

Fast sampling from β -ensembles

GUILLAUME GAUTIER^{*,1,2} RÉMI BARDENET¹ and MICHAL VALKO^{2,3}

¹*Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRISTAL, 59651 Villeneuve d'Ascq, France.*

²*INRIA Lille - Nord Europe, 40, avenue Hally 59650, Villeneuve d'Ascq, France.*

³*DeepMind Paris, 14 Rue de Londres, 75009 Paris, France.*

E-mails: *guillaume.gga@gmail.com; remi.bardenet@gmail.com; valkom@deepmind.com.

We study sampling algorithms for β -ensembles with time complexity less than cubic in the cardinality of the ensemble. Following Dumitriu & Edelman (2002), we see the ensemble as the eigenvalues of a random tridiagonal matrix, namely a random Jacobi matrix. First, we provide a unifying and elementary treatment of the tridiagonal models associated to the three classical Hermite, Laguerre and Jacobi ensembles. For this purpose, we use simple changes of variables between successive reparametrizations of the coefficients defining the tridiagonal matrix. Second, we derive an approximate sampler for the simulation of β -ensembles, and illustrate how fast it can be for polynomial potentials. This method combines a Gibbs sampler on Jacobi matrices and the diagonalization of these matrices. In practice, even for large ensembles, only a few Gibbs passes suffice for the marginal distribution of the eigenvalues to fit the expected theoretical distribution. When the conditionals in the Gibbs sampler can be simulated exactly, the same fast empirical convergence is observed for the fluctuations of the largest eigenvalue. Our experimental results support a conjecture by Krishnapur et al. (2016), that the Gibbs chain on Jacobi matrices of size N mixes in $\mathcal{O}(\log N)$.

Keywords: β -ensembles, tridiagonal random matrices, orthogonal polynomials, Gibbs sampling.

MSC 2010 subject classifications: Primary: 60K35; secondary: 65C40, 60B20, 33C45.

1. Introduction

β -ensembles are probability distributions of the form

$$|\Delta(x_1, \dots, x_N)|^\beta Z^{-1} \prod_{n=1}^N e^{-V(x_n)} dx_n, \quad x_1, \dots, x_N \in \mathbb{R}, \quad (1.1)$$

where $\Delta(x_1, \dots, x_N) = \prod_{i < j} (x_j - x_i)$ is the Vandermonde determinant, $\beta > 0$ is akin to an inverse temperature in statistical physics, and $V : \mathbb{R} \rightarrow \mathbb{R}$ is called the *potential*. Loosely speaking, one can think of (1.1) as representing the position of N particles living on the real line, confined by the potential V , and repelling each other through the Vandermonde determinant. As this interpretation suggests, β -ensembles arise as models in statistical physics (Forrester, 2010, Chapters 1 to 3). They are also famous as the distribution of the eigenvalues of some of the classical models of random matrices. The particular values $\beta \in \{1, 2, 4\}$ respectively appear when considering specific random

matrices with real, complex, or quaternionic Gaussian entries; see e.g., [Forrester \(2010\)](#) again or [Anderson, Guionnet, and Zeitouni \(2009, Chapter 4\)](#).

The case $\beta = 2$ is of particular interest, since the distribution of $\{x_1, \dots, x_N\}$ then becomes a particular determinantal point process (DPP), called an *orthogonal polynomial ensemble* (OPE, [Konstantopoulos, 2005](#)). Originally introduced as models in fermionic optics by [Macchi \(1975\)](#), DPPs are comparatively easier to analyze than other repulsive distributions. Moreover, there exists a generic algorithm to sample from DPPs ([Hough et al., 2006](#)). Together with their analytic tractability, the existence of sampling algorithms has sparked the study of Monte Carlo integration using an OPE for the quadrature nodes ([Bardenet and Hardy, 2019](#); [Gautier, Bardenet, and Valko, 2019b](#); [Belhadji, Bardenet, and Chainais, 2019](#)).

Besides Monte Carlo integration, numerical procedures to generate samples from β -ensembles are also needed to establish conjectures in statistical physics or random matrix theory. For instance, using a tailored version of the generic DPP sampler of [Hough et al. \(2006\)](#), [Olshanski et al. \(2015\)](#) explore so-called *universality properties* in random matrix theory, and make conjectures on the law of $\max x_i$ when $\beta = 2$ and V is a polynomial of degree 4. [Chafaï & Ferré \(2018\)](#) rather use Hamiltonian Monte Carlo to approximately sample from various *Coulomb gases*, including (1.1) with $\beta = 2$ and $V(x) = x^4/4$, and investigate their limiting features when $N \rightarrow \infty$. From a different perspective, [Lieb & Mattis \(2013\)](#) view (1.1) as the equilibrium distribution for the Dyson Brownian motion associated to the potential V . When $\beta = 2$, they generate approximate samples by discretizing the corresponding stochastic differential equation.

Algorithms to sample from β -ensembles come in three different guises, which we describe in increasing order of complexity. First, when $\beta > 0$ and V is the negative logarithm of a Gaussian, gamma, or beta pdf, we speak of the Hermite, Laguerre, and Jacobi β -ensemble, respectively. [Dumitriu & Edelman \(2002\)](#) showed that the Hermite and Laguerre β -ensembles can be characterized as the eigenvalue distribution of a random tridiagonal matrix with easy-to-sample independent entries. This gives a $\mathcal{O}(N^2)$ sampling algorithm. [Dumitriu & Edelman \(2002\)](#) expected the same to hold for the Jacobi β -ensemble, which was later proved by [Killip & Nenciu \(2004\)](#).

Second, when $\beta = 2$, the generic projection DPP sampler of [Hough et al. \(2006\)](#) applies. That there actually exists an exact sampler is maybe surprising, and it is a particular feature of DPPs among interacting particle systems. The procedure remains costly, though. It has at least a cubic cost in N , with the total cost further depending on rejection sampling subroutines, the cost of which is case-dependent and has been left uninvestigated. Additionally, it is required in this procedure to numerically evaluate the first N orthonormal polynomials $p_k, k = 0, \dots, N - 1$ with respect to $e^{-V(x)}dx$. This is traditionally done using their recurrence relation

$$\sqrt{b_{k-1}}p_{k-1}(x) + a_k p_k(x) + \sqrt{b_k}p_{k+1}(x) = x p_k(x), \quad (1.2)$$

see e.g., [Gautschi \(2004\)](#). In the Hermite, Laguerre, and Jacobi case, the recurrence coefficients a_k, b_k are known, but as we just saw, these three cases are already covered by a computationally more efficient tridiagonal matrix model. When the coefficients in (1.2)

are not known, one can either rely on the Stieltjes algorithm (Gautschi, 2004, Section 2.2) or numerically solve a Riemann-Hilbert problem (Olver, 2011). The latter is theoretically only an $\mathcal{O}(N)$ overcost.

A third algorithm is Markov chain Monte Carlo (MCMC, see e.g., Rokhlin & Caflisch, 2004), which is in principle valid for any $\beta > 0$ and any V that gives a well-defined distribution in (1.1). MCMC only requires to evaluate the pdf in (1.1) pointwise and up to a constant, but it only delivers approximate samples of (1.1), in the sense that it outputs a sample from a Markov chain with (1.1) as its limiting distribution. The issue is that the performance of MCMC samplers – the mixing time of the Markov chain – deteriorates when $N \gg 1$, which is typically the regime of interest for conjectures in random matrix theory or statistical physics. Hybrid Monte Carlo (HMC, Dumitriu et al., 1987; Neal, 2011) is an MCMC sampler that has demonstrated good mixing in high-dimensional problems, provided one can evaluate the gradient of the pdf in (1.1). For β -ensembles with $\beta = 2$, Chafaï & Ferré (2018) provide empirical evidence that the output of HMC successfully reproduces known limiting features of the large N regime, and they raise new conjectures. The main limitation of this approach is the large number of MCMC iterations required by HMC: Chafaï & Ferré (2018) require at least 10^4 iterations and are restricted to $N \leq 50$.

In this paper, we further investigate fast samplers of β -ensembles. Our contributions are twofold. First we gather existing tools from different communities to give an elementary proof of the tridiagonal models for the Hermite, Laguerre, and Jacobi β -ensembles. This proof crucially relies on successive reparametrizations of the recurrence coefficients in (1.2) and unifies the treatment of tridiagonal models for the three classical β -ensembles, pioneered with two different methods by Dumitriu & Edelman (2002) and Killip & Nenciu (2004). We take no credit for the originality of the proof: the credit should go – among others cited below – to Dette & Nagel (2012), who studied distributions on the space of moments, and recognized these three β -ensembles as corresponding to natural distributions over moments. We rather take credit for a stand-alone and elementary version of this unifying proof, using only basic facts on orthogonal polynomials and linear algebra.

Our second contribution is an MCMC sampler that applies to polynomial potentials. For V of degree at most 6, we give experimental evidence that the resulting Markov chain mixes extremely fast, which confirms an intuition of Krishnapur et al. (2016, Section 2). On a variety of potentials, we demonstrate that our simple Gibbs Markov kernel yields a much cheaper (although approximate) sampler than the exact procedure of Hough et al. (2006, for $\beta = 2$). Importantly, our Markov kernel outperforms the HMC approach of Chafaï & Ferré (2018) in the particular case of β -ensembles. To give an idea, we are able to reproduce known features of (1.1) for values of N in the hundreds, using only a few Gibbs sweeps, totaling a few seconds on a modern laptop: it takes roughly 10s for $N = 200$ points and less than a minute for $N = 1000$ points. That such a basic Gibbs kernel can outperform HMC may seem surprising. The key is that we exploit the structure of β -ensembles by defining a Markov chain on the recurrence coefficients of orthogonal polynomials. These recurrence coefficients are defined similarly to (1.2), but this time using the orthogonal polynomials with respect to a random discrete measure, the support of which is the β -ensemble. Intuitively, in that new parametrization, the in-

teraction between variables is short-range compared to the interaction among particles in (1.1), and Gibbs sampling thus becomes easier. In this sense, our MCMC kernel extends the tridiagonal models of the three classical β -ensembles. Finally, we note that all experiments can be reproduced using our DPPy toolbox (Gautier, Bardenet, and Valko, 2019a, <https://github.com/guilgautier/DPPy>), which features all samplers described here.

The rest of the paper is organized as follows. In Section 2, we survey existing results on tridiagonal models for β -ensembles. Known exact sampling results actually take the form of diagonalizing random Jacobi matrices, that is, tridiagonal matrices whose coefficients are the recurrence coefficients of a sequence of orthogonal polynomials. We introduce the necessary background on orthogonal polynomials in Section 3. In Section 4, we perform the change of variables between the points of a β -ensemble augmented with weights and the entries of a Jacobi matrix. In Section 5, we give an elementary proof of the known results on tridiagonal models. Finally, in Section 6, we demonstrate a simple MCMC scheme based on a Gibbs kernel, to sample Jacobi matrices corresponding to β -ensembles with polynomial potentials.

2. Classical β -ensembles and their tridiagonal models

The Hermite, Laguerre and Jacobi β -ensembles were originally defined for $\beta \in \{1, 2, 4\}$, as the eigenvalue distribution of some random full matrices; see e.g., Anderson et al. (2009). The latter matrices are symmetrizations of matrices filled with i.i.d. real, complex, or quaternionic Gaussian variables when β is respectively 1, 2, and 4. In this section, we recall the seminal results of Dumitriu & Edelman (2002) and Killip & Nenciu (2004) regarding the construction of real-symmetric tridiagonal random matrices, whose eigenvalues follow the classical Hermite, Laguerre and Jacobi β -ensembles. These results actually allow any $\beta \in (0, +\infty)$, and can be interpreted as samplers with $\mathcal{O}(N^2)$ time complexity, by simply diagonalizing the proposed tridiagonal matrices.

We note $\mathbf{a} \triangleq (a_1, \dots, a_N) \in \mathbb{R}^N$, $\mathbf{b} \triangleq (b_1, \dots, b_{N-1}) \in (0, +\infty)^{N-1}$, and define the tridiagonal matrix

$$J_{\mathbf{a}, \mathbf{b}} \triangleq \begin{bmatrix} a_1 & \sqrt{b_1} & & (0) \\ \sqrt{b_1} & a_2 & \ddots & \\ \ddots & \ddots & \ddots & \sqrt{b_{N-1}} \\ (0) & \sqrt{b_{N-1}} & & a_N \end{bmatrix}. \quad (2.1)$$

Such a matrix is called a *Jacobi matrix*. As we will see in Section 3, Jacobi matrices naturally arise in the study of orthogonal polynomials.

To build the random tridiagonal matrix model for the Hermite β -ensemble, Dumitriu & Edelman (2002) started from the original random full matrix model defining the Hermite ensemble with $\beta = 1$. More specifically, they considered the symmetric part of a random matrix filled with i.i.d. unit Gaussians, and applied Householder transformations to reduce it to tridiagonal form, as in, e.g., Golub & Van Loan (2013, Section 5.4.8).

Theorem 2.1 (Dumitriu & Edelman, 2002, II C, for $\mu = 0$ and $\sigma = 1$). *The Hermite β -ensemble, defined as (1.1) with potential $V(x) = \frac{1}{2\sigma^2}(x - \mu)^2$, corresponds to the eigenvalue distribution of the tridiagonal matrix $J_{\mathbf{a}, \mathbf{b}}$ in (2.1), with entries drawn independently as*

$$a_n \sim \mathcal{N}(\mu, \sigma^2), \quad \text{and} \quad b_n \sim \Gamma\left(\frac{\beta}{2}(N - n), \sigma^2\right). \quad (2.2)$$

For the Laguerre β -ensemble, Dumitriu & Edelman (2002) used the same linear algebra techniques starting from the original full matrix model defining the Laguerre β -ensemble for $\beta = 1$. The latter corresponds to the eigenvalue distribution of the covariance matrix XX^\top of i.i.d. $\mathcal{N}(0, I)$ vectors. More specifically, they reduced the matrix X to bidiagonal form, see, e.g., Golub & Van Loan (2013, Section 8.3.1).

