Support properties of spectra of polynomials in Wigner matrices

Greg W. Anderson
University of Minnesota
www.math.umn.edu/∼gwanders
gwanders@umn.edu

Advances in Random Matrix Theory

Institute for Mathematics and its Applications
University of Minnesota
June 18–29, 2012

Lecture 1 (June 18, 2012)

The plan

The plan for the course is

- to fill in background for and to state the result of Haagerup, Schultz and Thorbjørnsen for polynomials in GUE matrices,
- to prove this result taking care to explain the most important (because re-usable) “tricks,”
- to discuss variants, generalizations and related results in the literature,
- to prove an extension of HST to polynomials in Wigner matrices (not the best and sharpest possible) using methods worth learning about in their own right and finally,
- to indicate in passing various open questions in the area.

“No eigenvalues outside the support...” (Bai-Silverstein)

A very important result on spectral support is due to Bai-Silverstein [Bai-Silverstein 1998]. It is the archetype for all the results to be discussed in this course but it is still not completely encompassed by the theory it has inspired.

We will state the result in the next several frames.

Bai-Silverstein setup

First we describe the basic source of randomness.

- For each positive integer N fix a positive integer $n = n(N)$ and put $c^{(N)} = N/n$.
- Fix a constant $1 < c < \infty$ and assume that $\lim_{N\to\infty} c^{(N)} = c$.
- (We are lazily skipping the case $0 < c \leq 1$ which is not much different.)
- Let $\{\xi_{ij}\}_{i,j=1}^{\infty}$ be a i.i.d. family of $\mathbb{C}$-valued random variables with $\mathbf{E}|\xi_{ij}|^4 < \infty$, $\mathbf{E}|\xi_{ij}|^2 = 1$ and $\mathbf{E}\xi_{ij} = 0$.
- Let $\Xi^{(N)}$ be the N-by-n matrix with entries $\Xi^{(N)}[i,j] = \xi_{ij}$.

Empirical distributions of eigenvalues

Recall that for an N-by-N hermitian matrix A the *empirical distribution of eigenvalues* of A is the measure $\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$ on the real line where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of A.

Bai-Silverstein setup (concluded)

Next we describe the model.

- For each positive integer N fix a (deterministic) positive semidefinite N-by-N hermitian matrix $T^{(N)}$ and assume that $\sup_N [\![T^{(N)}]\!] < \infty$.
- Assume that the empirical distribution $\tau^{(N)}$ of eigenvalues of $T^{(N)}$ converges weakly to a limit τ.
- Put $B^{(N)} = \frac{1}{n}(\Xi^{(N)})^* T^{(N)} \Xi^{(N)}$ for each N. This is our model.
- Let $\beta^{(N)}$ be the empirical distribution of the eigenvalues of $B^{(N)}$.
- One takes for granted that $\beta^{(N)}$ converges almost surely to a limit $\beta = \beta_{\tau^{(N)}, c^{(N)}}$.
- (The latter was already known at time of writing of Bai-Silverstein.)

The Stieltjes transform of β

Let

$$S_\beta(z) = \int \frac{\beta(dt)}{t-z}$$

be the Stieltjes transform of β.

(We will review Stieltjes transforms more systematically later.)

One knows that for $\Im z > 0$ one has

$$S_\beta(z) = \int \frac{\tau(dt)}{t(1-c-czS_\beta(z))} \text{ and } \Im\left(-\frac{1-c}{z} + cS_\beta(z)\right) > 0,$$

and that these conditions uniquely characterize $S_\beta(z) = S_{\beta_{\tau,c}}(z)$.

(From the point of view to be cultivated below the preceding is an instance of the *Schwinger-Dyson equation*.)

In the statement of the following theorem we will need not only the limiting eigenvalue distribution β but also the distributions $\beta_{\tau^{(N)},c^{(N)}}$ whose Stieltjes transforms are characterized as above, with τ replaced by $\tau^{(N)}$ and c replaced by $c^{(N)}$.

Statement of the Bai-Silverstein result

Theorem ([Bai-Silverstein 1998])

Fix a closed interval $I \subset (0, \infty)$ disjoint from the support of β. Assume furthermore that I is disjoint from the support of $\beta_{\tau^{(N)}, c^{(N)}}$ for all sufficiently large N. Then there exists a random positive integer N_0 such that the support of $\beta^{(N)}$ is disjoint from I for $N \geq N_0$.

In short, the eigenvalues for large N "stick" to the support of the limiting distribution.

The Bai-Silverstein result has yet to be fully understood within the circle of ideas to be discussed in this course. It remains a touchstone.

Largest eigenvalues of (single) Wigner matrices

The "state of the art" for single Wigner matrices concerning convergence of the largest eigenvalue (best in terms of having lowest moment assumptions) is the following result of Bai-Yin.

Largest eigenvalues of Wigner matrices

- Let $\{\{\xi_{ij}\}\}_{1\leq i\leq j<\infty}$ be a family of independent $\mathbb{C}$-valued random variables.
- Assume the following:
  - The law of ξ_{ij} depends only on $\mathbf{1}_{i<j}$.
  - ξ_{11} is real-valued.
  - $\mathbf{E}\xi_{11} = 0$ and $\mathbf{E}\xi_{12} = 0$.
- Put $\sigma^2 = \mathbf{E}|\xi_{12}|^2$.
- Let $\Xi^{(N)}$ be the N-by-N random hermitian matrix with above-diagonal entries $\Xi^{(N)}[i,j] = \xi_{ij}$ for $1 \leq i \leq j \leq N$.

Largest eigenvalues of Wigner matrices (continued)

Theorem ([Bai-Yin 1988])

Notation and assumptions are as above. The largest eigenvalue of $\frac{\Xi^{(N)}}{\sqrt{N}}$ converges almost surely to 2σ.

In the same paper a complementary assertion was also proved showing that (without changing any other feature of the setup) the fourth moment hypothesis cannot be lowered without destroying convergence. In that sense fourth moments are best possible.

- In language generalizing well to the polynomials-in-independent-Wigner-matrices setting, we can reformulate the Bai-Yin result as follows.
- Let $\mu^{(N)}$ be the empirical distribution of the eigenvalues of $\underline{\Xi}^{(N)}$.
- Let μ be the centered semicircular law of variance σ^2, i.e., $\frac{d\mu}{dx} = \frac{1}{2\pi\sigma^2}\mathbf{1}_{|x|<2\sigma}\sqrt{4\sigma^2 - x^2}$.
- Note that $[-2\sigma, 2\sigma]$ is the support of μ.

Largest eigenvalues of Wigner matrices (concluded)

- By definition the ϵ-*neighborhood* of a set $K \subset \mathbb{R}$ is $\cup_{x\in K}(x-\epsilon, x+\epsilon)$.
- Almost sure weak convergence $\mu^{(N)} \Rightarrow \mu$ taken for granted, the result of Bai-Yin is equivalent to the assertion that for every $\epsilon > 0$, the support of $\mu^{(N)}$ is in the ϵ-neighborhood of the support of μ for $N \gg 0$, almost surely.
- We will be interested in statements of similar form for more complicated matrix models built by taking matrices of noncommutative polynomials in several independent Wigner matrices.

Apology for GUE matrices

Before discussing Wigner matrices we will be discussing GUE matrices for a long while. (And probably we will not get very far into the study of Wigner matrices anyway.) This is not actually a big loss of generality because there is a core of operator-theoretic and complex-analytic technique in this area which while undoubtedly motivated by random matrix applications has little intrinsically to do with random matrices. We might as well learn about these techniques while working in the “easy” GUE case.

More on GUE matrices

To clarify normalizations we make the following definition.

- We say that an N-by-N hermitian random matrix X is a *GUE matrix* if the random variables

$$\{X[i,i]\}_{i=1}^{N} \cup \{\sqrt{2}\Re X[i,j], \sqrt{2}\Im X[i,j]\}_{1\le i<j\le N}$$

  are i.i.d. standard Gaussian.

- Here and below $\Re =$ "real part" and $\Im =$ "imaginary part".

Fake GUE matrices

In order to explain how one pushes away from GUE hypotheses to handle more general classes of Wigner matrices without weighing the audience down by a lot of tedious hypotheses, we will consider for didactic purposes the following classes of random variables and random matrices.

- We say that a random variable X is *fake standard Gaussian* if $|X| \leq 100$ and $\mathbf{E}X^n = \mathbf{E}Z^n$ for $n = 1, \ldots, 5$ where Z is a standard Gaussian random variable.
- (As a nice application of noncommutative probability ideas we will construct many nontrivial examples.)
- We say that a random hermitian matrix X is *fake GUE* if the family of random variables

$$\{X[i,i]\}_{i=1}^N \cup \{\sqrt{2}\Re X[i,j], \sqrt{2}\Im X[i,j]\}_{1 \leq i < j \leq N}$$

  is independent and fake standard Gaussian. (We do not require these random variables to be identically distributed.)

Largest eigenvalues of GUE matrices and universality

In this frame and the next we open up a brief side-discussion.

- Suppose that one is given for each N an N-by-N GUE matrix $\Xi^{(N)}$. The limiting behavior of the largest eigenvalue $\lambda_1^{(N)}$ of $\Xi^{(N)}$ is known in very fine detail.

- Not only does one know that $\frac{\lambda_1^{(N)}}{\sqrt{N}} \to_{N\to\infty} 2$ almost surely, but furthermore, for example, one knows that $N^{2/3}(\frac{\lambda_1^{(N)}}{\sqrt{N}} - 2)$ converges in distribution to the Tracy-Widom law. In particular the scale of fluctuations is known.

- Edge-universality results for Wigner matrices lead to similar conclusions for largest eigenvalues of Wigner matrices in wide classes of such.

Spectral support results versus edge-universality results

- Even before we start to discuss results about spectral support of polynomials in independent Wigner matrices, it is clear that such results would be crude in comparison to the edge-universality results that one reasonably expects to hold.
- This course should thus be viewed not so much as a catalog of results in a certain area of RMT as it is an attempt to introduce polynomials in independent Wigner matrices as an area ripe for investigation with respect to universality questions.
- Actually universality questions are by and large unresolved for polynomials in independent GUE matrices, let alone polynomials in independent Wigner matrices.
- It seems a good bet that at least some of the tools used in this course for study of spectral support will retain their value for study of universality.

Our next goal is...

...to motivate by example and then to formulate the theorem of Haagerup-Schultz-Thorbjørnsen (HST for short) ([Haagerup-Schultz-Thorbjørnsen 2006]) and its precursor, namely the theorem of Haagerup-Thorbjørnsen ([Haagerup-Thorbjørnsen 2005]) (HT for short).

These theorems assert a strong sort of convergence of the empirical distribution of eigenvalues of a polynomial in independent GUE matrices to its limit.

Let's have a look at a special case accurately conveying the flavor of HST.

For each positive integer N fix independent N-by-N GUE matrices $\Xi_1^{(N)}$ and $\Xi_2^{(N)}$. Consider the hermitian random matrix

$$\Upsilon^{(N)} = \mathrm{i}(\Xi_1^{(N)}\Xi_2^{(N)} - \Xi_2^{(N)}\Xi_1^{(N)})$$

and let $\mu^{(N)}$ denote the empirical distribution of the eigenvalues of $\frac{\Upsilon^{(N)}}{N}$.

It is known [Nica-Speicher 1998] that in a suitable sense $\mu^{(N)}$ converges to a measure μ which has a density with respect Lebesgue measure satisfying $\operatorname{supp}\mu = [-r, r]$ where $r = \sqrt{(11 + 5\sqrt{5})/2}$.

Sample special case of HST (continued)

It then follows by HST that the support of $\mu^{(N)}$ tends to the support of μ and thus that the largest eigenvalue of $\frac{\Upsilon^{(N)}}{N}$ tends almost surely to $\sqrt{(11+5\sqrt{5})/2}$.

Sample special case of HST (complicated density)

In case you were wondering, here is an explicit description of the measure μ. We have

$$\frac{d\mu(t)}{dt} = \frac{\sqrt{3}}{2\pi|t|}\left(\frac{3t^2+1}{9h(t)} - h(t)\right), \quad |t| \leq \sqrt{(11+5\sqrt{5})/2},$$

where

$$h(t) = \sqrt[3]{\frac{18t^2+1}{27} + \sqrt{\frac{t^2(1+11t^2-t^4)}{27}}}.$$

This particular distribution has recently attracted some interest in connection with stochastic calculus, both classical and free. See for example [Deya-Nourdin 2011] and references therein.

Sample special case of HST (continued)

This example gives one pause since the explicit description of the limiting spectral distribution is very complicated and yet the commutator is the simplest self-adjoint noncommutative polynomial you can think of.

Fortunately, to prove HST, one gives up on using an explicit description of the limit in favor of an inexplicit operator-theoretic description which still gives excellent control of the situation.

Interlude for notation: Kronecker products

Given any matrix X we invariably denote its entry in row i and column j by $X[i,j]$.

Given a k-by-ℓ matrix X and an m-by-n matrix Y with (say) complex entries, we define the *Kronecker product* as usual by

$$X \otimes Y = \left[\begin{array}{ccc} X[1,1]Y & \dots & X[1,\ell]Y \\ \vdots & & \vdots \\ X[k,1]Y & \dots & X[k,\ell]Y \end{array} \right].$$

Let $\mathbf{I}_n$ denote the n-by-n identity matrix.

Sample special case of HST (continued)

At the expense of having to repeat ourselves a bit later, we now quickly sketch the construction of an operator suitable for describing the almost sure weak limit of the empirical distribution of eigenvalues of $\frac{\Upsilon^{(N)}}{N}$.

Sample special case of HST (continued)

Let $\mathcal{H}$ be a Hilbert space equipped with an orthonormal basis

$$\{v(i)\}_{i=0}^{\infty}.$$

For $\ell = 1, 2$ let

$$\theta_\ell \in B(\mathcal{H})$$

be the partial isometry given by the strange rule

$$\theta_\ell v(i) = \sum_{j=0}^{\infty} \mathbf{1}_{2^j - 1 \leq i < 2^{j+1} - 1} v(i + \ell 2^j).$$

We then define

$$\Xi_\ell = \theta_\ell + \theta_\ell^* \in B(\mathcal{H}).$$

By construction $[\![\Xi_\ell]\!] = 2$.

(Here and below $[\![\cdot]\!]$ denotes the operator norm for bounded operators on Hilbert space.)

Sample special case of HST (continued)

In the next frame we explain why the family

$$\{\theta_{\ell_1 \cdots \ell_k} v(0)\}$$

indexed by all sequence $\ell_1 \cdots \ell_k$ in $\{1, 2\}$ including the empty sequence is merely a relabeling of the canonical basis $\{v(i)\}$.

Sample special case of HST (continued)

The strange rule is actually less strange than it appears if one adopts a modification of the usual binary notation for nonnegative integers.

Instead of counting

$$0, 1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, 1100, \ldots$$

we count as follows, using digits 1 and 2 instead of digits 0 and 1, but otherwise still using place value notation:

$$\emptyset, 1, 2, 11, 12, 21, 22, 111, 112, 121, 122, 211, 212, 221, 222, 1111, \ldots$$

From the latter perspective, the operator θ_ℓ maps $v(i)$ to $v(i')$ where i' is (so to speak) the result of tacking the digit ℓ on the left side of i, e.g., $\theta_2 v(21) = v(221)$.

Sample special case of HST (continued)

Let us now think of $\mathcal{H}$ as a space of column vectors with entries indexed by the nonnegative integers. Correspondingly, let us identify $B(\mathcal{H})$ with a set of matrices with rows and columns indexed by the nonnegative integers in the obvious way.

Let $x = [x_1 \; x_2]$ with x_1 and x_2 real. We then have

$$x_1\Xi_1 + x_2\Xi_2 = \begin{bmatrix} & x & & & & \\ x^* & & x \otimes \mathbf{I}_2 & & & \\ & x^* \otimes \mathbf{I}_2 & & x \otimes \mathbf{I}_4 & & \\ & & x^* \otimes \mathbf{I}_4 & & \ddots & \\ & & & \ddots & & \end{bmatrix}.$$

Thus the operators Ξ_1 and Ξ_2 are "self-similar" in what turns out to be a useful way.

Sample special case of HST (continued)

Anyway, let us now put

$$\Upsilon = \mathrm{i}(\Xi_1\Xi_2 - \Xi_2\Xi_1) \in B(\mathcal{H}).$$

Since the operators Ξ_1 and Ξ_2 are self-adjoint, so is Υ.

Sample special case of HST (continued)

We next explain how to manufacture a probability measure from Υ.

Given a continuous function

$$\varphi : \mathrm{Spec}(\Upsilon) \to \mathbb{C},$$

fix any sequence

$$\varphi_n(\mathbf{X}) \in \mathbb{C}[\mathbf{X}]$$

of polynomials such that

$$\lim_{n\to\infty} \varphi_n(x) = \varphi(x)$$

uniformly on $\mathrm{Spec}(\Upsilon)$,

and then take

$$\varphi(A) = \lim_{n\to\infty} \varphi_n(A) \in B(\mathcal{H}).$$

The sequence is norm convergent and its limit $\varphi(A)$ is a normal operator.

Sample special case of HST (continued)

The construction leading to $\varphi(A)$ is the *functional calculus* at A,

a.k.a. the inverse Gelfand transform.

The Riesz representation theorem then yields a unique probability measure

$$\mu_\Upsilon$$

such that

$$\int \varphi \, d\mu_\Upsilon = (v(0), (\varphi|_{\mathrm{Spec}(\Upsilon)}(\Upsilon))v(0))$$

for all bounded continuous test-functions $\varphi : \mathbb{R} \to \mathbb{C}$.

Sample special case of HST (continued)

The measure μ_Υ so constructed turns out to be the almost sure weak limit of the empirical distribution of eigenvalues of $\frac{\Upsilon^{(N)}}{N}$.

Remark The method for proving HST yields as a byproduct a proof of this almost sure weak convergence.

Remark The probability measures analogously associated to $x_1\Xi_1 + x_2\Xi_2$ for real x_1 and x_2 such that $x_1^2 + x_2^2 = 1$ is the *semicircle law*, i.e., the measure having density

$$\frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{|x|<2}$$

with respect to Lebesgue measure on the real line.

Sample special case of HST (concluded)

Forward-looking remark If you replace $\Xi_1^{(N)}$ and $\Xi_2^{(N)}$ by, say, independent "Bai-Yin-Wigner" matrices, then the largest eigenvalue of $\frac{\Upsilon^{(N)}}{N}$ still converges to $\sqrt{(11+5\sqrt{5})/2}$ by the generalization of HST proved by the speaker [Anderson, Ann. Probab., to appear]. We discuss this and other generalizations of HST only later in the course and perhaps not in very much depth.

We now have to spend some time acquiring language and fixing notation for HST.

Algebras

- In this course *algebras* $\mathcal{A}$ are associative (not usually commutative) with complex scalars.

- The main examples of algebras used here are

  - $\mathcal{A} = \mathrm{Mat}_n = \mathrm{Mat}_n(\mathbb{C})$ (see below),
  - $\mathcal{A} = \mathbb{C}\langle \mathbf{X}_{1\dots m} \rangle$ (see below),
  - $\mathcal{A} = B(H)$ where H is a Hilbert space (see below) and
  - algebras of matrices with entries in any of these algebras.
  - More generally we use C^*-algebras (see below).

Units

- In this course we confine our attention to unital algebras, i.e., those possessing units.
- This restriction is convenient but it is definitely a loss of generality since sometimes, for example, one might want to study the C^*-algebra of compact operators on an infinite-dimensional Hilbert space. But not here...
- The unit of an algebra $\mathcal{A}$ is denoted $1_{\mathcal{A}}$ or simply 1, context permitting; other notation such as $\mathbf{I}_n$ for the n-by-n identity matrix may be used in the matrix algebra context (see below).
- Let $\mathcal{A}^\times$ denote the group of invertible elements of $\mathcal{A}$, i.e.,

$$\mathcal{A}^\times = \{x \in \mathcal{A} \mid \exists y \in \mathcal{A} \text{ s.t. } xy = yx = 1_{\mathcal{A}}\}.$$

One has $xy = 1 \Rightarrow yx = 1$ automatically if $\mathcal{A}$ is finite-dimensional over $\mathbb{C}$ but otherwise not necessarily.

Involutions

- Given an algebra $\mathcal{A}$, an *involution* $*$ is a $\mathbb{C}$-conjugate-linear map
$$a \mapsto a^* : \mathcal{A} \to \mathcal{A}$$
satisfying
$$(ab)^* = b^* a^* \ \text{ and } \ (a^*)^* = a.$$
Necessarily $1_{\mathcal{A}}^* = 1_{\mathcal{A}}$.

- The main example of an involution to keep in mind is the transpose-conjugate operation on the algebra of N-by-N matrices with entries in $\mathbb{C}$.

- More generally one should keep in mind the operation of taking adjoint in $B(H)$ where H is a Hilbert space.

- A $*$-*algebra* is an algebra equipped with an involution $*$.
- The main examples of $*$-algebras used here are
  - $\mathcal{A} = \mathrm{Mat}_n = \mathrm{Mat}_n(\mathbb{C})$ (see below),
  - $\mathcal{A} = \mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle$ (see below)
  - $\mathcal{A} = B(H)$ where H is a Hilbert space (see below),
  - C^*-algebras (see below) and
  - $*$-algebras of matrices with entries in any of the algebras above (see below).
- In fact, almost all algebras considered below are $*$-algebras (see below).
- If $\mathcal{A}$ is a $*$-algebra, let

$$\mathcal{A}_{\mathrm{sa}} = \{a \in \mathcal{A} \mid a = a^*\}.$$

(Here sa stands for *self-adjoint*.)

Matrices with algebra entries

- Let $\mathrm{Mat}_N(\mathcal{A})$ denote the algebra of N-by-N matrices with entries in an algebra $\mathcal{A}$.
- More generally, let $\mathrm{Mat}_{k\times\ell}(\mathcal{A})$ denote the space of k-by-ℓ matrices with entries in $\mathcal{A}$.
- When $\mathcal{A} = \mathbb{C}$ we just write $\mathrm{Mat}_N = \mathrm{Mat}_N(\mathbb{C})$ and $\mathrm{Mat}_{k\times\ell} = \mathrm{Mat}_{k\times\ell}(\mathbb{C})$.
- $X[i,j]$ invariably denotes the entry of a matrix X in row i and column j.

- For $X \in \mathrm{Mat}_{k \times \ell}(\mathcal{A})$ and $\mathcal{A}$ a $*$-algebra, the rule $X^*[j, i] = X[i, j]^*$ defines $X^* \in \mathrm{Mat}_{\ell \times k}(\mathcal{A})$.
- $\mathrm{Mat}_N(\mathcal{A})$ is thus automatically a $*$-algebra whenever $\mathcal{A}$ is.

Kronecker flattening

If $\mathcal{A} = \mathrm{Mat}_N(\mathcal{B})$ we identify $\mathrm{Mat}_{k\times\ell}(\mathcal{A})$ with $\mathrm{Mat}_{kN\times\ell N}(\mathcal{B})$ by viewing elements of the latter as k-by-ℓ arrays of N-by-N blocks.

In order to call this often-used rule to mind briefly, we will refer to it as *Kronecker flattening* in the sequel.

$$\left[\begin{array}{ccc} \left[\begin{array}{cc} 1 & 2 \\ 3 & 4 \end{array}\right] & \left[\begin{array}{cc} 5 & 6 \\ 7 & 8 \end{array}\right] & \left[\begin{array}{cc} 9 & 10 \\ 11 & 12 \end{array}\right] \\ \\ \left[\begin{array}{cc} 13 & 14 \\ 15 & 16 \end{array}\right] & \left[\begin{array}{cc} 17 & 18 \\ 19 & 20 \end{array}\right] & \left[\begin{array}{cc} 21 & 22 \\ 23 & 24 \end{array}\right] \end{array}\right]$$

$$\overset{\substack{\text{Kronecker} \\ \text{flattening rule}}}{=} \left[\begin{array}{cccccc} 1 & 2 & 5 & 6 & 9 & 10 \\ 3 & 4 & 7 & 8 & 11 & 12 \\ 13 & 14 & 17 & 18 & 21 & 22 \\ 15 & 16 & 19 & 20 & 23 & 24 \end{array}\right].$$

Kronecker products (redux)

Let $\mathcal{A}$ and $\mathcal{B}$ be algebras where at most one of the algebra differs from $\mathbb{C}$ and let $\mathcal{C}$ be the larger of the two algebras.

Given $X \in \mathrm{Mat}_{k\times\ell}(\mathcal{A})$ and $Y \in \mathrm{Mat}_{m\times n}(\mathcal{B})$ we define

$$X \otimes Y \in \mathrm{Mat}_{km\times\ell n}(\mathcal{C})$$

by the rule

$$X \otimes a = \begin{bmatrix} X[1,1]Y & \dots & X[1,\ell]Y \\ \vdots & & \vdots \\ X[k,1]Y & & X[k,\ell]Y \end{bmatrix}.$$

When $\mathcal{C} = \mathbb{C}$ this is just the usual Kronecker product.

Note that if $\mathcal{C}$ is a $*$-algebra, then

$$(X \otimes Y)^* = X^* \otimes Y^*.$$

Noncommutative polynomials

- $\mathbb{C}\langle \mathbf{X}_1, \ldots, \mathbf{X}_m \rangle = \mathbb{C}\langle \mathbf{X}_{1\ldots m} \rangle$ denotes the algebra of noncommutative polynomials in noncommuting variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$.

- Informally, a noncommutative polynomial is an expression into which one sensibly plugs square matrices of the same size.

- By definition a Hamel basis for $\mathbb{C}\langle \mathbf{X}_{1\ldots m} \rangle$ over $\mathbb{C}$ consists of all *monomials* in the variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$, i.e., expressions of the form $\mathbf{X}_{i_1} \cdots \mathbf{X}_{i_k}$ for nonnegative integers k and integers $i_1, \ldots, i_k \in \{1, \ldots, m\}$.

- In particular, $1_{\mathbb{C}\langle \mathbf{X}_{1\ldots m} \rangle}$ is the empty monomial.

- On the basis consisting of monomials, multiplication is given simply by juxtaposition.

$*$-algebra structure on $\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle$

- We equip $\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle$ with $*$-algebra structure by enforcing the rule $\mathbf{X}_i^* = \mathbf{X}_i$ for $i = 1, \ldots, m$ along with the rules obeyed by involutions.

Example:

$$\begin{aligned}
(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3 + \mathrm{i}\mathbf{X}_4\mathbf{X}_5)^* &= (\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3)^* - \mathrm{i}(\mathbf{X}_4\mathbf{X}_5)^* \\
&\quad \text{(by antilinearity of } *\text{)} \\
&= \mathbf{X}_3^*\mathbf{X}_2^*\mathbf{X}_1^* - \mathrm{i}\mathbf{X}_5^*\mathbf{X}_4^* \\
&\quad \text{(because } (ab)^* = b^*a^*\text{)} \\
&= \mathbf{X}_3\mathbf{X}_2\mathbf{X}_1 - \mathrm{i}\mathbf{X}_5\mathbf{X}_4 \\
&\quad \text{(because we stipulate } \mathbf{X}_i^* = \mathbf{X}_i\text{)}
\end{aligned}$$

Lecture 2 (June 19, 2012)

- Given $f = f(\mathbf{X}_{1\dots m}) = f(\mathbf{X}_1, \dots, \mathbf{X}_m) \in \mathbb{C}\langle \mathbf{X}_1, \dots, \mathbf{X}_m \rangle$, an algebra $\mathcal{A}$ and an m-tuple $a = (a_1, \dots, a_m) \in \mathcal{A}^m$, let $f(a) = f(a_1, \dots, a_m) \in \mathcal{A}$ denote the result of evaluating f at $\mathbf{X}_i = a_i$ for $i = 1, \dots, m$.
- E.g., if $f = \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3 + \mathrm{i}\mathbf{X}_4\mathbf{X}_5$ then $f(a) = a_1a_2a_3 + \mathrm{i}a_4a_5$.
- Note that if $a \in \mathcal{A}_{\mathrm{sa}}^m$ then $f^*(a) = f(a)^*$.
- E.g., if $f = \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3 + \mathrm{i}\mathbf{X}_4\mathbf{X}_5$ then $f(a)^* = (a_1a_2a_3 + \mathrm{i}a_4a_5)^* = a_3a_2a_1 - \mathrm{i}a_5a_4 = f^*(a)$.
- In particular if $f^* = f$ and $a \in \mathcal{A}_{\mathrm{sa}}^m$ then $f(a)^* = f^*(a) = f(a) \in \mathcal{A}_{\mathrm{sa}}$.

Matrices with polynomial entries

- Given $f \in \mathrm{Mat}_{k\times\ell}(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$, we define $f(a_1,\ldots,a_m) \in \mathrm{Mat}_{k\times\ell}(\mathcal{A})$ by $f(a_1,\ldots,a_m)[i,j] = f[i,j](a_1,\ldots,a_m)$, i.e., we evaluate entry-by-entry.
- Note that if $\mathcal{A}$ is a $*$-algebra and $a \in \mathcal{A}_{\mathrm{sa}}^m$, then $f^*(a) = f(a)^*$ and in particular if $f^* = f$ then $f(a) \in \mathcal{A}_{\mathrm{sa}}$.
- If $\mathcal{A} = \mathrm{Mat}_N$, then we view $f(a)$ as an element of $\mathrm{Mat}_{kN\times\ell N}$ by the Kronecker flattening rule.

