Monte Carlo with Determinantal Point Processes

Rémi Bardenet, Adrien Hardy

arXiv:1605.00361v1 [math.PR] 2 May 2016

Monte Carlo with Determinantal Point Processes

Rémi Bardenet* Adrien Hardy†

May 3, 2016

Abstract

We show that repulsive random variables can yield Monte Carlo methods with faster convergence rates than the typical $N^{-1/2}$, where N is the number of integrand evaluations. More precisely, we propose stochastic numerical quadratures involving determinantal point processes associated with multivariate orthogonal polynomials, and we obtain root mean square errors that decrease as $N^{-(1+1/d)/2}$, where d is the dimension of the ambient space. First, we prove a central limit theorem (CLT) for the linear statistics of a class of determinantal point processes, when the reference measure is a product measure supported on a hypercube, which satisfies the Nevai-class regularity condition; a result which may be of independent interest. Next, we introduce a Monte Carlo method based on these determinantal point processes, and prove a CLT with explicit limiting variance for the quadrature error, when the reference measure satisfies a stronger regularity condition. As a corollary, by taking a specific reference measure and using a construction similar to importance sampling, we obtain a general Monte Carlo method, which applies to any measure with continuously derivable density. Loosely speaking, our method can be interpreted as a stochastic counterpart to Gaussian quadrature, which, at the price of some convergence rate, is easily generalizable to any dimension and has a more explicit error term.

Contents

*CNRS & CRIStAL, Université de Lille, Cité Scientifique, 59655 Villeneuve d'Ascq Cedex, France. Email: remi.bardenet@gmail.com

†Laboratoire Paul Painlevé, Université de Lille, Cité Scientifique, 59655 Villeneuve d'Ascq Cedex, France. Email: adrien.hardy@math.univ-lille1.fr

1 Introduction

Numerical integration, or quadrature, refers to algorithms that approximate integrals

$$\int f(x)\mu(\mathrm{d}x), \tag{1.1}$$

where μ is a finite positive Borel reference measure, and where f ranges over some class of test functions $\mathscr{C}$. We assume for convenience that the support of μ is included in the d-dimensional hypercube $I^d = [-1,1]^d$, since one can recover this setting in most applications by means of appropriate transformations. For any given N, a quadrature algorithm outputs N nodes $\mathbf{x}_1, \dots, \mathbf{x}_N \in I^d$ and weights $w_1, \dots, w_N \in \mathbb{R}$ so that the approximation

$$\sum_{i=1}^{N} w_i f(\mathbf{x}_i) \approx \int f(x)\mu(\mathrm{d}x) \tag{1.2}$$

is reasonable for every $f \in \mathscr{C}$. The nodes and weights depend on N, μ, and can be realizations of random variables, but they are not allowed to depend on f. The quality of a quadrature algorithm is assessed through the approximation error

$$\mathscr{E}_N(f) = \sum_{i=1}^{N} w_i f(\mathbf{x}_i) - \int f(x)\mu(\mathrm{d}x) \tag{1.3}$$

and specifically its behaviour as $N \to \infty$. Many quadrature algorithms have been developed: variations on Riemann summation [Davis and Rabinowitz, 1984], Gaussian quadrature [Gautschi, 2004], Monte Carlo methods [Robert and Casella, 2004], etc. In the remaining of Section 1, we quickly review three families of such methods to provide context for our contribution, which we then introduce in Section 1.4.

1.1 Gaussian quadrature

Let us first assume $d = 1$, so that μ is supported on $I = [-1, 1]$. Let $(\varphi_k)_{k\in\mathbb{N}}$ be the orthonormal polynomials associated with this measure, that is, the family of polynomials such that φ_k has degree k, positive leading coefficient, and $\int \varphi_k(x)\varphi_\ell(x)\mu(\mathrm{d}x) = \delta_{k\ell}$ for every $k, \ell \in \mathbb{N}$. *Gaussian quadrature*, see e.g. [Davis and Rabinowitz, 1984, Gautschi, 2004, Brass and Petras, 2011] for general references, then corresponds to taking for nodes $\mathbf{x}_1, \dots, \mathbf{x}_N$ the zeros of the Nth degree orthonormal polynomial $\varphi_N(x)$, which are real and simple. As for the weights, Gaussian quadrature corresponds to

$$w_i = \frac{1}{K_N(\mathbf{x}_i, \mathbf{x}_i)}\,, \tag{1.4}$$

where we introduced the Nth Christoffel-Darboux kernel associated with μ,

$$K_N(x, y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y). \tag{1.5}$$

This celebrated method is characterized by the property to be exact, i.e. $\mathscr{E}_N(f) = 0$, for every polynomial function f of degree up to $2N - 1$. This is the highest possible degree such that this holds. Gaussian quadrature is thus particularly suitable when the test functions f look like polynomials. For instance, $\mathscr{E}_N(f)$ decays exponentially fast when f is analytic [Gautschi and Varga, 1983]. However, although Gaussian quadrature is now two centuries old [Gauss, 1815], optimal rates of decay for the error $\mathscr{E}_N(f)$ do not seem to be known for less regular test functions, say $f \in \mathscr{C}^1$, in general. By using Jackson's approximation theorem for algebraic polynomials, one can see that $\mathscr{E}_N(f) = \mathcal{O}(1/N)$ when $f \in \mathscr{C}^1$. Optimal decays have been recently investigated in the particular case of the Gauss-Legendre quadrature [Xiang and Bornemann, 2012, Xiang, 2016]. However, even in the familiar Gauss-Jacobi quadrature, optimal rates are only conjectured.

Efficient computation of the nodes and weights in Gaussian quadrature has been an active topic of research. Classical approaches are based on the QR algorithm, such as the Golub-Welsch algorithm, see e.g. [Gautschi, 2004, Section 3.5] for a discussion. The computational cost of these QR approaches usually scales as $\mathcal{O}(N^2)$. More recently, $\mathcal{O}(N)$ approaches have been proposed for specific choices of the reference measure [Glaser et al., 2007, Hale and Townsend, 2013], with parallelizable methods [Bogaert, 2014] further taking down costs.

Let us stress that Gaussian quadrature is intrinsically a one-dimensional method. Indeed, in the higher-dimensional setting where $\mathrm{Supp}(\mu) \subset I^d$, although one may define multivariate orthonormal polynomials associated with μ, it is not possible to take for nodes the zeros of a multivariate polynomial. However, if μ is a product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ with each μ_j supported on I, one could build a grid of nodes using d one-dimensional Gaussian quadratures. But this has for consequence to rise up the one-dimensional error estimate for $\mathcal{E}_N(f)$ to a power $1/d$, which essentially makes Gaussian quadrature ineffective in higher dimensions than one or two. In fact, the same phenomenon arises for any other grid-like product of one-dimensional quadratures; this is commonly referred to as the curse of dimensionality.

1.2 Monte Carlo methods

Monte Carlo methods [Robert and Casella, 2004] correspond to picking up the N nodes $(\mathbf{x}_i)$ in (1.2) as the realizations of random variables in I^d. For instance, assuming μ in (1.2) has a density ω with respect to the Lebesgue measure, *importance sampling* refers to taking the $(\mathbf{x}_i)$ to be i.i.d. realizations with a so-called *proposal* density q, and the weights to be

$$w_i = \frac{1}{N}\frac{\omega(\mathbf{x}_i)}{q(\mathbf{x}_i)}.$$

That way, $\mathscr{E}_N(f)$ has mean zero. Provided that

$$\mathbb{V}\mathrm{ar}\left[\frac{f(X)\omega(X)}{q(X)}\right] < \infty \tag{1.6}$$

where X has density q, $\mathscr{E}_N(f)$ has a standard deviation decreasing as $N^{-1/2}$, and satisfies the classical central limit theorem:

$$\sqrt{N}\mathscr{E}_N(f) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \sigma_f^2),$$

where σ_f^2 equals (1.6). Let us stress that the cost of importance sampling is virtually $\mathcal{O}(1)$, as draws can be made in parallel.

When the ambient dimension d becomes large, practitioners typically prefer *Markov chain Monte Carlo* (MCMC) methods over importance sampling. This means taking $w_i = 1/N$ and nodes $(\mathbf{x}_i)$ to be the realization of a Markov chain with stationary distribution μ, such as the Metropolis-Hastings chain. Under general conditions on the Markov chain and for f in L^1 of an appropriate measure related to the Markov kernel, $\sqrt{N}\mathscr{E}_N(f)$ then converges in distribution to a centered Gaussian variable [Douc et al., 2014, Theorem 7.32]. The limiting variance grows more slowly with the dimension d than for importance sampling, a posteriori justifying the preferential use of MCMC for large d. In any case, the typical order of magnitude of the error $\mathscr{E}_N(f)$ for Monte Carlo methods is $N^{-1/2}$, which is often deemed a rather slow decay.

Recently, Delyon and Portier [2016] proposed a variant of importance sampling that takes nodes as independent draws from some proposal density q, but takes weights to be

$$w_i = \frac{1}{N}\frac{\omega(\mathbf{x}_i)}{\breve{q}^{-i}(\mathbf{x}_i)},$$

where $\check{q}^{-i}$ is the so-called leave-one-out kernel estimator of the density q of the nodes. Perhaps surprisingly, for smooth enough products $f\omega$ and the right tuning of kernel parameters, $\sqrt{N}\mathscr{E}_N(f)$ then converges in probability to zero. Exact rates are investigated by Delyon and Portier [2016], and a central limit theorem is proven. We further discuss their results in Section 7.

1.3 Quasi-Monte Carlo methods

Quasi-Monte Carlo methods (QMC; [Dick and Pilichshammer, 2010, Dick et al., 2013]) are deterministic constructions that focus on the uniform case, $\mu(\mathrm{d}x) = \mathrm{d}x$ in (1.2). The cornerstone of classical QMC is the Koksma-Hlawka inequality [Dick et al., 2013, Equation 3.15]. This inequality bounds the error $\mathscr{E}_N(f)$ in (1.3) by the product of the star discrepancy of the nodes and the Hardy-Krause variation of f. The star discrepancy measures the departure of the empirical measure of the N nodes from the uniform measure. Classical QMC methods aim at proposing efficient node constructions that minimize this star discrepancy. Some constructions guarantee a star discrepancy that asymptotically decreases as fast as $N^{-1}\log^{d-1} N$. This implies the same rate for $\mathscr{E}_N(f)$ provided f has finite Hardy-Krause variation. While this seems faster than typical Monte Carlo methods in Section 1.2, the rate as a function of N does not decrease until N is exponential in d. Moreover, the Hardy-Krause variation is hard to manipulate in practice.

Modern QMC methods come up with more practical rates [Dick et al., 2013]. For example, scrambled nets [Owen, 1997, 2008] are randomized QMC methods, meaning that a stochastic perturbation is applied to a deterministic QMC construction. The perturbation is built so that $\mathscr{E}_N(f)$ has mean 0. Owen [1997] shows that only assuming f is L^2, the standard deviation of $\mathscr{E}_N(f)$ is $o(N^{-1/2})$, that is, converges to zero faster than the traditional Monte Carlo rate. When f is smooth enough, which requires at least that all mixed partial derivatives of f of order less than d are continuous, Owen [2008] further shows that the standard deviation is $\mathcal{O}(N^{-3/2-1/d}\log^{(d-1)/2} N)$. Again, this rate decreases only when N is exponential in the dimension, but Owen [1997] shows that for finite N, randomized QMC cannot perform significantly worse than Monte Carlo.

1.4 Our contribution

Our main goal is to leverage repulsive particle systems to build a Monte Carlo method with standard deviation of the error decaying as $o(N^{-1/2})$. More precisely, the idea is to use correlated random variables for the quadrature nodes, interacting as strongly repulsive particles. Our motivation comes from specific models in random matrix theory (see Section 2.2 for references), for which the linear statistic $\sum f(x_i)$ converges in distribution to a Gaussian, without requiring any normalizing factor. In this work, we focus on *determinantal point processes* (DPPs), which have received much attention recently in probability and related fields, see [Hough et al., 2006] for a survey.

In any dimension d, we construct DPPs generating the nodes $\mathbf{x}_1, \dots, \mathbf{x}_N$ and appropriate weights w_i's so that the error $\mathscr{E}_N(f)$ in (1.3) decreases rapidly, as $N \to \infty$. We obtain precise rates

$$\mathbb{V}\mathrm{ar}[\mathscr{E}_N(f)] \sim \frac{\sigma_f^2}{N^{1+1/d}} \tag{1.7}$$

with explicit σ_f. In fact, we prove a central limit theorem

$$\sqrt{N^{1+1/d}}\, \mathscr{E}_N(f) \xrightarrow[N\to\infty]{law} \mathcal{N}(0,\sigma_f^2), \tag{1.8}$$

for a certain class of measures μ and any f essentially $\mathscr{C}^1$, see Theorems 2.7 and 2.9. We also provide experimental evidence that convergence in (1.8) happens quite fast.

It turns out that, when $d = 1$, our method is formally very similar to Gaussian quadrature described in Section 1.1. We basically replace the zeros of orthogonal polynomials by particles sampled from *orthogonal polynomial ensembles* (OP Ensembles), DPPs whose building blocks are orthogonal polynomials. Our contribution also has the advantage of generalizing more naturally to higher dimensions than Gaussian quadrature through multivariate OP Ensembles.

Monte Carlo with DPPs is to be classified somewhere between classical Monte Carlo methods and QMC methods, respectively described in Sections 1.2 and 1.3. It is very much similar to importance sampling, but with negatively correlated nodes. Simultaneously, it is more Monte Carlo than scrambled nets, as it does not randomize *a posteriori* a low discrepancy deterministic set of points, but rather incorporate the low discrepancy constraint into the randomization procedure. We further comment on this in Section 7.

The rest of the paper is organized as follows. In Section 2, we state our quadrature rules and theoretical results on the convergence of its error. In Section 3, we demonstrate our results in a simple experimental setting. In Section 4, we introduce key notions and give the outline of our proofs, the technical parts of the proofs being detailed in Sections 5 and 6. We conclude with some perspectives in Section 7.

2 Statement of the results

Notation. All along this work, we write for convenience $I = [-1, 1]$ and $I^d = [-1, 1]^d$. Also, for any $0 < \varepsilon < 1$, we set $I_\varepsilon = [-1 + \varepsilon, 1 - \varepsilon]$ and $I_\varepsilon^d = [-1 + \varepsilon, 1 - \varepsilon]^d$. Finally, except when specified otherwise, a *reference measure* is a positive finite Borel measure with support inside I^d.

2.1 Determinantal point processes and multivariate OP Ensembles

In Section 2.1, we introduce the necessary background in order to state the main results.

2.1.1 Point processes and determinantal correlation functions

A *simple point process* (hereafter *point process*) on I^d is a probability distribution $\mathbb{P}$ on finite subsets $\mathbf{S}$ of I^d; see [Daley and Vere-Jones, 2003] for a general reference. Given a reference measure μ, a point process has a *n-correlation function* ρ_n if one has

$$\mathbb{E}\left[\sum_{\mathbf{x}_{i_1}\neq\cdots\neq\mathbf{x}_{i_n}} \varphi(\mathbf{x}_{i_1},\dots,\mathbf{x}_{i_n})\right] = \int_{(I^d)^n} \varphi(x_1,\dots,x_n)\rho_k(x_1,\dots,x_n)\mu^{\otimes n}(\mathrm{d}x_1,\dots,\mathrm{d}x_n) \tag{2.1}$$

for every bounded Borel function $\varphi : I^n \to \mathbb{R}$, where the sum in (2.1) ranges over all pairwise distinct k-uplets of the random finite subset $\mathbf{S}$. The function ρ_n, provided it exists, thus encodes the correlations between distinct n-uplets of the random set $\mathbf{S}$.

A point process is *determinantal* (DPP) if there exists an appropriate kernel $K : I^d \times I^d \to \mathbb{R}$ such that the n-correlation function exists for every n and reads

$$\rho_n(x_1, \dots, x_n) = \det \Big[K(x_i, x_\ell)\Big]_{i,\ell=1}^{n}, \qquad x_1, \dots, x_n \in I^d. \tag{2.2}$$

In particular such a kernel has to be positive definite, in the sense that the right-hand side of (2.2) is always non-negative. The kernel of a DPP thus encodes how the points in the random configurations interact.

A canonical way to construct DPPs generating configurations of N points $\mathbb{P}$-almost surely, i.e. $\mathbf{S} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, is the following. Consider N orthonormal functions $\varphi_0, \dots, \varphi_{N-1}$ in $L^2(\mu)$, namely satisfying $\int \varphi_k(x)\varphi_\ell(x)\mu(\mathrm{d}x) = \delta_{k\ell}$, and take for kernel

$$K_N(x, y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y). \tag{2.3}$$

In this setting, it turns out the (permutation invariant) random variables $\mathbf{x}_1, \dots, \mathbf{x}_N$ with joint probability distribution

$$\frac{1}{N!} \det \Big[K_N(x_i, x_\ell)\Big]_{i,\ell=1}^{N} \prod_{i=1}^{N} \mu(\mathrm{d}x_i) \tag{2.4}$$

generate a DPP with kernel $K_N(x, y)$. For further information on determinantal point processes, we refer the reader to [Hough et al., 2006, Johansson, 2006, Soshnikov, 2000, Lyons, 2003, Lavancier et al., 2014, 2015].