Theorem 2.2 (Dumitriu & Edelman, 2002, III B, for $k = \frac{\beta}{2}(M - N + 1)$ and $\theta = 2$). *The Laguerre β -ensemble, defined as (1.1) with potential $V(x) = -(k - 1)\log(x) + \frac{x}{\theta}$, corresponds to the eigenvalue distribution of the tridiagonal matrix $J_{\mathbf{a}, \mathbf{b}}$ in (2.1) parametrized by*

$$\begin{aligned} a_1 &= \xi_1, \quad \text{and} \quad a_n = \xi_{2n-2} + \xi_{2n-1}, \quad \text{for } 2 \leq n \leq N, \quad \text{and} \\ b_n &= \xi_{2n-1}\xi_{2n}, \quad \text{for } 1 \leq n \leq N - 1, \end{aligned} \quad (2.3)$$

with independent coefficients

$$\xi_{2n-1} \sim \Gamma\left(\frac{\beta}{2}(N - n) + k, \theta\right), \quad \text{and} \quad \xi_{2n} \sim \Gamma\left(\frac{\beta}{2}(N - n), \theta\right). \quad (2.4)$$

Dumitriu & Edelman (2002) left the construction of a tridiagonal model for the Jacobi β -ensemble as an open problem. Killip & Nenciu (2004) found such a model as a byproduct of their study the Circular β -ensemble. The latter ensemble is originally defined, for $\beta \in \{1, 2, 4\}$, as the eigenvalue distribution of orthogonal, unitary and symplectic matrices drawn uniformly at random from the corresponding Haar measures. First, Killip & Nenciu (2004) applied Householder transformations to reduce to quidiagonal form a unitary matrix drawn uniformly at random. Second, they projected the resulting eigenvalues onto the real line to obtain the tridiagonal model for the Jacobi β -ensemble.

Theorem 2.3 (Killip & Nenciu, 2004, Theorem 2). *The Jacobi β -ensemble, defined as (1.1) with potential $V(x) = -[(a - 1)\log(x) + (b - 1)\log(1 - x)]$, corresponds to the eigenvalue distribution of the tridiagonal matrix $J_{\mathbf{a}, \mathbf{b}}$ in (2.1) parametrized by*

$$\begin{aligned} a_1 &= c_1, & a_n &= (1 - c_{2n-3})c_{2n-2} + (1 - c_{2n-2})c_{2n-1}, \quad \text{for } 2 \leq n \leq N, \\ b_1 &= c_1(1 - c_1)c_2, & b_n &= (1 - c_{2n-2})c_{2n-1}(1 - c_{2n-1})c_{2n}, \quad \text{for } 2 \leq n \leq N - 1, \end{aligned} \quad (2.5)$$

with independent coefficients

$$\begin{aligned} c_{2n-1} &\sim \text{Beta}\left(\frac{\beta}{2}(N - n) + a, \frac{\beta}{2}(N - n) + b\right), \quad \text{and} \\ c_{2n} &\sim \text{Beta}\left(\frac{\beta}{2}(N - n), \frac{\beta}{2}(N - n - 1) + a + b\right). \end{aligned} \quad (2.6)$$

Observe how the stars align for these three special β -ensembles: Hermite, Laguerre, and Jacobi. The coefficients in successive parameterizations of the Jacobi matrix $J_{\mathbf{a},\mathbf{b}}$ are independent with easy-to-sample distributions. From a practical point of view, for any $\beta > 0$, the computation of the eigenvalues of these random real-symmetric tridiagonal matrices can be seen as a $\mathcal{O}(N^2)$ sampler for each of the model; see [Coakley & Rokhlin \(2013\)](#) for practical approaches to diagonalizing such matrices that can even run in quasi-linear time.

Studying distributions over the space of moments, [Dette & Nagel \(2012\)](#) elegantly derived the three classical tridiagonal models as the supports of random atomic measures corresponding to natural moment distributions. On our side, we provide a unified treatment of these three classical models using a more pedestrian, sampling-motivated approach. To do this, we consider an atomic measure $\mu = \sum_{n=1}^N \omega_n \delta_{x_n}$, whose support points are distributed as a target β -ensemble, and take the Jacobi matrix $J_{\mathbf{a},\mathbf{b}}$ in [\(2.1\)](#) with coefficients the recurrence coefficients [\(1.2\)](#) of the orthonormal polynomials w.r.t. μ . We shall see in Section 3 that the recurrence coefficients are a suitable reparametrization of the atomic measure μ . In particular, the support of μ actually coincides with the eigenvalues of $J_{\mathbf{a},\mathbf{b}}$, so that a tridiagonal model for the support of μ follows from knowing how to randomize $J_{\mathbf{a},\mathbf{b}}$.

The first step of our proof will be to rederive Theorem [2.4](#), which allows changing variables from the nodes and weights of an atomic measure μ to the recurrence coefficients defining the Jacobi matrix $J_{\mathbf{a},\mathbf{b}}$. Note that the specific choice of distribution on the weights is simply of mathematical convenience.

Theorem 2.4 ([Krishnapur et al., 2016](#), Proposition 2). *Consider a random atomic measure $\mu = \sum_{n=1}^N \omega_n \delta_{x_n}$, with nodes and weights independently distributed according to a β -ensemble with potential V [\(1.1\)](#) and a Dirichlet $\text{Dir}(\beta/2)$, respectively. Otherly put, the joint distribution of $(x_1, \dots, x_N, w_1, \dots, w_N)$ is proportional to*

$$|\Delta(x_1, \dots, x_N)|^\beta e^{-\sum_{n=1}^N V(x_n)} dx_{1:N} \prod_{n=1}^N w_n^{\frac{\beta}{2}-1} \mathbb{1}_{w_n \geq 0} \mathbb{1}_{\sum_{n=1}^N w_n = 1} dw_{1:N-1}. \quad (2.7)$$

Then, the recurrence coefficients $(a_1, \dots, a_N, b_1, \dots, b_{N-1})$ of μ have joint distribution proportional to

$$\prod_{n=1}^{N-1} b_n^{\frac{\beta}{2}(N-n)-1} e^{-\text{Tr}[V(J_{\mathbf{a},\mathbf{b}})]} da_{1:N} db_{1:N-1}. \quad (2.8)$$

In Section 4, we first re-prove that the change of variables underlying Theorem [2.4](#) is valid. Then, in Section 5, we obtain the three classical tridiagonal models of Theorems [2.1](#), [2.2](#), and [2.3](#) as instances of this result, using further smart-but-simple changes of variables. Before delving into the proof, we first survey how Jacobi matrices naturally appear in the theory of orthogonal polynomials.

3. Atomic measures, moments and Jacobi matrices

Throughout this section, we let $\mu = \sum_{n=1}^N w_n \delta_{x_n}$ be a discrete probability measure on \mathbb{R} with N distinct atoms x_1, \dots, x_N and positive weights w_1, \dots, w_N . We further denote its moments by

$$m_k \triangleq \sum_{n=1}^N w_n x_n^k, \quad k \geq 0.$$

3.1. Orthogonal polynomials and Jacobi matrices

This section closely follows [Simon \(2011, Section 1.3\)](#), to which we refer for details. Applying the Gram-Schmidt procedure in $L^2(\mu)$ to the monomials $(x \mapsto x^k)_{k=0}^{N-1}$ yields monic polynomials $(P_k)_{k=0}^{N-1}$ with $\deg P_k = k$ and

$$\langle P_k, P_\ell \rangle_\mu \triangleq \sum_{n=1}^N w_n P_k(x_n) P_\ell(x_n) = 0, \quad k \neq \ell. \quad (3.1)$$

These polynomials are called the *monic orthogonal polynomials* (monic OPs, in short) with respect to μ . We *define* the N -th monic OP as

$$P_N(x) = \prod_{n=1}^N (x - x_n).$$

Since $\|P_N\|_\mu \triangleq \langle P_N, P_N \rangle_\mu = 0$, P_N is the zero vector of $L^2(\mu)$: it is orthogonal to all P_k with $k \leq N-1$.

Furthermore, for any $n < N$, since $\langle x P_n, P_k \rangle_\mu = \langle P_n, x P_k \rangle_\mu = 0$ for $k < n-1$, the polynomial $x P_n$ can be uniquely expressed using only P_{n-1} , P_n and P_{n+1} . This is usually phrased as follows. The monic OPs satisfy a three-term recurrence relation involving two sequences of *recurrence coefficients*, namely

$$\begin{aligned} P_{-1} &\equiv 0, P_0 \equiv 1 \text{ and} \\ x P_n(x) &= b_n P_{n-1}(x) + a_{n+1} P_n(x) + P_{n+1}(x), \quad \forall 0 \leq n < N, \end{aligned} \quad (3.2)$$

where $\mathbf{a} = a_{1:N} = (a_n) \in \mathbb{R}^N$, and $\mathbf{b} = b_{1:N-1} = (b_n) \in (0, +\infty)^{N-1}$. These relations can be written in matrix form as

$$\begin{bmatrix} a_1 & 1 & & (0) \\ b_1 & a_2 & \ddots & \\ & \ddots & \ddots & 1 \\ (0) & & b_{N-1} & a_N \end{bmatrix} \begin{bmatrix} P_0(x) \\ \vdots \\ P_{N-2}(x) \\ P_{N-1}(x) \end{bmatrix} = x \begin{bmatrix} P_0(x) \\ \vdots \\ P_{N-2}(x) \\ P_{N-1}(x) \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 0 \\ P_N(x) \end{bmatrix}. \quad (3.3)$$

From (3.3), it is clear that the roots of P_N are also eigenvalues of the tridiagonal matrix $T_{\mathbf{a}, \mathbf{b}}$ appearing on the left-hand side. Roots and eigenvalues actually coincide since P_N has N distinct roots by definition.

A lot more can be said on the links between OPs and their recurrence coefficients. For instance, Proposition 3.1 will be of use later on.

Proposition 3.1. *The squared norms of the monic polynomials $(P_n)_{n=0}^{N-1}$ can be expressed as*

$$\|P_0\|_\mu^2 = 1, \quad \|P_k\|_\mu^2 = \prod_{n=1}^k b_n, \quad \forall 1 \leq k \leq N-1. \quad (3.4)$$

Proof. For $k = 0$, $\|P_0\|_\mu^2 = \sum_{n=1}^N w_n = 1$. Then, for any $1 \leq k \leq N-1$,

$$\begin{aligned} \langle (3.2), P_{k-1} \rangle_\mu &\iff \langle xP_k, P_{k-1} \rangle_\mu = \langle b_k P_{k-1}, P_{k-1} \rangle_\mu \\ &\iff \langle P_k, xP_{k-1} \rangle_\mu = b_k \langle P_{k-1}, P_{k-1} \rangle_\mu \\ &\iff \langle P_k, x^n \rangle_\mu = b_k \|P_{k-1}\|_\mu^2 \\ &\iff \|P_k\|_\mu^2 = b_k \|P_{k-1}\|_\mu^2, \end{aligned}$$

and a simple recursion provides $\|P_k\|_\mu^2 = \prod_{n=1}^k b_n > 0$. \square

Denoting by $D = \text{diag}(\|P_0\|, \dots, \|P_{N-1}\|)$, Proposition 3.1 yields $J_{\mathbf{a}, \mathbf{b}} = D^{-1} T_{\mathbf{a}, \mathbf{b}} D$, where we recall that the Jacobi matrix $J_{\mathbf{a}, \mathbf{b}}$ was defined in (2.1). This yields the following proposition.

Proposition 3.2. *The atoms of μ , coincide with the eigenvalues of $J_{\mathbf{a}, \mathbf{b}}$, where the coefficients of the matrix are taken to be the recurrence coefficients of the monic OPs with respect to μ .*

Proposition 3.2 already gives a tentative $\mathcal{O}(N^2)$ sampling algorithm for β -ensembles: find a distribution over Jacobi matrices such that the eigenvalues form the desired β -ensemble. This is precisely what the tridiagonal models of Dumitriu & Edelman (2002) do; see Theorem 2.1. To give a complete elementary proof, we need to perform a change of variables from the atoms and weights of μ to the recurrence coefficients. The rest of this section introduces the tools needed for this change of variables, which is then performed in Section 4.

So far, we have explained how to obtain a Jacobi matrix from an atomic measure with finite support. The reverse construction is also possible and elementary. This is called Favard's theorem for atomic measures with finite support. To save space and because our proof would be a simple copy of Simon's book, we only give a reference. We have used the same notation as Simon throughout this section, for ease of reference.

Theorem 3.1 (Simon, 2011, Theorem 1.3.3). *Let*

$$\mathbb{R}_>^N \triangleq \{x_1, \dots, x_N \in \mathbb{R} \mid x_1 > \dots > x_N\} \quad \text{and} \quad S_N \triangleq \left\{ \omega_1, \dots, \omega_{N-1} > 0 \mid \sum_{n=1}^{N-1} \omega_n < 1 \right\}. \quad (3.5)$$

Favard's map

$$\begin{aligned} \psi : \quad \mathbb{R}_>^N \times S_N &\longrightarrow \mathbb{R}^N \times (0, +\infty)^{N-1} \\ (x_{1:N}, w_{1:N-1}) &\longmapsto (a_{1:N}, b_{1:N-1}) \end{aligned} \quad (3.6)$$

linking the nodes and weights of $\mu = \sum_{n=1}^N w_n \delta_{x_n}$ with the entries of the corresponding Jacobi matrix $J_{\mathbf{a}, \mathbf{b}}$ defined in (2.1), is one-to-one and onto.

Note that whenever $w_{1:N-1} \in S_N$, we always set $w_N = 1 - \sum_{n=1}^{N-1} \omega_n$, so that μ is a probability measure. As a side remark, the weights $w_{1:N}$ of μ can also be expressed using evaluations of the monic OPs on the support of μ (Simon, 2011, Proposition 1.3.1): for all $n = 1, \dots, N$,

$$w_n = \frac{1}{K_N(x_n, x_n)}, \quad \text{with } K_N(x, y) = \sum_{k=0}^{N-1} \frac{P_k(x)P_k(y)}{\|P_k\|_\mu^2}. \quad (3.7)$$

These weights are reminiscent of Gaussian quadrature (Gautschi, 2004, Section 1.4.2), where the OPs are usually w.r.t. a non-atomic measure.

3.2. Orthogonal polynomials and moments

We know from Theorem 3.1 that the change of variables ψ is a bijection. In order to prove that ϕ is a C^1 -diffeomorphism and compute its Jacobian in Section 4, we pause to introduce an intermediate parametrization through moments. Intuitively, the moments are responsible for the Vandermonde determinant in (1.1).

The monic orthogonal polynomials $(P_n)_{n=0}^N$ w.r.t. μ can also be expressed in terms of the moments (m_k) of μ . First, define the following *moment matrices*, see, e.g., Dette & Studden (1997, Equation 1.4.3).