Example of evaluation

For example, given

$$f(\mathbf{X}_{1\dots3}) = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2\mathbf{X}_3 \\ \mathbf{X}_3\mathbf{X}_2 & \mathbf{X}_2 + 10\mathbf{X}_3 \end{bmatrix} \in \mathrm{Mat}_2(\mathbb{C}\langle \mathbf{X}_{1\dots3}\rangle)_{\mathrm{sa}}$$

then

$$f\left(\begin{bmatrix} 1 & 2\mathrm{i} \\ -2\mathrm{i} & 3 \end{bmatrix}, \begin{bmatrix} 4 & -5\mathrm{i} \\ 5\mathrm{i} & 4 \end{bmatrix}, \begin{bmatrix} 7 & 8\mathrm{i} \\ -8\mathrm{i} & 9 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 1 & 2\mathrm{i} & -12 & -13\mathrm{i} \\ -2\mathrm{i} & 3 & 3\mathrm{i} & -4 \\ -12 & -3\mathrm{i} & 74 & 75\mathrm{i} \\ 13\mathrm{i} & -4\mathrm{i} & -75\mathrm{i} & 94 \end{bmatrix} \in (\mathrm{Mat}_4)_{\mathrm{sa}}.$$

Thus, in general, elements $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\dots m}\rangle)_{\mathrm{sa}}$ can be viewed as "machines" for manufacturing nN-by-nN hermitian matrices from m given N-by-N hermitian matrices.

The standard orthonormal basis of Herm_N

- Let $\mathrm{Herm}_N \subset \mathrm{Mat}_N$ denote the space of N-by-N hermitian matrices.
- Rationale for notation: fewer subscripts than $(\mathrm{Mat}_N)_{\mathrm{sa}}$.
- Make Herm_N into a real Hilbert space by the trace-pairing $(A, B)_{\mathrm{Herm}_N} = \mathrm{tr}(AB)$.
- Let $e_{ij} \in \mathrm{Mat}_N$ be the elementary matrix $e_{ij}[i', j'] = \delta_{ii'}\delta_{jj'}$.
- Put

  $$\hat{e}_{ij} = \begin{cases} \frac{e_{ij}+e_{ji}}{\sqrt{2}} & \text{if } i < j, \\ e_{ii} & \text{if } i = j, \\ \frac{e_{ij}-e_{ji}}{\mathrm{i}\sqrt{2}} & \text{if } i > j, \end{cases}$$

  thus defining the *standard orthonormal basis* of Herm_N.
- The standard orthonormal basis is the means we routinely use in the sequel to identify Herm_N with $\mathbb{R}^{N^2}$.

GUE matrices

- Recall that random $X \in \mathrm{Herm}_N$ is a *GUE matrix* if the family

$$\{X[i,i]\}_{i=1}^N \cup \{\sqrt{2}\Re X[i,j], \sqrt{2}\Im X[i,j]\}_{1\leq i<j\leq N}$$

  is i.i.d. standard Gaussian.

- Equivalently, a random matrix $X \in \mathrm{Herm}_N$ is a GUE matrix if the random variables $(X, \hat{e}_{ij})$ for $i, j = 1, \ldots, N$ are i.i.d. standard Gaussian.

- In other words, identifying Herm_N with $\mathbb{R}^{N^2}$ by means of the standard orthonormal basis, a GUE matrix is the same thing as a random vector in $\mathbb{R}^{N^2}$ with i.i.d. standard Gaussian entries.

- The latter point of view expedites many calculations below.

Fake GUE matrices

Similarly, a random matrix $X \in \mathrm{Herm}_N$ is a fake GUE matrix if the random variables $(X, \hat{e}_{ij})$ for $i, j = 1, \ldots, N$ are independent fake standard Gaussian random variables. (They do not have to be identically distributed.)

Supports

- In general *support* means " complement of the largest open set on which there is identical vanishing".
- So supports are by definition closed sets.
- In particular, given a probability measure μ on the real line, its support is denoted $\operatorname{supp} \mu$.
- Similarly, given a function $\varphi : \mathbb{R} \to \mathbb{C}$ on the real line, its support is denoted $\operatorname{supp} \varphi$.

Empirical distributions of eigenvalues

Given $X \in \mathrm{Herm}_N$, and writing

$$X = U \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix} U^* \quad (U \in \mathrm{Mat}_N : \text{ unitary}),$$

the *empirical distribution* of the eigenvalues of X is by definition

$$\frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i},$$

which is a probability measure on the real line.

More on empirical distributions

Given a (Borel-measurable) function $\varphi : \mathbb{R} \to \mathbb{C}$ and $X \in \mathrm{Herm}_N$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ and empirical distribution μ, we have by definition

$$\int \varphi d\mu = \frac{1}{N} \sum_{i=1}^{N} \varphi(\lambda_i).$$

Note that $\operatorname{supp} \mu$ is just the set of eigenvalues of X.

The empirical distribution $\mu_f^{(N)}$

Now we get close to formulating HST.

Given $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$,

we let
$$\mu_f^{(N)}$$

be the empirical distribution of the eigenvalues

of the random matrix
$$f\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right) = f\left(\frac{\Xi_1^{(N)}}{\sqrt{N}}, \ldots, \frac{\Xi_m^{(N)}}{\sqrt{N}}\right) \in \mathrm{Herm}_{Nn},$$

where
$$\Xi^{(N)} = (\Xi_1^{(N)}, \ldots, \Xi_m^{(N)})$$

is an m-tuple of independent N-by-N GUE matrices.

Note that $\mu_f^{(N)}$ is actually a random measure.

The measure μ_f

Presently we will work up an operator-theoretic construction of a certain measure μ_f which we will later show to be the almost sure weak limit of $\mu_f^{(N)}$. For now this is just a black box, and we issue an IOU for the definition.

The HST theorem

With the understanding that we owe a definition for μ_f, we can state HST.

Theorem ([Haagerup-Schultz-Thorbjørnsen 2006])

Notation and assumptions are as above. For every $\epsilon > 0$, $\operatorname{supp} \mu_f^{(N)}$ *is contained in the* ϵ*-neighborhood of* $\operatorname{supp} \mu_f$ *for* $N \gg 0$*, almost surely.*

In other words, for every $\epsilon > 0$, there exists a (random) positive integer N_0 such that for every $N \geq N_0$, every eigenvalue of the random matrix

$$f\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right) = f\left(\frac{\Xi_1^{(N)}}{\sqrt{N}}, \ldots, \frac{\Xi_m^{(N)}}{\sqrt{N}}\right) \in \mathrm{Herm}_{Nn}$$

is within distance ϵ of some point of $\operatorname{supp} \mu_f$.

“Fake” HST

We may of course ask whether HST remains true if we keep exactly the same setup EXCEPT for replacing $\Xi^{(N)}$ by an m-tuple of fake N-by-N GUE matrices. This variant of HST we will call *fake* HST just to have a short catchphrase. Fake HST is in principle covered by [Anderson, Ann. Probab., to appear]. We will indicate below a much more palatable proof of this special case.

Remark

The HST theorem was a byproduct of a much larger project in operator theory, one ultimately motivated by the still unsolved problem to distinguish the von Neumann algebras associated to free groups according to the number of generators of the group. This in large part accounts for the hybrid nature of HST and its proof, with one foot in RMT and one foot in operator theory.

The operator theory goals of HST were accomplished already with GUE matrices; it "only" probabilistic curiosity that drives the further investigation of HST-like theorems involving Wigner matrices.

Remark

Suppose, say, that

$$g = \begin{bmatrix} \mathbf{X}_1 + \mathrm{i}\mathbf{X}_2 & \mathbf{X}_3 + \mathrm{i}\mathbf{X}_4 & \mathbf{X}_5 + \mathrm{i}\mathbf{X}_6 \\ \mathbf{X}_7 + \mathrm{i}\mathbf{X}_8 & \mathbf{X}_9 + \mathrm{i}\mathbf{X}_{10} & \mathbf{X}_{11} + \mathrm{i}\mathbf{X}_{12} \end{bmatrix},$$

fix $A \in \mathrm{Herm}_3$ such that $A \geq 0$ and put

$$f = gAg^* \in \mathrm{Mat}_2(\mathbb{C}\langle \mathbf{X}_{1\ldots 12}\rangle)_{\mathrm{sa}}.$$

Then $f\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right)$ is a random sample covariance matrix of the sort to which the theorem of Bai-Silverstein applies. Thus some phenomena of the Bai-Silverstein type are treated in the HST setup, but of course the full Bai-Silverstein result itself is not covered.

Our next goal is...

is to pay off our debt by defining μ_f in suitable fashion. "Suitable fashion" means "in terms of concepts from operator theory and free probability theory."

Background on C^*-probability spaces

In the next several frames we fill in background on Banach algebras, C^*-algebras, spectral theory and states. Note that we always use $[\![\cdot]\!]$ for the norm in a Banach space, Banach algebra or C^*-algebra. (We follow this convention because we write $\|Z\|_p$ for the L^p-norm of a $\mathbb{C}$-valued random variable Z and we want to avoid collisions.)

A good reference for this background (exclusive of the C^*-probability spaces *per se*) is [Murphy].

For the noncommutative probability angle see, e.g., [Voiculescu-Dykema-Nica], [Nica-Speicher], [Anderson-Guionnet-Zeitouni].

Spectra

Let $\mathcal{A}$ be a Banach algebra.

- Given $x \in \mathcal{A}$, as usual the *spectrum* of x relative to $\mathcal{A}$ is defined to be the set of $\lambda \in \mathbb{C}$ such that $x - \lambda 1_{\mathcal{A}}$ fails to be invertible in $\mathcal{A}$.
- We denote the spectrum by $\mathrm{Spec}_{\mathcal{A}}(x)$.
- It is well-known that the spectrum is compact, and by the Banach-Mazur theorem, nonempty.
- The *spectral radius* of $x \in \mathcal{A}$ is by definition $\sup_{\lambda \in \mathrm{Spec}_{\mathcal{A}}(x)} |\lambda|$.
- One has
$$\lim_{n\to\infty} [\![x^n]\!]^{1/n} = \text{spectral radius of } x$$
(Beurling's theorem= "spectral radius theorem")

Spectral mapping

- For $x \in \mathcal{A}$ we have
$$\mathrm{Spec}_{\mathcal{A}}(p(x)) = \{p(\lambda) \mid \lambda \in \mathrm{Spec}_{\mathcal{A}}(x)\}$$
for polynomials $p(\mathbf{X}) \in \mathbb{C}[\mathbf{X}]$.
- If $x \in \mathcal{A}^{\times}$ then
$$\mathrm{Spec}_{\mathcal{A}}(x^{-1}) = \{\lambda^{-1} \mid \lambda \in \mathrm{Spec}_{\mathcal{A}}(x)\}.$$
- The preceding are simple examples of spectral mapping statements.
- Depending on the setup one may be able to use more general functions and still preserve the relationship between functions of operators and functions of spectra.

C^*-algebras (definition)

- A C^*-algebra $\mathcal{A}$ is a Banach $*$-algebra such that $[\![xx^*]\!] = [\![x]\!]^2$ (and hence $[\![x^*]\!] = [\![x]\!]$) for all $x \in \mathcal{A}$.
- We consider only unital C^*-algebras $\mathcal{A}$ in these talks. Necessarily $[\![1_{\mathcal{A}}]\!] = 1$.
- The premier examples of C^*-algebras for us are those of the form $B(H)$ for a Hilbert space H. This includes of course $\mathrm{Mat}_n = B(\mathbb{C}^n)$.
- (Here and below $B(\mathcal{V}, \mathcal{W})$ denotes the space of bounded linear maps from a Banach space $\mathcal{V}$ to a Banach space $\mathcal{W}$ normed by the usual operator norm and we write $B(\mathcal{V}) = B(\mathcal{V}, \mathcal{V})$.)

Given $A \in B(H)$, we have

$$\llbracket A \rrbracket = \sup_{\substack{v \in H \\ \text{s.t. } \llbracket v \rrbracket = 1}} \llbracket Av \rrbracket = \sup_{\substack{v \in H \\ \text{s.t. } \llbracket v \rrbracket = 1}} |(Av, Av)|^{1/2},$$

$$\begin{aligned} \llbracket A \rrbracket &= \sup_{\substack{v \in H \\ \text{s.t. } \llbracket v \rrbracket = 1}} \llbracket Av \rrbracket = \sup_{\substack{v, w \in H \\ \text{s.t. } \llbracket v \rrbracket = \llbracket w \rrbracket = 1}} |(w, Av)| \\ &= \sup_{\substack{v, w \in H \\ \text{s.t. } \llbracket v \rrbracket = \llbracket w \rrbracket = 1}} |(A^* w, v)| = \llbracket A^* \rrbracket, \end{aligned}$$

$$\begin{aligned} \llbracket A \rrbracket &= \sup_{\substack{v, w \in H \\ \text{s.t. } \llbracket v \rrbracket = \llbracket w \rrbracket = 1}} |(Aw, Av)|^{1/2} \\ &= \sup_{\substack{v, w \in H \\ \text{s.t. } \llbracket v \rrbracket = \llbracket w \rrbracket = 1}} |(w, A^* Av)|^{1/2} = \llbracket A^* A \rrbracket^{1/2}. \end{aligned}$$

Commutative C^*-algebras

- The space of $\mathbb{C}$-valued continuous functions on a given compact Hausdorff space,
    - under the supremum norm, and
    - equipped with an involution by the rule $f^*(x) = f(x)^*$

  is a C^*-algebra.
- All commutative C^*-algebras are of this form, as is established by the theory of the Gelfand transform.

C^*-subalgebras and GNS

Let $\mathcal{A}$ be a C^*-algebra.

- We say that a closed subspace $\mathcal{A}_0 \subset \mathcal{A}$ is a C^**-subalgebra* if $\mathcal{A}_0^* = \mathcal{A}_0$, $\mathcal{A}_0\mathcal{A}_0 \subset \mathcal{A}_0$ and $1_{\mathcal{A}_0} = 1_{\mathcal{A}}$.
- Up to isomorphism, $\mathcal{A}$ is a C^*-subalgebra of $B(H)$ for some Hilbert space H, via the Gelfand-Naimark-Segal (GNS) construction.
- (This is only a very rough description of GNS. We supply a better one later.)
- The upshot is that you really do not lose any generality in assuming that a given C^*-algebra is embedded as a closed $*$-subalgebra of $B(H)$ for some Hilbert space H.

Let $\mathcal{A}$ be a C^*-algebra. Given $x \in \mathcal{A}$, we write simply $\mathrm{Spec}(x)$, omitting reference to $\mathcal{A}$, because for inclusions of C^*-algebras the spectrum does not change.

This is in contrast to the situation for general Banach algebras, where the spectrum can change under inclusion.

Example

- Let $\mathcal{A}$ be the C^*-algebra of continuous $\mathbb{C}$-valued functions defined on the circle $\{|z| = 1\}$.
- Let $\mathcal{A}_0 \subset \mathcal{A}$ be the closed subalgebra (NOT a C^*-subalgebra) consisting of boundary values of continuous functions on the closed disc $\{|z| \leq 1\}$ analytic in the open disc $\{|z| < 1\}$.
- Then the identity map of the circle to itself is invertible in $\mathcal{A}$ but not invertible in $\mathcal{A}_0$.

Let $\mathcal{A}$ be a C^*-algebra.

- The spectrum of $x \in \mathcal{A}_{\mathrm{sa}}$ is contained in $\mathbb{R}$.
- A *normal* operator $x \in \mathcal{A}$ is one for which $xx^* = x^*x$.

Spectra in C^*-algebras (continued)

Lemma (Norms of normal elements of C^*-algebras)

Let $\mathcal{A}$ be a C^*-algebra. For a normal operator $x \in \mathcal{A}$ the spectral radius of x equals the norm $[\![x]\!]$.

In particular, since $[\![a]\!]$ is the spectral radius of aa^*, there is only one way to put a norm on a $*$-algebra to make it into a C^*-algebra.

Proof Put

$$x_n = x^{2^n} \text{ and } y_n = (xx^*)^{2^n}.$$

Then

$$[\![x_n]\!]^4 = [\![x_n x_n^*]\!]^2 = [\![y_n]\!]^2 = \left[\!\!\left[y_n^2\right]\!\!\right] = [\![y_{n+1}]\!] = \left[\!\!\left[x_{n+1}x_{n+1}^*\right]\!\!\right] = [\![x_{n+1}]\!]^2$$

since x and x^* commute.The result now follows by the spectral radius formula. $\square$

Let $\mathcal{A}$ be a C^*-algebra.

- We say that $x \in \mathcal{A}$ is *positive* and write $x \geq 0$ if x is self-adjoint and $\mathrm{Spec}(x) \subset [0, \infty)$.
- We also write $x > 0$ if $x \geq 0$ and $x \in \mathcal{A}^\times$ (but we do not know of any word to use for this notion).
- If $\mathcal{A} = \mathrm{Mat}_n$ then $x \in \mathcal{A}$ satisfies $x \geq 0$ if and only if $x \in \mathrm{Herm}_N$ and all eigenvalues of x are nonnegative.
- If $\mathcal{A}$ is the C^*-algebra of continuous $\mathbb{C}$-valued functions on a given compact Hausdorff space, then $f \in \mathcal{A}$ satisfies $f \geq 0$ if and only f is a nonnegative function in the naive sense.

More on positivity

Let $\mathcal{A}$ be a C^*-algebra.

- It is well-known that the set of positive elements in $\mathcal{A}$ forms a convex cone, i.e., given $s, t \in \mathbb{R}$ and $x, y \in \mathcal{A}$ such that $s \geq 0$, $t \geq 0$, $x \geq 0$ and $y \geq 0$, again $sx + ty \geq 0$.
- For all $x, y \in \mathcal{A}$ such that $y \geq 0$, again $xyx^* \geq 0$.
- Each positive $x \in \mathcal{A}$ has a unique positive square root $x^{1/2} \in \mathcal{A}$.

Functional calculus

- Let $\mathcal{A}$ be a C^*-algebra and fix $A \in \mathcal{A}_{\mathrm{sa}}$.
- Given continuous $\varphi : \mathrm{Spec}(A) \to \mathbb{C}$, one defines a normal operator $\varphi(A) \in \mathcal{A}$ by the rule

$$\lim_{n\to\infty} \varphi_n(A) = \varphi(A)$$

  where $\{\varphi_n\}$ is any sequence of polynomial functions converging uniformly to φ on the spectrum $\mathrm{Spec}(A)$.
- Here one evaluates $\varphi_n(A)$ by plugging A into the corresponding polynomial, and the limit is taken with respect to the operator norm $[\![\cdot]\!]$.
- We refer to this standard construction by the catchphrase *functional calculus*.

States

Again let $\mathcal{A}$ be a C^*-algebra.

- A bounded linear functional $\phi : \mathcal{A} \to \mathbb{C}$ of norm 1 satisfying $\phi(1_{\mathcal{A}}) = 1$ is called a *state* of $\mathcal{A}$.
- (Equivalently: $\varphi(A) \geq 0$ for all $A \geq 0$ and $\phi(1_{\mathcal{A}}) = 1$.)
- If $\mathcal{A} = \mathrm{Mat}_n$ then the states $\phi : \mathrm{Mat}_n \to \mathbb{C}$ are the linear functionals of the form $\phi(A) = \mathrm{tr}(AB)$ where $B \geq 0$ and $\mathrm{tr}B = 1$.
  - For example both $A \mapsto \frac{1}{n}\mathrm{tr}(A)$ and $A \mapsto A[1,1]$ are states.
- If $\mathcal{A}$ is the space of $\mathbb{C}$-valued continuous functions on a given compact Hausdorff space, then the states are the linear functionals represented by probability measures on the space.

GNS in finer detail

As above let $\mathcal{A}$ be a C^*-algebra.

- The GNS construction, in finer detail, says that that for each state $\phi : \mathcal{A} \to \mathbb{C}$ there is a way of identifying $\mathcal{A}$ with a C^*-subalgebra of $B(H)$ for some Hilbert space H and a distinguished unit vector $1_H \in H$ such that
$$\phi(A) = (1_H, A1_H)$$
for all $A \in \mathcal{A}$.
- (Here and below Hilbert space inner products are always taken antilinear on the left and linear on the right.)
- For example, if $\mathcal{A} = \mathrm{Mat}_N$ and the given state $\phi : \mathrm{Mat}_N \to \mathbb{C}$ is of the form $\varphi(A) = \mathrm{tr}(AB)$ for some $B \in \mathrm{Mat}_N$ such that $B \geq 0$ and $\mathrm{tr}(B) = 1$, one can take $\mathcal{H} = \mathrm{Mat}_N$ under the Hilbert space inner product $(X, Y) = \mathrm{tr}(X^* Y)$ with $\mathcal{A}$ acting on the left of $\mathcal{H}$ in the obvious way and $1_{\mathcal{H}} = B^{1/2}$.

C^*-probability spaces

- A pair $(\mathcal{A}, \phi)$ with ϕ a state of $\mathcal{A}$ will be called a *C^*-probability space*.

- Typically below we take

$$(\mathcal{A}, \phi) = (B(H), A \mapsto (1_H, A1_H))$$

  for a Hilbert space H with distinguished unit vector $1_H \in H$.

- By GNS there is essentially no loss of generality in doing so.

Laws of single operators

- Let $(\mathcal{A}, \phi)$ be a C^*-probability space.
- Let $A \in \mathcal{A}_{\mathrm{sa}}$ be a self-adjoint operator, which in the present context we view as a *noncommutative random variable* belonging to $(\mathcal{A}, \phi)$.
- The Riesz representation theorem yields a unique probability measure μ_A on $\mathbb{R}$ supported in $\mathrm{Spec}(A)$ such that

$$\int \psi d\mu_A = \phi\left(\underbrace{(\psi|_{\mathrm{Spec}(A)})(A)}_{\text{defined by functional calculus in } \mathcal{A}}\right)$$

  for all continuous $\psi : \mathbb{R} \to \mathbb{C}$.
- We call μ_A the *law* of A.

Laws of single matrices

- If $(\mathcal{A}, \phi) = (\mathrm{Mat}_n, \frac{1}{n}\mathrm{tr})$ and $A \in \mathrm{Herm}_n$ then the law μ_A is none other than the empirical distribution of eigenvalues of A.
- If $(\mathcal{A}, \phi) = (\mathrm{Mat}_n, \text{evaluation of upper left entry})$ and $A \in \mathrm{Herm}_n$, then the law μ_A is the unique probability measure on the real line with moments $\int x^k \mu_A(dx) = A^k[1,1]$.
- Looking more closely at the example immediately above, and writing $AU = UD$ where U is unitary and D is diagonal, we have $\int x^k \mu_A(dx) = \sum_{j=1}^n U[1,j]^2 D[j,j]^k$ and thus $\mu_A(\{D[j,j]\}) = |U[1,j]|^2$ for $j = 1, \ldots, n$.
  - In other words, the "table" describing the probability distribution μ_A is the same as the list of absolute values squared of the top row of U.

Offbeat example: Gaussian quadrature

Let

$$(\mathcal{A}, \phi) = (\mathrm{Mat}_{n+1}, A \mapsto A[1,1])$$

and

$$A = \left[\begin{array}{cccccc} & \sqrt{1} & & & & \\ \sqrt{1} & & \sqrt{2} & & & \\ & \sqrt{2} & & \sqrt{3} & & \\ & & \sqrt{3} & & \ddots & \\ & & & \ddots & & \sqrt{n} \\ & & & & \sqrt{n} & \end{array}\right] \in \mathrm{Herm}_{n+1} = \mathcal{A}_{\mathrm{sa}}.$$

(The matrix above extended infinitely in evident fashion represents the three-term recurrence for suitably normalized Hermite polynomials.)

By definition of the law μ_A, the moments of μ_A are the numbers $A^k[1,1]$ for positive integers k.

The example considered here is cooked up so that the first $2n+1$ moments of μ_A are the same as those of a standard normal random variable, i.e.,

$$A^k[1,1] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2}\, dx \quad \text{for } k = 1, \ldots, 2n+1,$$

e.g.,

$$\begin{bmatrix} & \sqrt{1} & & \\ \sqrt{1} & & \sqrt{2} & \\ & \sqrt{2} & & \sqrt{3} \\ & & \sqrt{3} & \end{bmatrix}^4 = \begin{bmatrix} 3.00 & 0 & 8.48 & 0 \\ 0 & 15.00 & 0 & 14.69 \\ 8.48 & 0 & 27.00 & 0 \\ 0 & 14.69 & 0 & 15.00 \end{bmatrix}$$

and indeed, as claimed, the upper left entry is 3.

To find an explicit random variable X with the law μ_A one computes as follows.

For suitable orthogonal V and diagonal D we have

$$AV = VD, \quad D = \mathrm{diag}(x_1, \ldots, x_{n+1}), \quad x_1 < \cdots < x_{n+1}.$$

The random variable X with the same law as μ_A is then described by the rule

$$\Pr(X = x_j) = V[1, j]^2 \ \text{ for } j = 1, \ldots, n+1,$$

e.g., for $n = 3$,

$$\begin{array}{rcccc} x: & -2.3344 & -0.7420 & 0.7420 & 2.3344 \\ \Pr(X = x): & 0.0459 & 0.4541 & 0.4541 & 0.0459 \end{array}$$

Check:

$$\begin{aligned} & (-2.3344)^4(0.0459) + (-0.7420)^4(0.4541) \\ & +(0.7420)^4(0.4541) + (2.3344)^4(0.0459) \\ = \;& 3. \end{aligned}$$

This example is just a look at the classical topic of Gaussian quadrature through the lens of noncommutative probability.

Example: The semicircular law

Let $\mathcal{H}$ be a Hilbert space equipped with an orthonormal basis and "ground state"

$$\{v(i)\}_{i=0}^{\infty}, \quad 1_{\mathcal{H}} = v(0).$$

Let $\theta \in B(\mathcal{H})$ be the unilateral shift operator

$$\theta v(i) = v(i+1)$$

the adjoint θ^* of which satisfies

$$\theta v(i) = \mathbf{1}_{i>0} v(i-1).$$

Let

$$\Xi = \theta + \theta^* \in B(\mathcal{H})_{\mathrm{sa}}.$$

Note that

$$\theta^*\theta = 1_{B(\mathcal{H})}, \;\; \theta\theta^* = 1_{B(\mathcal{H})} - p_{\mathcal{H}}$$

where $p_{\mathcal{H}}$ is orthogonal projection to the span of the ground state.

Example: The semicircular law (concluded)

Then

$$(1_{\mathcal{H}}, \Xi^k 1_{\mathcal{H}})$$

is the number of Bernoulli walks of k steps starting at the origin which stay nonnegative and return to the origin.

For k odd there are no such walks and for k even the number of such is the *Catalan number*

$$\frac{1}{k/2+1}\begin{pmatrix} k \\ k/2 \end{pmatrix}.$$

Thus the law of Ξ has no choice but to be the semicircle law, i.e., the probability measure on the real line with p.d.f.

$$\frac{1}{2\pi}\mathbf{1}_{|x|<2}\sqrt{4-x^2}.$$

Laws of single matrices of operators

- Let $(\mathcal{A}, \phi)$ be a C^*-probability space and let $A \in \mathrm{Mat}_n(\mathcal{A})_{\mathrm{sa}}$ be a self-adjoint operator.

- Note that $(\mathrm{Mat}_n(\mathcal{A}), A \mapsto \frac{1}{n}\sum_{i=1}^n \phi(A[i,i]))$ is a C^*-probability space.

- Thus, just repeating the construction of the law of a noncommutative random variable, we get a unique probability measure μ_A on $\mathbb{R}$ supported in $\mathrm{Spec}(A)$ such that

$$\int \psi d\mu_A = \frac{1}{n}\sum_{i=1}^n \phi\left(\underbrace{(\psi|_{\mathrm{Spec}(A)})(A)}_{\substack{\text{defined by functional} \\ \text{calculus in } \mathrm{Mat}_n(\mathcal{A})}}[i,i]\right)$$

for all continuous $\psi : \mathbb{R} \to \mathbb{C}$.

Laws of single matrices of operators (concluded)

- In this more general setup we still call μ_A the *law* of $A \in \mathrm{Mat}_n(\mathcal{A})_{\mathrm{sa}}$ and
- we regard A as a noncommutative random variable defined on the noncommutative probability space $(\mathcal{A}, \phi)$.

The empirical distribution of eigenvalues of a matrix

$$A \in \mathrm{Herm}_{Nn}$$

is exactly the same thing as the law of A construed via "Kronecker fattening" as a noncommutative random variable defined on the C^*-probability space $(\mathrm{Mat}_N, \frac{1}{N}\mathrm{tr})$.

