Spiking the random matrix hard edge

Brian Rider (Temple University)

with José A. Ramírez (Universidad de Costa Rica)

Starting with the soft edge

We now know that the largest eigenvalue for fairly generic Wigner or sample covariance matrices converges to the Tracy-Widom law(s).

Here we will just consider the second case, and for now with Gaussian entries. So in particular, if $M = XX^\dagger$ where X is $n \times m$ with independent unit Gaussian entires then there are centerings/scalings $\mu_{n,m}$ and $\sigma_{n,m}$ such that

$$\sigma_{n,m}\Big(\lambda_{max}(M) - \mu_{n,m}\Big) \Rightarrow TW_\beta$$

as $n, m \to \infty$.

The $\beta = 1, 2, 4$ according to whether the entries are real, complex or quaternion Gaussians.

The limiting random variables ("Tracy-Widom(β)") have explicit distribution functions in terms of Painlevé II. Both $\mu_{n,m}$ and $\sigma_{n,m}$ are also explicit – for reference, in the classical setup that

$n \sim m$ it holds that $\mu_{n,m} \sim n$ and $\sigma_{n,m} \sim n^{-1/3}$.

Recalling the magic

In the above setup the joint density of eigenvalues (on $\mathbb{R}_+^n$ is proportional to

$$\prod_{i \neq j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^n w(\lambda_i), \quad w(\lambda) = \lambda^{\frac{\beta}{2}(m-n)-1} e^{-\frac{\beta}{2}\lambda}.$$

For $\beta = 2$ this determines a determinantal process; a Pfaffian process for $\beta = 1, 4$. And everything being explicit in terms of the Laguerre polynomials.

But from a different perspective we also know how to describe the limit law of the maximal point defined by the above density for any $\beta > 0$:

$$TW_\beta = \sup_{f \in L} \frac{2}{\sqrt{\beta}} \int_0^\infty f^2(x) db_x - \int_0^\infty [(f'(x))^2 + x f^2(x)] dx$$

where $L = \{f : \int_0^\infty f^2 = 1, f(0) = 0, \int_0^\infty [(f')^2 + xf^2] < \infty\}$. (Ramírez-R-Virág, 2007).

Tridiagonals

Edelman and Sutton had conjectured such a representation based on the tridiagonal models of Dumitriu-Edelman (2000).

In particular, if, for any $\beta > 0$, we set

$$B = \frac{1}{\sqrt{\beta}} \begin{bmatrix} \chi_{m\beta} & \chi_{(n-1)\beta} & & & \\ & \chi_{(m-1)\beta} & \chi_{(n-2)\beta} & & \\ & & \ddots & \ddots & \\ & & & \chi_{(m-n+2)\beta} & \chi_{\beta} \\ & & & & \chi_{(m-n+1)\beta} \end{bmatrix},$$

then the eigenvalues of the tridiagonal matrix $W = BB^T$ has the desired "general beta Laguerre" joint density.

What is proved is that the sequence of random operators $\sigma_{n,m}(\mu_{n,m} I - W)$ converges (in norm resolvent sense) to

$$\mathsf{SAO}_\beta = -\frac{d^2}{dx^2} + x + \frac{2}{\sqrt{\beta}} b'(x)$$

...on the half-line with Dirichlet conditions at the origin.

One consequences of SAO$_\beta$

Applying the classical Riccati map, one learns that the distribution $F_\beta(\lambda)$ of the Tracy-Widom(β) law is given by the probability that the diffusion

$$dq_t = \frac{2}{\sqrt{\beta}} db_t + (\lambda + t - q_t^2) dt$$

started at $+\infty$ never hits $-\infty$.

One can absorb the spectral parameter λ into a starting time. Hence, if we solve

$$\left(\frac{\partial}{\partial \lambda} + \frac{2}{\beta} \frac{\partial^2}{\partial w^2} + (\lambda - w^2) \frac{\partial}{\partial w} \right) u(\lambda, w) = 0,$$

with the right boundary conditions, then

$$\lim_{w \to \infty} u(\lambda, w) = F_\beta(\lambda).$$

Spiking

Back in the full random matrix set-up, one asks how the limit law of λ_{max} is deformed by altering the population covariance $X\Sigma X^\dagger$ for $\Sigma \neq I$.

What are referred to as the spiked ensembles is when one takes

$$\Sigma = \Sigma_r \oplus I_{m-r}$$

for a fixed r. Sufficient to have in mind that $\Sigma_r = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_\mathrm{r})$.

In 2005 Baik-Ben Arous-Peché found a phase transition phenomena for $\beta = 2$. With $r = 1$:

If $\sigma < \mathfrak{c}$: $\mathbb{P}\Big(\sigma_{n,m}(\lambda_{\mathrm{max}} - \mu_{n,m}) \leq \lambda\Big) \to F_2(\lambda)$.

