A TRANSPORTATION APPROACH TO UNIVERSALITY IN RANDOM MATRIX THEORY

A. FIGALLI[†]

ABSTRACT. In this note we discuss a new recent approach, based on transportation techniques, to obtain universality results in random matrix theory.

Large random matrices appear in many different fields, including quantum mechanics, quantum chaos, telecommunications, finance, and statistics. As such, understanding how the asymptotic properties of the spectrum depend on the fine details of the model, in particular on the distribution of the entries, soon appeared as a central question.

1. WIGNER MATRICES

1.1. **Empirical measures and the semicircle law.** An important model is the one of Wigner matrices, that is, Hermitian matrices with independent and identically distributed (i.i.d.) real or complex entries with mean zero and covariance $1/N$, N being the dimension of the matrix. In other words, our random matrices has the form

$$A_N = \begin{pmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,N} \\ X_{2,1} & X_{2,2} & \dots & X_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N,1} & X_{N,2} & \dots & X_{N,N} \end{pmatrix}$$

where the entries $\{X_{i,j}\}_{1\le i\le j\le N}$ are i.i.d., $\mathbb{E}[X_{i,j}]=0$, and $\mathbb{E}\big[|X_{i,j}|^2\big]=1/N$.[1] Also, the condition of our matrices being symmetric means that $X_{j,i}$ is equal to the complex conjugate of $X_{i,j}$.

Let $\lambda_1 \le \ldots \le \lambda_N$ be the eigenvalue of A_N, and consider the empirical measure

$$L_N := \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}.$$

Note that since the matrices A_N are Hermitian, their eigenvalues are real, hence L_N is a random probability measure on the real line.

Let $\rho_{\rm sc}(x)$ denote the semi-circle distribution on the real line, that is

$$\rho_{\rm sc}(x) := \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{\{|x|\le 2\}} \qquad \forall\, x\in\mathbb{R}.$$

It was shown by Wigner [Wig55] that the empirical measures L_N converge in probability to the semi-circle law $\mu_{\rm sc}(dx) := \rho_{\rm sc}(x)\,dx$. More precisely, if $\mathbb{P}^N$ denotes the law of the eigenvalues, then

[†] The University of Texas at Austin, Mathemtics Dept. RLM 8.100, 2515 Speedway Stop C1200, Austin, Texas 78712-1202, USA. email: figalli@math.utexas.edu.

[1]Actually, the exact definition of Wigner matrix is slightly more involved. Indeed, first of all one assume that all moments of $X_{i,j}$ are finite. Also, the assumptions on the diagonal coefficients $X_{i,i}$ are slightly different from the non-diagonal ones. Finally, some assumptions also depend on whether one considers real or complex coefficients. We refer to [AGZ10, Chapters 2.1 and 2.2] for more details.

for any $f : \mathbb{R} \to \mathbb{R}$ continuous and bounded, and for any $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}^N\big(|\langle f, L_N\rangle - \langle f, \mu_{\rm sc}\rangle| > \varepsilon\big) = 0, \tag{1.1}$$

where, given a continuous function g and a measure ν, we use the notation

$$\langle g, \nu\rangle := \int g\, d\nu.$$

In other words, even if the measures L_N are random, as $N \to \infty$ they all behave as the deterministic measure $\mu_{\rm sc}$.

1.2. **On the fluctuation of eigenvalues: a heuristic argument.** At least heuristically, Wigner's result can be used to get some insight about the fluctuation of consecutive eigenvalues.

Let us order the eigenvalues so that $\lambda_1 \leq \ldots \leq \lambda_N$. Then, since L_N behaves as $\mu_{\rm sc}$, given $i \in \{1, \ldots, N\}$ we have

$$\int_{-\infty}^{\lambda_i} \rho_{\rm sc}(x)\, dx \approx \int_{-\infty}^{\lambda_i} dL_N(x) = \frac{1}{N}\int_{-\infty}^{\lambda_i} \sum_{j=1}^N \delta_{\lambda_j}(dx) = \frac{1}{N}\#\{j \,:\, j \leq i\} = \frac{i}{N}. \tag{1.2}$$

To exploit this fact, define the function

$$F_{\rm sc}(x) := \int_{-\infty}^{x} \rho_{\rm sc}(y)\, dy \qquad \forall\, x \in \mathbb{R},$$

so that (1.2) can be rewritten as

$$F_{\rm sc}(\lambda_i) \approx \frac{i}{N}. \tag{1.3}$$

We note that $F_{\rm sc} : \mathbb{R} \to [0, 1]$ satisfies

$$\begin{cases} F_{\rm sc} \equiv 0 & \text{on } (-\infty, -2], \\ F_{\rm sc} \equiv 1 & \text{on } [2, \infty), \\ F_{\rm sc}' > 0 & \text{on } (-2, 2). \end{cases}$$

Then, it follows from (1.3) that

$$\lambda_{i+1} - \lambda_i \approx F_{\rm sc}^{-1}\Big(\frac{i+1}{N}\Big) - F_{\rm sc}^{-1}\Big(\frac{i}{N}\Big) \approx \big(F_{\rm sc}^{-1}\big)'\Big(\frac{i}{N}\Big)\frac{1}{N} = \frac{1}{\rho_{\rm sc} \circ F_{\rm sc}^{-1}\big(\frac{i}{N}\big)}\frac{1}{N} \tag{1.4}$$

Note now that if $i \in [\varepsilon N, (1-\varepsilon)N]$ for some $\varepsilon > 0$ small (in this case we say that λ_i belongs to the *bulk*), then

$$0 < c_\varepsilon := \min\big\{\rho_{\rm sc} \circ F_{\rm sc}^{-1}(\varepsilon), \rho_{\rm sc} \circ F_{\rm sc}^{-1}(1-\varepsilon)\big\} \leq \rho_{\rm sc} \circ F_{\rm sc}^{-1}\left(\frac{i}{N}\right) \leq \max_{[-2,2]} \rho_{\rm sc} = \frac{1}{\pi}.$$

that combined with (1.4) implies that $N(\lambda_{i+1} - \lambda_i)$ is of order 1.

On the other hand, if $1 \leq i \leq C$ for some fixed number C independent of N (in this case one says that λ_i belongs to the *edge*), since

$$\rho_{\rm sc} \approx (x+2)^{1/2} \qquad \text{and} \qquad F_{\rm sc}(x) \approx (x+2)^{3/2} \qquad \text{for } x \text{ close to } -2,$$

we get

$$\rho_{\rm sc} \circ F_{\rm sc}^{-1}\left(\frac{i}{N}\right) \approx \rho_{\rm sc}\left(\Big(\frac{i}{N}\Big)^{2/3} - 2\right) \approx \Big(\frac{i}{N}\Big)^{1/3},$$

so, recalling (1.4),

$$\lambda_{i+1} - \lambda_i \approx \Big(\frac{N}{i}\Big)^{1/3}\frac{1}{N} \approx \frac{1}{N^{2/3}}.$$

Thus, by this heuristic argument we have obtained the following "intuition":

The difference of two consecutive eigenvalues $\lambda_{i+1} - \lambda_i$ should fluctuate at scale N^{-1} in the bulk, and at scale $N^{-2/3}$ near the edge.

1.3. **On the fluctuation of eigenvalues: rigorous results.**

Making rigorous the intuition above took many years.

The first approach to the study of local fluctuations of the spectrum was based on exact models, namely the Gaussian models (i.e., the entries of our matrix are Gaussian). In this case the law of the eigenvalues is given by the measure

$$\mathbb{P}_G^N(d\lambda_1, \dots, d\lambda_n) := \frac{1}{Z_G^N} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-N\sum_i \lambda_i^2}\, d\lambda_1 \dots d\lambda_N,$$

where

$$Z_G^N := \int_{\mathbb{R}^N} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-N\sum_i \lambda_i^2}\, d\lambda_1 \dots d\lambda_N,$$

and

$$\beta = \begin{cases} 1 & \text{if the entries of the Gaussian matrices are real,} \\ 2 & \text{if the entries of the Gaussian matrices are complex.} \end{cases}$$

As predicted above, in this case the eigenvalues fluctuate near ± 2 at scale $N^{-2/3}$, and the limit distribution of these fluctuations is given by the so-called Tracy-Widom law [TW94a, TW94b]. On the other hand, inside the bulk the distance between two consecutive eigenvalues is of order N^{-1} and the fluctuations at this scale can be described by the sine-Kernel distribution.

In other words, if $\mathbf{P}_G^N$ denote the distribution of the increasingly ordered eigenvalues under $\mathbb{P}_G^N$, using the notation $\hat\lambda := (\lambda_1, \dots, \lambda_N) \in \mathbb{R}^N$, these results give a precise description for the limit as $N \to \infty$ of integrals of the form

$$\int_{\mathbb{R}^N} f\big(N(\lambda_{i+1} - \lambda_i)\big)\, d\mathbf{P}_G^N(\hat\lambda), \qquad \lambda_i \text{ in the bulk,}$$

$$\int_{\mathbb{R}^N} f\big(N^{2/3}(\lambda_i + 2)\big)\, d\mathbf{P}_G^N(\hat\lambda), \qquad \lambda_i \text{ at the edge,}$$

when $f : \mathbb{R} \to \mathbb{R}^+$ is a fixed test function.

Although these results were first obtained only for the Gaussian models, it was already envisioned by Wigner that these fluctuations should be *universal*, i.e., independent of the precise distribution of the entries. This deep fact was proved only recently in a series of remarkable papers [Erd10, EPR+10, ESYY12, EYY12, EY12b, TV12, TV11, TV10, Tao13].

2. Universality results for β-models

A related question to the one described above consists in studying universality for local fluctuations for the so-called β-models. These are laws of particles in interaction according to a Coulomb-gas potential to the power β and to a general potential V. More precisely, one is interested in understanding the fluctuations of the λ_i's when they are distributed according to the probability measure

$$\mathbb{P}_V^N(d\lambda_1, \dots, d\lambda_n) := \frac{1}{Z_V^N} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-N\sum_i V(\lambda_i)}\, d\lambda_1 \dots d\lambda_N,$$

where $\beta > 0$,

$$Z_V^N := \int_{\mathbb{R}^N} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-N\sum_i V(\lambda_i)}\, d\lambda_1 \dots d\lambda_N,$$

and $V : \mathbb{R} \to \mathbb{R}$ is a smooth interaction potential.