Definition 3.1. *Let*

$$\underline{H}_{2n} = [m_{i+j}]_{i,j=0}^n = \begin{bmatrix} m_0 & \cdots & m_n \\ \vdots & \ddots & \vdots \\ m_n & \cdots & m_{2n} \end{bmatrix} \quad (3.8)$$

$$\underline{H}_{2n+1} = [m_{i+j+1}]_{i,j=0}^n = \begin{bmatrix} m_1 & \cdots & m_{n+1} \\ \vdots & \ddots & \vdots \\ m_{n+1} & \cdots & m_{2n+1} \end{bmatrix} \quad (3.9)$$

$$\overline{H}_{2n+1} = [m_{i+j} - m_{i+j+1}]_{i,j=0}^n = \begin{bmatrix} m_0 - m_1 & \cdots & m_n - m_{n+1} \\ \vdots & \ddots & \vdots \\ m_n - m_{n+1} & \cdots & m_{2n} - m_{2n+1} \end{bmatrix}. \quad (3.10)$$

where H stands for Hankel matrix.

The determinant of the Vandermonde matrix

$$\Delta(x_1, \dots, x_n) \triangleq \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \\ \vdots & & \vdots \\ x_1^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}, \quad (3.11)$$

which appears in the definition of β -ensembles (1.1), comes out naturally when taking the determinant of moment matrices associated to discrete measures.

Lemma 3.1. *It holds that*

$$|\underline{H}_{2n-2}| \begin{cases} > 0, & \text{for any } 1 \leq n \leq N, \\ = |\Delta(x_1, \dots, x_N)|^2 \prod_{n=1}^N w_n, & \text{for } n = N, \\ = 0, & \text{for } n > N. \end{cases} \quad (3.12)$$

Moreover

$$|\underline{H}_{2N-1}| = |\underline{H}_{2N-2}| \prod_{n=1}^N x_n \quad \text{and} \quad |\overline{H}_{2N-1}| = |\underline{H}_{2N-2}| \prod_{n=1}^N (1 - x_n). \quad (3.13)$$

Proof. For any $1 \leq n \leq N$, the Cauchy-Binet formula yields

$$\begin{aligned} |\underline{H}_{2n-2}| &= \left| \sum_{k=1}^N w_k x_k^{i+j} \right|_{i,j=0}^{n-1} \\ &= \left| \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_N \\ \vdots & \vdots & \vdots \\ x_1^{n-1} & \cdots & x_N^{n-1} \end{bmatrix} \begin{bmatrix} w_1 & & & \\ & \ddots & & \\ & & w_N & \\ & & & \end{bmatrix} \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_N & \cdots & x_N^{n-1} \end{bmatrix} \right| \\ &= \sum_{\{i_1, \dots, i_n\} \subset [N]} |\Delta(x_{i_1}, \dots, x_{i_n})|^2 \prod_{k=1}^n w_{i_k} > 0. \end{aligned} \quad (3.14)$$

The particular case $n = N$ yields

$$|\underline{H}_{2N-2}| = |\Delta(x_1, \dots, x_N)|^2 \prod_{n=1}^N w_n.$$

In the same vein, the two other determinants are obtained starting from

$$|\underline{H}_{2N-1}| = \left| \sum_{n=1}^N w_n x_n^{i+j} x_n \right|_{i,j=0}^{N-1} \quad \text{and} \quad |\overline{H}_{2N-1}| = \left| \sum_{n=1}^N w_n x_n^{i+j} (1-x_n) \right|_{i,j=0}^{N-1}.$$

For $n > N$, (3.14) clearly shows that \underline{H}_{2n-2} is rank deficient. \square

Moment matrices also provide an alternative description of orthogonal polynomials.

Proposition 3.3. *The monic polynomials $(P_n)_{n=0}^N$ orthogonal with respect to $\mu = \sum_{n=1}^N w_n \delta_{x_n}$ admit the following expression*

$$P_0 = 1 \quad \text{and} \quad P_n(x) = \frac{1}{|\underline{H}_{2n-2}|} \begin{vmatrix} & \underline{H}_{2n-2} & 1 \\ & \vdots & \vdots \\ m_n & \cdots & m_{2n-1} & x^n \end{vmatrix}, \quad \forall 1 \leq n \leq N. \quad (3.15)$$

Besides,

$$\|P_0\|_\mu^2 = 1 \quad \text{and} \quad \|P_n\|_\mu^2 = \frac{|\underline{H}_{2n}|}{|\underline{H}_{2n-2}|}, \quad \forall 1 \leq n \leq N. \quad (3.16)$$

In particular, $P_N(x) = \prod_{n=1}^N (x - x_n)$ is the zero vector of $L^2(\mu)$.

Proof. The previous Lemma 3.1 validates the definition of $(P_n)_{n=0}^N$ as a sequence of monic polynomials with $\deg P_n = n$ since the denominator $|\underline{H}_{2n-2}| > 0$. They are also mutually orthogonal. To see this, let $1 \leq n \leq N$, then

$$\langle P_n, x^k \rangle_\mu = \frac{1}{|\underline{H}_{2n-2}|} \begin{vmatrix} m_0 & \cdots & m_{n-1} & m_k \\ \vdots & & \vdots & \vdots \\ m_n & \cdots & m_{2n-1} & m_{n+k} \end{vmatrix} = 0, \quad \forall k < n.$$

Moreover, $\forall 1 \leq n \leq N$,

$$\|P_n\|_\mu^2 = \langle P_n, P_n \rangle_\mu = \langle P_n, x^n \rangle_\mu = \frac{|\underline{H}_{2n}|}{|\underline{H}_{2n-2}|}.$$

Then, Lemma 3.1 yields $\|P_N\|_\mu^2 = \frac{|\underline{H}_{2N}|}{|\underline{H}_{2N-2}|} = 0$. Thus, the distinct support points of μ are zeros of P_N . But the latter is monic with $\deg P_N = N$, hence $P_N = \prod_{n=1}^N (x - x_n)$. \square

The next result further relates moment matrices and the recurrence coefficients.

Lemma 3.2. *The moment matrix \underline{H}_{2N-2} associated to $\mu = \sum_{n=1}^N w_n \delta_{x_n}$ has determinant*

$$|\underline{H}_{2N-2}| = |\Delta(x_1, \dots, x_N)|^2 \prod_{n=1}^N w_n = \prod_{n=1}^{N-1} b_n^{N-n}. \quad (3.17)$$

Proof. The first equality was established in Lemma 3.1. The second results from a simple recursion combining Equations 3.4 and 3.16. For any $1 \leq k \leq N - 1$

$$\|P_k\|_\mu^2 = \frac{|\underline{H}_{2k}|}{|\underline{H}_{2k-2}|} = \prod_{n=1}^k b_n \implies |\underline{H}_{2k}| = \prod_{\ell=1}^k \prod_{n=1}^\ell b_n = \prod_{n=1}^k b_n^{k+1-n}. \quad (3.18)$$

□

From the point of view of sampling a β -ensemble, Lemma 3.2 already hints what tridiagonal models can achieve: if we see the β -ensemble as the support of a random atomic measure, which is parametrized by its recurrence coefficients, then the complex interaction term that is the Vandermonde determinant in (1.1) gets replaced by a simple product of powers of b_n s. This intuition, formalized in Theorem 2.4, requires to make explicit the change of variables between the nodes and weights of μ and the recurrence coefficients of the corresponding Jacobi matrix $J_{\mathbf{a}, \mathbf{b}}$.

4. Making the change of variables

To compute the Jacobian of Favard's map $(x_{1:N}, \omega_{1:N}) \mapsto (a_{1:N}, b_{1:N-1})$, defined in Theorem 3.1, we first compute the Jacobian of the moment map $(x_{1:N}, \omega_{1:N}) \mapsto (m_{1:2N-1})$, and then use the lattice path construction of [Hardy \(2018\)](#) to express the Jacobian of $(m_{1:2N-1}) \mapsto (a_{1:N}, b_{1:N-1})$. We mention that the overall Jacobian has already been derived, in a more concise style, by [Forrester & Rains \(2006\)](#) and [Krishnapur et al. \(2016\)](#). Our contribution in this section is to give all details while remaining as elementary as possible. In particular, we only rely on Favard's theorem for atomic measures, and the proof of Theorem 2.4 boils down to checking that the changes of variables are C^1 -diffeomorphisms.

Let $\phi : \mathbb{R}_>^N \times S_N \rightarrow \mathbb{R}^{2N-1}$ map a set of N distinct atoms and $N - 1$ positive weights to their moments (m_k) . Let $\mathcal{M} \subset \mathbb{R}^{2N-1}$ be the image of ϕ .

Proposition 4.1 (From atomic measures to moments). $\mathcal{M} \subset \mathbb{R}^{2N-1}$ is open, ϕ is a C^1 -diffeomorphism from $\mathbb{R}_>^N \times S_N$ onto \mathcal{M} , and

$$\left| \frac{\partial m_{1:2N-1}}{\partial x_{1:N}, w_{1:N-1}} \right| = |\Delta(x_1, \dots, x_N)|^4 \prod_{n=1}^N w_n = \frac{|\underline{H}_{2N-2}|^2}{\prod_{n=1}^N w_n}, \quad (4.1)$$

where the Hankel matrix \underline{H}_{2N-2} is defined by (3.8).

Proof. Moments define monic OPs; see Proposition 3.15. By Favard's Theorem 3.1, monic OPs in turn define the atoms and weights of μ uniquely. Thus, ϕ is injective. Moreover $\mathbb{R}_>^N \times S_N \subset \mathbb{R}^{2N-1}$ is open, and ϕ is C^1 . By the classical inverse function theorem, see e.g., [Carlitz \(1971, Corollary 4.2.2\)](#), it is thus enough to show that the Jacobian of ϕ never vanishes.

The i -th moment of μ can be written in two forms

$$m_i = \sum_{j=1}^N w_j x_j^i = \sum_{j=1}^{N-1} w_j (x_j^i - x_N^i) + x_N^i, \quad (4.2)$$

so that

$$\frac{\partial m_i}{\partial x_j} = i w_j x_j^{i-1} \quad \text{and} \quad \frac{\partial m_i}{\partial w_j} = x_j^i - x_N^i. \quad (4.3)$$

Thus,

$$\begin{aligned} \left| \frac{\partial m_{1:2N-1}}{\partial x_{1:N}, w_{1:N-1}} \right| &= \left| \begin{bmatrix} \left[\frac{\partial m_i}{\partial x_j} \frac{\partial m_i}{\partial w_j} \right]_{j=1}^{N-1} & \left[\frac{\partial m_i}{\partial x_N} \right]_{i=1}^{2N-1} \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} [iw_j x_j^{i-1} \ x_j^i - x_N^i]_{j=1}^{N-1} & iw_N x_N^{i-1} \end{bmatrix}_{i=1}^{2N-1} \right| \\ &= \left| \begin{bmatrix} [ix_j^{i-1} \ x_j^i - x_N^i]_{j=1}^{N-1} & ix_N^{i-1} \end{bmatrix}_{i=1}^{2N-1} \right| \times \prod_{n=1}^N w_n \\ &= \left| \begin{bmatrix} [(i-1)x_j^{i-2} \ x_j^{i-1} - x_N^{i-1}]_{j=1}^{N-1} & (i-1)x_N^{i-2} \ x_N^{i-1} \end{bmatrix}_{i=1}^{2N} \right| \times \prod_{n=1}^N w_n \\ &= \left| \begin{bmatrix} (i-1)x_j^{i-2} \ x_j^{i-1} \end{bmatrix}_{i=1, j=1}^{2N, N} \right| \prod_{n=1}^N w_n. \end{aligned} \quad (4.4)$$

The last equality is obtained by adding the last column to all other even columns. The determinant in (4.4) is called a *confluent* Vandermonde determinant. Its value is given, e.g., by [Hausdorff & Givens \(1980\)](#), Corollary 1, with $\eta_i \equiv 2$)

$$\left| \begin{bmatrix} (i-1)x_j^{i-2} \ x_j^{i-1} \end{bmatrix}_{i=1, j=1}^{2N, N} \right| = \prod_{1 \leq i < j \leq N} (x_j - x_i)^{2 \times 2} = |\Delta(x_1, \dots, x_N)|^4.$$

In particular, (4.4) never vanishes on $\mathbb{R}_>^N \times S_N$. \square

Let us now consider the map

$$\rho : \mathcal{M} \rightarrow \mathbb{R}^N \times (0, +\infty)^{N-1}, \quad (4.5)$$

that takes moments $m_{1:2N-1}$ and returns the recurrence coefficients $(a_{1:N}, b_{1:N-1})$.

Proposition 4.2 (From recurrence coefficients to moments). ρ is a C^1 -diffeomorphism from \mathcal{M} onto $\mathbb{R}^N \times (0, +\infty)^{N-1}$, and

$$\left| \frac{\partial m_{1:2N-1}}{\partial a_{1:N}, b_{1:N-1}} \right| = \prod_{n=1}^{N-1} b_n^{2(N-n)-1} = \frac{|\underline{H}_{2N-2}|^2}{\prod_{n=1}^{N-1} b_n}, \quad (4.6)$$

where the Hankel matrix \underline{H}_{2N-2} is defined by (3.8).

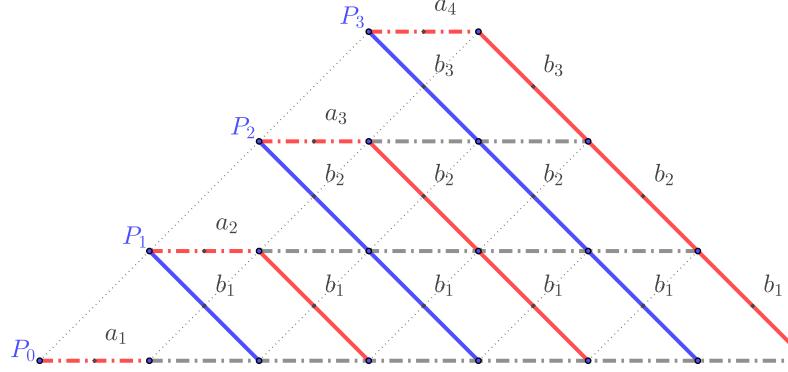
Proof. Using Theorem 3.1 and Proposition 4.1, $\rho = \psi \circ \phi^{-1}$, so that ρ is bijective. As in the proof of Proposition 4.1, we apply the inverse function theorem (Carlitz, 1971, Corollary 4.2.2), but this time to ρ^{-1} . We first note that $\mathbb{R}^N \times (0, +\infty)^{N-1} \subset \mathbb{R}^{2N-1}$ is open. It is thus enough to show that ρ^{-1} is C^1 and that its Jacobian never vanishes. To this end, we borrow an elegant lattice path representation of the recurrence relations for OPs from Hardy (2018, Equation 1.8). This allows us to express the successive moments as polynomials in the recurrence coefficients.