Matrix spaces over C^*-algebras

Again let $\mathcal{A}$ be a C^*-algebra.

- The $*$-algebra $\mathrm{Mat}_n(\mathcal{A})$ has of course at most one norm making it into a C^*-algebra, but it also has at least one, namely any one gotten by an embedding $\mathrm{Mat}_n(\mathcal{A}) \subset B(H^n)$ provided by GNS.

- **We always use that norm.**

- More generally, $\mathrm{Mat}_{k\times\ell}(\mathcal{A})$ is normed by the rule $[\![A]\!]^2 = [\![AA^*]\!]$. In particular, this rule puts the usual largest-singular-value norm on $\mathrm{Mat}_{k\times\ell}$.

- We have $[\![AB]\!] \leq [\![A]\!][\![B]\!]$ for matrices A and B with entries in $\mathcal{A}$ for which the product AB is defined.

- For $A \in \mathrm{Mat}_{k\times\ell}(\mathcal{A})$ we have $\max_{i,j} [\![A[i,j]]\!] \leq [\![A]\!] \leq \sum_k \max_{i-j=k} [\![A[i,j]]\!]$.

Our next goal is. . .

. . . to briefly review Stieltjes transforms and to bring that familiar notion into contact with the operator-theoretic point of view being emphasized here.

Stieltjes transforms

In general, given a probability measure μ on the real line, recall that the *Stieltjes transform* is defined by the formula

$$S_\mu(z) = \int \frac{\mu(dt)}{t-z} \text{ for } z \in \mathbb{C} \setminus \operatorname{supp} \mu.$$

The facts

$$\Im z > 0 \Rightarrow \Im S_\mu(z) > 0, \;\; S_\mu(z^*) = S_\mu(z)^* \text{ and } |S_\mu(z)\Im z| \leq 1$$

are taken for granted in the sequel.

We often restrict attention to the upper half-plane when studying Stieltjes transforms. We use the notation

$$\mathfrak{h} = \{z \in \mathbb{C} \mid \Im z > 0\}$$

for the classical upper half-plane.

Stieltjes transforms of empirical distributions

For $A \in \mathrm{Herm}_N$, the empirical distribution μ_A of eigenvalues of A has the Stieltjes transform

$$S_{\mu_A}(z) = \frac{1}{N}\mathrm{tr}(A - z\mathbf{I}_N)^{-1}.$$

This is utterly familiar. This relationship generalizes naturally in the context of C^*-probability spaces as we explain in the next frame.

Stieltjes transforms of laws of noncommutative random variables

Let $(\mathcal{A}, \phi)$ be a C^*-probability space. Let $A \in \mathrm{Mat}_n(\mathcal{A})_{\mathrm{sa}}$ be a noncommutative random variable and let μ_A be its law. Then we have

$$S_{\mu_A}(z) = \frac{1}{n} \sum_{i=1}^{n} \phi((A - z1_{\mathcal{A}})^{-1}[i,i])$$

for $z \in \mathbb{C} \setminus \mathrm{Spec}(A)$ by simply plugging the appropriate test-function into the definition of the law μ_A. This is a fundamental tool in the sequel.

Semicircular example

$$\text{upper left corner of } \begin{bmatrix} -z & 1 & & & \\ 1 & -z & 1 & & \\ & 1 & -z & 1 & \\ & & 1 & \ddots & \ddots \\ & & & \ddots & \end{bmatrix}^{-1}$$

$$= \quad \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4-t^2}}{t-z}\, dt.$$

Let

$$A = \begin{bmatrix} & \sqrt{1} & & & & \\ \sqrt{1} & & \sqrt{2} & & & \\ & \sqrt{2} & & \sqrt{3} & & \\ & & \sqrt{3} & & \ddots & \\ & & & \ddots & & \sqrt{n} \\ & & & & \sqrt{n} & \end{bmatrix} \in \mathrm{Herm}_{n+1}.$$

Then

$$(A - z\mathbf{I}_{n+1})^{-1}[1,1]$$

is the Stieltjes transform of a probability measure on the real line the first $2n+1$ moments of which are the same as those of a standard normal random variable.

Boltzmann-Fock space and the operator-theoretic characterization of μ_f

We turn now to the task of characterizing the probability measure μ_f figuring in HST by means of a certain explicitly constructed C^*-probability space and a certain explicitly constructed operator, thus paying off our IOU. Along the way to defining μ_f we explain why (but do not yet prove that) μ_f turns out to be the almost sure weak limit of $\mu_f^{(N)}$.

Boltzmann-Fock space setup

- Let $\mathcal{H}$ be a Hilbert space equipped with an orthonormal basis

$$\{v(i)\}_{i=0}^{\infty}$$

  indexed by the nonnegative integers.

- Put $1_{\mathcal{H}} = v(0)$.

- We equip $B(\mathcal{H})$ with the state

$$(A \mapsto (1_{\mathcal{H}}, A1_{\mathcal{H}}))$$

  (to which we do not bother to give a name) thus making it into a noncommutative probability space.

- Hereafter we can speak of laws of operators $A \in \mathrm{Mat}_n(B(\mathcal{H}))_{\mathrm{sa}}$.

For $\ell = 1, \ldots, m$ we define

$$\theta_\ell \in B(\mathcal{H})$$

by the awful formula

$$\theta_\ell v(i) = \sum_{j=0}^{\infty} \mathbf{1}_{\frac{m^j - 1}{m-1} \leq i < \frac{m^{j+1}-1}{m-1}} v(i + m^j \ell).$$

By construction θ_ℓ is a partial isometry. We call θ_ℓ a *raising operator*. The adjoint θ_ℓ^* is called a *lowering operator*.

Modified place-value notation again

The definition of θ_ℓ is less awful than it seems. One can show easily enough that the family

$$\{\theta_{\ell_1} \cdots \theta_{\ell_k} v(0)\}$$

indexed by all finite sequences $\ell_1, \ldots, \ell_k \in \{1, \ldots, m\}$ including the empty sequence is simply a relabeling of the canonical orthonormal basis $\{v(i)\}_{i=0}^{\infty}$.

The noncommutative random variables $\Xi_1, \ldots, \Xi_m$

For $\ell = 1, \ldots, m$ we put

$$\Xi_\ell = \theta_\ell + \theta_\ell^* \in B(\mathcal{H})$$

which is self-adjoint and satisfies $[\![\Xi_\ell]\!] = 2$.

The operators Ξ_ℓ can also be characterized by the formula

$$\sum_{\ell=1}^{m} x_\ell \Xi_\ell = \begin{bmatrix} & x & & & & \\ x^* & & x \otimes \mathbf{I}_m & & & \\ & x^* \otimes \mathbf{I}_m & & x \otimes \mathbf{I}_{m^2} & & \\ & & x^* \otimes \mathbf{I}_{m^2} & & \ddots & \\ & & & \ddots & & \end{bmatrix}$$

which holds for all row vectors $x = (x_1, \cdots, x_m)$ with real entries.

Here and below we identify $\mathcal{H}$ with a space of column vectors with entries indexed by the nonnegative integers and we identify $B(\mathcal{H})$ with a space of matrices having rows and columns indexed by the nonnegative integers.

Self-similarity

$$\sum_{\ell=1}^{m} x_\ell \Xi_\ell = \begin{bmatrix} & x & & & \\ x^* & & x \otimes \mathbf{I}_m & & \\ & x^* \otimes \mathbf{I}_m & & x \otimes \mathbf{I}_{m^2} & \\ & & x^* \otimes \mathbf{I}_{m^2} & & \ddots \\ & & & \ddots & \end{bmatrix}.$$

Thus, with $\tilde{x} = [x \; 0 \; \ldots]$, we have

$$\sum_{\ell=1}^{m} x_\ell \Xi_\ell = \begin{bmatrix} & \tilde{x} \\ \tilde{x}^* & \left(\sum_{\ell=1}^{m} x_\ell \Xi_\ell\right) \otimes \mathbf{I}_m \end{bmatrix}.$$

The joint distribution of the family $\{\Xi_\ell\}_{\ell=1}^m$

- Each operator Ξ_ℓ has the semicircle law and more generally, for $x_1, \ldots, x_n$ real such that $\sum x_i^2 = 1$ the noncommutative random variable $\sum_{\ell=1}^m x_\ell \Xi_\ell$ has the semicircle law.

- The family
$$\{\Xi_\ell\}_{\ell=1}^m$$
has in the parlance of free probability a *free semicircular joint distribution.*

- More precisely, for any sequence $\ell_1, \ldots, \ell_k \in \{1, \ldots, m\}$ the matrix element
$$(1_{\mathcal{H}}, \Xi_{\ell_1} \cdots \Xi_{\ell_k} 1_{\mathcal{H}})$$
is a nonnegative integer having several equivalent (and aesthetically appealing) combinatorial descriptions, one of which we describe in the next frame.

Digression: combinatorial interpretation of $(1_{\mathcal{H}}, \Xi_{\ell_1} \cdots \Xi_{\ell_k} 1_{\mathcal{H}})$

- Given a permutation $\sigma \in S_k$, let $c(\sigma)$ denote the number of σ-cycles into which $\{1, \ldots, k\}$ is decomposed.
- Let $\eta = (12 \cdots k) \in S_k$.
- Then
$$(1_{\mathcal{H}}, \Xi_{\ell_1} \cdots \Xi_{\ell_k} 1_{\mathcal{H}})$$
is the number of fixed-point-free elements $\sigma \in S_k$ such that
$$\sigma^2 = 1, \quad c(\sigma\eta) = k/2 + 1$$
and
$$\ell_{\sigma(1)} = \ell_1, \;\; \ell_{\sigma(2)} = \ell_2, \;\; \ldots, \;\; \ell_{\sigma(k)} = \ell_k.$$

The preceding characterization of $(1_{\mathcal{H}}, \Xi_{\ell_1} \cdots \Xi_{\ell_k} 1_{\mathcal{H}})$ will not be used in the sequel.

The law μ_f

Now we pay off the IOU.

- Let
$$\Xi = (\Xi_1, \ldots, \Xi_m) \in B(\mathcal{H})^m_{\mathrm{sa}}.$$
- For $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$, consider the operator
$$f(\Xi) = f(\Xi_1, \ldots, \Xi_m) \in \mathrm{Mat}_n(B(\mathcal{H}))_{\mathrm{sa}}.$$
- Let
$$\mu_f$$
be the law of $f(\Xi)$.

The HST theorem (redux)

Let us look again at HST to see if it makes any more sense.

Theorem (HST)

For every $\epsilon > 0$, the support of $\mu_f^{(N)}$ is contained in the ϵ-neighborhood of support of μ_f for $N \gg 0$, almost surely.

Nagging doubts such as the possibility $\operatorname{supp} \mu_f = \mathbb{R}$ are now dispelled, since $\operatorname{supp} \mu_f$ is clearly compact.

We emphasize that since almost sure weak convergence does not care about a few stray eigenvalues, HST really has some content.

In the next several frames we recall on related recent results, variants and generalizations.

Lecture 3 (June 20, 2012)

The spectrum and the support of the law of a noncommutative random variable

Let $(\mathcal{A}, \phi)$ be a noncommutative probability space.

Let ϕ be *faithful*, i.e., suppose that for all $A \in \mathcal{A}$ if $A \geq 0$ and $A \neq 0$, then $\phi(A) > 0$.

Proposition (Support equals spectrum in presence of faithfulness)

Notation and assumptions as are above. For every noncommutative random variable $A \in \mathrm{Mat}_n(\mathcal{A})_{\mathrm{sa}}$ the support of the law of A and the spectrum of A are equal.

We present a proof in the next couple of frames.

Proof of "support equals spectrum in presence of faithfulness"

Lemma

The linear functional $\phi_n = (A \mapsto \frac{1}{n}\sum_{i=1}^n \phi(A[i,i]))$ *on* $\mathrm{Mat}_n(\mathcal{A})$ *is a faithful state.*

Proof Fix $A \in \mathrm{Mat}_n(\mathcal{A})$ such that $A \geq 0$ and $A \neq 0$.

Then

$$\phi_n(A) = \frac{1}{n}\sum_{i=1}^n \phi(A^{1/2}[i,j]A^{1/2}[j,i]) \geq 0$$

and moreover at least one term on the right does not vanish by faithfulness of ϕ. □

Proof of "support equals spectrum in presence of faithfulness" (continued)

Lemma

Fix $A \in \mathcal{A}_{\mathrm{sa}}$ and let $\mathcal{A}_0 \subset \mathcal{A}$ be the C^*-subalgebra generated by A. Then every $B \in \mathcal{A}_0$ is of the form $B = f(A)$ for some continuous function $f : \mathrm{Spec}(A) \to \mathbb{C}$ and moreover $[\![B]\!] = \sup_{x \in \mathrm{Spec}(A)} |f(x)|$. (In other words, the functional calculus is an isomorphism from the C^*-algebra of continuous $\mathbb{C}$-valued functions on $\mathrm{Spec}(A)$ to $\mathcal{A}_0$.)

Proof This is a basic consequence of the theory of the Gelfand transform. $\square$

Proof of "support equals spectrum in presence of faithfulness" (concluded)

The preceding two lemmas granted, we may assume that $\mathcal{A}$ is the space of continuous $\mathbb{C}$-valued functions defined on on compact subset $K \subset \mathbb{R}$ and that A is the identity map $K \to K$. Then ϕ is represented by a probability measure on $\mathrm{Spec}(A)$. Faithfulness of ϕ implies via Urysohn's lemma that $\operatorname{supp} \mu = \mathrm{Spec}(A)$. $\square$

Theorem (Voiculescu)

For each $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\dots m}\rangle)_{\mathrm{sa}}$ the empirical distribution $\mu_f^{(N)}$ "converges momentwise" to μ_f.

The scare-quotes here are because we are actually talking about a simpler kind of convergence, namely:

for any finite sequence $\ell_1, \dots, \ell_k \in \{1, \dots, m\}$,

$$\lim_{N\to\infty} \frac{1}{N}\mathbf{E}\mathrm{tr}\left(\frac{\Xi_{\ell_1}^{(N)}}{\sqrt{N}} \cdots \frac{\Xi_{\ell_k}^{(N)}}{\sqrt{N}}\right) = (1_{\mathcal{H}}, \Xi_{\ell_1} \cdots \Xi_{i_k} 1_{\mathcal{H}}).$$

The latter statement is the foundational example for free probability theory. It is a far-reaching generalization of Wigner's semicircle law.

Amplification of Voiculescu's theorem

More probabilistic versions of Voiculescu's result can be proved without a great deal of trouble. In particular, in the setup for HST one can prove the following statement.

Theorem (Amplification of Voiculescu's theorem)

$\mu_f^{(N)}$ *converges weakly to* μ_f*, almost surely.*

The paper [Male 2010] contains a proof of this, and of a more general result. The paper [Meckes-Szarek 2011] gives an interesting treatment of concentration phenomena related to this result.

Proposition (Folklore about faithfulness)

The restriction of the state

$$(A \mapsto (1_{\mathcal{H}}, A1_{\mathcal{H}})) : B(\mathcal{H}) \to \mathbb{C}$$

to the C^*-subalgebra generated by the free semicircular family $\{\Xi_i\}_{i=1}^m$ is faithful and thus for every $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ we have $\mathrm{supp}\,\mu_f = \mathrm{Spec}(f(\Xi))$.

I thank K. Dykema for explaining this important point to me. It is quite elementary but hard to pin down in the literature.

This is advantageous because the spectrum of $f(\Xi)$ is a more tractable object to study than the support of μ_f.

We give the following proof as an excuse simply to better acquaint the audience with the structure of Boltzmann-Fock space.

Proof

Put

$$w(\ell_1 \cdots \ell_k) = \theta_{\ell_1} \cdots \theta_{\ell_k} v(0)$$

for all finite sequences $\ell_1 \cdots \ell_k$ in $\{1, \ldots, m\}$, including the empty sequence.

Then, as previously noted, $\{w(\ell_1 \cdots \ell_k)\}$ is just a relabeling of the canonical orthonormal basis $\{v(i)\}$.

For example,

$$v(0) = w(\emptyset),\ w(1) = v(1), \ldots, w(m) = v(m), w(11) = v(m+1), \ldots,$$

Proof (continued)

Thus we have

$$\theta_\ell w(\ell_1 \cdots \ell_k) = w(\ell \ell_1 \cdots \ell_k), \quad \theta_\ell^* w(\ell_1 \cdots \ell_k) = \mathbf{1}_{\ell_1 = \ell} w(\ell_2 \cdots \ell_k).$$

Recall that

$$\Xi_\ell = \theta_\ell + \theta_\ell^*.$$

Analogously, we now define $\hat{\theta}_\ell, \hat{\Xi}_\ell \in B(\mathcal{H})$ "on the right side" by

$$\hat{\theta}_\ell w(\ell_1 \cdots \ell_k) = w(\ell_1 \cdots \ell_k \ell), \quad \hat{\theta}_\ell^* w(\ell_1 \cdots \ell_k) = \mathbf{1}_{\ell = \ell_k} w(\ell_1 \cdots \ell_{k-1})$$

and we put

$$\hat{\Xi}_\ell = \hat{\theta}_\ell + \hat{\theta}_\ell^*.$$

Proof (continued)

We have commutation relations

$$\theta_i \hat{\theta}_j = \hat{\theta}_j \theta_i, \ \ \theta_i^* \hat{\theta}_j^* = \hat{\theta}_j^* \theta_i^*,$$

$$\theta_i \hat{\theta}_j^* - \hat{\theta}_j^* \theta_i = -\delta_{ij} p_{\mathcal{H}}, \quad \theta_i^* \hat{\theta}_j - \hat{\theta}_j \theta_i^* = \delta_{ij} p_{\mathcal{H}},$$

where $p_{\mathcal{H}}$ denotes orthogonal projection to the groundstate $1_{\mathcal{H}} = v(0)$

and thus

$$\Xi_i \hat{\Xi}_j = \hat{\Xi}_j \Xi_i.$$

Proof (continued)

Let $\mathcal{A} \subset B(\mathcal{H})$ (resp., $\hat{\mathcal{A}} \subset B(\mathcal{H})$) be the C^*-subalgebra generated by the Ξ_i (resp., by the $\hat{\Xi}_i$).

Clearly $\mathcal{A}$ and $\hat{\mathcal{A}}$ commute with each other.

Furthermore, it is not difficult to verify that $\hat{\mathcal{A}} 1_{\mathcal{H}}$ is dense in $\mathcal{H}$.

Now let $A \in \mathcal{A}$ satisfy $A \geq 0$ and $A \neq 0$.

Clearly there exists $h \in \mathcal{H}$ such that $(h, A^{1/2}h) > 0$.

Since $\hat{\mathcal{A}}1_{\mathcal{H}}$ is dense in $\mathcal{H}$, we may assume that $h = \hat{A}1_{\mathcal{H}}$ for some $\hat{A} \in \hat{\mathcal{A}}$.

Then temporarily writing $\phi = (A \mapsto (1_{\mathcal{H}}, A1_{\mathcal{H}}))$ to give the state of $B(\mathcal{H})$ a name, we have

$$0 < (\hat{A}1_{\mathcal{H}}, A^{1/2}\hat{A}1_{\mathcal{H}}) = (1_{\mathcal{H}}, \hat{A}^*A^{1/2}\hat{A}1_{\mathcal{H}}) = \phi(\hat{A}^*A^{1/2}\hat{A}) = \phi(\hat{A}^*\hat{A}A^{1/2}).$$

Proof (concluded)

Making further use of the hypothesis that operators in $\mathcal{A}$ commute with operators in $\hat{\mathcal{A}}$, we have

$$0 \leq \phi((\sqrt{t}\hat{A}^*\hat{A}-A^{1/2}/\sqrt{t})^2) = t\phi((\hat{A}^*\hat{A})^2)+\phi(A)/t-2\phi(\hat{A}^*\hat{A}A^{1/2})$$

for $t > 0$.

The last inequality forces $\phi(A) > 0$.

Thus $\phi|_{\mathcal{A}}$ is indeed faithful. $\square$

The precursor to HST was the following result, which we are now in a position to state.

Theorem ([Haagerup-Thorbjørnsen 2005])

Notation and assumptions are as for HST. For every $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ *(not necessarily self-adjoint)*

$$\lim_{N\to\infty} \left[\left[f\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right)\right]\right] = [\![f(\Xi)]\!] \ \ a.s.$$

Recovery of HT from HST and amplification of Voiculescu's theorem After replacing f by ff^* we may assume that f is self-adjoint. The upper bound follows from HST and $\mathrm{supp}(f(\Xi)) = \mathrm{supp}(\mu_f)$. The lower bound follows from the amplified version of Voiculescu's theorem. □

Results of [Schultz 2005]

In the cited paper Schultz proves the analogue of HT with GUE replaced by GOE or GSE. The new difficulty emerging in the proof is that there are correction terms which must be dealt with.

Later when more technical details of the proof of HST have been discussed, we can say more precisely what this new difficulty is.

Results of [Capitaine-Donati-Martin 2007]

In the cited paper Capitaine and Donati-Martin prove the analogue of HT with the independent GUE matrices replaced by independent matrices of the following class.

Fix a random variable Z with a symmetric distribution satisfying a Poincaré inequality, i.e., $\mathbf{Var}(f(Z)) \leq c\mathbf{E}|f'(Z)|^2$ for all nice enough functions $f : \mathbb{R} \to \mathbb{C}$, for a constant c depending only on Z. Let μ be the law of Z. Consider random matrices $X \in \mathrm{Herm}_N$ such that the family of random variables $(X, \hat{e}_{ij})$ is i.i.d., each with law μ.

The algebra no longer works out quite so nicely so a new tool is needed. Capitaine and Donati-Martin use a certain form of "integration by parts" introduced in the influential paper [Khorunzhy-Khoruzhenko-Pastur]. We will discuss the latter circle of ideas at a later point in the course.

Results of [Male 2010]

A natural question raised by HST is to consider what happens for polynomials in independent GUE matrices and some other independent matrices where the other matrices already are known to have a convergent joint distribution in the sense of noncommutative probability. This was studied in [Male 2010], and the exact analogue of HST was obtained in this more general setting, as well as the exact analogue of the amplification of Voiculescu's theorem. The exact formulation Male's result is too difficult to attempt here since it requires more background on free probability than we have supplied. It remains an open question to generalize Male's results by replacing GUE matrices by Wigner matrices. Doing so, one would to a large extent succeed in recovering and generalizing results of Bai-Silverstein type.

Results of [Collins-Male 2011]

A natural question raised by Male's result is to consider what happens if one replaces the GUE matrices by unitary matrices. This was studied in [Collins-Male 2011] and the exact analogue of HT was recovered in this case. The method of proof is very clever, efficiently reducing the proof to the previous results in [Male 2010].

In [Anderson, Ann. Probab., to appear] the GUE matrices $\Xi_\ell^{(N)}$ in HST are generalized to the following collection of matrices. Let

$$\{\{\xi_\ell(i,j)\}_{1\le i\le j\le m}\}_{\ell=1}^\infty$$

be an array of independent $\mathbb{C}$-valued random variables with finite absolute fourth moments and zero means such that the law of $\xi_\ell(i,j)$ depends only on ℓ and $\mathbf{1}_{i<j}$. Assume furthermore that $\xi_\ell(1,1)$ is real-valued almost surely, that real and imaginary parts of $\xi_\ell(1,2)$ are independent and that $\mathbf{E}|\xi_\ell(1,2)|^2 = 1$ for all ℓ. In the setting of HST consider random $\Xi^{(N)} \in \mathrm{Herm}_N$ with entries $\Xi_\ell^{(N)}[i,j] = \xi_\ell(i,j)$ for $\ell = 1,\ldots,m$ and $i,j = 1,\ldots,N$. Then with these matrices $\Xi_\ell^{(N)}$ instead of GUE matrices, one can still draw the same conclusion as in HST. One also has the analogue of the amplification of Voiculescu's theorem.

Notation

Given a $\mathbb{C}$-valued random variable Z, let $\|Z\|_\infty$ denote the essential supremum of $|Z|$ and let $\|Z\|_p = (\mathbf{E}|Z|^p)^{1/p}$ for $p \in [1, \infty)$.

Smooth means infinitely differentiable.

Occasionally when brevity demands it we write $\wedge$ and $\vee$ for min and max, respectively.

The matrix norms $[\![\cdot]\!]_p$

- Given $A \in \mathrm{Mat}_{k\times\ell}$, let $\mu_1 \geq \cdots \geq \mu_{k\wedge\ell} \geq 0$ be the singular values of A, let $[\![A]\!] = [\![A]\!]_\infty = \mu_1$ and more generally for $p \in [1,\infty)$ let $[\![A]\!]_p = \left(\sum_{i=1}^{k\wedge\ell} \mu_i^p\right)^{1/p}$.
- $[\![\cdot]\!]_p$ is a norm on $\mathrm{Mat}_{k\times\ell}$.
- $[\![UXV]\!]_p = [\![X]\!]_p$ for unitary $U \in \mathrm{Mat}_k$, unitary $V \in \mathrm{Mat}_\ell$ and any $X \in \mathrm{Mat}_{k\times\ell}$.
- $\sum_{i=1}^{k\wedge\ell} |A[i,i]|^p \leq [\![A]\!]_p^p$ for $p \in [1,\infty)$.
- $\lim_{p\to\infty} [\![A]\!]_p = [\![A]\!]$.
- $[\![AB]\!]_r \leq [\![A]\!]_p [\![B]\!]_q$ whenever AB is defined and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$.
- For the canonical basis $\{\hat{e}_{ij}\}_{i,j=1}^N$ of Herm_N one has $\max_{i,j=1}^N [\![\hat{e}_{ij}]\!]_1 \leq 1$.

See [Simon] for a good introduction to this material as well as applications in mathematical physics. Or see [Horn-Johnson].

Rough control of eigenvalues

The following result is a key component of the proof of HST.

Proposition (Rough control of eigenvalues)

On a common probability space, for each positive integer N, let $\Xi^{(N)}$ be an N-by-N GUE matrix. Then we have

$$\left\| \limsup_{N\to\infty} \left[\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\right] \right\|_\infty < \infty, \quad \sup_{p\in[1,\infty)} \limsup_{N\to\infty} \left\| \left[\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\right] \right\|_p < \infty.$$

This is very old news and far weaker than what is known. In fact both quantities in question equal 2.

It is striking that the HST method does not require knowledge of the exact constant in order to work.

Toward rough control

We reduce the proposition below to a well-known combinatorial estimate of Füredi-Komlós. We will do so by general methods so that the proposition can be extended to handle a wide class of sequences of Wigner-type matrices.

Random matrices of Wigner type

- We say that random $X \in \mathrm{Herm}_N$ is of *Wigner type* if the real random variables $(\hat{e}_{ij}, X)$ are independent, have absolute moments of all orders, and have mean zero.
- Equivalently: random $X \in \mathrm{Herm}_N$ is of Wigner type if the entries $X[i,j]$ for $1 \leq i \leq j \leq N$ are independent, have absolute moments of all orders, have mean zero, and have independent real and imaginary parts.
- If $X \in \mathrm{Herm}_N$ is of Wigner type we write $X \in \mathrm{Wig}_N$.
- To maintain flexibility we do not insist on normalizing the variance of entries of a Wigner-type matrix.

Weak Wigner words

Let $\mathcal{W}_N(k)$ denote the set consisting of $(2k+1)$-tuples

$$(i_0, \ldots, i_{2k}) = (i_\alpha)_{\alpha=0}^{2k} \in \{1, \ldots, N\}^{2k+1}$$

with the following two properties:

- $i_0 = i_{2k}$.
- Every set on the list $\{i_0, i_1\}, \ldots, \{i_{2k-1}, i_{2k}\}$ appears there at least twice.

We call such $(2k+1)$-tuples *weak Wigner words* of *length* $2k+1$.

“Opening the brackets”

For $X \in \mathrm{Wig}_N$ and positive integers k one has

$$\mathbf{E}[\![X]\!]^{2k} \leq \mathbf{E}\mathrm{tr}X^{2k} = \sum_{(i_\alpha)_{\alpha=0}^{2k} \in \mathcal{W}_N(k)} \mathbf{E} \prod_{\alpha=1}^{2k} X[i_{\alpha-1}, i_\alpha].$$

The omitted terms vanish since they are of the form

$$\mathbf{E}(UV), \quad \mathbf{E}U = 0, \quad U \text{ and } V \text{ independent.}$$

This is the standard first move in the combinatorial analysis of largest eigenvalues of Wigner matrices.

The "norm" $[\![\![\cdot]\!]\!]_{2k}$

Let k be a positive integer. For $X \in \mathrm{Herm}_N$ (yes, deterministic, not random) we define

$$[\![\![X]\!]\!]_{2k} = \left(\sum_{(i_\alpha)_{\alpha=0}^{2k} \in \mathcal{W}_N(k)} \prod_{\alpha=1}^{2k} |X[i_{\alpha-1}, i_\alpha]| \right)^{\frac{1}{2k}}.$$

For $X \in \mathrm{Wig}_N$ the relation

$$\| [\![X]\!] \|_{2k} \leq \| [\![X]\!]_{2k} \|_{2k} = \left(\mathbf{E} \mathrm{tr} X^{2k} \right)^{\frac{1}{2k}} \leq \| [\![\![X]\!]\!]_{2k} \|_{2k}$$

is just a rewrite (in admittedly eccentric notation) of the "standard first move" recalled in the previous frame.

