Sharp total variation rates of convergence for fluctuations of linear statistics of β-ensembles

Jürgen Angst, Ronan Herry, Dominique Malicet, Guillaume Poly

SHARP TOTAL VARIATION RATES OF CONVERGENCE FOR FLUCTUATIONS OF LINEAR STATISTICS OF β-ENSEMBLES

JÜRGEN ANGST, RONAN HERRY, DOMINIQUE MALICET, AND GUILLAUME POLY

Abstract. In this article, we revisit the question of fluctuations of linear statistics of beta ensembles in the single cut and non-critical regime for general potentials V under mild regularity and growth assumptions. Our main objective is to establish sharp quantitative Central Limit Theorems (CLT) for strong distances, such as the *total variation* distance, which to the best of our knowledge, is new for general potentials, even qualitatively. Namely, setting μ_V the equilibrium measure, for a test function $\xi \in \mathscr{C}^{14}$, we establish the convergence in total variation of $X_n = \sum_{i=1}^n \xi(\lambda_i) - n\langle \xi, \mu_V\rangle$ to an explicit Gaussian variable at the sharp speed $1/n$. Under the same assumptions, we also establish multivariate CLTs for vectors of linear statistics in $p-$Wasserstein distances for any $p \geq 1$, with the optimal rate $1/n$, a result which already in dimension one sharpens the speed of convergence established in the recent contribution [26] as well as the required regularity on the test functions. A second objective of this paper, in a more qualitative direction, is to establish the so-called *super-convergence* of linear statistics, that is to say the convergence of all derivatives of the densities of X_n uniformly on $\mathbb{R}$, provided that $\xi \in \mathscr{C}^\infty(\mathbb{R})$ and is not too degenerated in some sense.

As far as quantitative results are concerned, our approach shares similar tools with the latter reference as we also rely on Stein's method and integration by part formalism provided by the *Dyson generator* L. Nevertheless, we adopt a different and innovative strategy which consists in proving that, up to centering, the above linear statistics X_n can be written in the form $\mathsf{L}F_n + Z_n$ where Z_n is some small remainder and $\Gamma[X_n, F_n]$ converges to some constant as direct consequences of the law of large numbers. We emphasize that this approach bypasses the costly requirement of the invertibility of the Markov operator which is central in *Malliavin-Stein's* method and seems robust enough to be implemented for various models of Gibbs measures. On the qualitative side, the proof of the super-convergence of densities, for its part, is a consequence of the existence of negative moments for the carré du champ of linear statistics, associated with integration by parts techniques.

Contents

1. Introduction and statement of the results

1.1. **Overview of our contributions.** Given a parameter $\beta > 0$ interpreted as an inverse temperature, the *β-ensemble*, with n particles associated with a continuous potential $V\colon \mathbb{R} \to \mathbb{R}$ such that $\liminf_{x\to\pm\infty} V(x) - \log|x| > 0$, is the Gibbs probability distribution on $\mathbb{R}^n$ associated with the *energy*

$$H_n(\lambda) := \sum_{i<j} \log \frac{1}{|\lambda_i - \lambda_j|} + n\sum_i V(\lambda_i).$$

Namely, the β-ensemble is the probability measure

$$\mathbf{P}(\mathtt{d}\lambda) = \mathbf{P}_{n,\beta}(\mathtt{d}\lambda) := \frac{1}{Z_{n,\beta}} \mathrm{e}^{-\beta H_n(\lambda)} \mathtt{d}\lambda, \tag{1.1}$$

where $Z_{n,\beta} := \int \mathrm{e}^{-\beta H_n} \mathtt{d}\lambda$ is the *partition function*, and $\mathtt{d}\lambda$ stands for the Lebesgue measure on $\mathbb{R}^n$. In this paper, we establish both qualitative and quantitative Central Limit Theorems (CLT) for *linear statistics* of β-ensemble, which are random variables of the form

$$X_n := \sum_{i=1}^{n} \xi(\lambda_i) - n \int \xi \mathtt{d}\mu_V, \qquad n \in \mathbb{N},$$

for some non-constant test function $\xi \colon \mathbb{R} \to \mathbb{R}$ satisfying some mild regularity assumptions, where μ_V is the so-called *equilibrium measure*, and where the random vector $(\lambda_1, \dots, \lambda_n)$ is drawn from the Gibbs measure $\mathbf{P}$.

Our first main results quantifies precisely the fluctuations of such linear statistics under mild assumptions on the potential V and the test function ξ. Whenever $V \in \mathscr{C}^7(\mathbb{R})$ is semi-convex, regular, and that we are in the single-cut regime (see Assumption 1 below), we indeed obtain a CLT for linear statistics, in both total variation and $p-$Wasserstein distances ($1 \leq p < \infty$), with the following rates:

- (Theorem 1.2) If $\xi \in \mathscr{C}^{14}(\mathbb{R})$, we obtain the optimal rate $\frac{1}{n}$.
- (Theorem 1.1) If $\xi \in \mathscr{C}^{6}(\mathbb{R})$, we obtain the almost optimal rate $\frac{1}{n^{1-\alpha}}$ for all $\alpha > 0$.
- (Theorem 1.3) If $\xi \in \mathscr{C}^{1,\gamma}(\mathbb{R})$ with $\gamma \in (0,1)$, we obtain an explicit rate of convergence depending on the Hölder exponent γ.

Our bounds in $p-$Wasserstein distance are furthermore obtained in the multivariate case. In fact, we obtain precise rates of convergence, with explicit upper bounds involving n, β and norms of derivatives of ξ. This could allow for limit theorems with varying temperature and/or with a test functions ξ also depending on n.

Our next main result Theorem 1.4 is of qualitative nature and show that the convergence in law of a linear statistic to a Gaussian automatically upgrades to a very strong form of convergence. Namely, if the potential V and the test function ξ are of class $\mathscr{C}^\infty$ and if the latter is not degenerate is some sense, then (X_n) *super-converges* to a Gaussian (see Section 1.3.3 below for definition of the super-convergence). Informally it means that the density of (X_n) converges in the $\mathscr{C}^\infty$-topology to that of a Gaussian. In particular, it entails that the convergence of linear statistics holds in relative entropy.

1.2. **Motivations.** The β-ensembles, also known as $1d$*-log gas*, are related to random matrix theory. We refer to the monographs [33, 20] for details on the subject, as well as [36] for a broader introduction to Coulomb gases. Indeed, in the case $\beta \in \{1, 2, 4\}$ and $V(x) = x^2$, the corresponding β-ensemble describes the joint law of the spectrum of an $n \times n$ random matrix whose density is proportional to $\exp(-\frac{\beta n}{4} \operatorname{Tr} V(M))\mathtt{d}M$, where $\mathrm{d}M$ is the Haar measure on the sets of symmetric, hermitian, or symplectic matrices respectively. The observation that certain β-ensembles relate to eigenvalues of Gaussian ensembles goes back at least to [17].
The spectral macroscopic properties of large Gaussian matrices first observed by [41], actually extend to more general β-ensembles, and it is by now well-understood, see for instance [33, Thm. 11.1.2], that

$$\frac{1}{n} \sum_{i=1}^{n} \xi(\lambda_i) \xrightarrow[n\to\infty]{a.s} \int \xi \mathtt{d}\mu_V,$$

where the non-random *equilibrium measure* μ_V is the unique minimizer of the *mean-field energy*

$$\mathcal{I}_V(\mu) := \int V \mathtt{d}\mu - \frac{1}{2} \iint \log|x-y| \mu(\mathtt{d}x)\mu(\mathtt{d}y).$$

In this article, our goal is to study the fluctuations associated with the above "law of large numbers", with different motivations that we detail below.

1.2.1. *Sharp quantitative Central Limit Theorems.* The seminal work of [24] shows that, in the matrix case and under suitable assumptions on the potential V and the test function ξ, the fluctuations of $\sum_{i=1}^n \xi(\lambda_i)$ around the value $n\int \xi \mathtt{d}\mu_V$ are Gaussian. The quantitative aspect of this convergence has attracted a lot of attention in the last decade: *i*) in [13], total variation bounds are provided in the matrix case via second order Poincaré inequalities; *ii*) in [3], still in the matrix case, near optimal rates of convergence are derived for Kolmogorov distance; *iii*) for general β-ensembles, polynomial rates of convergence are obtained for 2−Wasserstein distance in [26] using a variant of Stein's method due to Meckes. However, in the last reference, the rates obtained are at most $O(n^{-2/3+\varepsilon})$ for general potential and of near optimal order $O(n^{-1+\varepsilon})$ only in the quadratic case. Our primary motivation in this work is to derive total variation estimates, which is a much stronger distance than the Wasserstein ones, but also to get optimal rate of convergence of order $O(n^{-1})$. We also provide sharp rates of convergence in Wasserstein metric in both univariate and multivariate settings, and our findings require less regularity on the test functions than previous results on this question, see [26, 5].

1.2.2. *Implementing a robust Stein's approach for β-ensembles.* Historically, Stein's method [38] has proved its remarkable efficientcy to obtain quantitative CLTs in strong probabilistic metrics such as total variation or relative entropy. Despite numerous applications to various probabilistic models such as: (a) Erdös–Rényi random graphs [35]; (b) infinite-dimensional Gaussian fields [32]; (c) Poisson point processes [34]; (d) free probability [19]; to the best of our knowledge, Stein's method to study β-ensemble has only been implemented in [26, 21], through a variant due to Meckes [30], which in this context consists in approaching linear statistics by infinite sums of approximate eigenvectors of the Dyson generator. This demanding task requires strong regularity assumptions on the test function ξ and prevents one to both get optimal rates and to handle total variation metric. A second motivation for this works is to develop a Stein's method for β-ensembles, sufficiently robust and general to be possibly applied to other models in statistical physics involving Gibbs measures.

1.2.3. *Super-convergence phenomenon.* For sums of independent and identically distributed random variables, due to the convolution structure and provided the common law has some small initial regularity, the regularity in fact improves along the classical CLT. In particular, it is possible to reinforce the classical convergence in law to the get the $\mathscr{C}^\infty$ convergence of densities, see [28]. This particular behavior also appears in the free CLT [8], from which the terminology of *super-convergence* proceeds. Recently, three of the authors have revisited this regularization phenomenon on Wiener chaoses [23] and for quadratic forms [22]. In the setting of β-ensembles, despite the linear nature of the statistics in consideration, no convolution structure arises due to the interaction between the particles. Establishing better-than-expected limit theorems in this dependent setting serves as the third motivation of this article.

1.3. **Statement of the main results.** Let us now precise our assumptions, fix our notations and state formally our main results.

1.3.1. *Assumption and notations.* In the rest of the paper, we will always assume that the potential V satisfies the following conditions. Note that the latter are classical, in particular they coincide with the assumptions in [26].