To provide intuition, we first compute the first few moments by hand, recursively applying the recurrence relation (3.2). It comes

$$\begin{aligned}
m_1 &= \langle xP_0, P_0 \rangle = 1 \cdot \langle P_1, P_0 \rangle + a_1 \cdot \langle P_0, P_0 \rangle + 0 = \textcolor{red}{a}_1, \\
m_2 &= \langle x^2 P_0, P_0 \rangle = 1 \cdot \langle xP_1, P_0 \rangle + a_1 \cdot \langle xP_0, P_0 \rangle + 0 \\
&= 1 \cdot (1 \cdot \langle P_2, P_0 \rangle + a_2 \cdot \langle P_1, P_0 \rangle + b_1 \langle P_0, P_0 \rangle) \\
&\quad + a_1 \cdot (1 \cdot \langle P_1, P_0 \rangle + a_1 \cdot \langle P_0, P_0 \rangle + 0) \\
&= \textcolor{blue}{1} \cdot \textcolor{blue}{b}_1 + a_1 \cdot a_1, \\
m_3 &= \langle x^3 P_0, P_0 \rangle = 1 \cdot \langle x^2 P_1, P_0 \rangle + a_1 \cdot \langle x^2 P_0, P_0 \rangle + 0 \\
&= 1 \cdot (1 \cdot \langle xP_2, P_0 \rangle + a_2 \cdot \langle xP_1, P_0 \rangle + b_1 \cdot \langle xP_0, P_0 \rangle) \\
&\quad + a_1 \cdot (1 \cdot \langle xP_1, P_0 \rangle + a_1 \cdot \langle xP_0, P_0 \rangle + 0) \\
&= 1 \cdot 1 \cdot (1 \cdot \langle P_3, P_0 \rangle + a_3 \cdot \langle P_2, P_0 \rangle + b_3 \cdot \langle P_1, P_0 \rangle) \\
&\quad + 1 \cdot a_2 \cdot (1 \cdot \langle P_2, P_0 \rangle + a_2 \cdot \langle P_1, P_0 \rangle + b_1 \cdot \langle P_0, P_0 \rangle) \\
&\quad + 1 \cdot b_1 \cdot (1 \cdot \langle P_1, P_0 \rangle + a_1 \cdot \langle P_0, P_0 \rangle + 0) \\
&\quad + a_1 \cdot 1 \cdot (1 \cdot \langle P_2, P_0 \rangle + a_2 \cdot \langle P_1, P_0 \rangle + b_1 \cdot \langle P_0, P_0 \rangle) \\
&\quad + a_1 \cdot a_1 \cdot (1 \cdot \langle P_1, P_0 \rangle + a_1 \cdot \langle P_0, P_0 \rangle + 0) \\
&= \textcolor{red}{1} \cdot \textcolor{red}{a}_2 \cdot \textcolor{blue}{b}_1 + a_1 \cdot 1 \cdot b_1 + 1 \cdot b_2 \cdot a_1 + a_1 \cdot a_1 \cdot a_1. \tag{4.7}
\end{aligned}$$

More generally, when computing $m_k = \langle x^k P_0, P_0 \rangle$, the recursive application of the recurrence relation (3.2) allows to decrease the power of x from k to 0 until each term in the development is proportional to the inner product of $P_0 = 1$ with another monic OP. The only nonzero such inner product is $\langle P_0, P_0 \rangle = 1$. Consequently, each nonzero term in the final development of m_k corresponds to a path of length at most k that leaves from the lower left corner of the graph in Figure 1(a) and ends up on the bottom row. In between, the path has to remain above the bottom row, and can only move North-East, East, or South-East. Each edge corresponds to picking one of the three terms in the recurrence relation (3.2). For example, the development of m_3 in (4.7) corresponds to three such paths, shown in green in Figure 1(a). The product of the coefficients along each path forms the resulting term in the development.

In the end, odd moments m_{2i-1} , resp. even moments m_{2i} , are the sum of the weights of the paths below the i -th red, respectively blue path, counting from the bottom left.



(a) The North-East, East, South-East edges associated to weights 1, a_n , b_n are respectively represented as dashed, dash-dotted and solid lines. Note that on each dashed and dash-dotted line, the weight is constant.

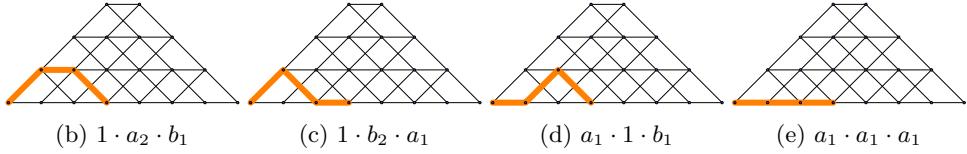


Figure 1: The lattice path of [Hardy \(2018\)](#) used to compute $m_n = \langle x^n P_0, P_0 \rangle$ is displayed in (a). The paths used for the computation of m_3 (4.7) are highlighted in (b)-(e) with the corresponding weight as caption.

More precisely,

$$m_{2i-1} = a_i \prod_{k=1}^{i-1} b_k + f_1(a_{1:i-1}, b_{1:i-2}) \quad \text{and} \quad m_{2i} = \prod_{k=1}^i b_k + f_2(a_{1:i}, b_{1:i-1}). \quad (4.8)$$

Thus, the Jacobian is the determinant of a triangular matrix

$$\left| \frac{\partial m_{1:2N-1}}{\partial a_{1:N}, b_{1:N-1}} \right| = \left| \begin{bmatrix} \frac{\partial m_{2i-1}}{\partial a_j} & \frac{\partial m_{2i-1}}{\partial b_j} \\ \frac{\partial m_{2i}}{\partial a_j} & \frac{\partial m_{2i}}{\partial b_j} \\ \frac{\partial m_{2N-1}}{\partial a_j} & \frac{\partial m_{2N-1}}{\partial b_j} \end{bmatrix} \begin{bmatrix} \frac{\partial m_{2i-1}}{\partial a_N} \\ \frac{\partial m_{2i}}{\partial a_N} \\ \frac{\partial m_{2N-1}}{\partial a_N} \end{bmatrix} \right|_{i,j=1}^{N-1} = \prod_{i=1}^N \frac{\partial m_{2i-1}}{\partial a_i} \prod_{i=1}^{N-1} \frac{\partial m_{2i}}{\partial b_i}.$$

The formulation (4.8) yields

$$\frac{\partial m_{2i-1}}{\partial a_i} = \prod_{k=1}^{i-1} b_k \quad \text{and} \quad \frac{\partial m_{2i}}{\partial b_i} = \prod_{k=1}^{i-1} b_k.$$

Finally, we obtain

$$\left| \frac{\partial m_{1:2N-1}}{\partial a_{1:N}, b_{1:N-1}} \right| = \prod_{i=1}^N \prod_{k=1}^{i-1} b_k \prod_{i=1}^{N-1} \prod_{k=1}^{i-1} b_k = \frac{\left[\prod_{i=1}^N \prod_{k=1}^{i-1} b_k \right]^2}{\prod_{k=1}^{N-1} b_k} = \frac{\left[\prod_{n=1}^{N-1} b_n^{N-n} \right]^2}{\prod_{n=1}^{N-1} b_n},$$

which does not vanish since all b_n s are positive by construction. Finally, the last equality in (4.6) follows from Lemma 3.2. \square

Propositions 4.1 and 4.2 now allow us to conclude that Favard's map $\psi = \rho \circ \phi$ (cf. Theorem 3.1) is a C^1 -diffeomorphism, and compute its Jacobian.

Proposition 4.3. *Favard's map ψ is a C^1 -diffeomorphism from $\mathbb{R}_>^N \times S_N$ onto $\mathbb{R}^N \times (0, +\infty)^{N-1}$, and*

$$\left| \frac{\partial x_{1:N}, w_{1:N-1}}{\partial a_{1:N}, b_{1:N-1}} \right| = \prod_{n=1}^{N-1} b_n^{-1} \prod_{n=1}^N w_n. \quad (4.9)$$

We now have all the ingredients to give an explicit proof of Theorem 2.4, of which the three classical tridiagonal models of Section 5 will be seen to be corollaries.

Proof of Theorem 2.4. For simplicity we drop the indicator functions and rewrite the density of the nodes and weights as

$$(2.7) = \left(|\Delta(x_1, \dots, x_N)|^2 \prod_{n=1}^N w_n \right)^{\frac{\beta}{2}} e^{-\text{Tr}[V(\text{diag}(x_1, \dots, x_N))]} \prod_{n=1}^N w_n^{-1} dx_{1:N} dw_{1:N-1}$$

Combining Lemma 3.2, and the fact that x_1, \dots, x_N are the eigenvalues of $J_{\mathbf{a}, \mathbf{b}}$, the change of variables provided by Proposition 4.3 yields

$$\begin{aligned} (2.7) &= \left(\prod_{n=1}^{N-1} b_n^{N-n} \right)^{\frac{\beta}{2}} e^{-\text{Tr}[V(J_{\mathbf{a}, \mathbf{b}})]} \prod_{n=1}^N w_n^{-1} \left| \frac{\partial x_{1:N}, w_{1:N-1}}{\partial a_{1:N}, b_{1:N-1}} \right| da_{1:N} db_{1:N-1} \\ &= \prod_{n=1}^{N-1} b_n^{\frac{\beta}{2}(N-n)-1} e^{-\text{Tr}[V(J_{\mathbf{a}, \mathbf{b}})]} da_{1:N} db_{1:N-1}, \end{aligned}$$

where the last equality follows from (4.9). \square

5. Proving the three classical tridiagonal models

Theorem 2.4 gives the distribution over recurrence coefficients, from which one has to sample, in order for the atoms of the corresponding atomic measure to follow a given β -ensemble. When the potential of the β -ensemble is taken among three particular forms, the recurrence coefficients turn out to be independent with simple distributions. In particular, the recurrence coefficients are much simpler to sample than the complex joint distribution (1.1) of the atoms.

5.1. The $H\beta E$ and its tridiagonal model

The tridiagonal model associated to the Hermite β -ensemble, cf. Theorem 2.1, follows from a direct application of Theorem 2.4 and the following immediate lemma.

Lemma 5.1. *Let $J_{\mathbf{a},\mathbf{b}}$ be a Jacobi matrix as defined by (2.1), with eigenvalues x_1, \dots, x_N . It holds that*

$$\sum_{n=1}^N x_n = \text{Tr } J_{\mathbf{a},\mathbf{b}} = \sum_{n=1}^N a_n \quad \text{and} \quad \sum_{n=1}^N x_n^2 = \text{Tr } J_{\mathbf{a},\mathbf{b}}^2 = \sum_{n=1}^N a_n^2 + 2 \sum_{n=1}^{N-1} b_n. \quad (5.1)$$

Proof of Theorem 2.1. Starting from Theorem 2.4 it remains to express the term $\text{Tr } V(J_{\mathbf{a},\mathbf{b}})$, where $V(x) = \frac{(x-\mu)^2}{2\sigma^2} = \frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2)$. To this end, Lemma 5.1 yields

$$\text{Tr } V(J_{\mathbf{a},\mathbf{b}}) = \frac{1}{2\sigma^2} [\text{Tr } J_{\mathbf{a},\mathbf{b}}^2 - 2\mu \text{Tr } J_{\mathbf{a},\mathbf{b}} + N\mu^2] = \frac{1}{2\sigma^2} \sum_{n=1}^N (a_n - \mu)^2 + \frac{1}{\sigma^2} \sum_{n=1}^{N-1} b_n.$$

Finally, we can plug this expression back into (2.7) to see that the entries of $J_{\mathbf{a},\mathbf{b}}$ are independently distributed, with joint distribution proportional to

$$\prod_{n=1}^{N-1} b_n^{\frac{\beta}{2}(N-n)-1} e^{-\frac{1}{\sigma^2}b_n} db_n \prod_{n=1}^N e^{-\frac{1}{2\sigma^2}(a_n-\mu)^2} da_n. \quad (5.2)$$

□

Note that when μ is still supported on \mathbb{R} , but the potential is a more general polynomial, the recurrence parameters are no longer independent, but the interaction remains short range. This is what we later exploit in Section 6, where we derive a fast approximate sampler for various β -ensembles with polynomial potentials.

5.2. The $L\beta E$ and its tridiagonal model

When the target β -ensemble is supported on $(0, +\infty)$, there is a natural reparametrization of the recurrence coefficients of μ , which allows to express other quantities than those in Lemma 5.1. This leads to the tridiagonal model of Theorem 2.2 for the Laguerre β -ensemble.

The reparametrization, denoted by $\xi_1, \dots, \xi_{2N-1} > 0$, arose in the work of [Stieltjes \(1894\)](#) on continued fractions; see also [Chihara \(1978, 1971\)](#), Equation 9.12 in Corollary of Theorem 9.1; Equation 2) in the context of three-term recurrence relations. To introduce these new parameters, first note that since μ is now supported on $(0, +\infty)$, the recurrence relation (3.2) implies

$$a_n = \|P_{n-1}\|^{-2} \langle xP_{n-1}, P_{n-1} \rangle = \|P_{n-1}\|^{-2} \int_0^\infty xP_{n-1}^2(x) \mu(dx) > 0, \quad n = 1, \dots, N. \quad (5.3)$$

Now, we set

$$\begin{aligned} a_1 &= \xi_1, \quad a_n = \xi_{2n-2} + \xi_{2n-1}, \quad \text{for } 2 \leq n \leq N, \\ \text{and} \quad b_n &= \xi_{2n-1}\xi_{2n}, \quad \text{for } 1 \leq n \leq N-1. \end{aligned} \tag{5.4}$$

Equivalently, the new parameters correspond to the Cholesky factorization

$$J_{\mathbf{a}, \mathbf{b}} = \Xi \Xi^T, \quad \text{where} \quad \Xi = \begin{pmatrix} \sqrt{\xi_1} & & & (0) \\ \sqrt{\xi_2} & \sqrt{\xi_3} & & \\ \ddots & \ddots & & \\ (0) & & \sqrt{\xi_{2N-2}} & \sqrt{\xi_{2N-1}} \end{pmatrix}. \tag{5.5}$$

Note that this bidiagonal transformation is reminiscent of the construction of the tridiagonal model for the L β E, where [Dumitriu & Edelman \(2002\)](#) bidiagonalize a random Gaussian matrix. The following proposition shows that the change of variables replacing the recurrence coefficients by $\xi_{1:2N-1}$ is valid.