We use the scare-quotes because $[\![\![\cdot]\!]\!]_{2k}$ is only approximately a norm.

The main reason for introducing $[\![\![\cdot]\!]\!]_{2k}$ is that it has an obvious monotonicity property which $[\![\cdot]\!]_{2k}$ does not.

The approximate norm property of $[\![\cdot]\!]_{2k}$

Lemma

Given $X_1, \ldots, X_m \in \mathrm{Herm}_N$ and a positive integer k we have

$$\left[\!\!\left[\sum_{\ell=1}^{m} X_\ell\right]\!\!\right]_{2k} \leq \frac{1+\sqrt{5}}{2} \sum_{\ell=1}^{m} [\![X_\ell]\!]_{2k}.$$

The lemma is perhaps not so important but the proof is entertaining.

Proof

We say that a real random variable χ has the *Fibonacci law* if it is bounded, of mean zero and satisfies

$$\chi^2 = \chi + 1,$$

in which case

$$\mathbf{E}\chi^n = f_{n-1} \geq 1$$

for integers $n \geq 2$, where f_n is the n^{th} Fibonacci number and

$$\|\chi\|_\infty = \frac{1+\sqrt{5}}{2}.$$

Proof (continued)

Now let $\{\chi_{ij}\}_{1\leq i\leq j\leq N}$ be an i.i.d. family of Fibonacci random variables. Let $\{e_{ij}\}_{i,j=1}^{N}$ be the family of elementary N-by-N matrices. Put

$$\begin{aligned}
\tilde{X}_\ell &= \sum_{i=1}^{N} \chi_{ii}|X_\ell[i,i]| + \sum_{1\leq i<j\leq N} \chi_{ij}|X_\ell[i,j]|(e_{ij}+e_{ji}), \\
Z &= \sum_{\ell=1}^{m} X_\ell, \;\; \tilde{Z} = \sum_{\ell=1}^{N} \tilde{X}_\ell,
\end{aligned}$$

noting that $\tilde{X}_\ell, \tilde{Z} \in \mathrm{Wig}_N$ and $Z \in \mathrm{Herm}_N$.

Proof (concluded)

Clearly we have

$$\left\| \left[\!\left[\widetilde{X_\ell}\right]\!\right]_{2k} \right\|_{2k} \leq \frac{1+\sqrt{5}}{2} [\![X_\ell]\!]_{2k}.$$

It is easy to see that for $(i_\alpha)_{\alpha=0}^{2k} \in \mathcal{W}_N(k)$ we have

$$\prod_{\alpha=1}^{2k} |Z[i_{\alpha-1}, i_\alpha]| \leq \mathbf{E} \prod_{\alpha=1}^{2k} \widetilde{Z}[i_{\alpha-1}, i_\alpha]$$

and thus

$$[\![Z]\!]_{2k} \leq \left\| \left[\!\left[\tilde{Z}\right]\!\right]_{2k} \right\|_{2k}.$$

Since $\|[\![\cdot]\!]_{2k}\|_{2k}$ is truly a norm, the result follows. $\square$

Lemma (Rough control via $[\![\cdot]\!]_{2k}$)

For each sufficiently large positive integer N let there be given $X^{(N)} \in \mathrm{Wig}_N$ and an even positive integer q_N. Fix constants $C \in [0,\infty)$ and $\epsilon \in (0,\infty]$. Assume that

$$\limsup_{N\to\infty} \left\| [\![[\![X^{(N)}]\!]]\!]_{q_N} \right\|_{q_N} \le C \quad \text{and} \quad \liminf_{N\to\infty} q_N/\log N \ge \epsilon.$$

Then

$$\sup_{q\in[1,\infty)} \limsup_{N\to\infty} \left\| [\![X^{(N)}]\!] \right\|_q \le K \quad \text{and} \quad \left\| \limsup_{N\to\infty} [\![X^{(N)}]\!] \right\|_\infty \le K$$

for a constant $K \in [0,\infty)$ depending only on C and ϵ.

Proof Pick $0 < \eta < \epsilon$ and $C < D < \infty$ arbitrarily. For some positive integer N_0 and for all $N \geq N_0$ the random matrix $X^{(N)}$ is defined and we have

$$e^{q_N/\eta} \geq N \text{ and } \left\| [\![X^{(N)}]\!] \right\|_{q_N} \leq \left\| [\![\![X^{(N)}]\!]\!]_{q_N} \right\|_{q_N} \leq D,$$

which already proves the first claim with $K = D$, and we have

$$\sum_{N \geq N_0} \Pr\left([\![X^{(N)}]\!] > De^{2/\eta}\right) \leq \sum_{N \geq N_0} \mathbf{E}\left(\frac{X^{(N)}}{De^{2/\eta}}\right)^{q_N} \leq \sum_{N \geq N_0} \frac{1}{N^2} < \infty.$$

The latter statement proves the second claim with $K = De^{2/\eta}$ by the Borel-Cantelli lemma.

After optimizing over D and η we can take, say, $K = Ce^{2/\epsilon}$. □

Füredi-Komlós enumeration

Let

$$\mathcal{W}_N(k,w) \subset \mathcal{W}_N(k)$$

denote the subset consisting of $(2k+1)$-tuples in which exactly w distinct integers appear.

Theorem ([Füredi-Komlós 1981])

For positive integers k, w and N we have

$$|\mathcal{W}_N(k,w)| \leq 2^{2k}(2k)^{6(k-w+1)}N^w\mathbf{1}_{w-1\leq k}.$$

See for example [Anderson-Guionnet-Zeitouni] for treatment and application of this theorem at length.

Remark: rooted planar trees

It is well-known that

$$|\mathcal{W}_N(k, k{+}1)| = \frac{1}{k+1}\left(\begin{array}{c} 2k \\ k \end{array}\right) N(N{-}1)\cdots(N{-}w{+}1) \le 2^{2k-1}N^w.$$

This follows, say, by interpreting elements of the set in question as rooted planar trees with $k+1$ vertices each bearing a distinct label chosen from $\{1,\ldots,N\}$. These are the least "singular" of the elements of $\mathcal{W}_N(k)$.

This is the main point in the combinatorial moment-method proof of the semicircle law.

Idea of proof of FK estimate

We carry the explanation just far enough to see how the form of the estimate comes about.

Fix

$$\mathbf{i} = (i_\alpha)_{\alpha=0}^{2k} \in \mathcal{W}_N(k,w).$$

Put

$$I_1 = \{\alpha \in \{1,\dots,2k\} \mid i_\alpha \notin \{i_0,\dots,i_{\alpha-1}\}\}.$$

Consider the tree $T = (V,E)$ where

$V = \{i_\alpha \in \{1,\dots,N\} \mid \alpha = 0,\dots,2k\}$ and $E = \{\{i_{\alpha-1}, i_\alpha\} \mid \alpha \in I_1\}$.

Note that

$$|V| = w, \quad |I_1| = w-1.$$

The sequence $\mathbf{i} = (i_\alpha)_{\alpha=0}^{2k}$ is only approximately a walk on T.

Sometimes it takes steps along edges of T but other times it takes a gratuitous leap.

Idea of proof of FK estimate (continued)

Let

$$I_2 \subset \{1, \ldots, 2k\}$$

be the set of "times" α at which the walk-with-jumps $\mathbf{i}$ completes a visit to an edge of T it has visited exactly once before.

Note that

$$|I_2| = |I_1| = w - 1 \leq k.$$

by the definition of a weak Wigner word.

Idea of proof of FK estimate (continued)

Let $\mathcal{A}$ be the partition of the set $\{0, \ldots, 2k\}$ into the vertex sets of the connected components of the graph

$$(\{0, \ldots, 2k\}, \{\{\alpha - 1, \alpha\} \mid \alpha \in I_1 \cup I_2\}).$$

Each block of the partition $\mathcal{A}$ is a set of consecutive integers.

Note that

$$|\mathcal{A}| = 2k - 2w + 3.$$

The point of this "parsing" operation is that for each $A \in \mathcal{A}$ the subsequence

$$\mathbf{i}_A = \{i_\alpha\}_{\alpha \in A}$$

is an honest walk on T.

Idea of the proof of FK estimate (concluded)

It can be shown that $\mathbf{i}$ is uniquely determined from knowledge of

- I_1;
- I_2;
- i_α for each $\alpha \in I_1 \cup \{0\}$;
- $i_{\min A}$ for each $0 \notin A \in \mathcal{A}$, and
- $i_{\mathrm{crit} A}$ for each $0 \notin A \in \mathcal{A}$ where $\mathrm{crit} A$ is the time of the last visit of the walk $\mathbf{i}_A$ to the set $\{i_\alpha \mid \alpha \leq \min A\}$.

Counting these choices somewhat crudely we get the bound

$$\frac{(2k!)}{(2w-2)!(2k-2w+2)!} \cdot \frac{(2w-2)!}{(w-1)!(w-1)!} N^w w^{4(k-w+1)}$$

whence after further simplification the announced bound.

Technical lemma for applying FK enumeration

Lemma

Fix $X \in \mathrm{Wig}_N$ *and* $(i_\alpha)_{\alpha=0}^{2k} \in \mathcal{W}_N(k)$. *If*

$$\sup_{q \in [2,\infty)} \max_{i,j=1}^{N} q^{-c} \|X[i,j]\|_q \leq 1.$$

then

$$\mathbf{E} \prod_{\alpha=1}^{2k} |X[i_{\alpha-1}, i_\alpha]| \leq 2^{2ck} (2k)^{2c(k-w+1)}.$$

Roughly, in words, the hypothesis says that you have polynomial control of L^q-norms of entries.

Proof of technical lemma

Consider the (unoriented) graph $G = (V, E)$ where

$$V = \{i_\alpha \mid \alpha = 0, \dots, 2k\} \text{ and } E = \{\{i_{\alpha-1}, i_\alpha\} \mid \alpha = 1, \dots, 2k\}.$$

Since this time, in contrast to the procedure used in the proof of the FK bound, no edges were discarded,

$$\mathbf{i} = (i_\alpha)_{\alpha=0}^{2k}$$

as an honest walk on G.

Proof of technical lemma (concluded)

Let p_n be the number of edges of G visited n or more times by $\mathbf{i}$. Clearly $p_1 = p_2$ (this is just a re-iteration of the definition of a weak Wigner word) and $\sum_n p_n = \sum_{n=1}^{2k} p_n = 2k$.

Furthermore since G is connected $p_1 + 1 \leq w$. It follows that $\sum_{n\geq 3} p_n \leq 2(k-w+1)$.

Finally we have

$$\mathbf{E}\prod_{\alpha=1}^{2k} |X[i_{\alpha-1}, i_\alpha]| \leq \prod_{n=2}^{\infty} e^{cn\log(n)(p_n - p_{n+1})}$$

$$= \quad 2^{2cp_2}\prod_{n=3}^{2k} e^{c(n\log(n)-(n-1)\log(n-1))p_n} \leq 2^{2ck}\prod_{n=3}^{2k} e^{c\log(n)p_n}$$

$$\leq \quad 2^{2ck}\prod_{n=3}^{2k}(2k)^{cp_n} \leq 2^{2ck}(2k)^{2c(k-w+1)}. \quad \square$$

Corollary

Let $X \in \mathrm{Wig}_N$ and a constant $c \in [0,\infty)$ be given such that

$$\sup_{q \in [2,\infty)} \max_{i,j=1}^{N} q^{-c} \|X[i,j]\|_q \leq 1.$$

Then for positive integers k we have

$$2(2k)^{6+2c} \leq N \Rightarrow \left\| \left[\left[\left[\frac{X}{2^{c+1}\sqrt{N}} \right]\right]\right]_{2k} \right\|_{2k} \leq (2N)^{\frac{1}{2k}}.$$

Proof

$$
\begin{aligned}
\left\| \left[\!\left[\!\left[\frac{X}{\sqrt{N}} \right]\!\right]\!\right]_{2k} \right\|_{2k}^{2k} &\leq \sum_{w=1}^{k+1} |\mathcal{W}_N(k,w)| \frac{2^{2ck}(2k)^{2c(k-w+1)}}{N^k} \\
&\leq \sum_{w=1}^{k+1} 2^{2k}(2k)^{6(k-w+1)} N^2 \frac{2^{2ck}(2k)^{2c(k-w+1)}}{N^k} \\
&\leq 2^{(1+c)(2k)} N \sum_{w=1}^{k+1} \frac{(2k)^{(6+2c)(k-w+1)}}{N^{k-w+1}} \\
&\leq 2^{1+(1+c)(2k)} N.
\end{aligned}
$$

$\square$

Rough control of eigenvalues of GUE matrices

Recall that to get rough control of eigenvalues of GUE matrices we needed to prove the following statement.

Proposition

On a common probability space, for each positive integer N, let $\Xi^{(N)}$ be an N-by-N GUE matrix. Then we have

$$\left\|\limsup_{N\to\infty}\left[\left[\frac{\Xi^{(N)}}{\sqrt{N}}\right]\right]\right\|_\infty<\infty,\quad \sup_{p\in[1,\infty)}\limsup_{N\to\infty}\left\|\left[\left[\frac{\Xi^{(N)}}{\sqrt{N}}\right]\right]\right\|_p<\infty.$$

Rough control of eigenvalues of more general Wigner matrices

The next result vastly generalizes the previous one since for a standard normal random variable Z we have $\|Z\|_q = O(\sqrt{q})$. It also handles fake GUE matrices.

Proposition

Fix a constant $c \geq 0$. On a common probability space, for each positive integer N, let there be given $\Xi^{(N)} \in \mathrm{Wig}_N$ such that

$$\sup_{q\in[2,\infty)} \sup_{N=1}^{\infty} \max_{i,j=1}^{N} q^{-c} \left\| \Xi^{(N)}[i,j] \right\|_q \leq 1.$$

Then for an absolute constant K depending only on c we have

$$\left\| \limsup_{N\to\infty} \left[\!\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\!\right] \right\|_\infty \leq K, \quad \sup_{p\in[1,\infty)} \limsup_{N\to\infty} \left\| \left[\!\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\!\right] \right\|_p \leq K.$$

Proof The preceding corollary along with the rough control in terms of $\|[\![\cdot]\!]_{2k}\|_{2k}$ together prove this.

A few more words on rough control

We have analyzed the proof of Bai-Yin and extracted a combinatorial estimate from it. Derivation of the estimate is too technical to go into. But the estimate itself is relatively simple to state and it interesting to compare it to the FK estimate.

The Bai-Yin parameter

Let

$$\mathbf{i} = (i_\alpha)_{\alpha=0}^{2k} \in \mathcal{W}_N(k, w)$$

be a weak Wigner word. Consider again the graph

$$G = (\{i_\alpha \mid \alpha = 0, \ldots, 2k\}, \{\{i_{\alpha-1}, i_\alpha\} \mid \alpha = 1, \ldots, 2k\}).$$

As before, let p_n denote the number of edges of G visited at least n times by $\mathbf{i}$.

We define the *Bai-Yin parameter* of $\mathbf{i}$ to be the quantity

$$p_1 + 1 - w + p_3.$$

In words this parameter is the number of edges of G not needed to make G connected plus the number of edges visited three or more times.

More on the Bai-Yin parameter

The Bai-Yin parameter is a rough measure of the "badness" of G and the walk $\mathbf{i}$ on G which is sensitive to third moments.

The importance of this parameter is something one can learn by attending to the proofs in Bai-Yin.

Refinement of Füredi-Komlós enumeration

Let $\mathcal{W}_N(k,w,t)$ denote the subset of $\mathcal{W}_N(k,w)$ consisting of weak Wigner words with Bai-Yin parameter equal to t.

Theorem (FKBY estimate)

For integers $k,w,t,N>0$ we have

$$|\mathcal{W}_N(k,w,t)| \leq 2^{6k} w^{3t} t^{2(k-w+1)} N^w \mathbf{1}_{w\leq k} \mathbf{1}_{t\leq 2(k-w+1)}.$$

For comparison recall that the FK estimate is

$$|\mathcal{W}_N(k,w)| \leq 2^{2k}(2k)^{6(k-w+1)} N^w \mathbf{1}_{w\leq k+1}.$$

The FKBY estimate can be proved by only a light modification of FK arguments. It does not require some alternative development of graph-theoretical tools.

Sample application of FKBY

The FKBY estimate is good enough to prove the following estimate which cannot so far as we know be proved using the FK estimate.

Proposition

On a common probability space, for each positive integer N, let there be given $\Xi^{(N)} \in \mathrm{Wig}_N$ such that

$$\sup_{N=1}^{\infty} \max_{i,j=1}^{N} \left\| \Xi^{(N)}[i,j] \right\|_3 \leq 1 \quad \textit{and} \quad \sup_{N=1}^{\infty} \max_{i,j=1}^{N} \left\| \Xi^{(N)}[i,j] \right\|_\infty \leq \sqrt{N}.$$

Then for some absolute constant K we have

$$\left\| \limsup_{N\to\infty} \left[\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\right] \right\|_\infty < K \quad \textit{and} \quad \sup_{p\in[1,\infty)} \limsup_{N\to\infty} \left\| \left[\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\right] \right\|_p < K.$$

This result is strong enough to permit deduction of the theorem of Bai-Yin from the preceding results on rough control derived from the FK bound.

Our next goal is. . .

. . . is to briefly explain by example the role of truncation techniques.

Recollection of the theorem of Bai-Yin

In order to show how one can exploit truncation techniques, we carry out an exercise. Recall the theorem of Bai-Yin.

Theorem (Bai-Yin [Bai-Yin 1988])

Let $\{\xi_{ij}\}_{1\leq i\leq j<\infty}$ be an independent family of $\mathbb{C}$-valued random variables such that the law of ξ_{ij} depends only on $\mathbf{1}_{i<j}$, ξ_{11} is real almost surely, $\|\xi_{11}\|_4 < \infty$ and $\|\xi_{12}\|_4) < \infty$, and $\mathbf{E}\xi_{11} = 0 = \mathbf{E}\xi_{12}$. Let $\Xi^{(N)}$ be the random N-by-N hermitian matrix with entries $\Xi^{(N)}[i,j] = \xi_{ij}$ for $i \leq j$. Then $\lim_{N\to\infty}\left[\left[\frac{\Xi^{(N)}}{\sqrt{N}}\right]\right] = 2\|\xi_{12}\|_2$ almost surely.

Separation of issues in Bai-Yin

In order to separate important issues from less important ones, consider the following three statements.

Proposition (A)

Hypotheses are those of the theorem strengthened so that $\|\xi_{12}\|_2 = 0$. *Then* $\lim_{N\to\infty} \left[\!\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\!\right] = 0$ *almost surely.*

Proposition (B)

Hypotheses are those of the theorem strengthened so that $\max(\|\xi_{11}\|_\infty, \|\xi_{12}\|_\infty) < \infty$. *Then* $\lim_{N\to\infty} \left[\!\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\!\right] = 2\|\xi_{12}\|_2$ *almost surely.*

Proposition (C)

Hypotheses are exactly as in the theorem but one concludes only that $\limsup_{N\to\infty} \left[\!\left[\frac{\Xi^{(N)}}{\sqrt{N}} \right]\!\right] \le K\|\xi_{12}\|_4$ *for an absolute constant* K.

“ABC implies Bai-Yin”

In order to complete our separation of issues, we will prove the following logical relationship.

Proposition (“A,B,C implies Bai-Yin”)

Propositions A,B,C together imply the theorem.

Before starting the proof we comment on the proofs of statements A, B and C.

Comment on the proof of A

Proposition A asserts that the theorem of Bai-Yin holds when the Wigner matrices in question are diagonal. To prove this is an exercise in the Borel-Cantelli lemma. To get the necessary convergence of a series one exploits the integration identity

$$\mathbf{E}Z^4 = \int_0^\infty \Pr(Z^4 > u)\,du = 2\int_0^\infty t\,\Pr(Z > \sqrt{t})\,dt.$$

Proposition B asserts the weaker version of the theorem of Bai-Yin that was proved earlier by Füredi-Komlós. Most of the arguments needed for that proof have been rehearsed above. For the remaining details see, for example, [Anderson-Guionnet-Zeitouni].

Comment on the proof of C

The FKBY criterion stated above along with the application following it are strong enough to prove statement C but do not make the constant K explicit.

Proof of "A,B,C implies Bai-Yin"

By Proposition A we may assume that that diagonal entries of $\Xi^{(N)}$ vanish identically. For any constant $c > 0$, let $\Xi^{(N)}_{\leq c}$ be the N-by-N random hermitian matrix with entries $\Xi^{(N)}_{\leq c}[i,j] = \xi_{ij}\mathbf{1}_{|\xi_{ij}|\leq c}$ for $i \leq j$ and put $\widehat{\Xi}^{(N)}_{\leq c} = \Xi^{(N)}_{\leq c} - \mathbf{E}\Xi^{(N)}_{\leq c}$. Then for any $c > 0$ we have almost surely

$$\lim_{N\to\infty}\left[\left[\frac{\widehat{\Xi}^{(N)}_{\leq c}}{\sqrt{N}}\right]\right] = 2\left\|\xi_{12}\mathbf{1}_{|\xi_{12}|\leq c} - \mathbf{E}\xi_{12}\mathbf{1}_{|\xi_{12}|\leq c}\right\|_2$$

and

$$\limsup_{N\to\infty}\left[\left[\frac{\Xi^{(N)} - \widehat{\Xi}^{(N)}_{\leq c}}{\sqrt{N}}\right]\right] \leq K\left\|\xi_{12}\mathbf{1}_{|\xi_{12}|> c} - \mathbf{E}\xi_{12}\mathbf{1}_{|\xi_{12}|> c}\right\|_4$$

by Propositions B and C, respectively. By dominated convergence the preceding statements imply the desired result. □

Upshot

- Bai-Silverstein in [Bai-Silverstein 1998] start their proof with a truncation step similar in form to the example considered above.
- Anderson in [Anderson, Ann. Probab., to appear] imitated the truncation procedure of Bai-Silverstein.
- It seems that quite generally support questions involving Wigner matrices satisfying parsimonious moment conditions can be reduced to the study of Wigner matrices with bounded entries.

Our next goal is...

...to recall the Poincaré inequality for Gaussian random variables along with a carefully organized proof of it designed to expedite a certain generalization to be considered later.

Setup for the Poincaré inequality

- Let $Z \in \mathbb{R}^n$ be a random vector with i.i.d. centered Gaussian entries, all of variance σ^2.
- Let $f : \mathbb{R}^n \to \mathbb{C}$ be infinitely differentiable and suppose that partial derivatives of all orders have polynomial growth; hereafter we call such functions *tame*.
- Let D_i denote the operation of differentiation of functions on $\mathbb{R}^n$ with respect to the i^{th} coordinate.

Statement of the Poincaré inequality

Theorem (The Poincaré inequality for Gaussian random variables)

Notation and assumptions are as above. We have

$$\mathbf{E}|f(Z)|^2 - |\mathbf{E}f(Z)|^2 \leq \sigma^2 \mathbf{E} \sum_{i=1}^{n} |D_i[f](Z)|^2.$$

For simplicity we have given ourselves assumptions concerning the regularity of f that are much stronger than necessary but which are no trouble to verify in the applications we have in mind.

In the next several frames we give a proof by way of a lemma which we will use later to generalize the Poincaré inequality in a useful way.

Setup for a technical lemma

For the moment we assume of the random vector $Z \in \mathbb{R}^n$ only that its entries have absolute moments of all orders so as to rule out integrability issues, and we let σ^2 denote an arbitrary positive constant.

Let $X, Y \in \mathbb{R}^n$ be i.i.d. copies of Z and for $t \in (0,1)$ put

$$X^{(t)} = \sqrt{t}X + \sqrt{1-t}Y.$$

We introduce the auxiliary functions

$$h_i^{(y,t)} : \mathbb{R}^n \to \mathbb{C} \text{ for } y \in \mathbb{R}^n,\ 0 < t < 1 \text{ and } i = 1, \ldots, n$$

defined by

$$h_i^{(y,t)}(x) = (f^*(x) - f^*(y))f_i(\sqrt{t}\,x + \sqrt{1-t}\,y).$$

Finally, let M_i denote the operation of multiplication of functions on $\mathbb{R}^n$ by the i^{th} coordinate.

Statement of the technical lemma

Lemma (The pre-Poincaré identity)

Assumptions and notation are as above. We have an integration identity

$$\begin{aligned} & \mathbf{E}|f(Z)|^2 - |\mathbf{E}f(Z)|^2 - \sigma^2 \mathbf{E} \int_0^1 \sum_{i=1}^n D_i[f^*](X) D_i[f](X^{(t)})\, d\sqrt{t} \\ = \ & \sum_{i=1}^n \mathbf{E} \int_0^1 (M_i - \sigma^2 D_i)[h_i^{(Y,t)}](X)\, d\sqrt{t}. \end{aligned}$$

Later we will present a nice method for bounding the right side.

Proof of the technical lemma

We start by using the fundamental theorem of calculus.

$$
\begin{aligned}
\mathbf{E}|f(Z)|^2 - |\mathbf{E} f(Z)|^2 &= \mathbf{E} f^*(X)(f(X) - f(Y)) \\
= & \int_0^1 \frac{d}{dt} \mathbf{E} f^*(X) f(X^{(t)})\, dt \\
= & \mathbf{E} \int_0^1 \sum_{i=1}^n f^*(X) \left(\frac{X_i}{2\sqrt{t}} - \frac{Y_i}{2\sqrt{1-t}} \right) D_i[f](X^{(t)})\, dt.
\end{aligned}
$$

Proof of the technical lemma (continuation)

Then we observe a symmetry.

$$\begin{aligned} & \mathbf{E}\int_0^1 \sum_{i=1}^n f^*(X)\frac{Y_i}{2\sqrt{1-t}}D_i[f](\sqrt{t}\,X+\sqrt{1-t}\,Y)\,dt \\ = \quad & \mathbf{E}\int_0^1 \sum_{i=1}^n f^*(Y)\frac{X_i}{2\sqrt{t}}D_i[f](\sqrt{1-t}\,Y+\sqrt{t}\,X)\,dt. \end{aligned}$$

Thus we have

$$\begin{aligned} & \mathbf{E}|f(Z)|^2-|\mathbf{E}f(Z)|^2 \\ = \quad & \mathbf{E}\int_0^1 \sum_{i=1}^n X_i(f^*(X)-f^*(Y))D_i[f](X^{(t)})\,d\sqrt{t} \\ = \quad & \sum_{i=1}^n \mathbf{E}\int_0^1 M_i[h_i^{(Y,t)}](X)\,d\sqrt{t}. \end{aligned}$$

Proof of the technical lemma (conclusion)

For $i = 1, \ldots, n$ we have

$$\begin{aligned} D_i[h_i^{(y,t)}](x) &= D_i[f^*](x) D_i[f](\sqrt{t}\, x + \sqrt{1-t}\, y) \\ &\quad + (f^*(x) - f^*(y)) D_i^2[f](\sqrt{t} x + \sqrt{1-t}\, y)\sqrt{t} \end{aligned}$$

and thus

$$\mathbf{E} \int_0^1 D_i[h_i^{(Y,t)}](X) d\sqrt{t} = \mathbf{E} \int_0^1 D_i[f^*](X) D_i[f](X^{(t)})\, d\sqrt{t}.$$

The right side has only the one term because the other is killed by odd symmetry. The proof of the lemma is complete. $\square$

Proof of the Poincaré inequality (conclusion)

For a standard normal random variable ζ and tame function $g : \mathbb{R} \to \mathbb{C}$ we have

$$\mathbf{E}\zeta g(\zeta) = \sigma^2 \mathbf{E} g'(\zeta).$$

Let $X, Y \in \mathbb{R}^n$ be i.i.d. copies of a Gaussian random vector $Z \in \mathbb{R}^n$ with i.i.d. centered entries of variance σ^2.

Plugging into the technical lemma we can now kill the error on the right side, thus obtaining the (very well-known) identity

$$\begin{aligned} & \mathbf{E}|f(Z)|^2 - |\mathbf{E}f(Z)|^2 \\ = \quad & \sigma^2 \mathbf{E} \int_0^1 \sum_{i=1}^n D_i[f^*](X) D_i[f](tX + \sqrt{1-t^2}Y)\, dt \end{aligned}$$

whence the result via Cauchy-Schwarz. The proof of the theorem is complete. $\square$

Our next goal. . .

. . . is to introduce a technique for approximately integrating by parts and then by means of this technique to prove a variant of the Poincaré inequality.

The papers [Khorunzhy-Khoruzhenko-Pastur] and [Lytova-Pastur 2009 A] are the sources of this excellent idea.

It has been taken up elsewhere in RMT, e.g., it is used in [O'Rourke-Renfrew-Soshnikov 2011 A] and [O'Rourke-Renfrew-Soshnikov 2011 B] and called there the *decoupling formula*. See also [Lytova-Pastur 2009 B].

