Classification (2020). 60B20, 15B52.

1 Introduction

A covariance matrix records the variances of the coordinates of a random vector with finite second moments and the covariances between those coordinates. Its eigenvalues give the variances along orthogonal principal directions. In principal component analysis, the directions associated with the largest eigenvalues are used to summarize the dominant variation in the data [2]. Since the population covariance matrix is typically unknown, it is estimated from observations. Understanding how sampling affects the spectrum of the resulting sample covariance matrix is therefore important for covariance estimation and dimension reduction.

For each n , set $p = p_n$ and consider a random matrix $X_n = (x_1, \dots, x_n) \in \mathbb{R}^{p \times n}$, together with a deterministic positive semidefinite matrix $T_n \in \mathbb{R}^{p \times p}$. We study the eigenvalues of

$$B_n = \frac{1}{n} T_n^{1/2} X_n X_n^* T_n^{1/2}$$

when p and n grow at comparable rates. A basic problem in this regime is to identify a deterministic measure that approximates the empirical spectral distribution of B_n . Such convergence alone does

not rule out finitely many eigenvalues lying in a compact set disjoint from the deterministic support. The question addressed here is whether eigenvalues can occur in such a set.

For sample covariance matrices formed from data matrices with independent entries, Bai and Silverstein proved, under suitable hypotheses, that fixed compact subsets of the complement of the deterministic limiting support eventually contain no eigenvalues [1, 2]. Their argument combines leave-one-column-out resolvent identities, high-moment resolvent estimates at spectral parameters $E + i\eta_n$ with $\eta_n \downarrow 0$, and an argument using a nonnegative kernel. For dependent data, quadratic-form criteria and martingale-type assumptions yield global spectral limits in several settings [4, 6, 7], while blockwise dependence can also be compatible with Marchenko–Pastur laws [3]. These global results do not supply the quantitative moment estimates at vanishing imaginary parts used in the spectral-exclusion argument.

In the present model, the sample columns are independent and the coordinates within each column form a martingale-difference sequence with conditional variance one. Deleting a column therefore leaves a minor resolvent that is independent of the omitted column, while dependence among the coordinates of that column is handled by a martingale quadratic-form estimate. To state the conditional form of this estimate, write

$$r_j = n^{-1/2} T_n^{1/2} x_j, \quad \mathcal{G}_j = \sigma(x_k : k \neq j).$$

Under conditions (IC3)–(IC5) of Assumptions 2.1, Proposition 3.2 shows that every $\mathcal{G}_j$ -measurable matrix $A_j \in \mathbb{C}^{p \times p}$ satisfies

$$\mathbb{E} \left(\left| r_j^* A_j r_j - \frac{1}{n} \text{Tr}(T_n A_j) \right|^q \middle| \mathcal{G}_j \right) \leq C_q n^{-q} \{ \text{Tr}(T_n A_j T_n A_j^*) \}^{q/2}, \quad 2 \leq q \leq 1313,$$

where C_q is independent of n , j , and A_j . The Hilbert–Schmidt dependence on A_j is the probabilistic input needed for the leave-one-column-out resolvent estimates.

Theorem 2.2 is the principal result under uniformly bounded coordinate moments. Under Assumptions 2.1, which include the uniform bound on the 2626th absolute coordinate moments in (IC5) and the deterministic support and stability conditions in (IC6), it gives a 164th-moment comparison of the random and deterministic Stieltjes transforms down to imaginary part $n^{-1/68}$ on a high-probability event. For every compact interval $\mathcal{I}$ considered in that theorem,

$$\mathbb{P}(\text{spec}(B_n) \cap \mathcal{I} \neq \emptyset) \leq C n^{-379/272},$$

so $\mathcal{I}$ is almost surely free of eigenvalues for all sufficiently large n . Theorem 2.4 permits the polynomially growing moment bound

$$\max_{i,j} \mathbb{E} |X_{ij}|^{8000} \leq C n^5$$

and yields the summable estimate

$$\mathbb{P}(\text{spec}(B_n) \cap \mathcal{I} \neq \emptyset) \leq C n^{-387/272}.$$

The resulting factor $n^{1/800}$ in the L^q -norm estimate of Proposition 3.3 enters the estimates in the second proof. The moment orders and exponents used below are sufficient for these arguments and are not claimed to be optimal.

The proof retains the rank-one resolvent identities, finite-dimensional self-consistent equation, deterministic stability argument, and use of a nonnegative kernel from the Bai–Silverstein method [1, 2]. The probabilistic input for the present model is developed in Section 3. There, a martingale

quadratic-form inequality is applied conditionally on all columns except the omitted one. Bounds for the resulting minor resolvents then permit a martingale estimate for the fluctuation of the normalized trace about its expectation. This controls the random fluctuation but does not yet compare the expectation with the deterministic transform. For this, we express the error in the self-consistent equation as centered quadratic forms and remainder terms, estimate its mean, and use the trace bound to control the random error. The deterministic stability assumptions, together with a preliminary continuation argument that locates the relevant solution, convert this error estimate into the bound for $m_n - m_{c,n}$ in Theorem 4.5. Finally, we combine this bound with the estimate for a compact family of test functions in Lemma 5.1 to control the integral of a nonnegative kernel near $\mathcal{I}$. An eigenvalue in $\mathcal{I}$ would contribute more than this bound allows.

Section 2 gives the model, assumptions, and main results. Sections 3–5 prove Theorem 2.2 through the sequence just described. Section 6 proves Theorem 2.4 by adapting these estimates to polynomially growing moments. Section 7 summarizes the results and discusses possible extensions.

2 Model and main results

The two main results concern the same sample covariance construction, with different bounds on coordinate moments. Subsection 2.1 introduces the notation and states the result for uniformly bounded moments; Subsection 2.2 allows a moment bound that grows polynomially with n . We then explain the deterministic spectral conditions in Subsection 2.3 and the dependence permitted among coordinates in Subsection 2.4. Subsection 2.5 verifies the assumptions in an i.i.d. special case.

2.1 Uniformly bounded moments

We first define the sample matrices and their comparison transforms. The notation introduced here will also be used for polynomially growing moments in Subsection 2.2.

For each n , fix a positive integer p_n , set $y_n = p_n/n$, and write $[n] = \{1, 2, \dots, n\}$. Consider a random matrix

$$X_n = (x_1^{(n)}, \dots, x_n^{(n)}) \in \mathbb{R}^{p_n \times n}, \quad x_j^{(n)} = (X_{1j}^{(n)}, \dots, X_{p_n j}^{(n)})^T,$$

together with a deterministic, real symmetric, positive semidefinite matrix $T_n \in \mathbb{R}^{p_n \times p_n}$. Its empirical spectral distribution is

$$H_n = \frac{1}{p_n} \sum_{a=1}^{p_n} \delta_{\lambda_a(T_n)}.$$

For $j \in [n]$ and $z \in \mathbb{C}^+ = \{w \in \mathbb{C} : \Im w > 0\}$, write

$$r_{n,j} = n^{-1/2} T_n^{1/2} x_j^{(n)}, \quad B_n = \sum_{j=1}^n r_{n,j} r_{n,j}^*, \quad R_n(z) = (B_n - z I_{p_n})^{-1}.$$

For $S \subset [n]$, use the notation

$$B_n^{(S)} = B_n - \sum_{j \in S} r_{n,j} r_{n,j}^*, \quad R_{n,S}(z) = (B_n^{(S)} - z I_{p_n})^{-1}.$$

The spectral measures in the two normalizations are

$$\nu_{n,S} = \frac{1}{p_n} \sum_{\ell=1}^{p_n} \delta_{\lambda_\ell(B_n^{(S)})}, \quad \mu_n^{(S)} = \frac{1}{n} \sum_{\ell=1}^{p_n} \delta_{\lambda_\ell(B_n^{(S)})}, \quad \mu_n = \mu_n^{(\emptyset)}.$$

Thus $\nu_{n,S}$ is a probability measure and $\mu_n^{(S)} = y_n \nu_{n,S}$ has total mass y_n . The latter normalization gives the Stieltjes transform $n^{-1} \text{Tr } R_{n,S}(z)$ and assigns mass $1/n$ to each eigenvalue, counted with multiplicity.

On the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$, define

$$\mathcal{G}_{n,j} = \sigma(x_k^{(n)} : k \neq j), \quad \mathcal{H}_{n,k} = \sigma(x_1^{(n)}, \dots, x_k^{(n)}), \quad \mathbb{E}_{n,k} = \mathbb{E}(\cdot \mid \mathcal{H}_{n,k}),$$

for $j \in [n]$ and $0 \leq k \leq n$, with $\mathcal{H}_{n,0} = \{\emptyset, \Omega\}$. The coordinate filtration in column j is

$$\mathcal{F}_{n,j,i} = \sigma(X_{1j}^{(n)}, \dots, X_{ij}^{(n)}), \quad 1 \leq i \leq p_n,$$

with $\mathcal{F}_{n,j,0} = \{\emptyset, \Omega\}$.

The normalized traces are

$$s_n(z) = \frac{1}{p_n} \text{Tr } R_n(z), \quad s_{n,S}(z) = \frac{1}{p_n} \text{Tr } R_{n,S}(z).$$

The corresponding transforms are

$$m_n(z) = y_n s_n(z) = \frac{1}{n} \text{Tr } R_n(z), \quad u_n(z) = -\frac{1 - y_n}{z} + m_n(z), \quad u_{n,S}(z) = -\frac{1 - y_n}{z} + y_n s_{n,S}(z).$$

For $v \in \mathbb{C}$ satisfying $v \neq 0$ and $1 + tv \neq 0$ for every $t \in \text{supp } H_n$, define

$$f_n(v) = -\frac{1}{v} + y_n \int \frac{t dH_n(t)}{1 + tv}.$$

For $z \in \mathbb{C}^+$, the deterministic comparison transform $u_{c,n}(z)$ is the solution in $\mathbb{C}^+$ of the finite-dimensional sample-covariance equation [2, (6.1.5)–(6.1.6)]

$$f_n(u_{c,n}(z)) = z.$$

Here z is the spectral parameter, whereas v is the argument of the function f_n . The deterministic transforms in the two trace normalizations are

$$s_{c,n}(z) = \frac{1}{y_n} \left(u_{c,n}(z) + \frac{1 - y_n}{z} \right), \quad m_{c,n}(z) = y_n s_{c,n}(z) = u_{c,n}(z) + \frac{1 - y_n}{z}.$$

Thus $s_{c,n}$ uses the normalization $p_n^{-1} \text{Tr } R_n$, whereas $m_{c,n}$ uses $n^{-1} \text{Tr } R_n$. Denote by $\mu_{c,n}$ the deterministic measure of total mass y_n whose Stieltjes transform is $m_{c,n}$. Multiplying a measure by the positive number y_n does not change its support, so the support condition below is equivalent in either normalization. For $v \in \mathbb{C}$ for which the inverse exists, write

$$P_n(v) = (I_{p_n} + vT_n)^{-1}.$$

When n is fixed, we suppress its index in the notation; for example, $p = p_n$, $x_j = x_j^{(n)}$, $X_{ij} = X_{ij}^{(n)}$, and $u_c = u_{c,n}$. We also write $R_j = R_{\{j\}}$, $R_{jk} = R_{\{j,k\}}$, $s_j = s_{\{j\}}$, and $u_j = u_{\{j\}}$. We use $\|\cdot\|_{\text{op}}$ and $\|\cdot\|_F$ for the operator and Frobenius norms, respectively, $\|Y\|_q = (\mathbb{E}|Y|^q)^{1/q}$ for the L^q -norm of a random variable, and $\mathbf{1}_{\mathcal{E}}$ for the indicator of an event $\mathcal{E}$.

Assumptions 2.1 (Uniformly bounded moments).

(IC1) There are constants $0 < d_- < d_+ < \infty$ such that $d_- \leq p_n/n \leq d_+$.

(IC2) $\sup_n \|T_n\|_{\text{op}} < \infty$.

(IC3) The columns $x_1^{(n)}, \dots, x_n^{(n)}$ are mutually independent; they need not be identically distributed.

(IC4) For the filtration $\mathcal{F}_{n,j,i}$ (as defined in Subsection 2.1),

$$\mathbb{E}(X_{ij}^{(n)} \mid \mathcal{F}_{n,j,i-1}) = 0, \quad \mathbb{E}((X_{ij}^{(n)})^2 \mid \mathcal{F}_{n,j,i-1}) = 1 \quad \text{a.s.}$$

(IC5) For $Q = 1313$,

$$\sup_n \max_{1 \leq i \leq p_n, 1 \leq j \leq n} \mathbb{E}|X_{ij}^{(n)}|^{2Q} < \infty.$$

(IC6) There are compact intervals $\mathcal{I}, \mathcal{J}, \mathcal{U}$ satisfying

$$\mathcal{I} \subset \text{int } \mathcal{J}, \quad \mathcal{J} \subset \text{int } \mathcal{U}, \quad \mathcal{U} \subset (0, \infty),$$

and constants $L_0, c_{\text{st}} > 0$ such that $\mathcal{U} \cap \text{supp } \mu_{c,n} = \emptyset$ for all large n , and, uniformly for $E \in \mathcal{U}$ and $0 \leq \eta \leq L_0$,

$$|u_{c,n}(E + i\eta)| \geq c_{\text{st}}, \quad \inf_{t \in \text{supp } H_n} |1 + tu_{c,n}(E + i\eta)| \geq c_{\text{st}},$$

$$\left| \frac{1}{u_{c,n}(E + i\eta)^2} - y_n \int \frac{t^2 dH_n(t)}{(1 + tu_{c,n}(E + i\eta))^2} \right| \geq c_{\text{st}}.$$

For the uniformly bounded moment regime, set

$$q_0 = 164, \quad \eta_* = n^{-1/68}, \quad \beta_* = \frac{277}{272}, \quad \eta_\diamond = \eta_*^6 = n^{-3/34}.$$

The following theorem is our first main result. It gives a moment estimate for the difference of the random and deterministic Stieltjes transforms on an event of high probability, and a summable bound on the probability that B_n has an eigenvalue in $\mathcal{I}$. The deterministic conditions that make this comparison possible are explained in Subsection 2.3.

Theorem 2.2 (Uniformly bounded moments). *Under Assumptions 2.1, there are events $\mathfrak{A}_n$ such that*

$$\mathbb{P}(\mathfrak{A}_n^c) \leq Cn^{-243/5}$$

and, uniformly for $E \in \mathcal{I}$ and $n^{-1/68} \leq \eta \leq 1$,

$$\mathbb{E} [|m_n(E + i\eta) - m_{c,n}(E + i\eta)|^{164} \mathbf{1}_{\mathfrak{A}_n}] \leq Cn^{-164}.$$

Moreover,

$$\mathbb{P}(\text{spec}(B_n) \cap \mathcal{I} \neq \emptyset) \leq Cn^{-379/272}, \quad (2.1)$$

and therefore

$$\text{spec}(B_n) \cap \mathcal{I} = \emptyset \quad \text{for all sufficiently large } n, \text{ almost surely.} \quad (2.2)$$

2.2 Polynomially growing moments

We next allow the bound on the coordinate moments to depend on n . The assumptions below use moment order 8000 and permit a bound of order n^5 . The resulting concentration estimate contains an additional power of n , whose effect on the estimate for $m_n - m_{c,n}$ and the resulting probability bound is quantified in the second main theorem.

Assumptions 2.3 (Polynomial moment growth).

(IG1) There are constants $0 < d_- < d_+ < \infty$ such that $d_- \leq p_n/n \leq d_+$.

(IG2) The matrices T_n are deterministic, real symmetric, and positive semidefinite, and $\sup_n \|T_n\|_{\text{op}} < \infty$.

(IG3) The columns $x_1^{(n)}, \dots, x_n^{(n)}$ are mutually independent; they need not be identically distributed.

(IG4) For the filtration $\mathcal{F}_{n,j,i}$ (as defined in Subsection 2.1),

$$\mathbb{E}(X_{ij}^{(n)} \mid \mathcal{F}_{n,j,i-1}) = 0, \quad \mathbb{E}((X_{ij}^{(n)})^2 \mid \mathcal{F}_{n,j,i-1}) = 1 \quad \text{a.s.}$$

(IG5) For $Q_g = 4000$,

$$M_{2Q_g, n} := \max_{1 \leq i \leq p_n, 1 \leq j \leq n} \mathbb{E}|X_{ij}^{(n)}|^{2Q_g} \leq Cn^5.$$

(IG6) There are intervals $\mathcal{I}, \mathcal{J}, \mathcal{U}$ with compact closures satisfying

$$\overline{\mathcal{I}} \subset \text{int } \mathcal{J}, \quad \overline{\mathcal{J}} \subset \text{int } \mathcal{U}, \quad \overline{\mathcal{U}} \subset (0, \infty),$$

and constants $L_0 > 1$ and $c_{\text{st}} > 0$ such that, for all large n , $\mathcal{U} \cap \text{supp } \mu_{c,n} = \emptyset$, and uniformly for $E \in \mathcal{U}$ and $0 \leq \eta \leq L_0$,

$$\begin{aligned} |u_{c,n}(E + i\eta)| &\geq c_{\text{st}}, \\ \inf_{t \in \text{supp } H_n} |1 + tu_{c,n}(E + i\eta)| &\geq c_{\text{st}}, \\ \left| \frac{1}{u_{c,n}(E + i\eta)^2} - y_n \int \frac{t^2 dH_n(t)}{(1 + tu_{c,n}(E + i\eta))^2} \right| &\geq c_{\text{st}}. \end{aligned}$$

Set

$$q_g = 200, \quad \delta = \frac{1}{800}, \quad \eta_* = n^{-1/68}, \quad \eta_\diamond = \eta_*^6 = n^{-3/34}, \quad \beta_* = \frac{277}{272}. \quad (2.3)$$

Under these assumptions, the next theorem gives a transform estimate that still yields a summable bound on the probability that B_n has an eigenvalue in $\mathcal{I}$, and hence eventual almost-sure exclusion. Its proof in Section 6 keeps track of the additional moment factor throughout the resolvent estimates.

Theorem 2.4 (Polynomially growing moments). *Under Assumptions 2.3, there are events $\mathfrak{A}_n^g$ such that*

$$\mathbb{P}((\mathfrak{A}_n^g)^c) \leq Cn^{-581}$$

and, uniformly for $E \in \mathcal{I}$ and $n^{-1/68} \leq \eta \leq 1$,

$$\mathbb{E} [|m_n(E + i\eta) - m_{c,n}(E + i\eta)|^{200} \mathbf{1}_{\mathfrak{A}_n^g}] \leq Cn^{-200+1/2}.$$

Moreover,

$$\mathbb{P}(\text{spec}(B_n) \cap \mathcal{I} \neq \emptyset) \leq Cn^{-387/272}, \quad (2.4)$$

and therefore

$$\text{spec}(B_n) \cap \mathcal{I} = \emptyset \quad \text{for all sufficiently large } n, \text{ almost surely.} \quad (2.5)$$

2.3 Deterministic spectral assumptions

Conditions (IC6) and (IG6) combine separation from the support of the comparison measure with bounds on the solution of the deterministic equation $f_n(u_{c,n}(z)) = z$ (recall its definition in Subsection 2.1). The interval $\mathcal{I}$ is the target for spectral exclusion; $\mathcal{J}$ provides a larger interval for the covering and cutoff arguments, and the analytic bounds hold on the still larger interval $\mathcal{U}$. Since $\mathcal{U}$ is disjoint from the support of $\mu_{c,n}$, the representation

$$m_{c,n}(z) = \int \frac{d\mu_{c,n}(\lambda)}{\lambda - z}$$

extends analytically across $\mathcal{U}$ and is real-valued on $\mathcal{U}$. The same holds for $u_{c,n}(z) = m_{c,n}(z) - (1 - y_n)/z$, because $\mathcal{U} \subset (0, \infty)$. At $\eta = 0$, the expressions in (IC6) and (IG6) are evaluated using these analytic continuations.