Assumption 1. *The potential V and associated equilibrium measure μ_V are such that*

- *(Smoothness) $V \in \mathscr{C}^7(\mathbb{R})$.*
- *(Single-cut) The equilibrium measure μ_V is supported on the interval $\Sigma_V := [-1, 1]$.*
- *(Semi-convexity) $\inf_{\mathbb{R}} V'' > -\infty$.*
- *(Regularity) The measure μ_V has a positive density with respect to the semi-circle law.*

Recall that the measure μ_V is here the unique minimizer of the mean-field energy $\mathcal{I}_V$. In particular, it satisfies the Euler–Lagrange equation, for some constant $C_V \in \mathbb{R}$:

$$V(x) - \int \log(|x-y|)\mu_V(\mathtt{d}y) = C_V, \qquad \forall x \in \Sigma_V.$$

Under the above assumptions, the equilibrium measure μ_V admits a density with respect to the semi-circular law $\mu_{sc}(\mathtt{d}x) := \frac{2}{\pi}\sqrt{1-x^2}\mathbb{1}_{[-1,1]}(x)\mathtt{d}x$, namely setting $\rho(\mathtt{d}x) := \frac{\mathtt{d}x}{\pi\sqrt{1-x^2}}$

$$\mu_V(\mathtt{d}x) = S(x)\mu_{sc}(\mathtt{d}x), \quad \text{where} \quad S(x) := \frac{1}{2}\int_{-1}^{1} \frac{V'(x)-V'(y)}{x-y}\rho(\mathtt{d}y).$$

As already mentioned, our main objective is to provide (multivariate) qualitative and quantitative CLTs for linear statistics of β-ensembles. We consider test functions $\xi_1, \dots, \xi_d \colon \mathbb{R} \to \mathbb{R}$ and we define the random vector $X = (X_1, \dots, X_d)$ by

$$X_k := \sum_{i=1}^{n} \xi_k(\lambda_i) - n\int \xi_k \mathtt{d}\mu_V, \quad 1 \le k \le d.$$

For such a vector X, as n goes to infinity, following Equations (1.10) and (1.11) in [26], the candidate limiting mean $m = (m_i)_{1\le i\le d}$ and covariance matrix $C = (c_{i,j})_{1\le i,j,d}$ are given by

$$m_i := \left(\frac{1}{2}-\frac{1}{\beta}\right)\left[\frac{\xi_i(-1)+\xi_i(1)}{2} - \int_{\Sigma_V}\xi_i(x)\rho(\mathtt{d}x) - \frac{1}{2}\int_{\Sigma_V^2}\frac{S'(x)}{S(x)}\frac{\xi_i(x)-\xi_i(y)}{x-y}\rho(\mathtt{d}y)\mu_{sc}(\mathtt{d}x)\right],$$

$$c_{ij} := \frac{1}{2\beta}\int \frac{\xi_i(x)-\xi_i(y)}{x-y}\frac{\xi_j(x)-\xi_j(y)}{x-y}(1-xy)\rho(\mathtt{d}x)\rho(\mathtt{d}y).$$

The above matrix C is well-defined as soon as the ξ_i's are $1/2$-Hölder continuous. We shall see in Section 4.1.1 below that C is indeed a covariance matrix as soon as

$$1, \xi_1, \dots, \xi_d \text{ are linearly independent.} \tag{1.2}$$

For an open set $U \subset \mathbb{R}$, we write $\mathscr{C}^r(U)$ for the space of r times continuously differentiable functions on U with bounded derivatives. We endow it with the Banach norm

$$\|\xi\|_{\mathscr{C}^r(U)} := \max_{r'\le r}\sup_{x\in U}|\xi^{(r')}(x)|.$$

When $\xi = (\xi_1, \dots, \xi_d)$, we also write

$$\|\xi\|_{\mathscr{C}^r(U)} = \sum_{k=1}^{d}\|\xi_k\|_{\mathscr{C}^r(U)}.$$

We write indifferently $|\cdot|$ for the absolute value of a real number, the Euclidean norm of a vector, or the Euclidean norm of a square matrix, also known as its Hilbert–Schmidt norm. For a square symmetric matrix A, we write $\|A\|_{op}$ for its *operator norm*, that is its largest singular value. Given two random variables X and Y, the *$p-$Wasserstein distance* is

$$\mathbf{W}_p(X,Y) := \inf\Big\{\mathbf{E}\big[|X'-Y'|^p\big] : X' \overset{\text{law}}{=} X, Y' \overset{\text{law}}{=} Y\Big\}.$$

The Wasserstein distances only depends on X and Y through their respective laws but we favor a probabilistic notation more suited for the approximation theorems we establish. The $p-$Wasserstein distance induces a topology corresponding to convergence in law together with convergence of the p-th moment [39, Thm. 6.9]. We also work with the total variation distance of real-valued random variables

$$\mathbf{TV}(X,Y) := \sup\{\mathbf{P}[X\in B] - \mathbf{P}[Y\in B] : B \text{ Borel}\}.$$

Let us finally define

$$A_\beta := \left(\frac{1}{\beta}\|C^{1/2}\|_{op}|C^{-1}|\|N\|_{L^p} + \left|\frac{1}{2}-\frac{1}{\beta}\right|\|C^{1/2}\|_{op}\right);$$

$$a_\beta := \left(\frac{1}{\beta}\frac{1}{\sigma^2} + \left|\frac{1}{2}-\frac{1}{\beta}\right|\frac{\sqrt{\pi}}{2}\right).$$

1.3.2. *Quantitative CLT in total variation and Wasserstein distances.* We can now state our main results concerning the quantitative behavior of the fluctuations of linear statistics of β-ensembles. The following bounds are optimal or nearly optimal depending on the regularity of the test functions.

Theorem 1.1. *Let us fix $1 \leq p < \infty$ and $\alpha > 0$. There exists $K_{p,\alpha} > 0$ such that for all test functions $\xi_1, \dots, \xi_d \in \mathscr{C}^6(\mathbb{R})$ satisfying (1.2), we have, with N a standard Gaussian in $\mathbb{R}^d$*

$$\mathbf{W}_p(X, C^{1/2}N + m) \leq K_{p,\alpha} A_\beta \frac{\|\xi\|_{\mathscr{C}^6(\mathbb{R})} + \|\xi\|^2_{\mathscr{C}^2(\mathbb{R})}}{n^{1-\alpha}}.$$

In the univariate case $d = 1$, with $\sigma := \sqrt{c_{11}}$ and $m = m_1$, we also have

$$\mathbf{TV}(X, \sigma N + m) \leq K_{1,\alpha} a_\beta \frac{\|\xi\|_{\mathscr{C}^6(\mathbb{R})} + \|\xi\|^2_{\mathscr{C}^2(\mathbb{R})}}{n^{1-\alpha}}.$$

Assuming more regularity on the test functions, we obtain an optimal rate of convergence.

Theorem 1.2. *For $1 \leq p < \infty$ and $\alpha > 0$, there exists $K_{p,\alpha} > 0$ such that for all test functions $\xi_1, \dots, \xi_d \in \mathscr{C}^{14}(\mathbb{R})$ satisfying (1.2), we have, with N a standard Gaussian in $\mathbb{R}^d$*

$$\mathbf{W}_p(X, C^{1/2}N + m) \leq K_p A_\beta \frac{\|\xi\|_{\mathscr{C}^{14}(\mathbb{R})} + \|\xi\|^2_{\mathscr{C}^7(\mathbb{R})}}{n}.$$

In the univariate case $d = 1$, with $\sigma := \sqrt{c_{11}}$ and $m = m_1$, we also have

$$\mathbf{TV}(X, \sigma N + m) \leq K_1 a_\beta \frac{\|\xi\|_{\mathscr{C}^{14}(\mathbb{R})} + \|\xi\|^2_{\mathscr{C}^7(\mathbb{R})}}{n}.$$

We also prove a theorem for functions of lower regularity. For $\gamma \in (0, 1)$, we write $\xi \in \mathscr{C}^{1,\gamma}(\mathbb{R})$, provided $\xi \in \mathscr{C}^1(\mathbb{R})$ and it satisfies

$$\|\xi\|_{\mathscr{C}^{1,\gamma}(\mathbb{R})} := \|\xi\|_{\mathscr{C}^1(\mathbb{R})} + \sup_{x \neq y} \frac{|\xi'(x) - \xi'(y)|}{|x - y|^\gamma} < \infty.$$

Theorem 1.3. *Let $\xi \in \mathscr{C}^{1,\gamma}(\mathbb{R})$ for some $\gamma \in (0, 1)$. Setting $\sigma := \sqrt{c_{11}}$ and $m = m_1$, and for any $a < \frac{\gamma}{6+\gamma}$, we have*

$$\mathbf{TV}(X, \sigma N + m) \leq K_a a_\beta \frac{\|\xi\|_{\mathscr{C}^{1,\gamma}(\mathbb{R})}}{n^a}.$$

1.3.3. *Super-convergence.* We now turn to our qualitative results and the reinforcement of the mode of convergence is the smooth case. We say that a sequence (X_n) *super-converges* to a non-degenerate Gaussian $N \sim \mathcal{N}(m, \sigma^2)$ with density φ_{m,σ^2} provided that, for all $r \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, the law of X_n admits a density $\varphi_n \in \mathscr{C}^r(\mathbb{R})$, and

$$\limsup_{n \to \infty} \|\varphi_n - \varphi_{m,\sigma^2}\|_{\mathscr{C}^r(\mathbb{R})} = 0.$$

Regarding linear statistics, we obtain that convergence in law can easily be upgraded to super-convergence. For $\alpha > 0$, we say that $\xi \in \mathscr{C}^1(\mathbb{R})$ is *α-regular* provided that

$$\mathsf{Leb}(x \in \mathbb{R}^d : |\xi'(x)| \leq \epsilon) \lesssim \epsilon^\alpha.$$

It is in particular true for any polynomial function ξ.

Theorem 1.4. *Let $\xi \in \mathscr{C}^\infty(\mathbb{R})$ be α-regular such that (X_n) converges in law to a Gaussian variable $N \sim \mathcal{N}(m, \sigma^2)$. Then, (X_n) super-converges to N.*

1.4. **Comparison with existing results.**

1.4.1. *Gaussian fluctuations for linear statistics.* The mathematical study of fluctuations of linear statistics around their equilibrium starts with the seminal work [24]. His method consists in writing the Laplace transform of a linear statistics as the ratio of a partition function associated with a perturbed potential by that associated with the initial potential. Interestingly, the perturbation involves the so-called master operator, noted Θ_V in the sequel, see section 3.1 for precise definition. This operators plays also a central role in our approach. The method developed in [24] has been continued in [25, 37, 9, 10] to obtain qualitative CLT with increasing levels of generality, allowing, for instance, results in the multi-cut regime. All the above results require the potential V to be analytic. Using the same approach on the Laplace transform, [5] proves a CLT when $V \in \mathscr{C}^5(\mathbb{R})$ and $\xi \in \mathscr{C}^3_c(\mathbb{R})$. To the best of our knowledge, these are the best results available regarding the regularity of V and ξ. In contrast, we obtain a CLT for $V \in \mathscr{C}^7(\mathbb{R})$ but allowing to lower the regularity to $\xi \in \mathscr{C}^{1,\gamma}(\mathbb{R})$. Our results are also limited to the single-cut case. We stress that in the multi-cut regime, non-Gaussian fluctuations are known, and additional compatibility conditions on ξ are required to ensure a CLT. We also mention [6] that derives a CLT for linear statistics at all meso-scales, that is for linear statistics where the particles are rescaled by a factor $n^{-\alpha}$ for $\alpha \in (0,1)$. Even if our main result allows to recover CLT at some meso-scale, we cannot reach their full range with our method.