As in the Gaussian case, also for β-models with "nice" interaction potentials (for instance, when V is uniformly convex) the empirical measure converge to a stationary distribution $\mu_V(dx) := \rho_V(x)\,dx$ which can be written as

$$\rho_V(x) := S_V(x)\sqrt{(x-a_V)(b_V-x)}\,\mathbf{1}_{\{a_V\le x\le b_V\}} \qquad \forall\, x\in\mathbb{R},$$

where $S_V : [a_V, b_V] \to \mathbb{R}$ is smooth and uniformly bounded away from zero. However, also in this case one would like to prove that the local fluctuations of the spectrum are ***universal***, that is, *independent* of the potential V.

Local fluctuations were first studied in the case $\beta = 2$ in [Meh04, PS97, DKM+99, LL08]. Then, in the cases $\beta = 1, 4$, universality was proved in [DG07b, DG07a, Shc09, KS10, Shc12] (see also [DG09] for a review). Finally, the local fluctuations of more general β-ensembles were only derived recently [VV09, RRV11] in the Gaussian case, and universality for general β-ensembles was obtained in [BEY14a, EY12a, BEY14b, KRV13, Shc13, BFG15, Bek15, FG16].

Our goal here is to describe the general transportation approach introduced in [BFG15] and further developed and improved in [FG16] to prove universality in this setting.

2.1. **Preliminary considerations.** The general strategy is to find a change of variable $T_N : \mathbb{R}^N \to \mathbb{R}^N$ that allows us to relate the fluctuation of the eigenvalues distributed according to $\mathbb{P}^N_V$ to the fluctuations with respect to $\mathbb{P}^N_G$.

More precisely, since we want to consider the fluctuations of consecutive eigenvalues after ordering, we have to consider the laws of the ordered eigenvalues rather than the original laws $\mathbb{P}^N_V$ and $\mathbb{P}^N_G$ (where the λ_i's are completely symmetric). Hence, let $\mathbf{P}^N_V$ denote the distribution of the increasingly ordered eigenvalues under $\mathbb{P}^N_V$, and let $\mathbf{P}^N_G$ denote the distribution of the increasingly ordered eigenvalues under $\mathbb{P}^N_G$. We want to find a "nice" map T_N such that $(T_N)_\#\mathbf{P}^N_G = \mathbf{P}^N_V$, that is,

$$\int_{\mathbb{R}^N}\psi\,d\mathbf{P}^N_V = \int_{\mathbb{R}^N}\psi\circ T_N\,d\mathbf{P}^N_G \qquad \forall\,\psi:\mathbb{R}^N\to\mathbb{R} \text{ continuous and bounded.}$$

Indeed, if $T_N = (T^1_N,\dots,T^N_N)$ is such a map, then it follows by the above condition applied with $\psi(\hat\lambda) := f(N(\lambda_{i+1}-\lambda_i))$ that

$$\int_{\mathbb{R}^N} f\big(N(\lambda_{i+1}-\lambda_i)\big)\,d\mathbf{P}^N_V(\hat\lambda) = \int_{\mathbb{R}^N} f\big(N\big(T^{i+1}_N(\hat\lambda)-T^i_N(\hat\lambda)\big)\big)\,d\mathbf{P}^N_G(\hat\lambda) \tag{2.1}$$

Hence, proving a universality result in the bulk corresponds to showing that the right hand side above behaves as in the Gaussian case, that is,

$$\int_{\mathbb{R}^N} f\big(N\big(T^{i+1}_N(\hat\lambda)-T^i_N(\hat\lambda)\big)\big)\,d\mathbf{P}^N_G(\hat\lambda) = \int_{\mathbb{R}^N} f\big(c_i\,N(\lambda_{i+1}-\lambda_i)\big)\,d\mathbf{P}^N_G(\hat\lambda) + o_N(1), \tag{2.2}$$

where $c_i > 0$ is a constant, and $o_N(1)$ is a quantity that goes to 0 as $N\to\infty$.

Thus, for this strategy to work, we need to find a transport map T^N such that the difference $T^{i+1}_N(\hat\lambda) - T^i_N(\hat\lambda)$ between two consecutive components, that a priori depends on all eigenvalues $\lambda_1,\dots\lambda_N$, actually behaves as $c_i(\lambda_{i+1}-\lambda_i)$ for some constant $c_i>0$, up to a small error.

As we shall see later, instead of finding a transport map from $\mathbf{P}^N_G$ to $\mathbf{P}^N_V$, it will be enough to construct a "nice" map that transports $\mathbb{P}^N_G$ onto $\mathbb{P}^N_V$. In other words, at the moment we do not need to worry about ordering the eigenvalues.

Before describing our construction, we first make a simple observation: let $\mu(dx) := f(x)\,dx$ and $\nu(dy) := g(y)\,dy$ be two probability measures in $\mathbb{R}^n$, and suppose that $T:\mathbb{R}^n\to\mathbb{R}^n$ is a smooth diffeomorphism transporting μ onto ν. Then it follows by the transport condition and the standard change of variable formula that

$$\int_{\mathbb{R}^n}\psi(T(x))\,f(x)\,dx = \int_{\mathbb{R}^n}\psi(y)\,g(y)\,dy = \int_{\mathbb{R}^n}\psi(T(x))\,g(T(x))\,\big|\mathrm{det}DT(x)\big|\,dx.$$

By the arbitrariness of ψ, this implies the validity of the Jacobian equation

$$f(x) = g(T(x))\,\big|\mathrm{det} DT(x))\big|. \tag{2.3}$$

2.2. **A possible flow construction for a transport map.** To build a transport map, we argue as follows: first of all, we construct a curve of measures connecting $\mathbb{P}_G^N$ to $\mathbb{P}_V^N$ by setting $V_t(x) := (1-t)x^2 + tV(x)$ and defining

$$\mathbb{P}_{V_t}^N(d\hat\lambda) := \sigma_t(\hat\lambda)\,d\hat\lambda, \qquad \sigma_t(\hat\lambda) := \frac{1}{Z_{V_t}^N}\prod_{i<j}|\lambda_i-\lambda_j|^\beta e^{-N\sum_i V_t(\lambda_i)}, \qquad \hat\lambda := (\lambda_1,\dots,\lambda_N).$$

Note that, with this definition,

$$\mathbb{P}_{V_0}^N = \mathbb{P}_G^N \qquad \text{and} \qquad \mathbb{P}_{V_1}^N = \mathbb{P}_V^N,$$

hence the curve $[0,1]\ni t\mapsto \mathbb{P}_{V_t}^N$ interpolates between our two measures.

We now observe that, given two probability measures $\mathbb{P}_{V_t}^N$ to $\mathbb{P}_{V_s}^N$ as above with $0\le t\le s\le 1$, a diffeomophism $T_{t,s}:\mathbb{R}^N\to\mathbb{R}^N$ is a transport from $\mathbb{P}_{V_t}^N$ to $\mathbb{P}_{V_s}^N$ if and only if

$$\big|\det(DT_{t,s})\big| = \frac{\sigma_t}{\sigma_s\circ T_{t,s}} \tag{2.4}$$

(see (2.3)). Also, when $t=s$ it is natural to take $T_{t,t}$ as the identity map. Hence, assuming that $T_{t,s}$ depends smoothly on the parameters t,s, there should exists a map $\mathbf{Y}_t:\mathbb{R}^N\to\mathbb{R}^N$ such that

$$T_{t,s} = \mathrm{Id} + (s-t)\mathbf{Y}_t + o(s-t). \tag{2.5}$$

If we plug this expression into (2.4) and we expand all terms (note that if $T_{t,s}$ is close to the identity map then $|\det(DT_{t,s})| = \det(DT_{t,s}))$ we get[2]

$$\begin{aligned} &1+(s-t)\,\mathrm{div}\mathbf{Y}_t + o(s-t)\\ &\qquad = 1+(s-t)\Big(c_t - \beta\sum_{i<j}\frac{\mathbf{Y}_t^i-\mathbf{Y}_t^j}{\lambda_i-\lambda_j} + N\sum_i W(\lambda_i) + N\sum_i V_t'(\lambda_i)\mathbf{Y}_t^i\Big) + o(s-t)\end{aligned}$$

where

$$W := V_1 - V_0 = V - x^2 \qquad \text{and} \qquad c_t := \frac{d}{dt}\left(\log Z_{V_t}^N\right).$$

Hence, by letting $s\to t$ we discover that $\mathbf{Y}_t = (\mathbf{Y}_t^1,\dots,\mathbf{Y}_t^N)$ should solve

$$\mathrm{div}\mathbf{Y}_t = c_t - \beta\sum_{i<j}\frac{\mathbf{Y}_t^i-\mathbf{Y}_t^j}{\lambda_i-\lambda_j} + N\sum_i W(\lambda_i) + N\sum_i V_t'(\lambda_i)\mathbf{Y}_t^i. \tag{2.6}$$

Although this is a formal argument, it suggests a way to construct maps $T_{0,t}:\mathbb{R}^N\to\mathbb{R}^N$ from $\mathbb{P}_{V_0}^N$ to $\mathbb{P}_{V_t}^N$: indeed, if $T_{0,t}$ sends $\mathbb{P}_{V_0}^N$ onto $\mathbb{P}_{V_t}^N$ then $T_{t,s}\circ T_{0,t}$ sends $\mathbb{P}_{V_0}^N$ onto $\mathbb{P}_{V_s}^N$. Hence, if we impose the condition

$$T_{0,s} = T_{t,s}\circ T_{0,t},$$

by differentiating this relation with respect to s and setting $s=t$ we obtain (recall that $Y_t = \partial_s T_{t,s}|_{s=t}$, see (2.5))

$$\partial_t T_{0,t} = \mathbf{Y}_t(T_{0,t}), \qquad T_{0,0} = \mathrm{Id}.$$

[2]Recall that, given a matrix $A\in\mathbb{R}^{n\times n}$,

$$\det(\mathrm{Id}+\varepsilon A) = 1+\varepsilon\,\mathrm{tr}(A) + o(\varepsilon),$$

hence, for any vector field $Y:\mathbb{R}^n\to\mathbb{R}^n$,

$$\det(\mathrm{Id}+\varepsilon\,DY) = 1+\varepsilon\,\mathrm{tr}(DY)+o(\varepsilon) = 1+\varepsilon\,\mathrm{div}Y + o(\varepsilon).$$

In other words, to construct a transport map from $\mathbb{P}^N_{V_0}$ onto $\mathbb{P}^N_{V_1}$ we can simply solve the ODE associated to the vector-field $\mathbf{Y}_t$, as the following lemma shows:

Lemma 2.1. *Let $\mathbf{Y}_t : \mathbb{R}^N \to \mathbb{R}^N$ solve* (2.6) *and define its flow*

$$\begin{cases} \dot{X}_t = \mathbf{Y}_t(X_t), \\ X_0 = \mathrm{Id}. \end{cases}$$

Then X_1 transports $\mathbb{P}^N_{V_0}$ onto $\mathbb{P}^N_{V_1}$.