The first lower bound controls $u_{c,n}^{-1}$, which appears in the deterministic equation and the residual identities. The second controls the inverse of $I_{p_n} + u_{c,n}T_n$. Indeed, diagonalizing T_n gives

$$\|(I_{p_n} + u_{c,n}T_n)^{-1}\|_{\text{op}} = \max_{t \in \text{supp } H_n} \frac{1}{|1 + tu_{c,n}|} \leq c_{\text{st}}^{-1}.$$

These two conditions keep the reciprocal terms in f_n bounded. Differentiation with respect to its argument v gives

$$f'_n(v) = \frac{1}{v^2} - y_n \int \frac{t^2 dH_n(t)}{(1 + tv)^2}.$$

The third lower bound is therefore $|f'_n(u_{c,n}(z))| \geq c_{\text{st}}$; it controls the local inverse of the deterministic equation. All three bounds hold with one positive constant independent of n , E , and η on the specified parameter set. Nonvanishing for each fixed n alone would not give this uniform control. Including $\eta = 0$ makes the bounds applicable as the imaginary parts in the proof tend to zero.

The role of the derivative can also be seen from the exact identity

$$f_n(v) - z = (v - u_{c,n}(z)) \left[\frac{1}{vu_{c,n}(z)} - y_n \int \frac{t^2 dH_n(t)}{(1 + tv)(1 + tu_{c,n}(z))} \right].$$

At $v = u_{c,n}(z)$, the factor in brackets equals $f'_n(u_{c,n}(z))$. The uniform denominator and derivative bounds keep this factor bounded away from zero in a sufficiently small neighborhood of $u_{c,n}(z)$. On the event Ω_n constructed in Lemma 4.1, the continuation argument keeps $u_{n,S}(E + i\eta)$ (as defined in Subsection 2.1) in that neighborhood for $E \in \mathcal{J}$, $\eta_* \leq \eta \leq 1$, and $|S| \leq 2$. Theorem 4.5 then uses this factorization to convert the residual estimate into a bound on $m_n - m_{c,n} = u_n - u_{c,n}$. Lemma 6.1 and Theorem 6.6 use the same deterministic argument under (IG6).

2.4 Dependent coordinates

We now turn from the deterministic spectral conditions to the dependence structure of the sample vectors. Conditions (IC3)–(IC4) and (IG3)–(IG4) distinguish independence across sample columns from dependence among the coordinates of a single column. For $k < i$, the martingale-difference property gives

$$\mathbb{E}(X_{ij}X_{kj}) = \mathbb{E}[X_{kj}\mathbb{E}(X_{ij} \mid \mathcal{F}_{j,i-1})] = 0,$$

while conditional standardization gives $\mathbb{E}X_{ij}^2 = 1$. Hence

$$\mathbb{E}(x_j x_j^*) = I_p, \quad \mathbb{E}(r_j r_j^*) = \frac{1}{n} T_n.$$

The conditional distribution of X_{ij} beyond its first two moments may nevertheless depend on $X_{1j}, \dots, X_{i-1,j}$; in particular, the coordinates within a column need not be independent or identically distributed.

The two principal uses of independence across columns are conditional quadratic-form concentration and the trace martingale. Conditioning on $\mathcal{G}_j$ leaves the law of x_j unchanged and makes a leave-one-column-out resolvent deterministic relative to that conditional law. The same independence makes $(\mathbb{E}_k - \mathbb{E}_{k-1})Y = 0$ whenever Y is integrable and measurable with respect to all columns except the k th. No independence among the coordinates of one column is used in Theorem 3.1.

2.5 The i.i.d. case

The probabilistic conditions simplify for independent entries. We verify the deterministic conditions explicitly in the square case with identity population covariance.

Let the entries of X_n be i.i.d. copies of a real random variable ξ satisfying

$$\mathbb{E}\xi = 0, \quad \mathbb{E}\xi^2 = 1.$$

Independence gives (IC3)–(IC4) and (IG3)–(IG4).

For this calculation, take $p_n = n$ and $T_n = I_n$; the general aspect-ratio conditions (IC1) and (IG1) require only $p_n \asymp n$. Then $y_n = 1$, $H_n = \delta_1$, and the deterministic equation becomes

$$z = -\frac{1}{u_{c,n}} + \frac{1}{1 + u_{c,n}} = -\frac{1}{u_{c,n}(1 + u_{c,n})}. \quad (2.6)$$

The corresponding measure is the Marchenko–Pastur law with aspect ratio one, whose support is $[0, 4]$ [2]. Choose compact intervals $\mathcal{I}, \mathcal{J}, \mathcal{U}$ with

$$\mathcal{I} \subset \text{int } \mathcal{J}, \quad \mathcal{J} \subset \text{int } \mathcal{U}, \quad \mathcal{U} \subset (4, \infty) \quad (2.7)$$

and fix $L_0 > 1$. On

$$\mathcal{Z} = \{E + i\eta : E \in \mathcal{U}, 0 \leq \eta \leq L_0\},$$

the solution of (2.6) is the continuous branch

$$u_{c,n}(z) = \frac{-1 + \sqrt{1 - 4/z}}{2}$$

that tends to zero as $z \rightarrow \infty$. It does not take the values 0 or -1 on $\mathcal{Z}$. Moreover,

$$f'_n(u_{c,n}) = \frac{1}{u_{c,n}^2} - \frac{1}{(1 + u_{c,n})^2} = \frac{1 + 2u_{c,n}}{u_{c,n}^2(1 + u_{c,n})^2}. \quad (2.8)$$

If the right-hand side vanished, then $u_{c,n} = -1/2$, and (2.6) would give $z = 4$, which does not belong to $\mathcal{Z}$. Since this solution is independent of n , compactness of $\mathcal{Z}$ gives a constant $c_{\text{st}} > 0$ for all three bounds in (IC6) and (IG6). The choices $p_n = n$, $T_n = I_n$, and (2.7) consequently satisfy (IC1), (IC2), (IC6), (IG1), (IG2), and (IG6).

If $\mathbb{E}|\xi|^{2626} < \infty$, then (IC5) holds and Theorem 2.2 applies with these choices of p_n and T_n . If $\mathbb{E}|\xi|^{8000} < \infty$, then

$$M_{8000,n} = \mathbb{E}|\xi|^{8000} \leq Cn^5,$$

so (IG5) holds and Theorem 2.4 applies as well.

3 Quadratic-form concentration

This section establishes the probabilistic estimate used in the subsequent resolvent arguments. Theorem 3.1 treats a deterministic coefficient matrix and depends only on the martingale-difference and moment conditions within one column. Proposition 3.2 conditions on all other columns to obtain the form used in Sections 4 and 5, while Subsection 3.1 retains the moment factor needed in Section 6.

Theorem 3.1 (Quadratic forms of martingale differences). *Let $Q \geq 2$ be an integer and let $x = (X_1, \dots, X_p)^T$ be a real random vector. For $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$, with $\mathcal{F}_0 = \{\emptyset, \Omega\}$, suppose*

$$\mathbb{E}(X_i \mid \mathcal{F}_{i-1}) = 0, \quad \mathbb{E}(X_i^2 \mid \mathcal{F}_{i-1}) = 1 \quad a.s.,$$

and put $M_{2Q} = \max_i \mathbb{E}|X_i|^{2Q} < \infty$. Then, for every deterministic $A \in \mathbb{C}^{p \times p}$ and every $2 \leq q \leq Q$,

$$\mathbb{E} |x^T A x - \text{Tr } A|^q \leq C_{q,Q} M_{2Q}^{q/Q} \|A\|_F^q. \quad (3.1)$$

The constant $C_{q,Q}$ depends only on q and Q ; in particular, it is independent of p , A , and the distribution of x .

Proof. It is enough to prove the estimates for $q = Q$, because Lyapunov's inequality then gives the result for every $q \leq Q$. We first treat a real symmetric matrix A normalized by $\|A\|_F = 1$. Write

$$x^T A x - \text{Tr } A = D + O,$$

where

$$D = \sum_{i=1}^p a_{ii}(X_i^2 - 1), \quad O = 2 \sum_{i=1}^p X_i L_i, \quad L_i = \sum_{k < i} a_{ik} X_k.$$

For the diagonal term, the summands $d_i = a_{ii}(X_i^2 - 1)$ form a martingale-difference sequence. The Burkholder–Rosenthal inequality for real-valued martingales gives

$$\mathbb{E}|D|^Q \leq C_Q \left[\mathbb{E} \left(\sum_i \mathbb{E}_{i-1} |d_i|^2 \right)^{Q/2} + \sum_i \mathbb{E} |d_i|^Q \right], \quad (3.2)$$

where $\mathbb{E}_{i-1} = \mathbb{E}(\cdot \mid \mathcal{F}_{i-1})$; see Bai–Silverstein, Lemma 2.13 [2]. Since $M_{2Q} \geq (\mathbb{E} X_i^2)^Q = 1$ and $\sum_i a_{ii}^2 \leq 1$, the weighted convexity inequality

$$\left(\sum_i w_i y_i \right)^{Q/2} \leq \sum_i w_i y_i^{Q/2}, \quad w_i, y_i \geq 0, \quad \sum_i w_i \leq 1,$$

followed by conditional Jensen gives

$$\begin{aligned} \mathbb{E} \left(\sum_i a_{ii}^2 \mathbb{E}_{i-1} |X_i^2 - 1|^2 \right)^{Q/2} &\leq \sum_i a_{ii}^2 \mathbb{E} |X_i^2 - 1|^Q \\ &\leq C_Q M_{2Q}. \end{aligned}$$

Also, because $Q \geq 2$ and $\sum_i a_{ii}^2 \leq 1$,

$$\sum_i \mathbb{E} |d_i|^Q \leq C_Q M_{2Q} \sum_i |a_{ii}|^Q \leq C_Q M_{2Q}.$$

Equation (3.2) therefore gives

$$\mathbb{E}|D|^Q \leq C_Q M_{2Q}. \quad (3.3)$$

For the off-diagonal term, L_i is $\mathcal{F}_{i-1}$ -measurable and $X_i L_i$ is a martingale difference. Burkholder–Rosenthal gives

$$\mathbb{E}|O|^Q \leq C_Q \left[\mathbb{E} \left(\sum_i L_i^2 \right)^{Q/2} + \sum_i \mathbb{E}|X_i L_i|^Q \right]. \quad (3.4)$$

Let $L = (L_1, \dots, L_p)^T$. It is the martingale transform

$$L = \sum_{k=1}^p X_k v_k, \quad v_k = (a_{1k} \mathbf{1}_{\{k < 1\}}, \dots, a_{pk} \mathbf{1}_{\{k < p\}})^T.$$

The vectors v_k are deterministic, so $(X_k v_k)$ is an $\mathbb{R}^p$ -valued martingale-difference sequence with respect to $(\mathcal{F}_k)$. Conditional standardization gives

$$\mathbb{E}_{k-1} \|X_k v_k\|_2^2 = \|v_k\|_2^2.$$

Apply the Hilbert-space case of [5, Theorem 4.1] at moment order Q , using

$$\mathbb{E} \max_k \|X_k v_k\|_2^Q \leq \sum_k \mathbb{E}|X_k|^Q \|v_k\|_2^Q.$$

The theorem bounds the maximum of the norms of the partial sums, and therefore also $\|L\|_2$. These substitutions give

$$\begin{aligned} \mathbb{E}\|L\|_2^Q &\leq C_Q \left[\left(\sum_k \|v_k\|_2^2 \right)^{Q/2} + \sum_k \mathbb{E}|X_k|^Q \|v_k\|_2^Q \right] \\ &\leq C_Q \left(1 + M_{2Q}^{1/2} \sum_k \|v_k\|_2^Q \right) \leq C_Q M_{2Q}^{1/2} \leq C_Q M_{2Q}. \end{aligned} \quad (3.5)$$

Here $\sum_k \|v_k\|_2^2 \leq \|A\|_F^2 = 1$, and $\sum_k \|v_k\|_2^Q \leq 1$. Thus the first term on the right of (3.4) is at most $C_Q M_{2Q}$.

For the second term in (3.4), Cauchy–Schwarz gives

$$\sum_i \mathbb{E}|X_i L_i|^Q \leq \sum_i (\mathbb{E}|X_i|^{2Q})^{1/2} (\mathbb{E}|L_i|^{2Q})^{1/2} \leq M_{2Q}^{1/2} \sum_i (\mathbb{E}|L_i|^{2Q})^{1/2}.$$

Apply the Burkholder–Rosenthal inequality at order $2Q$ to $L_i = \sum_{k < i} a_{ik} X_k$. If $r_i^2 = \sum_{k < i} a_{ik}^2$, then

$$\mathbb{E}|L_i|^{2Q} \leq C_Q \left[r_i^{2Q} + M_{2Q} \sum_{k < i} |a_{ik}|^{2Q} \right].$$

Consequently,

$$\begin{aligned} \sum_i (\mathbb{E}|L_i|^{2Q})^{1/2} &\leq C_Q \left[\sum_i r_i^Q + M_{2Q}^{1/2} \sum_i \left(\sum_{k < i} |a_{ik}|^{2Q} \right)^{1/2} \right] \\ &\leq C_Q \left[\left(\sum_i r_i^2 \right)^{Q/2} + M_{2Q}^{1/2} \sum_i r_i^Q \right] \\ &\leq C_Q M_{2Q}^{1/2}. \end{aligned}$$

Here $\|(a_{ik})_{k < i}\|_{2Q} \leq \|(a_{ik})_{k < i}\|_2 = r_i$, $\sum_i r_i^2 \leq 1$, and $\sum_i r_i^Q \leq (\sum_i r_i^2)^{Q/2} \leq 1$. The second term in (3.4) is therefore at most $C_Q M_{2Q}$, and

$$\mathbb{E}|O|^Q \leq C_Q M_{2Q}. \quad (3.6)$$

Combining (3.3) and (3.6) proves (3.1) at $q = Q$ for real symmetric A with $\|A\|_F = 1$.

For a real matrix A , $x^T A x = x^T (A + A^T) x / 2$, and $\|(A + A^T)/2\|_F \leq \|A\|_F$. If A is complex, write $A = A_1 + iA_2$ with real A_1, A_2 , replace each A_h by $(A_h + A_h^T)/2$, and apply the real estimate to the resulting two real symmetric matrices. Their Frobenius norms are bounded by a universal multiple of $\|A\|_F$. Homogeneity removes the normalization, and Lyapunov's inequality completes the estimate for $2 \leq q \leq Q$. $\square$

Conditioning on all columns except the omitted one fixes the coefficient matrix in a leave-one-column-out quadratic form. The next proposition records the resulting estimate under Assumptions 2.1.

Proposition 3.2. *Fix $j \in [n]$. If $A_j \in \mathbb{C}^{p \times p}$ is $\mathcal{G}_j$ -measurable, then, for every $2 \leq q \leq Q$ (recall $Q = 1313$ from (IC5)),*

$$\mathbb{E} \left(\left| r_j^* A_j r_j - \frac{1}{n} \text{Tr}(T A_j) \right|^q \middle| \mathcal{G}_j \right) \leq C_q n^{-q} \{ \text{Tr}(T A_j T A_j^*) \}^{q/2}. \quad (3.7)$$

The constant C_q is uniform in j .

Proof. Condition on $\mathcal{G}_j$. The matrix A_j is then fixed, while column independence leaves the law of x_j unchanged. Apply Theorem 3.1 to

$$n^{-1} T^{1/2} A_j T^{1/2}.$$

The identities

$$\|T^{1/2} A_j T^{1/2}\|_F^2 = \text{Tr}(T A_j T A_j^*), \quad \text{Tr}(T^{1/2} A_j T^{1/2}) = \text{Tr}(T A_j)$$

give the assertion. $\square$

3.1 Polynomial moment growth

Under Assumptions 2.3, the conditioning argument in Proposition 3.2 retains the factor contributed by $M_{2Q_g, n}$. The second conclusion below is the form used for the resolvent coefficients in Section 6.

Proposition 3.3. *Fix $j \in [n]$. Let $A_j \in \mathbb{C}^{p \times p}$ be $\mathcal{G}_j$ -measurable. For every $2 \leq q \leq Q_g$ (recall $Q_g = 4000$ from (IG5)),*

$$\mathbb{E} \left(\left| r_j^* A_j r_j - \frac{1}{n} \text{Tr}(T A_j) \right|^q \middle| \mathcal{G}_j \right) \leq C_q n^{-q} M_{2Q_g, n}^{q/Q_g} \{ \text{Tr}(T A_j T A_j^*) \}^{q/2}. \quad (3.8)$$

Consequently, if $\mathcal{E}_j \in \mathcal{G}_j$ and $\text{Tr}(T A_j T A_j^*) \leq Cn$ on $\mathcal{E}_j$, then, with $\delta = 1/800$ as in (2.3),

$$\left\| \left(r_j^* A_j r_j - \frac{1}{n} \text{Tr}(T A_j) \right) \mathbf{1}_{\mathcal{E}_j} \right\|_q \leq C_q n^{-1/2+\delta}. \quad (3.9)$$

Proof. Condition on $\mathcal{G}_j$. By (IG3), the conditional law of x_j is its unconditional law, while A_j is fixed. Theorem 3.1, applied with $C = n^{-1}T^{1/2}A_jT^{1/2}$, gives

$$\mathbb{E} \left(|x_j^* C x_j - \text{Tr } C|^q \mid \mathcal{G}_j \right) \leq C_q M_{2Q_g, n}^{q/Q_g} \|C\|_F^q.$$

Since

$$\|C\|_F^2 = n^{-2} \text{Tr}(T A_j T A_j^*),$$

this is (3.8). Multiplying (3.8) by $\mathbf{1}_{\mathcal{E}_j}$, taking expectations, and using the stated Hilbert–Schmidt bound gives

$$\mathbb{E} \left[|r_j^* A_j r_j - n^{-1} \text{Tr}(T A_j)|^q \mathbf{1}_{\mathcal{E}_j} \right] \leq C_q n^{-q/2} M_{2Q_g, n}^{q/Q_g}.$$

Taking the q -th root yields

$$\|(r_j^* A_j r_j - n^{-1} \text{Tr}(T A_j)) \mathbf{1}_{\mathcal{E}_j}\|_q \leq C_q n^{-1/2} M_{2Q_g, n}^{1/Q_g}.$$

Finally,

$$M_{2Q_g, n}^{1/Q_g} \leq C n^{5/4000} = C n^{1/800},$$

which proves (3.9). $\square$

4 Resolvent estimates

Under Assumptions 2.1, we apply the conditional quadratic-form estimate to bound $m_n - m_{c,n}$, as needed for spectral exclusion. Subsection 4.1 first bounds the minor transforms, weighted traces, and reciprocals of the rank-one denominators on a common event. These bounds allow the martingale argument in Subsection 4.2 to control the fluctuation of the trace about its expectation. To identify the deterministic approximation, we estimate the residual in the self-consistent equation in Subsection 4.3. Subsection 4.4 then uses deterministic stability to obtain the comparison with $m_{c,n}$.