1.4.2. *Stein's method and β-ensemble.* To the best of our knowledge, the reference [26] is the first to implement Stein's method in the setting of *general* β-ensembles and to provide quantitative bounds in $2-$Wasserstein distance which are near optimal for $V(x) = x^2$ but sub-optimal for more general potentials. Their method works for linear statistics that are close to be eigenvalues of the Dyson generator, and technically completing their requires, among others, the *tour de force* of diagonalizing the so-called master operator. Relying on the concept of exchangeable pairs which plays an important role in Stein's method (see e.g. [30]), quantitative CLTs for linear statistics of Haar distributed random variables on compact classical groups are established in [16, 15] and later on, in the context of *circular* β-ensembles, in [40] where exchangeable pairs are built through the n-dimensional circular Dyson Brownian motion. Finally, the contribution [21] which builds upon the techniques given in [26] to provide quantitative (near optimal) CLTs in the high temperature regime.

Let us also mention the seminal article [13]. There, the author introduced a variation around Stein's method, based on *second-order Poincaré inequality* in order to study fluctuations of eigenvalues of matrices with random coefficients, possibly not identically distributed, whose distributions admit suitable densities. It is somewhat complicated to compare the results there with ours, simply because for general potentials V we cannot a priori interpret β-ensembles as spectrum of random matrices with independent coefficients. Besides, [13, Thm. 4.2] obtains convergence to a Gaussian of random variables of the form $\operatorname{Tr} A_n^{p_n}$ where (A_n) are some random matrices as above, and $p_n = o(\log n)$. Note that, in comparison, in our context, it should be possible to get CLTs for $\xi_n(x) = x^{p_n}$ when p_n grows polynomially although the models are pretty different and only coincide for $V(x) = x^2$.

We would like to point out that despite the aforementioned articles and ours being inspired by the philosophy of Stein's method, they strongly differ in the way Stein's method is implemented. Indeed, [26] relies on ideas coined in [30] using exchangeable pairs while our builds on a novel refinement of the well established *Malliavin Stein's* approach, which allows to handle random variables which **are not** in the image of the underlying Markov generator. Indeed, whenever X belongs to the image of L one can directly provide a *Stein's kernel* by setting $\tau(x) = \mathbb{E}\left[\Gamma\left[X, -L^{-1}X\right]\right]$ which fulfils the equation $\mathbb{E}\left[\phi'(X)\tau(X) - X\phi(X)\right] = 0$. The great advantage of this point of view is to control the total variation distance (among others) by $\mathbb{E}\left[|1-\tau(X)|\right]$ but it requires the costly assumption that L is invertible which is not true in general. To overcome this obstacle, we notice that it is enough to show that X is near $\operatorname{Im}(Z)$ in the following sense: $X = LF + Z$ for Z small in probability. This simple remark, which is new to the best of our knowledge, increases considerably the applicability of the above Malliavin-Stein's approach and

still enables to provide total variation bounds but is a priori not sufficient to build a Stein's kernel. While this being noticed, our strategy then consists in proving that any linear statistics is close to the range of L, a step which proceeds from the invertibility of the master operator. As a result, our findings thus refine those of [26] by providing optimal rates of convergence for stronger metrics such as the total variation distance and we believe that our approach could also be successfully implemented in the framework of circular β-ensembles and Haar distributed matrices on compact groups.

1.4.3. *Super-convergence.* To the best of our knowledge, in the context of random matrices or β-ensembles, the question of establishing convergence in metrics stronger than total variation or Wasserstein distances has not been considered yet. We deploy here ideas that are classical in the framework of Malliavin calculus and stochastic analysis and which merely consists in establishing negative moments for the *carré-du-champ* operator applied to the considered linear statistics. Then, relying on integration by parts techniques, we are able to prove strong forms of convergence for the densities of the underlying sequences of random variables.

1.5. **Outline of the proofs an plan of the paper.** Let us give here more details on our strategy of proof and on the plan of the paper.

1.5.1. *The generator of the Dyson Brownian motion.* The overall strategy behind our quantitative and qualitative estimates leverage the characterization of the β-ensemble as the unique invariant distribution of the Dyson Brownian motion. Namely, consider the *generator of the Dyson Brownian motion*

$$\mathsf{L} = \mathsf{L}_{n,\beta,V} =:= \Delta - \beta\nabla H\cdot\nabla = \sum_{i=1}^n \partial_{\lambda_i}^2 - \beta n\sum_{i=1}^n V'(\lambda_i)\partial_{\lambda_i} + \frac{\beta}{2n}\sum_{i\neq j}\frac{\partial_{\lambda_i}-\partial_{\lambda_j}}{\lambda_i-\lambda_j}. \tag{1.3}$$

The operator L is the diffusive Markov generator canonically associated with the β-ensemble. We have indeed the following characterization of the β-ensemble Gibbs measure $\mathbf{P}$:

$$\tilde{\mathbf{P}} = \mathbf{P} \Leftrightarrow \Big(\tilde{\mathbf{E}}\mathsf{L}F = 0,\ \forall F\in\mathscr{C}_c^\infty(\mathbb{R}^n)\Big) \Leftrightarrow \Big(\tilde{\mathbf{E}}[F\mathsf{L}G] = \tilde{\mathbf{E}}[G\mathsf{L}F],\ \forall F,G\in\mathscr{C}_c^\infty(\mathbb{R}^n)\Big).$$

The differential structure induced by L on $\mathbb{R}^n$, technically called a *Dirichlet structure*, comes with a *carré du champ* operator

$$\Gamma[F,G] := \nabla F\cdot\nabla G, \qquad F,\,G\in\mathscr{C}^1(\mathbb{R}^n).$$

We have the following integration by parts formula

$$\mathbf{E}[\Gamma[F,G]] = -\,\mathbf{E}[F\mathsf{L}G], \qquad F,\,G\in\mathscr{C}_c^\infty(\mathbb{R}^n). \tag{1.4}$$

1.5.2. Γ-*Stein's method for β-ensemble.* At an informal level, the data of a diffusive Markov generator generally combines well with Stein's method to provide quantitative bounds for normal approximation. Provided $\ker\mathsf{L}$ is limited to constant functions, L is invertible on mean-zero functions, and the celebrated Γ-*Stein* approach [32, 34, 1, 27], or *Malliavin–Stein* approach in the setting of Gaussian fields or Poisson point processes, yields that the variance of $\Gamma[X,-\mathsf{L}^{-1}X]$ controls the Gaussian fluctuations of X, in total variation or Wasserstein distance. However, in the case β-ensembles, the operator L is in general *not invertible*.
To overcome this difficulty, we amend the classical Γ-Stein approach. Intuitively, whenever $X=\mathsf{L}F$ for some F, $\mathsf{L}^{-1}X$ makes sense despite the non-invertibility of L. In this case, it is natural to expect that $\Gamma[X,-\mathsf{L}^{-1}X]=\Gamma[X,-F]$ controls the Gaussian fluctuations of X. In the next Section 2 and more precisely in Theorem 2.1 below, we formalize this intuition by proving quantitative bounds for the normal approximation, in Wasserstein distance and total variation, of random variables of the form $X=\mathsf{L}F+Z$. Provided, Z is small and that $\Gamma[X,-F]$ is close to a constant σ^2, then X is close to a Gaussian with variance σ^2. This theorem should not come as a surprise to Stein's method aficionados and its proof is rather straightforward. However, this observation has, to the best of our knowledge, never been remarked and this "almost invertibility" decomposition is precisely what allows us to provide an efficient proof. We stress that our bounds

hold for generic abstract diffusive Markov generators and we believe it could prove useful in other statistical physics models where the Gibbs measure has an explicit density.

1.5.3. *The master equation.* The next Section 3 of the paper consists in checking that the Dyson generator is indeed in line with the global strategy developed in Section 2. In order to apply our abstract bound, we need to show that our linear statistics X is of the form $\mathsf{L}F + Z$. With the notations of Section 3, using mostly algebraic manipulations together with the minimality of μ_V with respect to the mean-field energy $\mathcal{I}_V$, we show, in Theorem 3.8, that whenever $F := \sum_{i=1}^n f(\lambda_i)$ is a linear statistic, setting $m_f := \left(\frac{1}{2} - \frac{1}{\beta}\right)\langle f''|\mu_V\rangle$ which is such that the term in parenthesis below is asymptotically centered, we have

$$\frac{1}{n\beta}\mathsf{L}F = \left(\sum_{i=1}^n (\Theta_V f')(\lambda_i) - n\int (\Theta_V f')\mathrm{d}\mu_V - m_f\right) + \frac{Z}{\beta},$$

where Z is a quadratic (which should be thought as a remainder) term and Θ_V is the so-called *master operator*. In particular, under Assumption 1, Θ_V is invertible in a neighborhood of $[-1,1]$. Namely, there exists an open neighborhood U of $[-1,1]$, such that given $\xi \in \mathscr{C}^6(\mathbb{R})$, we can find $\psi \in \mathscr{C}^5_c(\mathbb{R})$ and $c_\xi \in \mathbb{R}$ satisfying

$$(\Theta_V \psi)(x) = \xi(x) + c_\xi, \qquad x \in U.$$

Under Assumption 1, by [11] the (λ_i) enjoy a strong rigidity property, precisely recalled in Theorem 3.3. Informally, the probability that the λ_i's deviate from the grid given by the i/n-th quantiles of μ_V is overwhelmingly small. Note that this rigidity phenomenon is also used in [26]. Thus up to paying a loss factor, that is negligible compared to our bound, we can assume that all the λ_i's are localized in the neighborhood U. By the master equation, choosing f any primitive of ψ, we are thus left with establishing the quantitative CLT for the random variable

$$X = \frac{1}{n\beta}\mathsf{L}F - \frac{Z}{\beta},$$

which falls precisely under the scope of our Stein's bound.

1.5.4. *Controlling the carré du champ and the remainder.* In Section 4, we concretely implement our strategy and complete the proofs of our main results. We first establish the quantitative statements, namely Theorem 1.1, Theorem 1.2 and Theorem 1.3. In view of our strategy, we show that $(n\beta)^{-1}\Gamma[X,-F]$ is close to a constant and that $(\beta)^{-1}Z$ is small. We control the term involving the carré du champ by relying on a simple yet remarkable property of the carré du champ: it preserves linear statistics. Precisely, we have that

$$\Gamma\Big[\sum \varphi(\lambda_i), \sum \chi(\lambda_i)\Big] = \sum \varphi'(\lambda_i)\chi'(\lambda_i).$$

It follows that

$$\frac{1}{n}\Gamma[X,F] = \frac{1}{n}\sum \xi'(\lambda_i)f'(\lambda_i).$$

By the convergence to equilibrium, the above quantity converges to $\int \xi' f' \mathrm{d}\mu_V$. Using the minimality property of μ_V, it can be shown that this last integral coincides with limiting covariance given above. To derive quantitative bounds from there, we rely on a quantitative convergence to equilibrium which also follows from the rigidity estimate from [11].
Handling the quadratic remainder Z is more tedious. We use Fourier inversion to decompose Z into a product of linear statistics. By the aforementioned quantitative law, one of term goes to 0 at a given rate while the other stays bounded. This allows us to conclude for the proof Theorem 1.1. From there, Theorem 1.2 follows by a bootstrap argument: we essentially redo the same proof but instead of using the sub-optimal quantitative law of large numbers provided by [11], we use the CLT we just established that gives us a law of large numbers at speed $1/n$. To conclude for Theorem 1.3, we approximate $\xi \in \mathscr{C}^{1,\gamma}(\mathbb{R})$ by a sequence of smooth functional ξ_ε while choosing $\varepsilon \simeq n^a$, for a well-chosen a.