This result is a particular case of Lemma 2.2 below (just take $\mathcal{R}_t \equiv 0$ and $K_q = 0$ there).

Thanks to the Lemma 2.1, to construct a transport map T_N from $\mathbb{P}^N_{V_0}$ onto $\mathbb{P}^N_{V_1}$ we may try to use the following strategy:
- find a solution of (2.6);
- solve the ODE $\dot{X}_t = \mathbf{Y}_t(X_t)$ with $X_0 = \mathrm{Id}$;
- set $T_N := X_1$.

2.3. **Approximate solutions of** (2.6) **induce approximate transport maps.** Although the argument in the previous section has reduced the issue of finding a transport map to the "simpler" problem of solving the linear equation (2.6) for any $t \in [0,1]$, finding solutions of (2.6) that enjoy "nice" regularity estimates that are uniform in N seems extremely difficult.

However, if we go back to (2.1), we can observe that we do not need an exact equality there, but it is enough that the two terms in the equation are equal up to an error that goes to 0 as $N \to \infty$. As the next lemma tells us, if we are able to solve (2.6) up to a small error, then the flow of $\mathbf{Y}_t$ will produce an approximate transport map.

Lemma 2.2. *Set*

$$\mathcal{R}_t(\mathbf{Y}_t) := c_t - \beta \sum_{i<j} \frac{\mathbf{Y}^i_t - \mathbf{Y}^j_t}{\lambda_i - \lambda_j} + N \sum_i W(\lambda_i) + N \sum_i V'_t(\lambda_i)\mathbf{Y}^i_t - \mathrm{div}\mathbf{Y}_t, \tag{2.7}$$

and let X_t denote the flow of $\mathbf{Y}_t$. Assume that, for any $q < \infty$, there exists a constant K_q such that

$$\|\mathcal{R}_t(\mathbf{Y}_t)\|_{L^q(\mathbb{P}^N_{V_t})} \le K_q \frac{(\log N)^3}{N} \qquad \forall\, t \in [0,1], \tag{2.8}$$

and set $T^N := X_1$. Let $\chi : \mathbb{R}^N \to \mathbb{R}^+$ be a nonnegative measurable function satisfying $\|\chi\|_\infty \le N^k$ for some $k \ge 0$. Then, for any $\eta > 0$ there exists a constant $C_{k,\eta}$, depending only on k and η, such that

$$\left| \log\biggl(1 + \int_{\mathbb{R}^N} \chi \, d\mathbb{P}^N_{V_1}\biggr) - \log\biggl(1 + \int_{\mathbb{R}^N} \chi \circ T_N \, d\mathbb{P}^N_{V_0}\biggr) \right| \le C_{k,\eta}\, K_q\, N^{\eta-1}.$$

Remark 2.3. Observe that, because of the presence of a logarithm, the estimate provided by Lemma 2.2 does not imply that (2.1) holds up to a small error. However, as we shall see later, this is not a big issue: indeed, whenever a, b are positive numbers with $b \le C$ for some universal constant C, then the bound

$$|\log(1+a) - \log(1+b)| \le C\, N^{\eta-1}$$

implies that

$$|a - b| \le C\, N^{\eta-1}.$$

Proof. Recall that σ_t denotes the density of $P^N_{V_t}$ with respect to the Lebesgue measure. Then, by a direct computation one can check that σ_t, $\mathbf{Y}_t$, and $\mathcal{R}_t := \mathcal{R}_t(\mathbf{Y}_t)$ are related by the following formula:

$$\partial_t\sigma_t + \operatorname{div}(\mathbf{Y}_t\,\sigma_t) = \mathcal{R}_t\,\sigma_t. \tag{2.9}$$

Now, given a smooth function $\chi : \mathbb{R}^N \to \mathbb{R}^+$ satisfying $\|\chi\|_\infty \le N^k$, we define

$$\chi_t := \chi \circ X_1 \circ (X_t)^{-1} \qquad \forall\, t \in [0,1]. \tag{2.10}$$

Note that, with this definition, $\chi_1 = \chi$. Also, since by construction the function $\chi_t \circ X_t = \chi \circ X_1$ is constant in time, differentiating with respect to t we deduce that

$$0 = \frac{d}{dt}\big(\chi_t \circ X_t\big) = \Big(\partial_t\chi_t + \mathbf{Y}_t\cdot\nabla\chi_t\Big)\circ X_t,$$

where we used that $\partial_t X_t = \mathbf{Y}_t \circ X_t$. Hence, this proves that χ_t solves the transport equation

$$\partial_t\chi_t + \mathbf{Y}_t\cdot\nabla\chi_t = 0, \qquad \chi_1 = \chi. \tag{2.11}$$

Combining (2.9) and (2.11), and integrating the divergence by parts, we get

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^N}\chi_t\,\sigma_t &= \int_{\mathbb{R}^N}\partial_t\chi_t\,\sigma_t + \int_{\mathbb{R}^N}\chi_t\,\partial_t\sigma_t\\ &= -\int_{\mathbb{R}^N}\mathbf{Y}_t\cdot\nabla\chi_t\,\sigma_t - \int_{\mathbb{R}^N}\chi_t\operatorname{div}(\mathbf{Y}_t\,\sigma_t) + \int_{\mathbb{R}^N}\chi_t\,\mathcal{R}_t\,\sigma_t = \int_{\mathbb{R}^N}\chi_t\,\mathcal{R}_t\,\sigma_t.\end{aligned}$$

We now want to control the last term. To this aim we notice that, since $\|\chi\|_\infty \le N^k$, it follows immediately from (2.10) that $\|\chi_t\|_\infty \le N^k$ for any $t \in [0,1]$. Hence, using Hölder inequality and (2.8), for any $p > 1$ we can bound

$$\begin{aligned}\left|\int_{\mathbb{R}^N}\chi_t\,\mathcal{R}_t\,\sigma_t\right| &\le \|\chi_t\|_{L^p(\mathbb{P}^N_{V_t})}\|\mathcal{R}_t\|_{L^q(\mathbb{P}^N_{V_t})} \le \|\chi_t\|_\infty^{\frac{p-1}{p}}\,\|\chi_t\|^{1/p}_{L^1(\mathbb{P}^N_{V_t})}\|\mathcal{R}_t\|_{L^q(\mathbb{P}^N_{V_t})}\\ &\le N^{\frac{k(p-1)}{p}}\,\|\chi_t\|^{1/p}_{L^1(\mathbb{P}^N_{V_t})}\|\mathcal{R}_t\|_{L^q(\mathbb{P}^N_{V_t})} \le K_q\,\frac{N^{\frac{k(p-1)}{p}}(\log N)^3}{N}\,\|\chi_t\|^{1/p}_{L^1(\mathbb{P}^N_{V_t})},\end{aligned}$$

where $q := \frac{p}{p-1}$. Thus, given $\eta > 0$, we can choose $p := 1 + \frac{\eta}{2k}$ to obtain

$$\left|\int_{\mathbb{R}^N}\chi_t\,\mathcal{R}_t\,\sigma_t\right| \le C\,K_q\,N^{\eta-1}\|\chi_t\|^{1/p}_{L^1(\mathbb{P}^N_{V_t})} \le C\,K_q\,N^{\eta-1}\Big(1 + \|\chi_t\|_{L^1(\mathbb{P}^N_{V_t})}\Big),$$

where C depends only on k and η. Therefore, setting

$$Z(t) := \int_{\mathbb{R}^N}\chi_t\,\sigma_t = \|\chi_t\|_{L^1(\mathbb{P}^N_{V_t})}$$

(recall that $\chi_t \ge 0$), we proved that

$$|\dot Z(t)| \le C\,K_q\,N^{\eta-1}\big(1 + Z(t)\big) \qquad \forall\, t \in [0,1],$$

which implies that

$$\big|\log\big(1 + Z(1)\big) - \log\big(1 + Z(0)\big)\big| \le C\,K_q\,N^{\eta-1}.$$

Recalling that $\chi_0 = \chi \circ X_1$, $\chi_1 = \chi$, and $T_N = X_1$, this proves the desired result when χ is smooth. By approximation the result extends to all measurable functions $\chi : \mathbb{R}^N \to \mathbb{R}^+$ satisfying $\|\chi\|_\infty \le N^k$, concluding the proof. □

As explained before, thanks to this lemma we know that if (2.6) holds up to a small error, then the flow of $\mathbf{Y}_t$ provides an approximate transport map. Hence, the next step consists in finding a vector field $\mathbf{Y}_t$ satisfying (2.8).

2.4. **Construction of $\mathbf{Y}_t$.** Fix $t \in [0,1]$. To find an approximate solution of (2.6) we shall make an ansatz.

The idea is that at first order eigenvalues do not interact, then at order $1/N$ eigenvalues interact at most by pairs, and so on. As we shall see, to construct a vector field satisfying (2.8) it will be enough to consider at most pairwise interactions.