4.1 Estimates with at most three deleted columns

The Hilbert–Schmidt quantities on the right of Proposition 3.2 must be controlled when the coefficient matrix is a minor resolvent or one of its powers. The following lemma bounds these weighted traces and the reciprocals of the denominators in the rank-one resolvent identity on a common event, uniformly over the indicated column deletions and spectral parameters. We then use its probability estimate to construct the predictable indicators needed for the column martingale in § 4.1.1.

Recall $Q = 1313$ from (IC5) and $\eta_* = n^{-1/68}$, $\eta_\diamond = \eta_*^6$ from Subsection 2.1.

Lemma 4.1. *There is an event Ω_n such that*

$$\mathbb{P}(\Omega_n^c) \leq C n^{-D_{\text{reg}}}, \quad D_{\text{reg}} = \frac{Q}{5} - 13 = \frac{1248}{5}, \quad (4.1)$$

and on Ω_n , writing $z = E + i\eta$, uniformly for $E \in \mathcal{J}$, $\eta_* \leq \eta \leq 1$, and $|S| \leq 3$,

$$|s_S(z) - s_c(z)| \leq C n^{-23/85}, \quad (4.2)$$

$$\nu_S(\mathcal{J}) \leq C \eta_*^6. \quad (4.3)$$

Moreover, uniformly for $E \in \mathcal{I}$, $\eta_* \leq \eta \leq 1$, and $r = 1, 2, 3$,

$$\frac{1}{n} \operatorname{Tr}(TR_S^r T(R_S^*)^r) \leq C. \quad (4.4)$$

Under the same restrictions on E and η , for every $j \notin S$ with $|S| \leq 2$,

$$\left| 1 + \frac{1}{n} \operatorname{Tr}(TR_{S \cup \{j\}}) \right|^{-1} + |1 + r_j^* R_{S \cup \{j\}} r_j|^{-1} \leq C. \quad (4.5)$$

Proof. We first obtain simultaneous quadratic-form estimates on finite meshes. These estimates initiate a comparison of the minor transforms at a large imaginary part, which is then continued down to η_ϕ . The resulting comparison gives the spectral-mass and weighted-trace bounds, together with the reciprocal denominator bounds in the statement.

Choose deterministic meshes of $\mathcal{J}$ and $[\eta_\phi, L_0]$, each with at most Cn^5 points and mesh size at most n^{-5} . For $|S| \leq 2$, $j \notin S$, and a mesh point, apply Proposition 3.2 to the three $\mathcal{G}_j$ -measurable matrices

$$I_p, \quad R_{S \cup \{j\}}, \quad R_{S \cup \{j\}} P(u_{S \cup \{j\}}).$$

For each of these matrices A , set

$$Z_{j,S}(A) = \begin{cases} \frac{|x_j^* T^{1/2} A T^{1/2} x_j - \operatorname{Tr}(TA)|}{\|T^{1/2} A T^{1/2}\|_F}, & \|T^{1/2} A T^{1/2}\|_F > 0, \\ 0, & \|T^{1/2} A T^{1/2}\|_F = 0. \end{cases}$$

Since

$$r_j^* A r_j - n^{-1} \operatorname{Tr}(TA) = n^{-1} (x_j^* T^{1/2} A T^{1/2} x_j - \operatorname{Tr}(TA)),$$

Proposition 3.2, applied to A and conditioned on $\mathcal{G}_j$, gives

$$\mathbb{E} Z_{j,S}(A)^Q \leq C.$$

There are at most

$$Cn^2 \cdot n \cdot n^{10} = Cn^{13}$$

tests. Therefore the event on which all normalized tests are at most $n^{1/5}$ satisfies

$$\mathbb{P}(\Omega_n^c) \leq Cn^{13-Q/5} = Cn^{-1248/5}.$$

On the resulting event, Sherman–Morrison gives, for $j \notin S$,

$$R_S = R_{S \cup \{j\}} - \beta_{j,S} R_{S \cup \{j\}} r_j r_j^* R_{S \cup \{j\}}, \quad \beta_{j,S} = (1 + r_j^* R_{S \cup \{j\}} r_j)^{-1}. \quad (4.6)$$

The rank-one trace estimate is

$$|s_S - s_{S \cup \{j\}}| \leq \frac{1}{p\eta} \leq \frac{C}{n\eta}. \quad (4.7)$$

Also

$$\frac{1}{n} \sum_{j \notin S} \beta_{j,S} = -z u_S - \frac{|S|}{n}. \quad (4.8)$$

We first work at the mesh points on Ω_n . The normalized test with coefficient $R_{S \cup \{j\}}$ gives

$$\left| r_j^* R_{S \cup \{j\}} r_j - \frac{1}{n} \operatorname{Tr}(TR_{S \cup \{j\}}) \right| \leq Cn^{-3/10} \eta^{-1}. \quad (4.9)$$

Here we used $\|R_{S \cup \{j\}}\|_{\text{op}} \leq \eta^{-1}$. Deleting one column from the data matrix changes the sample covariance resolvent by a matrix of rank at most one, by (4.6), and of operator norm at most $2\eta^{-1}$. For this difference D ,

$$|\text{Tr}(TD)| \leq \|T\|_{\text{op}} \|D\|_{\text{op}} \leq 2\|T\|_{\text{op}} \eta^{-1}.$$

For distinct indices $j, k \notin S$, apply this bound to the two comparisons through $R_{S \cup \{j, k\}}$ to obtain

$$\left| \frac{1}{n} \text{Tr}(TR_{S \cup \{j\}}) - \frac{1}{n} \text{Tr}(TR_{S \cup \{k\}}) \right| \leq \frac{C}{n\eta}.$$

For $q_{j,S} = r_j^* R_{S \cup \{j\}} r_j$, the preceding two estimates imply

$$|q_{j,S} - q_{k,S}| \leq Cn^{-3/10}\eta^{-1} + Cn^{-1}\eta^{-1}.$$

Since

$$\beta_{j,S} - \beta_{k,S} = \beta_{j,S} \beta_{k,S} (q_{k,S} - q_{j,S})$$

and the resolvent bound gives $|\beta_{j,S}| \leq |z|/\eta$,

$$\max_{j,k \notin S} |\beta_{j,S} - \beta_{k,S}| \leq C\{n^{-3/10}\eta^{-3} + n^{-1}\eta^{-3}\}. \quad (4.10)$$

At $\eta = \eta_\circ$, the right-hand side equals $Cn^{-3/85} + Cn^{-25/34} = o(1)$. If $|u_S - u_c| \leq 1$, equation (4.8) bounds the mean of the $\beta_{j,S}$. Equation (4.10) bounds the difference between each $\beta_{j,S}$ and this mean; hence

$$\sup_{j \notin S} |\beta_{j,S}| \leq C \quad (4.11)$$

whenever $|u_S - u_c| \leq 1$.

For later use, define

$$\mathcal{K}(v, u_c) = \frac{1}{vu_c} - y \int \frac{t^2 dH_n(t)}{(1+tv)(1+tu_c)}.$$

Uniform continuity and (IC6) give a fixed $0 < \rho \leq 1$. For $z = E + i\eta$, where $E \in \mathcal{U}$ and $0 \leq \eta \leq L_0$, set

$$\mathcal{V}(z) = \{v \in \mathbb{C} : |v - u_c(z)| < \rho\}.$$

Uniformly over these values of z ,

$$v \in \overline{\mathcal{V}(z)} \implies \|P(v)\|_{\text{op}} \leq C, \quad |v| \geq c, \quad |\mathcal{K}(v, u_c)| \geq c. \quad (4.12)$$

We begin the continuation argument at a fixed value $\eta = L_0$. The stability bounds in (IC6) persist if L_0 is increased to another fixed value. To verify this, write $u_c = a + ib$ and take imaginary parts in its defining equation. Then

$$\eta = \frac{b}{|u_c|^2} - y \int \frac{t^2 b dH_n(t)}{|1 + tu_c|^2},$$

so $|u_c| \leq \eta^{-1}$. On any additional bounded interval of imaginary parts, the defining equation and $\|T\|_{\text{op}} \leq C$ bound $|u_c|$ away from zero; moreover

$$|1 + tu_c| \geq \eta|u_c|, \quad \left| \frac{1}{u_c^2} - y \int \frac{t^2 dH_n(t)}{(1 + tu_c)^2} \right| \geq \eta^2.$$

Thus we may and do assume that the L_0 used in the mesh above was chosen sufficiently large.

The radius ρ can be chosen together with L_0 . Fix a sufficiently large H , independently of n . For $\eta \geq H$, the defining equation gives $(2\eta)^{-1} \leq |u_c| \leq \eta^{-1}$. If $|v - u_c| \leq (8\eta)^{-1}$, then

$$\|P(v)\|_{\text{op}} \leq 2, \quad |\mathcal{K}(v, u_c)| \geq \frac{8\eta^2}{9} - 4y\|T\|_{\text{op}}^2 \geq \frac{4\eta^2}{9}.$$

Indeed, $|v| \leq 9/(8\eta)$, and both $|1 + tv|$ and $|1 + tu_c|$ are at least $1/2$ for $t \in \text{spec}(T)$, after increasing H if necessary. Let $\rho_H > 0$ be a uniform radius on $0 \leq \eta \leq H$. For $L_0 \geq H$, the choice $\rho = \min\{\rho_H, (8L_0)^{-1}\}$ therefore gives (4.12) on the whole interval $[0, L_0]$.

The test with $A = I_p$ and $\|T\|_F \leq C\sqrt{p}$ gives $x_j^* T x_j \leq Cn$ on this event. Consequently, $p^{-1} \text{Tr } B^{(S)} \leq C$ for $|S| \leq 2$, and at $z = E + iL_0$,

$$|s_S(z) + z^{-1}| = \left| \int \frac{\lambda d\nu_S(\lambda)}{z(\lambda - z)} \right| \leq \frac{C}{L_0^2}.$$

Hence $|u_S(z) + z^{-1}| \leq CL_0^{-2}$. If $h_c = y \int t(1 + tu_c)^{-1} dH_n(t)$, then $u_c = -(z - h_c)^{-1}$ and, for sufficiently large L_0 , $|h_c| \leq C$ and $|u_c(z) + z^{-1}| \leq CL_0^{-2}$, with constants independent of $L_0 \geq H$. Since $CL_0^{-2} \leq \frac{1}{4} \min\{\rho_H, (8L_0)^{-1}\}$ for sufficiently large fixed L_0 , this gives

$$\sup_{E \in \mathcal{J}, |S| \leq 2} |u_S(E + iL_0) - u_c(E + iL_0)| \leq \rho/4. \quad (4.13)$$

Let $P_S = P(u_S)$. Multiplication of $(B^{(S)} - zI)R_S = I$ by P_S , followed by a normalized trace, gives the exact identity

$$w_S = -\frac{1}{nz} \sum_{j \notin S} \beta_{j,S} d_{j,S} + \frac{|S|}{nzp} \text{Tr}(P_S T R_S), \quad (4.14)$$

where

$$w_S = -\frac{1}{zp} \text{Tr } P_S - \frac{1}{p} \text{Tr } R_S, \quad (4.15)$$

$$d_{j,S} = \frac{1}{p} x_j^* T^{1/2} R_{S \cup \{j\}} P_S T^{1/2} x_j - \frac{1}{p} \text{Tr}(P_S T R_S). \quad (4.16)$$

Put $P_{j,S} = P(u_{S \cup \{j\}})$. Replacing P_S by the $\mathcal{G}_j$ -measurable $P_{j,S}$, using

$$P(u_S) - P(u_{S \cup \{j\}}) = (u_{S \cup \{j\}} - u_S)P(u_S)TP(u_{S \cup \{j\}}),$$

gives the exact decomposition

$$\begin{aligned} d_{j,S} &= \frac{1}{p} \left[x_j^* T^{1/2} R_{S \cup \{j\}} P_{j,S} T^{1/2} x_j - \text{Tr}(P_{j,S} T R_{S \cup \{j\}}) \right] \\ &\quad + \frac{1}{p} x_j^* T^{1/2} R_{S \cup \{j\}} (P_S - P_{j,S}) T^{1/2} x_j \\ &\quad + \frac{1}{p} \text{Tr}((P_{j,S} - P_S) T R_{S \cup \{j\}}) + \frac{1}{p} \text{Tr}(P_S T (R_{S \cup \{j\}} - R_S)). \end{aligned} \quad (4.17)$$

To estimate these terms at a mesh point, suppose $u_S \in \overline{\mathcal{V}(z)}$. Equation (4.12) bounds P_S . Also, $|u_{S \cup \{j\}} - u_S| \leq C(n\eta)^{-1} = o(1)$, uniformly for $\eta \geq \eta_\diamond$, and

$$P_{j,S} = [I_p + (u_{S \cup \{j\}} - u_S)P_S T]^{-1} P_S.$$

For all large n , the matrix perturbing I_p has norm at most $1/2$, so $\|P_{j,S}\|_{\text{op}} \leq C$. The three quadratic-form tests and the rank-one resolvent bounds now bound the four terms in (4.17), respectively, by

$$Cn^{-3/10}\eta^{-1}, \quad Cn^{-1}\eta^{-2}, \quad Cn^{-1}\eta^{-2}, \quad Cn^{-1}\eta^{-1}.$$

The definitions of w_S and u_S give the identity

$$zyw_S = u_S \mathcal{D}_S(z),$$

where

$$\mathcal{D}_S(z) = -z - \frac{1}{u_S} + y \int \frac{t dH_n(t)}{1 + tu_S}.$$

Whenever $u_S \in \mathcal{V}(z)$, both u_S and z are bounded above and away from zero. Combining this identity with (4.14) and the preceding estimates gives

$$|\mathcal{D}_S(z)| \leq C\delta_{\text{reg}}(\eta), \quad \delta_{\text{reg}}(\eta) = n^{-3/10}\eta^{-2} + n^{-1}\eta^{-2}. \quad (4.18)$$

The residual admits the factorization

$$\mathcal{D}_S(z) = (u_S - u_c)\mathcal{K}(u_S, u_c). \quad (4.19)$$

The factor $\mathcal{K}(u_S, u_c)$ is bounded away from zero whenever $|u_S - u_c| \leq \rho$, by (4.12). At $\eta = \eta_\diamond$,

$$\frac{\delta_{\text{reg}}(\eta_\diamond)}{\eta_\diamond} \leq Cn^{-3/10}\eta_\diamond^{-3} + Cn^{-1}\eta_\diamond^{-3} = o(1).$$

Order the mesh in η from L_0 down to $\eta_\diamond$. Equation (4.13) places every u_S , $|S| \leq 2$, in the $\rho/4$ -neighborhood at $\eta = L_0$. For adjacent mesh points z, z' , resolvent continuity gives

$$\|R_S(z) - R_S(z')\|_{\text{op}} \leq |z - z'| \eta_\diamond^{-2},$$

and the same estimate holds for the deterministic transform. Thus

$$|u_S(z) - u_S(z')| + |u_c(z) - u_c(z')| \leq Cn^{-5}\eta_\diamond^{-2} = o(1).$$

Suppose that z is the first mesh point at which $|u_S(z) - u_c(z)| \geq \rho/2$ for some S with $|S| \leq 2$. The estimates at the preceding mesh point and the continuity bound imply $|u_S(z) - u_c(z)| < \rho$ for all large n . Equations (4.18) and (4.19) then give

$$|u_S - u_c| \leq C\delta_{\text{reg}}(\eta_\diamond) = C\{n^{-21/170} + n^{-14/17}\} = o(1),$$

contradicting the choice of that mesh point for all large n . Hence $|u_S - u_c| < \rho/2$ at every mesh point. Applying (4.18) and (4.19) there gives $|u_S - u_c| \leq C\delta_{\text{reg}}(\eta)$. For an arbitrary point $E + i\eta$ in the parameter rectangle, choose a nearest mesh point $E' + i\eta'$. Then $\eta' \asymp \eta$, and the continuity estimate yields

$$|u_S(E + i\eta) - u_c(E + i\eta)| \leq C\delta_{\text{reg}}(\eta) + Cn^{-5}\eta_\diamond^{-2}.$$

Here $n^{-5}\eta_\diamond^{-2} = n^{-82/17}$ is $o(\eta_\diamond)$ and $o(n^{-23/85})$. In particular, $|u_S - u_c| < \rho/2$ throughout the rectangle for all large n . At $\eta \geq \eta_*$,

$$n^{-3/10}\eta^{-2} \leq n^{-3/10+2/68} = n^{-23/85},$$

and $u_S - u_c = y(s_S - s_c)$. Since y is bounded away from zero, this proves (4.2) for $|S| \leq 2$. If $|S| = 3$, delete one index from S and apply (4.7); since $(n\eta_*)^{-1} = n^{-67/68} = o(n^{-23/85})$, the same estimate holds for matrices with three deleted columns.

At $\eta = \eta_\diamond = \eta_*^6$, support separation and the transform bound imply the following estimate for $|S| \leq 2$. The rank-one comparison extends it to $|S| = 3$, because $(n\eta_\diamond)^{-1} = n^{-31/34} = o(\eta_\diamond)$:

$$\sup_{E \in \mathcal{J}} \Im s_S(E + i\eta_\diamond) \leq C\eta_\diamond.$$

Cover $\mathcal{J}$ by at most $C\eta_\diamond^{-1}$ intervals of radius $\eta_\diamond$, centered at points $E_a \in \mathcal{J}$. If $|\lambda - E_a| \leq \eta_\diamond$, then

$$\frac{\eta_\diamond}{(\lambda - E_a)^2 + \eta_\diamond^2} \geq \frac{1}{2\eta_\diamond}.$$

The spectral representation of s_S therefore gives

$$\nu_S([E_a - \eta_\diamond, E_a + \eta_\diamond]) \leq 2\eta_\diamond \Im s_S(E_a + i\eta_\diamond) \leq C\eta_\diamond^2.$$

Summing over the covering intervals proves $\nu_S(\mathcal{J}) \leq C\eta_\diamond = C\eta_*^6$, which is (4.3). For $E \in \mathcal{I}$, the spectral theorem then yields, for $r = 1, 2, 3$,

$$\begin{aligned} \frac{1}{n} \operatorname{Tr}(T R_S^r T (R_S^*)^r) &\leq C \int \frac{d\nu_S(\lambda)}{((\lambda - E)^2 + \eta^2)^r} \\ &\leq C\eta_*^{6-2r} + C \leq C. \end{aligned}$$

This proves (4.4). To obtain (4.5), first work at a mesh point. The continuation just proved allows us to apply (4.11), so $|1 + q_{j,S}| \geq c$. Equation (4.9) then gives

$$\left| 1 + \frac{1}{n} \operatorname{Tr}(T R_{S \cup \{j\}}) \right| \geq |1 + q_{j,S}| - \left| q_{j,S} - \frac{1}{n} \operatorname{Tr}(T R_{S \cup \{j\}}) \right| \geq c - o(1).$$

Both denominators vary by at most $C|z - z'|\eta_\diamond^{-2}$ between neighboring parameters. For the quadratic-form denominator this uses $\|r_j\|_2^2 \leq C$, supplied by the test with $A = I_p$; for the trace denominator it follows from $p \asymp n$ and $\|T\|_{\text{op}} \leq C$. The mesh spacing therefore extends both positive lower bounds to all $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$, proving (4.5). $\square$

4.1.1 Predictable indicators

Conditioning on all columns except x_k fixes the coefficient matrix in the quadratic-form estimate. The martingale estimate below instead uses multipliers measurable with respect to $\mathcal{H}_{k-1} = \sigma(x_1, \dots, x_{k-1})$ (as defined in Subsection 2.1). We use the conditional probabilities of the events below to construct such multipliers.