If $\sigma > \mathfrak{c}$: $\mathbb{P}\Big(\sigma'_{n,m}(\lambda_{\mathrm{max}} - \mu'_{n,m}) \leq \lambda\Big) \to \int_{-\infty}^{t} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$.

If $\sigma = \mathfrak{c} - wn^{-1/3}$: $\mathbb{P}\Big(\sigma_{n,m}(\lambda_{\mathrm{max}} - \mu_{n,m}) \leq \lambda\Big) \to F(\lambda, w) = F_2(\lambda) f(\lambda, w)$ where f can again be described in terms of Painlevé II.

Spiking and tridiagonals

Bloemendal-Virág (2011) noticed that, when $r = 1$ and $\beta = 1, 2, 4$ no problem with classical tridiagonalization procedure.

The corresponding bidiagonal B becomes:

$$B = \frac{1}{\sqrt{\beta}} \begin{bmatrix} \sqrt{\sigma}\chi_{m\beta} & \chi_{(n-1)\beta} & & \\ & \chi_{(m-1)\beta} & \chi_{(n-2)\beta} & \\ & & \ddots & \ddots \end{bmatrix},$$

and afterwards can declare this the general β one-spiked model.

The show the operator limit is still SAO_β, but with the boundary condition

$$f'(0) = wf(0) \quad \text{replacing} \quad f(0) = 0.$$

For Riccati start at w rather than $+\infty$. And so $F(\lambda, w) = P(TW_{\beta,w} \le \lambda)$ solves:

$$\left(\frac{\partial}{\partial \lambda} + \frac{2}{\beta} \frac{\partial^2}{\partial w^2} + (\lambda - w^2) \frac{\partial}{\partial w} \right) F(\lambda, w) = 0.$$

(see Rumanov arXix:1408.3779 for an application to $\beta = 6 \neq 1, 2, 4$!)

More than one spike

If you alter the variance (spike) of more one row/column of X the standard tri/bi-diagonalization procedure gets fouled up.

But you can bi-diagonalize in blocks (of size $r =$ number of spikes): the kth diagonal block having the form,

$$(D_k)_{ij} = \begin{cases} \frac{1}{\sqrt{\beta}} \chi_{\beta(n+a-r(k-1)-i+1)} & \text{if } i = j, \\ g_{ij} & \text{if } j > i, \\ 0 & \text{otherwise,} \end{cases}$$

With similar O_k's. No obvious generalization to $\beta \neq 1, 2, 4$.

Leads to a (vector) generalization of SAO_β:

$$SAO_{\beta,r} = -\frac{d^2}{dx^2} + rx + \sqrt{2}B'(x), \text{ with boundary condition } f'(0) = Wf(0).$$

Here $W = diag(w_1, \ldots, w_r)$ and B is a $\beta = 1, 2, 4$ "Dyson Brownian motion".

The hard edge

Again consider the Gaussian $XX^\dagger$ tuned so that the counting measure of eigenvalues converges: if also $\frac{m}{n} \to \gamma$,

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k/n}(\lambda) \to \frac{\sqrt{(\lambda - a_-)(a_+ - \lambda)}}{2\pi\lambda} \, \mathbf{1}_{[a_-,a_+]} d\lambda$$

where $a_\pm = (1 \pm \sqrt{\gamma})^2$.

When $\gamma > 1$ both edges are "soft", and we have Tracy-Widom fluctuations.

When $\gamma = 1$, then $a_- = 0$ and sees a different type of phenomenon as the eigenvalues now feel the "hard edge" of the origin.

In fact, if $m = n + a$ as $n \uparrow \infty$ there is a one-parameter family of limit laws for λ_{min} indexed by a (first discovered by Tracy-Widom at $\beta = 2$ in terms of Painlevé IV).

As $a \to \infty$ after the fact intuitive that one recovers the (soft-edge) Tracy-Widom laws (and known by way of the explicit formulas at $\beta = 1, 2, 4$.

Hard edge operator

Back in 2009 Ramírez-R. showed the general beta hard edge can be described via the limits of the inverse:

The operator limit of $(nBB^{\dagger})^{-1}$ has the same eigenvalues as that of

$$(\mathfrak{G}f)(x) = \int_0^\infty \int_0^{x \wedge y} s(dz) f(y) m(dy),$$

where

$$m(dx) = e^{-(a+1)x + \frac{2}{\sqrt{\beta}} b(x)} dx$$

and

$$s(dx) = e^{ax + \frac{2}{\sqrt{\beta}} b(x)} dx$$

for a standard Brownian motion $x \mapsto b(x)$.