Note that, by symmetry, it is clear that the interaction of λ_i with λ_j, $i \neq j$, should be equal to the one of λ_i with λ_k for any other index $k \neq i$. Hence, the term involving pairwise interactions can be written as an interaction of λ_i with the empirical measure $L_N = \frac{1}{N}\sum_j \delta_{\lambda_j}$. Recalling that

$$L_N \rightharpoonup^* \mu_{V_t} := \rho_{V_t}\, dx \qquad \text{as } N \to \infty,$$

it will actually be convenient to rewrite that pairwise interaction in term of the measure $L_N - \mu_{V_t}$ rather than L_N. Moreover, for normalization purposes, we will actually consider the measure $M_N := N(L_N - \mu_{V_t})$.

These considerations suggest us the following ansatz for the components of $\mathbf{Y}_t = (\mathbf{Y}_t^1, \dots, \mathbf{Y}_t^N)$:

$$\mathbf{Y}_t^i(\lambda_1, \dots, \lambda_N) := \mathbf{y}_{0,t}(\lambda_i) + \frac{1}{N}\mathbf{y}_{1,t}(\lambda_i) + \frac{1}{N}\boldsymbol{\xi}_t(\lambda_i, M_N) \qquad \forall\, i = 1 \dots, N, \tag{2.12}$$

where

$$\boldsymbol{\xi}_t(x, M_N) := \int_{\mathbb{R}} \mathbf{z}_t(x,y)\, dM_N(y) = \sum_j \mathbf{z}_t(x, \lambda_j) - N \int_{\mathbb{R}} \mathbf{z}_t(x,y)\, d\mu_{V_t}(y),$$

and $\mathbf{y}_{0,t} : \mathbb{R} \to \mathbb{R}$, $\mathbf{y}_{1,t} : \mathbb{R} \to \mathbb{R}$, and $\mathbf{z}_t : \mathbb{R}^2 \to \mathbb{R}$ are functions to be determined.

To be able to compute $\mathcal{R}_t(\mathbf{Y}_t)$, it will be important to use some explicit properties of the measure μ_{V_t}. More precisely, as shown in [AG97, AGZ10], the measure μ_{V_t} is characterized as the unique minimizer, among probability measures on $\mathbb{R}$, of the functional

$$\mu \mapsto I_{V_t}(\mu) := \frac{1}{2} \iint_{\mathbb{R}^2} \big(V_t(x) + V_t(y) - \beta \log|x-y|\big)\, d\mu(x)\, d\mu(y).$$

Hence, given a smooth function $h : \mathbb{R} \to \mathbb{R}$, considering the family of measures $\mu_{t,\varepsilon} := (\mathrm{Id} + \varepsilon h)_{\#}\mu_t$ and writing that $I_{V_t}(\mu_{t,\varepsilon}) \geq I_{V_t}(\mu_t)$, by taking the derivative with respect to ε and setting $\varepsilon = 0$, we get

$$\int_{\mathbb{R}} V_t'(x)\, h(x)\, d\mu_t(x) = \frac{\beta}{2} \iint_{\mathbb{R}^2} \frac{h(x) - h(y)}{x-y}\, d\mu_t(x)\, d\mu_t(y) \tag{2.13}$$

for all smooth (say, C^1) functions $h : \mathbb{R} \to \mathbb{R}$.

Therefore, using the convention that

$$\iint_{\mathbb{R}^2} \frac{h(x)-h(y)}{x-y}\, \delta_{\lambda_i}(dx) \otimes \delta_{\lambda_j}(dy) = \begin{cases} \frac{h(\lambda_i) - h(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ h'(\lambda_i) & \text{if } \lambda_i = \lambda_j, \end{cases}$$

and defining the operator

$$\Xi_t : C^1(\mathbb{R}) \to C^0(\mathbb{R}), \qquad \Xi_t[h](x) := -\beta \int_{\mathbb{R}} \frac{h(x)-h(y)}{x-y}\, d\mu_{V_t}(y) + V_t'(x)\, h(x), \tag{2.14}$$

it follows by a direct computation using (2.13) that

$$\begin{aligned} \mathcal{R}_t(\mathbf{Y}_t) = {} & N \int_{\mathbb{R}} \Big(\Xi_t[\mathbf{y}_{0,t}](x) + W(x)\Big)\, dM_N(x) \\ & + \int_{\mathbb{R}} \left(\Xi_t[\mathbf{y}_{1,t}](x) + \Big(\frac{\beta}{2} - 1\Big)\left[\mathbf{y}_{0,t}'(x) + \int_{\mathbb{R}} \partial_1 \mathbf{z}_t(z,x) d\mu_{V_t}(z)\right]\right) dM_N(x) \\ & + \iint_{\mathbb{R}^2} dM_N(x)\, dM_N(y) \left(\Xi_t[\mathbf{z}_t(\cdot, y)](x) - \frac{\beta}{2}\frac{\mathbf{y}_{0,t}(x) - \mathbf{y}_{0,t}(y)}{x-y}\right) + C_t^N + E_N, \end{aligned}$$

where C_t^N is a constant, and E_N is a remainder that we will prove to be negligible:

$$
\begin{aligned}
E_N := &-\frac{1}{N}\int_{\mathbb{R}} \partial_2 \mathbf{z}_t(x,x)\, dM_N(x) - \frac{1}{N}\Big(1-\frac{\beta}{2}\Big)\int_{\mathbb{R}} \mathbf{y}'_{1,t}(x)\, dM_N(x) \\
&- \frac{1}{N}\Big(1-\frac{\beta}{2}\Big)\iint_{\mathbb{R}^2} \partial_1 \mathbf{z}_t(x,y)\, dM_N(x)\, dM_N(y) \\
&- \frac{\beta}{2N}\iint_{\mathbb{R}^2} \frac{\mathbf{y}_{1,t}(x)-\mathbf{y}_{1,t}(y)}{x-y}\, dM_N(x)\, dM_N(y) \\
&- \frac{\beta}{2N}\iiint_{\mathbb{R}^3} \frac{\mathbf{z}_t(x,y)-\mathbf{z}_t(\tilde{x},y)}{x-\tilde{x}}\, dM_N(x)\, dM_N(\tilde{x})\, dM_N(y)
\end{aligned}
\tag{2.15}
$$

(see [BFG15, Section 3] for more details).

Notice that, since

$$\int_{\mathbb{R}} dM_N(x) = \sum_{i=1}^N \int_{\mathbb{R}} \delta_{\lambda_i}(dx) - N\int_{\mathbb{R}} d\mu_{V_t}(x) = 0,$$

all constants integrate to zero against M_N. Hence, for $\mathcal{R}_t^N$ to be small we want to impose

$$
\begin{cases}
\Xi_t[\mathbf{y}_{0,t}] &= -W + c,\\
\Xi_t[\mathbf{z}_t(\cdot,y)](x) &= -\frac{\beta}{2}\frac{\mathbf{y}_{0,t}(x)-\mathbf{y}_{0,t}(y)}{x-y} + c(y),\\
\Xi_t[\mathbf{y}_{1,t}] &= -\Big(\frac{\beta}{2}-1\Big)\Big[\mathbf{y}'_{0,t} + \int_{\mathbb{R}} \partial_1 \mathbf{z}_t(z,\cdot)\, d\mu_{V_t}(z)\Big] + c',
\end{cases}
\tag{2.16}
$$

where c, c' are some constant to be fixed later, and $c(y)$ does not depend on x.

This shows that, to construct $\mathbf{y}_{0,t}$, $\mathbf{y}_{1,t}$, and $\mathbf{z}_t$, we need to invert the operator Ξ_t. The following lemma is proved in [BFG15, Lemma 3.2]:

Lemma 2.4. *Let $g : \mathbb{R} \to \mathbb{R}$ be a function function of class C^k, and assume that $V_t \in C^r(\mathbb{R})$. Then there exists a unique constant c_g such that the equation*

$$\Xi_t[f](x) = g(x) + c_g$$

has a solution $f : \mathbb{R} \to \mathbb{R}$ of class $C^{(k-2)\wedge(r-3)}$.

Recalling that $V_t(x) = (1-t)x^2 + tV(x)$, $W = V - x^2$, and that Ξ_t is defined on C^1 functions (see (2.14)), as an immediate consequence of Lemma 2.4 we obtain the following corollary (given a function of two variables, $\psi \in C^{s,v}(\mathbb{R}^2)$ means that ψ is s times continuously differentiable with respect to the first variable and v times with respect to the second).

Corollary 2.5. *Let $r \geq 10$. If $V \in C^r(\mathbb{R})$ then we can find $\mathbf{y}_{0,t} \in C^{r-3}(\mathbb{R})$, $\mathbf{z}_t \in C^{s,v}(\mathbb{R}^2)$ for $s + v \leq r - 6$, and $\mathbf{y}_{1,t} \in C^{r-9}(\mathbb{R})$ solving* (2.16).

We now need to estimate the rest E_N defined in (2.15) and prove that it is negligible. Note that

$$\int_{\mathbb{R}} d|M_N|(x) = \sum_{i=1}^N \int_{\mathbb{R}} \delta_{\lambda_i}(dx) + N\int_{\mathbb{R}} d\mu_{V_t}(x) = 2N,$$

that is, M_N is a measure of total mass $2N$. So, a priori, $E_N \approx \frac{1}{N}N^3 \approx N^2 \gg 1$.

Luckily, when considering $\int f\, dM_N$ for some smooth function f, one may expect this integral to be much smaller. Just to make a comparison remember that, given i.i.d. random variables X_i, by the central limit theorem

$$\frac{\sum_i X_i - N\,\mathbb{E}[X_i]}{N^{1/2}}$$

converge to a Gaussian law, hence $\sum_i X_i - N\,\mathbb{E}[X_i]$ is usually of size $N^{1/2}$.

In our case this would not be enough since in E_N we have triple integrals appearing, so from an estimate of the form $\int f\,dM_N \approx N^{1/2}$ we would get $E_N \approx \frac{1}{N}(N^{1/2})^3 = N^{1/2}$, which is not infinitesimal. Hence we need to improve this bound.