1.5.5. *Super-convergence.* The proof of our qualitative result Theorem 1.4 is finally given in Section 4.4. The derivation of the super-convergence relies on a lemma, classical in Dirichlet forms / Malliavin calculus theory, stating that negative moments of $\Gamma[X,X]$ control the Sobolev norms of the density of X. We rely again on the fact that Γ preserves linear statistics. In this precise case,

$$\Gamma[X,X] = \sum (\xi'(\lambda_i))^2.$$

We obtain negative moments through a direct control on $\mathbf{P}\big[\sum(\xi'(\lambda_i))^2 \leq \varepsilon\big]$.

2. Stein's method for Markov diffusive operators

In order to carry out our program, we need to establish quantitative bounds in total variation for random variables which are close in some sense to the range of a given Markov diffusive operator L. To the best of our knowledge, the following estimates seem to be new in the well-studied area of Malliavin–Stein's method and are of independent interest. We stress that all the results of this section are valid for any diffusive Markov operator L associated with Γ the so-called carré du champ and $\mathbf{P}$ the invariant measure for L. We refer to [2] for definitions in this abstract setting. Indeed, the only properties used in our proofs are the chain rule and the integration by parts which hold in full generality. The cornerstone of our method relies on the following Theorem which will proved in Section 2.3 based on the content of Sections 2.1 and 2.2.

In case of vector-valued random variables F and G, we extend our definition to the *matrix-valued carré du champ*

$$\Gamma[F,G]_{ij} := \Gamma[F_i, G_j], \qquad i,j = 1,\dots,d.$$

Theorem 2.1. *Take $F_1,\dots,F_d \in \mathbb{D}\mathrm{om}\,\mathsf{L}$, $Z_1,\dots,Z_d \in L^p$, and $X := (\mathsf{L}F_1 + Z_1, \dots, \mathsf{L}F_d + Z_d)$. Let Σ be a positive symmetric matrix, $C = \Sigma^2$ and N be standard Gaussian in $\mathbb{R}^d$. Then, we have*

$$\begin{aligned} \mathbf{W}_p(X, \Sigma N) &\leq \|\Sigma\|_{op} \|N\|_{L^p} \|\mathrm{id} - \Sigma^{-1}\Gamma[X,-F]\Sigma^{-1}\|_{L^p} + \|\Sigma\|_{op}\|Z\|_{L^p} \\ &\leq \|\Sigma\|_{op}\|C^{-1}\|\|N\|_{L^p}\|C - \Gamma[X,-F]\|_{L^p} + \|\Sigma\|_{op}\|Z\|_{L^p}. \end{aligned} \tag{2.1}$$

Moreover, in the univariate case $d=1$, setting $\sigma = \Sigma_{11}$, we obtain

$$\mathbf{TV}(X, \sigma N) = \mathbf{TV}(\sigma^{-1}X, N) \leq \frac{2}{\sigma^2}\|\sigma - \Gamma[X,-F]\|_{L^1} + \frac{\sqrt{\pi}}{2}\|Z\|_{L^1}. \tag{2.2}$$

2.1. **Reminders on Stein kernels.** Let us first recall important results regarding Stein kernels. Here we work on a general probability space $(\Omega, \mathfrak{W}, \mathbf{P})$. Given a multivariate random variable $X = (X_1,\dots,X_d)$, we say that a matrix-valued random variable τ, measurable with respect to X, is a *Stein kernel* for X provided

$$\mathbf{E}[X \cdot \nabla\varphi(X)] = \mathbf{E}\big[\tau \cdot \nabla^2\varphi(X)\big], \qquad \varphi \in \mathscr{C}^\infty(\mathbb{R}^d).$$

Heuristically, Stein kernels are relevant for normal approximation, since X is a standard multivariate normal variable if and only if it admits $\tau = \mathrm{id}$ as a Stein kernel. A key observation at the heart of Stein's method are the following quantitative normal approximation inequalities.

Lemma 2.2 (Stein kernel bounds [27, 18, 31])**.** *Let $X = (X_1,\dots,X_d)$, let N be a standard normal variable on $\mathbb{R}^d$, and τ be Stein kernel for X.*

(i) Let $p \in [1,\infty)$. If $X \in L^p(\Omega)$, then

$$\mathbf{W}_p(X,N)^p \leq \mathbf{E}[|N|^p]\,\mathbf{E}[\|\tau - \mathrm{id}\|^p].$$

(ii) In the univariate setting $d=1$, we have

$$\mathbf{TV}(X,N) \leq 2\,\mathbf{E}[|\tau - 1|].$$

2.2. **Computations of Stein kernels on** $\operatorname{Im}\mathsf{L}$. The celebrated Malliavin–Stein method [32] puts forward a strikingly efficient way to compute a Stein kernel for sufficiently smooth functionals of an infinite-dimensional Gaussian field.

Lemma 2.3. *Take* $F_1, \dots, F_d \in \mathbb{D}\mathrm{om}\,\mathsf{L}$. *Let* $X := (\mathsf{L}F_1, \dots, \mathsf{L}F_d)$, *and define the matrix-valued random variable* $\tau = (\tau_{ij})$, *where*

$$\tau_{ij} := \mathbf{E}[\Gamma[X_i, -F_j] \mid X], \qquad i, j = 1, \dots, d.$$

Then, τ *is a Stein kernel for* X.

Proof. Take $\varphi \in \mathscr{C}^\infty(\mathbb{R}^d)$. Then, using the integration by parts, and then the chain rule, we get

$$\mathbf{E}[X \cdot \nabla\varphi(X)] = \sum_{i=1}^{l} \mathbf{E}[\mathsf{L}F_i \partial_i\varphi(X)] = -\sum_{i,j} \mathbf{E}[\partial_{ij}\varphi(X)\Gamma[F_i, X_j]].$$

This concludes the proof in view of the definition of a Stein kernel. □

Combining Lemmas 2.2 and 2.3, we obtain the following Γ-Stein bound.

Theorem 2.4. *Take* $F_1, \dots, F_d \in \mathbb{D}\mathrm{om}\,\mathsf{L}$, *and* $X := (\mathsf{L}F_1, \dots, \mathsf{L}F_d)$. *Then*

$$\mathbf{W}_p(X, N)^p \le \mathbf{E}[|N|^p]\,\mathbf{E}[\|\mathrm{id} - \Gamma[X, -F]\|^p].$$

Moreover, in the univariate case $d = 1$, *we find*

$$\mathbf{TV}(X, N) \le 2\,\mathbf{E}[|1 - \Gamma[X, -F]|].$$

Remark 2.5. The classical approach on the Wiener space relies on the invertibility of the generator of the Ornstein–Uhlenbeck process. In this case, it is known that

$$\tau_{ij} := \mathbf{E}\big[\Gamma[X_i, -\mathsf{L}^{-1}X_j] \mid X\big], \qquad i, j = 1, \dots, d,$$

is a Stein kernel for $X \in L^2(\Omega)$ with $\mathbf{E}\,X = 0$. This strategy has also been applied in the setting of other diffusive Markov generators with discrete spectrum, see for instance [1, 27]. This actually could be further generalized to any diffusive Markov generators that is invertible away from constants. Note that, the generator of the Dyson Brownian motion L is in general not invertible, in particular it is a priori not true that a given linear statistics belongs to the range of L. Therefore, we need to extend the above techniques for random variables not belonging to the range of L which is precisely the object of the next section.

2.3. **Normal approximation away from** $\operatorname{Im}\mathsf{L}$. Given X, finding F such that $\mathsf{L}F = X$ is a demanding task. Indeed as already mentioned, L may be not invertible. Nevertheless, as we now demonstrate, the bounds obtained by the Stein's method can be amended in order to cover the case where X is close to the range of L that is to say of the form $X = \mathsf{L}F + Z$ with small perturbation Z.

Proof of Theorem 2.1. We first prove the claim in the case $\Sigma = \mathrm{id}$, and then explain how to handle the case of general non-degenerate covariance. For the upper bound (2.1), by the triangle inequality, we have that

$$\mathbf{W}_p(X, N) \le \mathbf{W}_p(\mathsf{L}F, N) + \mathbf{W}_p(\mathsf{L}F, \mathsf{L}F + Z).$$

We apply Lemma 2.3 on the first term on the right-hand side. For the second term by choosing the coupling $(\mathsf{L}F, \mathsf{L}F + Z)$ in the minimization problem defining $\mathbf{W}_p$, we see that

$$\mathbf{W}_p(\mathsf{L}F, \mathsf{L}F + Z)^p \le \mathbf{E}[\|\mathsf{L}F - \mathsf{L}F + Z\|^p] = \mathbf{E}[\|Z\|^p].$$

This concludes the proof of (2.1).
We now prove (2.2). Since, by the chain rule an integration by parts, we have

$$\mathbf{E}\big[\varphi'(X)\Gamma[X, -F]\big] = \mathbf{E}[\varphi(X)\mathsf{L}F] = \mathbf{E}[\varphi(X)X] - \mathbf{E}[\varphi(X)Z],$$

we find that

$$\mathbf{E}\big[\varphi'(X) - X\varphi(X)\big] = \mathbf{E}\big[\varphi'(X)(1 - \Gamma[X, -F])\big] - \mathbf{E}[\varphi(X)Z].$$

Recalling that, by [32, Thm. 3.3.1],

$$\mathbf{TV}(X,N) \leq \sup\Big\{\mathbf{E}\big[\varphi'(X)\big] - \mathbf{E}[X\varphi(X)] : \|\varphi\|_\infty \leq \sqrt{\frac{\pi}{2}},\ \|\varphi'\|_\infty \leq 2\Big\},$$

this concludes the proof in the case where $\Sigma = \mathrm{id}$. Now we take an general positive, hence invertible, covariance matrix Σ. On the one hand, by the stability of Wasserstein distances under the Lipschitz map $\mathbb{R}^n \ni v \mapsto \Sigma v$, we find that

$$\mathbf{W}_p(X, \Sigma N) \leq \|\Sigma\|_{op} \mathbf{W}_p(\Sigma^{-1}X, N).$$

Recall that $\|\cdot\|_{op}$ is the operator norm with respect to the Euclidean norm on $\mathbb{R}^n$. On the other hand, a routine computation yields

$$\Gamma[\Sigma^{-1}X, -\Sigma^{-1}F] = \Sigma^{-1}\Gamma[X,-F]\Sigma^{-1}.$$

Thus by (2.1), we get

$$\begin{aligned}\mathbf{W}_p(X,\Sigma N) &\leq \|\Sigma\|_{op}\|N\|_{L^p}\|\mathrm{id} - \Sigma^{-1}\Gamma[X,-F]\Sigma^{-1}\|_{L^p} + \|\Sigma\|_{op}\|Z\|_{L^p} \\ &\leq \|\Sigma\|_{op}\|C^{-1}\|\|N\|_{L^p}\|C - \Gamma[X,-F]\|_{L^p} + \|\Sigma\|_{op}\|Z\|_{L^p}.\end{aligned} \tag{2.3}$$

Similarly, in the univariate case $d = 1$, we obtain the bound (2.2). □

3. The master equation for linear statistics

We now apply the content of the previous section to the specific operator L which is the generator of the Dyson Brownian motion which was given in equation (1.3).