We reconfigure the idea a bit to handle the new situation of polynomials in Wigner matrices. The idea is quite robust so this turns out not to be especially difficult.

Setup for an integration by parts identity

- Let Z be a real random variable with absolute moments of all orders.
- Let $f : \mathbb{R} \to \mathbb{C}$ be an infinitely differentiable function such that derivatives of all orders have polynomial growth. (Hereafter such is called *tame*.)
- Let $U_1, \ldots, U_k$ be real random variables independent of Z such that U_i has the beta distribution of parameters 1 and i.
- Recall that this means $\mathbf{E}\varphi(U_k) = k \int_0^1 (1-t)^{k-1}\varphi(t)\,dt$.
- Let $\kappa_i(Z)$ denote the i^{th} cumulant of Z.
- Recall that $\sum_{n=1}^{\infty} \frac{\kappa_n(Z)}{n!} t^n = \log\left(\sum_{n=0}^{\infty} \frac{\mathbf{E}Z^n}{n!} t^n\right)$.

Integration by parts identity

Proposition (Integration by parts identity)

Assumptions and notation are as above. We have

$$\begin{aligned} & \mathbf{E}\left(Zf(Z) - \sum_{i=0}^{k-1} \frac{\kappa_{i+1}(Z)}{i!} f^{(i)}(Z)\right) \\ = \; & \mathbf{E}\left(\frac{Z^{k+1}}{k!} f^{(k)}(U_k Z) - \sum_{i=0}^{k-1} \frac{\kappa_{i+1}(Z)}{i!} \frac{Z^{k-i}}{(k-i)!} f^{(k)}(U_{k-i} Z)\right). \end{aligned}$$

Things are set up so that there are no issues of integrability to contend with.

We review relevant definitions and then prove the proposition in the next several slides.

By definition we have a formal power series identity

$$\sum_{n=1}^{\infty} \frac{\kappa_n(Z)}{n!} t^n = \log \left(\sum_{n=0}^{\infty} \frac{\mathbf{E} Z^n}{n!} t^n \right).$$

Applying $\frac{d}{dt}$ on both sides and rearranging, we get a factorization

$$\left(\sum_{n=0}^{\infty} \frac{\mathbf{E} Z^{n+1}}{n!} t^n \right) = \left(\sum_{n=0}^{\infty} \frac{\kappa_{n+1}(Z)}{n!} t^n \right) \left(\sum_{n=0}^{\infty} \frac{\mathbf{E} Z^n}{n!} t^n \right)$$

Cumulant moment relations

... and thus relations

$$\frac{\mathbf{E} Z^{n+1}}{n!} = \sum_{i=0}^{n} \frac{\kappa_{i+1}(Z)}{i!} \frac{\mathbf{E} Z^{n-i}}{(n-i)!}$$

for integers $n \geq 0$.

Taylor's formula (toward proof of IBPI)

Recall that Taylor's formula says that

$$\begin{aligned} f(x) &= \sum_{n=0}^{k-1} \frac{x^n}{n!} f^{(n)}(0) + \frac{x^k}{(k-1)!} \int_0^1 f^{(k)}(tx)(1-t)^{k-1}\, dt \\ &= \sum_{n=0}^{k-1} \frac{x^n}{n!} f^{(n)}(0) + \frac{x^k}{k!} \mathbf{E} f^{(k)}(U_k x). \end{aligned}$$

Simple consequences of the preceding identities (toward proof of IBPI)

We have

$$\sum_{n=0}^{k-1} \frac{\mathbf{E} Z^{n+1}}{n!} f^{(n)}(0) = \sum_{i=0}^{k-1} \sum_{n=i}^{k-1} \frac{\kappa_{i+1}(Z)}{i!} \frac{\mathbf{E} Z^{n-i}}{(n-i)!} f^{(n)}(0),$$

and

$$\begin{aligned} \mathbf{E} Z f(Z) &= \sum_{n=0}^{k-1} \frac{\mathbf{E} Z^{n+1}}{n!} f^{(n)}(0) + \mathbf{E} \frac{Z^{k+1}}{k!} f^{(k)}(U_k Z), \\ \mathbf{E} f^{(i)}(Z) &= \sum_{n=i}^{k-1} \frac{\mathbf{E} Z^{n-i}}{(n-i)!} f^{(n)}(0) + \mathbf{E} \frac{Z^{k-i}}{(k-i)!} \mathbf{E} f^{(k)}(U_{k-i} Z) \end{aligned}$$

for $i = 0, \ldots, k-1$.

Completion of the proof of IBPI

One simply combines the identities on the preceding frame appropriately to get the claimed formula. $\square$

Drastic simplification of IBPI

Going forward, we will make use of the preceding theory exclusively through use of a drastically simplified version of the preceding identity. Needless to say it does not exhaust the possibilities inherent in IBPI.

Drastic simplification of IBPI (statement)

- As before let Z be a real random variable with absolute moments of all orders.
- Fix an integer $k \geq 2$ and assume that Z has the same first k moments as a centered Gaussian random variable. (Hereafter such is called *Gaussian through the k^{th} order*.)
- As before let $f : \mathbb{R} \to \mathbb{C}$ be tame. Put $\mathcal{M}^k[f](x) = \sup_{t \in [0,1]} |f^{(k)}(tx)|$.

Proposition (Drastically simplified IBPI)

Assumptions and notation are as above. We have

$$\left|\mathbf{E}\left(Zf(Z) - \mathbf{Var}(Z)f'(Z)\right)\right| \leq 2\|Z\|_{2(k+1)}^{k+1}\left\|\mathcal{M}^k[f](Z)\right\|_2.$$

Proof of drastically simplified IBPI

By hypothesis

$$\kappa_j(Z) = \delta_{j2}\mathbf{Var}(Z) \text{ for } j = 1, \ldots, k.$$

Thus IBPI simplifies in the present case to

$$\begin{aligned} & \mathbf{E}(Zf(Z) - f'(Z)) \\ = \ & \mathbf{E}\left(\frac{Z^{k+1}}{k!}f^{(k)}(U_k Z) - \mathbf{Var}(Z)\frac{Z^{k-1}}{(k-1)!}f^{(k)}(U_{k-1}Z)\right). \end{aligned}$$

Now make the obvious application of the Hölder inequality and the definitions. $\square$

Lecture 4 (June 21, 2012)

Soft pitch proof of semicircle for GUE

$$\Xi^{(N)}\text{: } N\text{-by-}N \text{ GUE}$$

$$F^{(N)}(z) = \frac{1}{N}\mathrm{tr}\left(\frac{\Xi^{(N)}}{\sqrt{N}} - z\mathbf{I}_N\right)^{-1} = S_{\mu^{(N)}}(z)$$

$$G(z) = \frac{1}{2\pi}\int_{-2}^{2}\frac{\sqrt{4-t^2}}{t-z}\,dt = S_\mu(z)$$

$$\text{Recall:}\quad 1 + (z + G(z))G(z) = 0.$$

How to make $F^{(N)}(z)$ get close to $G(z)$? Roughly: Make

$$\left|1 + (z + \mathbf{E}F^{(N)}(z))(\mathbf{E}F^{(N)}(z))\right|$$

small.

Soft pitch (concluded)

In (some) more detail.

Integrate by parts:

$$\mathbf{E}\left(1 + (z + F^{(N)}(z))F^{(N)}(z)\right) = 0.$$

Use Poincaré:

$$\begin{aligned}
&|\mathbf{E}F^{(N)}(z)^2 - (\mathbf{E}F^{(N)}(z))^2| \\
&\leq \mathbf{E}|F^{(N)}(z) - \mathbf{E}F^{(N)}(z)|^2 \leq \frac{1}{N^2(\Im z)^4}
\end{aligned}$$

$$|1 + (z + (\mathbf{E}F^{(N)}(z)))\mathbf{E}F^{(N)}(z)| \leq \frac{1}{N^2(\Im z)^4}.$$

Some complex analysis around $G(z)$ gives:

$$|\mathbf{E}F^{(N)}(z) - G(z)| \leq 64\left(1 + \frac{1}{(\Im z)^4}\right)\frac{1}{N^2(\Im z)^4}$$

Enough for $\mu^{(N)} \Rightarrow \mu$ weakly almost surely.

Our next goal is. . .

...to present the general theory of the Schwinger-Dyson equation.

This theory has two main parts.

(a) Operator-theoretic construction of solutions.

(b) Approximation of solutions.

Besides of course [Haagerup-Thorbjørnsen 2005] and [Haagerup-Schultz-Thorbjørnsen 2006], the papers [Hachem-Loubaton-Najim 2007] and [Hachem-Loubaton-Najim 2008] pushing forward Girko's notion deterministic equivalent and the paper [Helton-Rashidi Far-Speicher 2007] studying the Schwinger-Dyson equation from both analytical and numerical perspectives are important sources of influence for the ideas presented in this section. We have to review some simple facts of operator theory and algebra first.

The $\Im$-lemma

For an element A of a $*$-algebra, let

$$\Re A = \frac{A + A^*}{2} \in \mathcal{A}_{\mathrm{sa}} \ \text{ and } \ \Im A = \frac{A - A^*}{2\mathrm{i}} \in \mathcal{A}_{\mathrm{sa}}.$$

Lemma ($\Im$-lemma)

Let $\mathcal{A}$ be a C^*-algebra. If $A \in \mathcal{A}$ satisfies $\Im A > 0$ then $A \in \mathcal{A}^\times$ and $[\![A^{-1}]\!] \leq [\![(\Im A)^{-1}]\!]$.

This statement is an important technical tool.

Proof of the $\Im$-lemma

- Let $X = \Re A$ and $Y = \Im A$ so that $A = X + \mathrm{i}Y$.
- By hypothesis $Y > 0$. Put $W = Y^{-1/2}XY^{-1/2}$.
- Then W is self-adjoint and thus $W + \mathrm{i}$ is normal.
- Since $\mathrm{Spec}(W + \mathrm{i}) \subset \{x + \mathrm{i} \mid x \in \mathbb{R}\}$, we have

$$\mathrm{Spec}((W + \mathrm{i})^{-1}) \subset \{\frac{1}{x + \mathrm{i}} \mid x \in \mathbb{R}\},$$

  hence the spectral radius of $(W + \mathrm{i})^{-1}$ is ≤ 1, and
- hence $[\![(W + \mathrm{i})^{-1}]\!] \leq 1$.
- Since we have $A = Y^{1/2}(W + \mathrm{i})Y^{1/2}$, in fact A is invertible.
- Finally, we have

$$\left[\!\left[A^{-1}\right]\!\right] \leq \left[\!\left[Y^{-1/2}\right]\!\right]^2 \left[\!\left[(W + \mathrm{i})^{-1}\right]\!\right] \leq \left[\!\left[(\Im A)^{-1}\right]\!\right],$$

  as claimed. □

The jack-knife lemma

Let $\mathcal{A}$ be a C^*-algebra. For $A \in \mathcal{A}^\times$ put $\rho(A) = \left[\!\left[A^{-1}\right]\!\right]^{-1}$.

Lemma (The jack-knife)

Assumptions and notation are as above. We have

$$\mathcal{A}^\times = \{\Lambda \in \mathcal{A} \mid \Im\Lambda > 0\} \cup \bigcup_{\Lambda \in \mathcal{A}^\times} \{\Lambda' \in \mathcal{A} \mid [\![\Lambda' - \Lambda]\!] < \rho(\Lambda)\}.$$

Furthermore, the following statements hold for all $A, B \in \mathcal{A}^\times$:

$$\left[\!\left[A^{-1}\right]\!\right] \leq \frac{1}{\rho(A)}, \quad \left[\!\left[A^{-1} - B^{-1}\right]\!\right] \leq \frac{[\![A-B]\!]}{\rho(A)\rho(B)},$$

$$\left[\!\left[A^{-1} - B^{-1} + A^{-1}(A-B)A^{-1}\right]\!\right] \leq \frac{[\![A-B]\!]^2}{\rho(A)^2\rho(B)},$$

$$|\rho(A) - \rho(B)| \leq [\![A-B]\!], \quad \Im A > 0 \Rightarrow \rho(\Im A) \leq \rho(A),$$

$$A\text{: normal } \Rightarrow \rho(A) = \inf_{z \in \mathrm{Spec}(A)} |z|.$$

Quick proof of the jack-knife lemma

- The last two statements and the inclusion $\{\Lambda \in \mathcal{A} \mid \Im\Lambda > 0\} \subset \mathcal{A}^\times$ are restatements of facts already presented in the previous two lemmas.
- The rest of the lemma except $|\rho(A) - \rho(B)| \leq [\![A - B]\!]$ is proved by applying the inversion formula

  $$B^{-1} = A^{-1} + A^{-1}(A-B)A^{-1} + A^{-1}(A-B)A^{-1}(A-B)A^{-1} + \ldots$$

  which is valid when $[\![A - B]\!]\,[\![A^{-1}]\!] < 1$, i.e., $[\![A - B]\!] < \rho(A)$, and...
- ...the identities

$$B^{-1} - A^{-1} = A^{-1}(A - B)B^{-1},$$

$$B^{-1} - A^{-1} - A^{-1}(A - B)A^{-1} = (A^{-1}(A - B))^2 B^{-1}.$$

Finally,

$$\begin{aligned}
|\rho(A) - \rho(B)| &= \left| \left[\!\left[A^{-1} \right]\!\right]^{-1} - \left[\!\left[B^{-1} \right]\!\right]^{-1} \right| \\
&\leq \rho(A)\rho(B) \left| \left[\!\left[A^{-1} \right]\!\right] - \left[\!\left[B^{-1} \right]\!\right] \right| \\
&\leq \rho(A)\rho(B) \left[\!\left[A^{-1} - B^{-1} \right]\!\right] \\
&\leq \rho(A)\rho(B) \frac{[\![A - B]\!]}{\rho(A)\rho(B)} = [\![A - B]\!].
\end{aligned}$$

Interlude: Review of Schur complements

Given a matrix $A = \left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$ of finite size with entries in some algebra, broken down into blocks with a and d square and d invertible, we have a factorization

$$\left[\begin{array}{cc} a & b \\ c & d \end{array}\right] = \left[\begin{array}{cc} 1 & bd^{-1} \\ 0 & 1 \end{array}\right]\left[\begin{array}{cc} a-bd^{-1}c & 0 \\ 0 & d \end{array}\right]\left[\begin{array}{cc} 1 & 0 \\ d^{-1}c & 1 \end{array}\right].$$

In this setup A is invertible if and only if the *Schur complement*

$$a - bd^{-1}c$$

is invertible, and under these equivalent conditions we have...

Interlude: Inversion formulas involving the Schur complement

...an inversion formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -d^{-1}c & 1 \end{bmatrix} \begin{bmatrix} (a - bd^{-1}c)^{-1} & 0 \\ 0 & d^{-1} \end{bmatrix} \begin{bmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{bmatrix}$$

and another equivalent formula

$$\begin{bmatrix} (a - bd^{-1}c)^{-1} & 0 \\ 0 & d^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ d^{-1}c & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} 1 & bd^{-1} \\ 0 & 1 \end{bmatrix},$$

both of which are useful.

Let

$$\mathcal{S} = \mathrm{Mat}_s \ \text{(to abbreviate notation slightly)}$$

and fix

$$L_1, \ldots, L_m \in \mathrm{Herm}_s.$$

Put

$$\begin{aligned}
\Phi &= \left(A \mapsto \sum_{\ell=1}^{m} L_\ell A L_\ell\right) \in B(\mathrm{Mat}_s),\\
L &= \sum_{\ell=1}^{m} L_\ell \otimes \Xi_\ell \in \mathrm{Mat}_s\left(B(\mathcal{H})\right)_{\mathrm{sa}},\\
\mathcal{D} &= \{\Lambda \in \mathrm{Mat}_s \mid L - \Lambda \otimes 1_{B(\mathcal{H})} \in \mathrm{GL}_s(B(\mathcal{H}))\},\\
\rho &= \left(\Lambda \mapsto \left[\!\left[(L - \Lambda \otimes 1_{B(\mathcal{H})})^{-1}\right]\!\right]^{-1}\right) : \mathcal{D} \to (0, \infty).
\end{aligned}$$

Finally for $\Lambda \in \mathcal{D}$ we define $G(\Lambda) \in \mathrm{Mat}_s$ by

$$G(\Lambda)[i,j] = (1_{\mathcal{H}}, (L - \Lambda \otimes 1_{B(\mathcal{H})})^{-1}[i,j]1_{\mathcal{H}}).$$

The semicircular example

- $\mathcal{S} = \mathbb{C}$,
- $L_1 = 1 \in \mathrm{Herm}_1 = \mathbb{R}$,
- $\Phi = 1 \in B(\mathcal{S}) = \mathbb{C}$.
- $\mathcal{D} = \mathbb{C} \setminus \mathrm{Spec}(\Xi_1) = \mathbb{C} \setminus [-2, 2]$.
- $\rho(z) = \inf_{t \in [-2,2]} |z - t| = \begin{cases} |z+1| & \text{if } \Re(z) \leq -1, \\ \Im z & \text{if } |\Re(z)| < 1, \\ |z-1| & \text{if } \Re(z) \geq 1. \end{cases}$
- $G(z) = (1_{\mathcal{H}}, (\Xi_1 - z 1_{B(\mathcal{H})})^{-1} 1_{\mathcal{H}}) = \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4-t^2}}{t-z}\, dt$.

Properties of $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$

The quintuple $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$ satisfies

$$\mathcal{D} = \{\Lambda \in \mathcal{S} \mid \Im\Lambda > 0\} \cup \bigcup_{\Lambda \in \mathcal{D}} \left\{\Lambda' \in \mathcal{S} \middle| [\![\Lambda' - \Lambda]\!] < \rho(\Lambda)\right\}$$

and furthermore for $\Lambda, \Lambda' \in \mathcal{D}$ satisfy

$$1_{\mathcal{S}} + (\Lambda + \Phi(G(\Lambda)))G(\Lambda) = 0 \quad \text{(Schwinger-Dyson equation)},$$

$$[\![G(\Lambda)]\!] \leq \frac{1}{\rho(\Lambda)} \text{ and } [\![G(\Lambda) - G(\Lambda')]\!] \leq \frac{[\![\Lambda - \Lambda']\!]}{\rho(\Lambda)\rho(\Lambda')},$$

$$|\rho(\Lambda) - \rho(\Lambda')| \leq [\![\Lambda' - \Lambda]\!] \quad \text{and} \quad \Im\Lambda > 0 \Rightarrow \frac{1}{\rho(\Lambda)} \leq \left[\!\left[(\Im\Lambda)^{-1}\right]\!\right].$$

All these properties with the exception of the Schwinger-Dyson equation are clear by the jack-knife lemma.

We have the following fundamental result which completes the verification that $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$ is a Schwinger-Dyson approximation scheme.

Theorem (Boltzmann-Fock space construction solutions of the Schwinger-Dyson equation)

For every $\Lambda \in \mathcal{D}$ we have

$$(\star)_{SD} \qquad \mathbf{I}_s + (\Lambda + \Phi(G(\Lambda)))G(\Lambda) = 0.$$

To do so we fix $\Lambda \in \mathcal{D}$ and study the operator

$$A = L_1 \otimes \Xi_1 + \cdots + L_m \otimes \Xi_m - \Lambda \otimes 1_{B(\mathcal{H})} \in \mathrm{Mat}_s(B(\mathcal{H})).$$

Semicircular specialization

Changing notation temporarily, let

$$G(z) = \text{upper left entry of } \begin{bmatrix} -z & 1 & \\ 1 & \ddots & \ddots \\ & \ddots & \end{bmatrix}^{-1},$$

which is the Stieltjes transform of the semicircle law. We have

$$\begin{aligned} G(z) &= \left(-z - \begin{bmatrix} 1 & 0 & \dots \end{bmatrix} \begin{bmatrix} -z & 1 & \\ 1 & \ddots & \ddots \\ & \ddots & \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix} \right)^{-1} \\ &= (-z - G(z))^{-1}, \text{ i.e., } 1 + (z + G(z))G(z) = 0, \end{aligned}$$

by the Schur complement formalism. "Self-similarity" is the key.

Proof of $(\star)_{SD}$

The proof is based on a certain block decomposition of operators in the C^*-algebra $\mathrm{Mat}_s(B(\mathcal{H}))$. First of all, we will identify $\mathrm{Mat}_s(B(\mathcal{H}))$ with an array of s-by-s matrices with rows and columns indexed by the nonnegative integers. More precisely, given $X \in \mathrm{Mat}_s(B(\mathcal{H}))$ we define the block $X\langle i,j\rangle \in \mathrm{Mat}_s$ by the rule

$$X\langle i,j\rangle[i',j'] = (v(i), X[i,'j']v(j))_{\mathcal{H}}.$$

In this notation we have

$$\begin{array}{rcl} A\langle i,i\rangle & = & -\Lambda \;\text{ for } i \geq 0, \\ G(\Lambda) & = & A^{-1}\langle 0,0\rangle, \\ A\langle 0,i\rangle = A\langle i,0\rangle & = & \left\{ \begin{array}{ll} L_i & \text{for } i = 1,\ldots,m, \\ 0 & \text{for } i > m. \end{array} \right. \end{array}$$

Proof of $(\star)_{SD}$

Let

$$L_\bullet = [\ L_1 \quad \dots \quad L_m\ ] \in \mathrm{Mat}_{s\times sm}.$$

The operator

$$A = L_1 \otimes \Xi_1 + \cdots + L_m \otimes \Xi_m - \Lambda \otimes 1_{B(\mathcal{H})}$$

then has the following remarkable self-similar structure:

$$A = \begin{bmatrix} -\Lambda & L_\bullet & & & \\ L_\bullet^* & -\Lambda \otimes \mathbf{I}_m & L_\bullet \otimes \mathbf{I}_m & & \\ & L_\bullet^* \otimes \mathbf{I}_m & -\Lambda \otimes \mathbf{I}_{m^2} & L_\bullet \otimes \mathbf{I}_{m^2} & \\ & & L_\bullet^* \otimes \mathbf{I}_{m^2} & \ddots & \ddots \\ & & & \ddots & \end{bmatrix}.$$

Proof of $(\star)_{SD}$ (completion)

Since the operator A is by hypothesis invertible and the operator

$$\begin{bmatrix} A\langle 1,1\rangle & \dots \\ \vdots & \ddots \end{bmatrix}^{-1}$$

is invertible by self-similarity, $G(\Lambda)$ is invertible and using Schur complements we have

$$\begin{aligned} & G(\Lambda) = A^{-1}\langle 0,0\rangle \\ = & \left(A\langle 0,0\rangle - \begin{bmatrix} A\langle 0,1\rangle & \dots \end{bmatrix} \begin{bmatrix} A\langle 1,1\rangle & \dots \\ \vdots & \ddots \end{bmatrix}^{-1} \begin{bmatrix} A\langle 1,0\rangle \\ \vdots \end{bmatrix} \right)^{-1}. \end{aligned}$$

Using the self-similar structure noted above we get the equation

$$G(\Lambda) = \left(-\Lambda - \sum_{\ell=1}^{m} L_\ell G(\Lambda) L_\ell \right)^{-1}$$

which just a rewrite of the Schwinger-Dyson equation. $\square$

Crossing t's and dotting i's

The Schur complement formalism can be rigorized in the operator-theoretic framework. So while the explanation above admittedly falls short of an i's-crossed-and-t's-dotted proof, it can be turned into an honest proof with only a modest amount of work. Alternatively see the textbook by Anderson-Guionnet-Zeitouni for another approach.

The Schwinger-Dyson equation part (b)

We now have a look at approximation theory for solutions of the Schwinger-Dyson equation. We take a formal axiomatic approach.

Schwinger-Dyson approximation schemes

- Let $\mathcal{S}$ be a finite-dimensional C^*-algebra.
- Let $\Phi \in B(\mathcal{S})$ be a linear map.
- Let $\mathcal{D} \subset \mathcal{S}$ be an open set.
- Let $\rho : \mathcal{D} \to (0, \infty)$ be a function.
- Let $G : \mathcal{D} \to \mathcal{S}$ be an analytic function.

We call $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$ a *Schwinger-Dyson approximation scheme* under the following conditions.

The quintuple $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$ must satisfy

$$\mathcal{D} = \{\Lambda \in \mathcal{S} \mid \Im\Lambda > 0\} \cup \bigcup_{\Lambda \in \mathcal{D}} \left\{\Lambda' \in \mathcal{S} \middle| [[\Lambda' - \Lambda]] < \rho(\Lambda)\right\}$$

and furthermore for $\Lambda, \Lambda' \in \mathcal{D}$ satisfy

$$1_{\mathcal{S}} + (\Lambda + \Phi(G(\Lambda)))G(\Lambda) = 0 \ \ \text{(Schwinger-Dyson equation)},$$
$$[\![G(\Lambda)]\!] \leq \frac{1}{\rho(\Lambda)} \text{ and } [[G(\Lambda) - G(\Lambda')]] \leq \frac{[\![\Lambda - \Lambda']\!]}{\rho(\Lambda)\rho(\Lambda')},$$
$$|\rho(\Lambda) - \rho(\Lambda')| \leq [[\Lambda' - \Lambda]] \ \text{ and } \ \Im\Lambda > 0 \Rightarrow \frac{1}{\rho(\Lambda)} \leq \left[\left[(\Im\Lambda)^{-1}\right]\right].$$

The semicircular example of a Schwinger-Dyson approximation scheme

- $\mathcal{S} = \mathbb{C}$.
- $\Phi = 1 \in B(\mathcal{S}) = \mathbb{C}$.
- $\mathcal{D} = \mathbb{C} \setminus [-2, 2]$.
- $\rho(z) = \inf_{t \in [-2,2]} |z - t| = \begin{cases} |z+1| & \text{if } \Re(z) \le -1, \\ \Im z & \text{if } |\Re(z)| < 1, \\ |z-1| & \text{if } \Re(z) \ge 1. \end{cases}$
- $G(z) = \frac{1}{2\pi} \int_{-2}^{2} \frac{\sqrt{4-t^2}}{t-z}\, dt =$ Stieltjes transform of semicircle.

In this example all the axioms are trivial to check except for the Schwinger-Dyson equation itself (and anyhow we already know that the latter holds).

In advance of proving a general approximation result, we first state what the bound specializes to in the semicircular case.

The tunnel estimate (preview in semicircle case)

As on the previous frame let

$$G(z) = \frac{1}{2\pi}\int_{-2}^{2} \frac{\sqrt{4-t^2}}{t-z}\,dt \quad (z \in \mathbb{C} \setminus [-2,2])$$

noting that $1 + (z + G(z))G(z) = 0$. We can specialize and simplify the general approximation result to be derived below, as follows. Let $F(z)$ be an analytic function defined in $\mathfrak{h}$ satisfying

$$|F(z)\Im z| \leq 1.$$

Define a function $E(z)$ analytic in $\mathfrak{h}$ by the formula

$$E(z) = 1 + (z + F(z))F(z).$$

Then for all $z \in \mathfrak{h}$ we have

$$\begin{aligned} & [\![F(z) - G(z)]\!] \\ \leq\ & 64\left(1 + \frac{1}{\Im z}\right)^2 \left(1 + \frac{1}{\inf_{t\in[-2,2]}|z-t|}\right)^2 \sup_{t\in[0,\infty)} |E(z + \mathrm{i}t)|. \end{aligned}$$

Trivial consequences of the axioms

Recall that we have

$$\llbracket G(\Lambda) \rrbracket \leq \frac{1}{\rho(\Lambda)} \text{ and } \llbracket G(\Lambda) - G(\Lambda') \rrbracket \leq \frac{\llbracket \Lambda - \Lambda' \rrbracket}{\rho(\Lambda)\rho(\Lambda')},$$
$$|\rho(\Lambda) - \rho(\Lambda')| \leq \llbracket \Lambda' - \Lambda \rrbracket$$

Consequently we have

$$\llbracket \Lambda - \Lambda' \rrbracket < \frac{\rho(\Lambda)}{2} \Rightarrow \rho(\Lambda') > \frac{\rho(\Lambda)}{2}$$
$$\Rightarrow \llbracket G(\Lambda') - G(\Lambda) \rrbracket \leq \frac{2\llbracket \Lambda' - \Lambda \rrbracket}{\rho(\Lambda)^2} < \frac{1}{\rho(\Lambda)} \text{ and } \llbracket G(\Lambda') \rrbracket < \frac{2}{\rho(\Lambda)}.$$

The "tunnel" $\mathcal{T}$ attached to $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$

Consider the following open subset of $\mathcal{S} \times \mathcal{S}$.

$$\mathcal{T} = \left\{ (F, \Lambda) \in \mathcal{S} \times \mathcal{D} \;\middle|\; \begin{array}{l} \Im\Lambda > 0 \Rightarrow [\![F]\!] < 2\left[\!\left[(\Im\Lambda)^{-1}\right]\!\right] \\ \text{and furthermore} \\ E = 1_{\mathcal{S}} + (\Lambda + \Phi(F))F \text{ satisfies} \\ [\![E]\!] < 1/2 \text{ and } 4[\![\Phi]\!][\![F]\!][\![E]\!]/\rho(\Lambda) < 1. \end{array} \right\}.$$

The set $\mathcal{T}$ is the *tunnel* naturally associated to the Schwinger-Dyson approximation scheme $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$.