Following [BG13b, BG13a, BGK14] (see also [MMS14] and [Shc09]), in [BFG15, FG16] we do this in two steps: we first derive a rough estimate which provides a bound of order $N^{1/2}$, and then in a second step we use the so-called "loop equations" to get a bound of order $\log N$. In this way, we can prove that if $V \in C^{36}(\mathbb{R})$ then there exist positive constant C, c such that

$$|\mathcal{R}_t(\mathbf{Y}_t)| \le C\,\frac{(\log N)^3}{N} \qquad \text{for all } \hat\lambda \in G_t, \tag{2.17}$$

where $G_t \subset \mathbb{R}^N$ is a set satisfying $\mathbb{P}^N_{V_t}(G_t) \ge 1 - N^{-cN}$.

Since $\mathcal{R}_t(\mathbf{Y}_t)$ is trivially bounded by CN^2 everywhere (being the sum of $O(N^2)$ bounded terms, see (2.7)), (2.17) implies that

$$\begin{aligned}\|\mathcal{R}_t(\mathbf{Y}_t)\|_{L^q(Q_t^{N,aV})} &\le \left(\int_{G_t}\left(C\,\frac{(\log N)^3}{N}\right)^q d\mathbb{P}^N_{V_t} + \int_{\mathbb{R}^n\setminus G_t}(C\,N^2)^q\,d\mathbb{P}^N_{V_t}\right)^{1/q}\\ &\le C\left(\left(\frac{(\log N)^3}{N}\right)^q + N^{2q-cN}\right)^{1/q} \le C\,\frac{(\log N)^3}{N},\end{aligned}$$

which proves (2.8).

2.5. **Construction of the flow map.** In the previous section we constructed a vector field $\mathbf{Y}_t$ satisfying (2.8). Thanks to Lemma 2.2 this guarantees that if we solve the ODE

$$\dot X_t = \mathbf{Y}_t(X_t), \qquad X_0 = \mathrm{Id}, \tag{2.18}$$

then $T_N := X_1$ is an approximate transport map. However, as mentioned before, to hope that (2.2) holds true we need T_N to have a very special structure.

Luckily, since $\mathbf{Y}_t$ has a very simple form (see (2.12)) it is natural to expect that we can find a nice expansion for X_t in powers of $1/N$. More precisely, let us define the flow of $\mathbf{y}_{0,t}$:

$$X_{0,t}:\mathbb{R}\to\mathbb{R}, \qquad \dot X_{0,t} = \mathbf{y}_{0,t}(X_{0,t}), \quad X_{0,0} = \mathrm{Id}. \tag{2.19}$$

Then, since at first order $\mathbf{Y}_t(\lambda_1,\ldots,\lambda_N) \approx \big(y_{0,t}(\lambda_1),\ldots,y_{0,t}(\lambda_N)\big)$, we expect that

$$X_t(\lambda_1,\ldots,\lambda_N) \approx \big(X_{0,t}(\lambda_1),\ldots,X_{0,t}(\lambda_N)\big).$$

To find the next order term in the expansion of X_t, we set

$$X_{0,t}^{\otimes N}(\lambda_1,\ldots,\lambda_N) := \big(X_{0,t}(\lambda_1),\ldots,X_{0,t}(\lambda_N)\big)$$

and write

$$X_t = X_{0,t}^{\otimes N} + \frac{1}{N}X_{1,t} + \frac{1}{N^2}X_{2,t}.$$

If we plug this expansion into (2.18), a direct computation using (2.12) shows that $X_{1,t} = (X_{1,t}^1,\ldots,X_{1,t}^N)$ should solve the linear ODE

$$\begin{aligned}\dot X^k_{1,t}(\lambda_1,\ldots,\lambda_N) &= \mathbf{y}'_{0,t}(X_{0,t}(\lambda_k))\cdot X^k_{1,t}(\lambda_1,\ldots,\lambda_N) + \mathbf{y}_{1,t}(X_{0,t}(\lambda_k))\\ &\quad + \int_{\mathbb{R}} \mathbf{z}_t(X_{0,t}(\lambda_k),y)\,dM_N^{X_{0,t}}(y)\\ &\quad + \frac{1}{N}\sum_{j=1}^N \partial_2\mathbf{z}_t\Big(X_{0,t}(\lambda_k),X_{0,t}(\lambda_j)\Big)\cdot X^j_{1,t}(\lambda_1,\ldots,\lambda_N)\end{aligned} \tag{2.20}$$

with the initial condition $X_{1,0} = 0$, where $M_N^{X_{0,t}}$ is defined as

$$\int_{\mathbb{R}} f(y)\, dM_N^{X_{0,t}}(y) := \sum_{i=1}^{N} f(X_{0,t}(\lambda_i)) - N \int_{\mathbb{R}} f\, d\mu_t \qquad \forall\, f \in C_c(\mathbb{R}).$$

With these definitions, the following result holds.

Proposition 2.6. *Assume that $V \in C^r$ for some $r \geq 16$. Then the flow $X_t = (X_t^1, \ldots, X_t^N) : \mathbb{R}^N \to \mathbb{R}^N$ is of class C^{r-9} and the following properties hold.*

Let $X_{0,t}$ and $X_{1,t}$ be as in (2.19) and (2.20) above, and define $X_{2,t} : \mathbb{R}^N \to \mathbb{R}^N$ via the identity

$$X_t = X_{0,t}^{\otimes N} + \frac{1}{N} X_{1,t} + \frac{1}{N^2} X_{2,t}\,.$$

Then

$$\max_{1\leq k\leq N} \|X_{1,t}^k\|_{L^4(\mathbb{P}_V^N)} \leq C\, \log N \qquad \text{and} \qquad \max_{1\leq k\leq N} \|X_{2,t}^k\|_{L^2(\mathbb{P}_V^N)} \leq C\, (\log N)^2. \tag{2.21}$$

In addition, there exist constants $C, c > 0$ such that, with probability greater than $1 - C\, e^{-c(\log N)^2}$,

$$\sup_{t\in[0,1]} \max_{1\leq k\leq N} |X_{1,t}^k| \leq C\, \log N\, N^{1/(r-14)}, \qquad \sup_{t\in[0,1]} \max_{1\leq k\leq N} |X_{2,t}^k| \leq C\, (\log N)^2\, N^{2/(r-15)}, \tag{2.22}$$

$$\sup_{t\in[0,1]} \max_{1\leq k,k'\leq N} \big|X_{1,t}^k(\hat\lambda) - X_{1,t}^{k'}(\hat\lambda)\big| \leq C\, \log N\, N^{1/(r-15)} |\lambda_k - \lambda_{k'}|, \tag{2.23}$$

and

$$\sup_{t\in[0,1]} \max_{1\leq k,k'\leq N} \big|\partial_{\lambda_{k'}} X_{1,t}^k\big|(\hat\lambda) \leq C\, \log N\, N^{1/(r-15)}. \tag{2.24}$$

The proof of this result is rather involved and we refer to the proof of [FG16, Proposition 4.13] for more detail. Here, we just explain the reason for the presence of terms of the form $N^{1/(r-14)}$ and $N^{1/(r-15)}$ in the estimates above.

Remember that, as explained before, to prove (2.17) we needed to show that, with very large probability, if f is smooth enough then

$$\left| \int f\, dM_N \right| \leq C\, \log N.$$

From this fact we can deduce that, for any $t \in [0,1]$, there exists a set $H_t \subset \mathbb{R}^N$ such that $\mathbb{P}_{V_0}^N(H_t) \geq 1 - N^{-cN}$ and the term

$$\int \mathbf{z}_t(X_{0,t}(\lambda_k), y)\, dM_N^{X_{0,t}}(y)$$

appearing in (2.20) is of size $O(\log N)$ whenever $\hat\lambda \in H_t$. Unfortunately this "good set" H_t depends on t, and to construct our approximate transport map we need to integrate $\int \mathbf{z}_t(X_{0,t}(\lambda_k), y)\, dM_N^{X_{0,t}}(y)$ with respect to $t \in [0,1]$. So, we need to find a "universal good set" H such that, fixed $\hat\lambda \in H$, the term $\int \mathbf{z}_t(X_{0,t}(\lambda_k), y)\, dM_N^{X_{0,t}}(y)$ can be controlled *for all* $t \in [0,1]$.

As shown in [FG16, Lemma 4.12], this is indeed possible, and the smoother is the test function against which we integrate M_N the better the bound we get. Recalling that the regularity of $\mathbf{z}_t$ and $X_{0,t}$ depends on the regularity of V, since $V \in C^r$ it follows by [FG16, Proposition 4.13] that there exists a set H such that $\mathbb{P}_{V_0}^N(H) \geq 1 - N^{-c(\log N)^2}$ and

$$\sup_{t\in[0,1]} \left| \int \mathbf{z}_t(X_{0,t}(\lambda_k), y)\, dM_N^{X_{0,t}}(y) \right| \leq C\, \log N\, N^{1/(r-14)} \qquad \forall\, \hat\lambda \in H,\ \forall\, k = 1, \ldots, N. \tag{2.25}$$

In particular, setting

$$A_{1,t} := \max_{1\le k\le N} |X^k_{1,t}|,$$

(2.20) and (2.25) imply that

$$\frac{d}{dt} A_{1,t} \le C\, A_{1,t} + C\, \log N\, N^{1/(r-14)},$$

so the first bound in (2.22) follows by integration, noticing that $A_{1,0} = 0$.

2.6. **On the leading order term in the transport map.** The goal of this section is to show that the leading order term $X_{0,1}$ of our approximate transport map (see Proposition 2.6) can be defined in terms of the stationary measures $\mu_{\rm sc}$ and μ_V. More precisely, we claim that $X_{0,1} : \mathbb{R} \to \mathbb{R}$ is the monotone rearrangement of $\mu_{\rm sc}$ onto μ_V, that is, $X_{0,1}$ coincides with the unique monotone increasing map $T_0 : \mathbb{R} \to \mathbb{R}$ satisfying $(T_0)_\#\mu_{\rm sc} = \mu_V$.