2.1.2 Multivariate OP Ensembles

In the one-dimensional setting, we can for instance build a DPP using (2.4) with $\varphi_0, \dots, \varphi_{N-1}$ the N lowest degree orthonormal polynomials associated with the reference measure μ. Such DPPs are known as *OP Ensembles* and have been popularized by random matrix theory, see e.g. [Köning, 2005] for an overview.

Our contribution involves a higher-dimensional generalization of OP Ensembles, relying on multivariate orthonormal polynomials, which we now introduce. Given a reference measure μ, assume it has well-defined multivariate orthonormal polynomials, meaning that $\int P^2(x)\mu(\mathrm{d}x) > 0$ for every non-trivial polynomial P. This is for instance true if $\mu(A) > 0$ for some non-empty open set $A \subset I^d$. Now choose an ordering for the multi-indices $(\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, that is, pick a bijection $\mathfrak{b} : \mathbb{N} \to \mathbb{N}^d$. This gives an ordering of the monomial functions $(x_1, \dots, x_d) \mapsto x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, to which one applies the Gram-Schmidt algorithm. This yields a sequence of orthonormal polynomial functions $(\varphi_k)_{k\in\mathbb{N}}$, the multivariate orthonormal polynomials. In this work, we use a specific bijection $\mathfrak{b}$ defined in Section 2.1.3.

Equipped with this sequence $(\varphi_k)_{k\in\mathbb{N}}$ of multivariate orthonormal polynomials, we finally consider for every N the DPP associated with the associated kernel (2.3), that we refer to as the *multivariate OP Ensemble* associated with a reference measure μ. When $d = 1$, it reduces to the classical OP Ensemble.

2.1.3 The graded lexicographic order and the bijection $\mathfrak{b}$

We consider the bijection $\mathfrak{b}$ associated with the graded (with respect to the sup norm) alphabetic order on $\mathbb{N}^d$. We start with the usual lexicographic order on $\mathbb{N}^d$, defined by saying that $(\alpha_1,\dots,\alpha_d) <_{\text{lex}} (\beta_1,\dots,\beta_d)$ if there exists $j \in \{1,\dots,d\}$ such that $\alpha_i = \beta_j$ for every $i < j$ and $\alpha_j < \beta_j$. Now we define the *graded lexicographic order* as follows. We say that $(\alpha_1,\dots,\alpha_d) < (\beta_1,\dots,\beta_d)$ if either $\max\{\alpha_1,\dots,\alpha_d\} < \max\{\beta_1,\dots,\beta_d\}$ or $\max\{\alpha_1,\dots,\alpha_d\} = \max\{\beta_1,\dots,\beta_d\}$ and $(\alpha_1,\dots,\alpha_d) <_{\text{lex}} (\beta_1,\dots,\beta_d)$. Moreover, from now on we specify the bijection $\mathfrak{b}$ to be the unique bijection $\mathbb{N} \to \mathbb{N}^d$ increasing for this order. Otherly put, set $\mathfrak{b}(0) = (0,\dots,0)$ and $\mathfrak{b}(n) = \min\ \mathbb{N}^d \setminus \{\mathfrak{b}(0),\dots,\mathfrak{b}(n-1)\}$ by induction, where the minimum refers to the graded lexicographic order. An important feature of this ordering on which our proofs rely is that, for every $M \geqslant 1$, the set of the first M^d indices $\{\mathfrak{b}(0),\dots,\mathfrak{b}(M^d-1)\}$ matches the discrete hypercube

$$\mathcal{C}_M = \Big\{\boldsymbol{n} \in \mathbb{N}^d :\ 0 \leqslant n_1,\dots,n_d \leqslant M-1\Big\}. \tag{2.5}$$

The indices between $\mathfrak{b}(M^d-1)$ and $\mathfrak{b}(M^{d+1}-1)$ then fill the layer $\mathcal{C}_{M+1}\setminus\mathcal{C}_M$ by following the usual lexicographic order.

We are now in position to state our first result on multivariate OP Ensembles, which is the cornerstone for the Monte Carlo methods we introduce later in Section 2.3.

2.2 A central limit theorem for multivariate OP Ensembles

Several central limit theorems (CLTs) have been obtained for determinantal point processes and related models in random matrix theory, but only when the random configurations lie in a one- or two-dimensional domain. See for instance [Johansson, 1997, 1998, Diaconis and Evans, 2001, Soshnikov, 2002, Pastur, 2006, Rider and Virág, 2007, Popescu, 2009, Kriecherbauer and Shcherbina, 2010, Ameur et al., 2011, 2015, Berman, 2012, Shcherbina, 2013, Breuer and Duits, 2013, 2014, Johansson and Lambert, 2015, Lambert, 2015a,b] for a non-exhaustive list. Although DPPs on higher-dimensional supports have attracted attention in complex geometry [Berman, 2009a,b, 2013, 2014], in statistics [Lavancier et al., 2014, 2015, Møller et al., 2015], and in physics [Torquato et al., 2008, Scardicchio et al., 2009], it seems no CLT has been established yet when $d \geqslant 3$.

Our first result for multivariate OP Ensembles is a CLT for $\mathscr{C}^1$ test functions when the reference measure μ is a product of d Nevai-class probability measures on I. The exact definition of the Nevai class is postponed until Definition 4.1, but we now give a simple sufficient condition. As a consequence of Denisov–Rakhmanov's theorem (see Theorem 4.2), if a measure on I has for Lebesgue decomposition $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x + \mu_s$ (where μ_s is orthogonal to the Lebesgue measure) with $\omega(x) > 0$ almost everywhere, then μ is Nevai-class. Denote by $(T_k)_{k\in\mathbb{N}}$ the normalized Chebyshev polynomials, defined on I by

$$T_0 = 1, \qquad T_k(\cos\theta) = \sqrt{2}\cos(k\theta), \qquad k \geqslant 1.$$

Theorem 2.1. *Let μ be a reference measure supported inside I^d, and assume $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ where each μ_j is Nevai class (see Definition 4.1). If $\mathbf{x}_1,\dots,\mathbf{x}_N$ stands for the associated multivariate OP Ensemble, then for every $f \in \mathscr{C}^1(I^d,\mathbb{R})$, we have*

$$\frac{1}{\sqrt{N^{1-1/d}}}\left(\sum_{i=1}^N f(\mathbf{x}_i) - \mathbb{E}\left[\sum_{i=1}^N f(\mathbf{x}_i)\right]\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0,\sigma_f^2),$$

where

$$\sigma_f^2 = \frac{1}{2} \sum_{k_1,\dots,k_d=0}^{\infty} (k_1+\cdots+k_d) \hat{f}(k_1,\dots,k_d)^2 \tag{2.6}$$

and

$$\hat{f}(k_1,\dots,k_d) = \int_{I^d} f(x_1,\dots,x_d) \prod_{j=1}^{d} T_{k_j}(x_j) \frac{\mathrm{d}x_j}{\pi\sqrt{1-x_j^2}}. \tag{2.7}$$

When $d=1$, Theorem 2.1 was obtained by Breuer and Duits [2013], see also [Lambert, 2015b] for an alternative proof, but the higher-dimensional case $d \geqslant 2$ is novel. We shall restrict to $d \geqslant 2$ for the proof of the theorem, which is deferred to Section 5. Let us now make a few remarks concerning the statement of Theorem 2.1.

Remark 2.2. The limiting variance σ_f^2 does not depend on the reference measure μ.

Remark 2.3. By making the change of variables $x_j = \cos\theta_j$, we obtain

$$\hat{f}(k_1,\dots,k_d) = \frac{(\sqrt{2})^{|\{j:\, k_j\neq 0\}|}}{\pi^d} \int_{[0,\pi]^d} f(\cos\theta_1,\dots,\cos\theta_d) \prod_{j=1}^{d} \cos(k_j\theta_j)\mathrm{d}\theta_j,$$

which is, up to a multiplicative factor, a usual Fourier coefficient.

Next, we obtain that the limiting variance in Theorem 2.1 is dominated by a Dirichlet energy.

Proposition 2.4. *For any $f \in \mathscr{C}^1(I^d,\mathbb{R})$, we have the inequality*

$$\sigma_f^2 \leqslant \frac{1}{2} \sum_{\alpha=1}^{d} \int_{I^d} \left(\sqrt{1-x_\alpha^2}\, \partial_\alpha f(x_1,\dots,x_d)\right)^2 \prod_{j=1}^{d} \frac{\mathrm{d}x_j}{\pi\sqrt{1-x_j^2}}. \tag{2.8}$$

It will appear from the proof we provide in Section 4.3 that this inequality is sharp, since equality holds whenever f is a linear combination of monomials $x_1^{\alpha_1}\cdots x_d^{\alpha_d}$ with $\alpha_j \in \{0,1\}$; see (4.26).

Remark 2.5. After the change of variables $x_j = \cos\theta_j$, we see the right hand side of (2.8) reads

$$\frac{1}{2\pi^d} \sum_{\alpha=1}^{d} \int_{[0,\pi]^d} \partial_\alpha f(\cos\theta_1,\dots,\cos\theta_d)^2 \sin^2\theta_\alpha \prod_{j=1}^{d} \mathrm{d}\theta_j. \tag{2.9}$$

Setting for convenience $\widetilde{f}(\theta_1,\dots,\theta_d) = f(\cos\theta_1,\dots,\cos\theta_d)$, one can interpret (2.9) as a Dirichlet energy since it equals

$$\frac{1}{2\pi^d} \int_{[0,\pi]^d} \left\|\nabla \widetilde{f}(\theta)\right\|^2 \mathrm{d}\theta,$$

where $\|\cdot\|$ stands for the usual Euclidean norm of $\mathbb{R}^d$.

We now turn to Monte Carlo methods based on Theorem 2.1.

2.3 Monte Carlo methods based on determinantal point processes

Consider a reference measure μ with support inside I^d, having well-defined multivariate orthonormal polynomials (say, $\mu(A) > 0$ for some open set $A \subset I^d$). Let $K_N(x,y)$ be the Nth Christoffel-Darboux kernel for the associated multivariate OP Ensemble, namely

$$K_N(x,y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y), \tag{2.10}$$

where $(\varphi_k)_{k\in\mathbb{N}}$ is the sequence of multivariate orthonormal polynomials associated with μ and the graded lexicographic order, see Section 2.1.3. Our quadrature rule is as follows: take for nodes $\mathbf{x}_1, \dots, \mathbf{x}_N$ the random points coming from the multivariate OP Ensemble, namely with joint density (2.4), and for weights $w_i = 1/K_N(\mathbf{x}_i, \mathbf{x}_i)$. Thus, for any μ-integrable function f, our estimator of $\int f(x)\mu(\mathrm{d}x)$ reads

$$\sum_{i=1}^{N} \frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i, \mathbf{x}_i)}. \tag{2.11}$$

One can readily see by taking $k = 1$ in (2.1)–(2.3) that the estimator (2.11) is unbiased,

$$\mathbb{E}\left[\sum_{i=1}^{N} \frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i, \mathbf{x}_i)}\right] = \int f(x)\mu(\mathrm{d}x). \tag{2.12}$$

Remark 2.6. For $d = 1$, comparing (2.11) to (1.4)–(1.5) yields that our method matches Gaussian quadrature except for the nodes, since we replace the zeros of the univariate orthogonal polynomial φ_N by random points drawn from an OP Ensemble. In fact, this replacement is not aberrant since zeros of orthogonal polynomials and particles of associated OP Ensembles get arbitrarily close with high probability as $N \to \infty$, see [Hardy, 2015] for further information and generalizations. Notice however that our quadrature rule makes sense in any dimension.

Our next result is a CLT for (2.11), thus giving a precise rate of decay for the error made in the approximation, provided we make regularity assumptions on μ and on the class $\mathscr{C}$ of test functions f. More precisely, recalling the notation $I_\varepsilon^d = [-1+\varepsilon, 1-\varepsilon]^d$, we consider

$$\mathscr{C} = \Big\{ f \in \mathscr{C}^1(I^d, \mathbb{R}) :\ \mathrm{Supp}(f) \subset I_\varepsilon^d \text{ for some } \varepsilon > 0 \Big\}. \tag{2.13}$$

As for the reference measure, we shall assume μ is a product measure with a density which is $\mathscr{C}^1$ and positive on the open set $(-1,1)^d$. We also need an extra technical assumption, Assumption 1, that we introduce and comment in Section 4.4. Set for convenience

$$\omega_{eq}^{\otimes d}(x) = \prod_{j=1}^{d} \frac{1}{\pi\sqrt{1-x_j^2}}\,, \qquad x \in I^d. \tag{2.14}$$

Theorem 2.7. *Let $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x$ with $\omega(x) = \omega_1(x_1)\cdots\omega_d(x_d)$ be a product reference measure supported inside I^d. Assume its density ω is $\mathscr{C}^1$ and positive on the open* *set* $(-1,1)^d$*, and satisfies Assumption 1. If* $\mathbf{x}_1,\dots,\mathbf{x}_N$ *stands for the multivariate OP Ensemble associated with* μ*, then for every* $f\in\mathscr{C}$*,*

$$\sqrt{N^{1+1/d}}\left(\sum_{i=1}^N\frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)}-\int f(x)\mu(\mathrm{d}x)\right)\xrightarrow[N\to\infty]{law}\mathcal{N}(0,\Omega^2_{f,\omega}),$$

where, see (2.7)*,*

$$\Omega^2_{f,\omega}=\frac{1}{2}\sum_{k_1,\dots,k_d=0}^{\infty}(k_1+\cdots+k_d)\widehat{\left(\frac{f\omega}{\omega_{eq}^{\otimes d}}\right)}(k_1,\dots,k_d)^2. \tag{2.15}$$

In particular, we have for the mean square error of the estimator,

$$\lim_{N\to\infty}N^{1+1/d}\,\mathbb{E}\left[\left(\sum_{i=1}^N\frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)}-\int f(x)\mu(\mathrm{d}x)\right)^2\right]=\Omega^2_{f,\omega}\,. \tag{2.16}$$

We will discuss the assumptions of Theorem 2.7 in Section 4.4, but let us already state that the condition that μ is a product measure is not imposed by the method, as it is the case for Gaussian quadrature. However, our particular proof relies on this factorization. Also, anticipating the discussion on Assumption 1, we prove the following result in Section 4.4.

Proposition 2.8. *Given any parameters* $\alpha_1,\beta_1,\dots,\alpha_d,\beta_d>-1$*, the reference measure*

$$\mu(\mathrm{d}x)=\prod_{j=1}^d(1-x_j)^{\alpha_j}(1+x_j)^{\beta_j}\mathbf{1}_I(x_j)\mathrm{d}x_j, \tag{2.17}$$

satisfies the assumptions of Theorem 2.7.

Hereafter, we call measures of the form (2.17) *Jacobi measures.* From a practical point of view, Theorem 2.7 requires knowledge on the measure μ, in particular all its moments should be known, since we need the corresponding orthonormal polynomials. This is the case for most applications of Gaussian quadrature, where the reference measure is such that orthonormal polynomials are computable, like Jacobi measures (2.17) for instance.

To bypass these restrictions on μ, we provide an importance sampling result.

Theorem 2.9. *Let* $\mu(\mathrm{d}x)=\omega(x)\mathrm{d}x$ *be a reference measure on* I^d *with a* $\mathscr{C}^1$ *density* ω *on the open set* $(-1,1)^d$*. Consider a measure* $q(x)\mathrm{d}x$ *satisfying the assumptions of Theorem 2.7, let* $K_N(x,y)$ *be the* N*th Christoffel-Darboux kernel associated with* $q(x)\mathrm{d}x$*, and* $\mathbf{x}_1,\dots,\mathbf{x}_N$ *the associated multivariate OP Ensemble. Then, for every* $f\in\mathscr{C}$*, we have*

$$\mathbb{E}\left[\sum_{i=1}^N\frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)}\frac{\omega(\mathbf{x}_i)}{q(\mathbf{x}_i)}\right]=\int f(x)\mu(\mathrm{d}x), \tag{2.18}$$

and moreover,

$$\sqrt{N^{1+1/d}}\left(\sum_{i=1}^N\frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)}\frac{\omega(\mathbf{x}_i)}{q(\mathbf{x}_i)}-\int f(x)\mu(\mathrm{d}x)\right)\xrightarrow[N\to\infty]{law}\mathcal{N}(0,\Omega^2_{f,\omega}), \tag{2.19}$$

where $\Omega^2_{f,\omega}$ is the same as (2.15). In particular, we have for the mean square error of the estimator,

$$\lim_{N\to\infty} N^{1+1/d}\, \mathbb{E}\left[\left(\sum_{i=1}^{N} \frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)} \frac{\omega(\mathbf{x}_i)}{q(\mathbf{x}_i)} - \int f(x)\mu(\mathrm{d}x)\right)^2\right] = \Omega^2_{f,\omega}\ . \tag{2.20}$$

Indeed, Theorem 2.9 follows from Theorem 2.7 by taking $f\omega/q$ for test function with $f \in \mathscr{C}$ and $q(x)\mathrm{d}x$ for reference measure.