For each $(F, \Lambda) \in \mathcal{T}$ we automatically have that

- $H = F(1 - E)^{-1} = -(\Lambda + \Phi(F))^{-1}$ is defined,
- $H - F = F(1 - E)^{-1}E$,
- $[\![H]\!] < 2[\![F]\!]$,
- $\Lambda + \Phi(F - H) = \Lambda - \Phi(F(1 - E)^{-1}E) \in \mathcal{D}$ and
- $[\![G(\Lambda + \Phi(F - H))]\!] < \frac{2}{\rho(\Lambda)}$.

The tunnel lemma

Lemma (Tunnel lemma)

Let $(F,\Lambda) \in \mathcal{T}$ satisfy $\Im\Lambda > 0$ and $8[\![\Phi]\!]\left[\!\left[(\Im\Lambda)^{-1}\right]\!\right]^2 < 1$. Put $H = -(\Lambda + \Phi(F))^{-1}$ and $\tilde{H} = G(\Lambda + \Phi(F - H))$. Then $H = \tilde{H}$.

Proof Using the Schwinger-Dyson equation and the definitions we have

$$H-\tilde{H} = -(\Lambda+\Phi(F))^{-1}+(\Lambda+\Phi(F-H)+\Phi(\tilde{H}))^{-1} = H\Phi(H-\tilde{H})\tilde{H}.$$

Since

$$[\![\Phi]\!][\![H]\!]\left[\!\left[\tilde{H}\right]\!\right] \leq [\![\Phi]\!](4\left[\!\left[(\Im\Lambda)^{-1}\right]\!\right])\left(\frac{2}{\rho(\Lambda)}\right) \leq 8[\![\Phi]\!]\left[\!\left[(\Im\Lambda)^{-1}\right]\!\right]^2 < 1,$$

we indeed have $H = \tilde{H}$. $\square$

The tunnel estimate (setup)

We continue working with the given Schwinger-Dyson approximation scheme $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$. We now fix a point

$$\Lambda_0 \in \mathcal{D}$$

and assume that

$$\Lambda_t := \Lambda_0 + \mathrm{i}t1_{\mathcal{S}} \in \mathcal{D}$$

for $t \in [0, \infty)$. Let

$$(t \mapsto F_t) : [0, \infty) \to \mathcal{S}$$

be a continuous map and assume that

$$\Im\Lambda_t > 0 \Rightarrow [\![F(\Lambda_t)]\!] \leq \left[\!\left[(\Im\Lambda_t)^{-1}\right]\!\right]$$

for $t \in [0, \infty)$.

The tunnel estimate (statement)

Put

$$E_t = 1_{\mathcal{S}} + (\Lambda_t + \Phi(F_t))F_t \in \mathcal{S}$$

for $t \in [0, \infty)$. Put

$$\begin{aligned} \mathfrak{C} &= 4(1 + [\![\Phi]\!]), \quad \mathfrak{G} = 1 + \sup_{t \in [0,\infty)} \frac{1}{\rho(\Lambda_t)}, \\ \mathfrak{F} &= 1 + \sup_{t \in [0,\infty)} [\![F_t]\!], \quad \mathfrak{E} = \sup_{t \in [0,\infty)} [\![E_t]\!]. \end{aligned}$$

All constants $\mathfrak{C}$, $\mathfrak{G}$, $\mathfrak{F}$ and $\mathfrak{E}$ are finite under our hypotheses.

Proposition (Tunnel bound for the Schwinger-Dyson equation)

Notation and assumptions are as above. We have

$$[\![F_0 - G(\Lambda_0)]\!] \leq \mathfrak{C}^2\mathfrak{F}^2\mathfrak{G}^2\mathfrak{E}.$$

Since otherwise the asserted bound holds trivially,we may assume that

$$\mathfrak{C}\mathfrak{F}\mathfrak{G}\mathfrak{E} < 1$$

in which case $(F_t, \Lambda_t) \in \mathcal{T}$ for all $t \in [0, \infty)$.By the tunnel lemma and our hypotheses, for $t \gg 0$ the pair (F_t, Λ_t) belongs to the connected component of $\mathcal{T}$ consisting of pairs (F, Λ) satisfying

$$-(\Lambda + \Phi(F))^{-1} = G(\Lambda + \Phi(F + (\Lambda + \Phi(F))^{-1})),$$

and thus (F_t, Λ_t) belongs to that connected component for all $t \in [0, \infty)$. It remains now only to estimate the difference

$$F_0 - G(\Lambda_0) = G(\Lambda_0 - \Phi(F_0(1 - E_0)E_0^{-1}))(1 - E_0) - G(\Lambda_0)$$

in order to get the result.

Concluding calculation

We have

$$
\begin{aligned}
& \left[\!\left[G(\Lambda_0 - \Phi(F_0(1-E_0)^{-1}E_0))(1-E_0) - G(\Lambda_0) \right]\!\right] \\
\leq \; & \left[\!\left[G(\Lambda_0 - \Phi(F_0(1-E_0)^{-1}E_0)) \right]\!\right] [\![E_0]\!] \\
& + \left[\!\left[G(\Lambda_0 - \Phi(F_0(1-E_0)^{-1}E_0)) - G(\Lambda_0) \right]\!\right] \\
\leq \; & \frac{2[\![E_0]\!]}{\rho(\Lambda_0)} + \frac{4[\![\Phi]\!][\![F_0]\!][\![E_0]\!]}{\rho(\Lambda_0)^2} \\
\leq \; & 2\mathfrak{E}\mathfrak{G} + 4[\![\Phi]\!]\mathfrak{F}\mathfrak{E}\mathfrak{G}^2 \leq \mathfrak{C}^2\mathfrak{F}^2\mathfrak{G}^2\mathfrak{E},
\end{aligned}
$$

which finishes the proof. □

Remark There is plenty of room for sharpening this result, but as it stands it is enough for proving HST.

Notation: Tensor products etc.

- For an algebra $\mathcal{A}$, let $\mathrm{GL}_n(\mathcal{A})$ denote the group of invertible elements of $\mathrm{Mat}_n(\mathcal{A})$.
  - GL is short for the *general linear group*.
- Given $X \in \mathrm{Mat}_{k\times\ell}$ and $a \in \mathcal{A}$, let $X \otimes a \in \mathrm{Mat}_{k\times\ell}$ be defined by $(X \otimes a)[i,j] = X[i,j]a$.
- Note that $(X \otimes a)^* = X^* \otimes a^*$.
- Note also that if $a \in \mathrm{Mat}_N$, then $X \otimes a$ is the usual Kronecker product of matrices, i.e.,

$$X \otimes a = \begin{bmatrix} X[1,1]a & \dots & X[1,\ell]a \\ \vdots & & \vdots \\ X[k,1]a & \dots & X[k,\ell]a \end{bmatrix}.$$

The linearization trick

We next discuss the *linearization trick*. Strikingly enough, Schur complements are as important for setting up the linearization trick as they were for proving that the Schwinger-Dyson equation holds.

Proposition (The linearization trick)

Fix any $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$. (i) Then there exists an integer $s > n$ and matrices $L_0, L_1, \ldots, L_m \in \mathrm{Mat}_s$ such that we have a block decomposition

$$L_0 + L_1 \otimes \mathbf{X}_1 + \cdots + L_m \otimes \mathbf{X}_m = \begin{bmatrix} 0_n & P \\ Q & U \end{bmatrix}$$

where the upper left zero block is n-by-n, the lower $(s-n)$-by-$(s-n)$ block U belongs to $\mathrm{GL}_{s-n}(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ and $f = -PU^{-1}Q$.

We call $L_0 + \sum_{\ell=1}^{m} L_\ell \otimes X_\ell \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ a *linearization* of f. We present the proof below after examples and amplification.

What is the point?

The point of the definition is that one automatically has a factorization

$$\begin{bmatrix} 0_n & P \\ Q & U \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & PU^{-1} \\ 0 & \mathbf{I}_{s-n} \end{bmatrix} \begin{bmatrix} f & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & 0 \\ U^{-1}Q & \mathbf{I}_{s-n} \end{bmatrix}.$$

Example

$$\begin{bmatrix} 0 & \mathbf{X}_1 & \mathbf{X}_2 \\ \mathbf{X}_1 & 0 & -\mathrm{i} \\ \mathbf{X}_2 & \mathrm{i} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \otimes \mathbf{X}_1 + \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \otimes \mathbf{X}_2$$

is a linearization of

$$\mathrm{i}(\mathbf{X}_1\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1) = -\begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

Thus

$$\begin{bmatrix} 0 & \mathbf{X}_1 & \mathbf{X}_2 \\ \mathbf{X}_1 & 0 & -\mathrm{i} \\ \mathbf{X}_2 & \mathrm{i} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \mathrm{i}\mathbf{X}_2 & -\mathrm{i}\mathbf{X}_1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{i}(\mathbf{X}_1\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1) & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -\mathrm{i}\mathbf{X}_2 & 1 & 0 \\ \mathrm{i}\mathbf{X}_1 & 0 & 1 \end{bmatrix}.$$

Amplification: preservation of self-adjointness

Corollary

Every self-adjoint $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ *has a self-adjoint linearization.*

Proof Write $f = g + g^*$.Let

$$\begin{bmatrix} 0 & P \\ Q & U \end{bmatrix}$$

be a linearization of g. Then

$$\begin{bmatrix} 0 & P & Q^* \\ P^* & 0 & U^* \\ Q & U & 0 \end{bmatrix}$$

is a linearization of $f = g + g^*$. $\square$

Check

$$
\begin{aligned}
g &= -PU^{-1}Q, \\
& -\begin{bmatrix} P & Q^* \end{bmatrix} \begin{bmatrix} 0 & U^* \\ U & 0 \end{bmatrix}^{-1} \begin{bmatrix} P^* \\ Q \end{bmatrix} \\
&= -\begin{bmatrix} P & Q^* \end{bmatrix} \begin{bmatrix} 0 & U^{-1} \\ (U^*)^{-1} & 0 \end{bmatrix} \begin{bmatrix} P^* \\ Q \end{bmatrix} \\
&= -PU^{-1}Q - Q^*(U^*)^{-1}P^* = g + g^* = f.
\end{aligned}
$$

Corollary (Schur complement $\leftrightarrow$ linearization relations)

Fix $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\dots m}\rangle)$ *and a linearization* $\tilde{f} = \begin{bmatrix} 0_n & P \\ Q & U \end{bmatrix} \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}_{1\dots m}\rangle)$ *thereof. Put* $\overline{P} = -PU^{-1}$, $\overline{Q} = -U^{-1}Q$ *and* $\overline{U} = U^{-1}$. *Fix an algebra* $\mathcal{A}$, *an* m*-tuple* $a \in \mathcal{A}^m$ *and a matrix* $\Lambda \in \mathrm{Mat}_n(\mathcal{A})$. *Put* $\tilde{\Lambda} = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \in \mathrm{Mat}_s(\mathcal{A})$. *Then we have* $f(a) - \Lambda \in \mathrm{GL}_n(\mathcal{A})$ *if and only if* $\tilde{f}(a) - \tilde{\Lambda} \in \mathrm{GL}_s(\mathcal{A})$, *and under these equivalent conditions we further have*

$$\left(\tilde{f}(a) - \tilde{\Lambda}\right)^{-1} = \begin{bmatrix} 1 & 0 \\ \overline{Q}(a) & 1 \end{bmatrix} \begin{bmatrix} (f(a) - \Lambda)^{-1} & 0 \\ 0 & \overline{U}(a) \end{bmatrix} \begin{bmatrix} 1 & \overline{P}(a) \\ 0 & 1 \end{bmatrix}.$$

Proof Just specialize the Schur complement formulas given above.

Corollary (Linearization bounds)

In the setup of the proposition, suppose that $\mathcal{A}$ is a C^*-algebra. Then we have

$$\begin{aligned}
\left[\!\left[(\tilde{f}(a) - \tilde{\Lambda})^{-1}\right]\!\right] &\leq c_1 \left(1 + \sum_{\ell=1}^{m} [\![a_\ell]\!]\right)^{c_2} \left(1 + \left[\!\left[(f(a) - \Lambda)^{-1}\right]\!\right]\right), \\
\left[\!\left[(f(a) - \Lambda)^{-1}\right]\!\right] &\leq c_1 \left(1 + \sum_{\ell=1}^{m} [\![a_\ell]\!]\right)^{c_2} \left[\!\left[(\tilde{f}(a) - \tilde{\Lambda})^{-1}\right]\!\right]
\end{aligned}$$

for constants c_1 and c_2 depending only on f and $\tilde{f}$.

- Let $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ and a self-adjoint linearization $L_0 + \sum_{\ell=1}^m L_\ell \otimes \mathbf{X}_\ell \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ be given.

- Put $\Lambda^{(z)} = \left[\begin{array}{cc} z\mathbf{I}_n & 0 \\ 0 & 0 \end{array}\right] - L_0 \in \mathrm{Mat}_s$ and $L = \sum_{\ell=1}^m L_\ell \otimes \mathbf{X}_\ell$.

- Let $(\mathcal{A}, \phi)$ be a C^*-probability space and fix $a = (a_1, \ldots, a_m) \in \mathcal{A}_{\mathrm{sa}}^m$. Put $A = f(a) \in \mathrm{Mat}_n(\mathcal{A})_{\mathrm{sa}}$.

- Then we have the *linearized representation* of the Stieltjes transform of the law μ_A, namely

$$S_{\mu_A}(z) = \frac{1}{n}\sum_{i=1}^n \phi\left((L(a) - \Lambda^{(z)} \otimes 1_{\mathcal{A}})^{-1}[i,i]\right).$$

Linearized representation of Stieltjes transforms of empirical distributions

- Let $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ and a self-adjoint linearization $L_0 + \sum_{\ell=1}^m L_\ell \otimes \mathbf{X}_\ell \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ be given.
- Put $\Lambda^{(z)} = \begin{bmatrix} z\mathbf{I}_n & 0 \\ 0 & 0 \end{bmatrix} - L_0 \in \mathrm{Mat}_s$ and $L = \sum_{\ell=1}^m L_\ell \otimes \mathbf{X}_\ell$.
- Fix $a = (a_1, \ldots, a_n) \in \mathrm{Herm}_N^m$ and put $A = f(a) \in \mathrm{Herm}_{nN}$.
- Then we have the *linearized representation* of the Stieltjes transform of the empirical distribution μ_A of eigenvalues of A, namely

$$S_{\mu_A}(z) = \frac{1}{nN} \sum_{i=1}^{nN} (L(a) - \Lambda^{(z)} \otimes \mathbf{I}_N)^{-1}[i,i].$$

Proof of the proposition (existence of linearizations)

We break the proof into two main steps and a short concluding argument. The proof is simple and it yields a very tractable algorithm for generating linearizations.

Additive construction of linearizations

Suppose for $j = 1, \ldots, k$ we are given $f_j \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ along with linearizations

$$\tilde{f}_j = \begin{bmatrix} 0 & P_j \\ Q_j & U_j \end{bmatrix} \in \mathrm{Mat}_{s_j}(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$$

thereof.

Then

$$\begin{bmatrix} & P_1 & \ldots & P_k \\ Q_1 & U_1 & & \\ \vdots & & \ddots & \\ Q_k & & & U_k \end{bmatrix}$$

is a linearization of

$$f_1 + \cdots + f_k \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle).$$

Let's check

$$
-\begin{bmatrix} P_1 & \dots & P_k \end{bmatrix}
\begin{bmatrix} U_1^{-1} & & \\ & \ddots & \\ & & U_k^{-1} \end{bmatrix}
\begin{bmatrix} Q_1 \\ \vdots \\ Q_k \end{bmatrix}
$$

$$
= -\sum_{i=1}^{k} P_i U_i^{-1} Q_i = \sum_{i=1}^{k} f_i = f
$$

Multiplicative construction of linearizations

Again, suppose for $j = 1, \ldots, k$ we are given $f_j \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ which are linear, i.e., have all entries of degree 1 or less in $\mathbf{X}_1, \ldots, \mathbf{X}_m$. Suppose furthermore that $k \geq 2$.

Then

$$\begin{bmatrix} & & & f_1 \\ & & f_2 & -\mathbf{I}_n \\ & \iddots & \iddots & \\ f_k & -\mathbf{I}_n & & \end{bmatrix}$$

is a linearization of $f_1 \cdots f_k$.

Let's check

$$\begin{bmatrix} & & f_2 & -\mathbf{I}_n \\ & \iddots & \iddots & \\ f_{k-1} & \iddots & & \\ -\mathbf{I}_n & & & \end{bmatrix}^{-1} = \begin{bmatrix} & & & -\mathbf{I}_n \\ & & \iddots & -f_{k-1} \\ & \iddots & \iddots & \vdots \\ -\mathbf{I}_n & -f_2 & \dots & -f_2 \cdots f_{k-1} \end{bmatrix}$$

End of the proof

The algebra $\mathrm{Mat}_{\boldsymbol{n}}(\mathbb{C}\langle \mathbf{X} \rangle)_{\mathrm{sa}}$ is generated over $\mathbb{C}$ by its elements of the form e_{ij} and $\mathbf{I}_{\boldsymbol{n}} \otimes \mathbf{X}_\ell$. Thus the additive and multiplicative constructions together prove existence of linearizations in general. $\square$

Amplification: More general multiplicative construction of linearizations

Again, suppose for $j = 1, \ldots, k$ we are given $f_j \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ along with linearizations $\tilde{f}_j = \left[\begin{array}{cc} 0 & P_j \\ Q_j & U_j \end{array}\right] \in \mathrm{Mat}_{s_j}(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ thereof.

Then

$$\left[\begin{array}{cccccc} & & & & & P_1 \\ & & & & Q_1 & U_1 \\ & & & P_2 & -\mathbf{I}_n & \\ & & Q_2 & U_2 & & \\ & \iddots & \iddots & & & \\ & P_k & -\mathbf{I}_n & & & \\ Q_k & U_k & & & & \end{array}\right]$$

is a linearization of

$$f_1 \cdots f_k \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle).$$

Let's check

By induction we just have to check $k = 2$.

$$\begin{bmatrix} 0 & Q_1 & U_1 \\ P_2 & -1 & 0 \\ U_2 & 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} & & U_2^{-1} \\ & -1 & P_2 U_2^{-1} \\ U_1^{-1} & U_1^{-1} Q_1 & -U_1^{-1} Q_1 P_2 U_2^{-1} \end{bmatrix}$$

Example of linearization

Let us find a self-adjoint linearization of

$$(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3 - w)(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_4 - w)^*$$

where w is a fixed complex number.

Well,

$$\begin{bmatrix} 0 & 0 & \mathbf{X}_1 \\ 0 & \mathbf{X}_2 & -1 \\ \mathbf{X}_3 & -1 & 0 \end{bmatrix}$$

is a linearization of $\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$, and

$$\begin{bmatrix} 0 & 1 \\ w & -1 \end{bmatrix}$$

is a linearization of w, hence...

Example of linearization (continued)

... the matrix

$$\begin{bmatrix} 0 & 1 & 0 & \mathbf{X}_1 \\ w & -1 & 0 & 0 \\ 0 & 0 & \mathbf{X}_2 & -1 \\ \mathbf{X}_3 & 0 & -1 & 0 \end{bmatrix}$$

is a linearization of $\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3 - w$ and finally...

Example of linearization (concluded)

the matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{X}_1 \\ 0 & 0 & 0 & 0 & w & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{X}_2 & -1 \\ 0 & 0 & 0 & 0 & \mathbf{X}_3 & 0 & -1 & 0 \\ 0 & w^* & 0 & \mathbf{X}_3 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{X}_2 & -1 & 0 & 0 & 0 & 0 \\ \mathbf{X}_1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is a self-adjoint linearization of

$$(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3 - w)(\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3 - w)^*$$

as desired.

The two-variable estimate (setup)

Let $\mathcal{A}$ be a C^*-algebra. Fix $L_0, \ldots, L_m \in \mathrm{Herm}_s$ such that

$$L_0 + \sum_{\ell=1}^{m} L_\ell \otimes \mathbf{X}_\ell = \begin{bmatrix} 0_n & P \\ Q & U \end{bmatrix}$$

is a self-adjoint linearization of

$$f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}.$$

Put

$$L = \sum_{\ell=1}^{m} L_\ell \otimes \mathbf{X}_\ell$$

and for $z \in \mathfrak{h}$ and $t \geq 0$ put

$$\Lambda^{(z,t)} = \begin{bmatrix} z\mathbf{I}_n & 0 \\ 0 & 0 \end{bmatrix} + t\mathbf{I}_s - L_0.$$

The two-variable estimate (statement of result)

Proposition

Let $\mathcal{A}$ be any C^*-algebra and let $(a_1, \ldots, a_m) \in \mathcal{A}_{\mathrm{sa}}^m$ be any m-tuple of self-adjoint elements. Then

$$L(a) - \Lambda^{(z,t)} \otimes 1_{\mathcal{A}} \in \mathrm{GL}_s(\mathcal{A})$$

and we have

$$\left[\!\left[\left(L(a) - \Lambda^{(z,t)} \otimes 1_{\mathcal{A}}\right)^{-1}\right]\!\right] \leq \frac{1}{t} \wedge c_1 \left(1 + \sum_{\ell=1}^{m} [\![a_\ell]\!]\right)^{c_2} \left(1 + \frac{1}{\Im z}\right)$$

for constants c_1 and c_2 depending only on f and L.

By abuse of notation often repeated below $\frac{1}{t}\big|_{t=0} = +\infty$.

Proof of two-variable estimate

For $t = 0$ we have invertibility by Schur complement$\leftrightarrow$linearization relationship and we have

$$\left[\!\left[\left(L(a) - \Lambda^{(z,0)} \otimes 1_{\mathcal{A}}\right)^{-1}\right]\!\right] \leq \frac{c_1}{2}\left(1 + \sum_{\ell=1}^{m} [\![a_\ell]\!]\right)^{c_2}\left(1 + \frac{1}{\Im z}\right)$$

by the linearization bounds and the $\Im$-lemma. For $t > 0$ we have invertibility of

$$L(a) - \Lambda^{(z,t)} \otimes 1_{\mathcal{A}}$$

and the bound

$$\left[\!\left[\left(L(a) - \Lambda^{(z,t)} \otimes 1_{\mathcal{A}}\right)^{-1}\right]\!\right] \leq \frac{1}{t}.$$

by the $\Im$-lemma.

Proof of two-variable estimate (concluded)

For

$$tc_1 \left(1 + \sum_{\ell=1}^{m} [\![a_\ell]\!]\right)^{c_2} \left(1 + \frac{1}{\Im z}\right) < 1$$

we have invertibility of

$$L(a) - \Lambda^{(z,t)} \otimes 1_{\mathcal{A}}$$

and the bound

$$\left[\!\left[\left(L(a) - \Lambda^{(z,t)} \otimes 1_{\mathcal{A}}\right)^{-1}\right]\!\right] \leq c_1 \left(1 + \sum_{\ell=1}^{m} [\![a_\ell]\!]\right)^{c_2} \left(1 + \frac{1}{\Im z}\right)$$

by the jack-knife. In the remaining case there is actually nothing left to prove.

Our next goal is. . .

. . . is to present the $\bar{\partial}$-trick, which is a technical device borrowed from the Helffer-Sjöstrand functional calculus [Helffer-Sjöstrand 1989] and then to apply it in various ways so as to illuminate the properties of a certain nice class of functions of several hermitian matrices. Ideas from the Helffer-Sjöstrand functional calculus have been showing up in other (perhaps not so distant?) parts of RMT lately. See for example the papers [O'Rourke-Renfrew-Soshnikov 2011 A] and [O'Rourke-Renfrew-Soshnikov 2011 B] which have to do with entries of regular functions of Wigner matrices.

Technical lemma: the raw $\bar{\partial}$-trick

Lemma

Fix a smooth function $\varphi : \mathbb{R} \to \mathbb{C}$ *along with an integer* $\gamma > 0$ *and a constant* $\epsilon > 0$*. Then there exists a function*

$$\Gamma = \Gamma_{\varphi,\gamma,\epsilon} : \mathbb{R}^2 \to \mathbb{C}$$

such that (i) Γ *is smooth, (ii)* Γ *is supported in the set* $(\operatorname{supp}\varphi) \times [-\epsilon, \epsilon]$*, (iii) if* φ *is real-valued then* Γ *has the symmetry* $\Gamma(x, y)^* \equiv \Gamma(x, -y)$*, (iv)* Γ *factors as* y^γ *times a smooth function on* $\mathbb{R}^2$ *and (v)* $\Gamma = -\frac{1}{2\pi}\left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right)\widetilde{\Gamma}$ *for some smooth function* $\tilde{\Gamma} : \mathbb{R}^2 \to \mathbb{C}$ *also supported in* $(\operatorname{supp}\varphi) \times [-\epsilon, \epsilon]$ *satisfying* $\widetilde{\Gamma}(x, 0) \equiv \varphi(x)$*.*

Proof Let $\theta : \mathbb{R} \to [0,1]$ be an even smooth function supported in $[-\epsilon, \epsilon]$ and identically equal to 1 on $[-\epsilon/2, \epsilon/2]$. Put

$$\widetilde{\Gamma}(x,y) = \theta(y) \sum_{j=0}^{\gamma} \frac{(\mathrm{i}y)^j}{j!} \varphi^{(j)}(x).$$

Let

$$\begin{aligned} -2\pi\Gamma(x,y) &= \left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right)\widetilde{\Gamma}(x,y) \quad \text{(Remark: } 2\bar{\partial} = \tfrac{\partial}{\partial x} + \mathrm{i}\tfrac{\partial}{\partial y}\text{)} \\ &= \theta(y)\frac{(\mathrm{i}y)^\gamma}{\gamma!}\varphi^{(\gamma+1)}(x) + \mathrm{i}\theta'(y)\sum_{j=0}^{\gamma}\frac{(\mathrm{i}y)^j}{j!}\varphi^{(j)}(x). \end{aligned}$$

The functions Γ and $\tilde{\Gamma}$ clearly have all the desired properties. $\square$

Proposition

With φ, γ, ϵ and Γ as above, for any open set $D \supset (\operatorname{supp} \varphi) \times [-\epsilon, \epsilon]$ and solution $b : D \to \mathbb{C}$ of the Cauchy-Riemann equations, we have $\iint_D b(x, y)\Gamma(x, y)\, dx\, dy = 0$.

Since b is a solution of the Cauchy-Riemann equations, it is an analytic function under the usual identification $\mathbb{R}^2 = \mathbb{C}$. The proposition is proved by a straightforward application of Green's theorem. We omit the proof because it is similar to the proof of the next proposition.

$\bar{\partial}$-trick (continued)

Proposition

With φ, γ, ϵ and Γ as above, for $t \in \mathbb{R}$ and $0 \le k < \gamma$, we have

$$\frac{\varphi^{(k)}(t)}{k!} = \frac{1}{2\pi \mathrm{i}} \int \frac{\Gamma(x,y)}{(x + \mathrm{i}y - t)^k} \, dx \, dy.$$

Proof We may assume $k = 1$ because differentiation under the integral is permissible. We may also assume that $t = 0$.

Fix $C \gg 0$ such that

$$\operatorname{supp}\varphi \times [-\epsilon, \epsilon] \subset \{(x, y) \mid \sqrt{x^2 + y^2} < C\}.$$

For small $c > 0$ let

$$T_c = \{(x, y) \in \mathbb{R}^2 \mid c \leq \sqrt{x^2 + y^2} \leq C\}.$$

Then we have

$$\begin{aligned}
\iint \frac{\Gamma(x, y)}{x + \mathrm{i}y}\, dx\, dy \;&=\; \lim_{c \to 0} \iint_{T_c} \frac{\Gamma(x, y)}{x + \mathrm{i}y}\, dx\, dy \\
&\overset{\text{Green's theorem}}{=} \lim_{c \to 0} -\frac{1}{2\pi\mathrm{i}} \oint_{\partial T_c} \widetilde{\Gamma} \frac{dx + \mathrm{i}\, dy}{x + \mathrm{i}y} \\
&= \lim_{c \to \infty} \frac{1}{2\pi} \int_0^{2\pi} \widetilde{\Gamma}(c\cos\theta, c\sin\theta)\, d\theta = \varphi(0).
\end{aligned}$$

$\square$

The $\bar{\partial}$-trick (continued)

The $\bar{\partial}$-trick gives us the following method to reconstruct probability measures from their Stieltjes transforms. We will use this reconstruction method heavily in the sequel.

Proposition (Reconstruction of probability measures from their Stieltjes transforms)

Let φ, γ, ϵ and Γ be as above, but now assume that φ is real-valued. For any probability measure μ on the real line we have $\int \varphi \, d\mu = -2\Re \iint_{y>0} S_\mu(x+\mathrm{i}y)\Gamma(x,y)\, dx\, dy$.