To prove this fact, we first show that $X_{0,1}$ is a monotone increasing function. Note that, since $X_{0,t} : \mathbb{R} \to \mathbb{R}$ is obtained as the flow of the Lipschitz function $\mathbf{y}_{0,t}$, differentiating (2.19) with respect to $x \in \mathbb{R}$ we get

$$\dot X'_{0,t} = \mathbf{y}'_{0,t}(X_{0,t})\, X'_{0,t}, \qquad X'_{0,0} = 1,$$

thus

$$\frac{d}{dt}\big|X'_{0,t}\big| \le \big|\mathbf{y}'_{0,t}(X_{0,t})\big|\,\big|X'_{0,t}\big| \le L\,\big|X'_{0,t}\big|, \qquad X'_{0,0} = 1,$$

and Gronwall's inequality gives the bound

$$e^{-Lt} \le \big|X'_{0,t}\big| \le e^{Lt}.$$

Since $X'_{0,0} = 1$, it follows by continuity that $X'_{0,t}$ must remain positive for all time and it satisfies

$$e^{-Lt} \le X'_{0,t} \le e^{Lt}, \tag{2.26}$$

from which we deduce that

$$e^{-Lt}(x-y) \le X_{0,t}(x) - X_{0,t}(y) \le e^{Lt}(x-y) \qquad \forall\, y < x, \quad t \in [0,1].$$

In particular,

$$e^{-L}(x-y) \le X_{0,1}(x) - X_{0,1}(y) \le e^{L}(x-y) \qquad \forall\, y < x, \tag{2.27}$$

which proves that $X_{0,1}$ is monotone increasing.

We now claim that $X_{0,1}$ transports $\mu_{\rm sc}$ onto μ_V. For this, we fix a smooth compactly supported function $\varphi : \mathbb{R} \to \mathbb{R}^+$ and we apply Lemma 2.2 with

$$\chi(\hat\lambda) := \langle \varphi, L_N\rangle = \frac{1}{N}\sum_{i=1}^N \varphi(\lambda_i).$$

Then

$$\chi \circ T_N(\hat\lambda) = \frac{1}{N}\sum_{i=1}^N \varphi\big(T^i_N(\hat\lambda)\big)$$

and

$$\left|\log\biggl(1 + \int_{\mathbb{R}^N} \langle \varphi, L_N\rangle\, d\mathbb{P}^N_V\biggr) - \log\biggl(1 + \int_{\mathbb{R}^N} \frac{1}{N}\sum_{i=1}^N \varphi\big(T^i_N(\hat\lambda)\big)\, d\mathbb{P}^N_G\biggr)\right| \le C_\eta\, N^{\eta-1}.$$

Noticing that

$$\frac{1}{N}\sum_{i=1}^N \varphi\big(T_N^i(\hat\lambda)\big) = \frac{1}{N}\sum_{i=1}^N \varphi\big(X_{0,1}(\lambda_i)\big)\big) + O\bigg(\|\varphi'\|_\infty \frac{1}{N}\sum_{i=1}^N\Big(\frac{|X_{1,1}^i|}{N}+\frac{|X_{2,1}^i|}{N^2}\Big)\bigg)$$
$$= \langle \varphi\circ X_{0,1}, L_N\rangle + O\bigg(\|\varphi'\|_\infty \frac{1}{N}\sum_{i=1}^N\Big(\frac{|X_{1,1}^i|}{N}+\frac{|X_{2,1}^i|}{N^2}\Big)\bigg)$$

(see Proposition 2.6), it follows from (2.21) that

$$\int_{\mathbb{R}^N}\frac{1}{N}\sum_{i=1}^N \varphi\big(T_N^i(\hat\lambda)\big)\,d\mathbb{P}_G^N = \int_{\mathbb{R}^N}\langle \varphi\circ X_{0,1}, L_N\rangle\,d\mathbb{P}_G^N + O\bigg(\|\varphi'\|_\infty \frac{\log N}{N}+\frac{(\log N)^2}{N^2}\bigg).$$

Therefore

$$\bigg|\log\bigg(1+\int_{\mathbb{R}^N}\langle\varphi, L_N\rangle\,d\mathbb{P}_V^N\bigg) - \log\bigg(1+\int_{\mathbb{R}^N}\langle\varphi\circ X_{0,1}, L_N\rangle\,d\mathbb{P}_G^N\bigg)\bigg| \le C_\eta\, N^{\eta-1}.$$

Recalling that $L_N \rightharpoonup^* \mu_V$ under $\mathbb{P}_V^N$ (resp. $L_N \rightharpoonup^* \mu_{\rm sc}$ under $\mathbb{P}_G^N$) (see (1.1) and the beginning of Section 2), letting $N\to\infty$ in the formula above we obtain

$$\log\bigg(1+\int_{\mathbb{R}}\varphi\,d\mu_V\bigg) = \log\bigg(1+\int_{\mathbb{R}}\varphi\circ X_{0,1}\,d\mu_{\rm sc}\bigg),$$

that is,

$$\int_{\mathbb{R}}\varphi\,d\mu_V = \int_{\mathbb{R}}\varphi\circ X_{0,1}\,d\mu_{\rm sc}.$$

By the arbitrariness of φ, this proves that $(X_{0,1})_\#\mu_{\rm sc} = \mu_V$. Since $X_{0,1}$ is monotone increasing (see (2.27)), this shows that $X_{0,1}$ is the monotone rearrangement of $\mu_{\rm sc}$ onto μ_V, as desired.

2.7. **Universality in the bulk.** Let $T_N := X_1$, where X_t is the flow of $\mathbf{Y}_t$. The results in the previous sections show that T_N is an approximate transport map having a very special structure. As we shall explain now, this allows us to show that (2.1) and (2.2) hold.

We begin by noticing that (2.1) and (2.2) involve the laws of the ordered eigenvalues $\mathbf{P}_G^N$ and $\mathbf{P}_V^N$, while our previous construction of the transport map was done with $\mathbb{P}_G^N$ and $\mathbb{P}_V^N$. To fix this, we observe that $\mathbf{P}_G^N$ and $\mathbf{P}_V^N$ can be defined as

$$\mathbf{P}_G^N := \mathcal{R}_\#\mathbb{P}_G^N \qquad \text{and} \qquad \mathbf{P}_V^N := \mathcal{R}_\#\mathbb{P}_V^N,$$

where $\mathcal{R}:\mathbb{R}^N\to\mathbb{R}^N$ is given by

$$[\mathcal{R}(x_1,\ldots,x_N)]_i := \min_{\#J=i}\max_{j\in J} x_j \qquad \forall\, i=1,\ldots,N. \tag{2.28}$$

Note that $\mathcal{R}$ is 1-Lipschitz for the sup norm. In particular, if $\chi:\mathbb{R}^N\to\mathbb{R}$ is a L-Lipschitz function depending only on m variables, then $\chi\circ\mathcal{R}$ is $\sqrt{m}L$-Lipschitz.

Now, let $f:\mathbb{R}\to\mathbb{R}^+$ be a nonnegative Lipschitz function with compact support (say, $\mathrm{supp}(f)\subset[-M,M]$) and set $\chi(\hat\lambda) := f(N(\lambda_{i+1}-\lambda_i))$. Then

$$\int_{\mathbb{R}^N}\chi\,d\mathbf{P}_V^N = \int_{\mathbb{R}^N}\chi\circ\mathcal{R}\,d\mathbb{P}_V^N. \tag{2.29}$$

We now apply Lemma 2.2 with $\frac{1}{\|f\|_\infty}\chi\circ\mathcal{R}$ in place of χ to deduce that, for any $\eta>0$,

$$\bigg|\log\bigg(1+\frac{1}{\|f\|_\infty}\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\,d\mathbb{P}_V^N\bigg) - \log\bigg(1+\frac{1}{\|f\|_\infty}\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_N\,d\mathbb{P}_G^N\bigg)\bigg| \le C_\eta\, N^{\eta-1}.$$

In particular, choosing $\eta<1$ (so that the right hand side is infinitesimal), since $\frac{1}{\|f\|_\infty}\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\,d\mathbb{P}_V^N$ and $\frac{1}{\|f\|_\infty}\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_N\,d\mathbb{P}_G^N$ are both bounded by $\frac{\|\chi\|_\infty}{\|f\|_\infty}=1$, it follows by Remark 2.3 that

$$\frac{1}{\|f\|_\infty}\left|\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\,d\mathbb{P}_V^N-\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_N\,d\mathbb{P}_G^N\right|\leq C_\eta\,N^{\eta-1}. \tag{2.30}$$

Let us define

$$T_{N,1}:=X_{0,1}^{\otimes N}+\frac{1}{N}X_{1,1} \tag{2.31}$$

where $X_{0,t}$ and $X_{1,t}$ are as in Proposition 2.6. Then, since $T_N-T_{N,1}=\frac{1}{N^2}X_{2,1}$,

$$\left|\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_N\,d\mathbb{P}_G^N-\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_{N,1}\,d\mathbb{P}_G^N\right|\leq\|\nabla(\chi\circ\mathcal{R})\|_\infty\frac{1}{N^2}\int_{\mathbb{R}^N}|X_{2,1}|\,d\mathbb{P}_G^N,$$

where

$$|X_{2,1}|:=\left(\sum_{k=1}^N|X_{2,1}^k|^2\right)^{1/2}.$$

Noticing that $\|\nabla(\chi\circ\mathcal{R})\|_\infty\leq\sqrt{2}\,N\,\|f'\|_\infty$ and

$$\int_{\mathbb{R}^N}|X_{2,1}|\,d\mathbb{P}_G^N\leq\left(\sum_{k=1}^N\int_{\mathbb{R}^N}|X_{2,1}^k|^2d\mathbb{P}_G^N\right)^{1/2}\leq C\,(\log N)^2\,N^{1/2}$$

(see (2.21)), we deduce that

$$\left|\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_N\,d\mathbb{P}_G^N-\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_{N,1}\,d\mathbb{P}_G^N\right|\leq C\,(\log N)^2\,N^{-1/2}\,\|f'\|_\infty. \tag{2.32}$$

We now claim that $\hat X_1^N$ preserves the order of the eigenvalues with large probability. Indeed, since

$$\left|\frac{1}{N}X_{1,1}^k(\hat\lambda)-\frac{1}{N}X_{1,1}^{k'}(\hat\lambda)\right|\leq C\,\frac{\log N\,N^{1/(r-15)}}{N}\,|\lambda_k-\lambda_{k'}|\ll|\lambda_k-\lambda_{k'}|$$

with probability greater than $1-C\,e^{-c(\log N)^2}$ (see (2.23)), it follows by (2.31) and (2.27) that

$$e^{-2L}\big(\lambda_k-\lambda_{k'}\big)\leq T_{N,1}^k(\hat\lambda)-T_{N,1}^{k'}(\hat\lambda)\leq e^{2L}\big(\lambda_k-\lambda_{k'}\big)\qquad\forall\,\lambda_{k'}<\lambda_k$$

with probability greater than $1-C\,e^{-c(\log N)^2}$, as desired.