Remark 2.10. From a classical importance sampling perspective, it is surprising that the limiting variance in (2.19) does not depend on the proposal density q.

Remark 2.11. Proposition 2.4 also yields that

$$\Omega^2_{f,\omega} \leqslant \frac{1}{2\pi} \sum_{\alpha=1}^{d} \int_{I^d} \left(\sqrt{1-x_\alpha^2}\,\partial_\alpha \left(\frac{f\omega}{\omega_{eq}^{\otimes d}}\right)(x_1,\dots,x_d)\right)^2 \prod_{j=1}^{d} \frac{\mathrm{d}x_j}{\pi\sqrt{1-x_j^2}} \tag{2.21}$$

or equivalently, see Remark 2.5,

$$\Omega^2_{f,\omega} \leqslant \frac{1}{2\pi^{d+1}} \int_{[0,\pi]^d} \left\| \nabla \widetilde{\left[\frac{f\omega}{\omega_{eq}^{\otimes d}}\right]}(\theta) \right\|^2 \mathrm{d}\theta. \tag{2.22}$$

2.4 Sampling a multivariate OP Ensemble

For Monte Carlo with DPPs to be a practical tool, we need to be able to sample realizations of the random variables $\mathbf{x}_1,\dots,\mathbf{x}_N$ with joint density (2.4). Hough et al. [2006] give a generic algorithm for sampling DPPs, which we use here; see also [Scardicchio et al., 2009] and [Lavancier et al., 2014] for more details.

The algorithm is based on the fact that the chain rule for the joint distribution (2.4) is available as

$$\frac{1}{N!} \det\left[K_N(x_i,x_\ell)\right]_{i,\ell=1}^{N} \prod_{i=1}^{N} \mu(\mathrm{d}x_i) = \prod_{i=1}^{N} \frac{1}{N-i+1} \left\| P_{H_{i-1}} K_N(x_i,\cdot) \right\|^2_{L^2(\mu)} \mu(\mathrm{d}x_i). \tag{2.23}$$

In (2.23), P_H is the orthogonal projection onto a subspace H of $L^2(\mu)$,

$$H_0 = \mathrm{Span}(\varphi_0,\dots,\varphi_{N-1}),$$

and H_{i-1} is the orthocomplement in H_0 of

$$\mathrm{Span}\left(K_N(x_\ell,\cdot),\ 1 \leqslant \ell \leqslant i-1\right)$$

for every $i > 1$. In particular, all the terms in the product of the RHS of (2.23) are probability measures [Hough et al., 2006, Proposition 19]. Notice that the factorization (2.23) is the equivalent of the "base times height" formula that computes the squared volume of the parallelotope generated by the vectors $(\varphi_0(x_i),\dots,\varphi_{N-1}(x_i))$ for $1 \leqslant i \leqslant N$.

Using the normal equations, we can also rewrite each term in the product (2.23)

$$\left\| P_{H_{i-1}} K_N(x_i, \cdot) \right\|^2_{L^2(\mu)} = \begin{cases} K(x_1, x_1) & \text{if } i = 1, \\ K_N(x_i, x_i) - \mathbf{k}_{1:i-1}(x_i)^T \mathbf{K}_{1:i-1}^{-1} \mathbf{k}_{1:i-1}(x_i) & \text{else,} \end{cases} \tag{2.24}$$

where

$$\mathbf{k}_{1:i-1}(\cdot) = (K_N(x_1, \cdot), \ldots, K_N(x_{i-1}, \cdot))^T$$

and

$$\mathbf{K}_{1:i-1} = \Big[K_N(x_k, x_\ell) \Big]_{1 \leqslant k, \ell \leqslant i-1}.$$

Remark 2.12. Equation (2.24) will be familiar to users of Gaussian processes (GPs; Rasmussen and Williams 2006): the unnormalized conditional densities (2.24) are the incremental posterior variances in a GP model with the same kernel.

Sampling from the joint distribution in (2.23) is achieved by sequentially sampling of each term in the product, using rejection sampling [Robert and Casella, 2004, Section 2.3]. This requires proposal densities $(q_i)_{1 \leqslant i \leqslant N}$ and tight bounds on the density ratios

$$\frac{\left\| P_{H_{i-1}} K_N(x, \cdot) \right\|^2_{L^2(\mu)} \omega(x)}{q_i(x)}, \quad 1 \leqslant i \leqslant N, \tag{2.25}$$

when $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x$. Theorem 4.8 suggests choosing

$$q_i(x) = q(x) = \omega_{eq}^{\otimes d}(x) = \prod_{j=1}^{d} \frac{1}{\pi \sqrt{1 - x_j^2}} \mathbf{1}_{[-1,1]}(x_j), \quad 1 \leqslant i \leqslant N.$$

To bound (2.25), it is enough to bound $K_N(x,x)\omega(x)/\omega_{eq}^{\otimes d}(x)$ since K_N is a positive definite kernel. Obtaining tight bounds is problem-dependent. For Jacobi measures, such bounds can be found for instance in [Gautschi, 2009].

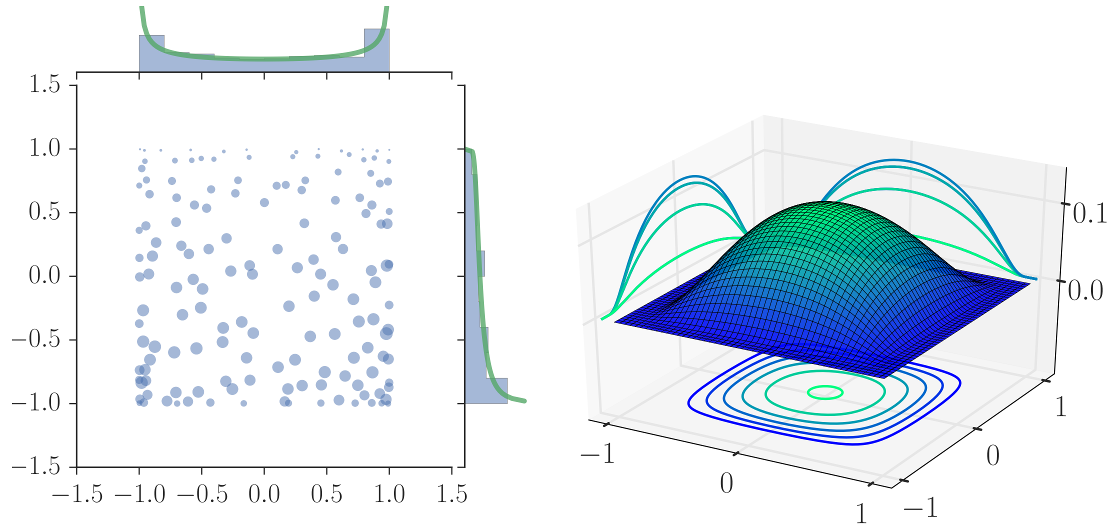
Overall, the cost of the sampling algorithm described in this section is the Achille's heel of Monte Carlo with DPPs for now. Without taking into account the evaluation of orthogonal polynomials nor rejection sampling, the number of basic operations is as much as for Gram-Schmidt orthogonalization of N vectors of dimension N, that is of order N^3 [Golub and Van Loan, 2012, Section 5.2]. Improving this cost is out of the scope of this paper, but we further comment on this issue in Section 7.

3 Experimental illustration

In this section, we illustrate Theorem 2.7 with a toy experiment[1]. In particular, we investigate how fast the Gaussian limit appears.

We take the reference measure $\mu(\mathrm{d}x)$ to be the product Jacobi measure in (2.17) with $\alpha_1 = \beta_1 = -1/2$, and α_j, β_j drawn i.i.d. uniformly on $[-1/2, 1/2]$ for $1 < j \leqslant d$. As proposed in Section 2.4, we use ω_{eq} for density proposal in the rejection sampling steps, and the bounds in [Gautschi, 2009]. Figure 1(a) depicts a weighted sample of the associated

[1]Pending publication, please email the authors to obtain Python code to reproduce the experiments.



(a) 2D sample

(b) test function

Figure 1: 1(a) A weighted sample of the DPP in Theorem 2.7 with Jacobi base measure, see Section 3 for details. 1(b) The test function (3.1) when $d = 2$.

multivariate OP Ensemble when $d = 2$ and $N = 150$. Each disk is centered at a node $\mathbf{x}_i$ in the sample, and the area of the disk is proportional to the weight $1/K_N(\mathbf{x}_i, \mathbf{x}_i)$. The marginal plots on each axis depict the marginal histograms of the weighted sample, with a green curve indicating the density of the marginal Jacobi measures corresponding to $j = 1, 2$ in (2.17). Good agreement is observed for the marginals, as expected from the unbiasedness in (2.12).

We define a simple "bump" test function that is $\mathscr{C}^\infty$ on $I^d = [-1, 1]^d$ and vanishes outside $I_\varepsilon^d = [-1 + \varepsilon, 1 - \varepsilon]^d$,

$$f(x) = \mathbf{1}_{I_\varepsilon^d}(x) \prod_{j=1}^{d} \exp\left(-\frac{1}{1 - \varepsilon - x_j^2} \right), \tag{3.1}$$

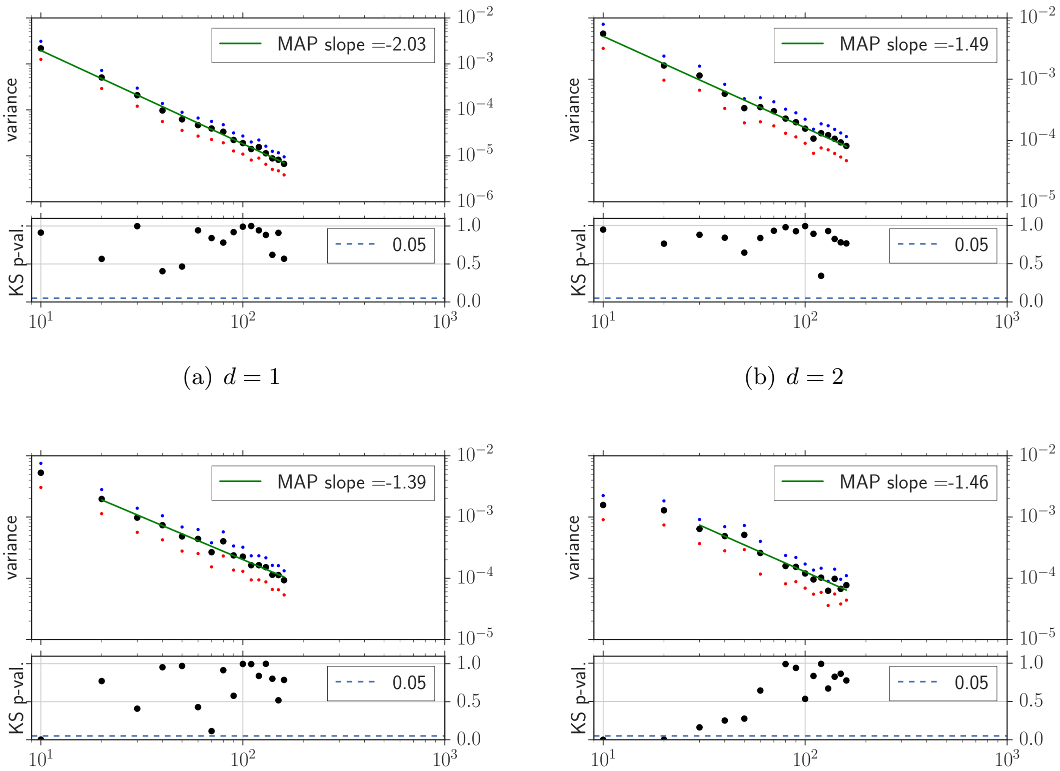
so as to satisfy the assumptions of Theorem 2.7. We set $\varepsilon = 0.05$ and plot f for $d = 2$ in Figure 1(b).

For various $N \in [10, 150]$ and each dimension $d \in \{1, 2, 3, 4\}$, we sample $N_{\text{repeat}} = 100$ independent realizations of $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. We plot the results for each dimension d in Figure 2. On each quadrant and for each N, we plot the sample variance of

$$\sum_{i=1}^{N} \frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i, \mathbf{x}_i)},$$

computed over the N_{repeat} realizations. Blue and red dots indicate standard confidence intervals, for indication only.

For a given dimension d, we want to infer the rate of decay of the variance, in order to confirm the rate in the CLT of Theorem 2.7. We proceed as follows. We first select the values of N for which the N_{repeat} realizations give a p-value larger than 0.05 in a Kolmogorov-Smirnov test of Gaussianity. This is meant to eliminate the small values of N



(a) $d = 1$

(b) $d = 2$

(c) $d = 3$

(d) $d = 4$

Figure 2: Summary of the results.

for which the Gaussian in the CLT (2.19) is a bad approximation for our samples. We do not claim to perform any multiple testing, but rather use the p-value as a loose indicator of Gaussianity. The bottom plot of each quadrant of Figure 2 shows the p-values as a function of N. Note how Gaussianity is hinted even for small N in $d = 1, 2$, while for larger dimensions, it takes larger N to kick in. Then, we perform a standard Bayesian linear regression on the selected log variances vs. $\log(N)$, with wide Gaussian priors for the slope and intercept. For visualization, we plot on each quadrant the maximum a posteriori (MAP) line in green and indicate its slope in the legend. We now summarize the obtained posteriors by giving central 95%-credible intervals in Table 3, that is an interval such that the posterior puts 2.5% of its mass to its left and another 2.5% to its right.

d	credible interval C	theoretical slope $s_{\text{th}} = -1 - 1/d$	$s_{\text{th}} \in C$?
1	$[-2.12, -1.93]$	-2	✓
2	$[-1.61, -1.38]$	-1.5	✓
3	$[-1.55, -1.23]$	-1.33	✓
4	$[-1.74, -1.20]$	-1.25	✓

Table 1: Posterior credible intervals for the variance decay.

Table 3 shows very good agreement between experimental results and the CLT in Theorem 2.7 for each dimension d. As shown by the MAP slope and the credible intervals in Table 3, the CLT approximation is strikingly accurate for $d = 1, 2$, even for small N. For $d = 3, 4$, the credible intervals are relatively large, and the MAP values are quite off the theoretical slopes. This larger uncertainty is due to less points satisfying our Gaussianity requirement, in the bottom panels of Figures 2(c) and 2(d), and a more erratic behaviour in the top panel of Figure 2(d). Unsurprisingly, convergence to a Gaussian distribution is slower when the dimension increases.

4 Orthogonal polynomials and outlines for the proofs

In this section, we provide some general background on orthogonal polynomials and outlines for the proofs of the main theorems.

4.1 Orthogonal polynomials and the Nevai class

In the following, we use the equilibrium measure μ_{eq} of I, defined by

$$\mu_{eq}(\mathrm{d}x) = \omega_{eq}(x)\mathrm{d}x, \qquad \omega_{eq}(x) = \frac{1}{\pi\sqrt{1-x^2}}\,\mathbf{1}_I(x). \tag{4.1}$$

The name comes from its characterization as the unique minimizer of the logarithmic energy $\iint \log|x-y|^{-1}\mu(\mathrm{d}x)\mu(\mathrm{d}y)$ over Borel probability measures μ on I [Saff and Totik, 1997]. It is also the image of the uniform measure on the unit circle through the map $e^{i\theta} \mapsto x = \cos\theta$. The associated orthonormal polynomials are the normalized Chebyshev polynomials of the first kind, defined on I by

$$T_k(\cos\theta) = \begin{cases} \sqrt{2}\cos(k\theta) & \text{if } k \geqslant 1 \\ 1 & \text{if } k = 0 \end{cases}, \qquad \theta \in [0, \pi]. \tag{4.2}$$

They satisfy the three-term recurrence relation

$$xT_k(x) = a_k^* T_{k+1}(x) + b_k^* T_k(x) + a_{k-1}^* T_{k-1}(x), \qquad k \in \mathbb{N}, \tag{4.3}$$

where

$$a_k^* = \begin{cases} 0 & \text{if } k = -1 \\ 1/\sqrt{2} & \text{if } k = 0 \\ 1/2 & \text{if } k \geqslant 1 \end{cases} \qquad \text{and} \qquad b_k^* = 0. \tag{4.4}$$

More generally, given a reference measure μ on I with orthonormal polynomials (φ_k), we always have the three-term recurrence relation

$$x\varphi_k(x) = a_k \varphi_{k+1}(x) + b_k \varphi_k(x) + a_{k-1} \varphi_{k-1}(x), \qquad k \in \mathbb{N}, \tag{4.5}$$

where $a_{-1} = 0$ and $a_k > 0$ and $b_k \in \mathbb{R}$ for every $k \geqslant 0$. The existence of the recurrence coefficients $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ follows by decomposing the polynomial $x\varphi_k$ into the orthonormal family $(\varphi_\ell)_{\ell=0}^{k+1}$ of $L^2(\mu)$ and observing that $\langle x\varphi_k, \varphi_\ell\rangle = \langle \varphi_k, x\varphi_\ell\rangle = 0$ as soon as $\ell < k-1$ by orthogonality.