For $z \in \mathcal{I} + i[\eta_*, 1]$, define

$$\begin{aligned} t_k(z) &= \frac{1}{n} \operatorname{Tr}(T R_k(z)), & \bar{\beta}_k(z) &= (1 + t_k(z))^{-1}, \\ \Phi_{r,k}(z) &= \frac{1}{n} \operatorname{Tr}(T R_k(z)^r T (R_k(z)^*)^r), & r &= 1, 2. \end{aligned}$$

Choose a constant K larger than the constants in Lemma 4.1, and let $\mathcal{C}_k \in \mathcal{G}_k$ be the event on which, uniformly over this parameter set,

$$|\bar{\beta}_k| + \Phi_{1,k} + \Phi_{2,k} + \|P(u_k)\|_{\text{op}} \leq K, \quad |u_k - u_c| \leq \frac{\rho}{2}. \quad (4.20)$$

Here ρ is the stability radius selected in (4.12). Every quantity in (4.20) is a function of the columns other than x_k , so $\mathcal{C}_k \in \mathcal{G}_k$. Lemma 4.1 gives

$$\Omega_n \subseteq \bigcap_{k=1}^n \mathcal{C}_k, \quad \mathbb{P}(\mathcal{C}_k^c) \leq Cn^{-1248/5}. \quad (4.21)$$

Set $D_0 = 200$ and define

$$B_k = \mathbf{1}_{\{\mathbb{E}(\mathbf{1}_{\mathcal{C}_k^c} | \mathcal{H}_{k-1}) \leq n^{-D_0}\}}, \quad \mathcal{B}_n = \bigcap_{k=1}^n \{B_k = 1\}. \quad (4.22)$$

Each B_k is $\mathcal{H}_{k-1}$ -measurable. Markov's inequality and (4.21) give

$$\begin{aligned} \mathbb{P}(\mathcal{B}_n^c) &\leq \sum_{k=1}^n \mathbb{P}(\mathbb{E}(\mathbf{1}_{\mathcal{C}_k^c} | \mathcal{H}_{k-1}) > n^{-200}) \\ &\leq n^{201} \max_k \mathbb{P}(\mathcal{C}_k^c) \leq Cn^{-243/5}. \end{aligned} \quad (4.23)$$

Define the event

$$\mathfrak{A}_n = \Omega_n \cap \mathcal{B}_n. \quad (4.24)$$

Then

$$\mathbb{P}(\mathfrak{A}_n^c) \leq Cn^{-243/5}. \quad (4.25)$$

The sequence (B_k) is predictable with respect to the column filtration, since B_k is $\mathcal{H}_{k-1}$ -measurable. In contrast, the event $\mathcal{C}_k$ may depend on future columns and need not belong to $\mathcal{H}_{k-1}$.

4.2 Trace fluctuation

On $\mathfrak{A}_n$, defined in (4.24), we now estimate $s - \mathbb{E}s$. The rank-one identity in § 4.2.1 gives a martingale decomposition with respect to the column filtration. § 4.2.2 estimates its terms using the predictable indicators B_k . The second- and higher-moment bounds are then combined in § 4.2.3 to prove Theorem 4.2.

4.2.1 Exact decomposition

The trace difference caused by deleting one column can be written as a quadratic form. Column independence then removes the terms that depend only on the remaining columns from the martingale differences.

For fixed $z = E + i\eta$, define

$$\beta_k = (1 + r_k^* R_k r_k)^{-1}, \quad \widehat{\gamma}_k = r_k^* R_k r_k - t_k, \quad (4.26)$$

$$a_k = \frac{1}{n} \text{Tr}(T R_k^2), \quad \alpha_k = r_k^* R_k^2 r_k - a_k. \quad (4.27)$$

Here $\mathbb{E}_k = \mathbb{E}(\cdot | \mathcal{H}_k)$, with $\mathcal{H}_k = \sigma(x_1, \dots, x_k)$ (recall the notation in Subsection 2.1). Sherman–Morrison gives

$$\text{Tr}(R - R_k) = -\beta_k r_k^* R_k^2 r_k. \quad (4.28)$$

If Y is integrable and $\mathcal{G}_k$ -measurable, column independence implies

$$(\mathbb{E}_k - \mathbb{E}_{k-1})Y = 0. \quad (4.29)$$

Since

$$\beta_k = \bar{\beta}_k - \bar{\beta}_k \beta_k \hat{\gamma}_k, \quad (4.30)$$

equations (4.28) and (4.29) yield the exact identity

$$s - \mathbb{E}s = -\frac{1}{p} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) Z_{1k} + \frac{1}{p} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) Z_{2k}, \quad (4.31)$$

where

$$Z_{1k} = \alpha_k \beta_k, \quad Z_{2k} = a_k \hat{\gamma}_k \bar{\beta}_k \beta_k. \quad (4.32)$$

4.2.2 Pointwise and conditional estimates

We estimate the terms in (4.31) separately on $\mathcal{C}_k$ and its complement. On $\mathcal{C}_k$, the weighted-trace bounds give the required conditional moments; on its complement, global resolvent bounds are combined with the small conditional probability in the definition of B_k .

Two global estimates will be used on the complement of $\mathcal{C}_k$. First, (4.7), with $S = \emptyset$ and $j = k$, and the identity (4.28) give

$$|\beta_k r_k^* R_k^2 r_k| \leq \eta^{-1}. \quad (4.33)$$

Moreover,

$$|a_k| \leq C\eta^{-2}, \quad |\beta_k| \leq \frac{|z|}{\eta}.$$

Hence

$$|Z_{1k}| \leq C\eta^{-3}. \quad (4.34)$$

Let $(v_\ell)_{\ell=1}^p$ be an orthonormal eigenbasis of $B^{(\{k\})}$, with corresponding eigenvalues λ_ℓ . Then

$$t_k = \frac{1}{n} \sum_{\ell=1}^p \frac{v_\ell^* T v_\ell}{\lambda_\ell - z}, \quad a_k = \frac{1}{n} \sum_{\ell=1}^p \frac{v_\ell^* T v_\ell}{(\lambda_\ell - z)^2}.$$

Since $v_\ell^* T v_\ell \geq 0$, it follows that

$$|\bar{\beta}_k| \Im t_k \leq 1, \quad |a_k| \leq \eta^{-1} \Im t_k.$$

Therefore $|a_k \bar{\beta}_k| \leq \eta^{-1}$. Since $\hat{\gamma}_k \bar{\beta}_k \beta_k = \bar{\beta}_k - \beta_k$,

$$|Z_{2k}| \leq |a_k \bar{\beta}_k| + |a_k \beta_k| \leq C\eta^{-3}. \quad (4.35)$$

On $\mathcal{C}_k$, Proposition 3.2 gives, for $2 \leq r \leq Q$,

$$\mathbb{E}(|\hat{\gamma}_k|^r \mid \mathcal{G}_k) \mathbf{1}_{\mathcal{C}_k} \leq C_r n^{-r/2}, \quad (4.36)$$

$$\mathbb{E}(|\alpha_k|^r \mid \mathcal{G}_k) \mathbf{1}_{\mathcal{C}_k} \leq C_r n^{-r/2}. \quad (4.37)$$

Indeed, the two Hilbert–Schmidt quantities in (3.7) are $n\Phi_{1,k}$ and $n\Phi_{2,k}$, respectively. On $\mathcal{C}_k$, $|\bar{\beta}_k| \leq K$, and

$$|a_k| \leq \frac{\sqrt{p}}{n} \left\{ \text{Tr}(T R_k^2 T (R_k^*)^2) \right\}^{1/2} = \sqrt{\frac{p}{n}} \Phi_{2,k}^{1/2} \leq C. \quad (4.38)$$

Define

$$\mathcal{S}_k = \{|\hat{\gamma}_k| \leq (2K)^{-1}\}. \quad (4.39)$$

On $\mathcal{C}_k \cap \mathcal{S}_k$, one has $|\bar{\beta}_k| \leq K$ and $|\bar{\beta}_k \hat{\gamma}_k| \leq 1/2$. Equation (4.30) therefore gives

$$|\beta_k| = \left| \frac{\bar{\beta}_k}{1 + \bar{\beta}_k \hat{\gamma}_k} \right| \leq 2K.$$

Consequently, on the same event,

$$|Z_{1k}| \leq C|\alpha_k|, \quad |Z_{2k}| \leq C|\hat{\gamma}_k|. \quad (4.40)$$

At order Q , (4.36) also gives

$$\mathbb{P}(|\hat{\gamma}_k| > (2K)^{-1} \mid \mathcal{G}_k) \mathbf{1}_{\mathcal{C}_k} \leq Cn^{-Q/2}. \quad (4.41)$$

Combining (4.22), (4.34)–(4.41), for $h = 1, 2$, gives the conditional second-moment estimate

$$B_k \mathbb{E}_{k-1} |Z_{hk}|^2 \leq Cn^{-1} + C\eta^{-6}(n^{-Q/2} + n^{-D_0}). \quad (4.42)$$

For the portions supported on $\mathcal{C}_k$, first condition on $\mathcal{G}_k$, where the matrices and the event $\mathcal{C}_k$ are fixed and (4.36)–(4.41) apply. Then condition on $\mathcal{H}_{k-1}$ and use the tower property. The portion supported on $\mathcal{C}_k^c$ is controlled directly by the definition of the $\mathcal{H}_{k-1}$ -measurable variable B_k . To see the three contributions explicitly, split

$$|Z_{hk}|^r = |Z_{hk}|^r \mathbf{1}_{\mathcal{C}_k \cap \mathcal{S}_k} + |Z_{hk}|^r \mathbf{1}_{\mathcal{C}_k \cap \mathcal{S}_k^c} + |Z_{hk}|^r \mathbf{1}_{\mathcal{C}_k^c}.$$

For $r = 2$, the conditional expectation of the first term is at most Cn^{-1} by (4.40), (4.36), and (4.37). The second is at most $C\eta^{-6}n^{-Q/2}$ by the global estimates and (4.41). After multiplication by B_k , the third is at most

$$C\eta^{-6} B_k \mathbb{E}(\mathbf{1}_{\mathcal{C}_k^c} \mid \mathcal{H}_{k-1}) \leq C\eta^{-6} n^{-D_0}.$$

This proves (4.42). The same splitting at order $q_0 = 164$ (recall the parameters in Subsection 2.1) gives

$$\mathbb{E}|B_k Z_{hk}|^{q_0} \leq Cn^{-q_0/2} + C\eta^{-3q_0}(n^{-Q/2} + n^{-D_0}). \quad (4.43)$$

The constants in (4.42) and (4.43) are uniform in k .

4.2.3 High-moment trace estimate

The preceding bounds control the conditional second moments and the q_0 -th moments needed in the martingale inequality. Applying that inequality to (4.31), with the predictable multipliers B_k , gives the following trace estimate.

Theorem 4.2. *Uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$\mathbb{E}[|s(E + i\eta) - \mathbb{E}s(E + i\eta)|^{q_0} \mathbf{1}_{\mathfrak{A}_n}] \leq Cn^{-q_0}. \quad (4.44)$$

Proof. Because B_k is $\mathcal{H}_{k-1}$ -measurable,

$$(\mathbb{E}_k - \mathbb{E}_{k-1})(B_k Z_{hk}) = B_k(\mathbb{E}_k - \mathbb{E}_{k-1})Z_{hk}.$$

On $\mathcal{B}_n$, every B_k equals one. Thus

$$|s - \mathbb{E}s| \mathbf{1}_{\mathcal{B}_n} \leq \frac{1}{p} \sum_{h=1}^2 \left| \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1})(B_k Z_{hk}) \right|. \quad (4.45)$$

Lemma 2.13 of Bai–Silverstein [2] is the complex Burkholder–Rosenthal inequality: for a complex martingale-difference sequence (D_k) and $q \geq 2$,

$$\mathbb{E} \left| \sum_k D_k \right|^q \leq C_q \left[\mathbb{E} \left(\sum_k \mathbb{E}_{k-1} |D_k|^2 \right)^{q/2} + \sum_k \mathbb{E} |D_k|^q \right]. \quad (4.46)$$

Put $Y_k = B_k Z_{hk}$ and $D_k = (\mathbb{E}_k - \mathbb{E}_{k-1})Y_k$. Conditional Jensen's inequality and the tower property give

$$\mathbb{E}_{k-1} |D_k|^2 \leq \mathbb{E}_{k-1} |Y_k|^2, \quad \mathbb{E} |D_k|^{q_0} \leq 2^{q_0} \mathbb{E} |Y_k|^{q_0}. \quad (4.47)$$

Indeed, the tower property gives $\mathbb{E}_{k-1} \mathbb{E}_k Y_k = \mathbb{E}_{k-1} Y_k$. Expanding the square and applying conditional Jensen's inequality therefore yields

$$\begin{aligned} \mathbb{E}_{k-1} |D_k|^2 &= \mathbb{E}_{k-1} |\mathbb{E}_k Y_k|^2 - |\mathbb{E}_{k-1} Y_k|^2 \\ &\leq \mathbb{E}_{k-1} |\mathbb{E}_k Y_k|^2 \\ &\leq \mathbb{E}_{k-1} \mathbb{E}_k |Y_k|^2 = \mathbb{E}_{k-1} |Y_k|^2. \end{aligned}$$

The second inequality in (4.47) follows by applying $|a - b|^{q_0} \leq 2^{q_0-1}(|a|^{q_0} + |b|^{q_0})$ and conditional Jensen to $\mathbb{E}_k Y_k$ and $\mathbb{E}_{k-1} Y_k$.

Since $\eta \geq n^{-1/68}$, one has $\eta^{-6} \leq n^{3/34}$. Using $Q = 1313$, equation (4.42) gives

$$\begin{aligned} \sum_{k=1}^n B_k \mathbb{E}_{k-1} |Z_{hk}|^2 &\leq C + Cn\eta^{-6}(n^{-Q/2} + n^{-200}) \\ &\leq C \left(1 + n^{1+3/34-1313/2} + n^{1+3/34-200} \right) \\ &= C \left(1 + n^{-11142/17} + n^{-6763/34} \right) \leq C'. \end{aligned}$$

Likewise, $q_0 = 164$ gives $n^{1-q_0/2} = n^{-81}$ and $\eta^{-3q_0} \leq n^{3q_0/68} = n^{123/17}$. Thus (4.43) yields

$$\begin{aligned} \sum_{k=1}^n \mathbb{E} |B_k Z_{hk}|^{q_0} &\leq Cn^{1-q_0/2} + Cn\eta^{-3q_0}(n^{-Q/2} + n^{-200}) \\ &\leq C \left(n^{-81} + n^{1+123/17-1313/2} + n^{1+123/17-200} \right) \\ &= C \left(n^{-81} + n^{-22041/34} + n^{-3260/17} \right) \leq C'. \end{aligned}$$

Every power of n in these final expressions is at most one for $n \geq 1$, so both bounds hold with a constant C' independent of n , E , and η . Equation (4.46) therefore gives, for $h = 1, 2$,

$$\mathbb{E} \left| \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1})(B_k Z_{hk}) \right|^{q_0} \leq C.$$

Since $p \asymp n$, (4.45) proves

$$\mathbb{E} [|s - \mathbb{E}s|^{q_0} \mathbf{1}_{\mathcal{B}_n}] \leq Cn^{-q_0}.$$

Finally, $\mathfrak{A}_n \subseteq \mathcal{B}_n$, which proves (4.44). $\square$

4.3 Approximate self-consistent equation

Theorem 4.2 bounds $s - \mathbb{E}s$ on $\mathfrak{A}_n$, but does not yet compare $\mathbb{E}s$ with s_c . We next rewrite the resolvent equation as an approximate self-consistent equation. Lemma 4.3 separates its centered leading term from lower-order remainders in § 4.3.1. § 4.3.2 uses this expansion to control the equation at the expected trace, then combines that control with (4.44) to prove Theorem 4.4.

Throughout this subsection, fix $z = E + i\eta$ with $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$.

4.3.1 Exact residual identity

The rank-one identities below isolate a quadratic form centered with respect to $\mathcal{G}_j$. The other terms are estimated using the minor bounds from Lemma 4.1.

Let

$$q_j = p^{-1/2}x_j, \quad P = P(u), \quad P_j = P(u_j),$$

and define

$$d_j = q_j^* T^{1/2} R_j P T^{1/2} q_j - \frac{1}{p} \text{Tr}(P T R), \quad (4.48)$$

$$w(z) = -\frac{1}{zp} \text{Tr} P - s(z). \quad (4.49)$$

The exact residual identity is

$$w(z) = -\frac{1}{n} \sum_{j=1}^n \frac{d_j}{z(1 + r_j^* R_j r_j)}. \quad (4.50)$$

Split $d_j = d_j^{(1)} + \dots + d_j^{(4)}$, where

$$\begin{aligned} d_j^{(1)} &= q_j^* T^{1/2} R_j (P - P_j) T^{1/2} q_j, \\ d_j^{(2)} &= q_j^* T^{1/2} R_j P_j T^{1/2} q_j - \frac{1}{p} \text{Tr}(P_j T R_j), \\ d_j^{(3)} &= \frac{1}{p} \text{Tr}(P_j T (R_j - R)), \\ d_j^{(4)} &= \frac{1}{p} \text{Tr}((P_j - P) T R). \end{aligned}$$

Also set

$$b_j = 1 + \frac{1}{n} \text{Tr}(T R_j), \quad \varepsilon_j = r_j^* R_j r_j - \frac{1}{n} \text{Tr}(T R_j).$$

Then

$$\frac{1}{b_j + \varepsilon_j} = \frac{1}{b_j} - \frac{\varepsilon_j}{b_j(b_j + \varepsilon_j)}. \quad (4.51)$$

Define the leading centered term by

$$Y_j = \frac{1}{z b_j} d_j^{(2)}. \quad (4.52)$$

The coefficient in $d_j^{(2)}$, as well as b_j , is $\mathcal{G}_j$ -measurable. Column independence and the covariance calculation in Subsection 2.4 give $\mathbb{E}(q_j q_j^* \mid \mathcal{G}_j) = p^{-1} I_p$. Consequently,

$$\mathbb{E}(Y_j \mid \mathcal{G}_j) = 0. \quad (4.53)$$

Multiplication by $\mathbf{1}_{\mathfrak{A}_n}$ need not preserve the conditional centering in (4.53), because $\mathfrak{A}_n$ need not be $\mathcal{G}_j$ -measurable. To estimate the expectation contributed by $\mathfrak{A}_n^c$, we replace P_j and b_j^{-1} by bounded $\mathcal{G}_j$ -measurable functions that agree with them on $\mathfrak{A}_n$. Choose a smooth map $\varrho_z : \mathbb{C} \rightarrow \mathbb{C}$ that equals the identity on $\{v : |v - u_c(z)| \leq \rho/2\}$ and whose image is contained in $\mathcal{V}(z)$ (as defined in the proof of Lemma 4.1). Set

$$\tilde{P}_j = (I_p + \varrho_z(u_j)T)^{-1},$$

so $\|\tilde{P}_j\|_{\text{op}} \leq C$ everywhere. Let $c_b > 0$ be a lower bound for $|b_j|$ on $\mathfrak{A}_n$. Choose $\chi_b \in C^\infty([0, \infty); [0, 1])$ with $\chi_b(t) = 0$ for $t \leq c_b/2$ and $\chi_b(t) = 1$ for $t \geq c_b$, and define the bounded smooth function

$$\mathfrak{I}(b) = \begin{cases} \chi_b(|b|)b^{-1}, & b \neq 0, \\ 0, & b = 0. \end{cases}$$

Thus $\mathfrak{I}(b) = b^{-1}$ whenever $|b| \geq c_b$. Define

$$\hat{Y}_j = \frac{\mathfrak{I}(b_j)}{z} \left[q_j^* T^{1/2} R_j \tilde{P}_j T^{1/2} q_j - \frac{1}{p} \text{Tr}(\tilde{P}_j T R_j) \right]. \quad (4.54)$$

The matrix $T^{1/2} R_j \tilde{P}_j T^{1/2}$, its normalized trace, and the factor $\mathfrak{I}(b_j)/z$ are $\mathcal{G}_j$ -measurable. The same conditional second-moment identity gives

$$\mathbb{E}(\hat{Y}_j \mid \mathcal{G}_j) = 0, \quad \mathbb{E}\hat{Y}_j = 0, \quad \|\hat{Y}_j\|_Q \leq C\eta^{-1}. \quad (4.55)$$

The last bound follows from Theorem 3.1 and $\|T^{1/2} R_j \tilde{P}_j T^{1/2}\|_F \leq C\sqrt{p}\eta^{-1}$. On $\mathfrak{A}_n$, one has $\varrho_z(u_j) = u_j$ and $\mathfrak{I}(b_j) = b_j^{-1}$, so $\hat{Y}_j = Y_j$.