Thanks to this fact we obtain that $\mathcal{R}\circ T_{N,1}=T_{N,1}\circ\mathcal{R}$ outside a set of probability bounded by $C\,e^{-c(\log N)^2}$, therefore

$$\left|\int_{\mathbb{R}^N}\chi\circ\mathcal{R}\circ T_{N,1}\,d\mathbb{P}_G^N-\int_{\mathbb{R}^N}\chi\circ T_{N,1}\circ\mathcal{R}\,d\mathbb{P}_G^N\right|\leq C\,e^{-c(\log N)^2}\,\|f\|_\infty. \tag{2.33}$$

Thus, recalling that $\chi(\hat\lambda):=f(N(\lambda_{i+1}-\lambda_i))$, combining (2.29), (2.30), (2.32), and (2.33) we get

$$\begin{aligned}\left|\int_{\mathbb{R}^N}f\big(N(\lambda_{i+1}-\lambda_i)\big)\,d\mathbf{P}_V^N-\int_{\mathbb{R}^N}f\Big(N\big(T_{N,1}^{i+1}(\hat\lambda)-T_{N,1}^i(\hat\lambda)\big)\Big)\,d\mathbf{P}_G^N\right|&\\ \leq C\,N^{\eta-1}\,\|f\|_\infty+C\,(\log N)^2\,N^{-1/2}\,\|f'\|_\infty.&\end{aligned} \tag{2.34}$$

Observe now that, since $X_{0,1}$ is of class C^2 and $X_{0,1}'(\lambda_i)\geq e^{-L}$ (see (2.27)),

$$X_{0,1}(\lambda_{i+1})-X_{0,1}(\lambda_i)=X_{0,1}'(\lambda_i)\,(\lambda_{i+1}-\lambda_i)+O\big(\big|X_{0,1}'(\lambda_i)\,(\lambda_{i+k}-\lambda_i)\big|^2\big).$$

Also, using again that $X_{0,1}'(\lambda_i)\geq e^{-L}$, it follows by (2.23) that, outside a set of probability bounded by $C\,e^{-c(\log N)^2}$,

$$\big|X_{1,1}^{i+1}(\hat\lambda)-X_{1,1}^i(\hat\lambda)\big|\leq C\,\log N\,N^{1/(r-15)}\,\big|X_{0,1}'(\lambda_i)\,(\lambda_{i+1}-\lambda_i)\big|.$$

Thus, recalling (2.31), we get
(2.35)
$$T^{i+1}_{N,1}(\hat\lambda) - T^{i}_{N,1}(\hat\lambda) = X'_{0,1}(\lambda_i)\,(\lambda_{i+1}-\lambda_i)\Big[1 + O\big(\log N\, N^{1/(r-15)-1}\big) + O\big(\big|X'_{0,1}(\lambda_i)\,(\lambda_{i+1}-\lambda_i)\big|\big)\Big]$$

with probability greater than $1 - C\,e^{-c(\log N)^2}$.

Since we assume f supported in $[-M,M]$, when computing
$$\int_{\mathbb{R}^N} f\Big(N\big(T^{i+1}_{N,1}(\hat\lambda) - T^{i}_{N,1}(\hat\lambda)\big)\Big)\,d\mathbf{P}^N_G$$
we can restrict the domain of integration to those $\hat\lambda$ such that $\big|NX'_{0,t}(\lambda_i)\,(\lambda_{i+1}-\lambda_i)\big|$ is bounded by M. Hence, using (2.35) and a Taylor expansion, we note that for such $\hat\lambda$ we have
$$f\Big(N\big(T^{i+1}_{N,1}(\hat\lambda) - T^{i}_{N,1}(\hat\lambda)\big)\Big) = f\big(X'_{0,1}(\lambda_i)\,N(\lambda_{i+1}-\lambda_i)\big) \\ + O\bigg(M\log N\, N^{1/(r-15)-1} + M^2\,N^{-1}\bigg)\,\|f'\|_\infty.$$

Combining this bound with (2.34), recalling Section 2.6 and the fact that (2.17) is valid under the assumption that $V\in C^{36}$, we proved the following universality result:

Theorem 2.7 (Universality in the bulk)**.** *Assume that $V\in C^{36}$ is a uniformly convex function. Denote by $\mathbf{P}^N_V$ (resp. $\mathbf{P}^N_G$) the distribution of the increasingly ordered eigenvalues λ_i under $\mathbb{P}^N_V$ (resp. $\mathbb{P}^N_G$). Also, let T_0 denote the monotone rearrangement from μ_{sc} to μ_V. Fix $\varepsilon,\eta>0$. There exists a constant $\hat C>0$, independent of N, such that the following holds:*

Let $M\in(0,\infty)$. Then, for any nonnegative Lipschitz function $f:\mathbb{R}\to\mathbb{R}^+$ supported inside $[-M,M]$ and for any $i\in[\varepsilon N,(1-\varepsilon)N]$,
$$\bigg|\int_{\mathbb{R}^N} f\big(N(\lambda_{i+1}-\lambda_i)\big)\,d\mathbf{P}^N_V - \int_{\mathbb{R}^N} f\big(T'_0(\lambda_i)\,N(\lambda_{i+k}-\lambda_i)\big)\,d\mathbf{P}^N_G\bigg| \\ \le \hat C\,N^{\eta-1}\,\|f\|_\infty + \hat C\,\Big((\log N)^2\,N^{-1/2} + M^2\,N^{-1}\Big)\,\|f'\|_\infty.$$

2.8. **Universality at the edge.** In the previous section we showed how our approximate transport maps allow us to prove universality of the law of $N(\lambda_{i+1}-\lambda_i)$ when $i\in[\varepsilon N,(1-\varepsilon)N]$. We note that the same strategy also proves universality at the edge.

More precisely, recalling that $\mathrm{supp}(\mu_{sc})=[-2,2]$ and $\mathrm{supp}(\mu_V)=[a_V,b_V]$, since $X_{0,1}=T_0$ is the monotone rearrangement of μ_{sc} onto μ_V (see Section 2.6), we deduce that $X_{0,1}(-2)=a_V$. Then, arguing as in the previous section, we see that
$$\bigg|\int_{\mathbb{R}^N} f\big(N^{2/3}(\lambda_1-a_V)\big)\,d\mathbf{P}^N_V - \int_{\mathbb{R}^N} f\Big(N^{2/3}\big(T_{N,1}(\hat\lambda) - X_{0,1}(-2)\big)\Big)\,d\mathbf{P}^N_G\bigg| \\ \le C_\eta\,N^{\eta-1}\,\|f\|_\infty + C\,\frac{(\log N)^2}{N^{5/6}}\,\|f'\|_\infty.$$

Since, by (2.21),
$$T^i_{N,1}(\lambda) = X_{0,1}(\lambda_i) + O_{L^4(\mathbb{P}^N_G)}\bigg(\frac{\log N}{N}\bigg) \\ = X_{0,1}(-2) + X'_{0,1}(-2)\,(\lambda_i+2) + O\big(|\lambda_i+2|^2\big) + O_{L^4(\mathbb{P}^N_G)}\bigg(\frac{\log N}{N}\bigg),$$

one can conclude that the following holds:

Theorem 2.8 (Universality at the edge). *Under the same assumptions of Theorem 2.7, given $\eta > 0$, there exists a constant $\hat{C} > 0$, independent of N, such that the following holds:*

Let $M \in (0,\infty)$. Then, for any nonnegative Lipschitz function $f : \mathbb{R}^m \to \mathbb{R}^+$ supported inside $[-M, M]$, we have

$$\left| \int_{\mathbb{R}^N} f\big(N^{2/3}(\lambda_1 - a_V)\big)\, d\mathbf{P}_V^N - \int_{\mathbb{R}^N} f\big(T_0'(-2)\, N^{2/3}(\lambda_1 + 2)\big)\, d\mathbf{P}_G^N \right|$$
$$\leq \hat{C}\, N^{\eta-1}\, \|f\|_\infty + \hat{C}\left(\log N\, N^{-1/3} + M^2\, N^{-4/3}\right) \|f'\|_\infty.$$

It is worth noticing that the very same argument as the one used above allows one to show universality when replacing the test functions

$$f\big(N(\lambda_{i+1} - \lambda_i)\big) \qquad \text{and} \qquad f\big(N^{2/3}(\lambda_1 - a_V)\big)$$

by

$$f\big(N(\lambda_{i+1} - \lambda_i), \dots, N(\lambda_{i+m} - \lambda_i)\big) \qquad \text{and} \qquad f\big(N^{2/3}(\lambda_1 - a_V, \dots, N^{2/3}(\lambda_m - a_V)\big),$$

where $f : \mathbb{R}^m \to \mathbb{R}^+$ is a Lipschitz and compactly supported function, and m can even be allowed to increase with N (see [BFG15, FG16] for more details).