Note $\mathfrak{G}$ is almost surely trace class on $L^2[m]$.

Hard edge diffusion

There is a differential form of $\mathfrak{G}$. Formally,

$$\mathfrak{G}^{-1} = -e^{x}\left(\frac{d^2}{dx^2} - (a + \frac{2}{\sqrt{\beta}}b'(x))\frac{d}{dx}\right)$$

This gives a Riccati picture.

With $\Lambda(\beta, a)$ the limiting minimal eigenvalue, $P(\Lambda(\beta, a) > \lambda)$ is given by the probability that

$$dp_t = \frac{2}{\sqrt{\beta}}p_t db_t + ((a + \frac{2}{\beta})p_t - p_t^2 - \lambda e^{-t})dt$$

never hits *zero* when started from infinity.

This rescales to the soft-edge diffusion described above, giving the general beta hard-to-soft transition.

Spiking (down) the hard edge

Return now to the spiked tridiagonal model, tuned to see the hard edge: for any $a > -1$,

$$B = \frac{1}{\sqrt{\beta}} \begin{bmatrix} \sqrt{\sigma}\chi_{(n+a)\beta} & \chi_{(n-1)\beta} & & & \\ & \chi_{(n+a-1)\beta} & \chi_{(n-2)\beta} & & \\ & & \ddots & \ddots & \\ & & & \chi_{(a+2)\beta} & \chi_{\beta} \\ & & & & \chi_{(a+1)\beta} \end{bmatrix}.$$

We want to understand the limit law for $\lambda_{min}(BB^{\dagger})$, tuning $\sigma \downarrow 0$ as $n \uparrow \infty$.

Right scaling turns out to be $n\sigma \to c$.

One-spiked hard edge operator

What we show is that the eigenvalues of $(nBB^\dagger)^{-1}$, after an appropriate L^2 embedding, converge to those of the (compact) integral operator:

$$(\mathfrak{G}f)(x) = \int_0^\infty \int_0^{x \wedge y} s(dz) f(y) m(dy) + \frac{1}{c} \int_0^\infty f(y) m(dy),$$

where again: b is a Brownian motion and

$$\mathfrak{m}(dx) = e^{-(a+1)x - \frac{2}{\sqrt{\beta}} b(x)} dx, \quad \mathfrak{s}(dx) = e^{ax + \frac{2}{\sqrt{\beta}} b(x)} dx.$$

This is the resolvent for the diffusion $t \mapsto X_t$ with speed measure $m(dx)$, scale function $\int_0^x s(dx')$, and *killing measure* $c\delta_0(x)$.

Take $t \mapsto \bar{X}_t$ with the same speed and scale, but with simple reflection at the origin. With L_t the local time of $\bar{X}_t$ at the origin, X_t equals $\bar{X}_t$ up to time T defined by

$$\mathbb{P}(T > t \,|\, \bar{X}_\cdot) = e^{-cL_t},$$

at which point the path is killed.

r-spiked hard edge spiking

Recall the setup: $\Sigma = \Sigma_r \oplus I_{m-r}$ and assume $n\Sigma_r \to C$ as $n \to \infty$ ($m = n + a$).

Define the $r \times r$ matrix processes:

$$dA_x = A_x dB_x + (-\frac{a}{2} + \frac{1}{2\beta})A_x dx, \quad A_0 = I_r,$$

with an appropriate $\beta = 1, 2, 4$ matrix Brownian motion $x \mapsto B_x$. Set

$$M_x = e^{-rx} A_x A_x^\dagger, \text{ and } S_x = (A_x A_x^\dagger)^{-1}.$$

Then, the limiting (inverse) spiked hard edge operator reads

$$\mathfrak{G}_r f(x) = \int_0^\infty \left(\int_0^{x \wedge y} S_z dz \right) M_y f(y) dy + C^{-1} \int_0^\infty M_y f(y) dy.$$

So structurally identical, though no nice diffusion description for the differential form.

Aside on the supercritical regime

For $r = 1$, letting $\Lambda(\beta, a, c)$ be the limiting smallest

$$\frac{1}{c}\Lambda(\beta, a, c) \Rightarrow \frac{1}{\beta}\chi^2_{\beta(a+1)} \text{ as } c \to 0.$$

This is the analog of the Gaussian limit at the supercritically spiked soft-edge, and follows from a simple perturbation argument: clearly the limit law is described by (the inverse of)

$$\int_0^\infty m(dx) = \int_0^\infty e^{-(a+1)x - \frac{2}{\sqrt{\beta}} b(x)}\, dx,$$

but that has the same distribution as $\frac{\beta}{\chi^2_{\beta(a+1)}}$ by an old result of Dufresne.