Lemma 4.3. *Uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$w\mathbf{1}_{\mathfrak{A}_n} = -\frac{1}{n} \sum_{j=1}^n \hat{Y}_j \mathbf{1}_{\mathfrak{A}_n} + \mathcal{R}_n \mathbf{1}_{\mathfrak{A}_n}, \quad \|\mathcal{R}_n \mathbf{1}_{\mathfrak{A}_n}\|_{q_0} \leq Cn^{-1}. \quad (4.56)$$

Proof. On $\mathfrak{A}_n$, Lemma 4.1 bounds P, P_j, b_j^{-1} , and $(b_j + \varepsilon_j)^{-1}$. For the conditional quadratic-form estimates, first work on $\mathcal{C}_j \in \mathcal{G}_j$, defined in (4.20). Its bounds on $P_j, \Phi_{1,j}$, and $\Phi_{2,j}$ control the coefficient norms. Take expectations and then use $\mathfrak{A}_n \subseteq \mathcal{C}_j$ to restrict to $\mathfrak{A}_n$. Proposition 3.2 therefore gives, for $2 \leq r \leq 4q_0 < Q$,

$$\|d_j^{(2)} \mathbf{1}_{\mathfrak{A}_n}\|_r + \|\varepsilon_j \mathbf{1}_{\mathfrak{A}_n}\|_r \leq C_r n^{-1/2}. \quad (4.57)$$

Therefore Hölder's inequality and (4.51) give

$$\left\| \frac{1}{n} \sum_j \frac{d_j^{(2)} \varepsilon_j}{z b_j (b_j + \varepsilon_j)} \mathbf{1}_{\mathfrak{A}_n} \right\|_{q_0} \leq Cn^{-1}. \quad (4.58)$$

Recall that $\beta_j = (b_j + \varepsilon_j)^{-1}$, by (4.26). For the remaining terms, the exact identities are

$$\begin{aligned} u - u_j &= -\frac{y\beta_j}{n} q_j^* T^{1/2} R_j^2 T^{1/2} q_j, \\ P - P_j &= (u_j - u) P T P_j, \\ R_j - R &= \beta_j R_j r_j r_j^* R_j. \end{aligned}$$

Lemma 4.1 and Proposition 3.2 imply

$$\|(u - u_j)\mathbf{1}_{\mathfrak{A}_n}\|_r \leq C_r n^{-1}. \quad (4.59)$$

They also give

$$\|d_j^{(h)}\mathbf{1}_{\mathfrak{A}_n}\|_{q_0} \leq Cn^{-1}, \quad h = 1, 3, 4. \quad (4.60)$$

For the third term, put $K_j = T^{1/2}R_jT^{1/2}$ and $A_{3j} = K_jP_jK_j$. The rank-one resolvent identity gives

$$d_j^{(3)} = \frac{y\beta_j}{p}q_j^*A_{3j}q_j.$$

On $\mathfrak{A}_n$, the weighted Hilbert–Schmidt estimate (4.4) with $r = 1$ and $\|R_j\|_{\text{op}} \leq \eta^{-1}$ give

$$|\text{Tr } A_{3j}| \leq Cn, \quad \|A_{3j}\|_F \leq C\eta^{-1}\sqrt{n}.$$

These bounds also hold on $\mathcal{C}_j$. Apply Theorem 3.1 conditionally on $\mathcal{G}_j$ with coefficient $p^{-1}A_{3j}$. Since $q_j = p^{-1/2}x_j$, this gives

$$\|(q_j^*A_{3j}q_j - p^{-1}\text{Tr } A_{3j})\mathbf{1}_{\mathcal{C}_j}\|_{q_0} \leq Cn^{-1/2}\eta^{-1} \leq C.$$

Together with the normalized trace bound and $\mathfrak{A}_n \subseteq \mathcal{C}_j$, this proves $\|q_j^*A_{3j}q_j\mathbf{1}_{\mathfrak{A}_n}\|_{q_0} \leq C$, and hence the claim for $d_j^{(3)}$.

For $d_j^{(4)}$, the inverse identity gives

$$d_j^{(4)} = (u - u_j)\frac{1}{p}\text{Tr}(PTP_jTR).$$

On $\mathfrak{A}_n$, Cauchy–Schwarz and (4.4) with $S = \emptyset$ and $r = 1$ bound this normalized trace by C . Equation (4.59) therefore proves the bound for $d_j^{(4)}$.

Finally, put $\Delta_j = u - u_j$ and $A_{1j} = T^{1/2}R_jP_jTP_jT^{1/2}$. Substituting $P = P_j - \Delta_jPTP_j$ gives

$$d_j^{(1)} = -\Delta_jq_j^*A_{1j}q_j + e_j, \quad |e_j|\mathbf{1}_{\mathfrak{A}_n} \leq C\eta^{-1}|\Delta_j|^2q_j^*Tq_j\mathbf{1}_{\mathfrak{A}_n}.$$

On $\mathcal{C}_j$, $\|A_{1j}\|_F \leq C\sqrt{n}$ and $|\text{Tr } A_{1j}| \leq Cn$. The conditional application of Theorem 3.1, now with coefficient $p^{-1}A_{1j}$, gives $\|q_j^*A_{1j}q_j\mathbf{1}_{\mathcal{C}_j}\|_{2q_0} \leq C$. The same theorem with coefficient $p^{-1}T$ gives $\|q_j^*Tq_j\|_{2q_0} \leq C$. Hölder’s inequality and (4.59), at orders $2q_0$ and $4q_0$, yield

$$\begin{aligned} \|\Delta_jq_j^*A_{1j}q_j\mathbf{1}_{\mathfrak{A}_n}\|_{q_0} &\leq Cn^{-1}, \\ \|e_j\mathbf{1}_{\mathfrak{A}_n}\|_{q_0} &\leq C\eta^{-1}\|\Delta_j\mathbf{1}_{\mathfrak{A}_n}\|_{4q_0}^2\|q_j^*Tq_j\|_{2q_0} \leq Cn^{-2+1/68}. \end{aligned}$$

Since $4q_0 < Q$, all these moment orders are available. This proves the remaining bound in (4.60).

Insert (4.51) and $d_j = \sum_{h=1}^4 d_j^{(h)}$ into (4.50). Equations (4.58) and (4.60), followed by the triangle inequality for the normalized sum, prove (4.56). Indeed, $Y_j = \widehat{Y}_j$ on $\mathfrak{A}_n$, so the leading sum from the expansion is the one in (4.56). $\square$

4.3.2 Control of the deterministic residual

We use the centered expansion in Lemma 4.3 to estimate the self-consistent equation at $\mathbb{E}s$. Theorem 4.2, which bounds $s - \mathbb{E}s$ on $\mathfrak{A}_n$, then transfers this estimate to the random trace.

Write

$$\bar{s}(z) = \mathbb{E}s(z), \quad \bar{u}(z) = -\frac{1-y}{z} + y\bar{s}(z).$$

For $\zeta \in \mathbb{C}$ such that $I_p + [-(1-y)/z + y\zeta]T$ is invertible, define

$$F_z(\zeta) = -\frac{1}{zp} \operatorname{Tr} \left(I_p + \left[-\frac{1-y}{z} + y\zeta \right] T \right)^{-1} - \zeta. \quad (4.61)$$

The nonzero eigenvalues of $n^{-1}X^*TX$ and B , counted with multiplicity, coincide, so

$$u(z) = \frac{1}{n} \operatorname{Tr}(n^{-1}X^*TX - zI_n)^{-1}.$$

For $\Im z > 0$, it follows that $\Im u(z) > 0$ and $\Im \bar{u}(z) > 0$. Since T is positive semidefinite, both $s(z)$ and $\bar{s}(z)$ lie in the domain of F_z . In particular, $w(z) = F_z(s(z))$.

Lemma 4.1 and (4.25) imply

$$\begin{aligned} |\bar{s}(z) - s_c(z)| &\leq \mathbb{E}(|s - s_c| \mathbf{1}_{\mathfrak{A}_n}) + \mathbb{E}(|s - s_c| \mathbf{1}_{\mathfrak{A}_n^c}) \\ &\leq Cn^{-23/85} + C\eta^{-1}\mathbb{P}(\mathfrak{A}_n^c) \leq Cn^{-23/85}. \end{aligned} \quad (4.62)$$

For all large n , Lemma 4.1 places $u(z) \in \mathcal{V}(z)$ on $\mathfrak{A}_n$. Equation (4.62) also gives $\bar{u}(z) \in \mathcal{V}(z)$. If $\zeta_1, \zeta_2 \in \mathbb{C}$ satisfy

$$-\frac{1-y}{z} + y\zeta_h \in \mathcal{V}(z), \quad h = 1, 2,$$

then the resolvent identity for $P(v)$ gives

$$|F_z(\zeta_1) - F_z(\zeta_2)| \leq C|\zeta_1 - \zeta_2|. \quad (4.63)$$

Indeed,

$$P(v_1) - P(v_2) = (v_2 - v_1)P(v_1)TP(v_2),$$

the normalized trace of the product is bounded, and $|v_1 - v_2| = y|\zeta_1 - \zeta_2|$.

Taking expectations in (4.56) and using (4.55) gives

$$\begin{aligned} |\mathbb{E}(w \mathbf{1}_{\mathfrak{A}_n})| &\leq Cn^{-1} + \frac{1}{n} \sum_{j=1}^n |\mathbb{E}(\widehat{Y}_j \mathbf{1}_{\mathfrak{A}_n})| \\ &= Cn^{-1} + \frac{1}{n} \sum_{j=1}^n |\mathbb{E}(\widehat{Y}_j \mathbf{1}_{\mathfrak{A}_n^c})| \\ &\leq Cn^{-1} + C\eta^{-1}\mathbb{P}(\mathfrak{A}_n^c)^{1-1/Q} \leq Cn^{-1}. \end{aligned} \quad (4.64)$$

The equality follows from $\mathbb{E}\widehat{Y}_j = 0$.

Let $\bar{w}(z) = F_z(\bar{s}(z))$. By (4.63), Theorem 4.2, and (4.64),

$$\begin{aligned} |\bar{w}| \mathbb{P}(\mathfrak{A}_n) &= |\mathbb{E}(\bar{w} \mathbf{1}_{\mathfrak{A}_n})| \\ &\leq |\mathbb{E}(w \mathbf{1}_{\mathfrak{A}_n})| + \mathbb{E}(|\bar{w} - w| \mathbf{1}_{\mathfrak{A}_n}) \\ &\leq Cn^{-1} + C\mathbb{E}(|\bar{s} - s| \mathbf{1}_{\mathfrak{A}_n}) \leq Cn^{-1}. \end{aligned}$$

By (4.25), $\mathbb{P}(\mathfrak{A}_n) \geq 1/2$ for all large n , and hence

$$|\bar{w}(z)| \leq Cn^{-1}. \quad (4.65)$$

Theorem 4.4. *Uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$\mathbb{E} [|w(E + i\eta)|^{q_0} \mathbf{1}_{\mathfrak{A}_n}] \leq Cn^{-q_0}. \quad (4.66)$$

Proof. On $\mathfrak{A}_n$, equations (4.63) and (4.65) give

$$|w| \leq |F_z(s) - F_z(\bar{s})| + |\bar{w}| \leq C|s - \bar{s}| + Cn^{-1}.$$

Raise this inequality to the q_0 -th power, take expectations, and apply Theorem 4.2. $\square$

4.4 Comparison of Stieltjes transforms

The residual estimate (4.66) can now be converted into a comparison with the deterministic transform. The preliminary continuation in Lemma 4.1 places the random transform in the neighborhood where the deterministic equation is stable. We first recall the trace normalizations used in this comparison:

$$m_n(z) = \frac{1}{n} \text{Tr } R(z) = ys(z), \quad (4.67)$$

and

$$u(z) = -\frac{1-y}{z} + m_n(z). \quad (4.68)$$

The deterministic transforms satisfy

$$s_{c,n}(z) = s_c(z) = \frac{1}{y} \left(u_c(z) + \frac{1-y}{z} \right), \quad m_{c,n}(z) = m_c(z) = ys_{c,n}(z), \quad (4.69)$$

where

$$z = -\frac{1}{u_c(z)} + y \int \frac{t dH_n(t)}{1 + tu_c(z)}. \quad (4.70)$$

Below, w_n denotes the residual w defined in (4.49).

Theorem 4.5. *Under Assumptions 2.1, uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$\mathbb{E} [|m_n(E + i\eta) - m_{c,n}(E + i\eta)|^{q_0} \mathbf{1}_{\mathfrak{A}_n}] \leq Cn^{-q_0}. \quad (4.71)$$

Equivalently,

$$\|(m_n(E + i\eta) - m_{c,n}(E + i\eta)) \mathbf{1}_{\mathfrak{A}_n}\|_{q_0} \leq Cn^{-1}.$$

Proof. Fix $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$, and write $z = E + i\eta$. For brevity put

$$u = u(z), \quad u_c = u_c(z), \quad H = H_n.$$

Define the residual in the self-consistent equation by

$$\mathcal{D}_n(z) = -z - \frac{1}{u} + y \int \frac{t dH(t)}{1 + tu}. \quad (4.72)$$

We first derive its exact relation to w_n . Since

$$\frac{1}{p} \text{Tr } P(u) = \int \frac{dH(t)}{1 + tu},$$

equations (4.68) and (4.49) give

$$\begin{aligned} zyw_n &= -y \int \frac{dH(t)}{1+tu} - zys \\ &= -y \int \frac{dH(t)}{1+tu} - zu - (1-y). \end{aligned}$$

The elementary identity

$$\frac{1}{1+tu} = 1 - \frac{tu}{1+tu}$$

implies

$$y \int \frac{dH(t)}{1+tu} = y - yu \int \frac{t dH(t)}{1+tu}.$$

Substitution yields

$$\begin{aligned} zyw_n &= -y + yu \int \frac{t dH(t)}{1+tu} - zu - (1-y) \\ &= -1 - zu + yu \int \frac{t dH(t)}{1+tu} \\ &= u\mathcal{D}_n(z). \end{aligned}$$

Thus

$$\mathcal{D}_n(z) = \frac{zy}{u} w_n(z). \quad (4.73)$$

Equation (4.70) states that $\mathcal{D}_n$ vanishes when $u = u_c$. Subtracting this value from (4.72) gives

$$\mathcal{D}_n(z) = (u - u_c) \left[\frac{1}{uu_c} - y \int \frac{t^2 dH(t)}{(1+tu)(1+tu_c)} \right]. \quad (4.74)$$

Indeed,

$$-\frac{1}{u} + \frac{1}{u_c} = \frac{u - u_c}{uu_c},$$

and

$$\frac{t}{1+tu} - \frac{t}{1+tu_c} = -\frac{t^2(u - u_c)}{(1+tu)(1+tu_c)}.$$

On $\mathfrak{A}_n$, equations (4.2) and (4.69) give

$$|u - u_c| = y|s - s_{c,n}| \leq Cn^{-23/85}.$$

The preceding estimate places u in $\mathcal{V}(z)$ for all large n . The uniform bounds in (4.12) therefore give

$$|u| \geq c, \quad \left| \frac{1}{uu_c} - y \int \frac{t^2 dH(t)}{(1+tu)(1+tu_c)} \right| \geq c \quad \text{on } \mathfrak{A}_n \quad (4.75)$$

for a constant $c > 0$. Since z , y , and u^{-1} are uniformly bounded on the parameter set, equations (4.73)–(4.75) give

$$|u - u_c| \mathbf{1}_{\mathfrak{A}_n} \leq C|w_n| \mathbf{1}_{\mathfrak{A}_n}.$$

Finally,

$$m_n - m_{c,n} = u - u_c$$

by (4.68) and (4.69). Raising the last inequality to the q_0 -th power, taking expectations, and invoking (4.66) proves (4.71). $\square$

5 Spectral exclusion

Theorem 4.5 gives a pointwise transform estimate at positive imaginary parts. Lemma 5.1 first extends this control to a compact family of test functions that vanish on a neighborhood of $\mathcal{I}$. This bounds the contribution away from the target interval in the kernel argument of Subsection 5.1. There, estimates for $m_n - m_{c,n}$ at finitely many imaginary parts rule out an eigenvalue once this contribution has been controlled. Since the location of a possible eigenvalue is unknown, we also extend the pointwise estimates to a finite net in $\mathcal{I}$ and use resolvent continuity between net points. A union bound then completes the proof of Theorem 2.2.

Choose a continuously differentiable function $0 \leq \chi \leq 1$ satisfying

$$\chi = 0 \text{ on } \mathcal{J}, \quad \chi = 1 \text{ on } [0, \infty) \setminus \mathcal{U}. \quad (5.1)$$

Such a function exists because $\overline{\mathcal{J}}$ is a compact subset of $\text{int } \mathcal{U}$. Fix $\eta_0 \in (0, 1]$. For $u \in \mathcal{I}$ and $0 \leq \eta \leq \eta_0$, define

$$F_{\eta,u}(\lambda) = \begin{cases} 0, & \lambda \in \mathcal{J}, \\ \chi(\lambda) \prod_{k=1}^{34} \frac{1}{(u - \lambda)^2 + k\eta^2}, & \lambda \in [0, \infty) \setminus \mathcal{J}. \end{cases} \quad (5.2)$$

Let

$$\mathfrak{F}_{\text{out}} = \{F_{\eta,u} : u \in \mathcal{I}, 0 \leq \eta \leq \eta_0\}.$$

Lemma 5.1. *For every fixed $\delta > 0$,*

$$\mathbb{P} \left(\sup_{f \in \mathfrak{F}_{\text{out}}} \left| \int f d(\mu_n - \mu_{c,n}) \right| > \delta \right) \leq C_\delta \left(n^{-q_0} + n^{-243/5} \right). \quad (5.3)$$

Proof. Set

$$d_0 = \text{dist}(\overline{\mathcal{I}}, \mathcal{J}^c) > 0.$$

For $\lambda \notin \mathcal{J}$ and $u \in \mathcal{I}$, $|u - \lambda| \geq d_0$, so

$$|F_{\eta,u}(\lambda)| \leq d_0^{-68}.$$

The two formulas in (5.2) agree on the boundary of $\mathcal{J}$, and the functions vanish uniformly as $\lambda \rightarrow \infty$. Hence

$$(u, \eta) \mapsto F_{\eta,u}$$

is continuous from the compact set $\overline{\mathcal{I}} \times [0, \eta_0]$ into $C_0([0, \infty))$, equipped with the uniform norm. Therefore $\mathfrak{F}_{\text{out}}$ is totally bounded.