Proposition 5.1. *Consider μ supported on $(0, +\infty)$, then the corresponding Jacobi matrix (2.1) factorizes uniquely as $J_{\mathbf{a}, \mathbf{b}} = \Xi \Xi^T$, where Ξ is given by (5.5). Moreover, the mapping*

$$(\xi_1, \dots, \xi_{2N-1}) \mapsto (a_{1:N}, b_{1:N-1}), \tag{5.6}$$

defined by (5.4) is a C^1 -diffeomorphism of $(0, +\infty)^{2N-1}$ onto itself, and its Jacobian reads

$$\left| \frac{\partial a_{1:N}, b_{1:N-1}}{\partial \xi_{1:2N-1}} \right| = \prod_{i=1}^{N-1} \xi_{2i-1}. \tag{5.7}$$

Proof. Given that $J_{\mathbf{a}, \mathbf{b}}$ is symmetric with positive eigenvalues, the Cholesky factorization $J_{\mathbf{a}, \mathbf{b}} = \Xi \Xi^T$ is unique, see, e.g., [Golub & Van Loan \(2013, Theorem 4.2.7\)](#). Moreover, since $J_{\mathbf{a}, \mathbf{b}}$ is tridiagonal, the factor Ξ can only be bidiagonal. Hence, the mapping (5.6) is injective (and even bijective) and C^1 because it is polynomial. Finally, by definition of the transformation (5.4), the Jacobian reads as the determinant of a triangular matrix

$$\left| \frac{\partial a_{1:N}, b_{1:N-1}}{\partial \xi_{1:2N-1}} \right| = \left| \begin{bmatrix} \left[\begin{array}{cc} \frac{\partial a_i}{\partial \xi_{2j-1}} & \frac{\partial a_i}{\partial b_i} \\ \frac{\partial b_i}{\partial \xi_{2j-1}} & \frac{\partial b_i}{\partial \xi_{2j}} \end{array} \right]_{i,j=1}^{N-1} & \left[\begin{array}{c} \frac{\partial a_i}{\partial \xi_{2N-1}} \\ \frac{\partial b_i}{\partial \xi_{2N-1}} \end{array} \right]_{i=1}^{N-1} \\ \left[\begin{array}{cc} \frac{\partial a_N}{\partial \xi_{2j-1}} & \frac{\partial a_N}{\partial b_j} \\ \frac{\partial b_j}{\partial \xi_{2j-1}} & \frac{\partial b_j}{\partial \xi_{2j}} \end{array} \right]_{j=1}^{N-1} & \frac{\partial a_N}{\partial \xi_{2N-1}} \end{bmatrix} \right| = \prod_{i=1}^N \underbrace{\frac{\partial a_i}{\partial \xi_{2i-1}}}_{=1} \prod_{i=1}^{N-1} \underbrace{\frac{\partial b_i}{\partial \xi_{2i}}}_{=\xi_{2i-1}}.$$

□

For our purpose, the Cholesky factorization (5.5) is ideal to express the key quantities that appear in the L β E. The proof of the corresponding tridiagonal model, cf. Theorem 2.2, follows from a direct application of Theorem 2.4 and the following immediate lemma.

Lemma 5.2. *Let $J_{\mathbf{a},\mathbf{b}} = \Xi \Xi^\top$ as in (5.5) and note x_1, \dots, x_N its eigenvalues. Then,*

$$\sum_{n=1}^N x_n = \text{Tr } J_{\mathbf{a},\mathbf{b}} = \sum_{n=1}^{2N-1} \xi_n \quad \text{and} \quad \prod_{n=1}^N x_n = \det J_{\mathbf{a},\mathbf{b}} = \prod_{n=1}^N \xi_{2n-1}. \quad (5.8)$$

Proof of Theorem 2.2. Applying Lemma 5.2 to the $V(x) = -(k-1) \log(x) + \frac{x}{\theta}$ yields

$$\begin{aligned} \exp[-\text{Tr } V(J_{\mathbf{a},\mathbf{b}})] &= (\det J_{\mathbf{a},\mathbf{b}})^{k-1} \exp\left(-\frac{1}{\theta} \text{Tr } J_{\mathbf{a},\mathbf{b}}\right) \\ &\stackrel{(5.8)}{=} \prod_{n=1}^N \xi_{2n-1}^{k-1} \exp\left(-\frac{1}{\theta} \sum_{n=1}^{2N-1} \xi_n\right). \end{aligned} \quad (5.9)$$

Starting from (2.8), Proposition 5.1 gives the joint distribution of the underlying $\xi_{1:2N-1}$ parameters as proportional to

$$\begin{aligned} &\prod_{n=1}^{N-1} b_n^{\frac{\beta}{2}(N-n)-1} e^{-\text{Tr } V(J_{\mathbf{a},\mathbf{b}})} da_{1:N} db_{1:N-1} \\ &\stackrel{(5.4)}{=} \prod_{n=1}^{N-1} (\xi_{2n-1} \xi_{2n})^{\frac{\beta}{2}(N-n)-1} e^{-\text{Tr } V(J_{\mathbf{a},\mathbf{b}})} \left| \frac{\partial a_{1:N}, b_{1:N-1}}{\partial \xi_{1:2N-1}} \right| d\xi_{1:2N-1} \\ &\stackrel{(5.7)}{=} \prod_{n=1}^{N-1} \xi_{2n-1}^{\frac{\beta}{2}(N-n)-1} \xi_{2n}^{\frac{\beta}{2}(N-n)-1} e^{-\text{Tr } V(J_{\mathbf{a},\mathbf{b}})} \prod_{n=1}^{N-1} \cancel{\xi_{2n-1}} d\xi_{1:2N-1} \\ &\stackrel{(5.9)}{=} \prod_{n=1}^N \xi_{2n-1}^{\frac{\beta}{2}(N-n)} \prod_{n=1}^{N-1} \xi_{2n}^{\frac{\beta}{2}(N-n)-1} \left(\prod_{n=1}^N \xi_{2n-1} \right)^{k-1} e^{-\frac{1}{\theta} \sum_{n=1}^{2N-1} \xi_n} d\xi_{1:2N-1} \end{aligned} \quad (5.10)$$

□

In the next section, we introduce another reparametrization of the recurrence coefficients, this time when μ is supported in a compact interval: the canonical moments of Dette & Studden (1997).

5.3. The $J\beta E$ and its tridiagonal model

Finding a tridiagonal model for the $J\beta E$ was left as an open problem by Dumitriu & Edelman (2002, IV B). The latter was addressed by Killip & Nenciu (2004, Theorem 2) in their study of the quidiagonal model associated to the circular ensemble. However, the authors acknowledged that lifting the points on the unit circle to apply their result represents a winding detour to prove the $J\beta E$. Besides, the Jacobian required by this method was obtained by indirect means by Killip & Nenciu (2007, Lemma 4.3). Subsequently, Forrester & Rains (2006, Theorem 2) obtained the Jacobian more directly.

We can actually prove the tridiagonal model of Theorem 2.3 by reparametrizing the Jacobi matrix $J_{\mathbf{a}, \mathbf{b}}$ again, this time using *canonical moments* (Dette & Studden, 1997, Chapter 1). In essence, the result can be found in the work of Gautier & Rokhlin (2010) and Dette & Nagel (2012), but we rephrase it as just another consequence of Theorem 2.4.

Before formally introducing them, let us mention that canonical moments and their complex counterpart were successfully used to investigate the connection between randomized moments problems, orthogonal polynomials, and optimal design (Dette & Studden, 1997) and in random matrix theory (Gautier & Rokhlin, 2010; Gautier et al., 2016). In particular, canonical moments can be thought of as a reparametrization of the moments, where $0 < c_n < 1$ represents the relative position of the n -th moment m_n in the range of all possible moments associated to measure with compatible previous moments m_1, \dots, m_{n-1} , see Dette & Studden (1997).

Throughout this section, we assume that the N -atomic measure μ is supported on $(0, 1)$. In particular, with (ξ_n) the parameters introduced in Section 5.2, it comes

$$0 < \xi_{2n-2} + \xi_{2n-1} = a_n = \|P_{n-1}\|^{-2} \langle xP_{n-1}, P_{n-1} \rangle < 1, \quad n = 2, \dots, N.$$

Similarly, $\xi_1 = a_1 \in (0, 1)$. This implies $0 < \xi_n < 1$ for all $1 \leq n \leq 2N - 1$. Following the work of Wall (1940) on chain sequences and continued fractions, we introduce a new parametrization of the recurrence coefficients.

Lemma 5.3 (Wall). *Assume μ is supported on $(0, 1)$, there exist a sequence $(c_n) \in (0, 1)^\mathbb{N}$ such that*

$$\xi_1 = c_1 \quad \text{and} \quad \xi_n = (1 - c_{n-1})c_n, \quad \forall 2 \leq n \leq 2N - 1. \quad (5.11)$$

We do not prove Lemma 5.3 and refer to Wall (1940, Theorem 6.1); see also Chihara (1978, Chapter 3) for more details on chain sequences. We simply note that defining (c_n) in (5.11) is straightforward, the nontrivial part of the lemma is that $0 < c_n < 1$ for all n . We also note that the c_n s are today known as the *canonical moments* of μ ; see the monograph of Dette & Studden (1997).

The following proposition shows that the change of variables replacing ξ by c is valid.

Proposition 5.2. *Consider μ supported on $(0, 1)$, then the corresponding Jacobi matrix (2.1) can be parametrized in terms of the canonical moments following (5.4) and (5.11). Moreover, the mapping*

$$(c_1, \dots, c_{2N-1}) \mapsto (\xi_1, \dots, \xi_{2N-1}), \quad (5.12)$$

defined by (5.11) is a C^1 -diffeomorphism of $(0, 1)^{2N-1}$ onto itself, and its Jacobian reads

$$\left| \frac{\partial \xi_{1:2N-1}}{\partial c_{1:2N-1}} \right| = \prod_{n=1}^{2N-2} (1 - c_n). \quad (5.13)$$

Proof. The map (5.12) is a bijection by definition and Lemma 5.3, and C^1 because it is polynomial. Then, by definition of the transformation (5.11), the Jacobian is the determinant of a triangular matrix

$$\left| \frac{\partial \xi_{1:2N-1}}{\partial c_{1:2N-1}} \right| = \prod_{n=1}^{2N-1} \frac{\partial \xi_n}{\partial c_n} = 1 \cdot \prod_{n=2}^{2N-1} (1 - c_{n-1}) = \prod_{n=1}^{2N-2} (1 - c_n).$$

□

For our purpose, the canonical moment parametrization is ideal to express the key quantities that appear in the $J\beta E$. The proof of the corresponding tridiagonal model in Theorem 2.3 is again a direct application of Proposition 5.2 and the following lemma.

Lemma 5.4. *It holds that*

$$\prod_{n=1}^N x_n = \det J_{\mathbf{a}, \mathbf{b}} = \prod_{n=1}^N c_{2n-1} \prod_{n=1}^{N-1} (1 - c_{2n}) \quad \text{and} \quad \prod_{n=1}^N (1 - x_n) = \det [I_N - J] = \prod_{n=1}^{2N-1} (1 - c_n). \quad (5.14)$$

Proof. First, combine the result of Lemma 5.2 and the definition of the canonical moments in (5.11) to get

$$\prod_{n=1}^N x_n = \xi_1 \prod_{n=2}^N \xi_{2n-1} = c_1 \prod_{n=2}^N (1 - c_{2n-2}) c_{2n-1} = \prod_{n=1}^N c_{2n-1} (1 - c_{2n}). \quad (5.15)$$

Then, Lemma 3.1 yields

$$\prod_{n=1}^N (1 - x_n) = \frac{|\bar{H}_{2N-1}|}{|\underline{H}_{2N-2}|}. \quad (5.16)$$

The denominator can be expressed in terms of the $\xi_{1:2N-1}$ parameters

$$|\underline{H}_{2N-2}|^2 \stackrel{(3.17)}{=} \prod_{n=1}^{N-1} b_n^{N-n} \stackrel{(5.4)}{=} \prod_{n=1}^{N-1} [\xi_{2n-1} \xi_{2n}]^{N-n}. \quad (5.17)$$

For the numerator, we follow Dette & Studden (1997, Theorem 1.4.10) who introduced additional quantities $\gamma_{1:2N-1}$ to get

$$|\bar{H}_{2N-1}| = \gamma_1^N \prod_{n=1}^{N-1} [\gamma_{2n} \gamma_{2n+1}]^{N-n}, \quad (5.18)$$

where

$$\begin{cases} \xi_1 = c_1 \\ \gamma_1 = 1 - c_1 \end{cases} \quad \text{and} \quad \begin{cases} \xi_n = (1 - c_{n-1}) c_n \\ \gamma_n = c_{n-1} (1 - c_n) \end{cases} \quad \forall 2 \leq n \leq 2N-1. \quad (5.19)$$

We plug these results back into (5.16), and conclude that

$$\begin{aligned}
\prod_{n=1}^N (1 - x_n) &= \frac{|\bar{H}_{2N-1}|}{|\underline{H}_{2N-2}|} = \gamma_1^N \prod_{n=1}^{N-1} \left[\frac{\gamma_{2n} \gamma_{2n+1}}{\xi_{2n-1} \xi_{2n}} \right]^{N-n} \\
&= (1 - c_1)^N \left[\frac{\cancel{c_1}(1 - c_2) \cancel{c_2}(1 - c_3)}{\cancel{c_1}(1 - c_1) \cancel{c_2}} \right]^{N-1} \prod_{n=2}^{N-1} \left[\frac{\cancel{c_{2n-1}}(1 - c_{2n}) \cancel{c_{2n}}(1 - c_{2n+1})}{(1 - c_{2n-2}) \cancel{c_{2n-1}}(1 - c_{2n-1}) \cancel{c_{2n}}} \right]^{N-n} \\
&= (1 - c_1)^N \prod_{n=1}^{N-1} \left[\frac{1 - c_{2n+1}}{1 - c_{2n-1}} \right]^{N-n} (1 - c_2)^{N-1} \prod_{n=2}^{N-1} \left[\frac{1 - c_{2n}}{1 - c_{2n-2}} \right]^{N-n} \\
&= \prod_{n=1}^N (1 - c_{2n-1}) \prod_{n=1}^{N-1} (1 - c_{2n}).
\end{aligned}$$