2.9. **Generalization and extensions.** Theorem 2.7 concerns the universality of fluctuations for the difference of consecutive eigenvalues. Another natural and important question concerns the universality of fluctuations around some fixed "energy" value E in the bulk. This corresponds to consider as test functions m-points correlation functions of the form

$$\sum_{i_1 \neq \dots \neq i_m} f\big(N(\lambda_{i_1} - E), \dots, N(\lambda_{i_m} - E)\big),$$

where E belongs to the bulk of the spectrum. Note that, since we are considering a sum over a set of indices of cardinality of order N^m, these test functions have L^∞ norm of size N^m. Hence, to attack this problem, it is crucial that Lemma 2.2 applies to this class of functions.

By exploiting the estimates on the approximate transport maps stated in Proposition 2.6, as shown in [FG16] we can prove universality for test functions obtained as an average with respect E over a very small interval. Here and in the following, we use $⨍_I$ to denote the averaged integral over an interval $I \subset \mathbb{R}$, namely $⨍_I = \frac{1}{|I|}\int_I$.

Corollary 2.9. *Under the same assumptions of Theorem 2.7, fix $m \in \mathbb{N}$ and $\zeta \in (0,1)$. Then, given $E \in (-2,2)$, $\theta \in (0, \min\{\zeta, 1-\zeta\})$, and $f : \mathbb{R}^m \to \mathbb{R}^+$ a nonnegative Lipschitz function with compact support, there exists a constant $\hat{C} > 0$, independent of N, such that the following holds true:*

$$\bigg| \int \bigg[⨍_{T_0(E) - N^{-\zeta}\, T_0'(E)}^{T_0(E)+N^{-\zeta}\, T_0'(E_1)} \sum_{i_1 \neq \dots \neq i_m} f\big(N(\lambda_{i_1} - \tilde{E}), \dots, N(\lambda_{i_m} - \tilde{E})\big)\, d\tilde{E} \bigg]\, d\mathbb{P}_V^N$$
$$- \int \bigg[⨍_{E-N^{-\zeta}}^{E+N^{-\zeta}} \sum_{i_1 \neq \dots \neq i_m} f\big(T_0'(E)\ N(\lambda_{i_1} - \tilde{E}), \dots, T_0'(E)\, N(\lambda_{i_m} - \tilde{E})\big)\, d\tilde{E} \bigg]\, d\mathbb{P}_G^N \bigg|$$
$$\leq \hat{C}\left(N^{\theta+\zeta-1} + N^{\theta-\zeta}\right).$$

All these universality results are contained in [FG16]. As mentioned at the beginning of Section 2, universality for β-models has also been obtained by other techniques. However, the approach described above has the advantage of being extremely robust. In particular, as shown in [FG16] it can be used also in multi-matrix models. For instance, a corollary of the results in [FG16] is the following theorem about the universality of fluctuation of matrices obtained as the image of independent Guassian matrices via a polynomial close to the identity.

We recall that a polynomial on $\mathbb{R}^{N\times N}$ with complex coefficients is said to be self-adjoint if it preserves the space of Hermitian matrices.

Theorem 2.10. *Let $P_1,\dots,P_d:\mathbb{R}^{N\times N}\to\mathbb{R}^{N\times N}$ be self-adjoint polynomials. There exists $\epsilon_0>0$ such that the following holds: Let $\{X_\ell\}_{1\le\ell\le d}$ be independent Gaussian matrices and set*

$$Y_\ell := X_\ell + \epsilon\, P_\ell(X_1,\dots,X_d) \qquad \forall\, \ell = 1,\dots,d.$$

Then, for $\epsilon\in[-\epsilon_0,\epsilon_0]$, the eigenvalues of the matrices $\{Y_\ell\}_{1\le\ell\le d}$ fluctuate both in the bulk and at the edge as when $\epsilon=0$, up to rescaling. In other words, the fluctuations of $\{Y_\ell\}_{1\le\ell\le d}$ follow the sine-kernel law inside the bulk and the Tracy-Widom law at the edge.

Acknowledgments. AF has been partially supported by NSF Grant DMS-1262411 and NSF Grant DMS-1361122.

References

[AG97] G. Ben Arous and A. Guionnet, *Large deviations for Wigner's law and Voiculescu's non-commutative entropy*, Probab. Theory Related Fields **108** (1997), no. 4, 517–542.

[AGZ10] G. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge University Press, 2010.

[Bek15] F. Bekerman, *Transport maps for β-matrix models in the multi-cut matrix regime*, arXiv:1512.00302 (2015).

[BEY14a] P. Bourgade, L. Erdös, and H.-T. Yau, *Edge universality of beta ensembles*, Comm. Math. Phys. **332** (2014), no. 1, 261–353. MR 3253704

[BEY14b] ———, *Universality of general β-ensembles*, Duke Math. J. **163** (2014), no. 6, 1127–1190. MR 3192527

[BFG15] F. Bekerman, A. Figalli, and A. Guionnet, *Transport maps for β-matrix models and universality*, Comm. Math. Phys. **338** (2015), no. 2, 589–619. MR 3351052

[BG13a] G. Borot and A. Guionnet, *Asymptotic Expansion of β Matrix Models in the multi-cut Regime*, arXiv:1303.1045 (2013).

[BG13b] G. Borot and A. Guionnet, *Asymptotic Expansion of β Matrix Models in the one-cut Regime*, Comm. Math. Phys. **317** (2013), no. 2, 447–483.

[BGK14] G. Borot, A. Guionnet, and K. Kozlowski, *Large-n asymptotic expansion for mean field models with coulomb gas interaction*, arXiv:1312.6664 (2014).

[DG07a] P.A. Deift and D. Gioev, *"Universality at the edge of the spectrum for unitary, orthogonal and symplectic ensembles of random matrices."*, Comm. Pure Appl. Math. **60** (2007), 867–910.

[DG07b] ———, *"Universality in Random Matrix Theory for orthogonal and symplectic ensembles."*, Int. Math. Res. Pap. (2007), ArtID rpm007, 116pp.

[DG09] P. Deift and D. Gioev, *Random matrix theory: invariant ensembles and universality*, Courant Lecture Notes in Mathematics, vol. 18, Courant Institute of Mathematical Sciences, New York, 2009.

[DKM+99] P. Deift, T. Kriecherbauer, K.T.-R. McLaughlin, S. Venakides, and X. Zhou, *Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory*, Comm. Pure Appl. Math. **52** (1999), no. 11, 1335–1425.

[EPR+10] L. Erdös, S. Péché, J. A. Ramírez, B. Schlein, and H.-T. Yau, *Bulk universality for Wigner matrices*, Comm. Pure Appl. Math. **63** (2010), no. 7, 895–925.

[Erd10] L. Erdös, *Universality of Wigner random matrices*, XVIth International Congress on Mathematical Physics, World Sci. Publ., Hackensack, NJ, 2010, pp. 86–105.

[ESYY12] L. Erdös, B. Schlein, H.-T. Yau, and J. Yin, *The local relaxation flow approach to universality of the local statistics for random matrices*, Annales Inst. H. Poincaré(B) **48** (2012), 1–46.

[EY12a] L. Erdös and H.-T. Yau, *Gap universality of generalized wigner and beta-ensembles*, arXiv: 1211.3786 (2012).

[EY12b] L. Erdös and H.-T. Yau, *Universality of local spectral statistics of random matrices*, Bull. Amer. Math. Soc. (N.S.) **49** (2012), no. 3, 377–414.

[EYY12] L. Erdös, H.-T. Yau, and J. Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. **229** (2012), no. 3, 1435–1515.

[FG16] A. Figalli and A. Guionnet, *Universality in several-matrix models via approximate transport maps*, arxiv:1407.2759 (2016).

[KRV13] M. Krishnapur, B. Rider, and B. Virag, *Universality of the stochastic airy operator*, arXiv:1306.4832 (2013).

[KS10] T. Kriecherbauer and M. Shcherbina, *Fluctuations of eigenvalues of matrix models and their applications*, arXiv:1003.6121 (2010).

[LL08] E. Levin and D. S. Lubinsky, *Universality limits in the bulk for varying measures*, Adv. Math. **219** (2008), no. 3, 743–779. MR 2442052 (2010a:60009)

[Meh04] M. L. Mehta, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004. MR 2129906 (2006b:82001)

[MMS14] M. Maïda and E. Maurel-Segala, *Free transport-entropy inequalities for non-convex potentials and application to concentration for random matrices*, Probab. Theory Related Fields **159** (2014), no. 1-2, 329–356.

[PS97] L. Pastur and M. Shcherbina, *Universality of the local eigenvalue statistics for a class of unitary invariant matrix ensembles*, J. Stat. Phys **86** (1997), 109–147.

[RRV11] J. Ramirez, B. Rider, and B. Virag, *Beta ensembles, stochastic airy spectrum, and a diffusion*, J. Amer. Math. Soc. **24** (2011), 919–944.

[Shc09] M. Shcherbina, *On universality for orthogona ensembles of random matrices and their applications*, Commun. Math. Phys. **285** (2009), 957–974.

[Shc12] ———, *Fluctuations of linear eigenvalue statistics of β matrix models in the multi-cut regime*, arxiv.org/abs/1205.7062, (2012).

[Shc13] ———, *Change of variables as a method to study general β-models: bulk universality*, arXiv:1310.7835 (2013).

[Tao13] T. Tao, *The asymptotic distribution of a single eigenvalue gap of a Wigner matrix*, Probab. Theory Related Fields **157** (2013), no. 1-2, 81–106.

[TV10] T. Tao and V. Vu, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. **38** (2010), no. 5, 2023–2065, with an appendix by Manjunath Krishnapur.

[TV11] ———, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1, 127–204.

[TV12] ———, *Random covariance matrices: universality of local statistics of eigenvalues*, Ann. Probab. **40** (2012), no. 3, 1285–1315.

[TW94a] C. A. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys. **159** (1994), no. 1, 151–174.

[TW94b] ———, *Level spacing distributions and the Bessel kernel*, Comm. Math. Phys. **161** (1994), 289–309.

[VV09] B. Valkò and B. Viràg, *Continuum limits of random matrices and the brownian carousel.*, Invent. Math. **177** (2009), 463–508.

[Wig55] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564.