Assume $\mathcal{I} \neq \emptyset$, since otherwise the assertion is immediate. Fix $E_0 \in \mathcal{I}$ and $z_0 = E_0 + i/2$. The nonunital self-adjoint algebra generated by

$$\lambda \mapsto (\lambda - z_0)^{-1}$$

separates points of $[0, \infty)$ and vanishes nowhere. By the Stone–Weierstrass theorem, it is dense in $C_0([0, \infty))$. For integers $a, b \geq 1$, partial fractions give

$$\frac{1}{(\lambda - z_0)^a (\lambda - \bar{z}_0)^b} = \sum_{j=1}^a \frac{A_j}{(\lambda - z_0)^j} + \sum_{j=1}^b \frac{B_j}{(\lambda - \bar{z}_0)^j}.$$

Indeed, subtracting the right side after choosing the two principal parts leaves an entire rational function that tends to zero at infinity. Thus every element of the generated algebra is a finite linear combination of powers at z_0 and $\bar{z}_0$. Moreover,

$$(\lambda - z)^{-k} = \frac{1}{(k-1)!} \partial_z^{k-1} (\lambda - z)^{-1}.$$

The case $k = 1$ is already a resolvent. For $k \geq 2$, put $\phi_\lambda(t) = (\lambda - E_0 - it)^{-1}$ and $r = k - 1 \geq 1$. Write $\Delta_h \phi(t) = \phi(t + h) - \phi(t)$. Applying this operator r times gives the r th forward difference

$$\Delta_h^r \phi(t) = \sum_{\ell=0}^r (-1)^{r-\ell} \binom{r}{\ell} \phi(t + \ell h).$$

Then

$$\frac{\Delta_h^r \phi_\lambda(1/2)}{i^r r! h^r} \longrightarrow (\lambda - z_0)^{-k}$$

uniformly for $\lambda \geq 0$. Indeed, repeated application of the fundamental theorem of calculus gives

$$\Delta_h^r \phi_\lambda(1/2) = \int_{[0, h]^r} \phi_\lambda^{(r)}(1/2 + s_1 + \cdots + s_r) ds_1 \cdots ds_r.$$

Since

$$\phi_\lambda^{(r)}(t) = i^r r! (\lambda - E_0 - it)^{-r-1}, \quad |\phi_\lambda^{(r+1)}(t)| \leq (r+1)! 2^{r+2} \quad (t \geq 1/2),$$

comparison with $h^r \phi_\lambda^{(r)}(1/2)$ bounds the error by $r(r+1)2^{r+2}h$. If $0 < h \leq (4r)^{-1}$, all poles used by the difference have imaginary parts in $[1/2, 3/4]$. Conjugation gives the same assertion for powers at $\bar{z}_0$. Consequently, finite linear combinations of the resolvents

$$\lambda \longmapsto (\lambda - E_0 - i\eta_\ell)^{-1}, \quad \eta_\ell \in [1/2, 3/4],$$

and their complex conjugates are dense in $C_0([0, \infty))$.

For all sufficiently large n , $\eta_* \leq 1/2$, so the heights in these resolvents lie in the range of Theorem 4.5. For each fixed linear combination g of this form, that theorem and Minkowski's inequality imply

$$\mathbb{E} \left[\left| \int g d(\mu_n - \mu_{c,n}) \right|^{q_0} \mathbf{1}_{\mathfrak{A}_n} \right] \leq C_g n^{-q_0}. \quad (5.4)$$

The measures μ_n and $\mu_{c,n}$ both have total mass $y \leq d_+$. Thus $|\int g d(\mu_n - \mu_{c,n})| \leq 2d_+ \|g\|_\infty$, and increasing C_g extends (5.4) to the finitely many remaining values of n . By total boundedness and density, there are finitely many $g_1, \dots, g_K$, with K independent of n , such that every $f \in \mathfrak{F}_{\text{out}}$ satisfies

$$\|f - g_\ell\|_\infty \leq \frac{\delta}{4d_+}$$

for some ℓ . The contribution of $f - g_\ell$ to integration against $\mu_n - \mu_{c,n}$ is then at most $\delta/2$. Markov's inequality applied to (5.4), a finite union bound, and (4.25) prove (5.3). $\square$

5.1 Excluding eigenvalues with a nonnegative kernel

The preceding lemma controls integrals against $\mu_n - \mu_{c,n}$ after multiplication by the cutoff. Combined with the transform estimates, this gives an upper bound for the remaining, nonnegative integral against μ_n . If an eigenvalue lies in $\mathcal{I}$, we may center the kernel at that eigenvalue. The corresponding

atom of μ_n has mass at least $1/n$ and provides a lower bound. The next lemma shows that these bounds are incompatible at $\eta = \eta_*$ when the two comparison errors are sufficiently small.

For $\eta > 0$ and $u \in \mathcal{I}$, define

$$\mathcal{K}_{\eta,u}(\lambda) = \prod_{k=1}^{34} \frac{1}{(u - \lambda)^2 + k\eta^2}.$$

Lemma 5.2. *Let $\eta = \eta_*$. There are constants $c_D > 0$ and $\delta_D > 0$, independent of n , such that, for all large n , the following two conditions imply $\text{spec}(B_n) \cap \mathcal{I} = \emptyset$:*

$$\sup_{\substack{u \in \mathcal{I} \\ 1 \leq k \leq 34}} |m_n(u + i\sqrt{k}\eta) - m_{c,n}(u + i\sqrt{k}\eta)| \leq \frac{c_D}{n\eta}, \quad (5.5)$$

$$\sup_{f \in \mathfrak{F}_{\text{out}}} \left| \int f d(\mu_n - \mu_{c,n}) \right| \leq \delta_D. \quad (5.6)$$

Proof. For $x \geq 0$, $h > 0$, and an integer $M \geq 1$, the partial-fraction identity

$$\sum_{k=1}^M \frac{(-1)^{k-1} \binom{M-1}{k-1}}{x + kh} = \frac{(M-1)! h^{M-1}}{\prod_{k=1}^M (x + kh)} \quad (5.7)$$

follows by multiplying both sides by $\prod_{k=1}^M (x + kh)$ and evaluating the resulting polynomials at $x = -jh$, $1 \leq j \leq M$.

Write $\nu_n = \mu_n - \mu_{c,n}$. Equation (5.5) and $|\Im(m_n - m_{c,n})| \leq |m_n - m_{c,n}|$, followed by division by $\sqrt{k}\eta$, give

$$\left| \int \frac{d\nu_n(\lambda)}{(u - \lambda)^2 + k\eta^2} \right| \leq \frac{C c_D}{n\eta^2}.$$

Apply (5.7) with $M = 34$, $h = \eta^2$, and $x = (u - \lambda)^2$, then integrate. Since $n^{-1} = \eta^{68}$,

$$\sup_{u \in \mathcal{I}} \left| \int \mathcal{K}_{\eta,u} d\nu_n \right| \leq C_{34} c_D. \quad (5.8)$$

Assumption 2.1(IC6) and (5.1) imply $(1 - \chi)\mu_{c,n} = 0$. Therefore

$$\int \mathcal{K}_{\eta,u} d\nu_n = \int \chi \mathcal{K}_{\eta,u} d\nu_n + \int (1 - \chi) \mathcal{K}_{\eta,u} d\mu_n. \quad (5.9)$$

The last integral is nonnegative. For all sufficiently large n , $\eta = \eta_* \leq \eta_0$, so $\chi \mathcal{K}_{\eta,u} = F_{\eta,u} \in \mathfrak{F}_{\text{out}}$. Conditions (5.6) and (5.8) give

$$\sup_{u \in \mathcal{I}} \int (1 - \chi) \mathcal{K}_{\eta,u} d\mu_n \leq C_{34} c_D + \delta_D.$$

If B_n has an eigenvalue $\lambda_\ell \in \mathcal{I}$, choose $u = \lambda_\ell$. Since $\chi = 0$ on $\mathcal{I}$, the summand corresponding to λ_ℓ contributes

$$\frac{1/n}{\prod_{k=1}^{34} k\eta^2} = \frac{1}{34!}$$

to the preceding nonnegative integral. Choosing $C_{34} c_D + \delta_D < 1/34!$ produces a contradiction. $\square$

Proof of Theorem 2.2. The event and transform assertions follow from (4.25) and Theorem 4.5. It remains to prove the spectral-exclusion estimate.

If $\mathcal{I} = \emptyset$, there is nothing to prove. Assume $\mathcal{I} \neq \emptyset$, and let c_D and δ_D be supplied by Lemma 5.2. With $\beta_* = 277/272$ from Subsection 2.1, choose an $n^{-\beta_*}$ -net $\mathcal{S}_n$ of $\mathcal{I}$ with

$$|\mathcal{S}_n| \leq Cn^{\beta_*}. \quad (5.10)$$

The spectral representation of the Stieltjes transforms gives

$$|\partial_E m_n(E + i\eta)| + |\partial_E m_{c,n}(E + i\eta)| \leq C\eta^{-2}. \quad (5.11)$$

Indeed, both measures have mass at most d_+ , and $|\lambda - E - i\eta|^{-2} \leq \eta^{-2}$.

At $\eta = \eta_* = n^{-1/68}$, normalize the interpolation error by the threshold in Lemma 5.2:

$$\begin{aligned} \frac{Cn^{-\beta_*}\eta_*^{-2}}{(n\eta_*)^{-1}} &= Cn^{1-\beta_*}\eta_*^{-1} \\ &= Cn^{69/68-\beta_*} = Cn^{-1/272} \longrightarrow 0. \end{aligned} \quad (5.12)$$

Thus, for all large n ,

$$Cn^{-\beta_*}\eta_*^{-2} \leq \frac{c_D}{2n\eta_*}.$$

Also, $\sqrt{34}\eta_* \leq 1$ for all large n . Hence all 34 imaginary parts $\sqrt{k}\eta_*$ belong to the parameter range of Theorem 4.5.

Define

$$\mathcal{G}_n = \bigcap_{s \in \mathcal{S}_n} \bigcap_{k=1}^{34} \left\{ |m_n(s + i\sqrt{k}\eta_*) - m_{c,n}(s + i\sqrt{k}\eta_*)| \leq \frac{c_D}{2n\eta_*} \right\}.$$

Equations (5.11)–(5.12) imply that (5.5) holds on $\mathcal{G}_n$.

For each grid point and each imaginary part, Markov's inequality and Theorem 4.5 give

$$\begin{aligned} \mathbb{P} \left(|m_n(s + i\sqrt{k}\eta_*) - m_{c,n}(s + i\sqrt{k}\eta_*)| > \frac{c_D}{2n\eta_*}, \mathfrak{A}_n \right) \\ \leq \left(\frac{2n\eta_*}{c_D} \right)^{q_0} Cn^{-q_0} \leq Cn^{-q_0/68}. \end{aligned}$$

Using (5.10) and summing over the 34 imaginary parts,

$$\mathbb{P}(\mathcal{G}_n^c \cap \mathfrak{A}_n) \leq Cn^{\beta_*-q_0/68} = Cn^{-379/272}. \quad (5.13)$$

The arithmetic is

$$\frac{q_0}{68} = \frac{164}{68} = \frac{656}{272}, \quad \frac{277}{272} - \frac{656}{272} = -\frac{379}{272}.$$

Let

$$\mathcal{C}_n = \left\{ \sup_{f \in \mathfrak{F}_{\text{out}}} \left| \int f d(\mu_n - \mu_{c,n}) \right| \leq \delta_D \right\}.$$

On $\mathfrak{A}_n \cap \mathcal{G}_n \cap \mathcal{C}_n$, both hypotheses of Lemma 5.2 hold. Therefore

$$\{\text{spec}(B_n) \cap \mathcal{I} \neq \emptyset\} \subset \mathfrak{A}_n^c \cup (\mathcal{G}_n^c \cap \mathfrak{A}_n) \cup \mathcal{C}_n^c.$$

Equations (4.25), (5.13), and (5.3) imply

$$\begin{aligned}\mathbb{P}(\text{spec}(B_n) \cap \mathcal{I} \neq \emptyset) &\leq Cn^{-243/5} + Cn^{-379/272} + C(n^{-164} + n^{-243/5}) \\ &\leq Cn^{-379/272},\end{aligned}$$

because

$$\frac{243}{5} > \frac{379}{272}, \quad 164 > \frac{379}{272}.$$

This proves (2.1) for all sufficiently large n ; enlarging C absorbs the finitely many remaining values. Since

$$\frac{379}{272} = 1 + \frac{107}{272} > 1,$$

the series of the probabilities in (2.1) converges. The Borel–Cantelli lemma proves (2.2). $\square$

6 Proof under polynomial moment growth

We now prove Theorem 2.4 under Assumptions 2.3, with the parameters in (2.3). Proposition 3.3 replaces the conditional quadratic-form bound used in Sections 4–5. The additional factor $n^{1/800}$ in its L^q -norm estimate changes the moment estimates and the resulting probability bounds.

Subsection 6.1 first establishes the estimates for deleted columns and constructs the predictable indicators and the event used in the martingale argument. On this event, Subsection 6.2 bounds $s - \mathbb{E}s$. Subsection 6.3 then estimates the residual in the deterministic equation, including the comparison at $\mathbb{E}s$ needed to identify the deterministic approximation. These estimates yield a bound for $m_n - m_{c,n}$ and the exclusion probability in Subsection 6.4. Finally, Subsection 6.5 explains the choice of moment order in this argument.

6.1 Estimates for deleted columns

The next lemma supplies the same types of minor bounds as Lemma 4.1, with a probability estimate obtained from the growing-moment assumption. The deterministic identities and continuation argument apply under (IG6); the quadratic-form tail probabilities are recomputed below. In § 6.1.1, we then construct the predictable indicators and the event used in the trace estimate.

Lemma 6.1. *There is an event Ω_n^g , independent of E and η , such that*

$$\mathbb{P}((\Omega_n^g)^c) \leq Cn^{-782}. \quad (6.1)$$

On this event, writing $z = E + i\eta$, uniformly for $E \in \mathcal{J}$, $\eta_ \leq \eta \leq 1$, and $|S| \leq 3$,*

$$|s_S(z) - s_{c,n}(z)| \leq Cn^{-23/85}, \quad (6.2)$$

$$\mu_n^{(S)}(\mathcal{J}) \leq C\eta_*^6. \quad (6.3)$$

Moreover, uniformly for $E \in \mathcal{I}$, $\eta_ \leq \eta \leq 1$, and $r = 1, 2, 3$,*

$$\frac{1}{n} \text{Tr}(TR_S^r T(R_S^*)^r) \leq C. \quad (6.4)$$

Under the same restrictions on E and η , for every $j \notin S$ with $|S| \leq 2$,

$$\left| 1 + \frac{1}{n} \text{Tr}(TR_{S \cup \{j\}}) \right|^{-1} + |1 + r_j^* R_{S \cup \{j\}} r_j|^{-1} \leq C. \quad (6.5)$$

Proof. It suffices to consider all sufficiently large n ; for the finitely many remaining values, take $\Omega_n^g = \emptyset$ and increase C . Recall $\eta_\diamond = n^{-3/34}$ from (2.3) and L_0 from (IG6). Choose deterministic meshes of $\mathcal{J}$ and $[\eta_\diamond, L_0]$, each of cardinality at most Cn^5 and mesh size at most n^{-5} . For $|S| \leq 2$, $j \notin S$, and every mesh point, apply Proposition 3.3 to the three $\mathcal{G}_j$ -measurable matrices

$$I_p, \quad R_{S \cup \{j\}}, \quad R_{S \cup \{j\}} P(u_{S \cup \{j\}}).$$

For each of these matrices A , define

$$Z_{j,S}(A) = \begin{cases} \frac{|x_j^* T^{1/2} A T^{1/2} x_j - \text{Tr}(TA)|}{\|T^{1/2} A T^{1/2}\|_F}, & \|T^{1/2} A T^{1/2}\|_F > 0, \\ 0, & \|T^{1/2} A T^{1/2}\|_F = 0. \end{cases}$$

Applying Proposition 3.3 to A , using $r_j = n^{-1/2} T^{1/2} x_j$, and conditioning on $\mathcal{G}_j$ gives

$$\mathbb{E} Z_{j,S}(A)^{Q_g} \leq C M_{2Q_g, n} \leq C n^5.$$

The number of tests is at most

$$C n^2 \cdot n \cdot n^{10} = C n^{13}.$$

Markov's inequality at threshold $n^{1/5}$ and a union bound therefore give

$$\mathbb{P}\left(\max Z_{j,S}(A) > n^{1/5}\right) \leq C n^{13} n^5 n^{-Q_g/5} = C n^{18-800} = C n^{-782}.$$

Let Ω_n^g be the event that every normalized test is at most $n^{1/5}$. The remainder of the proof of Lemma 4.1 uses only this threshold, the mesh, the rank-one identities, and the deterministic stability assumption. At each mesh point on Ω_n^g , the bound $\|T^{1/2} R_S T^{1/2}\|_F \leq C \sqrt{n} \eta^{-1}$ gives

$$|r_j^* R_{S \cup \{j\}} r_j - n^{-1} \text{Tr}(T R_{S \cup \{j\}})| \leq C n^{-1} \sqrt{n} n^{1/5} \eta^{-1} = C n^{-3/10} \eta^{-1}.$$

This tends to zero uniformly for $\eta \geq \eta_\diamond = n^{-3/34}$. The estimates (4.10)–(4.11) for $\beta_{j,S}$, the bound at $\eta = L_0$ in (4.13), and the continuation argument therefore apply with (IG6) in place of (IC6). Continuing from L_0 to $\eta_\diamond$, and then applying the interval-covering and weighted-trace estimates in the proof of Lemma 4.1, gives (6.2)–(6.5). The bound on $\mathbb{P}((\Omega_n^g)^c)$ is (6.1). $\square$

6.1.1 Predictable indicators

As in § 4.1.1, we use the bounds in Lemma 6.1 to construct indicators measurable with respect to the column filtration. Recall the factor $\mathcal{K}(v, u_c)$ defined in the proof of Lemma 4.1:

$$\mathcal{K}(v, u_c) = \frac{1}{v u_c} - y \int \frac{t^2 dH_n(t)}{(1 + tv)(1 + t u_c)}.$$

The lower bounds in (IG6) and the bounds on y and $\|T\|_{\text{op}}$ give, for v sufficiently close to u_c , uniformly in n and the stated spectral parameters,

$$|\mathcal{K}(v, u_c) - \mathcal{K}(u_c, u_c)| \leq C |v - u_c|.$$

Indeed, subtracting the reciprocal factors bounds the left-hand side by

$$\frac{|v - u_c|}{|v| |u_c|^2} + y |v - u_c| \int \frac{t^3 dH_n(t)}{|1 + tv| |1 + t u_c|^2}.$$

For a sufficiently small fixed neighborhood, all denominators here are bounded away from zero. Since $\mathcal{K}(u_c, u_c) = f'_n(u_c)$, one can therefore choose a fixed $0 < \rho_g \leq 1$. For $z = E + i\eta$, where $E \in \mathcal{U}$ and $0 \leq \eta \leq L_0$, set

$$\mathcal{V}_g(z) = \{v \in \mathbb{C} : |v - u_c(z)| < \rho_g\}.$$

Uniformly over these values of z ,

$$v \in \overline{\mathcal{V}_g(z)} \implies \|P(v)\|_{\text{op}} \leq C, \quad |v| \geq c, \quad |\mathcal{K}(v, u_c)| \geq c.$$

Choose K_g large enough to bound the sum in (6.6) using Lemma 6.1 and the preceding bound for $P(v)$. For each k , let $\mathcal{C}_k^g \in \mathcal{G}_k$ be the event on which, uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,

$$\left|1 + \frac{1}{n} \text{Tr}(TR_k)\right|^{-1} + \sum_{r=1}^2 \frac{1}{n} \text{Tr}(TR_k^r T(R_k^*)^r) + \|P(u_k)\|_{\text{op}} \leq K_g, \quad |u_k - u_c| \leq \frac{\rho_g}{2}. \quad (6.6)$$

Every quantity in (6.6) is $\mathcal{G}_k$ -measurable, and Lemma 6.1 gives

$$\Omega_n^g \subseteq \bigcap_{k=1}^n \mathcal{C}_k^g, \quad \mathbb{P}((\mathcal{C}_k^g)^c) \leq Cn^{-782}.$$

Set $D_0 = 200$ and

$$B_k^g = \mathbf{1}_{\{\mathbb{E}(\mathbf{1}_{(\mathcal{C}_k^g)^c} | \mathcal{H}_{k-1}) \leq n^{-D_0}\}}, \quad \mathcal{B}_n^g = \bigcap_{k=1}^n \{B_k^g = 1\}. \quad (6.7)$$

Each B_k^g is $\mathcal{H}_{k-1}$ -measurable, and Markov's inequality gives

$$\begin{aligned} \mathbb{P}((\mathcal{B}_n^g)^c) &\leq \sum_{k=1}^n \mathbb{P}\left(\mathbb{E}(\mathbf{1}_{(\mathcal{C}_k^g)^c} | \mathcal{H}_{k-1}) > n^{-200}\right) \\ &\leq n^{201} \max_k \mathbb{P}((\mathcal{C}_k^g)^c) \leq Cn^{-581}. \end{aligned} \quad (6.8)$$

Define

$$\mathfrak{A}_n^g = \Omega_n^g \cap \mathcal{B}_n^g. \quad (6.9)$$

Equations (6.1) and (6.8) imply

$$\mathbb{P}((\mathfrak{A}_n^g)^c) \leq Cn^{-581}. \quad (6.10)$$

6.2 Trace fluctuation

On $\mathfrak{A}_n^g$, defined in (6.9), the minor bounds hold and every $B_k^g = 1$. Column independence gives the decomposition (4.31), now estimated under Assumptions 2.3. Lemma 6.2 gives the second-moment and higher-moment estimates. Theorem 6.3 then bounds the fluctuation $s - \mathbb{E}s$ on $\mathfrak{A}_n^g$ by applying the Burkholder–Rosenthal inequality.