Definition 4.1. A measure μ supported on I is *Nevai-class* if the recurrence coefficients for the associated orthonormal polynomials satisfy

$$\lim_{k\to\infty} a_k = 1/2, \qquad \lim_{k\to\infty} b_k = 0.$$

Notice the respective limits of the a_k's and b_k's for Nevai class measures are the recurrence coefficients (4.4) of the measure μ_{eq} when $k \geqslant 1$.

The next theorem gives a sufficient condition for a measure to be Nevai class [Simon, 2011, Theorem 1.4.2].

Theorem 4.2. (Denisov-Rakhmanov) *Let μ be a reference measure on I with Lebesgue decomposition $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x + \mu_s$. If $\omega(x) > 0$ almost everywhere, then μ is Nevai-class.*

Consider now the Christoffel-Darboux kernel

$$K_N(x,y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y), \tag{4.6}$$

and notice $\frac{1}{N}K_N(x,x)\mu(\mathrm{d}x)$ is a probability measure. One of the interesting properties of Nevai-class measures is that this probability measure has μ_{eq} for weak limit as $N \to \infty$ [Stahl and Totik, 1992].

Theorem 4.3. *Assume μ supported on I is Nevai-class. Then, for every $f \in \mathscr{C}^0(I,\mathbb{R})$,*

$$\int f(x)\frac{1}{N}K_N(x,x)\mu(\mathrm{d}x) \xrightarrow[N\to\infty]{} \int f(x)\mu_{eq}(\mathrm{d}x).$$

Now, consider instead a reference measure μ on I^d with associated multivariate orthogonal polynomials $(\varphi_k)_{k\in\mathbb{N}}$ (see Section 2.1) and Christoffel-Darboux kernel $K_N(x,y)$ defined as in (4.6). Assume further that $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ is a product of d measures

μ_j on I, and denote by $\varphi_k^{(j)}$ and $K_N^{(j)}(x,y)$ the respective orthogonal polynomials and Christoffel-Darboux kernel associated with μ_j. Then, we have

$$\varphi_k(x) = \varphi_{k_1}^{(1)}(x_1)\cdots\varphi_{k_d}^{(d)}(x_d) \tag{4.7}$$

where $(k_1,\ldots,k_d) = \mathfrak{b}(k)$. Moreover,

$$K_{M^d}(x,y) = \prod_{j=1}^d K_M^{(j)}(x_j,y_j). \tag{4.8}$$

As a consequence, Theorem 4.3 easily yields the following.

Corollary 4.4. *Let $\mu = \mu_1\otimes\cdots\otimes\mu_d$ with μ_j supported on I and Nevai-class. Then, for every $f\in\mathscr{C}^0(I^d,\mathbb{R})$,*

$$\int f(x)\frac{1}{N}K_N(x,x)\mu(\mathrm{d}x) \xrightarrow[N\to\infty]{} \int f(x)\mu_{eq}^{\otimes d}(\mathrm{d}x). \tag{4.9}$$

Proof. By the Stone-Weierstrass theorem, it is enough to show (4.9) when $f(x) = \prod_{j=1}^d f_j(x_j)$ with $f_j\in\mathscr{C}^0(I,\mathbb{R})$. Without loss of generality, one can further assume the functions f_j are non-negative. Let $M = \lfloor N^{1/d}\rfloor$ be the unique integer satisfying $M^d \leqslant N < (M+1)^d$. Since we have $K_{M^d}(x,x)\leqslant K_N(x,x)\leqslant K_{(M+1)^d}(x,x)$ and, by (4.8),

$$\begin{aligned}&\frac{M^d}{N}\prod_{j=1}^d\int f_j(x)\frac{1}{M}K_M^{(j)}(x,x)\mu_j(\mathrm{d}x)\\ &\quad\leqslant \int f(x)\frac{1}{N}K_N(x,x)\mu(\mathrm{d}x)\leqslant\frac{(M+1)^d}{N}\prod_{j=1}^d\int f_j(x)\frac{1}{M+1}K_{M+1}^{(j)}(x,x)\mu_j(\mathrm{d}x),\end{aligned} \tag{4.10}$$

Corollary 4.4 follows from Theorem 4.3. □

The next lemma is yet another aspect of Nevai-class measures that is relevant to our proofs, and may be of independent interest.

Lemma 4.5. *Assume μ supported on I is Nevai-class. We have the weak convergence of*

$$Q_N(\mathrm{d}x,\mathrm{d}y) = (x-y)^2K_N(x,y)^2\mu(\mathrm{d}x)\mu(\mathrm{d}y) \tag{4.11}$$

towards

$$L(\mathrm{d}x,\mathrm{d}y) = \frac{1}{2}(1-xy)\mu_{eq}(\mathrm{d}x)\mu_{eq}(\mathrm{d}y). \tag{4.12}$$

Proof. First, the Christoffel-Darboux formula reads

$$(x-y)^2K_N(x,y)^2 = a_N^2\big(\varphi_N(x)\varphi_{N-1}(y)-\varphi_{N-1}(x)\varphi_N(y)\big)^2, \tag{4.13}$$

and thus, by the orthonormality conditions, we see $\iint Q_N(\mathrm{d}x,\mathrm{d}y) = 2a_N^2$. Since μ is Nevai-class, the former converges to $1/2 = \iint L(\mathrm{d}x,\mathrm{d}y)$. This allows us to use the usual weak topology (i.e. the topology coming by duality with respect to the continuous functions) for bounded Borel measures.

Step 1. We first prove the lemma when $\mu = \mu_{eq}$, so that the φ_k's are the Chebyshev polynomials T_k, see (4.2). By (4.13), the push-forward of (4.11) by the map $(x,y) \mapsto (\cos\theta, \cos\eta)$, where $\theta, \eta \in [0,\pi]$, reads

$$\frac{1}{\pi^2}\Big\{\cos(N\theta)\cos((N-1)\eta) - \cos((N-1)\theta\cos(N\eta)\Big\}^2 \mathrm{d}\theta\mathrm{d}\eta. \tag{4.14}$$

This measure has for Fourier transform

$$\begin{aligned}&\frac{1}{\pi^2}\int_0^\pi\int_0^\pi e^{i(\theta u+\eta v)}\Big\{\cos(N\theta)\cos((N-1)\eta) - \cos((N-1)\theta)\cos(N\eta)\Big\}^2 \mathrm{d}\theta\mathrm{d}\eta\\ &\quad= \frac{1}{\pi^2}\int_0^\pi\int_0^\pi \cos(\theta u+\eta v)\Big\{\cos(N\theta)\cos((N-1)\eta) - \cos((N-1)\theta)\cos(N\eta)\Big\}^2 \mathrm{d}\theta\mathrm{d}\eta.\end{aligned}$$

By developing the square in the integrand and linearizing the products of cosines, we see that the non-vanishing contribution as $N \to \infty$ of the Fourier transform are the terms which are independent on N. Indeed, the N-dependent terms come up with a factor $1/N$ after integration. Thus, the Fourier transform equals, up to $\mathcal{O}(1/N)$, to

$$\frac{1}{2\pi^2}\int_0^\pi\int_0^\pi \cos(\theta u+\eta v)\big(1-\cos\theta\cos\eta\big)\mathrm{d}\theta\mathrm{d}\eta.$$

This yields the weak convergence of (4.14) towards $(2\pi^2)^{-1}(1-\cos\theta\cos\eta)\mathrm{d}\theta\mathrm{d}\eta$, and the lemma follows, in the case where $\mu = \mu_{eq}$, by taking the image of the measures by the inverse map $(\cos\theta, \cos\eta) \mapsto (x,y)$.

Step 2. We now prove the lemma for a general Nevai-class measure μ on I. Let us denote by Q_N^μ the measure (4.11) in order to stress the dependence on μ. Thanks to Step 1, it is enough to prove that for every $m,n \in \mathbb{N}$, we have

$$\lim_{N\to\infty}\left|\iint x^m y^n Q_N^\mu(\mathrm{d}x,\mathrm{d}y) - \iint x^m y^n Q_N^{\mu_{eq}}(\mathrm{d}x,\mathrm{d}y)\right| = 0,$$

in order to complete the proof of the lemma. Recalling (4.11), (4.13), and that $a_N \to 1/2$, it is enough to show that for every $m \in \mathbb{N}$,

$$\lim_{N\to\infty}\left|\int x^m\varphi_N^2(x)\mu(\mathrm{d}x) - \int x^m T_N^2(x)\mu_{eq}(\mathrm{d}x)\right| = 0 \tag{4.15}$$

and

$$\lim_{N\to\infty}\left|\int x^m\varphi_N(x)\varphi_{N-1}(x)\mu(\mathrm{d}x) - \int x^m T_N(x)T_{N-1}(x)\mu_{eq}(\mathrm{d}x)\right| = 0. \tag{4.16}$$

To do so, we first complete for convenience the sequences of recurrence coefficients $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ introduced in (4.5) as bi-infinite sequences $(a_n)_{n\in\mathbb{Z}}$, $(b_n)_{n\in\mathbb{Z}}$, where we set $a_n = b_n = 0$ for every $n < 0$. It follows inductively from the three-term recurrence relation (4.5) that for every $k, \ell, m \in \mathbb{N}$,

$$\int x^m\varphi_k(x)\varphi_\ell(x)\mu(\mathrm{d}x) = \sum_{\gamma:(0,k)\to(m,\ell)}\prod_{e\in\gamma}\omega(e)_{\{(a_n),(b_n)\}}, \tag{4.17}$$

where the sum ranges over all the paths γ lying on the oriented graph with vertices $\mathbb{Z}^2$ and edges $(i,j) \to (i+1, j+1)$, $(i,j) \to (i+1, j)$ and $(i,j) \to (i+1, j-1)$ for $(i,j) \in \mathbb{Z}^2$, starting from $(0,k)$ and ending at (m,ℓ). For every edge e of $\mathbb{Z}^2$, we introduced the weight associated with the sequences $(a_n) = (a_n)_{n\in\mathbb{Z}}$, $(b_n) = (b_n)_{n\in\mathbb{Z}}$ defined by

$$\omega(e)_{\{(a_n),\,(b_n)\}} = \begin{cases} a_j & \text{if } e = (i,j) \to (i+1, j+1) \\ b_j & \text{if } e = (i,j) \to (i+1, j) \\ a_{j-1} & \text{if } e = (i,j) \to (i+1, j-1), \end{cases} \tag{4.18}$$

see also [Hardy, 2015]. Now, observe that the set of all paths γ satisfying $\gamma : (0,k) \to (m,\ell)$ only depends on k, ℓ through $|k-\ell|$ and is empty as soon as $|k-\ell| > m$. Thus it is a finite set, and moreover, by translation of the indices, for every $k, \ell, m \in \mathbb{N}$ we have

$$\int x^m \varphi_k(x)\varphi_\ell(x)\mu(\mathrm{d}x) = \mathbf{1}_{|k-\ell|\leqslant m} \sum_{\gamma:(0,k-\ell)\to(m,0)} \prod_{e\in\gamma} \omega(e)_{\{(a_{n+\ell}),\,(b_{n+\ell})\}}. \tag{4.19}$$

In particular, see (4.3)–(4.4),

$$\int x^m T_k(x) T_\ell(x) \mu_{eq}(\mathrm{d}x) = \mathbf{1}_{|k-\ell|\leqslant m} \sum_{\gamma:(0,k-\ell)\to(m,0)} \prod_{e\in\gamma} \omega(e)_{\{(a^*_{n+\ell}),\,(b^*_{n+\ell})\}}. \tag{4.20}$$

Finally, by combining (4.19) and (4.20), we obtain

$$\begin{aligned} &\left| \int x^m \varphi_k(x)\varphi_\ell(x)\mu(\mathrm{d}x) - \int x^m T_k(x) T_\ell(x)\mu_{eq}(\mathrm{d}x) \right| \\ &\qquad \leqslant \sum_{\gamma:(0,k-\ell)\to(m,0)} \left| \prod_{e\in\gamma} \omega(e)_{\{(a_{n+\ell}),\,(b_{n+\ell})\}} - \prod_{e\in\gamma} \omega(e)_{\{(a^*_{n+\ell}),\,(b^*_{n+\ell})\}} \right|. \end{aligned} \tag{4.21}$$

Together with the Nevai-class assumption for μ, which states that $a_n - a_n^* \to 0$ and $b_n - b_n^* \to 0$ as $n \to \infty$, it follows that (4.15) and (4.16) hold true by taking $k = \ell = N$, or $k = N$ and $\ell = N-1$, in (4.21). This completes the proof of Lemma 4.5.

□

4.2 Sketch of the proof of Theorem 2.1

4.2.1 Reduction to probability reference measures

First, in the statement of Theorem 2.1, we can assume the reference measure μ is a probability measure without loss of generality. This will simplify notation in the proof of Theorem 2.1.

Indeed, for any positive measure μ on I^d with (multivariate) orthonormal polynomials φ_k and any $\alpha > 0$, the orthonormal polynomials associated with $\alpha\mu$ are $\varphi_k/\sqrt{\alpha}$. Thus, if we momentarily denote by $K_N(\mu; x, y)$ the Nth Christoffel-Darboux kernel associated with a measure μ, we have $K_N(\alpha\mu; x, y) = K_N(\mu; x, y)/\alpha$. As a consequence, for every $n \geqslant 1$, the correlation measures

$$\det\Big[K_N(\mu; x_i, x_\ell)\Big]_{i,\ell=1}^n \prod_{i=1}^n \mu(\mathrm{d}x_i),$$

remain unchanged if we replace μ by $\alpha\mu$ for any $\alpha > 0$. Hence, multivariate OP Ensembles are invariant under $\mu \mapsto \alpha\mu$.

4.2.2 Soshnikov's key theorem

As stated previously, Theorem 2.1 has already been proven when $d = 1$ by Breuer and Duits [2013], as a consequence of a generalized strong Szegő theorem they obtained. The difficulty in proving Theorem 2.1 when $d \geqslant 2$ turns out to be of different nature than the one-dimensional setting. Indeed, the next result due to Soshnikov essentially states that the cumulants of order three and more of the linear statistic $\sum f(\mathbf{x}_i)$ decays to zero as $N \to \infty$ as soon as its variance goes to infinity. The latter condition turns out to be true if and only if $d \geqslant 2$. Thus, a CLT follows easily as soon as one can obtain asymptotic estimates on the variance. However, if obtaining such variance estimates is relatively easy when $d = 1$, the task becomes more involved in higher dimension.

More precisely, the general result [Soshnikov, 2002, Theorem 1] has the following consequence.

Theorem 4.6. (Soshnikov) *Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ form a multivariate OP Ensemble with respect to a given reference measure μ on I^d. Consider a sequence (f_N) of uniformly bounded and measurable real-valued functions on I^d satisfying, as $N \to \infty$,*

$$\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} f_N(\mathbf{x}_i)\right] \longrightarrow \infty, \tag{4.22}$$

and, for some $\delta > 0$,

$$\mathbb{E}\left[\sum_{i=1}^{N} \big|f_N(\mathbf{x}_i)\big|\right] = O\left(\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} f_N(\mathbf{x}_i)\right]^{\delta}\right). \tag{4.23}$$

Then, we have

$$\frac{\displaystyle\sum_{i=1}^{N} f_N(\mathbf{x}_i) - \mathbb{E}\left[\sum_{i=1}^{N} f_N(\mathbf{x}_i)\right]}{\sqrt{\mathbb{V}\mathrm{ar}\left[\displaystyle\sum_{i=1}^{N} f_N(\mathbf{x}_i)\right]}} \xrightarrow[N\to\infty]{law} \mathcal{N}(0,1).$$

4.2.3 Variance asymptotics

In order to prove Theorem 2.1 it is enough to show the following asymptotics.