3.1. **Notations and introduction of relevant operators.** We show here that linear statistics associated with β-ensembles can indeed be expressed in the form $\mathsf{L}F + Z$ thus making them good candidates to apply our Theorem 2.1. For a more concise exposition of our results and proofs, we adopt the following notations related to the empirical measure:

$$\nu_n := \sum_{i=1}^n \delta_{\lambda_i}, \text{ and } \quad \bar{\nu}_n := \nu_n - n\mu_V;$$
$$\mu_n := \frac{\nu_n}{n}, \text{ and } \quad \bar{\mu}_n := \mu_n - \mu_V.$$

We will write $\langle f|\mu\rangle$ for the integration of a function f against a measure μ. We consider the following operators, acting on functions $f \in \mathscr{C}^2(\mathbb{R})$:

$$T_V(f)(x) := \int \frac{f(x)-f(y)}{x-y}\mathtt{d}\mu_V(y) = \int\int_0^1 f'((1-u)x+uy)\mathtt{d}u\mu_V(\mathtt{d}y);$$
$$\Theta_V(f) := -V'f + T_V(f);$$
$$T_n(f)(x) := \int\int_0^1 f''((1-u)x+uy)\mathtt{d}u\mu_n(\mathtt{d}y) = \frac{1}{n}\sum_{i=1}^n \frac{f'(x)-f'(\lambda_i)}{x-\lambda_i}1_{x\neq\lambda_i} + f''(\lambda_i)1_{x=\lambda_i};$$
$$\Theta_n(f) := -V'f' + T_n(f).$$

To some extend, T_n and Θ_n are the empirical counterparts of T_V and Θ_V. Since μ_V is compactly supported and since f is chosen $\mathscr{C}^2$, the integrals above are all convergent. This shows in particular that the operators T_V and Θ_V are well-defined. We recall an *invertibility result* for Θ_V taken from [5].

Lemma 3.1 ([5, Lem. 3.3])**.** *There exists $\delta > 0$, such that if $U := (-1-\delta, 1+\delta)$, for all $\xi \in \mathscr{C}^6(\mathbb{R})$, there exist a constant c_ξ and $\psi \in \mathscr{C}^5_c(\mathbb{R})$ such that*

$$(\Theta_V\psi)(x) = \xi(x) + c_\xi, \qquad x \in U, \tag{3.1}$$
$$\|\psi\|_{\mathscr{C}^5(\mathbb{R})} \leq C\|\xi\|_{\mathscr{C}^6(\mathbb{R})}. \tag{3.2}$$

For the rest of the paper, we fix the margin $\delta > 0$ and the associated neighborhood U given by Lemma 3.1 above.

Remark 3.2. (a) A slight variant of this invertibility result can be found in [4, Lem. 3.2].
(b) We do not assume [5, (H3)], however as noticed in this last reference, this assumption is not used in the proof of [5, Lem. 3.3].
(c) Since we are in the single-cut case, with no singular density, with the notations of [5], we have here $\mathsf{k} = 0$ and $\mathsf{m} = 0$.

3.2. **The rigidity estimates and its consequences.** Many ingredients in our proof proceed from a rigidity result from [11] which is also used in [26].

Theorem 3.3 ([11, Thm 2.4])**.** *Let $\rho_0 := -1$, and ρ_j be the j/n-quantile of μ_V for $j \in \{1, \cdots, n\}$. Set $\hat{\jmath} := \min(j, n-j+1)$ and write $(\lambda_{(j)})$ for the increasingly-ordered vector of λ_j's. For any $\varepsilon > 0$, there exist $c_\varepsilon > 0$ and $N_\varepsilon > 0$ such that for all $n \geq N_\varepsilon$*

$$\mathbf{P}\Big[\exists j \in [1,n] : |\lambda_{(j)} - \rho_j| > \hat{\jmath}^{-\frac{1}{3}} n^{-\frac{2}{3}+\varepsilon}\Big] \leq \exp(-n^{c_\varepsilon}). \tag{3.3}$$

As anticipated, we repeatedly use in our proofs, the following corollary on the control of the outliers. Philosophically, it allows us to work on $U = (-1-\delta, 1+\delta)$ and to discard the rest up to accepting a negligible loss. The next result follows directly from the Theorem 3.3 and fact that the locations $(\rho_j)_{1\leq j\leq n}$ belong to $[-1,1]$.

Lemma 3.4. *There exists $C > 0$ and $c > 0$ such that for n large enough*

$$\mathbf{P}\Big[\max_{1\leq i\leq n} |\lambda_i| \geq 1+\delta\Big] \leq C\exp(-n^c). \tag{3.4}$$

Remark 3.5. Similar results are numerous in the literature: even with weaker assumptions on the potential V than the ones in [11] for which the rigidity is not yet known, one can derive better rate of convergence, such as exponential decay. Such exponential decay follows, for instance, from the large deviations principle for the extreme positions, as in [9, Prop. 2.1]. Under strong assumptions on the potential V, the reference [29, Lem. 4] offers an exponential decay with an explicit dependence on δ, which, of course we do not need since our margin δ is fixed. See also [12, Thm. 1.12] for similar a similar result under weaker assumptions in dimension greater than two. As the rigidity estimate given by Theorem 3.3 is in fact used to quantify the speed of convergence to equilibrium in the next Lemma 3.6 and since the associated bound (3.4) is sufficient for our purpose, we have decided to use the rigidity also to control the outliers.

As proved in [26], Theorem 3.3 provides a polynomial speed of convergence to equilibrium. Namely, they obtain the following lemma that we reproduce below for completeness.

Lemma 3.6 ([26, Lemma 5.3])**.** *Let $\alpha > 0$ and $p \geq 1$. Then, there exists $K_{p,\alpha} > 0$ such that for any bounded and Lipschitz function f we have*

$$\|\langle f|\mu_n - \mu_V\rangle\|_{L^p} \leq \frac{1}{n^{1-\alpha}} K_{p,\alpha}(\|f\|_\infty + \|f'\|_\infty). \tag{3.5}$$

Remark 3.7. As we shall see below, our findings allow to upgrade the rate of convergence given in Lemma 3.6 to get an optimal rate $O(n^{-1})$ for regular enough test functions f.

3.3. **The master equation.** In relation to the Dyson generator L, the master operator Θ_V allows us to derive the following *master equation*, which is at the heart of our argument.

Theorem 3.8. *Consider a test function $f \in \mathscr{C}^2(\mathbb{R})$ and define $F := \langle f|\bar{\nu}_n\rangle$. Then, we have the following decomposition*

$$\frac{\mathsf{L}F}{n} = \frac{2-\beta}{2}\langle f''|\mu_n\rangle + \beta\langle \Theta_V f'|\bar{\nu}_n\rangle + \frac{\beta}{2}\langle T_n f' - T_V f'|\bar{\nu}_n\rangle. \tag{3.6}$$

Proof. Recalling the definitions of L and T_n, we get that

$$\begin{aligned}\frac{\mathsf{L}F}{n} &= \langle f''|\mu_n\rangle - \beta\langle V'f'|\nu_n\rangle + \frac{\beta}{2n}\sum_{i\neq j}^{n}\frac{f'(\lambda_i)-f'(\lambda_j)}{\lambda_i-\lambda_j}\\ &= \frac{2-\beta}{2}\langle f''|\mu_n\rangle - \beta\langle V'f'|\nu_n\rangle + \frac{\beta}{2}\langle T_n f'|\nu_n\rangle\\ &= \frac{2-\beta}{2}\langle f''|\mu_n\rangle - \beta\langle V'f'|\bar\nu_n\rangle + \frac{\beta}{2}\langle T_n f'|\nu_n\rangle - n\beta\langle V'f'|\mu_V\rangle.\end{aligned} \tag{3.7}$$

Recall that the Euler–Lagrange equation for the minimality of μ_V with respect to the energy reads, for some $C_V\in\mathbb{R}$:

$$V(x) - \int \log(|x-y|)\mu_V(\mathtt{d}y) = C_V, \qquad x\in\Sigma_V.$$

Differentiating with respect to x leads to

$$V'(x) - \int \frac{1}{x-y}\mu_V(\mathtt{d}y) = 0, \qquad x\in\Sigma_V.$$

From there, multiplying by $f'(x)$, integrating with respect to $\mu_V(\mathtt{d}x)$, and symmetrizing yields

$$\langle V'f'|\mu_V\rangle = \frac{1}{2}\int\frac{f'(x)-f'(y)}{x-y}\mu_V(\mathtt{d}x)\mu_V(\mathtt{d}y) = \frac{1}{2}\langle T_V f'|\mu_V\rangle.$$

Substituting the latter in (3.7), one gets

$$\begin{aligned}\frac{\mathsf{L}F}{n} &= \frac{2-\beta}{2}\langle f''|\mu_n\rangle - \beta\langle V'f'|\bar\nu_n\rangle + \frac{\beta}{2}\langle T_n f'|\nu_n\rangle - \frac{\beta}{2}\langle T_V f'|n\mu_V\rangle\\ &= \frac{2-\beta}{2}\langle f''|\mu_n\rangle - \beta\langle V'f'|\bar\nu_n\rangle - \frac{\beta}{2}\langle T_V f'|n\mu_V\rangle\\ &\quad + \frac{\beta}{2}\langle T_n f' - T_V f'|\bar\nu_n\rangle + \frac{\beta}{2}\langle T_n f'|n\mu_V\rangle + \frac{\beta}{2}\langle T_V f'|\bar\nu_n\rangle.\end{aligned} \tag{3.8}$$

By Fubini Theorem, we have that $\langle T_n f'|n\mu_V\rangle = \langle T_V f'|\nu_n\rangle$. This gives

$$\langle T_n f'|n\mu_V\rangle - \langle T_V f'|n\mu_V\rangle = \langle T_V f'|\bar\nu_n\rangle.$$

Plugging this equality in (3.8) leads to the announced result. □

4. Proofs of the main results

We now complete the proofs of our main results. Regarding quantitative normal approximation for linear statistics of β-ensembles, we follow the strategy described in Sections 2 and 3, applying the general quantitative bounds given by Theorem 2.1.

4.1. **Proof of Theorem 1.1.**

We first give the proof of Theorem 1.1, which gives a near optimal rate of convergence $O(n^{-1+\alpha})$ for all $\alpha>0$.

4.1.1. *Handling the covariance.* Let us recall the form of the limit covariance matrix C of the linear statistics and let us first prove that it is invertible. Following [26, §4.1, in particular Eq. (4.14)], we have

$$c_{ij} = \frac{1}{2\beta}\int\frac{\xi_i(x)-\xi_i(y)}{x-y}\frac{\xi_j(x)-\xi_j(y)}{x-y}(1-xy)\rho(\mathtt{d}x)\rho(\mathtt{d}y) =: \langle\xi_i,\xi_j\rangle_{\mathscr{H}^{1/2}}.$$

The bilinear symmetric form $\langle\xi_i,\xi_j\rangle_{\mathscr{H}^{1/2}}$ is not a scalar product, since $\langle g,g\rangle_{\mathscr{H}^{1/2}}=0$ if and only if g is constant. Nevertheless, the matrix $C:=(c_{ij})$ is a Gram matrix with respect to the semi-scalar product $\langle\cdot,\cdot\rangle_{\mathscr{H}^{1/2}}$. Thus under the freeness condition (1.2), the matrix C is symmetric definite positive, and we write Σ for its unique positive symmetric square root.

4.1.2. *Preparatory computations.* We now use the master equation (3.6) to decompose the linear statistics in a suitable form to apply Theorem 2.1.