Proof Combine the preceding result with Fubini's theorem and exploitation of symmetry. $\square$

Given $X \in \mathrm{Herm}_N$ and and writing

$$X = U \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix} U^* \quad (U \in \mathrm{Mat}_N : \text{ unitary})$$

one defines

$$\varphi(X) = U \begin{bmatrix} \varphi(\lambda_1) & & \\ & \ddots & \\ & & \varphi(\lambda_N) \end{bmatrix} U^*$$

for any continuous function $\varphi : \mathbb{R} \to \mathbb{C}$.This is merely a specialization of the functional calculus in C^*-algebras to Mat_N. The novelty in what follows is that we are going to differentiate these functions with respect to matrix entries.

Functional calculus in relation to empirical distributions

For functions $\varphi : \mathbb{R} \to \mathbb{C}$ and hermitian matrices $A \in \mathrm{Herm}_N$, letting μ_A denote the empirical distribution of eigenvalues of A, we have

$$\operatorname{tr} \varphi(A) = N \int \varphi d\mu_A.$$

This fact is used repeatedly in the sequel. The two notations express complementary points of view.

The $\bar{\partial}$-trick (continued)

The next simple application is exactly in the spirit of the Helffer-Sjöstrand functional calculus.

Proposition

With φ, γ, ϵ and Γ as above, for $X \in \mathrm{Herm}_N$ and integers $0 \leq j < \gamma$ we have

$$\frac{\varphi^{(j)}(X)}{j!} = \iint ((x + \mathrm{i}y)\mathbf{I}_N - X)^{-(j+1)}\,\Gamma(x,y)\,dx\,dy.$$

Proof We may assume that X is diagonal. Then there is nothing to prove. $\square$

The $\bar{\partial}$-trick (continued)

For $H, X \in \mathrm{Herm}_N$ and smooth $\Phi : \mathrm{Mat}_N \to \mathbb{C}$ put

$$\partial_H \Phi(X) = \frac{d}{dt}\Phi(X + tH)\Big|_{t=0}.$$

Proposition

For any smooth $\varphi : \mathbb{R} \to \mathbb{C}$ and $H, X \in \mathrm{Herm}_N$, we have $\partial_H(\mathrm{tr}\varphi(X)) = \mathrm{tr}(H\varphi'(X))$.

Proof We may assume that φ is compactly supported. Then use the previous proposition and differentiate under the integral. □

The $\bar{\partial}$-trick (continued)

Proposition

With φ, γ, ϵ and Γ as above, for integers $0 \le k < \gamma$ and points $t_0, \ldots, t_k \in \mathbb{R}$ we have

$$\iint \frac{\Gamma(x,y)}{\prod_{j=0}^{k}(x + \mathrm{i}y - t_j)}\,dx\,dy = \frac{1}{k!}\mathbf{E}\varphi^{(k)}\left(\sum_{j=0}^{k} t_j U_j\right)$$

where the random vector $(U_0, \ldots, U_k)$ is uniformly distributed in the k-simplex

$$\left\{(u_0, \ldots, u_k) \in \mathbb{R}^{k+1} \,\middle|\, u_0, \ldots, u_k > 0,\ \sum_{j=0}^{k} u_j = 1\right\}.$$

Proof

The expressions on both sides depend continuously on $t_0, \ldots, t_k$ by dominated convergence. We may therefore assume that $t_0, \ldots, t_k$ are distinct. Without loss of generality we may assume that only t_0 is contained in the support of φ. We have the following variant of Taylor's theorem with remainder:

$$\varphi(x) = \frac{1}{k!}\mathbf{E}\varphi^{(k)}\left(U_0 x + \sum_{i=1}^{k} U_i t_i\right)\prod_{i=1}^{k}(x - t_i).$$

This version of Taylor's theorem granted, the result follows by a partial fraction calculation and the previous proposition. We omit further details. (A messier but perhaps clearer direct proof is also possible by comparing recursions obeyed on both sides.)

The $\bar{\partial}$-trick (continued)

Proposition

Let $\varphi : \mathbb{R} \to \mathbb{C}$ be any smooth function. Fix $H_0 \in \mathrm{Herm}_N$ and let

$$\Phi : \mathrm{Herm}_N \to \mathbb{C}$$

be defined by

$$\Phi(X) = \mathrm{tr}(H_0 \varphi(X)).$$

Then Φ is a smooth function and we have

$$|(\partial_{H_1} \cdots \partial_{H_k} \Phi)(X)| \leq \sup_{|x| \leq [\![X]\!]} |\varphi^{(k)}(x)| \cdot \prod_{i=0}^{k} [\![H_i]\!]_1$$

for all integers $k > 0$ and all $X, H_1, \ldots, H_k \in \mathrm{Herm}_N$.

We need one lemma before giving the proof.

The $\bar{\partial}$-trick (continued)

Lemma

In the setting of the preceding proposition we have

$$\sum_{\substack{i_0,\dots,i_{k+1}\in\{1,\dots,N\}\\ s.t.\ i_0=i_{k+1}}} \prod_{\alpha=0}^{k} |H_\alpha[i_\alpha, i_{\alpha+1}]| \le \prod_{\alpha=0}^{k} [\![H_\alpha]\!]_1.$$

Proof Write $H_\alpha = U_\alpha \Lambda_\alpha U_\alpha^*$ where U_α is unitary and Λ_α is diagonal. Let D_α be the diagonal matrix with $|\Lambda_\alpha[i,i]| = D_\alpha[i,i]$.Let $W_\alpha[i,j] = \sum_{k=1}^{N} |U_\alpha[i,k]U_{\alpha+1}[k,j]|$ where we put $U_{k+1} = U_0$. All entries of W_α belong to the unit interval.Clearly $\mathrm{tr}(D_0 W_0 \cdots D_k W_k)$ dominates the left side of the claimed inequality and is dominated by the right side. $\square$

Proof of the proposition

Fix $\Gamma(x, y)$ as in the $\bar{\partial}$-trick with γ taken very large. We may assume without loss of generality that X is diagonal with diagonal entries $X[i, i] = \lambda_i$. For large γ differentiation under the integral is permissible and we obtain the representation

$$\begin{aligned} & (\partial_{H_1} \cdots \partial_{H_k} \Phi)(X) \\ = \; & \mathrm{tr} \iint \Gamma(x, y) H_0 (x + \mathrm{i}y - X)^{-1} \cdots H_k (x + \mathrm{i}y - X)^{-1} \, dx \, dy \end{aligned}$$

for the partial derivative in question. The formula can be used to check continuity of the partial derivatives. Since k and $H_1, \ldots, H_k$ are arbitrary, the function Φ is indeed smooth. Finally, one can develop the expression on the right side as a sum with terms of the form handled by the preceding proposition, and using the preceding lemma one can verify the claimed estimate. $\square$

Lecture 5 (June 22, 2012)

Remarks toward “fake” HST

We finish up the discussion of the “fake” Poincaré inequality.

Drastic simplification of IBPI (statement)

- As before let Z be a real random variable with absolute moments of all orders.
- Fix an integer $k \geq 2$ and assume that Z has the same first k moments as a centered Gaussian random variable. (Hereafter such is called *Gaussian through the k^{th} order*.)
- Let $f : \mathbb{R} \to \mathbb{C}$ be tame. Put $\mathcal{M}^k[f](x) = \sup_{t \in [0,1]} |f^{(k)}(tx)|$.

Proposition (Drastically simplified IBPI)

Assumptions and notation are as above. We have

$$\left|\mathbf{E}\left(Zf(Z) - \mathbf{Var}(Z)f'(Z)\right)\right| \leq 2\|Z\|_{2(k+1)}^{k+1}\left\|\mathcal{M}^k[f](Z)\right\|_2.$$

Proof of drastically simplified IBPI

By hypothesis

$$\kappa_j(Z) = \delta_{j2}\mathbf{Var}(Z) \text{ for } j = 1, \ldots, k.$$

Thus IBPI simplifies in the present case to

$$\begin{aligned} & \mathbf{E}(Zf(Z) - f'(Z)) \\ = \quad & \mathbf{E}\left(\frac{Z^{k+1}}{k!} f^{(k)}(U_k Z) - \mathbf{Var}(Z)\frac{Z^{k-1}}{(k-1)!} f^{(k)}(U_{k-1} Z)\right). \end{aligned}$$

Now make the obvious application of the Hölder inequality and the definitions. $\square$

Harmless multivariable generalization

We consider a similar setup but in several dimensions.

- Let $Z = (Z_1, \ldots, Z_n) \in \mathbb{R}^n$ be a random vector with independent entries all of which have absolute moments of all orders.
- Fix an integer $k \geq 2$ and assume that all entries of Z are Gaussian through the k^{th} order and furthermore, for simplicity, assume that all entries have the same variance σ^2.
- Let $f : \mathbb{R}^n \to \mathbb{C}$ be an infinitely differentiable function such that partial derivatives of all orders have polynomial growth. (Hereafter such is called *tame*.)
- Let D_i denote differentiation with respect to the i^{th} coordinate.
- Generalizing from the one-dimensional case, we define $\mathcal{M}^k[f] : \mathbb{R}^n \to [0, \infty)$ by the formula $\mathcal{M}^k[f](x) = \max_{i=1}^n \sup_{u \in [0,1]} |D_i^k[f](x_1, \ldots, x_{i-1}, ux_i, x_{i+1}, \ldots, x_n)|$.

Multivariate integration by parts

We then have the following result.

Proposition (Multivariate drastically simplified integration by parts)

Notation and assumptions are as above. We have

$$\max_{i=1}^{n} \left|\mathbf{E}(Z_i f(Z) - \sigma^2 D_i[f](Z))\right| \leq 2 \left(\max_{i=1}^{n} [\![Z_i]\!]_{2(k+1)}^{k+1} \right) \left[\!\left[\mathcal{M}^k[f](Z) \right]\!\right]_2 .$$

Proof The proof is just an exercise in conditional expectation. □

A Poincaré-type inequality (setup)

We work with a setup very similar to the previous one.

- Let $Z \in \mathbb{R}^n$ be a random vector with independent entries all of which have absolute moments of all orders and fix i.i.d. copies $X, Y \in \mathbb{R}^n$ of Z.
- For some integer $k \geq 2$ and constant $\sigma^2 > 0$, assume that all entries of Z are Gaussian through the k^{th} order and moreover have variance σ^2.
- Let $f : \mathbb{R}^n \to \mathbb{C}$ be tame.
- For integers $r \geq 0$ we define

$$\mathcal{Q}^r[f], \mathcal{P}^k[f] : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$$

by

$$\mathcal{Q}^r[f](x, y) = 2^r \sup_{u \in [0,1]} \sup_{t \in [0,1]} \max_{i=1}^{n}$$
$$|D_i^r[f](\sqrt{t}\,(x_1, \ldots, x_{i-1}, ux_i, x_{i+1}, \ldots, x_n) + \sqrt{1-t}\,y)|,$$
$$\mathcal{P}^k[f](x, y) = 2 \sum_{r=0}^{k} \mathcal{Q}^r[f](x, y) \mathcal{Q}^{k+1-r}[f](x, y).$$

A Poincaré-type inequality in the presence of k Gaussian moments

Proposition

Notation and assumptions are as above. We have

$$\left|\mathbf{E}|f(Z)-\mathbf{E}f(Z)|^2-\sigma^2\int_0^1 \mathbf{E}(\nabla f(X)\cdot\nabla f(tX+\sqrt{1-t^2}\,Y))\,dt\right|$$
$$\leq n\left(\max_{i=1}^{n}\|Z_i\|_{2(k+1)}^{k+1}\right)\left\|\mathcal{P}^k[f](X,Y)\right\|_2.$$

In general $\mathcal{P}^k[f]$ might be difficult to estimate but for tame functions there is no difficulty to do this.

Recall the pre-Poincaré identity

Lemma (The pre-Poincaré identity)

Assumptions and notation are as in the preceding frame. Let M_i denote the operation of multiplying a function on $\mathbb{R}^n$ by the i^{th} coordinate. We have an integration identity

$$\mathbf{E}|f(Z)|^2 - |\mathbf{E}f(Z)|^2 - \sigma^2 \mathbf{E} \int_0^1 \sum_{i=1}^n D_i[f^*](X) D_i[f](X^{(t)})\, d\sqrt{t}$$

$$= \sum_{i=1}^n \mathbf{E} \int_0^1 (M_i - \sigma^2 D_i)[h_i^{(Y,t)}](X)\, d\sqrt{t}.$$

In more detail, recall the auxiliary functions

$$h_i^{(y,t)} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C} \text{ for } y \in \mathbb{R}^n,\ 0 < t < 1 \text{ and } i = 1, \ldots, n$$

defined by

$$h_i^{(y,t)}(x) = (f^*(x) - f^*(y)) D_i[f](\sqrt{t}\, x + \sqrt{1-t}\, y).$$

In more detail, our task is to bound the quantity

$$\sum_{i=1}^{n} \int_0^1 \mathbf{E}(X_i h_i^{(Y,t)}(X) - \sigma^2 \widetilde{D}_i[h_i^{(Y,t)}](X))\, d\sqrt{t}$$

and for that we will use the drastically simplified IBPI, as follows.

Proof of Poincaré-type inequality (concluded)

We have

$$
\begin{aligned}
& D_i^k[h_i^{(y,t)}](x) \\
= \; & -f(y) D_i^{k+1}[f](\sqrt{t}\,x + \sqrt{1-t}\,y) t^{k/2} \\
& + \sum_{r=0}^{k} \binom{k}{r} D_i^r[f](x) D_i^{k-r+1}[f](\sqrt{t}\,x + \sqrt{1-t}\,y) t^{j/2}
\end{aligned}
$$

hence

$$
|D_i^k[h_i^{(y,t)}](x)| \leq 2 \sum_{r=0}^{k} \mathcal{Q}^r[f](x,y) \mathcal{Q}^{k+1-r}[f](x,y) = \mathcal{P}^k[f](x,y).
$$

The latter leads to the desired bound after plugging into the IBPI and using Fubini's theorem. □

Our next goal is...

...to show that to prove HST it is enough to prove the following statement.

Proposition (Reduction to study of Stieltjes transforms)

For some constant $c > 0$ we have

$$\sup_N N^2 \sup_{z \in \mathfrak{h}} \frac{|\mathbf{E} S_{\mu_f^{(N)}}(z) - S_{\mu_f}(z)|}{(1 + |z| + 1/\Im z)^c} < \infty.$$

No loitering in compact sets

In view of the results on rough control of eigenvalues reviewed above, HST is proved if one can prove the following statement which says roughly that "eigenvalues don't loiter in compact sets."

Proposition

Let $\varphi : \mathbb{R} \to [0,1]$ be any smooth compactly supported function with support disjoint from the support of μ_f. Then $\lim_{N\to\infty} N \int \varphi d\mu_f^{(N)} = 0$ *almost surely.*

Notice that we have multiplied by N so that we can see individual eigenvalues.

Not so fast...

Thus, after a simple Borel-Cantelli/Chebychev argument, HST will be proved if we can prove the following proposition.

Proposition

Let $\varphi : \mathbb{R} \to [0,1]$ be smooth and compactly supported, with support disjoint from the support of μ_f. Then $\lim_{N\to\infty} N^{2+\epsilon}\mathbf{E}\int \varphi d\mu_f^{(N)} = 0$ for some $\epsilon > 0$.

While this is undoubtedly true, it as far as I know not (yet) proved.

I conjecture that the factor N^2 appearing in the conclusion can be replaced here by any power of N. This conjecture is not much of a stretch given

- the general expectation that there would be Tracy-Widom-like tails at the spectrum edges (hence superpolynomial decay) and
- the results of [Haagerup-Thorbjørnsen 2010] which give a $\frac{1}{N}$-expansion for single GUE matrices to all orders of a form one dreams to have in the polynomial case.

A less direct approach

After an only slightly less simple Borel-Cantelli argument, HST will be proved if we can prove the following proposition.

Proposition

Let $\varphi : \mathbb{R} \to [0,1]$ *be smooth and compactly supported, with support disjoint from the support of* μ_f*. Then* $\lim_{N\to\infty} N\mathbf{E}\int \varphi d\mu_f^{(N)} = 0$ *and furthermore* $\sup_N N^{3+\epsilon}\mathbf{Var}(\int \varphi d\mu_f^{(N)}) < \infty$ *for some* $\epsilon > 0$*.*

Slightly less simple Borel-Cantelli argument

Here's the argument we have in mind, which we state because we want to use it again.

Lemma (Slightly less simple Borel-Cantelli argument)

Let $\{X_N\}_{N=1}^{\infty}$ *be a sequence of square-integrable random variables such that* $\lim_{N\to\infty} \mathbf{E}X_N$ *exists and* $\sup_{N=1}^{\infty} N^{1+\epsilon}\mathbf{Var}(X_N) < \infty$*. Then* $\lim_{N\to\infty} X_N = \lim_{N\to\infty} \mathbf{E}X_N$ *almost surely.*

Proof After centering we may assume that all the X_N are of mean zero. Fix $\eta > 0$ arbitrarily. Our hypotheses imply via the Chebychev inequality that $\sum_N \Pr(X_N^2 > \eta) < \infty$ and hence $\Pr(|X_n| > \eta \text{ i.o.}) = 0$ by the Borel-Cantelli lemma, whence the result since η is arbitrary. □

We claim that the following holds.

Proposition (Subtle deployment of the Poincaré inequality)

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be any smooth compactly supported function. Fix any constant $0 < \eta < 1$. Then we have

$$\mathbf{Var}\left(N \int \varphi d\mu_f^{(N)}\right) \leq c \left(\mathbf{E} \int |\varphi'|^2 d\mu_f^{(N)}\right)^\eta$$

for a constant $c = c(\eta)$.

We pause to explain why this is helpful, and then we prove it.

The benefit

Taking advantage of "subtle deployment" and the "slightly less simple Borel-Cantelli argument," we can see that to prove HST it is enough to prove the following statement.

Proposition

Let $\varphi : \mathbb{R} \to [0, \infty)$ be smooth and compactly supported, with support disjoint from the support of μ_f. Then $\sup_N N^2 \mathbf{E} \int \varphi d\mu_f^{(N)} < \infty$.

On its face this statement just misses proving HST but remarkably turns out to do so anyway.

We turn to the task of "subtle deployment."

Crude bounds

In the next several frames we have a close look at "nice" functions. We will be using crude bounds of the following type routinely.

Proposition

For every $g \in \mathrm{Mat}_{r\times s}(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ there exist constants c_1 and c_2 such that for all positive integers N and m-tuples $X = (X_1, \ldots, X_m) \in \mathrm{Herm}_N^m$ one has

$$[\![g(X)]\!] \leq c_1 \left(1 + \sum_{\ell=1}^{m} [\![X_\ell]\!]\right)^{c_2}.$$

Proof One need only check this for monomials in the variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$, in which case it is trivial. □

NB: We often use the same letters c_1 and c_2 in estimates. Their numerical values may change and usually do change line by line.

Bounds for derivatives of matrix polynomials

Proposition

Given $g \in \mathrm{Mat}_{r\times s}(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)$ *and a positive integer* k*, there exist constants* c_1 *and* c_2 *such that for all positive integers* N*, exponents* $p \in [1,\infty)$*, m-tuples* $X = (X_1,\ldots,X_m) \in \mathrm{Herm}_N^m$ *and m-tuples* $H = (H_1,\ldots,H_m) \in \mathrm{Herm}_N^m$ *one has*

$$\left[\!\left[\frac{d^k}{dt^k} g(X+tH)\Big|_{t=0}\right]\!\right]_p \leq c_1\,(1+[\![X_\ell]\!])^{c_2}\left(\sum_{\ell=1}^m [\![H_\ell]\!]_{kp}^p\right)^{k/p}.$$

Proof One need only check this for $r = s = 1$ and g a monomial in the $\mathbf{X}_\ell$, in which case it is obvious. □

A Hilbert space lemma

Lemma

Let V and W be finite-dimensional Hilbert spaces with V real and W complex. Fix $w \in W$. Let $\{v_j\}$ be any orthonormal basis for V. Fix $\mathbb{R}$-linear $L : V \to W$ and let $[\![L]\!]^2 = \sup_{v \neq 0} \frac{(L(v),L(v))}{(v,v)}$. Then

$$\sum_j |(L(v_j), w)|^2 \leq [\![L]\!]^2 (w, w).$$

Proof In coordinates this merely says that for any rectangular matrix $A \in \mathrm{Mat}_{k\times\ell}$ we have $\sum_{j=1}^{\ell} |A[1,j]|^2 \leq [\![A]\!]^2$. $\square$

We intend to apply this presently with $V = \mathrm{Herm}_N^m$ and $W = \mathrm{Mat}_N$, the latter equipped with Hilbert space structure by the rule $(A, B) = \mathrm{tr} A^* B$.

The next proposition is one of the most basic "tricks" in the proof of HST. It is "just algebra." Put

$$\hat{e}_{ij;\ell} = (\underbrace{0,\ldots,0}_{\ell-1}, \hat{e}_{ij}, \underbrace{0,\ldots,0}_{m-\ell}) \in \mathrm{Herm}_N^m$$

and for a smooth function $\Phi : \mathrm{Herm}_N^m \to \mathrm{Mat}_{k\times\ell}$ put

$$\hat{\partial}_{ij;\ell}\Phi(X) = \frac{d}{dt}\Phi(X + \hat{e}_{ij;\ell})\Big|_{t=0}.$$

Proposition

For each $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ there exist constants c_1 and c_2 such that for all positive integers N, m-tuples $X = (X_1,\ldots,X_m) \in \mathrm{Herm}_N^m$ and smooth functions $\varphi : \mathbb{R} \to \mathbb{C}$ one has

$$\sum_{\ell=1}^{m}\sum_{i,j=1}^{N}\left|\hat{\partial}_{ij;\ell}\mathrm{tr}\,\varphi(f(X))\right|^2 \le c_1\left(1+\sum_{\ell=1}^{m}[\![X_\ell]\!]\right)^{c_2}\mathrm{tr}(|\varphi'|^2(f(X))).$$

Proof of the proposition

Let $\mathbf{H} = (\mathbf{H}_1, \ldots, \mathbf{H}_m)$ be an m-tuple of noncommutative variables independent of the m-tuple $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$. We declare $\mathbf{H}_\ell^* = \mathbf{H}_\ell$ to give $*$-algebra structure to $\mathbb{C}\langle \mathbf{X}_{1\ldots m}, \mathbf{H}_{1\ldots m}\rangle$. Put

$$f'(\mathbf{X}_{1\ldots m}, \mathbf{H}_{1\ldots m}) = \frac{d}{dt} f(\mathbf{X} + t\mathbf{H})\Big|_{t=0} \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}, \mathbf{H}_{1\ldots m}\rangle)_{\mathrm{sa}}.$$

(Note that f' is self-adjoint.) For constants c_1 and c_2 depending only on f' we have

$$[\![f'(X,H)]\!]_2^2 \leq c_1 \left(1 + \sum_{\ell=1}^m [\![X_\ell]\!]\right)^{c_2} \left(\sum_{\ell=1}^m [\![H_\ell]\!]_2^2\right).$$

Thus we have

$$\begin{aligned} \sum_{\ell=1}^m \sum_{i,j=1}^N \left|\hat{\partial}_{ij;\ell} \mathrm{tr}\, \varphi(X)\right|^2 &= \sum_{\ell=1}^m \sum_{i,j=1}^N |\mathrm{tr}(f'(X, \hat{e}_{ij;\ell})\varphi'(X))|^2 \\ &\leq c_1 \left(1 + \sum_{\ell=1}^m [\![X_\ell]\!]\right)^{c_2} \mathrm{tr}(\varphi'(X)\varphi'(X)^*), \end{aligned}$$

which finishes the proof. $\square$

Application of the Poincaré inequality

Proposition

Given $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$, there exist constants c_1 and c_2 such that for every positive integer N, m-tuple $\Xi^{(N)} = (\Xi_1^{(N)}, \ldots, \Xi_m^{(N)})$ of independent GUE matrices and smooth compactly supported function $\varphi : \mathbb{R} \to \mathbb{R}$ one has

$$\begin{aligned} & \mathbf{Var}\left(\int \varphi \, d\mu_f^{(N)}\right) \\ \leq \quad & \frac{c_1}{N^2}\mathbf{E}\left(1 + \sum_{\ell=1}^{m} \left[\left[\frac{\Xi^{(N)}}{\sqrt{N}}\right]\right]\right)^{c_2} \int |\varphi'|^2 \, d\mu_f^{(N)}. \end{aligned}$$

Proof This proposition follows immediately from the previous proposition and the Poincaré inequality. □

Remark

With an eye toward proving "fake" HST we note the following.

Proposition

Fix $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ *and a positive integer* k*. Then there exist constants* c_1, c_2 *depending only on* f *and* k *such that for every smooth function* $\varphi : \mathbb{R} \to \mathbb{C}$*, and* m*-tuple* $X = (X_1, \ldots, X_m) \in \mathrm{Herm}_N^m$ *we have*

$$\begin{aligned} &\max_{i,j=1}^{N} \max_{\ell=1}^{m} \left| (\hat{\partial}_{ij;\ell})^k \mathrm{tr}(\varphi(f(X))) \right| \\ \leq \quad & c_1 \left(1 + \sum_{\ell=1}^{m} [\![X_\ell]\!] \right)^{c_2} \sup_{\substack{r=1,\ldots,k \\ |x| \leq [\![f(X)]\!]}} |\varphi^{(r)}(x)|. \end{aligned}$$

Using this proposition we can carry out the Poincaré inequality step for fake GUE matrices.

Almost done with “subtle deployment”

Recall our goal.

Proposition (Subtle deployment of the Poincaré inequality)

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be any smooth compactly supported function. Fix any constant $0 < \eta < 1$. Then we have

$$\mathbf{Var}\left(N \int \varphi d\mu_f^{(N)}\right) \leq c \left(\mathbf{E} \int |\varphi'|^2 d\mu_f^{(N)}\right)^{\eta}$$

a constant c independent of N.

Note that there is no restriction on the support of φ other than compactness.

Completion of “subtle deployment”

Proof By “baby interpolation” we have

$$\left\| \int |\varphi'|^2 d\mu_f^{(N)} \right\|_p \leq \left\| \int |\varphi'|^2 d\mu_f^{(N)} \right\|_1^{1/p} \left(\underbrace{\left\| \int |\varphi'|^2 d\mu_f^{(N)} \right\|_\infty}_{O_{N\to\infty}(1)} \right)^{1/q}$$

for any conjugate exponents $p, q \in (1, \infty)$. We then get the desired result from Step 1 and the previous proposition via the Hölder inequality. $\square$

We have boiled the proof of HST (and its fake version) down to that of the following statement.

Proposition

For every smooth compactly supported function $\varphi : \mathbb{R} \to [0, \infty)$ *with support disjoint from* $\mathrm{supp}\mu_f$ *we have* $\sup_N N^2 \mathbf{E} \int \varphi d\mu_f^{(N)} < \infty$.

Seemingly this statement falls just short of proving HST but remarkably it suffices anyhow.

Raising the bar a little

It turns out to be no harder to prove the following statement, and that is in fact what we will do.

Proposition

For every smooth compactly supported function $\varphi : \mathbb{R} \to \mathbb{R}$ *we have* $\sup_N N^2 |\mathbf{E} \int \varphi d\mu_f^{(N)} - \int \varphi d\mu_f| < \infty$.

The preceding proposition and rough control of eigenvalues lead not only to a proof of HST (and its "fake" version) but also yield as a byproduct the following statement.

Theorem (Amplification of Voiculescu's theorem)

$\mu_f^{(N)}$ *converges weakly to* μ_f, *almost surely.*

Proof Since we already have tightness by Step 1, it is enough to check almost sure convergence $\int \varphi d\mu_f^{(N)} \to_{N\to\infty} \int \varphi d\mu_f$ for sufficiently many test-functions φ. Certainly the class of smooth compactly supported test-functions is big enough. We get almost sure convergence by exploiting "subtle deployment" and the "slightly less simple Borel-Cantelli argument" yet again. □

Completion of the reduction to study of Stieltjes transforms

Using the $\bar{\partial}$-trick-based method of recovering probability measures from their Stieltjes transforms along with Fubini's theorem and dominated convergence, we are able now to reduce the proof of HST as well as that of the amplification of Voiculescu's theorem to the proof of the following statement.

Proposition (Reduction to study of Stieltjes transforms)

For some constant $c > 0$ one has

$$\sup_{N\to\infty} N^2 \sup_{z\in\mathfrak{h}} \frac{|\mathbf{E}S_{\mu_f^{(N)}}(z) - S_{\mu_f}(z)|}{(1+|z|+1/\Im z)^c} < \infty.$$

I conjecture that the proposition holds for any power of N in front with $c = c(N)$. This conjecture via the $\bar{\partial}$-trick implies the previously mentioned conjecture.

Our next goal is...

...is to finish the proof of HST by deploying the Schwinger-Dyson/linearization machine.