Proposition 4.7. *Assume μ and $\mathbf{x}_1, \dots, \mathbf{x}_N$ satisfy the hypothesis of Theorem 2.1. Then, for every $f \in \mathscr{C}^1(I^d, \mathbb{R})$, we have*

$$\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} f(\mathbf{x}_i)\right] = \sigma_f^2. \tag{4.24}$$

Indeed, for any $d \geqslant 2$ and any $f \in \mathscr{C}^1(I^d, \mathbb{R})$, Corollary 4.4 and Proposition 4.7 imply (4.22) and (4.23) with $f_N = f$ and $\delta = d/(d-1)$. Thus, we can apply Theorem 4.6 to obtain Theorem 2.1.

Proposition 4.7 is the main technical result of this work. Consider the d-fold product of the equilibrium measure (4.1), namely the probability measure on I^d given by

$$\mu_{eq}^{\otimes d}(\mathrm{d}x) = \omega_{eq}^{\otimes d}(x)\mathrm{d}x, \qquad \omega_{eq}^{\otimes d}(x) = \prod_{j=1}^{d} \frac{1}{\pi\sqrt{1-x_j^2}} \mathbf{1}_{I^d}(x). \tag{4.25}$$

In our proof of Proposition 4.7, we start by investigating the limit (4.24) when $\mu = \mu_{eq}^{\otimes d}$, since algebraic identities are available for this reference measure. Then, we use comparison estimates to prove (4.24) in the general case.

4.3 A bound on the limiting variance

As stated in Proposition 2.4, one can bound the limiting variance σ_f^2 by a Dirichlet energy. Besides providing some control on the amplitude of σ_f^2, we will need this inequality in the proof of Proposition 4.7. We now give a proof for this proposition.

Proof of Proposition 2.4. Let $\mu_{sc}(\mathrm{d}x) = \pi^{-1}\sqrt{1-x^2}\mathbf{1}_I(x)\mathrm{d}x$ be the semi-circle measure. The associated orthonormal polynomials are the so-called Chebyshev polynomials of the second kind

$$U_k(\cos\theta) = \sqrt{2}\,\frac{\sin((k+1)\theta)}{\sin\theta}.$$

For any $1 \leqslant j \leqslant d$, define the measure

$$\nu^j(\mathrm{d}x) = \mu_{eq}(\mathrm{d}x_1)\cdots\mu_{eq}(\mathrm{d}x_{j-1})\mu_{sc}(\mathrm{d}x_j)\mu_{eq}(\mathrm{d}x_{j+1})\cdots\mu_{eq}(\mathrm{d}x_d),$$

so that the RHS of (2.8) becomes

$$\frac{1}{2}\sum_{j=1}^{d}\int_{I^d}\Big(\partial_j f(x)\Big)^2\,\nu^j(\mathrm{d}x)$$

For any $\boldsymbol{k} = (k_1,\ldots,k_d) \in \mathbb{N}^d$, set $T_{\boldsymbol{k}}(x) = T_{k_1}(x_1)\cdots T_{k_d}(x_d)$, where T_k are the Chebyshev polynomials (4.2), and let

$$V_{\boldsymbol{k}}^j(x) = T_{k_1}(x_1)\cdots T_{k_{j-1}}(x_{j-1})U_{k_j}(x_j)T_{k_{j+1}}(x_{j+1})\cdots T_{k_d}(x_d).$$

Thus, $(T_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{N}^d}$ and $(V_{\boldsymbol{k}}^j)_{\boldsymbol{k}\in\mathbb{N}^d}$ respectively form an orthonormal Hilbert basis of $L^2(\mu_{eq}^{\otimes d})$ and $L^2(\nu^j)$. Let $f \in \mathscr{C}^1(I^d,\mathbb{R})$, so that $f = \sum_{\boldsymbol{k}\in\mathbb{N}^d}\hat{f}(\boldsymbol{k})T_{\boldsymbol{k}}$ where $\hat{f}(\boldsymbol{k})$ is as in (2.7). Using the identity $T_k' = kU_{k-1}$, it comes

$$\partial_j f(x) = \sum_{\boldsymbol{k}\in\mathbb{N}^d} k_j\hat{f}(\boldsymbol{k})V_{\boldsymbol{k}}^j(x).$$

Then, Parseval's identity in $L^2(\nu^j)$ yields

$$\int_{I^d}\Big(\partial_j f(x)\Big)^2\,\nu^j(\mathrm{d}x) = \sum_{\boldsymbol{k}\in\mathbb{N}^d} k_j^2\hat{f}(\boldsymbol{k})^2.$$

Summing over $1 \leqslant j \leqslant d$, the RHS of (2.8) equals

$$\frac{1}{2}\sum_{j=1}^{d}\int_{I^d}\Big(\partial_j f(x)\Big)^2\nu^j(\mathrm{d}x)=\frac{1}{2}\sum_{\boldsymbol{k}\in\mathbb{N}^d}(k_1^2+\cdots+k_d^2)\hat{f}(\boldsymbol{k})^2, \tag{4.26}$$

from which Proposition 2.4 easily follows.

□

4.4 Assumptions of Theorem 2.7 and outline of the proof

We now discuss the assumptions and proof of Theorem 2.7.

Assume the reference measure μ is a product of d measures on I, and also that μ has a density ω. Then, Corollary 4.4 suggests that, as $N\to\infty$,

$$\frac{N}{K_N(x,x)}\approx\frac{\omega(x)}{\omega_{eq}^{\otimes d}(x)}. \tag{4.27}$$

This heuristic would yield for the variance of the estimator (2.11),

$$\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)}\right]\approx\frac{1}{N^2}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N}f(\mathbf{x}_j)\frac{\omega(\mathbf{x}_j)}{\omega_{eq}^{\otimes d}(\mathbf{x}_j)}\right]\approx\frac{\Omega_{f,\omega}^2}{N^{1+1/d}}\,, \tag{4.28}$$

where for the last approximation we used Proposition 4.7 with test function $f\omega/\omega_{eq}^{\otimes d}$, recalling $\Omega_{f,\omega}$ was defined in (2.15). This would essentially yield the CLT in Theorem 2.7 by applying Theorem 4.6 to $f_N(x)=Nf(x)/K_N(x,x)$. To make the approximation (4.28) rigorous, we will need extra regularity assumptions on μ.

First, regarding the approximation (4.27), we have the following result.

Theorem 4.8. (Totik) *Assume $\mu(\mathrm{d}x)=\omega(x)\mathrm{d}x$ with $\omega(x)=\omega_1(x_1)\cdots\omega_d(x_d)$, and that ω_j is continuous and positive on I. Then, for every $\varepsilon>0$, we have*

$$\frac{N}{K_N(x,x)}\xrightarrow[N\to\infty]{}\frac{\omega(x)}{\omega_{eq}(x)} \tag{4.29}$$

uniformly for $x\in I_\varepsilon^d$.

For a proof of Theorem 4.8 when $d=1$, see [Simon, 2011, Section 3.11] and references therein. The case $d\geqslant 2$ follows by the same arguments as in the proof of Corollary 4.4.

Remark 4.9. It is because of Theorem 4.8 that we restrict $\mathscr{C}^1(I^d,\mathbb{R})$ to the class $\mathscr{C}$ defined in (2.13) in the assumptions of Theorem 2.7. Unfortunately, there are examples of reference measures μ on I such that the convergence (4.29) is not uniform on the whole of I. However, in order to extend $\mathscr{C}$ to $\mathscr{C}^1(I^d,\mathbb{R})$ in the statement of Theorem 2.7, it would be enough to have $\sup_{x\in I_{\varepsilon_N}^d}|N/K_N(x,x)-\omega(x)/\omega_{eq}^{\otimes d}(x)|\to 0$ for some sequence ε_N going to zero as $N\to\infty$, but we were not able to locate such a result in the literature.

Next, the first approximation in (4.28) requires a control on the rate of change of $N/K_N(x,x)$. To this end, we introduce an extra assumption on the reference measure μ. More precisely, let us denote

$$\mathscr{D}_N(x,y)=\frac{N/K_N(x,x)-N/K_N(y,y)}{\|x-y\|}, \tag{4.30}$$

and further consider the sequence of measures on $I^d \times I^d$

$$Q_N(\mathrm{d}x, \mathrm{d}y) = \frac{1}{N^{1-1/d}} \|x-y\|^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y). \tag{4.31}$$

Our extra assumption on μ is then the following.

Assumption 1. *The measure μ satisfies*

$$\lim_{C\to\infty} \limsup_{\delta\to 0} \limsup_{N\to\infty} \iint_{I_\varepsilon^d \times I_\varepsilon^d,\ \|x-y\|\leqslant\delta} \mathbf{1}_{|\mathscr{D}_N(x,y)|>C}\ \mathscr{D}_N(x,y)^2\ Q_N(\mathrm{d}x, \mathrm{d}y) = 0. \tag{4.32}$$

In plain words, this means the squared rate of change $\mathscr{D}_N(x,y)^2$ is uniformly integrable with respect to the measures Q_N, at least on the restricted domain where $\|x-y\|$ is small enough and where x and y are not allowed to reach the boundary of I^d.

Remark 4.10. When $d=1$, Lemma 4.5 states that if μ is Nevai-class then Q_N converges weakly as $N\to\infty$ towards

$$L(\mathrm{d}x, \mathrm{d}y) = \frac{1}{2\pi^2} \frac{1-xy}{\sqrt{1-x^2}\sqrt{1-y^2}}\ \mathbf{1}_{I\times I}(x,y)\mathrm{d}x\mathrm{d}y.$$

Because the density of L is smooth within $I_\varepsilon \times I_\varepsilon$ for every $\varepsilon > 0$, one may at least heuristically understand that (4.32) reduces to the uniform integrability of $\mathscr{D}_N(x,y)^2$ with respect to the Lebesgue measure instead. In higher dimension, a similar guess can be made, but we do not pursue this reasoning here.

We now discuss sufficient conditions for (4.32) to hold true.

Remark 4.11. Since, for any $\kappa > 0$, we have

$$\mathbf{1}_{|\mathscr{D}_N(x,y)|>C}\ \mathscr{D}_N(x,y)^2 \leqslant \frac{1}{C^\kappa} \left|\mathscr{D}_N(x,y)\right|^{2+\kappa},$$

we see that condition (4.32) holds true as soon as, for every $\varepsilon > 0$, there exists $\kappa, \delta > 0$ satisfying

$$\limsup_{N\to\infty} \iint_{I_\varepsilon^d \times I_\varepsilon^d,\ \|x-y\|\leqslant\delta} \left|\mathscr{D}_N(x,y)\right|^{2+\kappa}\ Q_N(\mathrm{d}x, \mathrm{d}y) < \infty.$$

Namely, condition (4.32) is satisfied if the $L^{2+\kappa}(Q_N)$ norm of the rate of change of $N/K_N(x,x)$ is bounded, at least on the restricted domain where $\|x-y\|$ is small enough and x, y away from the boundary of I^d.

The following assumption is much stronger than Assumption 1, but it is easier to check in practice.

Assumption 2. *The measure μ satisfies*

(a) $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x$ *with ω positive and continuous on* $(-1,1)^d$.

(b) *For every $\varepsilon > 0$, the sequence*

$$\frac{1}{N} \sup_{x\in I_\varepsilon^d} \left\|\nabla K_N(x,x)\right\|$$

is bounded.

Indeed, thanks to the rough upper bound

$$|\mathscr{D}_N(x,y)| \leqslant \sup_{x\in I_\varepsilon^d} \|\nabla(N/K_N(x,x))\|, \qquad x,y\in \mathbf{1}_{I_\varepsilon^d},$$

we see that Assumption 1 holds true as soon as for every $\varepsilon>0$, $\sup_{x\in I_\varepsilon^d}\|\nabla(N/K_N(x,x))\|$ is bounded. Under Assumption 2(a), the latter follows from Assumption 2(b). Indeed, Theorem 4.8 and Assumption 2(a) together yield that for every $\varepsilon>0$, there exists $c>0$ independent of N such that $\frac{1}{N}K_N(x,x)>c$ for every $x\in I_\varepsilon^d$.

We conclude this section by proving that Jacobi measures (2.17) satisfy Assumption 2, which proves our Proposition 2.8. We start with a general lemma.

Lemma 4.12. *Assume the measures $\mu_1,\dots,\mu_d$ on I satisfy Assumption 2. Then the measure $\mu_1\otimes\cdots\otimes\mu_d$ on I^d satisfies Assumption 2.*

Proof. We decompose the set $\Gamma_N=\{\mathfrak{b}(0),\dots,\mathfrak{b}(N-1)\}\subset\mathbb{N}^d$ in a convenient way. To do so, set $\sigma_j(\boldsymbol{k})=(k_1,\dots,k_{j-1},k_{j+1},\dots,k_d)$ and say that $\boldsymbol{k}\sim\boldsymbol{\ell}$ if and only if $\sigma_j(\boldsymbol{k})=\sigma_j(\boldsymbol{\ell})$, that is, they have same coordinates except maybe the jth one. We denote by $[\boldsymbol{k}]$ the equivalence class under this relation. Set $N_j([\boldsymbol{k}])=\max\{\ell_j:\boldsymbol{\ell}\in[\boldsymbol{k}]\cap\Gamma_N\}$. Using the notation introduced in (4.7) and (4.8), it comes

$$\begin{aligned}
\partial_j K_N(x,x) &= 2\sum_{\boldsymbol{k}=\mathfrak{b}(0)}^{\mathfrak{b}(N-1)} \varphi_{k_j}^{(j)}(x_j)\frac{\mathrm{d}}{\mathrm{d}x_j}\varphi_{k_j}^{(j)}(x_j)\prod_{\alpha\neq j}\varphi_{k_\alpha}^{(\alpha)}(x_\alpha)^2\\
&= 2\sum_{[\boldsymbol{k}]\in\Gamma_N/\sim}\ \prod_{\alpha\neq j}\varphi_{k_\alpha}^{(\alpha)}(x_\alpha)^2\sum_{k_j=0}^{N_j([\boldsymbol{k}])}\varphi_{k_j}^{(j)}(x_j)\frac{\mathrm{d}}{\mathrm{d}x_j}\varphi_{k_j}^{(j)}(x_j)\\
&= \sum_{[\boldsymbol{k}]\in\Gamma_N/\sim}\ \prod_{\alpha\neq j}\varphi_{k_\alpha}^{(\alpha)}(x_\alpha)^2\,\frac{\mathrm{d}}{\mathrm{d}x_j}\left[K^{(j)}_{N_j([\boldsymbol{k}])+1}(x_j,x_j)\right].
\end{aligned}\tag{4.33}$$

Let now $\varepsilon>0$. Since μ_j satisfies Assumption 2, there exists $C>0$ such that for all $x\in I_\varepsilon$ and $n\in\mathbb{N}$,

$$\left|\frac{\mathrm{d}}{\mathrm{d}x}\left[K_n^{(j)}(x,x)\right]\right|\leqslant Cn.$$

Let $M=\lfloor N^{1/d}\rfloor$, so that $\Gamma_N\subset\mathcal{C}_{M+1}$, see (2.5). Thus, $N_j([\boldsymbol{k}])\leqslant M$ for all $\boldsymbol{k}\in\Gamma_N$. By (4.33),

$$\begin{aligned}
|\partial_j K_N(x,x)| &\leqslant C(M+1)\sum_{[\boldsymbol{k}]\in\Gamma_N/\sim}\ \prod_{\alpha\neq j}\varphi_{k_\alpha}^{(\alpha)}(x_\alpha)^2\\
&\leqslant C(M+1)\sum_{[\boldsymbol{k}]\in\mathcal{C}_{M+1}}\ \prod_{\alpha\neq j}\varphi_{k_\alpha}^{(\alpha)}(x_\alpha)^2\\
&= C(M+1)\prod_{\alpha\neq j}K_{M+1}^{(\alpha)}(x_\alpha,x_\alpha).
\end{aligned}$$

Hence

$$\frac{1}{N}|\partial_j K_N(x,x)|\leqslant C\frac{M+1}{M}\prod_{\alpha\neq j}\frac{1}{M}K_{M+1}^{(\alpha)}(x_\alpha,x_\alpha),$$

and Theorem 4.8 concludes. $\square$

Lemma 4.13. *Let $\alpha, \beta > -1$, then the measure*

$$(1-x)^{\alpha}(1+x)^{\beta}\mathbf{1}_I(x)\mathrm{d}x$$

satisfies Assumption 2.

Proof. Let $\varepsilon > 0$ be fixed. For convenience, Section 4.2.1 allows us to work with the probability measure

$$\mu^{(\alpha,\beta)}(\mathrm{d}x) = \omega^{(\alpha,\beta)}(x)\mathrm{d}x, \qquad \omega^{(\alpha,\beta)}(x) = \frac{1}{c_{\alpha,\beta}}(1-x)^{\alpha}(1+x)^{\beta},$$

where the normalization constant reads

$$c_{\alpha,\beta} = 2^{\alpha+\beta+1}\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)}$$

and Γ is the Euler Gamma function.