Reduction to the case $\mathsf{L}F + Z$. Let us consider $\xi \in \mathscr{C}^6(\mathbb{R})$, $\psi \in \mathscr{C}^5_c(\mathbb{R})$ associated with ξ through Lemma 3.1 and let f be any primitive of ψ. We then have

$$\Theta_V f' = \xi + c_\xi + \big(\Theta_V f' - \xi - c_\xi\big) 1_{\mathbb{R}\setminus U}.$$

Since $\langle 1|\bar\nu_n\rangle = 0$, and that $\operatorname{supp}\mu_V \subset U$, we find that

$$\langle \xi|\bar\nu_n\rangle = \langle \Theta_V f'|\bar\nu_n\rangle + \sum_{i=1}^n 1_{|\lambda_i|>1+\delta}(\Theta_V f' - \xi - c_\xi).$$

By Lemma 3.4, we have that

$$\left\| \sum_{i=1}^n 1_{|\lambda_i|>1+\delta}(\Theta_V f' - \xi - c_\xi) \right\|_{L^p} = n\mathrm{e}^{-n^c} O\left(\|\xi\|_\infty + \|\Theta_V f'\|_\infty + 1\right).$$

From the explicit expression of Θ_V and (3.2), we find that

$$\|\Theta_V f'\|_\infty \le c(\|f'\|_\infty + \|f''\|_\infty) \le c\|\psi\|_{\mathscr{C}^5(\mathbb{R})} \le c\|\xi\|_{\mathscr{C}^6(\mathbb{R})}.$$

Combining those two estimates, and since $\langle 1|\bar\nu_n\rangle = 0$, it is sufficient to establish our bound for linear statistics of the form $\langle \Theta_V f'|\bar\nu_n\rangle$, where $f' = \psi \in \mathscr{C}^5_c(\mathbb{R})$ is associated with $\xi \in \mathscr{C}^6(\mathbb{R})$ by Lemma 3.1. Thus, we now study

$$X := \big(\langle \Theta_V f_1'|\bar\nu_n\rangle, \dots, \langle \Theta_V f_d'|\bar\nu_n\rangle\big).$$

By Theorem 3.8, we can then decompose X under the form

$$X = m + \frac{1}{n}\mathsf{L}F + Z,$$

where $m = (m_i)_{1\le i\le d}$ with $m_i := (1/2 - 1/\beta)\langle f_i''|\mu_V\rangle$ where $F_i := \frac{1}{\beta}\langle f_i|\bar\nu_n\rangle$, and

$$Z_i := \left(\frac{1}{2} - \frac{1}{\beta}\right)\langle f_i''|\mu_n\rangle - m_i - \frac{1}{2}\langle T_n f_i' - T_V f_i'|\bar\nu_n\rangle. \tag{4.1}$$

In the next section, we apply Theorem 2.1 to provide quantitative bounds.

Computation of $\Gamma[X, -F]$. We repeatedly use the simple yet remarkable observation that linear statistics are stable under the action of the carré du champ, namely

$$\Gamma\big[\langle\varphi|\nu_n\rangle, \langle\psi|\nu_n\rangle\big] = \langle\varphi'\psi'|\nu_n\rangle, \qquad \varphi,\psi \in \mathscr{C}^1(\mathbb{R}). \tag{4.2}$$

Thus, we find that for $1 \le i, j \le d$

$$\Gamma\left[X_i, -\frac{F_j}{n}\right] = \frac{1}{\beta}\langle (\Theta_V f_i')' f_j'|\mu_n\rangle. \tag{4.3}$$

Heuristics. As a consequence of the convergence to equilibrium, $\mu_n \to \mu_V$, we see that the sum of the first two terms on the right-hand side in (4.1) converges to zero. Similarly, by Equation (4.3), the term $\Gamma[X_i, -F_j/n]$ converges to $\frac{1}{\beta}\langle \xi_i' f_j'|\mu_V\rangle$ which will be the limit covariance of $(\langle\xi_i, \bar\nu_n\rangle, \langle\xi_j, \bar\nu_n\rangle)$. Relying on the expression of the limit variance given in [26] and by uniqueness one must therefore have

$$c_{ij} = \lim_n \operatorname{Cov}(\langle\xi_i, \bar\nu_n\rangle, \langle\xi_j, \bar\nu_n\rangle) = \frac{1}{\beta}\langle \xi_i' f_j'|\mu_V\rangle.$$

However, the determination of the asymptotics of all the terms involving the quadratic term $\langle (T_n - T_v) f_i'|\bar\nu_n\rangle$ is more delicate. Indeed, the same heuristic on the convergence to equilibrium shows that this term is of order $o(n)$ whereas one would expect $o(1)$. We handle this remainder in two steps:

(i) We split the quadratic term in a product of linear terms, using Fourier inversion. Indeed since $f_i' \in \mathscr{C}^5_c(\mathbb{R})$ is is in the image of the Fourier transform.

(ii) We control each of the linear term as before using convergence to equilibrium. This is where we need the finer quantitative estimates recalled in Lemma 3.6.

4.1.3. *Quantitative control and completion of the proof.* Le us now quantify the remainders mentioned in the last Section.

Splitting the remainder by Fourier inversion. For simplicity, we omit the indices here and write f for f_i, $1 \leq i \leq d$. Since $f' \in \mathscr{C}^5_c(\mathbb{R})$, we have

$$|\widehat{f''}(t)| \leq 2\frac{\|f^{(6)}\|_\infty}{(1+|t|)^4}, \qquad t \in \mathbb{R}. \tag{4.4}$$

In particular, $\widehat{f''} \in L^1(\mathbb{R})$. By the Fourier inversion Theorem, one finds

$$\begin{aligned} \langle (T_n - T_V)f' | \bar{\nu}_n \rangle &= n \iint \int_0^1 f''((1-u)x+uy)(\mu_n - \mu_V)(\mathtt{d}x)(\mu_n - \mu_V)(\mathtt{d}y) \\ &= n \int_{\mathbb{R}} \int_0^1 \widehat{f''}(t) \langle \mathrm{e}^{\mathrm{i}tu\bullet} | \mu_n - \mu_V \rangle \langle \mathrm{e}^{\mathrm{i}t(1-u)\bullet} | \mu_n - \mu_V \rangle \mathtt{d}u\mathtt{d}t. \end{aligned} \tag{4.5}$$

By Lemma 3.6 applied to the exponential functions and by Hölder's inequality, we have

$$n\|\langle \mathrm{e}^{\mathrm{i}tu\bullet} | \mu_n - \mu_V \rangle \langle \mathrm{e}^{\mathrm{i}t(1-u)\bullet} | \mu_n - \mu_V \rangle\|_{L^p} \leq \frac{1}{n^{1-2\alpha}} K_{p,\alpha}^2 (1+|t|+t^2).$$

Thus, reporting in (4.5), and up to changing the constants from line to line, we find that

$$\|\langle (T_n - T_V)f' | \bar{\nu}_n \rangle\|_{L^p} \leq \frac{1}{n^{1-2\alpha}} K_{p,\alpha} \int_{\mathbb{R}} |\widehat{f''}(t)| \big(1+|t|+t^2\big) \mathtt{d}t \leq \frac{1}{n^{1-2\alpha}} K_{p,\alpha} \|f^{(6)}\|_\infty. \tag{4.6}$$

Completion of the proof. We can finally complete the proof Theorem 1.1. Again, to simplify the expressions, we omit the indices $1 \leq i \leq d$. here. Recall the definition of the term Z given by Equation (4.1). By Lemma 3.6 and the above Equation (4.6), we find that

$$\|Z\|_{L^p} \leq \left(\frac{1}{2} - \frac{1}{\beta}\right) \|\langle f'' | \mu_n - \mu_V \rangle\|_{L^p} + \|\langle (T_n - T_V)f' | \bar{\nu}_n \rangle\|_{L^p} \leq \frac{1}{n^{1-\alpha}} K_{p,\alpha} \|f'\|_{\mathscr{C}^5(\mathbb{R})}.$$

Introducing the random matrix

$$C_n(i,j) := \langle \xi_i' f_i' | \mu_n \rangle,$$

we have by Equation (4.3)

$$\left\| \frac{1}{\beta} C - \Gamma[X, -F/n] \right\|_{L^p} = \frac{1}{\beta} \|C - C_n\|.$$

Invoking Lemma 3.6 to control $C - C_n$ yields, again with constants which may change from line to line,

$$\begin{aligned} \|C - C_n\|_{L^p} &\leq K_{p,\alpha} \frac{1}{n^{1-\alpha}} \big(\|\xi'\|_\infty \|f'\|_\infty + \|\xi'\|_\infty \|f''\|_\infty + \|\xi''\|_\infty \|f'\|_\infty \big) \\ &\leq \frac{K_{p,\alpha}}{n^{1-\alpha}} \|\xi\|_{\mathscr{C}^2(\mathbb{R})} \|f\|_{\mathscr{C}^2(\mathbb{R})}. \end{aligned} \tag{4.7}$$

Since by Lemma 3.1, we have $\|f\|_{\mathscr{C}^2(\mathbb{R})} \leq K \|\xi\|_{\mathscr{C}^2(\mathbb{R})}$, all the bounds above can thus be rewritten only in term of the norm $\|\xi\|_{\mathscr{C}^2(\mathbb{R})}$. Injecting these estimates in the general upper bound given by Theorem 2.1, we arrive at the announced result.

4.2. **Proof of Theorem 1.2.** We now give the proof of Theorem 1.2, which gives an optimal rate of convergence $O(n^{-1})$ in the case where the test function ξ is sufficiently regular. We work here in dimension $d = 1$ for simplicity, but the proof will extend verbatim to higher dimensions.

The strategy used in order to obtain the optimal rate of convergence consists in bootstrapping the argument used in the previous proof of Theorem 1.1. Except that instead of invoking Lemma 3.6, we use here a better law of large numbers provided by Theorem 1.1 itself. Namely, for $\xi \in \mathscr{C}^6(\mathbb{R})$, taking $\alpha := 1/2$ and using the p-Wasserstein bound for any $p \geq 1$ given by Theorem 1.1, we have that

$$\|\langle \xi | \bar{\nu}_n \rangle\|_{L^p} \leq K_p \left(\frac{\|\xi\|_{\mathscr{C}^6(\mathbb{R})}) + \|\xi\|^2_{\mathscr{C}^2(\mathbb{R})}}{n^{1/2}} \right) + \|\sigma N + m\|_{L^p},$$

where N is a standard Gaussian and both the mean $m = m_1$ and the standard deviation $\sigma = \sqrt{c_{11}}$ are controlled by $\|\xi\|_{\mathscr{C}^1(\mathbb{R})}$, see Section 1.3.1 where their explicit expressions are given. This shows that provided we can control the 6-th derivative, we can remove the polynomial loss in the quantitative law of large numbers given by Lemma 3.6. At the level of the upper bound (4.6), this implies that we have a t^{12} appearing in the integral, which leads to an upper bound controlled by $\|\xi\|_{\mathscr{C}^{14}(\mathbb{R})}$. Similarly, in the upper bound (4.7), the right hand side now yields a control of the form $\|\xi'\|_{\mathscr{C}^6(\mathbb{R})}\|f'\|_{\mathscr{C}^6(\mathbb{R})}$ and concludes the proof.

As a result and as already mentioned in Remark 3.7, this argument implies that any polynomial speed of convergence on the law of large numbers can be upgraded to the optimal speed $\frac{1}{n}$, the price to pay is to impose stronger regularity assumption on the test functions.

4.3. **Proof of Theorem 1.3.** We now show how Theorem 1.3 dealing with functions with Hölder derivative easily follows from Theorem 1.1 together with Lemma 3.6.