This prompts a multi-variate version of Dufresne's identity. The random matrix

$$\int_0^\infty M_x dx,$$

the running integral of a squared BM on GL_r should have the inverse Wishart law. Proved by R.-Valkó (2014).

r-spiked Riccati

Again there is a "hitting time" (Riccati) description (and so PDEs).

Take the process:

$$dq_{i,t} = \frac{2}{\sqrt{\beta}} q_{i,t} db_i + \left((a + \frac{2}{\beta}) q_{i,t} - q_{i,t}^2 - \lambda e^{-rt} + q_{i,t} \sum_{j \neq i} \frac{q_{i,t} + q_{j,t}}{q_{i,t} - q_{j,t}} \right) dt,$$

begun at

$$(q_{1,0}, q_{2,0}, \ldots, q_{r,0}) = (c_1, \ldots, c_r) = \mathbf{c}.$$

Then

$$P(\Lambda(\beta, a, \mathbf{c}) > \lambda) = P(\mathbf{q} \text{ never hits } 0).$$

"Full" hard-to-soft transition

There is (of course) a r-spiked soft edge diffusion (due to Bloemendal-Virág) which gives the distribution function of $TW_{\beta,\mathbf{w}}$ as the probability of explosion (to $-\infty$) of

$$dp_{i,t} = \frac{2}{\sqrt{\beta}} db_{i,t} + \left(\lambda + rt - p_{i,t}^2 + \sum_{k \neq i} \frac{2}{p_{i,t} - p_{k,t}} \right) dt$$

begun at $\mathbf{w} = (w_1, \dots, w_r)$.

One upshot being

$$\frac{a^2 - \Lambda\Big(\beta, 2a, \mathbf{c}(a)\Big)}{a^{4/3}} \Rightarrow TW_{\beta,\mathbf{w}}$$

as $a \to \infty$ granted that

$$\lim_{a \to \infty} a^{-2/3}(c_i(a) - a) = w_i \in (-\infty, \infty], \quad i \in [1, r].$$

Some curious questions

You can, for $\beta = 1, 2, 4$ employ the r-spiked format (r-block model) when there are less than r spikes.

For instance, for the classical beta the $r = 1$ hard edge operator

$$(\mathfrak{G} f)(x) = \int_0^\infty \left(\int_0^{x \wedge y} s_z \, dz \right) f(y) \, m_y dy + \frac{1}{c} \int_0^\infty f(y) m_y \, dy$$

has to have the same spectrum as the 2×2 version

$$(\mathfrak{G}_2) f(x) = \int_0^\infty \left(\int_0^{x \wedge y} S_z dz \right) M_y f(y) dy + \begin{pmatrix} c^{-1} & 0 \\ 0 & 0 \end{pmatrix} \int_0^\infty M_y f(y) dy.$$

Pretty opaque!

In the diffusion format

Again for example, the r-spiked and $(r-1)$-spiked hard edge diffusions are related as thus (here we must take $\beta = 1, 2$ or 4).

The first hitting time to zero of

$$dq_{i,t} = \frac{2}{\sqrt{\beta}} q_{i,x} db_i + \left((a + \frac{2}{\beta}) q_{i,t} - q_{i,t}^2 - \lambda e^{-rt} + q_{i,t} \sum_{j \neq i \in [1,r]} \frac{q_{i,t} + q_{j,t}}{q_{i,t} - q_{j,t}} \right) dt$$

started from

$$q_{1,0} = +\infty, \ q_{2,0} = c_1, \ \ldots, \ q_{r,0} = c_{r-1}$$

is the same as the process on $r-1$ points (with the appropriate substitution) started from

$$q_{1,0} = c_1, \ \ldots, \ q_{r-1,0} = c_{r-1}.$$

Back to the soft edge

The same issue (of course) was known to Bloemendal-Virág. To reiterate in this context, for any $r = 1, 2, 3, \ldots$ the explosion (to $-\infty$) probability of the joint process,

$$t \mapsto p_{1,t}, \ldots, p_{r,t} \qquad \text{with } p_{1,0} = \cdots = p_{r,0} = +\infty$$

given by

$$dp_{i,t} = \frac{2}{\sqrt{\beta}} db_{i,t} + \left(\lambda + rt - p_{i,t}^2 + \sum_{j \neq i \in [1,r]} \frac{2}{p_{i,t} - p_{j,t}} \right) dt$$

is the same, or $F_\beta(\lambda) =$ the distribution of Tracy-Widom(β).

Again, no direct proof. And only can claim this for $\beta = 1, 2, 4$ but is presumably the case for all β.