Denote by $(\varphi_n^{(\alpha,\beta)})_{n\in\mathbb{N}}$ the associated orthonormal polynomials. They satisfy

$$\varphi_n^{(\alpha,\beta)}(x) = \frac{P_n^{(\alpha,\beta)}(x)}{\sqrt{h_n^{(\alpha,\beta)}}},$$

where the $P_n^{(\alpha,\beta)}$'s are the Jacobi polynomials (we refer to [Szegő, 1974] for definitions and properties) and

$$h_n^{(\alpha,\beta)} = \|P_n^{(\alpha,\beta)}\|^2_{L^2(\mu^{(\alpha,\beta)})} = \frac{1}{n!(\alpha+\beta+2n+1)}\frac{\Gamma(\alpha+\beta+1)\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{\Gamma(\alpha+1)\Gamma(\beta+1)\Gamma(\alpha+\beta+n+1)},$$

and moreover

$$(\varphi_n^{(\alpha,\beta)})' = \frac{u_{\alpha,\beta}}{2}\sqrt{n(n+\alpha+\beta+1)}\;\varphi_{n-1}^{(\alpha+1,\beta+1)}, \qquad u_{\alpha,\beta} = \sqrt{\frac{(\alpha+\beta+1)(\alpha+\beta+2)}{(\alpha+1)(\beta+1)}}. \tag{4.34}$$

This yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}K_N(x,x) &= 2\sum_{k=1}^{N-1}\varphi_k^{(\alpha,\beta)}(x)(\varphi_k^{(\alpha,\beta)})'(x)\\ &= u_{\alpha,\beta}\sum_{k=1}^{N-1}\sqrt{k(k+\alpha+\beta+1)}\;\varphi_k^{(\alpha,\beta)}(x)\varphi_{k-1}^{(\alpha+1,\beta+1)}(x).\end{aligned} \tag{4.35}$$

By [Kuijlaars et al., 2004], we have as $k\to\infty$, uniformly in $x=\cos\theta\in I_\varepsilon$,

$$\varphi_k^{(\alpha,\beta)}(\cos\theta) = \sqrt{\frac{2}{\omega^{(\alpha,\beta)}(x)\pi\sqrt{1-x^2}}}\;\cos\Big(\big(k+\frac{1}{2}(\alpha+\beta+1)\big)\theta - \frac{\pi}{2}(\alpha+\frac{1}{2})\Big) + O(1/k). \tag{4.36}$$

As a consequence, we obtain in the same asymptotic regime,

$$\varphi_{k-1}^{(\alpha+1,\beta+1)}(\cos\theta) = \sqrt{\frac{2}{\omega^{(\alpha,\beta)}(x)\pi\sqrt{1-x^2}}}\;\sin\Big(\big(k+\frac{1}{2}(\alpha+\beta+1)\big)\theta - \frac{\pi}{2}(\alpha+\frac{1}{2})\Big)+O(1/k). \tag{4.37}$$

Now (4.36) implies that the $P_k^{(\alpha,\beta)}(x)$'s are bounded uniformly for $x \in I$ and $k \in \mathbb{N}$. Using moreover that $2\sin(u)\cos(u) = \sin(2u)$ and combining (4.35)–(4.37), we obtain for some $C_1, C_2 > 0$ that

$$\sup_{x \in I_\varepsilon} \left| \frac{\mathrm{d}}{\mathrm{d}x} K_N(x,x) \right| \leqslant C_1 \sup_{x \in I_\varepsilon} \left| \sum_{k=1}^{N-1} k \sin\Big((2k+\alpha+\beta+1)\theta - \pi(\alpha + \frac{1}{2}) \Big) \right| + C_2$$

where we recall the relation $x = \cos\theta$. Next, we write

$$\left| \sum_{k=1}^{N-1} k \sin\Big((2k+\alpha+\beta+1)\theta - \pi(\alpha + \frac{1}{2}) \Big) \right| \leqslant \left| \sum_{k=1}^{N-1} k e^{i((2k+\alpha+\beta+1)\theta - \pi(\alpha+\frac{1}{2}))} \right|^{1/2}$$

and then

$$\begin{aligned} \sum_{k=1}^{N-1} k e^{i((2k+\alpha+\beta+1)\theta - \pi(\alpha+\frac{1}{2}))} &= e^{i((\alpha+\beta+1)\theta - \pi(\alpha+\frac{1}{2}))} \sum_{k=1}^{N-1} k e^{2ik\theta} \\ &= \frac{1}{2i} e^{i((\alpha+\beta+1)\theta - \pi(\alpha+\frac{1}{2}))} \frac{\mathrm{d}}{\mathrm{d}\theta} \sum_{k=0}^{N-1} e^{2ik\theta} \\ &= \frac{1}{2i} e^{i((\alpha+\beta+1)\theta - \pi(\alpha+\frac{1}{2}))} \frac{\mathrm{d}}{\mathrm{d}\theta} \Big(\frac{e^{i(N-1)\theta} \sin(N\theta)}{\sin(\theta)} \Big). \end{aligned} \tag{4.38}$$

Since the absolute value of the right hand side of (4.38) is bounded by $CN/\sin^2(\theta)$ for some $C > 0$ independent on N and θ, the lemma follows.

□

Lemmas 4.12 and 4.13 combined yield Proposition 2.8.

5 CLT for multivariate OP Ensembles: proof of Theorem 2.1

In this section we prove Proposition 4.7. As explained in Section 4.2, Theorem 2.1 follows from this proposition.

5.1 A useful representation of the covariance

Lemma 5.1. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be random variables drawn from a multivariate OP Ensemble with reference measure μ. For any multivariate polynomials P, Q, we have*

$$\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N P(\mathbf{x}_i), \sum_{i=1}^N Q(\mathbf{x}_i) \right] = \sum_{n=0}^{N-1} \sum_{m=N}^{\infty} \langle P\varphi_n, \varphi_m \rangle \langle Q\varphi_n, \varphi_m \rangle,$$

where $\langle \cdot\,, \cdot \rangle$ refers to the scalar product of $L^2(\mu)$.

Proof. We start from the standard formula

$$\begin{aligned} &\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N P(\mathbf{x}_i), \sum_{j=1}^N Q(\mathbf{x}_i) \right] \\ &\qquad = \int P(x)Q(x)K_N(x,x)\mu(\mathrm{d}x) - \iint P(x)Q(y)K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y), \end{aligned} \tag{5.1}$$

which follows from (2.1)–(2.2) with $k = 1, 2$ and that $K_N(x, y)$ is symmetric. One the one hand, it follows from the definition of K_N that

$$\iint P(x)Q(y)K_N(x,y)^2\mu(\mathrm{d}x)\mu(\mathrm{d}y) = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\langle P\varphi_n, \varphi_m\rangle\langle Q\varphi_n, \varphi_m\rangle. \tag{5.2}$$

On the other hand, by using the decomposition (where the sum is finite since P is polynomial)

$$P\varphi_n = \sum_{m=0}^{\infty}\langle P\varphi_n, \varphi_m\rangle\varphi_m$$

together with the identity

$$\int P(x)Q(x)K_N(x,x)\mu(\mathrm{d}x) = \sum_{n=0}^{N-1}\langle P\varphi_n, Q\varphi_n\rangle,$$

we obtain

$$\int P(x)Q(x)K_N(x,x)\mu(\mathrm{d}x) = \sum_{n=0}^{N-1}\sum_{m=0}^{\infty}\langle P\varphi_n, \varphi_m\rangle\langle Q\varphi_n, \varphi_m\rangle. \tag{5.3}$$

Lemma 5.1 then follows by combining (5.1), (5.2) and (5.3). □

5.2 Covariance asymptotics: the Chebyshev case

We first investigate the case of the product measure $\mu_{eq}^{\otimes d}$, where μ_{eq} defined in (4.1) is the equilibrium measure of I. Recalling the definition (4.2), the multivariate Chebyshev polynomials

$$T_{\boldsymbol{k}}(x_1, \ldots, x_d) = T_{k_1}(x_1)\cdots T_{k_d}(x_d), \qquad \boldsymbol{k} = (k_1, \ldots k_d) \in \mathbb{N}^d, \tag{5.4}$$

satisfy the orthonormality conditions

$$\int T_{\boldsymbol{k}}(x)T_{\boldsymbol{\ell}}(x)\mu_{eq}^{\otimes d}(\mathrm{d}x) = \delta_{\boldsymbol{k}\boldsymbol{\ell}}, \qquad \boldsymbol{k}, \boldsymbol{\ell} \in \mathbb{N}^d.$$

We shall see that the family $(T_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{N}^d}$ diagonalizes the covariance structure associated with our point process.

Proposition 5.2. *Let $\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*$ be drawn according to the multivariate OP Ensemble associated with $\mu_{eq}^{\otimes d}$. Then, given any multi-indices $\boldsymbol{k} = (k_1, \ldots, k_d) \in \mathbb{N}^d$ and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d) \in \mathbb{N}^d$, we have*

$$\lim_{N\to\infty}\frac{1}{N^{1-1/d}}\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^{N}T_{\boldsymbol{k}}(\mathbf{x}_i^*), \sum_{i=1}^{N}T_{\boldsymbol{\ell}}(\mathbf{x}_i^*)\right] = \begin{cases}\frac{1}{2}(k_1 + \cdots + k_d) & \text{if } \boldsymbol{k} = \boldsymbol{\ell},\\ 0 & \text{if } \boldsymbol{k} \neq \boldsymbol{\ell}.\end{cases}$$

As a warm-up, let us first prove the proposition when $d = 1$.

Proof of Proposition 5.2 when $d = 1$. Throughout this proof, $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2(\mu_{eq}^{\otimes d})$. For every $k, \ell \in \mathbb{N}$, Lemma 5.1 provides

$$\mathbb{Cov}\left[\sum_{i=1}^{N} T_k(\mathbf{x}_i^*), \sum_{i=1}^{N} T_\ell(\mathbf{x}_i^*)\right] = \sum_{n=0}^{N-1}\sum_{m=N}^{\infty} \langle T_k T_n, T_m\rangle \langle T_\ell T_n, T_m\rangle. \tag{5.5}$$

First, notice that if k or ℓ is zero, then the right-hand side of (5.5) vanishes because $\langle T_n, T_m\rangle = \delta_{nm}$, and hence we can assume both k, ℓ are non-zero. Next, (4.2) yields the multiplication formula

$$T_k T_n = \frac{1}{\sqrt{2}} T_{n+k}\, \mathbf{1}_{kn\neq 0} + \left(\frac{1}{\sqrt{2}}\right)^{\mathbf{1}_{nk\neq 0}\mathbf{1}_{n\neq k}} T_{|n-k|}, \qquad k, n \in \mathbb{N}. \tag{5.6}$$

Combined with the orthonormality relations, this yields for any $n, m \in \mathbb{N}$ and $k > 0$

$$\langle T_k T_n, T_m\rangle = \frac{1}{\sqrt{2}} \mathbf{1}_{n+k=m}\mathbf{1}_{n\neq 0} + \left(\frac{1}{\sqrt{2}}\right)^{\mathbf{1}_{n\neq 0}\mathbf{1}_{n\neq k}} \mathbf{1}_{|n-k|=m}. \tag{5.7}$$

Hence, if $n, m \in \mathbb{N}$ moreover satisfy $n < m$ and $m > \max(k, \ell)$, then we have

$$\langle T_k T_n, T_m\rangle \langle T_\ell T_n, T_m\rangle = \frac{1}{2}\mathbf{1}_{n\neq 0}\mathbf{1}_{n+k=m}\mathbf{1}_{\ell+n=m}. \tag{5.8}$$

By plugging (5.8) into (5.5), we obtain for every $N > \max(k, \ell)$,

$$\begin{aligned}\mathbb{Cov}\left[\sum_{i=1}^{N} T_k(\mathbf{x}_i^*), \sum_{i=1}^{N} T_\ell(\mathbf{x}_i^*)\right] &= \frac{1}{2}\sum_{n=1}^{N-1}\sum_{m=N}^{\infty} \mathbf{1}_{k+n=m}\mathbf{1}_{\ell+n=m} \\ &= \frac{1}{2} k\, \mathbf{1}_{k=\ell},\end{aligned}$$

and the proposition follows when $d = 1$. $\square$

We now provide a proof for the higher-dimensional case. We also use the multiplication formula (5.6) in an essential way, although the setting is more involved. We recall that we introduced the bijection $\mathfrak{b} : \mathbb{N} \to \mathbb{N}^d$ associated with the graded lexicographic order in Section 2.1.3.

Proof of Proposition 5.2 when $d \geqslant 2$. Fix multi-indices $\boldsymbol{k} = (k_1, \ldots, k_d) \in \mathbb{N}^d$ and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d) \in \mathbb{N}^d$, and also set

$$S = \{j : \; k_j \neq 0\}, \qquad S' = \{j : \; \ell_j \neq 0\}.$$

Thanks to Lemma 5.1, we can write

$$\mathbb{Cov}\left[\sum_{i=1}^{N} T_{\boldsymbol{k}}(\mathbf{x}_i^*), \sum_{i=1}^{N} T_{\boldsymbol{\ell}}(\mathbf{x}_i^*)\right] = \sum_{(\boldsymbol{n},\boldsymbol{m})\in \mathbb{A}_N} \langle T_{\boldsymbol{k}} T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle \langle T_{\boldsymbol{\ell}} T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle, \tag{5.9}$$

where we introduced for convenience the set

$$\mathbb{A}_N = \Big\{(\boldsymbol{n}, \boldsymbol{m}) \in \mathbb{N}^d \times \mathbb{N}^d : \quad \boldsymbol{n} \leqslant \mathfrak{b}(N-1), \quad \boldsymbol{m} \geqslant \mathfrak{b}(N)\Big\}. \tag{5.10}$$

Next, using (5.4), the orthonormality relations for the Chebyshev polynomials and (5.7), we obtain

$$\begin{aligned}\langle T_{\boldsymbol{k}}T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle &= \langle T_{k_1}T_{n_1}, T_{m_1}\rangle_{L^2(\mu_{eq})}\cdots\langle T_{k_d}T_{n_d}, T_{m_d}\rangle_{L^2(\mu_{eq})} && (5.11)\\ &= \Big(\prod_{j\notin S}\mathbf{1}_{n_j=m_j}\Big)\sum_{P\subset S}\Big(\frac{1}{\sqrt{2}}\Big)^{|P|+|\{j\in S\setminus P:\, n_j\neq 0,\, n_j\neq k_j\}|}\\ &\qquad\times\Big(\prod_{j\in P}\mathbf{1}_{n_j+k_j=m_j}\mathbf{1}_{n_j\neq 0}\Big)\Big(\prod_{j\in S\setminus P}\mathbf{1}_{|n_j-k_j|=m_j}\Big), && (5.12)\end{aligned}$$

where $|A|$ stands for the cardinality of the set A.

First, notice that if $S\neq S'$ then the right hand side of (5.9) vanishes. Indeed, if $S\neq S'$, then there exists $\alpha\in\{1,\dots,d\}$ such that $k_\alpha=0$ and $\ell_\alpha\neq 0$ (or the other way around, but the argument is symmetric). It then follows from (5.12) that $\langle T_{\boldsymbol{k}}T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle$ vanishes except if $n_\alpha=m_\alpha$, and moreover that $\langle T_{\boldsymbol{\ell}}T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle$ vanishes except if $|n_\alpha\pm\ell_\alpha|=m_\alpha$. Since $\ell_\alpha\neq 0$, it holds $\langle T_{\boldsymbol{k}}T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle\langle T_{\boldsymbol{\ell}}T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle=0$ for every $(\boldsymbol{n},\boldsymbol{m})\in\mathbb{N}^d\times\mathbb{N}^d$, and our claim follows. Moreover, because $(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N$ yields the existence of $\alpha\in\{1,\dots,d\}$ such that $n_\alpha<m_\alpha$, one can see from (5.11) that $\langle T_{\boldsymbol{k}}T_{\boldsymbol{n}}, T_{\boldsymbol{m}}\rangle$ vanishes for every $(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N$ if $\boldsymbol{k}=(0,\dots,0)$. We henceforth assume that $S=S'\neq\varnothing$, for the covariance not to be trivial.