Amplification

Recall that we have so far reduced the proof of HST to the proof of the following statement:

Proposition

For some constant $c > 0$ one has

$$\sup_{N\to\infty} N^2 \sup_{z\in\mathfrak{h}} \frac{|\mathbf{E} S_{\mu_f^{(N)}}(z) - S_{\mu_f}(z)|}{(1+|z|+1/\Im z)^c} < \infty.$$

We will use the Schwinger-Dyson/linearization machine to identify a relatively simple "statistic" in terms of which the quantity considered in the proposition above may be controlled.

Notation: Tensor products etc.

- For an algebra $\mathcal{A}$, let $\mathrm{GL}_n(\mathcal{A})$ denote the group of invertible elements of $\mathrm{Mat}_n(\mathcal{A})$.
  - GL is short for the *general linear group*.
- Given $X \in \mathrm{Mat}_{k\times\ell}$ and $a \in \mathcal{A}$, let $X \otimes a \in \mathrm{Mat}_{k\times\ell}$ be defined by $(X \otimes a)[i,j] = X[i,j]a$.
- Note that $(X \otimes a)^* = X^* \otimes a^*$.
- Note also that if $a \in \mathrm{Mat}_N$, then $X \otimes a$ is the usual Kronecker product of matrices, i.e.,

$$X \otimes a = \begin{bmatrix} X[1,1]a & \dots & X[1,\ell]a \\ \vdots & & \vdots \\ X[k,1]a & \dots & X[k,\ell]a \end{bmatrix}.$$

Notation: Block trace

We define

$$\mathrm{tr}_n : \mathrm{Mat}_{nN} \to \mathrm{Mat}_n$$

to be the unique linear map such that

$$\mathrm{tr}_n(a \otimes A) = \mathrm{tr}(A)a$$

for all $a \in \mathrm{Mat}_n$ and $A \in \mathrm{Mat}_N$.

For example

$$\mathrm{tr}_3 \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & 18 \\ 19 & 20 & 21 & 22 & 23 & 24 \\ 25 & 26 & 27 & 28 & 29 & 30 \\ 31 & 32 & 33 & 34 & 35 & 36 \end{bmatrix} = \begin{bmatrix} 9 & 13 & 17 \\ 33 & 37 & 41 \\ 57 & 61 & 65 \end{bmatrix}.$$

Interaction of block trace with $[\![\cdot]\!]_1$

Lemma

For all $A \in \mathrm{Mat}_{Nn}$ *we have* $[\![\mathrm{tr}_n(A)]\!] \leq \frac{1}{N}[\![A]\!]$ *and* $[\![\mathrm{tr}_n(A)]\!]_1 \leq [\![A]\!]_1$.

Proof Write $A = \sum_{i,j=1}^n a_{ij} \otimes e_{ij}$ where the $e_{ij} \in \mathrm{Mat}_N$ are elementary and the $a_{ij} \in \mathrm{Mat}_n$ are uniquely determined. It is easy to see that $[\![a_{ij}]\!] \leq [\![A]\!]$ for all i and j, which proves the first inequality. To prove the second inequality we may without loss of generality assume that the diagonal blocks a_{ii} are themselves diagonal with nonnegative diagonal entries after left- and right-multiplying A by suitable unitary matrices. Then we have $[\![A]\!]_1 \geq \sum_{i=1}^{Nn} |A[i,i]| = [\![\mathrm{tr}_n(A)]\!]_1$, which proves the result. $\square$

Setup for linearization

For $f \in \mathrm{Mat}_n(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}$ as in the statement of HST, fix a self-adjoint linearization

$$L_0 + \sum_{\ell=1}^{m} L_\ell \otimes \mathbf{X}_\ell \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}} \ \ (L_0, L_1, \ldots, L_m \in \mathrm{Herm}_s).$$

Put

$$\begin{aligned} L &= \sum_{\ell=1}^{m} L_\ell \otimes \mathbf{X}_\ell \in \mathrm{Mat}_s(\mathbb{C}\langle \mathbf{X}_{1\ldots m}\rangle)_{\mathrm{sa}}, \\ \Lambda^{(z)} &= \left[\begin{array}{cc} z\mathbf{I}_n & 0 \\ 0 & 0 \end{array}\right] - L_0 \in \mathrm{Mat}_s \ \text{ for } z \in \mathbb{C}. \end{aligned}$$

Definition of $F^{(N,z)}$

Recall that $\Xi^{(N)} = (\Xi_1^{(N)}, \ldots, \Xi_m^{(N)})$ is an m-tuple of independent GUE matrices. (We may also consider fake GUE matrices.)

We put

$$F^{(N,z)} = \mathrm{tr}_s \left(\left(L\left(\frac{\Xi^{(N)}}{\sqrt{N}} \right) - \Lambda^{(z)} \otimes \mathbf{I}_N \right)^{-1} \right) \in \mathrm{Mat}_s$$

for $z \in \mathfrak{h}$. By the Schur complement $\leftrightarrow$ linearization relationship $F^{(N,z)}$ is almost surely well-defined.

$(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$-setup at our disposal

Recall that $\Xi = (\Xi_1, \ldots, \Xi_m)$ is an m-tuple of free semicircular operators on Boltzmann-Fock space.

Put

$$\begin{aligned}
\mathcal{S} &= \mathrm{Mat}_s, \\
\Phi &= \left(A \mapsto \sum_{\ell=1}^m L_\ell A L_\ell\right) \in B(\mathrm{Mat}_s), \\
L &= \sum_{\ell=1}^m L_\ell \otimes \Xi_\ell \in \mathrm{Mat}_s\left(B(\mathcal{H})\right)_{\mathrm{sa}}, \\
\mathcal{D} &= \{\Lambda \in \mathrm{Mat}_s \mid L - \Lambda \otimes 1_{B(\mathcal{H})} \in \mathrm{GL}_s(B(\mathcal{H}))\}, \\
\rho &= \left(\Lambda \mapsto \left[\!\left[(L - \Lambda \otimes 1_{B(\mathcal{H})})^{-1}\right]\!\right]^{-1}\right) : \mathcal{D} \to (0, \infty).
\end{aligned}$$

Finally for $\Lambda \in \mathcal{D}$ we define $G(\Lambda) \in \mathrm{Mat}_s$ by

$$G(\Lambda)[i,j] = (1_\mathcal{H}, (L - \Lambda \otimes 1_{B(\mathcal{H})})^{-1}[i,j]1_\mathcal{H}).$$

Then $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$ is a Schwinger-Dyson approximation scheme.

The Schur complement $\leftrightarrow$ linearization relationship shows that for all $z \in \mathbb{C}$ we have

$$\Lambda^{(z)} \in \mathcal{D} \Leftrightarrow z \in \mathbb{C} \setminus \mathrm{Spec}(f(\Xi)) \stackrel{\text{folklore}}{=} \mathbb{C} \setminus \mathrm{supp}\mu_f$$

in which case we also must have

$$S_{\mu_f}(z) = \frac{1}{n} \sum_{i=1}^{n} G(\Lambda^{(z)})[i,i].$$

In parallel to this, for $z \in \mathfrak{h}$ we also have that

$$S_{\mu_f^{(N)}}(z) = \frac{1}{n} \sum_{i=1}^{n} F^{(N,z)}[i,i].$$

The upshot is that HST is proved once we prove the following statement.

Proposition

For some constant $c > 0$ we have

$$\sup_{N\to\infty} N^2 \sup_{z\in\mathfrak{h}} \frac{\left[\!\left[G(\Lambda^{(z)}) - \mathbf{E}F^{(N,z)} \right]\!\right]}{(1+|z|+1/\Im z)^c} < \infty.$$

The integrability of $F^{(N,z)}$ needs to be clarified. More importantly, the fact that $G(\Lambda)$ satisfies the Schwinger-Dyson equation needs to be exploited. These points will be taken up next.

As in the "two-variable estimate" put

$$\Lambda^{(z,t)} = \Lambda^{(z)} + \mathrm{i}t\mathbf{I}_s \ \mathrm{Mat}_s \text{ for } z \in \mathfrak{h} \text{ and } t \geq 0.$$

For $z \in \mathfrak{h}$ and $t \geq 0$ we put

$$\begin{aligned} R^{(N,z,t)} &= \left(L\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right) - \Lambda^{(z)} \otimes \mathbf{I}_N \right)^{-1} \in \mathrm{Mat}_{sN}, \\ F^{(N,z,t)} &= \frac{1}{N}\mathrm{tr}_s R^{(N,z,t)} \in \mathrm{Mat}_s. \end{aligned}$$

By the two-variable estimate these objects are well-defined. Clearly we have

$$\left[\!\left[F^{(N,z,t)} \right]\!\right] \leq \left[\!\left[R^{(N,z,t)} \right]\!\right].$$

By the two-variable estimate we have an almost sure bound

$$\left[\!\left[R^{(N,z,t)}\right]\!\right] \leq \frac{1}{t} \wedge c_1 \left(1 + \sum_{\ell=1}^{m} \left[\!\left[\frac{\Xi_\ell^{(N)}}{\sqrt{N}}\right]\!\right]\right)^{c_2} \left(1 + \frac{1}{\Im z}\right)$$

for constants c_1 and c_2 depending only on f, L and L_0.

Similarly we have

$$\left[\!\left[\frac{d}{dt}R^{(N,z,t)}\right]\!\right]^{1/2} \leq \frac{1}{t} \wedge c_1 \left(1 + \sum_{\ell=1}^{m} \left[\!\left[\frac{\Xi_\ell^{(N)}}{\sqrt{N}}\right]\!\right]\right)^{c_2} \left(1 + \frac{1}{\Im z}\right)$$

almost surely, with the same constants here as in the previous estimate.

Bounds for and continuity of $\mathbf{E}F^{(N,z,t)}$

We evidently have

$$\mathbf{E}\left[\!\left[F^{(N,z,t)}\right]\!\right] \leq \frac{1}{t} \wedge c_3\left(1+\frac{1}{\Im z}\right)$$

for a constant c independent of N, z and t. Here of course we make use in a serious way of Step 1 of the reduction of HST to study of Stieltjes transforms.

Thus integrability of $F^{(N,z)}$ is now clarified and moreover it is clear that $\mathbf{E}F^{(N,z,t)}$ depends continuously on t.

Similarly the two-variable estimate gives

$$\Lambda^{(z,t)} \in \mathcal{D} \ \text{ and } \ \left[\!\left[G(\Lambda^{(z,t)}) \right]\!\right] \leq \frac{1}{\rho(\Lambda^{(z,t)})} \leq \frac{1}{t} \wedge c_4 \left(1 + \frac{1}{\Im z} \right)$$

for a constant c_4 independent of z and t.

Resumé of $(\mathcal{S}, \Phi, \mathcal{D}, \rho, G)$-axioms

Recall that we have

$$\mathcal{D} = \{\Lambda \in \mathrm{Mat}_s \mid \Im\Lambda > 0\} \cup \bigcup_{\Lambda \in \mathcal{D}} \left\{\Lambda' \in \mathrm{Mat}_s \middle| [\![\Lambda' - \Lambda]\!] < \rho(\Lambda)\right\}$$

and furthermore for $\Lambda, \Lambda' \in \mathcal{D}$ we have the following:

$$\mathbf{I}_s + (\Lambda + \Phi(G(\Lambda)))G(\Lambda) = 0. \text{ (Schwinger-Dyson equation)}$$

$$[\![G(\Lambda)]\!] \leq \frac{1}{\rho(\Lambda)} \text{ and } [\![G(\Lambda) - G(\Lambda')]\!] \leq \frac{[\![\Lambda - \Lambda']\!]}{\rho(\Lambda)\rho(\Lambda')}.$$

$$|\rho(\Lambda) - \rho(\Lambda')| \leq [\![\Lambda' - \Lambda]\!] \text{ and } \Im\Lambda > 0 \Rightarrow \frac{1}{\rho(\Lambda)} \leq \left[\!\!\left[(\Im\Lambda)^{-1}\right]\!\!\right].$$

Application of the tunnel bound

For $z \in \mathfrak{h}$ and $t \in [0, \infty)$ put

$$E^{(N,z,t)} = \mathbf{I}_s + (\Lambda + \Phi(\mathbf{E}F^{(N,z,t)}))\mathbf{E}F^{(N,z,t)}.$$

Plugging into the tunnel bound we get

$$\left[\!\left[G(\Lambda^{(z)}) - F^{(N,z)} \right]\!\right] \leq c_5 \left(1 + \frac{1}{\Im z}\right)^4 \sup_{t \in [0,\infty)} \left[\!\left[E^{(N,z,t)} \right]\!\right]$$

for a constant c_5 independent of N and z.

Reduction by means of the Schwinger-Dyson/linearization machine

We have thus reduced the proof of HST to the proof of the following statement.

Proposition

For some constant $c > 0$ we have

$$\sup_{N\to\infty} \sup_{\substack{z\in\mathfrak{h}\\ t\in[0,\infty)}} N^2 \frac{\left[\left[E^{(N,z,t)}\right]\right]}{(1+|z|+1/\Im z)^c} < \infty.$$

Our next goal is...

is to carry out the Gaussian integration by parts and second application of the Poincaré inequality needed to complete the proof of HST.

We will emphasize the re-usable algebraic parts of the calculation independent of the specific properties of Gaussian random variables so that modifications to get “fake” HST are evident.

Generalized resolvents (setup)

For now we float more or less free of the discussion of HST, starting (temporarily) anew.

- Fix positive integers s, m and N.
- Fix matrices $\Lambda, L_1, \ldots, L_m \in \mathrm{Mat}_s$.
- Assume that

$$-\Lambda \otimes \mathbf{I}_N + \sum_{\ell=1}^{m} L_\ell \otimes X_\ell \in \mathrm{GL}_{sN}$$

for all $X = (X_1, \ldots, X_m) \in \mathrm{Herm}_N^m$.

Consider the *generalized resolvent* and its *normalized block trace*

$$\begin{aligned} R(X) &= \left(-\Lambda \otimes \mathbf{I}_N + \sum_{\ell=1}^{m} L_\ell \otimes X_\ell\right)^{-1} \in \mathrm{GL}_{sN}, \\ F(X) &= \frac{1}{N}\mathrm{tr}_s R(X) \in \mathrm{Mat}_s, \end{aligned}$$

respectively, for $X = (X_1, \ldots, X_m) \in \mathrm{Herm}_N^m$.

Generalized resolvents (calculus notation)

Recall that $\{\hat{e}_{ij}\}_{i,j=1}^{N}$ is the canonical orthonormal basis for Herm_N.

For any smooth function Φ defined on Herm_N^m with values in a finite-dimensional real or complex vector space recall that we have defined

$$\hat{\partial}_{ij;\ell}\Phi(X) = \frac{d}{dt}\Phi(X_1, \dots, X_{\ell-1}, X_\ell + t\hat{e}_{ij}, X_{\ell+1}, \dots, X_m)|_{t=0}$$

for $\ell = 1, \dots, m$ and $i, j = 1, \dots, N$.

One verifies immediately that

$$\hat{\partial}_{ij;\ell}R(X) = -R(X)(L_\ell \otimes \hat{e}_{ij})R(X).$$

More on the standard orthonormal basis

We more generally equip $\mathrm{Mat}_{k\times\ell}$ with Hilbert space structure (antilinear on the left, linear on the right) by the rule $(A,B)_{\mathrm{Mat}_{k\times\ell}} = \mathrm{tr}(A^*B)$. Note that, naturally enough, the family $\{e_{ij}\}$ of elementary k-by-ℓ matrices is orthonormal.

For $A, B \in \mathrm{Mat}_N$ we have

$$\begin{aligned}\sum_{i,j=1}^N \mathrm{tr}(A\hat{e}_{ij})\mathrm{tr}(\hat{e}_{ij}B) &= \sum_{i,j=1}^N \mathrm{tr}(Ae_{ij})\mathrm{tr}(e_{ji}B) = \sum_{i,j=1}^N A[j,i]B[i,j]\\ &= \mathrm{tr}(AB)\end{aligned}$$

and

$$\begin{aligned}\sum_{i,j=1}^N \mathrm{tr}(A\hat{e}_{ij}B\hat{e}_{ij}) &= \sum_{i,j=1}^N \mathrm{tr}(Ae_{ij}Be_{ji}) = \sum_{i,j=1}^N A[i,i]B[j,j]\\ &= \mathrm{tr}(A)\mathrm{tr}(B).\end{aligned}$$

Generalized resolvents ("just algebra")

Proposition ("just algebra")

In the setting above, for all $X = (X_1, \dots, X_m) \in \mathrm{Herm}_N^m$, *we have*

$$\mathbf{I}_s + \left(\Lambda + \sum_{\ell=1}^{m} L_\ell F(X) L_\ell\right) F(X)$$

$$= \frac{1}{N} \sum_{\ell=1}^{m} \sum_{i,j=1}^{N} \left((\hat{e}_{ij}, X_\ell) - \frac{\hat{\partial}_{ij;\ell}}{N} \right) \mathrm{tr}_s((L_\ell \otimes \hat{e}_{ij}) R(X)),$$

$$\max_{i,j=1}^{k} \max_{i,j=1}^{N} \left[\!\left[(\hat{\partial}_{ij;\ell})^k \mathrm{tr}_s((L_\ell \otimes \hat{e}_{ij}) R(X)) \right]\!\right]_1$$

$$\leq \quad k! \left(\max_{\ell=1}^{m} [\![L_\ell]\!]_1 \right)^{k+1} [\![R(X)]\!]^{k+1},$$

$$\sum_{\ell=1}^{m} \sum_{i,j=1}^{N} \left[\!\left[\hat{\partial}_{ij;\ell} F(X) \right]\!\right]_2^2 \leq \frac{s^6}{N} \left(\sum_{\ell=1}^{m} [\![L_\ell]\!]^2 \right) [\![R(X)]\!]^4,$$

$$\max_{i,j=1}^{k} \max_{i,j=1}^{N} \left[\!\left[(\hat{\partial}_{ij;\ell})^k F(X) \right]\!\right]_1 \leq \frac{k!}{N} \left(\max_{\ell=1}^{m} [\![L_\ell]\!]_1 \right)^k [\![R(X)]\!]^{k+1}.$$

Proof of "just algebra"

Proof We have

$$(\hat{\partial}_{ij;\ell})^k F(X) = \frac{(-1)^k k!}{N}\mathrm{tr}_s(R(X)((L_\ell \otimes \hat{e}_{ij})R(X))^k),$$

which proves the fourth relation after bounding the right side in the obvious way. The proof of the second relation is very similar.

The right side of the first equation can be rewritten

$$\frac{1}{N}\mathrm{tr}_s(L(X)R(X)) + \frac{1}{N^2}\sum_{\ell=1}^{m}\sum_{i,j=1}^{N}\mathrm{tr}_s((L_\ell \otimes \hat{e}_{ij})R(X)(L_\ell \otimes \hat{e}_{ij})R(X)).$$

and trivially we have

$$\mathbf{I}_s + \Lambda F(X) = \frac{1}{N}\mathrm{tr}_s(L(X)R(X)).$$

The left side of the second equation can be rewritten

$$\frac{1}{N^2}\sum_{\ell=1}^{m}\sum_{i,j=1}^{N}[\![R(X)(L_\ell \otimes \hat{e}_{ij})R(X)]\!]_2^2.$$

Proof of “just algebra” concluded modulo algebra lemmas

Thus the two lemmas and two proofs in the next four frames finish the proof of the proposition. $\square$

First algebra lemma

Lemma

For any matrices $P, Q \in \mathrm{Mat}_{sN}$ *and matrices* $a, b \in \mathrm{Mat}_s$ *we have*

$$\sum_{i,j=1}^{N} \mathrm{tr}_s((a \otimes \hat{e}_{ij})P(b \otimes \hat{e}_{ij})Q) = a\,\mathrm{tr}_s(P)\,b\,\mathrm{tr}_s(Q).$$

Proof of the first lemma

We may assume that $P = c \otimes C$ and $Q = d \otimes D$ with $c, d \in \mathrm{Mat}_s$ and $C, D \in \mathrm{Mat}_N$. We then have

$$\begin{aligned}
& \sum_{i,j=1}^{N} \mathrm{tr}_s((a \otimes \hat{e}_{ij})P(b \otimes \hat{e}_{ij})Q) \\
= \quad & \sum_{i,j=1}^{N} \mathrm{tr}_s((a \otimes e_{ij})(c \otimes C)(b \otimes e_{ji})d \otimes D) \\
= \quad & \sum_{i,j=1}^{N} acbd\mathrm{tr}(e_{ij} C e_{ji} D) = abcd\mathrm{tr}(C)\mathrm{tr}(D) \\
= \quad & a\,\mathrm{tr}_s(P)\, b\,\mathrm{tr}_s(Q),
\end{aligned}$$

which finishes the proof. $\square$

Second algebra lemma

Lemma

For any matrix $a \in \mathrm{Mat}_s$ and matrix $R \in \mathrm{Mat}_{sN}$ we have

$$\sum_{i,j=1}^{N} [\![\mathrm{tr}_s(R(a \otimes \hat{e}_{ij})R)]\!]_2^2 \leq s^6 N [\![a]\!]^2 [\![R]\!]^4$$

Writing $R = \sum_{i,j=1}^{s} e_{ij} \otimes R_{ij}$ with $R_{ij} \in \mathrm{Mat}_N)$ we have

$$
\begin{aligned}
& \sum_{i,j=1}^{N} [\![\mathrm{tr}_s(R(a \otimes \hat{e}_{ij})R)]\!]_2^2 \\
= \quad & \sum_{i,j=1}^{N} \left[\!\!\left[\sum_{i_1,i_2,j_1,j_2=1}^{s} e_{i_1,j_1} a e_{i_2 j_2} \mathrm{tr}(R_{i_1 j_1} \hat{e}_{ij} R_{i_2 j_2}) \right]\!\!\right]_2^2 \\
= \quad & \sum_{i,j=1}^{N} \sum_{\substack{i_1,i_2,i_3,i_4=1 \\ j_1,j_2,j_3,j_4=1}}^{s} \mathrm{tr}(e_{i_1 j_1} a e_{i_2 j_2} e_{j_4 i_4} a^* e_{j_3 i_3}) \mathrm{tr}(R_{i_1 j_1} \hat{e}_{ij} R_{i_2 j_2}) \mathrm{tr}(R^*_{i_4,j_4} \hat{e}_{ij} R^*_{i_3,j_3}) \\
= \quad & \sum_{\substack{i_1,i_2,i_3,i_4=1 \\ j_1,j_2,j_3,j_4=1}}^{s} \mathrm{tr}(e_{i_1 j_1} a e_{i_2 j_2} e_{j_4 i_4} a^* e_{j_3,i_3}) \mathrm{tr}(R_{i_2 j_2} R_{i_1 j_1} R^*_{i_3,j_3} R^*_{i_4,j_4}) \\
\leq \quad & s^6 [\![a]\!]^2 N [\![R]\!]^4.
\end{aligned}
$$

$\square$

Generalized resolvents of GUE matrices (setup for application)

Let us now assume that we have a bound

$$[\![R(X)]\!] \le c_1 \left(1 + \sum_{\ell=1}^{m} [\![X_\ell]\!]\right)^{c_2}$$

for c_1 and c_2 independent of X in order to avoid integrability issues in the next statement.

The linearization trick will provide us automatically with such an *a priori* bound.

Proposition (Tool for final step of the proof of HST)

Notation and assumptions are as above. Let $\Xi^{(N)} = (\Xi_1^{(N)}, \ldots, \Xi_m^{(N)})$ *be a m-tuple of independent N-by-N GUE matrices. We have*

$$\begin{aligned}
&\left[\!\!\left[\mathbf{I}_s + \left(\Lambda + \sum_{\ell=1}^{m} L_\ell \mathbf{E} F\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right) L_\ell\right) \mathbf{E} F\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right)\right]\!\!\right] \\
\leq \;\; & \frac{s^6}{N^2}\left(\sum_{\ell=1}^{m} [\![L_\ell]\!]^2\right)^2 \mathbf{E}\left[\!\!\left[R\left(\frac{\Xi^{(N)}}{\sqrt{N}}\right)\right]\!\!\right]^4 < \infty.
\end{aligned}$$

Now we return to the proof of HST. According to previous bound, along with the bounds we already have for $\left[\!\left[R^{(N,z,t)} \right]\!\right]$, we know that

$$\left[\!\left[E^{(N,z,t)} \right]\!\right] \leq c_6 \left(1 + \frac{1}{\Im z} \right)^4$$

for a constant c_6 independent of N, z and t. So we are done! $\square$

Fake HST proved, too

Remark And the proof of fake HST goes through as well. We wonder if this method could be refined enough to give an easier proof of the results of [Anderson, Ann. Probab., to appear], and beyond that, whether it could lead to generalization of the results of [Male 2010] and to results generalizing those of [Bai-Silverstein 1998].

- Anderson, G. *Convergence of the largest singular value of a polynomial in independent Wigner matrices.* Ann. Probab., to appear.
- Anderson, G., Guionnet, A., and Zeitouni, O. (2010). *An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics* **118**. Cambridge University Press, Cambridge. MR2760897
- Bai, Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051
- Bai, Z. D. and Yin, Y. Q. (1988). *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix.* Ann. Probab. **16** 1729–1741. MR0958213

- Capitaine, M. and Donati-Martin, C. (2007). Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56** 767–803. MR2317545
- Benoit Collins, Camille Male *The strong asymptotic freeness of Haar and deterministic matrices* arXiv:1105.4345
- A. Deya, I. Nourdin, *Convergence of Wigner integrals to the tetilla law* arXiv:1107.3538
- Füredi, Z. and Komlós, J. (1981). *The eigenvalues of random symmetric matrices.* Combinatorica **1** 233–241. MR0637828
- Haagerup, U., Schultz, H. and Thorbjørnsen, S. (2006). A random matrix approach to the lack of projections in $C^*_{\mathrm{red}}(\mathbb{F}_2)$. *Adv. Math.* **204** 1–83. MR2233126
- Haagerup, U. and Thorbjørnsen, S. (2005). A new application of random matrices: $\mathrm{Ext}(C^*(\mathbb{F}_2))$ is not a group. *Annals Math.* **162** 711–775. MR2183281

Haagerup, U. and Thorbjørnsen, S. (2010). *Asymptotic expansions for the Gaussian Unitary Ensemble* `arXiv:1004.3479`, 22 June 2012.

Hachem, W., Loubaton, P. and Najim, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235

Hachem, W., Loubaton, P. and Najim, J. (2008). A CLT for information-theoretic statistics of gram random matrices with a given variance profile. *Ann. Appl. Probab.* **18** 2071–2130. MR2473651

B. Helffer and J. Sjöstrand, *Equation de Schrödinger avec champ magnétique et équation de Harper,* Springer Lecture Notes in Physics **345**(1989), 118-197.

Helton, J., Rashidi Far, R. and Speicher, R. (2007). Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints. Int. Math. Res. Not., no. 22, Art. ID rnm086. MR2376207

R. Horn and C. Johnson, *Matrix analysis. Corrected reprint of the 1985 original.* Cambridge University Press, Cambridge, 1990.

Khorunzhy, A., Khoruzhenko, B. and Pastur, L. *On the 1/N corrections to the Green functions of random matrices with independent entries.* J. Phys. A **28**(1995), L31L35. MR1325832

Lytova, A. and Pastur, L. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. Ann. Probab. **37**(2009), no. 5, 17781840. MR2561434 (2011c:60024)

Lytova A. and Pastur L., *Fluctuations of matrix elements of regular functions of Gaussian random matrices,* J. Stat. Phys., **134**(2009), 147–159.

Male, C. (2010). The norm of polynomials in large random and deterministic matrices. `arXiv:1004.4155v4` 3 Mar 2011.

Meckes, M. W. and Szarek, S. J. (2011). Concentration for noncommutative polynomials in random matrices. `arXiv:1101.1923v1` 10 Jan 2011.

Murphy, G. J. (1990). *C^*-algebras and operator theory.* Academic Press, Boston. MR1074574

A. Nica and R. Speicher, *Commutators of free random variables.* Duke Math. J. **92**(1998)553–592. MR1620518

A. Nica, A. and Speicher, R. (2006). *Lectures on the combinatorics of free probability.* London Mathematical Society Lecture Note Series **335**. Cambridge University Press, Cambridge. MR2266879

S. ORourke, D. Renfrew, A. Soshnikov, *On fluctuations of matrix entries of regular functions of Wigner matrices with non-identically distributed entries, arXiv:1104.1663v2 [math.PR]*

O'Rourke, S., Renfrew, D., Soshnikov, A.: *FLUCTUATIONS OF MATRIX ENTRIES OF REGULAR FUNCTIONS OF SAMPLE COVARIANCE RANDOM MATRICES* arXiv:1106.0320v1 [math.PR] 1 Jun 2011

Schultz, H. (2005). Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields* **131** 261–309. MR2117954

B. Simon. *Trace Ideals and Their Applications.* Second Edition. Mathematical Surveys and Monographs, Volume 120, Amer. Math. Soc., Providence, RI, 2005.

Voiculescu, D. V., *Limit laws for random matrices and free products.* Invent. math. **104**(1991), 201–220.

Voiculescu, D. V., Dykema, K. J. and Nica, A. (1992). *Free random variables. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups. CRM*

Monograph Series **1**. American Mathematical Society, Providence, RI. MR1217253