□

Proof of Theorem 2.3. Considering the potential $V(x) = -[(a-1)\log(x) + (b-1)\log(1-x)]$, Lemma 5.4 yields

$$\begin{aligned}
\exp[-\text{Tr } V(J_{\mathbf{a}, \mathbf{b}})] &= [\det J]^{a-1} [\det I_N - J]^{b-1} \\
&\stackrel{(5.14)}{=} \prod n = 1 N c_{2n-1}^{a-1} (1 - c_{2n-1})^{b-1} \prod n = 1 N - 1 (1 - c_{2n})^{a+b-2}.
\end{aligned} \tag{5.20}$$

Starting from (5.10), Proposition 5.2 allows us to express the joint distribution of the canonical moments as

$$\begin{aligned}
&\prod_{n=1}^{N-1} \frac{[\xi_{2n-1} \xi_{2n}]^{\frac{\beta}{2}(N-n)}}{\xi_{2n}} e^{-\text{Tr } V(J_{\mathbf{a}, \mathbf{b}})} \left| \frac{\partial \xi_{1:2N-1}}{\partial c_{1:2N-1}} \right| dc_{1:2N-1} \\
&\stackrel{(5.13)}{=} \prod_{n=1}^{N-1} \frac{[\xi_{2n-1} \xi_{2n}]^{\frac{\beta}{2}(N-n)}}{\xi_{2n}} \prod_{n=1}^{2N-2} (1 - c_n) e^{-\text{Tr } V(J_{\mathbf{a}, \mathbf{b}})} dc_{1:2N-1} \\
&\stackrel{(5.11)}{=} \frac{[c_1(1 - c_1)c_2]^{\frac{\beta}{2}(N-1)}}{(1 - \cancel{c_1})\cancel{c_2}} \prod_{n=2}^{N-1} \frac{[(1 - c_{2n-2})c_{2n-1}(1 - c_{2n-1})c_{2n}]^{\frac{\beta}{2}(N-n)}}{(1 - \cancel{c_{2n-1}})\cancel{c_{2n}}} \\
&\quad \prod_{n=1}^{N-1} \frac{(1 - \cancel{c_{2n-1}})(1 - c_{2n})}{(1 - \cancel{c_{2n-1}})} e^{-\text{Tr } V(J_{\mathbf{a}, \mathbf{b}})} dc_{1:2N-1} \\
&= \prod_{n=1}^N [c_{2n-1}(1 - c_{2n-1})]^{\frac{\beta}{2}(N-n)} \prod_{n=1}^{N-1} c_{2n}^{\frac{\beta}{2}(N-n)-1} (1 - c_{2n})^{\frac{\beta}{2}(N-n-1)+1} e^{-\text{Tr } V(J_{\mathbf{a}, \mathbf{b}})} dc_{1:2N-1} \\
&\stackrel{(5.20)}{=} \prod n = 1 N c_{2n-1}^{\frac{\beta}{2}(N-n)+a-1} (1 - c_{2n-1})^{\frac{\beta}{2}(N-n)+b-1} \prod n = 1 N - 1 c_{2n}^{\frac{\beta}{2}(N-n)-1} (1 - c_{2n})^{\frac{\beta}{2}(N-n-1)+a+b-1} dc_{1:2N-1}.
\end{aligned}$$

□

6. Gibbs sampling tridiagonal models associated to polynomial potentials

As seen in Section 5, for the specific potentials associated to the Hermite, Laguerre, and Jacobi β -ensembles, the successive parametrizations of the corresponding Jacobi matrix $J_{\mathbf{a},\mathbf{b}}$ yield independent coefficients with easy-to-sample distributions. Thus, computing the eigenvalues of the corresponding randomized tridiagonal Jacobi matrices gives $\mathcal{O}(N^2)$ exact samplers. However, when the potential V is generic, these Jacobi parameters may not be independent anymore. For polynomial potentials, this dependence remains mild, in the sense that each parameter remains independent from the rest conditionally on a few “neighboring” parameters. As we shall see in this section, simple Gibbs samplers in the space of these Jacobi parameters can provide surprisingly fast-mixing approximate samplers for β -ensembles. In short, we study a Gibbs sampler on tridiagonal matrices, which we can diagonalize in $\mathcal{O}(N^2)$ to obtain approximate samples from a given β -ensemble. This approach is in contrast with that of [Lieb & Mattis \(2013\)](#) and [Chafaï & Ferré \(2018\)](#), who used MCMC directly on the original space where the particles $\{x_1, \dots, x_N\}$ live.

Our starting point is Proposition 2 of [Krishnapur et al. \(2016\)](#), which we rederived as Theorem 2.4. In short, a Jacobi matrix $J_{\mathbf{a},\mathbf{b}}$ with coefficients distributed as

$$(a_1, b_1, \dots, a_{N-1}, b_{N-1}, a_N) \sim \prod_{i=1}^{N-1} b_i^{\frac{\beta}{2}(N-i)-1} \exp^{-\text{Tr } V(J_{\mathbf{a},\mathbf{b}})} da_{1:N}, b_{1:N-1}, \quad (6.1)$$

has eigenvalues distributed according to the β -ensemble (1.1) with potential V . [Krishnapur et al. \(2016\)](#) already mention their intuition that a Gibbs chain with invariant measure (6.1) and a polynomial potential would mix fast, in $\mathcal{O}(\log N)$, due to the short range interaction between the coefficients. From an algorithmic point of view, the explicit conditionals in (6.1) similarly invite to use a Gibbs sampler, which we investigate in this section. For the sake of presentation, we fix the potential to be a polynomial with even degree at most 6 and positive leading coefficient, i.e.

$$V(x) = g_6 x^6 + g_5 x^5 + g_4 x^4 + g_3 x^3 + g_2 x^2 + g_1 x. \quad (6.2)$$

The absence of a term of degree 5 in (6.2) comes from practical reasons detailed in Section 6.1. While the method applies more generally, we restrict ourselves to potentials of the form (6.2) because (i) it already goes beyond the numerical state-of-the-art, (ii) it is rich enough to require different sampling schemes for different conditionals depending on the coefficients in (6.2), and (iii) the theory of sextic potentials is advanced enough that we have means to empirically assess the convergence of our samplers.

The associated implementation is available in our DPPy toolbox. We also provide a companion Python notebook where we illustrate our sampler on various potentials, see <https://github.com/guilgautier/DPPy/tree/master/notebooks>.

Algorithm 1 Gibbs sampler to sample from (6.1) with $\beta > 0$ and V as in (6.2)

Input: inverse temperature β , potential V , number of MCMC steps T
 Initialize $a_1 = \dots = a_N = b_1 = \dots = b_{N-1} = 0$
for $t = 1$ **to** T **do**
 for $n = 1$ **to** N **do**
 Sample $a_n | \mathbf{a}_{\setminus n}, \mathbf{b}$
 if $n < N$ **then**
 Sample $b_n | \mathbf{a}, \mathbf{b}_{\setminus n}$
 end if
 end for
 $x_1^t, \dots, x_N^t = \text{eigvals}(J_{\mathbf{a}, \mathbf{b}})$
end for

6.1. Sampling from the conditionals

We implement a systematic scan Gibbs sampler (Rokhlin & Caflisch, 2004, Chapter 10) to approximately sample from the distribution (6.1) on the Jacobi coefficients. Writing the conditionals in closed form for the generic sextic potential (6.2) is cumbersome, but we do it for a specific instance in Example 6.1 below. The expansion of $\text{Tr } V(J_{\mathbf{a}, \mathbf{b}})$ in (6.1) reveals that the size of the Markov blanket of each coefficient grows with $\deg V$. Quoting Krishnapur et al. (2016, Section 1), *variables with indices that are $\deg V/2$ apart are conditionally independent given the variables in between*. In other words, the Jacobi coefficients $a_1, \dots, a_N, b_1, \dots, b_{N-1}$ have a more short-range interaction than the corresponding particles x_1, \dots, x_N . Gibbs sampling can leverage that property.

For $1 \leq i \leq N$, let $\mathbf{a}_{\setminus i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_N)$. Similarly, for $1 \leq j \leq N-1$, let $\mathbf{b}_{\setminus j} = (b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_{N-1})$. In practice, we define one complete Gibbs pass as sampling from $a_n | \mathbf{a}_{\setminus n}, \mathbf{b}$, and then $b_n | \mathbf{a}, \mathbf{b}_{\setminus n}$, for each n in turn. We avoid the term of degree 5 in (6.2) to make sure that the conditionals $b_n | \mathbf{a}, \mathbf{b}_{\setminus n}$ are always log-concave, while the conditionals $a_n | \mathbf{a}_{\setminus n}, \mathbf{b}$ are log-concave if $g_2 > 0$ and $g_3 = g_6 = 0$. Univariate log-concave densities are interesting from a sampling point of view, since they are usually amenable to efficient rejection sampling.

In our case, for every log-concave conditional, the mode of the corresponding density can be derived analytically. We can thus use the tailored rejection sampler of Devroye (2012), with an expected 5 rejection steps per draw; see Example 6.1 for details. The overall algorithm is given in Algorithm 1.

When the conditionals $a_n | \mathbf{a}_{\setminus n}, \mathbf{b}$ are not log-concave, we switch from a Gibbs algorithm to a Metropolis-within-Gibbs algorithm, and replace exact sampling of the corresponding conditional by a draw from a Metropolis-Hastings kernel. More specifically, since the log of the conditional densities $a_n | \mathbf{a}_{\setminus n}, \mathbf{b}$ are polynomials, they are easy to differentiate, and we use their gradient in a Metropolis-adjusted Langevin kernel (MALA, see, e.g., Rokhlin & Caflisch, 2004, Section 7.8.5).

Example 6.1 (Quartic potential). Let $V(x) = g_4x^4 + g_2x^2$. With the convention $a_0 = a_{N+1} = b_0 = b_N = 0$, the conditionals write as follows.

For all $n \in \{1, \dots, N\}$,

$$\begin{aligned} a_n | \mathbf{a}_{\setminus n}, \mathbf{b} \\ \sim \exp \left[- \left[g_4 a_n^4 + a_n^2 [g_2 + 4g_4(b_{n-1} + b_n)] + 4g_4 a_n (a_{n-1} b_{n-1} + a_{n+1} b_n) \right] \right], \end{aligned} \quad (6.3)$$

For all $n \in \{1, \dots, N-1\}$,

$$\begin{aligned} b_n | \mathbf{a}, \mathbf{b}_{\setminus n} \\ \sim b_n^{\frac{\beta}{2}(N-i)-1} \exp \left[-2 \left[g_4 b_n^2 + b_n [g_2 + 2g_4(a_n^2 + a_n a_{n+1} + a_{n+1}^2 + b_{n-1} + b_{n+1})] \right] \right]. \end{aligned} \quad (6.4)$$

In this case, for $g_2, g_4 > 0$, the conditionals given in Equations 6.3 and 6.4 are unnormalized and log-concave, with easy-to-find modes. Thus, the rejection sampling technique of Devroye (2012) applies, with an expected number of rejections equal to 5. Given an unnormalized and log-concave target density π with mode $m = \text{argmax}_y \pi(y)$, Devroye (2012) constructs a piecewise dominating function h comprising 3 plateaus and 2 exponential tails such that $\int h / \int \pi \leq 5$. The breakpoints $m+2u, m+u, m+v, m+2v$ are located on both sides of the mode, where $u < 0 < v$ satisfy $\pi(m+x) \geq \pi(m)/4 \geq \pi(m+2x)$. Such u and v can be found using a simple bisection method. In practice, we compute $u' < 0 < v'$ solutions of $\pi(m+x) = \pi(m)/4$ and assign $u = u'/2$ and $v = v'/2$, see Figure 2.

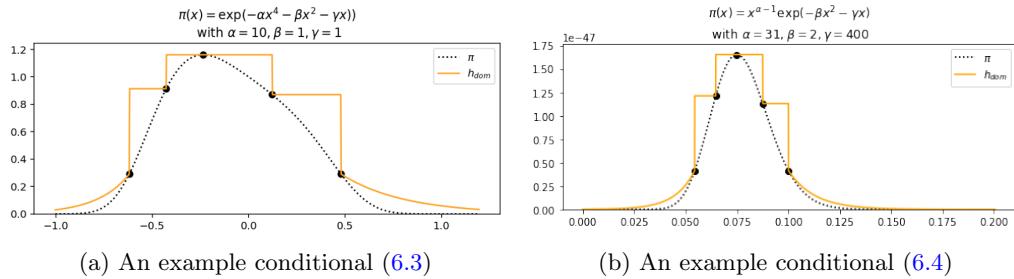


Figure 2: Construction of the dominating function h (solid line) of Devroye (2012) to perform rejection sampling with a log-concave target π (dashed line), the mode of which needs to be analytically tractable.

6.2. Example simulations and empirical study of the convergence

In this section, we investigate the convergence of the Gibbs sampler detailed in Section 6.1. We sample from β -ensembles with potential $W(x) = \frac{\beta N}{2} V(x)$, for various choices of V of the form (6.2). The rescaling in W is applied to capture the weak convergence of the empirical distribution of the particles towards the corresponding equilibrium measure μ_{eq} ; see e.g., Deift (2000, Section 6.1). Intuitively, the rescaling balances the effect of the Vandermonde determinant and that of the potential V in (1.1).

6.2.1. Convergence of the marginals

Let $(x_n^t)_{1 \leq n \leq N}$ be the vector of ordered particles after t full Gibbs passes, that is, after t outer iterations of Algorithm 1. A first quantity to monitor is how well the empirical distribution $\hat{\mu}_N^t = N^{-1} \sum_{n=1}^N \delta_{x_n^t}$ approximates, as t grows, the empirical distribution of the target β -ensemble

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=1}^N \delta_{x_n}, \quad \text{where } \{x_1, \dots, x_N\} \text{ is drawn from (1.1).}$$

It turns out that, under assumptions on the potential V that are satisfied by (6.2), the random measure $\hat{\mu}_N$ is itself well approximated when $N \gg 1$ by a (deterministic) measure μ_{eq} called the *equilibrium measure* of the potential. This statement can be made rigorous; see for instance the large deviation principle with fast rate $1/N^2$ in (Serre, 2015, Theorem 2.3).

Two observations are in order. First, the fast rate hints that the approximation should hold even for moderate values of N , as we shall confirm later on in our simulations. Second, μ_{eq} is known analytically for a few choices of polynomial potentials. Thus, for these potentials, we compare draws from $\hat{\mu}_N^t = N^{-1} \sum_{n=1}^N \delta_{x_n^t}$ with μ_{eq} , to assess convergence of our marginals $\hat{\mu}_N^t$ as t grows. This is in line with the experiments of Lieb & Mattis (2013), Olshanski et al. (2015) and Chafaï & Ferré (2018).