By combining (5.9) with (5.12), we obtain

$$\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N T_{\boldsymbol{k}}(\mathbf{x}_i^*), \sum_{i=1}^N T_{\boldsymbol{\ell}}(\mathbf{x}_i^*)\right] = \sum_{P,Q\subset S}\ \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N[P,Q]}\Big(\frac{1}{\sqrt{2}}\Big)^{\sigma[P,Q](\boldsymbol{n})},$$

where we introduced the subsets

$$\mathbb{A}_N[P,Q]=\left\{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N\ \middle|\ \begin{array}{lll} n_j+k_j=m_j, & n_j\neq 0, & \text{if } j\in P\\ n_j+\ell_j=m_j, & n_j\neq 0, & \text{if } j\in Q\end{array};\right.$$
$$\left.\begin{array}{ll}|n_j-k_j|=m_j, & \text{if } j\in S\setminus P\\ |n_j-\ell_j|=m_j, & \text{if } j\in S\setminus Q\end{array};\qquad n_j=m_j,\quad \text{if } j\notin S\right\} \tag{5.13}$$

and set for convenience

$$\begin{aligned}\sigma[P,Q](\boldsymbol{n}) = |P|+|Q|+|\{j\in S\setminus P:\ n_j\neq 0,\, n_j\neq k_j\}|&\\ +|\{j\in S\setminus Q:\ n_j\neq 0,\, n_j\neq \ell_j\}|.&\end{aligned}\tag{5.14}$$

Notice from (5.13) if $k_\alpha=\ell_\alpha\neq 0$ and $\mathbb{A}_N[P,Q]\neq\varnothing$ then necessarily $\alpha\in P\cap Q$ or $\alpha\in(S\setminus P)\cap(S\setminus Q)$. In particular, if $\boldsymbol{k}=\boldsymbol{\ell}$ then $\mathbb{A}_N[P,Q]=\varnothing$ unless $P=Q$. Thus,

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N T_{\boldsymbol{k}}(\mathbf{x}_i^*), \sum_{i=1}^N T_{\boldsymbol{\ell}}(\mathbf{x}_i^*)\right] &= \mathbf{1}_{\boldsymbol{k}=\boldsymbol{\ell}}\sum_{P\subset S}\ \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N[P,P]}\Big(\frac{1}{\sqrt{2}}\Big)^{\sigma[P,P](\boldsymbol{n})}\\ &\quad+\mathbf{1}_{\boldsymbol{k}\neq\boldsymbol{\ell}}\sum_{P,Q\subset S}\ \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N[P,Q]}\Big(\frac{1}{\sqrt{2}}\Big)^{\sigma[P,Q](\boldsymbol{n})}.\end{aligned}\tag{5.15}$$

Our goal is now to show that for every $P, Q \subset S$ the following holds true. As $N \to \infty$, if we assume $\boldsymbol{k} = \boldsymbol{\ell}$, then

$$\sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N[P,P]} \Big(\frac{1}{\sqrt{2}}\Big)^{\sigma[P,P](\boldsymbol{n})} = \Big(\frac{1}{2}\Big)^{|S|} \Big(\sum_{j\in P} k_j\Big) N^{1-1/d} + o(N^{1-1/d}), \tag{5.16}$$

and, if instead $\boldsymbol{k} \neq \boldsymbol{\ell}$, then

$$\big|\mathbb{A}_N[P,Q]\big| = o(N^{1-1/d}). \tag{5.17}$$

Since an easy rearrangement argument together with the definition of S yield

$$\begin{aligned}\sum_{P\subset S}\Big(\sum_{j\in P} k_j\Big) &= \frac{1}{2}\sum_{P\subset S}\Big(\sum_{j\in S} k_j\Big)\\ &= \Big(\sum_{j\subset S} k_j\Big)2^{|S|-1} = \Big(\sum_{j=1}^{d} k_j\Big)2^{|S|-1},\end{aligned}$$

Proposition 5.2 would then follow from (5.15)–(5.17).

We now turn to the proof of (5.16) and (5.17).

Truncated sets and consequences. Given distinct $\alpha_1, \dots, \alpha_p \in \{1, \dots, d\}$, we introduce the truncated sets

$$\begin{aligned}&\mathbb{A}_N[P,Q;\alpha_1,\dots,\alpha_p]\\ &\qquad = \mathbb{A}_N[P,Q] \cap \big\{n_{\alpha_1} \leqslant \max(k_{\alpha_1},\ell_{\alpha_1})\big\} \cap \dots \cap \big\{n_{\alpha_p} \leqslant \max(k_{\alpha_p},\ell_{\alpha_p})\big\}.\end{aligned} \tag{5.18}$$

By definition of $\mathfrak{b}$ and $\mathbb{A}_N$, if $N = M^d$ then $\mathbb{A}_{M^d} = \mathcal{C}_M \times (\mathbb{N}^d \setminus \mathcal{C}_M)$ where we recall

$$\mathcal{C}_M = \Big\{\boldsymbol{n} \in \mathbb{N}^d :\ 0 \leqslant n_1, \dots, n_d \leqslant M-1\Big\}. \tag{5.19}$$

Moreover, if for an arbitrary N we denote by $M = \lfloor N^{1/d} \rfloor$ the integer satisfying $M^d \leqslant N < (M+1)^d$, then $\mathfrak{b}(N) \in \mathcal{C}_{M+1}$ and thus, for any $(\boldsymbol{n}, \boldsymbol{m}) \in \mathbb{A}_N$, we have $\boldsymbol{n} \in \mathcal{C}_{M+1}$. As a consequence, for every $P, Q \subset S$, we have the rough upper bound $|\mathbb{A}_N[P,Q;\alpha_1,\dots,\alpha_p]| = \mathcal{O}(M^{d-p})$. In particular,

$$\big|\mathbb{A}_N[P,Q;\alpha_1,\dots,\alpha_p]\big| = o(N^{1-1/d}) \qquad \text{for every } p \geqslant 2. \tag{5.20}$$

In order to restrict ourselves to the easier setting where N is a power of d, we will use the following lemma. Its proof uses in a crucial way the graded lexicographic order we chose to equip $\mathbb{N}^d$ with, and it is deferred to the end of the present proof.

Lemma 5.3. *Assume $\boldsymbol{k} = \boldsymbol{\ell}$. For every $P \subset S$, $\alpha \in S \setminus P$ and for every $N > \max(k_1^d, \dots, k_d^d)$, we have*

(a) $\big|\mathbb{A}_N[P,P]\big| \leqslant \big|\mathbb{A}_{N+1}[P,P]\big|$,

(b) $\big|\mathbb{A}_N[P,P;\alpha]\big| \leqslant \big|\mathbb{A}_{N+1}[P,P;\alpha]\big|$.

We now provide a proof for (5.16).

The main contribution. Assume $\boldsymbol{k} = \boldsymbol{\ell}$. As a consequence of Lemma 5.3 (a), if we set $M = \lfloor N^{1/d} \rfloor$ then we have for every N large enough

$$\sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_{M^d}[P,P]} \left(\frac{1}{\sqrt{2}}\right)^{\sigma[P,P](\boldsymbol{n})} \leqslant \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_{N}[P,P]} \left(\frac{1}{\sqrt{2}}\right)^{\sigma[P,P](\boldsymbol{n})} \leqslant \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_{(M+1)^d}[P,P]} \left(\frac{1}{\sqrt{2}}\right)^{\sigma[P,P](\boldsymbol{n})}.$$

Thus, it is enough to prove that, as $M \to \infty$,

$$\sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_{M^d}[P,P]} \left(\frac{1}{\sqrt{2}}\right)^{\sigma[P,P](\boldsymbol{n})} = \left(\frac{1}{2}\right)^{|S|} \Big(\sum_{j\in P} k_j\Big) M^{d-1} + o(M^{d-1}), \tag{5.21}$$

in order to establish (5.16). To do so, for any $P \subset S$ and $\alpha \in S \setminus P$, we set

$$\mathbb{A}^*_{M^d}[P] = \mathbb{A}_{M^d}[P,P] \cap \big\{n_j > k_j \text{ for all } j \in S \setminus P\big\}, \tag{5.22}$$

$$\mathbb{A}^*_{M^d}[P;\alpha] = \mathbb{A}_{M^d}[P,P;\alpha] \cap \big\{n_j > k_j \text{ for all } j \in S \setminus (P \cup \{\alpha\})\big\}, \tag{5.23}$$

and use the following lemma; its proof is deferred to the end of the actual proof.

Lemma 5.4. *Assume $\boldsymbol{k} = \boldsymbol{\ell}$. For every $P \subset S$ and $\alpha \in S \setminus P$, we have as $M \to \infty$*

$$\big|\mathbb{A}^*_{M^d}[P]\big| = \Big(\sum_{j\in P} k_j\Big) M^{d-1} + o(M^{d-1}), \tag{5.24}$$

$$\big|\mathbb{A}^*_{M^d}[P;\alpha]\big| = o(M^{d-1}). \tag{5.25}$$

Next, as a consequence of the rough upper bound (5.20) and (5.25), we can write

$$\begin{aligned}
\big|\mathbb{A}_{M^d}[P,P]\big| &= \big|\mathbb{A}_{M^d}[P,P] \cap \big\{n_\alpha > k_\alpha \text{ for all } \alpha \in S \setminus P\big\}\big| \\
&\quad + \big|\mathbb{A}_{M^d}[P,P] \cap \big\{n_\alpha \leqslant k_\alpha \text{ for at least one } \alpha \in S \setminus P\big\}\big| \\
&= \big|\mathbb{A}^*_{M^d}[P]\big| + \sum_{\alpha\in S\setminus P} \big|\mathbb{A}^*_{M^d}[P;\alpha]\big| + o(M^{d-1}) \\
&= \big|\mathbb{A}^*_{M^d}[P]\big| + o(M^{d-1}).
\end{aligned} \tag{5.26}$$

Since for any $(\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}^*_{M^d}[P]$ we have $\sigma[P,P](\boldsymbol{n}) = 2|S|$, see (5.14) and (5.22), the estimate (5.21) follows from (5.26) and (5.24), and the proof of (5.16) is therefore complete.

We finally turn to the proof of (5.17).

The remaining contributions. Assume now that $\boldsymbol{k} \neq \boldsymbol{\ell}$. Since $\boldsymbol{k}$ and $\boldsymbol{\ell}$ have the same zero components, it follows that neither k_α nor ℓ_α is zero. Thus, (5.13) yields that if $k_\alpha \neq \ell_\alpha$ and $\mathbb{A}_N[P,Q] \neq \varnothing$, then either $\alpha \in P \cap (S \setminus Q)$ or $\alpha \in Q \cap (S \setminus P)$ and moreover, for any $(\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}_N[P,Q]$, we have

$$2n_\alpha = |k_\alpha - \ell_\alpha|, \qquad 2m_\alpha = k_\alpha + \ell_\alpha.$$

In particular $\mathbb{A}_{M^d}[P,Q] = \mathbb{A}_{M^d}[P,Q;\alpha]$. Thus, by virtue of the rough upper bound (5.20), we can assume in the proof of (5.17) that $\boldsymbol{k}$ and $\boldsymbol{\ell}$ differ by exactly one coordinate, namely there exists $\alpha \in \{1,\dots,d\}$ such that $k_\alpha \neq \ell_\alpha$ and $k_j = \ell_j$ for every $j \neq \alpha$. In this setting,

$\mathbb{A}_N[P,Q] \neq \varnothing$ then yields $P \setminus \{\alpha\} = Q \setminus \{\alpha\}$ and, if $(\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}_N[P,Q]$, then (n_α, m_α) satisfies the equations

$$\begin{cases} n_\alpha + k_\alpha = m_\alpha, & \ell_\alpha - n_\alpha = m_\alpha, \quad n_\alpha \neq 0, \quad \text{if } \alpha \in P \\ n_\alpha + \ell_\alpha = m_\alpha, & k_\alpha - n_\alpha = m_\alpha, \quad n_\alpha \neq 0, \quad \text{if } \alpha \in Q \end{cases}.$$

By weakening these constraints to

$$\begin{cases} |\ell_\alpha - n_\alpha| = m_\alpha, & n_\alpha \leqslant \ell_\alpha, \quad \text{if } \alpha \in P \\ |k_\alpha - n_\alpha| = m_\alpha, & n_\alpha \leqslant k_\alpha, \quad \text{if } \alpha \in Q \end{cases},$$

we obtain the upper bound

$$\big|\mathbb{A}_N[P,Q]\big| \leqslant \begin{cases} \big|\mathbb{A}_N^{(\boldsymbol{\ell},\boldsymbol{\ell})}[Q,Q;\alpha]\big| & \text{if } \alpha \in P \\ \big|\mathbb{A}_N^{(\boldsymbol{k},\boldsymbol{k})}[P,P;\alpha]\big| & \text{if } \alpha \in Q \end{cases} \tag{5.27}$$

where $\mathbb{A}_N^{(\boldsymbol{k},\boldsymbol{\ell})}[P,Q;\alpha]$ is defined as in (5.13),(5.18) but we emphasized the multi-indices $\boldsymbol{k}, \boldsymbol{\ell}$ which are involved. By setting $M = \lfloor N^{1/d} \rfloor + 1$, we obtain from Lemma 5.3 (b), the rough upper bound (5.20) and (5.23),(5.25) that, as $N \to \infty$,

$$\begin{aligned} &\big|\mathbb{A}_N^{(\boldsymbol{k},\boldsymbol{k})}[P,P;\alpha]\big| \\ &\quad \leqslant \big|\mathbb{A}_{M^d}^{(\boldsymbol{k},\boldsymbol{k})}[P,P;\alpha]\big| \\ &\quad = \big|\mathbb{A}_{M^d}^{(\boldsymbol{k},\boldsymbol{k})}[P,P;\alpha] \cap \big\{ n_j > k_j \text{ for all } j \in S \setminus (P \cup \{\alpha\}) \big\}\big| + o(N^{1-1/d}) \\ &\quad = o(N^{1-1/d}), \end{aligned} \tag{5.28}$$

and similarly

$$\big|\mathbb{A}_N^{(\boldsymbol{\ell},\boldsymbol{\ell})}[Q,Q;\alpha]\big| = o(N^{1-1/d}). \tag{5.29}$$

By combining (5.27)–(5.29), we have finally proved (5.17) and the proof of Proposition 5.2 is thus complete, up to the proof of Lemmas 5.3 and 5.4. $\square$

We now provide proofs for the remaining lemmas.

Proof of Lemma 5.3. Let $(\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}_N[P,P]$. Then $(\boldsymbol{n},\boldsymbol{m}) \notin \mathbb{A}_{N+1}[P,P]$ if and only if $\boldsymbol{m} = \mathfrak{b}(N)$. Since $\mathfrak{b}(N) \in \mathcal{C}_{M+1} \setminus \mathcal{C}_M$, where $M = \lfloor N^{1/d} \rfloor$ and $\mathcal{C}_M$ has been introduced in (5.19), there exists $j \in \{1,\dots,d\}$ such that $\mathfrak{b}(N)_j = M$; let j_* be the smallest j satisfying this property. Notice also $\boldsymbol{n} \in \mathcal{C}_{M+1}$ and $\boldsymbol{m} \in \mathbb{N}^d \setminus \mathcal{C}_M$. As soon as $M > \max(k_1,\dots,k_d)$, that we assume from now, the equality $\boldsymbol{m} = \mathfrak{b}(N)$ can only happen if $j_* \notin S \setminus P$. Indeed, if $j_* \in S \setminus P$, then $m_{j_*} = |n_{j_*} - k_{j_*}| \leqslant \max(n_{j_*} - 1, k_{j_*}) \leqslant M - 1$. As a consequence,

$$\big|\mathbb{A}_N[P,P]\big| \leqslant \big|\mathbb{A}_{N+1}[P,P]\big| \quad \text{if } j_* \in S \setminus P.$$

Next, assume that $j_* \in P$ or $j_* \notin S$. We claim that if we set

$$\tilde{m}_j = \begin{cases} m_j + k_j & \text{if } j \in P, \\ |m_j - k_j| & \text{if } j \in S \setminus P, \\ m_j & \text{if } j \notin S, \end{cases} \tag{5.30}$$

then $(\boldsymbol{m},\tilde{\boldsymbol{m}}) \in \mathbb{A}_{N+1}[P,P] \setminus \mathbb{A}_N[P,P]$. This would show in particular that

$$\big|\mathbb{A}_N[P,P]\big| \leqslant \big|\mathbb{A}_{N+1}[P,P]\big| \qquad \text{if either } j_* \in P \text{ or } j_* \notin S,$$

and thus complete the proof of (a). That $(\boldsymbol{m},\tilde{\boldsymbol{m}}) \in \mathbb{A}_{N+1}[P,P] \setminus \mathbb{A}_N[P,P]$ is by construction obvious provided one can show $\tilde{\boldsymbol{m}} \geqslant \mathfrak{b}(N+1)$.

If $j_* \in P$, then we have

$$\tilde{m}_{j_*} = m_{j_*} + k_{j_*} = M + k_{j_*} > M$$

and thus $\tilde{\boldsymbol{m}} \in \mathbb{N}^d \setminus \mathcal{C}_{M+1}$. As a consequence, there exists $m \geqslant (M+1)^d$ such that $\tilde{\boldsymbol{m}} = \mathfrak{b}(m)$ and, since $N+1 \leqslant (M+1)^d$, we have shown $\tilde{\boldsymbol{m}} \geqslant \mathfrak{b}(N+1)$.