Take η a smooth probability density supported on $[-1, 1]$ with finite first moment. Defining $\eta_\varepsilon := \frac{1}{\varepsilon}\eta(\frac{\bullet}{\varepsilon})$, and $\xi_\varepsilon := \xi * \eta_\varepsilon$, we see by direct computations that

$$\begin{aligned} \|\xi - \xi_\varepsilon\|_\infty &\leq \|\xi'\|_\infty \varepsilon \int |y|\eta(y), \\ \|\xi' - \xi'_\varepsilon\|_\infty &\leq \|\xi\|_{\mathscr{C}^{1,\gamma}(\mathbb{R})}\, \varepsilon^\gamma \int |y|^\gamma \eta(y). \end{aligned} \tag{4.8}$$

Thus by Lemma 3.6, for any $\alpha > 0$, we have

$$\|\langle \xi - \xi_\varepsilon | \bar{\nu}_n \rangle\|_1 \leq K_\alpha \varepsilon^\gamma n^\alpha \|\xi\|_{\mathscr{C}^{1,\gamma}(\mathbb{R})}.$$

Now we use the master equation (3.6) for the linear statistics ξ_ε and we write:

$$\frac{\mathsf{L}F_\varepsilon}{n} = \frac{2-\beta}{2}\langle f''_\varepsilon | \mu_n \rangle + \beta \langle \Theta_v f'_\varepsilon | \bar{\nu}_n \rangle + \frac{\beta}{2}\langle T_n f'_\varepsilon - T_V f'_\varepsilon | \bar{\nu}_n \rangle,$$

where $\Theta_v f'_\varepsilon = \xi_\varepsilon$ on the neighbourhood U. Mimicking the arguments developed in Section 4.1.2, we thus have:

$$\frac{\mathsf{L}F_\varepsilon}{n} = \frac{2-\beta}{2}\langle f''_\varepsilon | \mu_n \rangle + \beta \langle \xi_\varepsilon | \bar{\nu}_n \rangle + \frac{\beta}{2}\langle T_n f'_\varepsilon - T_V f'_\varepsilon | \bar{\nu}_n \rangle + O\left(n e^{-n^c} \|\xi_\varepsilon\|_{\mathscr{C}^6} \right).$$

As a result, we get

$$\frac{\mathsf{L}F_\varepsilon}{n} = \frac{2-\beta}{2}\langle f''_\varepsilon | \mu_n \rangle + \beta \langle \xi | \bar{\nu}_n \rangle + \frac{\beta}{2}\langle T_n f'_\varepsilon - T_V f'_\varepsilon | \bar{\nu}_n \rangle + O\left(n e^{-n^c} \|\xi_\varepsilon\|_{\mathscr{C}^6} \right) + \langle \xi - \xi_\varepsilon | \bar{\nu}_n \rangle.$$

Hence, we may decompose the linear statistics in the form $\langle \xi | \bar{\nu}_n \rangle = m_\varepsilon + \frac{\mathsf{L}F_\varepsilon}{n} + Z_\varepsilon$ as in the proof of Theorem 1.1 with this time

$$m_\varepsilon := \left(\frac{1}{2} - \frac{1}{\beta} \right) \langle f''_\varepsilon | \mu_V \rangle,$$

$$Z_\varepsilon := \frac{2-\beta}{2}\langle f''_\varepsilon | \mu_n - \mu_V \rangle + \frac{\beta}{2}\langle T_n f'_\varepsilon - T_V f'_\varepsilon | \bar{\nu}_n \rangle + O\left(n e^{-n^c} \|\xi_\varepsilon\|_{\mathscr{C}^6} \right) + \langle \xi - \xi_\varepsilon | \bar{\nu}_n \rangle.$$

Gathering the bounds given in the proof of Theorem 1.1 and the above Equation (4.8), we then obtain the following controls

$$\left\| \frac{2-\beta}{2} \langle f_\varepsilon'' | \mu_n - \mu_V \rangle + \frac{\beta}{2} \langle T_n f_\varepsilon' - T_V f_\varepsilon' | \bar{\nu}_n \rangle \right\|_1 \leq \frac{K_\alpha}{n^{1-2\alpha}} \left(\|\xi_\varepsilon\|_{\mathscr{C}^6} + \|\xi_\varepsilon\|_{\mathscr{C}^2}^2 \right),$$

$$\|\langle \xi - \xi_\varepsilon | \bar{\nu}_n \rangle\|_1 \leq K_\alpha \varepsilon^\gamma n^\alpha \|\xi\|_{\mathscr{C}^{1,\gamma}(\mathbb{R})},$$

$$\|C_\varepsilon - C_{n,\varepsilon}\|_{L^p} \leq \frac{K}{n^{1-\alpha}} \|\xi_\varepsilon\|_{\mathscr{C}^2}^2,$$

where C_ε denote the limit variance associated with the linear statistic ξ_ε and $C_{n,\varepsilon} := \langle \xi_\varepsilon' f_\varepsilon' | \mu_n \rangle$ is the empirical analogue. By Equation (2.2) in Theorem 2.1, combined with the three above bounds we may write for some constant K_α

$$\mathbf{TV}\left(\langle \xi | \bar{\nu}_n \rangle, m_\varepsilon + \sigma_\varepsilon N \right) \leq K_\alpha \left(\frac{1}{\varepsilon^6 n^{1-2\alpha}} + \varepsilon^\gamma n^\alpha \right) \|\xi\|_{\mathscr{C}^{1,\gamma}(\mathbb{R})}. \tag{4.9}$$

In order to complete the proof, one needs to show that m_ε and σ_ε converge towards m and σ and to estimate $mathbfTV\left(m_\varepsilon + \sigma_\varepsilon N, m + \sigma N\right)$ as ε goes to zero. First of all, relying on the explicit expressions of the limit means and variances, we have indeed as ε goes to zero

$$|m_\varepsilon - m| = O\left(\|\xi_\varepsilon - \xi\|_{\mathscr{C}^1} \right) = O\left(\varepsilon^\gamma \|\xi\|_{\mathscr{C}^{1,\gamma}} \right), \quad \text{as well as} \quad |\sigma_\varepsilon - \sigma| = O\left(\varepsilon^\gamma \|\xi\|_{\mathscr{C}^{1,\gamma}} \right).$$

Finally, using for instance Theorem 1.3 in [14], we derive that

$$\mathbf{TV}\left(m_\varepsilon + \sigma_\varepsilon N, m + \sigma N \right) \leq K \varepsilon^\gamma \|\xi\|_{\mathscr{C}^{1,\gamma}}.$$

Optimizing in the parameters in Equation (4.9), since α can be chosen arbitrarily small, we get the announced bound of $O(n^{-a})$ for any $a < \frac{\gamma}{\gamma+6}$.

4.4. **Proof of Theorem 1.4.** We finally give the proof of Theorem 1.4 stating the super-convergence of linear statistics of $\beta-$ensembles. If X denotes our linear statistics, the proof is based on the fact that $\Gamma[X,X]$ admit negative moments. This fact, combined with classical integration by parts techniques indeed ensure the convergence of densities.

4.4.1. *Control of the negative moments of* Γ. Let us first establish the following lemma.

Lemma 4.1. *Take $\xi \in \mathscr{C}^1(\mathbb{R})$ such that there exist $\alpha > 0$ such that*

$$\mathsf{Leb}\big(x \in \mathbb{R} : |\xi'(x)| \leq \varepsilon \big) \lesssim \varepsilon^\alpha, \qquad \varepsilon > 0.$$

Then, there exist $\gamma > 0$, such that for n large enough

$$\mathbf{P}\left[\frac{1}{n} \sum_{i=1}^n \xi'(\lambda_i)^2 \leq \varepsilon \right] \lesssim \varepsilon^{\gamma n}, \qquad \varepsilon > 0. \tag{4.10}$$

Proof of Lemma 4.1. To simplify the expressions in this proof, we set $g := (\xi')^2$ and $m := \langle g | \mu_V \rangle$. We distinguish here two regimes depending on the magnitude of ε.

The large deviation regime. Take $\varepsilon < \frac{1}{2} m$, then by the large deviations principle for β-ensembles [36, Thm. 2.3] (originally proved in [7]), there exist $c = c_g > 0$ such that, for n large enough

$$\mathbf{P}[\langle g | \mu_n \rangle \leq \varepsilon] \leq \mathbf{P}\left[|\langle g | \bar{\mu}_n \rangle| \geq \frac{m}{2} \right] \leq \exp\big(-cn^2 \big).$$

In this case, in the regime $\varepsilon \geq \mathrm{e}^{-kn}$ where k is some constant to be fixed later, we have

$$\mathrm{e}^{-cn^2} = (\mathrm{e}^{-kn})^{n\frac{c}{k}} \leq \varepsilon^{n\frac{c}{k}}.$$

The free energy regime. Let us work in the regime $\varepsilon \leq \mathrm{e}^{-kn}$, where we recall that we have the freedom to choose k. By the explicit form of the density for the β-ensemble, we find that

$$\begin{aligned}\mathbf{P}[\langle g|\mu_n\rangle \leq \varepsilon] &\leq \mathbf{P}[g(\lambda_i) \leq \varepsilon n,\, \forall i \in [1,n]] \\ &= \frac{1}{Z_{n,\beta}} \int_{\mathbb{R}^n} \prod_{i=1}^{n} \mathbf{1}_{\{g(\lambda_i)\leq n\varepsilon\}} \mathrm{e}^{-\beta H_n(\lambda_1,\dots,\lambda_n)} \mathtt{d}\lambda_1 \dots \mathtt{d}\lambda_n.\end{aligned}$$

Thus applying Cauchy–Schwarz inequality and using the $\alpha-$regular condition on ξ hence g, we obtain

$$\mathbf{P}[\langle g|\mu_n\rangle \leq \varepsilon] \leq \frac{Z_{n,2\beta}^{1/2}}{Z_{n,\beta}} (n\varepsilon)^{n\alpha/2}.$$

By the large deviations principle, see for example [36, Thm. 2.3], we know that

$$\frac{1}{n^2} \log Z_{n,\beta} \xrightarrow[n\to\infty]{} -\frac{\beta}{2} \mathcal{I}_V(\mu_V).$$

In particular, there exists a constant $a(\beta)$ such that

$$\frac{Z_{n,2\beta}^{1/2}}{Z_{n,\beta}} \leq \mathrm{e}^{n^2 a(\beta)}.$$

As a result, we have

$$\mathbf{P}[\langle g|\mu_n\rangle \leq \varepsilon] \leq (n\varepsilon)^{n\alpha/2} \mathrm{e}^{a(\beta)n^2}.$$

We can assume that $a(\beta) \geq 0$ otherwise the claim is trivial. In this regime, we have

$$\mathrm{e}^{a(\beta)n^2} \leq \varepsilon^{-a(\beta)n/k}, \qquad n^{n\alpha/2} \leq \varepsilon^{-\alpha/(2k)\log n}.$$

It follows that

$$\mathbf{P}[\langle g|\mu_n\rangle \leq \varepsilon] \leq \varepsilon^{n(\alpha/2 - a(\beta)/k) - \alpha/(2k)\log n}.$$

Choosing k large enough so that $\alpha/2 - a(\beta)/kr > 0$ then yields and upper bound for the probability of order ε^{cn}, for some $c > 0$. □

Recall that if $X = \sum_{i=1}^n \xi(\lambda_i)$ is a linear statistics, then we have $\Gamma[X,X] = \sum_{i=1}^n \xi'(\lambda_i)^2$. As a result, given an exponent $\alpha > 0$, we can write

$$\mathbf{E}\left[\frac{1}{\Gamma[X,X]^\alpha}\right] = \int_{\mathbb{R}^+} \mathbf{P}\left[\frac{1}{\Gamma[X,X]^\alpha} > t\right] dt = \int_{\mathbb{R}^+} \mathbf{P}\left[\Gamma[X,X] < \frac{1}{t^{1/\alpha}}\right] dt.$$

As a result, the last Lemma 4.1 indeed ensures that $\Gamma[X,X]$ admits negative moments.