The quartic potential. The equilibrium measure μ_{eq} is available in closed form for potentials proportional to x^{2d} (Deift, 2000, Proposition 6.156). We consider again $V(x) = \frac{1}{4}x^4$, as in Example 6.1. In this case all conditionals are log-concave, and can thus be sampled exactly, cf. Section 6.1. Figure 3 shows the aggregation of the marginal histograms of 1000 independent runs, after each of the first few Gibbs passes.

Observe that convergence to the equilibrium measure is extremely fast: beyond $t = 3$ Gibbs passes, the histograms are visually indistinguishable from the equilibrium measure. This observation is quantitatively monitored in Figure 3, where we plot the logarithm of the L_∞ distance between the empirical cdf of $\hat{\mu}_N^t$ and the cdf of μ_{eq} .

Other potentials of degree 4. We also consider the potential $V(x) = \frac{1}{20}x^4 - \frac{4}{15}x^3 + \frac{1}{5}x^2 + \frac{8}{5}x$ and potentials of the form $V(x) = g_2x^2 + \frac{1}{4}x^4$ where we vary g_2 . Except the case where $g_2 \geq 0$, the conditionals $a_n \mid \mathbf{a}_{\setminus n}, \mathbf{b}$ are not log concave and we sample from them using a few steps of MALA. This allows us to select various qualitative behaviors of μ_{eq} , which may become dissymmetric (Clarkson et al., 2009; Olshanski et al., 2015, Example 1.2; Section 3.2), or supported by more than one connected component (Moler, 2018, Figure 4). Our approach allows to simulate from the corresponding β -ensembles in regimes yet unexplored. Figure 4 shows good agreement of marginal histograms of a single sample of $N = 1000$ points with the equilibrium distribution after only $t = 10$ Gibbs passes.

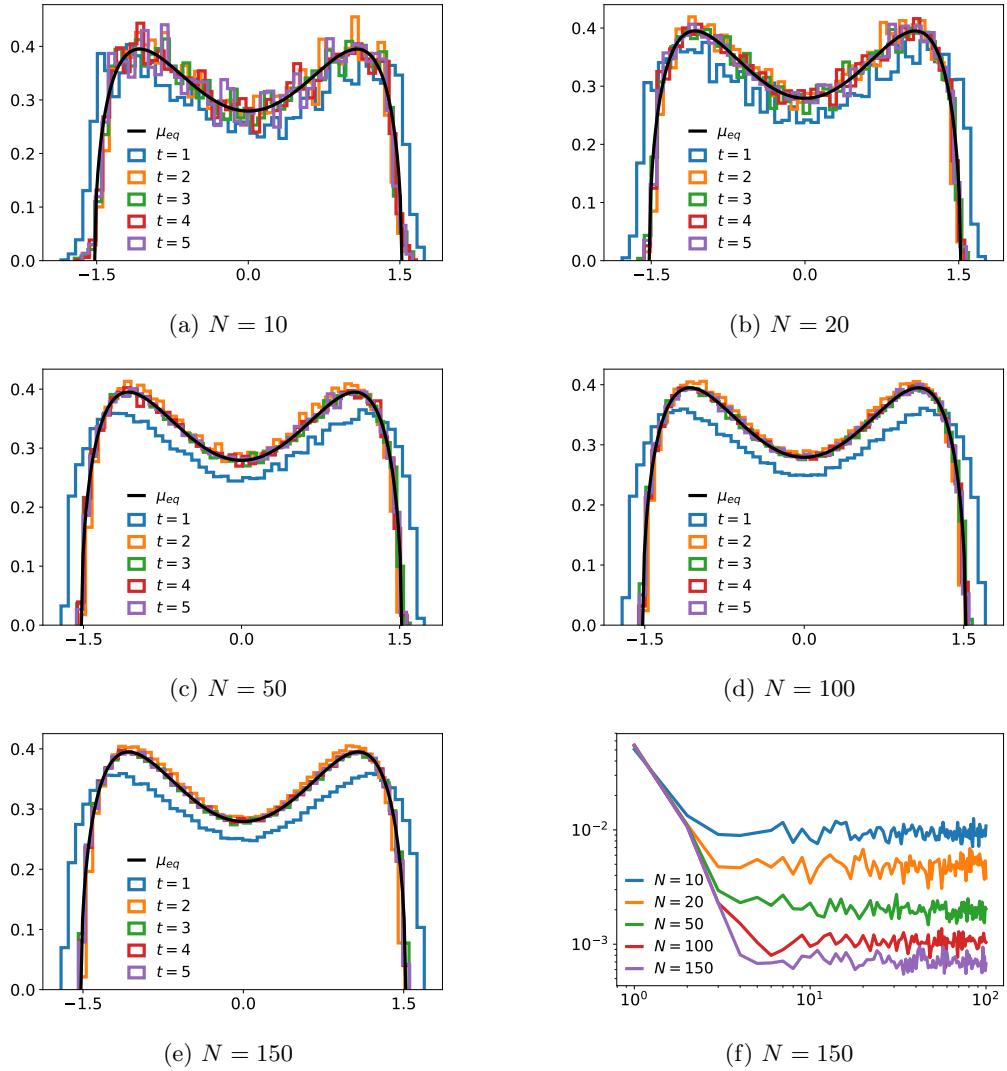


Figure 3: For $\beta = 2$ and $V(x) = \frac{1}{4}x^4$, panels (a)-(e) give a visual display of the convergence of the empirical marginal distribution μ_N^t of the eigenvalues constructed from 1000 independent chains. Each colored line corresponds to a Gibbs pass $t \in \{1, 2, 3, \dots\}$, while the equilibrium pdf is shown as a black line on each panel. Different panels correspond to increasing values of N . Panel (f) shows the supremum norm of the difference between the cdf of μ_{eq} and the cdf of $\hat{\mu}_N^t$ as a function of the number t of Gibbs passes.

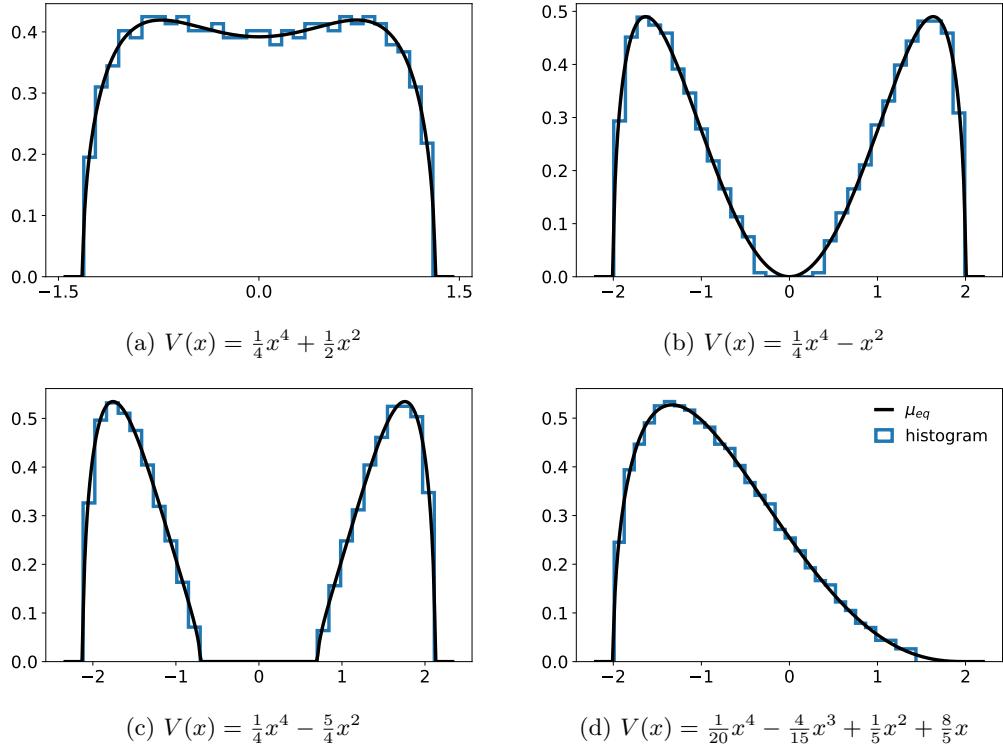


Figure 4: For various choices of potentials V of degree 4 and $\beta = 2$, each panel shows the histogram of a sample from μ_N^t , with $N = 1000$ points after $t = 10$ Gibbs passes. The corresponding equilibrium measures are superimposed in black.

The sextic potential. We extend the derivations of Example 6.1 and consider sampling from the β -ensemble with potential $V(x) = \frac{1}{6}x^6$. The corresponding equilibrium distribution can be derived from Deift (2000, Proposition 6.156). In this case, the conditionals $a_n \mid \mathbf{a}_{\setminus n}, \mathbf{b}$ are not log-concave and we cannot use the exact rejection sampler of Devroye (2012). Instead, we switch to a Metropolis-within-Gibbs sampler and make a few steps of MALA; see Section 6.1.

One free parameter is the number of MALA steps in one Gibbs pass. We empirically observed (not shown) that this number has an influence on the number of Gibbs passes needed to reach the plateau in Figure 5(d). Manually setting the number of MALA steps per Gibbs pass to 100 was enough for rapid overall convergence in our experiments, and larger values did not significantly influence the fit in Figure 5, which is already striking after less than 10 Gibbs passes.

6.2.2. Fluctuations of the largest eigenvalue

After looking at the global behavior of the eigenvalues, we zoom at the right edge of the support of μ to study the local behavior of our approximate samples. We do this for the quartic and sextic potentials. When $\beta = 2$, the target β -ensemble is determinantal and we can test the adequation of the largest atom of $\widehat{\mu}_N^t$ to the universal Tracy-Widom limiting distribution (Deift & Gioev, 2005, Corollary 1.3). We implemented the cumulative distribution function (cdf) of the Tracy-Widom law following the work of Borovkov (2009), and rescale the eigenvalues as Olshanski & Triebel (2014, Section 3.2).

For each potential, we run 1000 independent chains and record only the largest eigenvalue of each chain after each Gibbs pass. Figure 6 and Figure 7 show the histograms of the rescaled largest particles after a few Gibbs passes, respectively for the quartic and sextic potential. More quantitatively, in Figure 8 we monitor the convergence to the Tracy-Widom distribution across Gibbs passes, by computing the supremum distance between the empirical cdf of the largest eigenvalue and the cdf of the Tracy-Widom law.

For the quartic potential, we observe that the adequation with the cdf of the Tracy-Widom law gets tighter as N grows, and again only a few passes of the Gibbs sampler are sufficient to reach a plateau.

In contrast, for the sextic ensemble there seems to be an impassable gap, as if the rescaling was not adequate or the Tracy-Widom law was not the proper limiting distribution. This is despite the square root singularity at the right edge of the equilibrium distribution (Deift, 2000, Section 6.1). In particular, a simple Kolmogorov-Smirnov test at level 0.05 would reject the adequation to Tracy-Widom. Part of this effect might be due to the fact that the conditionals are not sampled exactly in the sextic case, though.

7. Conclusion

First, we wrote down the details of an elementary proof of the three classical tridiagonal models for β -ensembles. Most arguments of the proof already appeared in work by Dumitriu & Edelman (2002); Killip & Nenciu (2004); Forrester & Rains (2006); Gautier & Rokhlin (2010); Dette & Nagel (2012); Krishnapur et al. (2016) and we take no credit for the originality of the proof, only for a stand-alone and elementary version, akin to a survey. We hope that this version will help share the ideas of parametrizing a measure through its recurrence coefficients to computational scientists interested in interacting particle systems. Indeed, throughout the proof, we outline natural reparametrizations of β -ensembles through tridiagonal Jacobi matrices, in which the Vandermonde interaction disappears and leaves only a stream of easy-to-sample, independent matrix entries. Coupled with diagonalization of the underlying tridiagonal matrix, this gives a rejection-free, $\mathcal{O}(N^2)$ exact sampler for the three classical β -ensembles.

Second, when the potential is more generic, independence is lost, but the new interaction can be short-range. We exploited this property to implement a Gibbs kernel and a Metropolis-within-Gibbs variant, which sample β -ensembles with polynomial potentials. This leads to simple MCMC samplers that empirically mix much faster, even for a large

number of points, than more sophisticated MCMC kernels working in the original domain of the particles (Lieb & Mattis, 2013; Chafaï & Ferré, 2018) In particular, marginal behavior that matches known theoretical results can be obtained in a few Gibbs passes, totaling a few seconds on a laptop for hundreds of points. However, local behavior, such as the law of the largest particle in the β -ensemble, remains harder to approximate as the degree of the potential grows. Finally, to be fair, we note that the sampler of Chafaï & Ferré (2018) applies much more generally than ours, and in particular to multivariate β -ensembles.

Finally, we want to stress a third related approach, which we leave for future work. As we have seen, diagonalizing a random Jacobi matrix is equivalent to solving a randomized moment problem. One can thus cast sampling β -ensembles as a constrained optimization problem, namely a linear program, as in the work of Rybicki & Borysow (2015), but with randomized constraints. Our own interest in tridiagonal models actually came from trying to generalize a sampler for finite determinantal point processes (Gautier, Bardenet, and Valko, 2017) of this very form. It is then tempting to look for multivariate versions of the corresponding randomized linear program. We conjecture that the semidefinite relaxations of Lasserre (2010) of multivariate moment problems, with properly randomized constraints, would lead to efficient samplers for multidimensional β -ensembles. This is a technically difficult next step, both in mathematical and computational terms, but it would be useful for Monte Carlo integration (Bardenet & Hardy, 2019).

Acknowledgements

We acknowledge support from ANR grant BoB (ANR-16-CE23-0003) and ERC grant BLACKJACK (ERC-2019-STG-851866).