For fixed $z = E + i\eta$, define

$$\begin{aligned} t_k &= \frac{1}{n} \text{Tr}(TR_k), & \bar{\beta}_k &= (1 + t_k)^{-1}, \\ \beta_k &= (1 + r_k^* R_k r_k)^{-1}, & \hat{\gamma}_k &= r_k^* R_k r_k - t_k, \\ a_k &= \frac{1}{n} \text{Tr}(TR_k^2), & \alpha_k &= r_k^* R_k^2 r_k - a_k. \end{aligned}$$

Sherman–Morrison and column independence give the exact decomposition

$$s - \mathbb{E}s = -\frac{1}{p} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) Z_{1k} + \frac{1}{p} \sum_{k=1}^n (\mathbb{E}_k - \mathbb{E}_{k-1}) Z_{2k}, \quad (6.11)$$

where

$$Z_{1k} = \alpha_k \beta_k, \quad Z_{2k} = a_k \widehat{\gamma}_k \bar{\beta}_k \beta_k.$$

Lemma 6.2. *Uniformly for $E \in \mathcal{I}$, $\eta_* \leq \eta \leq 1$, $1 \leq k \leq n$, and $h = 1, 2$,*

$$B_k^g \mathbb{E}_{k-1} |Z_{hk}|^2 \leq Cn^{-1+1/400} + C\eta^{-6}(n^{-1995} + n^{-200}), \quad (6.12)$$

$$\mathbb{E} |B_k^g Z_{hk}|^{200} \leq Cn^{-100+1/4} + C\eta^{-600}(n^{-1995} + n^{-200}). \quad (6.13)$$

Proof. On $\mathcal{C}_k^g$, the weighted Hilbert–Schmidt bounds and Proposition 3.3 give, for $2 \leq r \leq Q_g$,

$$\mathbb{E}(|\widehat{\gamma}_k|^r \mid \mathcal{G}_k) \mathbf{1}_{\mathcal{C}_k^g} + \mathbb{E}(|\alpha_k|^r \mid \mathcal{G}_k) \mathbf{1}_{\mathcal{C}_k^g} \leq C_r n^{-r/2+r/800}. \quad (6.14)$$

Indeed, the squared Hilbert–Schmidt quantities are bounded by Cn , and the moment factor in (3.8) is bounded by $C_r n^{r/800}$.

Let K be a deterministic bound, on $\mathcal{C}_k^g$, for $|\bar{\beta}_k|$, $|a_k|$, and the quantities in (6.6), and set $\mathcal{S}_k = \{|\widehat{\gamma}_k| \leq (2K)^{-1}\}$. On $\mathcal{C}_k^g \cap \mathcal{S}_k$, the identity

$$\beta_k = \bar{\beta}_k - \bar{\beta}_k \beta_k \widehat{\gamma}_k$$

implies $|\beta_k| \leq 2K$, and hence

$$|Z_{1k}| \leq C|\alpha_k|, \quad |Z_{2k}| \leq C|\widehat{\gamma}_k|.$$

At moment order $Q_g = 4000$, (6.14) gives

$$\mathbb{P}(\mathcal{S}_k^c \mid \mathcal{G}_k) \mathbf{1}_{\mathcal{C}_k^g} \leq Cn^{-Q_g/2+Q_g/800} = Cn^{-2000+5} = Cn^{-1995}. \quad (6.15)$$

The estimates (4.34) and (4.35) give $|Z_{hk}| \leq C\eta^{-3}$. They follow from the rank-one identities and deterministic resolvent bounds, so they also hold under Assumptions 2.3.

For the portions supported on $\mathcal{C}_k^g$, first condition on $\mathcal{G}_k$, apply (6.14) and (6.15), and then condition on $\mathcal{H}_{k-1}$. The tower property transfers these estimates to the column filtration. On $(\mathcal{C}_k^g)^c$, the definition of B_k^g gives

$$B_k^g \mathbb{E}(\mathbf{1}_{(\mathcal{C}_k^g)^c} \mid \mathcal{H}_{k-1}) \leq n^{-200}.$$

Split each conditional moment over $\mathcal{C}_k^g \cap \mathcal{S}_k$, $\mathcal{C}_k^g \cap \mathcal{S}_k^c$, and $(\mathcal{C}_k^g)^c$. For the first part, (6.14) at $r = 2$ gives $Cn^{-1+2/800} = Cn^{-1+1/400}$. The second part is bounded by $C\eta^{-6}n^{-1995}$ using (6.15). After multiplication by B_k^g , the third part is at most $C\eta^{-6}n^{-200}$ by (6.7). This proves (6.12).

At order 200, the first part is

$$Cn^{-200/2+200/800} = Cn^{-100+1/4}.$$

The portion on $\mathcal{C}_k^g \cap \mathcal{S}_k^c$ is at most $C\eta^{-600}n^{-1995}$, and the portion on $(\mathcal{C}_k^g)^c$, after multiplication by B_k^g , is at most $C\eta^{-600}n^{-200}$. This proves (6.13). $\square$

We now apply the moment estimates in Lemma 6.2 to the martingale decomposition (6.11).

Theorem 6.3. *Uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$\mathbb{E}[|s(E + i\eta) - \mathbb{E}s(E + i\eta)|^{200} \mathbf{1}_{\mathfrak{A}_n^g}] \leq Cn^{-200+1/4}. \quad (6.16)$$

Proof. On $\mathcal{B}_n^g$, every $B_k^g = 1$. Since B_k^g is $\mathcal{H}_{k-1}$ -measurable, (6.11) bounds $|s - \mathbb{E}s|$ by $p^{-1} \sum_{h=1}^2 |\sum_{k=1}^n D_{hk}|$, where

$$D_{hk} = (\mathbb{E}_k - \mathbb{E}_{k-1})(B_k^g Z_{hk}), \quad h = 1, 2.$$

The estimates in (4.47), with $B_k Z_{hk}$ replaced by $B_k^g Z_{hk}$, bound the conditional second moments and the unconditional 200th moments of D_{hk} by the corresponding quantities in Lemma 6.2.

At $\eta \geq \eta_* = n^{-1/68}$,

$$\sum_{k=1}^n \mathbb{E}_{k-1} |D_{hk}|^2 \leq Cn^{1/400} + Cn\eta^{-6}(n^{-1995} + n^{-200}) \leq Cn^{1/400},$$

because

$$1 + \frac{6}{68} - 200 < 0, \quad 1 + \frac{6}{68} - 1995 < 0.$$

Likewise,

$$\sum_{k=1}^n \mathbb{E} |D_{hk}|^{200} \leq Cn^{1-100+1/4} + Cn\eta^{-600}(n^{-1995} + n^{-200}) \leq C,$$

since

$$1 + \frac{600}{68} - 200 < 0, \quad 1 + \frac{600}{68} - 1995 < 0.$$

The complex Burkholder–Rosenthal inequality (4.46), applied to $(D_{hk})_{k=1}^n$ at order 200, therefore gives

$$\begin{aligned} \mathbb{E} \left| \sum_{k=1}^n D_{hk} \right|^{200} &\leq C \left[\mathbb{E} \left(\sum_{k=1}^n \mathbb{E}_{k-1} |D_{hk}|^2 \right)^{100} + \sum_{k=1}^n \mathbb{E} |D_{hk}|^{200} \right] \\ &\leq C(n^{1/400})^{100} + C \leq Cn^{1/4}. \end{aligned}$$

Multiplication by $p^{-200} \leq Cn^{-200}$, summation over $h = 1, 2$, and $\mathfrak{A}_n^g \subseteq \mathcal{B}_n^g$ prove (6.16). $\square$

6.3 Approximate self-consistent equation

Theorem 6.3 controls $s - \mathbb{E}s$ on $\mathfrak{A}_n^g$. To compare s with s_c , we next estimate the residual w_n in (4.49). § 6.3.1 separates a sum of conditionally centered quadratic forms from a remainder, whose bound is given in Lemma 6.4. § 6.3.2 uses this expansion to estimate the equation at $\mathbb{E}s$, then combines that estimate with Theorem 6.3 to obtain the moment bound for w_n .

Throughout this subsection, fix $z = E + i\eta$ with $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$.

6.3.1 Expansion of the residual

We use the residual identity from § 4.3.1, retaining the moment factor from Proposition 3.3 when estimating its terms. The product of two quadratic-form errors accounts for the factor $n^{1/400}$ in the remainder bound (6.21). We first recall the definitions needed for the expansion.

Put

$$q_j = p^{-1/2}x_j, \quad s_j = \frac{1}{p} \operatorname{Tr} R_j, \quad u_j = -\frac{1-y}{z} + ys_j,$$

and define

$$P = P(u), \quad P_j = P(u_j).$$

Set

$$\begin{aligned} d_j &= q_j^* T^{1/2} R_j P T^{1/2} q_j - \frac{1}{p} \operatorname{Tr}(P T R), \\ w_n(z) &= -\frac{1}{zp} \operatorname{Tr} P - s(z), \\ b_j &= 1 + \frac{1}{n} \operatorname{Tr}(T R_j), \\ \varepsilon_j &= r_j^* R_j r_j - \frac{1}{n} \operatorname{Tr}(T R_j). \end{aligned}$$

The exact residual identity is

$$w_n(z) = -\frac{1}{n} \sum_{j=1}^n \frac{d_j}{z(b_j + \varepsilon_j)}. \quad (6.17)$$

The exact decomposition $d_j = d_j^{(1)} + \dots + d_j^{(4)}$ is

$$\begin{aligned} d_j^{(1)} &= q_j^* T^{1/2} R_j (P - P_j) T^{1/2} q_j, \\ d_j^{(2)} &= q_j^* T^{1/2} R_j P_j T^{1/2} q_j - \frac{1}{p} \operatorname{Tr}(P_j T R_j), \\ d_j^{(3)} &= \frac{1}{p} \operatorname{Tr}(P_j T (R_j - R)), \\ d_j^{(4)} &= \frac{1}{p} \operatorname{Tr}((P_j - P) T R). \end{aligned}$$

To estimate expectations on $(\mathfrak{A}_n^g)^c$, we use bounded $\mathcal{G}_j$ -measurable replacements for P_j and b_j^{-1} that agree with them on $\mathfrak{A}_n^g$, as in § 4.3.1. Choose a smooth map $\varrho_z^g : \mathbb{C} \rightarrow \mathbb{C}$ that equals the identity on $\{v : |v - u_c(z)| \leq \rho_g/2\}$ and whose image is contained in $\mathcal{V}_g(z)$ (as defined in § 6.1.1). Let

$$\tilde{P}_j^g = (I_p + \varrho_z^g(u_j)T)^{-1}.$$

Let $c_b^g > 0$ be the lower bound for $|b_j|$ on $\mathfrak{A}_n^g$, choose $\chi_b^g \in C^\infty([0, \infty); [0, 1])$ with $\chi_b^g(t) = 0$ for $t \leq c_b^g/2$ and $\chi_b^g(t) = 1$ for $t \geq c_b^g$, and put

$$\mathfrak{J}_g(b) = \begin{cases} \chi_b^g(|b|)b^{-1}, & b \neq 0, \\ 0, & b = 0. \end{cases}$$

Define

$$\widehat{Y}_j = \frac{\mathfrak{J}_g(b_j)}{z} \left[q_j^* T^{1/2} R_j \tilde{P}_j^g T^{1/2} q_j - \frac{1}{p} \operatorname{Tr}(\tilde{P}_j^g T R_j) \right]. \quad (6.18)$$

The matrix $T^{1/2} R_j \tilde{P}_j^g T^{1/2}$, its normalized trace, and the factor $\mathfrak{J}_g(b_j)/z$ are $\mathcal{G}_j$ -measurable. By (IG3)–(IG4), $\mathbb{E}(q_j q_j^* | \mathcal{G}_j) = p^{-1} I_p$. Together with Proposition 3.3, this gives

$$\mathbb{E}(\widehat{Y}_j | \mathcal{G}_j) = 0, \quad \mathbb{E}\widehat{Y}_j = 0, \quad \|\widehat{Y}_j\|_{Q_g} \leq C\eta^{-1}n^\delta. \quad (6.19)$$

The norm estimate uses $\|T^{1/2}R_j\tilde{P}_j^gT^{1/2}\|_F \leq C\sqrt{p}\eta^{-1}$ and $M_{2Q_g,n}^{1/Q_g} \leq Cn^\delta$, where $\delta = 1/800$ as in (2.3). On $\mathfrak{A}_n^g$, $\varrho_z^g(u_j) = u_j$ and $\mathfrak{I}_g(b_j) = b_j^{-1}$, so $\hat{Y}_j = (zb_j)^{-1}d_j^{(2)}$.

Lemma 6.4. *Uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$w_n \mathbf{1}_{\mathfrak{A}_n^g} = -\frac{1}{n} \sum_{j=1}^n \hat{Y}_j \mathbf{1}_{\mathfrak{A}_n^g} + \mathcal{R}_n \mathbf{1}_{\mathfrak{A}_n^g}, \quad (6.20)$$

where

$$\|\mathcal{R}_n \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq Cn^{-1+1/400}. \quad (6.21)$$

Proof. On $\mathfrak{A}_n^g$, Lemma 6.1 bounds P, P_j, b_j^{-1} , and $(b_j + \varepsilon_j)^{-1}$. For the conditional estimates, first work on $\mathcal{C}_j^g \in \mathcal{G}_j$, defined in (6.6). The coefficient bounds hold on this event. Apply Proposition 3.3, take expectations, and then use $\mathfrak{A}_n^g \subseteq \mathcal{C}_j^g$. At order 400, this gives

$$\|d_j^{(2)} \mathbf{1}_{\mathfrak{A}_n^g}\|_{400} + \|\varepsilon_j \mathbf{1}_{\mathfrak{A}_n^g}\|_{400} \leq Cn^{-1/2+1/800}. \quad (6.22)$$

The exact denominator identity

$$\frac{1}{b_j + \varepsilon_j} = \frac{1}{b_j} - \frac{\varepsilon_j}{b_j(b_j + \varepsilon_j)}$$

therefore produces

$$\begin{aligned} & \left\| \frac{1}{n} \sum_{j=1}^n \frac{d_j^{(2)} \varepsilon_j}{zb_j(b_j + \varepsilon_j)} \mathbf{1}_{\mathfrak{A}_n^g} \right\|_{200} \\ & \leq \frac{C}{n} \sum_{j=1}^n \|d_j^{(2)} \mathbf{1}_{\mathfrak{A}_n^g}\|_{400} \|\varepsilon_j \mathbf{1}_{\mathfrak{A}_n^g}\|_{400} \leq Cn^{-1+2/800}. \end{aligned} \quad (6.23)$$

Thus the sum in (6.23) has L^{200} -norm at most $Cn^{-1+1/400}$, which is the bound required in (6.21).

With $\beta_j = (b_j + \varepsilon_j)^{-1}$, as in Subsection 6.2, the remaining exact identities are

$$u - u_j = -\frac{y\beta_j}{n} q_j^* T^{1/2} R_j^2 T^{1/2} q_j,$$

$$P - P_j = (u_j - u) P T P_j, \quad R_j - R = \beta_j R_j r_j r_j^* R_j.$$

The quadratic form containing R_j^2 in the first identity, together with (6.4) and Proposition 3.3 at order $800 < Q_g$, gives

$$\|(u - u_j) \mathbf{1}_{\mathfrak{A}_n^g}\|_{800} \leq Cn^{-1}.$$

For $d_j^{(3)}$, use the representation

$$d_j^{(3)} = \frac{y\beta_j}{p} q_j^* (K_j P_j K_j) q_j, \quad K_j = T^{1/2} R_j T^{1/2},$$

and put $A_{3j} = K_j P_j K_j$. The coefficient satisfies

$$|\operatorname{Tr} A_{3j}| \leq Cn, \quad \|A_{3j}\|_F \leq C\eta^{-1}\sqrt{n} \quad \text{on } \mathcal{C}_j^g.$$

These are the same coefficient estimates used in the proof of Lemma 4.3. The conditional application of Theorem 3.1 with coefficient $p^{-1}A_{3j}$, as in the proof of Proposition 3.3, gives

$$\left\| (q_j^* A_{3j} q_j - p^{-1} \operatorname{Tr} A_{3j}) \mathbf{1}_{\mathcal{C}_j^g} \right\|_{200} \leq C n^{-1/2+1/800} \eta^{-1} \leq C.$$

The final inequality uses $\eta^{-1} \leq n^{1/68}$ and $-1/2 + 1/800 + 1/68 < 0$. The normalized trace is bounded, and β_j is bounded on $\mathfrak{A}_n^g$. Hence $\|d_j^{(3)} \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq C n^{-1}$.

For $d_j^{(4)}$, write

$$d_j^{(4)} = (u - u_j) \frac{1}{p} \operatorname{Tr}(PTP_j TR).$$

Cauchy–Schwarz and (6.4) bound this normalized trace on $\mathfrak{A}_n^g$. The preceding bound for $u - u_j$ therefore gives $\|d_j^{(4)} \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq C n^{-1}$.