4.4.2. *Regularity and super convergence.* The proof of Theorem 1.4 now follows from a well-known integration by parts procedure see [22, §2.1], or [23, §3.2.2], for details. We only recall the first part of the proof to highlight that, in the setting of β-ensemble, the correct quantities to control are the negative moments of $\Gamma[X,X]/n$, and not those of $\Gamma[X,X]$, as it is the case in [22, 23]. Take $\varphi \in \mathscr{C}_c^1(\mathbb{R})$ and write

$$\mathbf{E}\big[\varphi'(X)\big] = \mathbf{E}\left[\varphi(X)\left(\frac{\Gamma[X,\Gamma[X,X]]}{\Gamma[X,X]^2} - \frac{\mathsf{L}X}{\Gamma[X,X]}\right)\right]. \tag{4.11}$$

By Equation (4.2), we find that

$$\Gamma[X,X] = \langle (\xi')^2 | \nu_n\rangle, \qquad \Gamma[X,\Gamma[X,X]] = 2\langle (\xi')^2 \xi'' | \nu_n\rangle.$$

As a result, as n goes to infinity, we have

$$\frac{\Gamma[X,\Gamma[X,X]]}{\Gamma[X,X]^2} \to 0.$$

Now in view of the master equation (3.6), we have the decomposition

$$\Gamma\Big[\frac{\mathsf{L}X}{n},\frac{X}{n}\Big]=\frac{2-\beta}{2}\frac{1}{n}\langle\xi^{(3)}\xi'|\mu_n\rangle+\frac{1}{\beta}\langle(\Theta_V\xi')'\xi'|\mu_n\rangle - \frac{1}{2\beta}\Gamma\Big[\langle(T_V-T_n)\xi'|\nu_n\rangle,\frac{X}{n}\Big]. \tag{4.12}$$

Using again the Fourier splitting, by the chain rule, dominated convergence, and (4.2), we find that

$$\begin{aligned}R_n&:=\Gamma\big[\langle(T_n-T_V)\xi'|\bar\nu_n\rangle,\langle\psi|\mu_n\rangle\big]\\&=\mathrm{i}\int_{\mathbb{R}}\widehat{\xi''}(t)t\int_0^1 u\langle \mathrm{e}^{\mathrm{i}u\bullet}\psi'|\mu_n-\mu_V\rangle\langle \mathrm{e}^{\mathrm{i}t(1-u)\bullet}|\mu_n-\mu_V\rangle\mathrm{d}u\mathrm{d}t\\&+\mathrm{i}\int_{\mathbb{R}}\widehat{\xi''}(t)t\int_0^1 (1-u)\langle \mathrm{e}^{\mathrm{i}t(1-u)\bullet}\psi'|\mu_n-\mu_V\rangle\langle \mathrm{e}^{\mathrm{i}tu\bullet}|\mu_n-\mu_V\rangle\mathrm{d}u\mathrm{d}t.\end{aligned}$$

By Lemma 3.6, we then obtain that $R_n\to 0$ in L^2. By Lemma 2.3, this shows that $\mathsf{L}F/n$ converges to a Gaussian with respect to $\mathbf{W}_2$. In particular, its variance is of order 1. Thus re-writing, the last term in (4.11) as the product

$$\frac{\mathsf{L}X}{n}\frac{n}{\Gamma[X,X]},$$

the need to control the negative moments of $\Gamma[X,X]/n$ is more obvious. It implies in particular that for some constant $C>0$ and any $\phi\in\mathscr{C}^1$ we have $|\mathbb{E}[\phi'(X)]|\le C\|\phi\|_\infty$. This argument can be iterated and for any $p>1$ one would get similarly for n large enough that $\big|\mathbb{E}\big[\phi^{(p)}(X)\big]\big|\le C\|\phi\|_\infty$. The latter combined with the convergence in law proved in Theorem 1.1 entails the announced super-convergence, see [23] for more details.

References

[1] Ehsan Azmoodeh, Simon Campese, and Guillaume Poly. Fourth moment theorems for Markov diffusion generators. *J. Funct. Anal.*, 266(4):2341–2359, 2014.

[2] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators.* Number Vol. 348 in Grundlehren Math. Wiss. Springer, 2014.

[3] Zhigang Bao and Yukun He. Quantitative CLT for linear eigenvalue statistics of Wigner matrices. *The Annals of Applied Probability*, 33(6B):5171 – 5207, 2023.

[4] F. Bekerman, A. Figalli, and A. Guionnet. Transport maps for β-matrix models and universality. *Commun. Math. Phys.*, 338(2):589–619, 2015.

[5] Florent Bekerman, Thomas Leblé, and Sylvia Serfaty. CLT for fluctuations of β-ensembles with general potential. *Electron. J. Probab.*, 23:31, 2018. Id/No 115.

[6] Florent Bekerman and Asad Lodhia. Mesoscopic central limit theorem for general β-ensembles. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 54(4):1917–1938, 2018.

[7] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Relat. Fields*, 108(4):517–542, 1997.

[8] H. Bercovici and D. Voiculescu. Superconvergence to the central limit and failure of the Cramér theorem for free random variables. *Probab. Theory Relat. Fields*, 103(2):215–222, 1995.

[9] G. Borot and A. Guionnet. Asymptotic expansion of β-matrix models in the one-cut regime. *Commun. Math. Phys.*, 317(2):447–483, 2013.

[10] Gaëtan Borot and Alice Guionnet. Asymptotic expansion of β matrix models in the multi-cut regime. *Forum Math. Sigma*, 12:93, 2024. Id/No e13.

[11] Paul Bourgade, László Erdös, and Horng-Tzer Yau. Edge universality of beta ensembles. *Commun. Math. Phys.*, 332(1):261–353, 2014.

[12] Djalil Chafaï, Adrien Hardy, and Mylène Maïda. Concentration for Coulomb gases and Coulomb transport inequalities. *J. Funct. Anal.*, 275(6):1447–1483, 2018.

[13] Sourav Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Relat. Fields*, 143(1-2):1–40, 2009.

[14] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional gaussians with the same mean. *arXiv preprint arXiv:1810.08693*, 2018.

[15] Christian Döbler and Michael Stolz. Stein's method and the multivariate CLT for traces of powers on the classical compact groups. *Electron. J. Probab.*, 16:no. 86, 2375–2405, 2011.

[16] Christian Döbler and Michael Stolz. A quantitative central limit theorem for linear statistics of random matrix eigenvalues. *J. Theoret. Probab.*, 27(3):945–953, 2014.
[17] Freeman J. Dyson. Statistical theory of the energy levels of complex systems. I-III. *J. Math. Phys.*, 3:140–156, 157–165, 166–175, 1962.
[18] Max Fathi. Stein kernels and moment maps. *Ann. Probab.*, 47(4):2172–2185, 2019.
[19] Max Fathi and Brent Nelson. Free Stein kernels and an improvement of the free logarithmic Sobolev inequality. *Adv. Math.*, 317:193–223, 2017.
[20] Peter J. Forrester. *Log-gases and random matrices.* Number Vol. 34 in Lond. Math. Soc. Monogr. Ser. Princeton University Press, 2010.
[21] Adrien Hardy and Gaultier Lambert. CLT for circular beta-ensembles at high temperature. *Journal of Functional Analysis*, 280(7):108869, 2021.
[22] Ronan Herry, Dominique Malicet, and Guillaume Poly. Regularity of laws via dirichlet forms – application to quadratic forms in independent and identically distributed random variables, 2023.
[23] Ronan Herry, Dominique Malicet, and Guillaume Poly. Superconvergence phenomenon in wiener chaoses, 2024.
[24] Kurt Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.
[25] T. Kriecherbauer and M. Shcherbina. Fluctuations of eigenvalues of matrix models and their applications, 2010.
[26] Gaultier Lambert, Michel Ledoux, and Christian Webb. Quantitative normal approximation of linear statistics of β-ensembles. *Ann. Probab.*, 47(5):2619–2685, 2019.
[27] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.*, 25(1):256–306, 2015.
[28] Pierre Louis Lions and Giuseppe Toscani. A strengthened central limit theorem for smooth densities. *J. Funct. Anal.*, 129(1):148–167, 1995.
[29] Mylène Maïda and Édouard Maurel-Segala. Free transport-entropy inequalities for non-convex potentials and application to concentration for random matrices. *Probab. Theory Relat. Fields*, 159(1-2):329–356, 2014.
[30] Elizabeth Meckes. *On Stein's method for multivariate normal approximation*, pages 153–178. Number Vol 5 in Inst. Math. Stat. Collect. Institute of Mathematical Statistics, 2009.
[31] Ivan Nourdin and Giovanni Peccati. Stein's method on wiener chaos. *Probability Theory and Related Fields*, 145:75–118, 2009.
[32] Ivan Nourdin and Giovanni Peccati. *Normal approximations with Malliavin calculus. From Stein's method to universality.* Number Vol. 192 in Camb. Tracts Math. Cambridge University Press, 2012.
[33] Leonid Pastur and Mariya Shcherbina. *Eigenvalue distribution of large random matrices.* Number Vol. 171 in Math. Surv. Monogr. American Mathematical Society, 2011.
[34] Giovanni Peccati and Matthias Reitzner, editors. *Stochastic analysis for Poisson point processes. Malliavin calculus, Wiener-Itô chaos expansions and stochastic geometry.* Number Vol. 7 in Bocconi Springer Ser. Bocconi University Press; Springer, 2016.
[35] Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.
[36] Sylvia Serfaty. *Coulomb gases and Ginzburg-Landau vortices.* Zur. Lect. Adv. Math. European Mathematical Society, 2015.
[37] M. Shcherbina. Fluctuations of linear eigenvalue statistics of β matrix models in the multi-cut regime. *J. Stat. Phys.*, 151(6):1004–1034, 2013.
[38] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. 6th Berkeley Sympos. math. Statist. Probab., Univ. Calif. 1970, 2, 583-602., 1972.
[39] Cédric Villani. *Optimal transport. Old and new.* Number Vol. 338 in Grundlehren Math. Wiss. Springer, 2009.
[40] Christian Webb. Linear statistics of the circular β-ensemble, Stein's method, and circular Dyson Brownian motion. *Electron. J. Probab.*, 21:Paper No. 25, 16, 2016.
[41] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. Math. (2)*, 67:325–327, 1958.

IRMAR, Université de Rennes 1
Email address: jurgen.angst@univ-rennes.fr

IRMAR, Université de Rennes 1
Email address: ronan.herry@univ-rennes.fr
URL: https://orcid.org/0000-0001-6313-1372

LAMA, Université Gustave Eiffel
Email address: dominique.malicet@univ-eiffel.fr
URL: https://orcid.org/0000-0003-2768-0125

IRMAR, Université de Rennes 1
Email address: guillaume.poly@univ-rennes.fr