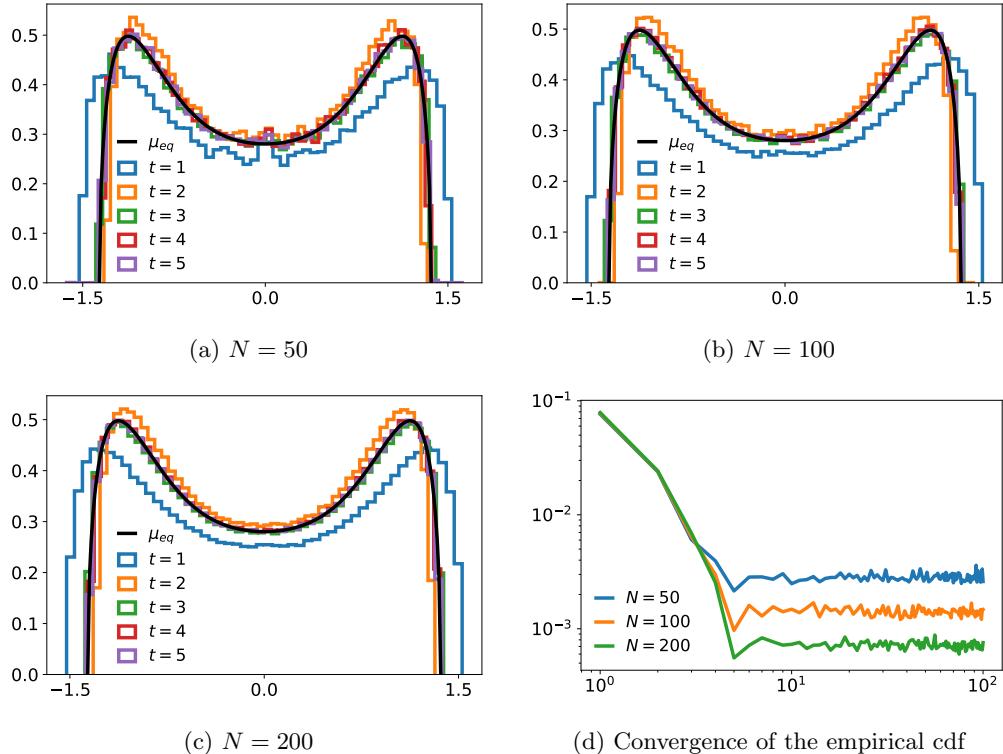


Figure 5: For $\beta = 2$ and $V(x) = \frac{1}{6}x^6$, panels (a)-(c) give a visual display of the convergence of the empirical marginal distribution μ_N^t of the eigenvalues constructed from 1000 independent chains. Each colored line corresponds to a Gibbs pass $t \in \{1, 2, 3, \dots\}$, while the equilibrium pdf is shown as a black line on each panel. Different panels correspond to increasing values of N . Panel (d) shows the supremum norm of the difference between the cdf of μ_{eq} and the cdf of $\hat{\mu}_N^t$ as a function of the number t of Gibbs passes.

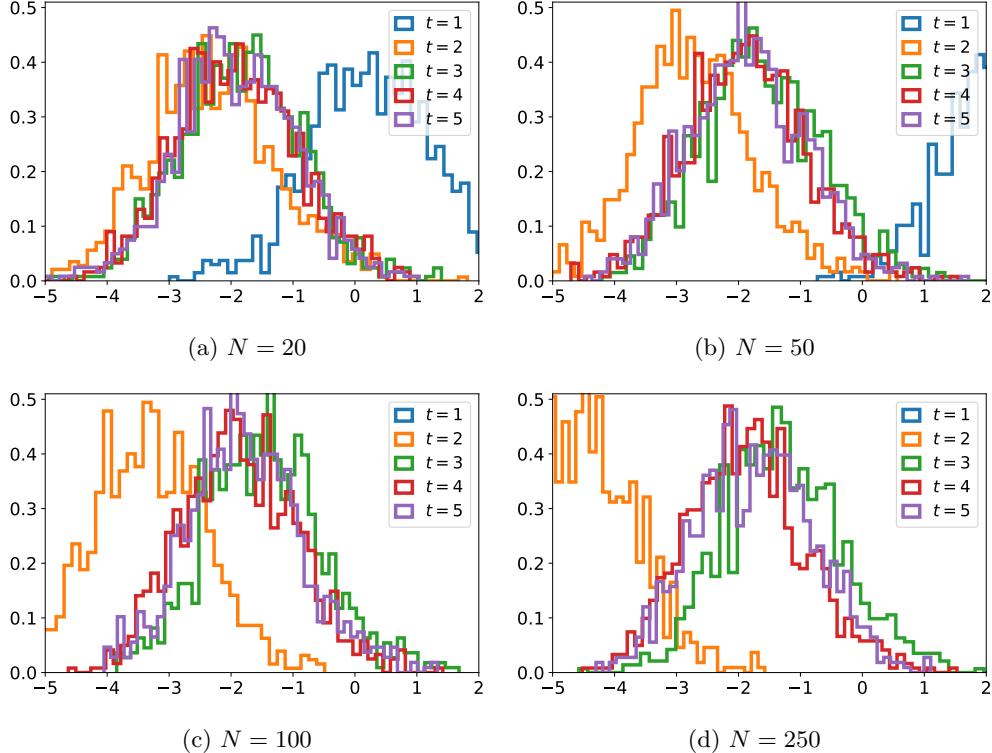


Figure 6: For $\beta = 2$ and $V(x) = \frac{1}{4}x^4$, we give a visual display of the convergence of the empirical distribution of the largest atom of μ_N^t constructed from 1000 independent chains to the Tracy-Widom distribution. Each colored line corresponds to a Gibbs pass $t \in \{1, 2, 3, \dots\}$. Different panels correspond to increasing values of N . As N increases the histogram of passes $t = 1$ and $t = 2$ are farther from matching the high density regions of the Tracy-Widom law but then the fit is good and fast; see also Figures 8(a)-(b) for a more quantitative monitoring.

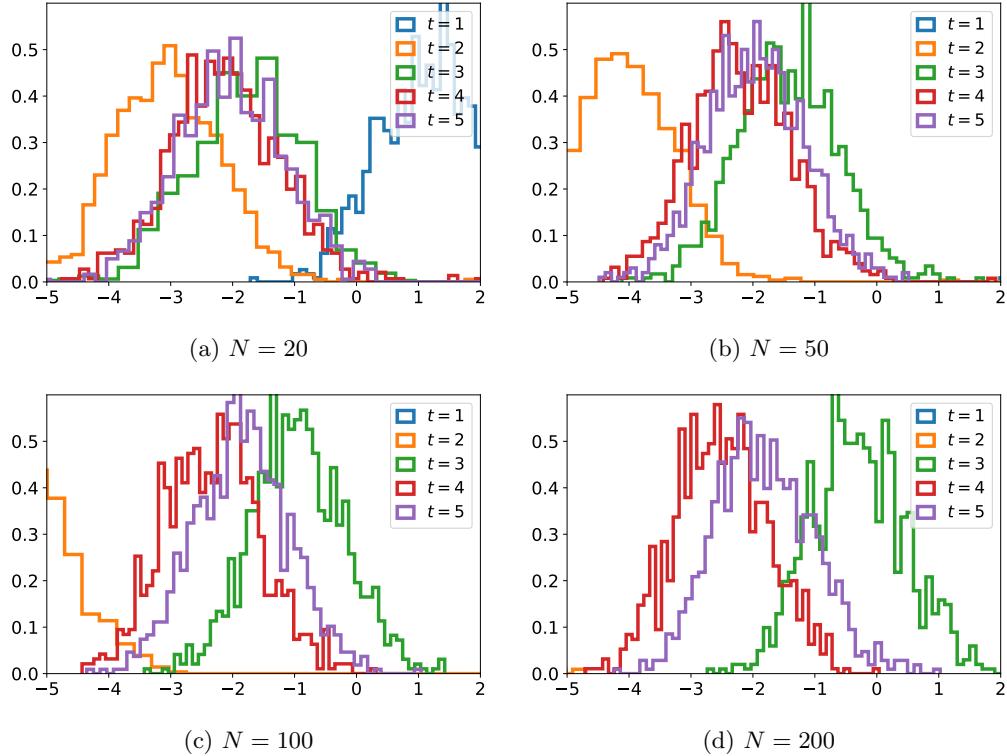


Figure 7: For $\beta = 2$ and $V(x) = \frac{1}{6}x^6$, we give a visual display of the convergence of the empirical distribution of the largest atom of μ_N^t constructed from 1000 independent chains to the Tracy-Widom distribution. Each colored line corresponds to a Gibbs pass $t \in \{1, 2, 3, \dots\}$, while the equilibrium pdf is shown as a black line on each panel. Different panels correspond to increasing values of N . As N increases the histogram of passes $t = 1$ and $t = 2$ are farther from matching the high density regions of the Tracy-Widom law and convergence is not as fast as for the quartic case; see also Figures 8(c)-(d) for a more quantitative monitoring.

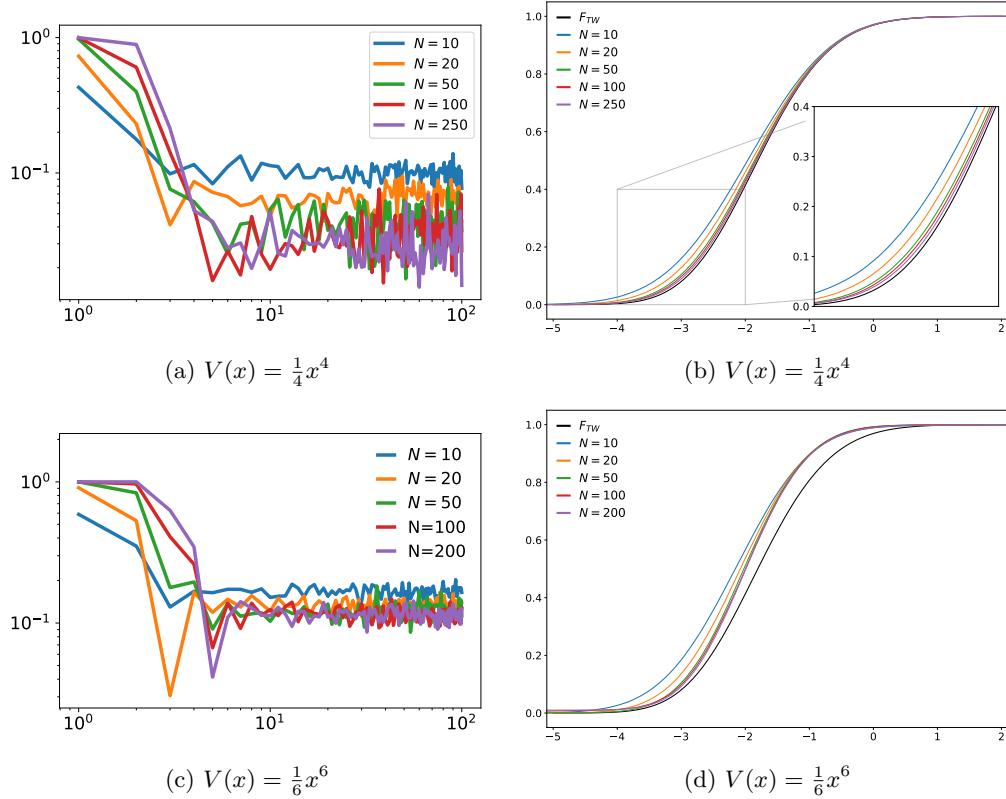


Figure 8: For $\beta = 2$ and $V(x) = \frac{1}{4}x^4$ and $\frac{1}{6}x^6$, we monitor the convergence of the empirical cdf of the largest atom of $\hat{\mu}_N^t$ to the expected Tracy-Widom distribution, for several values of the number of points N . Panels (a) and (c) show the supremum norm of the difference between the cdf of the Tracy-Widom distribution and the empirical cdf of the largest atom of $\hat{\mu}_N^t$, constructed from 1000 independent chains, as a function of the number $1 \leq t \leq 100$ of Gibbs passes. Panels (b) and (d) show the corresponding smoothed empirical cdf constructed from the aggregation of the 1000 independent chains over the $t = 100$ passes. Different panels correspond to increasing values of N .

References

Anderson, G. W., Guionnet, A., and Zeitouni, O. An introduction to random matrices. 2009.

Bardenet, R. and Hardy, A. Monte carlo with determinantal point processes. 2019.

Belhadji, A., Bardenet, R., and Chainais, P. Kernel quadrature with dpps. 2019.

Borovkov, A. Placeholder. 2009.

Carlitz, L. Placeholder. 1971.

Chafaï, D. and Ferré, G. Placeholder. 2018.

Chihara, T. S. Placeholder. 1971.

Chihara, T. S. Placeholder. 1978.

Clarkson, K. L., Kruskal, J. B., and Itoh, T. Placeholder. 2009.

Coakley, E. S. and Rokhlin, V. Placeholder. 2013.

Deift, P. Placeholder. 2000.

Deift, P. and Gioev, D. Placeholder. 2005.

Dette, H. and Nagel, J. Placeholder. 2012.

Dette, H. and Studden, W. J. Placeholder. 1997.

Devroye, L. Placeholder. 2012.

Dumitriu, I. and Edelman, A. Placeholder. 2002.

Dumitriu, I., Koev, P., Parlett, B. N., and Rokhlin, V. Placeholder. 1987.

Forrester, P. J. Placeholder. 2010.

Forrester, P. J. and Rains, E. M. Placeholder. 2006.

Gautier, G. and Rokhlin, V. Placeholder. 2010.

Gautier, G., Nagel, J., and Rokhlin, V. Placeholder. 2016.

Gautier, G., Bardenet, R., and Valko, M. Placeholder. 2017.

Gautier, G., Bardenet, R., and Valko, M. Placeholder. 2019a.

Gautier, G., Bardenet, R., and Valko, M. Placeholder. 2019b.

Gautschi, W. Placeholder. 2004.

Golub, G. H. and Van Loan, C. F. Placeholder. 2013.

Hardy, A. Placeholder. 2018.

Hausdorff, F. and Givens, C. R. Placeholder. 1980.

Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. Placeholder. 2006.

Killip, R. and Nenciu, I. Placeholder. 2004.

Killip, R. and Nenciu, I. Placeholder. 2007.

Konstantopoulos, T. Placeholder. 2005.

Krishnapur, M., Rider, B., and Virág, B. Placeholder. 2016.

Lasserre, J. B. Placeholder. 2010.

Lieb, E. H. and Mattis, D. C. Placeholder. 2013.

Macchi, O. Placeholder. 1975.

Moler, C. B. Placeholder. 2018.

Neal, R. M. Placeholder. 2011.

Olshanski, G. and Triebel, H. Placeholder. 2014.

Olshanski, G., Nazarov, F., and Triebel, H. Placeholder. 2015.

Olver, F. W. J. Placeholder. 2011.

Rokhlin, V. and Caflisch, R. Placeholder. 2004.
Rybicki, G. B. and Borysow, A. Placeholder. 2015.
Serre, D. Placeholder. 2015.
Simon, B. Placeholder. 2011.
Stieltjes, T. J. Placeholder. 1894.
Wall, H. S. Placeholder. 1940.