For $d_j^{(1)}$, put $\Delta_j = u - u_j$ and $A_{1j} = T^{1/2} R_j P_j T P_j T^{1/2}$. Substitution of $P = P_j - \Delta_j P T P_j$ gives

$$d_j^{(1)} = -\Delta_j q_j^* A_{1j} q_j + e_j, \quad |e_j| \mathbf{1}_{\mathfrak{A}_n^g} \leq C \eta^{-1} |\Delta_j|^2 q_j^* T q_j \mathbf{1}_{\mathfrak{A}_n^g}.$$

As in the proof of Lemma 4.3, $\|A_{1j}\|_F \leq C\sqrt{n}$ and $|\operatorname{Tr} A_{1j}| \leq Cn$ on $\mathcal{C}_j^g$. The conditional quadratic-form estimate at order 400, with its factor $n^{1/800}$, gives

$$\|q_j^* A_{1j} q_j \mathbf{1}_{\mathcal{C}_j^g}\|_{400} + \|q_j^* T q_j\|_{400} \leq C,$$

because the centered parts are bounded by $C n^{-1/2+1/800}$. Hölder's inequality now yields

$$\|\Delta_j q_j^* A_{1j} q_j \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq C n^{-1},$$

and

$$\|e_j \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq C \eta^{-1} \|\Delta_j \mathbf{1}_{\mathfrak{A}_n^g}\|_{800}^2 \|q_j^* T q_j\|_{400} \leq C n^{-2+1/68}.$$

Consequently,

$$\|d_j^{(h)} \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq C n^{-1}, \quad h = 1, 3, 4.$$

Inserting the denominator expansion and the four-term decomposition of d_j into (6.17), then combining these $O(n^{-1})$ estimates with (6.23) by the triangle inequality, proves (6.20)–(6.21). $\square$

6.3.2 Moment bound for the residual

Lemma 6.4 bounds the remainder in the expansion of w_n . Taking expectations uses the centering in (6.19) and the bound on $\mathbb{P}((\mathfrak{A}_n^g)^c)$ in (6.10). The resulting estimate at $\mathbb{E}s$, together with Theorem 6.3, gives the following bound.

Theorem 6.5. *Uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$\mathbb{E}[|w_n(E + i\eta)|^{200} \mathbf{1}_{\mathfrak{A}_n^g}] \leq C n^{-200+1/2}. \quad (6.24)$$

Proof. Let $\bar{s} = \mathbb{E}s$ and $\bar{u} = -(1-y)/z + y\bar{s}$. Recall from (4.61) that

$$F_z(\zeta) = -\frac{1}{zp} \operatorname{Tr} \left(I_p + \left[-\frac{1-y}{z} + y\zeta \right] T \right)^{-1} - \zeta,$$

whenever the inverse exists, and that $w_n = F_z(s)$. As shown after (4.61), the positive imaginary parts of u and $\bar{u}$ ensure that s and $\bar{s}$ lie in this domain. Lemma 6.1 places $u(z) \in \mathcal{V}_g(z)$ on $\mathfrak{A}_n^g$. As in

(4.62), split $\mathbb{E}(u - u_c)$ over $\mathfrak{A}_n^g$ and its complement. Lemma 6.1, the Stieltjes bound $|u| + |u_c| \leq C\eta^{-1}$, and (6.10) give

$$|\bar{u} - u_c| \leq Cn^{-23/85} + C\eta^{-1}n^{-581} \leq Cn^{-23/85} + Cn^{-581+1/68} \longrightarrow 0.$$

Thus $\bar{u}(z) \in \mathcal{V}_g(z)$ for all sufficiently large n , uniformly in the stated spectral parameters. If $\zeta_1, \zeta_2 \in \mathbb{C}$ satisfy

$$-\frac{1-y}{z} + y\zeta_h \in \mathcal{V}_g(z), \quad h = 1, 2,$$

then the resolvent identity for $P(v)$ gives

$$|F_z(\zeta_1) - F_z(\zeta_2)| \leq C|\zeta_1 - \zeta_2|. \quad (6.25)$$

Taking expectations in (6.20) and using $\mathbb{E}\hat{Y}_j = 0$ from (6.19) gives

$$\begin{aligned} |\mathbb{E}(w_n \mathbf{1}_{\mathfrak{A}_n^g})| &\leq Cn^{-1+1/400} + \frac{1}{n} \sum_{j=1}^n |\mathbb{E}(\hat{Y}_j \mathbf{1}_{(\mathfrak{A}_n^g)^c})| \\ &\leq Cn^{-1+1/400} + C\eta^{-1}n^{1/800}\mathbb{P}((\mathfrak{A}_n^g)^c)^{1-1/4000} \\ &\leq Cn^{-1+1/400}. \end{aligned}$$

The second term on the penultimate line is bounded, using $\eta^{-1} \leq n^{1/68}$ and (6.10), by $Cn^{1/68+1/800-581(1-1/4000)}$. Its exponent is less than -580 , which proves the final inequality.

Put $\bar{w}_n = F_z(\bar{s})$. Equations (6.25) and (6.16) imply

$$\begin{aligned} |\bar{w}_n| \mathbb{P}(\mathfrak{A}_n^g) &\leq |\mathbb{E}(w_n \mathbf{1}_{\mathfrak{A}_n^g})| + C\mathbb{E}(|s - \bar{s}| \mathbf{1}_{\mathfrak{A}_n^g}) \\ &\leq Cn^{-1+1/400} + Cn^{-1+1/800} \leq Cn^{-1+1/400}. \end{aligned}$$

Since $\mathbb{P}(\mathfrak{A}_n^g) \geq 1/2$ for all large n ,

$$|\bar{w}_n| \leq Cn^{-1+1/400}.$$

On $\mathfrak{A}_n^g$, (6.25) now gives

$$|w_n| \leq C|s - \bar{s}| + Cn^{-1+1/400}.$$

Taking the L^{200} -norm, invoking Theorem 6.3, and using

$$\|(s - \bar{s}) \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq Cn^{-1+1/800},$$

we obtain

$$\|w_n \mathbf{1}_{\mathfrak{A}_n^g}\|_{200} \leq Cn^{-1+1/400}.$$

Raising this estimate to the 200th power proves (6.24), because

$$200 \left(-1 + \frac{1}{400} \right) = -200 + \frac{1}{2}.$$

□

6.4 Stieltjes transforms and spectral exclusion

The residual estimate in Theorem 6.5, the bound for $|s - s_{c,n}|$ in Lemma 6.1, and the local stability condition (IG6) give the bound for $m_n - m_{c,n}$ in Theorem 6.6. Lemma 6.7 uses this estimate to bound integrals against the test functions uniformly over the family, while Lemma 6.8 gives a deterministic criterion for spectral exclusion. We combine these results with estimates on a finite net in $\mathcal{I}$ to complete the proof of Theorem 2.4.

Theorem 6.6. *Uniformly for $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$,*

$$\mathbb{E}[|m_n(E + i\eta) - m_{c,n}(E + i\eta)|^{200} \mathbf{1}_{\mathfrak{A}_n^g}] \leq Cn^{-200+1/2}. \quad (6.26)$$

Proof. Fix $E \in \mathcal{I}$ and $\eta_* \leq \eta \leq 1$, and write $z = E + i\eta$. Put $u = u(z)$ and $u_c = u_c(z)$. For $\mathcal{D}_n = f_n(u) - z$, the exact identities (4.73) and (4.74) give

$$\mathcal{D}_n(z) = \frac{zy}{u} w_n(z)$$

and

$$\mathcal{D}_n(z) = (u - u_c) \left[\frac{1}{uu_c} - y \int \frac{t^2 dH_n(t)}{(1 + tu)(1 + tu_c)} \right].$$

On $\mathfrak{A}_n^g$, Lemma 6.1 and (IG6) bound u^{-1} and the inverse of the factor in brackets. Therefore

$$|u - u_c| \mathbf{1}_{\mathfrak{A}_n^g} \leq C|w_n| \mathbf{1}_{\mathfrak{A}_n^g}.$$

By (4.68) and (4.69), $m_n - m_{c,n} = u - u_c$. Theorem 6.5 proves (6.26). $\square$

To use this transform bound in the exclusion argument, we first control integrals against the test functions from Lemma 5.1. We use the cutoff χ from (5.1), which is zero on $\mathcal{J}$ and one outside $\mathcal{U}$, and the family

$$\mathfrak{F}_{\text{out}} = \{F_{\eta,u} : u \in \mathcal{I}, 0 \leq \eta \leq \eta_0\}.$$

Here $\eta_0 \in (0, 1]$ is fixed. For $\eta > 0$, $F_{\eta,u} = \chi \mathcal{K}_{\eta,u}$, where

$$\mathcal{K}_{\eta,u}(\lambda) = \prod_{k=1}^{34} \frac{1}{(u - \lambda)^2 + k\eta^2}.$$

At $\eta = 0$, $F_{0,u}$ is given by the piecewise definition (5.2), which is zero on $\mathcal{J}$.

Lemma 6.7. *For every fixed $\varepsilon > 0$,*

$$\mathbb{P} \left(\sup_{f \in \mathfrak{F}_{\text{out}}} \left| \int f d(\mu_n - \mu_{c,n}) \right| > \varepsilon \right) \leq C_\varepsilon \left(n^{-399/2} + n^{-581} \right). \quad (6.27)$$

Proof. The total-boundedness and density arguments in the proof of Lemma 5.1 depend only on the deterministic family $\mathfrak{F}_{\text{out}}$. The resolvents used there have heights in $[1/2, 3/4]$, and $\eta_* \leq 1/2$ for all sufficiently large n . For the resulting finite collection of resolvent combinations $g_1, \dots, g_K$, Theorem 6.6 and Minkowski's inequality give

$$\mathbb{E} \left[\left| \int g_\ell d(\mu_n - \mu_{c,n}) \right|^{200} \mathbf{1}_{\mathfrak{A}_n^g} \right] \leq C_\ell n^{-200+1/2}.$$

The number K is independent of n , and both measures have total mass at most d_+ . The bound $|\int g_\ell d(\mu_n - \mu_{c,n})| \leq 2d_+ \|g_\ell\|_\infty$ allows the finitely many remaining n to be included by increasing C_ℓ . The uniform approximation with error $\varepsilon/(4d_+)$, as in Lemma 5.1, contributes at most $\varepsilon/2$ to the difference of the two integrals. Markov's inequality, the finite union bound, and (6.10) then prove (6.27), since $-200 + 1/2 = -399/2$. $\square$

Lemma 6.7 controls the contribution multiplied by χ . The following deterministic criterion combines this with the bounds for $m_n - m_{c,n}$ to show that B_n has no eigenvalues in $\mathcal{I}$.

Lemma 6.8. *There are constants $c_D, \varepsilon_D > 0$ such that, for all large n , the conditions*

$$\sup_{\substack{u \in \mathcal{I} \\ 1 \leq k \leq 34}} |m_n(u + i\sqrt{k}\eta_*) - m_{c,n}(u + i\sqrt{k}\eta_*)| \leq \frac{c_D}{n\eta_*},$$

$$\sup_{f \in \mathfrak{F}_{\text{out}}} \left| \int f d(\mu_n - \mu_{c,n}) \right| \leq \varepsilon_D$$

imply $\text{spec}(B_n) \cap \mathcal{I} = \emptyset$.

Proof. For $x \geq 0$ and $h > 0$, identity (5.7) at $M = 34$ reads

$$\sum_{k=1}^{34} \frac{(-1)^{k-1} \binom{33}{k-1}}{x + kh} = \frac{33! h^{33}}{\prod_{k=1}^{34} (x + kh)}.$$

With $x = (u - \lambda)^2$ and $h = \eta_*^2$, this identity expresses the product kernel as a linear combination of $((u - \lambda)^2 + k\eta_*^2)^{-1}$, $1 \leq k \leq 34$, with coefficients $\eta_*^{-66} (-1)^{k-1} \binom{33}{k-1} / 33!$. The first hypothesis and $|\Im(m_n - m_{c,n})| \leq |m_n - m_{c,n}|$ bound the imaginary part of the transform difference at each of these spectral parameters. Dividing the k -th bound by $\sqrt{k}\eta_*$, summing with the coefficients above, and using $n^{-1} = \eta_*^{68}$ gives

$$\sup_{u \in \mathcal{I}} \left| \int \left[\prod_{k=1}^{34} \frac{1}{(u - \lambda)^2 + k\eta_*^2} \right] d(\mu_n - \mu_{c,n})(\lambda) \right| \leq C c_D.$$

For all sufficiently large n , $\eta_* \leq \eta_0$, so $\chi \mathcal{K}_{\eta_*, u} = F_{\eta_*, u} \in \mathfrak{F}_{\text{out}}$. By (IG6) and (5.1), $(1 - \chi)\mu_{c,n} = 0$. The second hypothesis and the preceding bound give

$$0 \leq \int (1 - \chi) \mathcal{K}_{\eta_*, u} d\mu_n \leq C c_D + \varepsilon_D \quad (u \in \mathcal{I}).$$

If B_n had an eigenvalue $\lambda_\ell \in \mathcal{I}$, choosing $u = \lambda_\ell$ would make the summand corresponding to this eigenvalue contribute

$$\frac{1/n}{\prod_{k=1}^{34} k\eta_*^2} = \frac{1}{34!}.$$

Choosing $C c_D + \varepsilon_D < 1/34!$ gives a contradiction. $\square$

Proof of Theorem 2.4. The event and transform assertions follow from (6.10) and Theorem 6.6. To prove (2.4), we verify the two conditions in Lemma 6.8 using Theorem 6.6 and Lemma 6.7.

With $\beta_* = 277/272$ and $\eta_* = n^{-1/68}$ from (2.3), choose an $n^{-\beta_*}$ -net of $\mathcal{I}$, with at most $C n^{\beta_*}$ points, and use the 34 imaginary parts $\sqrt{k}\eta_*$. For all sufficiently large n , $\sqrt{34}\eta_* \leq 1$, so all these heights lie in the range of Theorem 6.6. The derivative bound (5.11) gives the interpolation estimate

$$\frac{C n^{-\beta_*} \eta_*^{-2}}{(n \eta_*)^{-1}} = C n^{69/68 - 277/272} = C n^{-1/272} \rightarrow 0.$$

For each grid point and each imaginary part, Markov's inequality and (6.26) give

$$\begin{aligned} \mathbb{P} \left(|m_n - m_{c,n}| > \frac{c_D}{2n\eta_*}, \mathfrak{A}_n^g \right) \\ \leq C(n\eta_*)^{200} n^{-200+1/2} = Cn^{-200/68+1/2}. \end{aligned}$$

After summing over the net and the 34 imaginary parts, the exponent is

$$\beta_* - \frac{200}{68} + \frac{1}{2} = \frac{277}{272} - \frac{800}{272} + \frac{136}{272} = -\frac{387}{272}. \quad (6.28)$$

The exponents -581 and $-399/2$, coming respectively from the event complement and the uniform test-function estimate, are both strictly smaller than $-387/272$. Lemma 6.8 therefore gives (2.4). Finally,

$$\frac{387}{272} = 1 + \frac{115}{272} > 1,$$

so Borel–Cantelli proves (2.5). $\square$

6.5 Choice of moment order

The preceding proof uses $q_g = 200$ in the transform estimate. To explain this choice, we compare the probability exponents at a general moment order.

At a fixed moment order $2 \leq q \leq Q_g/4$, the second-moment contribution to the trace estimate has the factor $n^{q\delta} = n^{q/800}$. The product $d_j^{(2)} \varepsilon_j$ in (6.23) contributes the larger factor $n^{2q\delta} = n^{q/400}$. If the residual and transform estimates are obtained at order q with these factors, the final finite-net calculation gives the probability exponent

$$\beta_* - \frac{q}{68} + \frac{q}{400} = \frac{277}{272} - q \frac{83}{6800}. \quad (6.29)$$

Summability requires this exponent to be less than -1 , equivalently

$$q > \frac{(1 + 277/272)6800}{83} = \frac{13725}{83} = 165.361 \dots$$

Thus any integer $166 \leq q \leq Q_g/4$ makes the final probability bound summable, provided the preceding moment estimates are proved at order q . The q -dependent residual estimates at $q_g = 200$ use quadratic-form moments only through $4q_g = 800 < Q_g$. The construction of $\mathfrak{A}_n^g$ and the tail estimate (6.15) use moment order $Q_g = 4000$.

The moment order $q_0 = 164$ from Theorem 2.2 does not give a summable bound by the same argument:

$$\frac{277}{272} - \frac{164}{68} + \frac{164}{400} = \frac{6925 - 16400 + 2788}{6800} = -\frac{6687}{6800} > -1.$$

7 Conclusion

Independent sample columns with martingale-difference coordinates satisfy, under the stated moment assumptions, the conditional Hilbert–Schmidt quadratic-form estimates used by the leave-one-column-out resolvent method. Under the assumptions of Theorem 2.2, the estimate

$$\|(m_n(E + i\eta) - m_{c,n}(E + i\eta)) \mathbf{1}_{\mathfrak{A}_n}\|_{164} = O(n^{-1})$$

holds uniformly for $E \in \mathcal{I}$ and $n^{-1/68} \leq \eta \leq 1$, where $\mathfrak{A}_n$ is the high-probability event in that theorem. The probability that B_n has an eigenvalue in $\mathcal{I}$ is summable. Theorem 2.4 permits the polynomial moment bound

$$\max_{i,j} \mathbb{E}|X_{ij}|^{8000} \leq Cn^5,$$

and gives the summable bound $Cn^{-387/272}$ for the probability that B_n has an eigenvalue in $\mathcal{I}$.

An extension of this proof to dependent sample columns would require control of the conditional expectations of the centered quadratic forms in Propositions 3.2 and 3.3, together with a replacement for the cancellation (4.29) in the martingale argument. A separate quantitative problem is to lower the coordinate moment orders while preserving a summable spectral-exclusion estimate.

References

- [1] Z. D. Bai and J. W. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices, *Ann. Probab.* **26** (1998), no. 1, 316–345. <https://doi.org/10.1214/aop/1022855421>.
- [2] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed., Springer Series in Statistics, Springer, New York, 2010. <https://doi.org/10.1007/978-1-4419-0661-8>.
- [3] J. Bryson, R. Vershynin, and H. Zhao, Marchenko–Pastur law with relaxed independence conditions, *Random Matrices Theory Appl.* **10** (2021), no. 4, 2150040, 28 pp. <https://doi.org/10.1142/S2010326321500404>.
- [4] F. Götze and A. N. Tikhomirov, Limit theorems for spectra of positive random matrices under dependence, *J. Math. Sci. (N.Y.)* **133** (2006), no. 3, 1257–1276. <https://doi.org/10.1007/s10958-006-0035-8>.
- [5] I. Pinelis, Optimum bounds for the distributions of martingales in Banach spaces, *Ann. Probab.* **22** (1994), no. 4, 1679–1706. <https://doi.org/10.1214/aop/1176988477>. Correction: *Ann. Probab.* **27** (1999), no. 4, 2119. <https://doi.org/10.1214/aop/1022874833>.
- [6] P. Yaskov, The universality principle for spectral distributions of sample covariance matrices, arXiv:1410.5190v3 [math.PR] (2014). <https://arxiv.org/abs/1410.5190>.
- [7] P. Yaskov, LLN for quadratic forms of long memory time series and its applications in random matrix theory, *J. Theoret. Probab.* **31** (2018), no. 4, 2032–2055. <https://doi.org/10.1007/s10959-017-0767-z>.