If $j_* \notin S$, we argue by contradiction and assume $\tilde{\boldsymbol{m}} \leqslant \mathfrak{b}(N) = \boldsymbol{m}$. We shall see this is not compatible with the graded lexicographic order. Indeed, since by construction $\tilde{\boldsymbol{m}} \neq \boldsymbol{m}$ and $\boldsymbol{n} \neq \boldsymbol{m}$ (because $\boldsymbol{k} \neq (0,\dots,0)$ by assumption), we actually have $\tilde{\boldsymbol{m}} < \boldsymbol{m}$ and $\boldsymbol{n} < \boldsymbol{m}$. Because $j_* \notin S$ by assumption, we moreover have $n_{j_*} = m_{j_*} = \tilde{m}_{j_*} = M$ and thus $\boldsymbol{n},\boldsymbol{m},\tilde{\boldsymbol{m}} \in \mathcal{C}_{M+1} \setminus \mathcal{C}_M$. As a consequence, $\boldsymbol{n} <_{\text{lex}} \boldsymbol{m}$ and $\tilde{\boldsymbol{m}} <_{\text{lex}} \boldsymbol{m}$ in the lexicographic order. This means there exists $\gamma \in \{1,\dots,d\}$ such that $n_i = m_i$ for every $i < \gamma$ and $n_\gamma < m_\gamma$, and equivalently $i \notin S$ when $i < \gamma$ and $\gamma \in P$. Similarly, there exists $\eta \in \{1,\dots,d\}$ such that $\tilde{m}_i = m_i$ for every $i < \eta$ and $\tilde{m}_\eta < m_\eta$, and thus $i \notin S$ when $i < \eta$ and $\eta \in S \setminus P$. But this is impossible and thus $\tilde{\boldsymbol{m}} \geqslant \mathfrak{b}(N+1)$, which completes the proof of (a).

Part (b) is proved by following the exact same line of arguments; in this setting one should also check that if $(\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}_N[P,P;\alpha]$ then $m_\alpha \leqslant k_\alpha$ in order to show $(\boldsymbol{m},\tilde{\boldsymbol{m}})$ (with $\tilde{\boldsymbol{m}}$ defined in (5.30)) actually belongs to $\mathbb{A}_{N+1}[P,P;\alpha]$. Recalling $\alpha \in S \setminus P$ by assumption this is clear, indeed $m_\alpha = |n_\alpha - k_\alpha|$ together with $n_\alpha \leqslant k_\alpha$ yield $m_\alpha = k_\alpha - n_\alpha \leqslant k_\alpha$. $\square$

Proof of Lemma 5.4. To prove (a), fix $P \subset S$ and assume $M > \max(k_1,\dots,k_d)$. It follows from the definitions (5.13),(5.22) that

$$\mathbb{A}^*_{M^d}[P] = \left\{ (\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}_{M^d} \;\middle|\; \begin{array}{rll} n_j + k_j = m_j, & n_j \neq 0, & \text{if } j \in P \\ n_j - k_j = m_j, & n_j > k_j, & \text{if } j \in S \setminus P \\ n_j = m_j, & & \text{if } j \notin S \end{array} \right\}. \tag{5.31}$$

Recall that $\mathbb{A}_{M^d} = \mathcal{C}_M \times (\mathbb{N}^d \setminus \mathcal{C}_M)$ where $\mathcal{C}_M$ is defined in (5.19). Clearly, if we set

$$\mathcal{C}_M[P] = \Big\{ \boldsymbol{n} \in \mathcal{C}_M : \text{ there exists } \boldsymbol{m} \in \mathbb{N}^d \setminus \mathcal{C}_M,\ (\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}^*_{M^d}[P] \Big\} \tag{5.32}$$

then $(\boldsymbol{n},\boldsymbol{m}) \mapsto \boldsymbol{n}$ is a bijection from $\mathbb{A}^*_{M^d}[P]$ to $\mathcal{C}_M[P]$.

We claim that if for any $p \in P$ we set

$$\mathcal{C}^{(p)}_M[P] = \left\{ \boldsymbol{n} \in \mathcal{C}_M \;\middle|\; \begin{array}{rl} 1 \leqslant n_j \leqslant M-1 & \text{if } j \in P \\ k_j < n_j \leqslant M-1 & \text{if } j \in S \setminus P \\ M - k_p \leqslant n_p \leqslant M-1 & \end{array} \right\}, \tag{5.33}$$

then we have

$$\mathcal{C}_M[P] = \bigcup_{p \in P} \mathcal{C}^{(p)}_M[P]. \tag{5.34}$$

Indeed, let $\boldsymbol{n} \in \mathcal{C}_M[P]$. By definition there exists $\boldsymbol{m} \in \mathbb{N}^d \setminus \mathcal{C}_M$ such that $(\boldsymbol{n},\boldsymbol{m}) \in \mathbb{A}^*_{M^d}[P]$. This provides, see (5.31), that $1 \leqslant n_j \leqslant M-1$ if $j \in P$, that $k_j < n_j \leqslant M-1$ if $j \in S \setminus P$, and the existence of $p \in \{1, \dots, d\}$ satisfying $m_p \geqslant M$. Since $\boldsymbol{n} \in \mathcal{C}_M$ then $n_p \leqslant M-1$ and thus $p \in P$ because otherwise $m_p \leqslant n_p$. Together with the equation $n_p + k_p = m_p$ this finally yields that $M - k_p \leqslant n_p \leqslant M-1$, namely $\boldsymbol{n} \in \mathcal{C}_M^{(p)}[P]$ for some $p \in P$. As for the reverse inclusion, if $\boldsymbol{n} \in \mathcal{C}_M^{(\alpha)}[P]$ for some $\alpha \in P$ then set

$$m_j = \begin{cases} n_j + k_j & \text{if } j \in P, \\ n_j - k_j & \text{if } j \in S \setminus P, \\ n_j & \text{if } j \notin S, \end{cases}$$

and observe that $\boldsymbol{m} \in \mathbb{N}^d \setminus \mathcal{C}_M$ since $m_p \geqslant M$ and $n_j - k_j \geqslant 0$ for every $j \in S \setminus P$. Thus, since clearly $(\boldsymbol{n}, \boldsymbol{m}) \in \mathbb{A}^*_{M^d}[P]$, we have shown $\boldsymbol{n} \in \mathcal{C}_M[P]$ and (5.34) is proved.

Next, since for every distinct $p_1, \dots, p_q \in P$ the definition (5.33) yields

$$\left| \mathcal{C}_M^{(p_1)}[P] \cap \cdots \cap \mathcal{C}_M^{(p_q)}[P] \right| = k_{p_1} \cdots k_{p_q} M^{d-q} + \mathcal{O}(M^{d-q-1}),$$

then (a) follows from (5.34) and the inclusion-exclusion principle.

We now turn to (b) and fix $\alpha \in S \setminus P$. Let $\mathcal{C}_M^{(d)}$ be the d-dimensional discrete hypercube of length M defined as in (5.19). We then set

$$\mathbb{A}^{(d-1)}_{M^{d-1}} = \mathcal{C}_M^{(d-1)} \times (\mathbb{N}^{d-1} \setminus \mathcal{C}_M^{(d-1)})$$

and introduce

$$\mathbb{A}^*_{M^{d-1}}[P]^{(d-1)} = \left\{ (\boldsymbol{n}, \boldsymbol{m}) \in \mathbb{A}^{(d-1)}_{M^{d-1}} \;\middle|\; \begin{array}{lll} n_j + k_j = m_j, & n_j \neq 0, & \text{if } j \in P \\ n_j - k_j = m_j, & n_j > k_j, & \text{if } j \in S \setminus (P \cup \{\alpha\}) \\ \quad\; n_j = m_j, & & \text{if } j \notin S \setminus \{\alpha\} \end{array} \right\}.$$

The statement (a) of the lemma applied in dimension $d-1$ then provides

$$\left| \mathbb{A}^*_{M^{d-1}}[P]^{(d-1)} \right| = o(M^{d-1}). \tag{5.35}$$

Consider the map $\mathfrak{p} : \mathbb{N}^d \to \mathbb{N}^{d-1}$ defined by

$$\mathfrak{p}(n_1, \dots, n_d) = (n_1, \dots, n_{\alpha-1}, n_{\alpha+1}, \dots, n_d).$$

Let $(\boldsymbol{n}, \boldsymbol{m}) \in \mathbb{A}^*_{M^d}[P; \alpha]$. Since $\boldsymbol{m} \in \mathbb{N}^d \setminus \mathcal{C}_M^{(d)}$ and $m_\alpha \leqslant k_\alpha < M$, there exists $j \neq \alpha$ such that $m_j \geqslant M$. It follows that $(\mathfrak{p}(\boldsymbol{n}), \mathfrak{p}(\boldsymbol{m})) \in \mathcal{C}_M^{(d-1)} \times (\mathbb{N}^{d-1} \setminus \mathcal{C}_M^{(d-1)})$ and thus $(\mathfrak{p}(\boldsymbol{n}), \mathfrak{p}(\boldsymbol{m})) \in \mathbb{A}^*_{M^{d-1}}[P]^{(d-1)}$. As a consequence, we have the upper bound

$$\left| A^*_{M^d}[P; \alpha] \right| \leqslant k_\alpha \left| (\mathfrak{p} \times \mathfrak{p}) \big(A^*_{M^d}[P; \alpha] \big) \right| \leqslant k_\alpha \left| A^*_{M^{d-1}}[P]^{(d-1)} \right|$$

and thus (b) follows from (5.35).

□

5.3 Covariance asymptotics: the general case

We now extend Proposition 5.2 to the general setting of measures satisfying the assumptions of Theorem 2.1. More precisely, we prove the following.

Proposition 5.5. *Let $\mu = \mu_1 \otimes \cdots \otimes \mu_d$, where the μ_j's are Nevai-class probability measures on I. Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and $\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*$ be random variables drawn from the multivariate OP Ensembles with respective reference measures μ and $\mu_{eq}^{\otimes d}$. Then, given any polynomial functions P, Q on $\mathbb{R}^d$,*

$$\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \left| \mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N P(\mathbf{x}_i), \sum_{j=1}^N Q(\mathbf{x}_i)\right] - \mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N P(\mathbf{x}_i^*), \sum_{j=1}^N Q(\mathbf{x}_i^*)\right]\right| = 0. \quad (5.36)$$

For the proof of the proposition, we use a few ingredients from the Step 2 of the proof of Lemma 4.5 to which we refer the reader to.

Proof of Proposition 5.5. By linearity, it is enough to prove the proposition with $P(x) = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ and $Q(x) = x_1^{\beta_1} \cdots x_d^{\beta_d}$ for any fixed $\alpha_1, \beta_1, \ldots, \alpha_d, \beta_d \in \mathbb{N}$. Lemma 5.1 then provides

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N P(\mathbf{x}_i), \sum_{j=1}^N Q(\mathbf{x}_i)\right] &= \sum_{n=0}^{N-1} \sum_{m=N}^{\infty} \langle x_1^{\alpha_1} \cdots x_d^{\alpha_d} \varphi_n, \varphi_m\rangle_{L^2(\mu)} \langle x_1^{\beta_1} \cdots x_d^{\beta_d} \varphi_n, \varphi_m\rangle_{L^2(\mu)} \\ &= \sum_{(\boldsymbol{n},\boldsymbol{m})\in \mathbb{A}_N} \prod_{j=1}^d \langle x^{\alpha_j} \varphi_{n_j}^{(j)}, \varphi_{m_j}^{(j)}\rangle_{L^2(\mu_j)} \langle x^{\beta_j} \varphi_{n_j}^{(j)}, \varphi_{m_j}^{(j)}\rangle_{L^2(\mu_j)}\end{aligned} \quad (5.37)$$

where we recall that

$$\mathbb{A}_N = \left\{ \big(\mathfrak{b}(n), \mathfrak{b}(m)\big) :\ n \leqslant N-1,\ m \geqslant N \right\} \subset \mathbb{N}^d \times \mathbb{N}^d.$$

In particular, by choosing $\mu = \mu_{eq}^{\otimes d}$ in (5.37), we obtain

$$\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N P(\mathbf{x}_i^*), \sum_{j=1}^N Q(\mathbf{x}_i^*)\right] = \sum_{(\boldsymbol{n},\boldsymbol{m})\in \mathbb{A}_N} \prod_{j=1}^d \langle x^{\alpha_j} T_{n_j}, T_{m_j}\rangle_{L^2(\mu_{eq})} \langle x^{\beta_j} T_{n_j}, T_{m_j}\rangle_{L^2(\mu_{eq})}. \quad (5.38)$$

Thus, by combining (5.37) and (5.38), we see that, if we set for convenience

$$\begin{aligned}E(\boldsymbol{n},\boldsymbol{m}) = \Bigg| &\prod_{j=1}^d \langle x^{\alpha_j} \varphi_{n_j}^{(j)}, \varphi_{m_j}^{(j)}\rangle_{L^2(\mu_j)} \langle x^{\beta_j} \varphi_{n_j}^{(j)}, \varphi_{m_j}^{(j)}\rangle_{L^2(\mu_j)} \\ &- \prod_{j=1}^d \langle x^{\alpha_j} T_{n_j}, T_{m_j}\rangle_{L^2(\mu_{eq})} \langle x^{\beta_j} T_{n_j}, T_{m_j}\rangle_{L^2(\mu_{eq})}\Bigg|,\end{aligned} \quad (5.39)$$

then proving the proposition amounts to showing that

$$\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \sum_{(\boldsymbol{n},\boldsymbol{m})\in \mathbb{A}_N} E(\boldsymbol{n},\boldsymbol{m}) = 0. \quad (5.40)$$

Next, for every $j \in \{1, \ldots, d\}$, the three-term recurrence relation reads

$$x\varphi_n^{(j)} = a_n^{(j)} \varphi_{n+1}^{(j)} + b_n^{(j)} \varphi_n^{(j)} + a_{n-1}^{(j)} \varphi_{n-1}^{(j)}, \qquad n \geqslant 0, \quad (5.41)$$

where we set $a_{-1}^{(j)} = 0$. As in Step 2 of the proof of Lemma 4.5, we complete the sequences of recurrence coefficients $(a_n^{(j)})_{n\in\mathbb{N}}$ and $(b_n^{(j)})_{n\in\mathbb{N}}$ introduced into sequences $(a_n^{(j)})_{n\in\mathbb{Z}}$, $(b_n^{(j)})_{n\in\mathbb{Z}}$, where we set $a_n^{(j)} = b_n^{(j)} = 0$ for every $n < 0$. We thus obtain the representations

$$\langle x^\alpha \varphi_{n_j}^{(j)}, \varphi_{m_j}^{(j)} \rangle_{L^2(\mu_j)} = \mathbf{1}_{|n_j - m_j| \leqslant \alpha} \sum_{\gamma:(0,n_j-m_j)\to(\alpha,0)} \prod_{e\in\gamma} \omega(e)_{\{(a_{n+m_j}^{(j)}),(b_{n+m_j}^{(j)})\}}, \tag{5.42}$$

and

$$\langle x^\alpha T_{n_j}, T_{m_j} \rangle_{L^2(\mu_{eq})} = \mathbf{1}_{|n_j - m_j| \leqslant \alpha} \sum_{\gamma:(0,n_j-m_j)\to(\alpha,0)} \prod_{e\in\gamma} \omega(e)_{\{(a_{n+m_j}^{*}),(b_{n+m_j}^{*})\}}. \tag{5.43}$$

Since the measures μ_j are Nevai-class by assumption, we have $a_n^{(j)} - a_n^* \to 0$ and $b_n^{(j)} - b_n^* \to 0$ as $n \to \infty$ for every $j \in \{1, \dots, d\}$. Notice that for every n_j, the right-hand side of (5.42) is a polynomial function of $a_{m_j-\alpha}^{(j)}, b_{m_j-\alpha}^{(j)}, \dots, a_{m_j+\alpha}^{(j)}, b_{m_j+\alpha}^{(j)}$ and does not depend on any other recurrence coefficients. Thus, we obtain for every fixed $\alpha \in \mathbb{N}$,

$$\sup_{n_j\in\mathbb{N}} \left| \langle x^\alpha \varphi_{n_j}^{(j)}, \varphi_{m_j}^{(j)} \rangle_{L^2(\mu_j)} - \langle x^\alpha T_{n_j}, T_{m_j} \rangle_{L^2(\mu_{eq})} \right| \xrightarrow[m_j\to\infty]{} 0. \tag{5.44}$$

Moreover, we see from (5.39), (5.42) and (5.43) that $E(\boldsymbol{n}, \boldsymbol{m}) = 0$ except when $|n_j - m_j| \leqslant \min(\alpha_j, \beta_j)$ for every $j \in \{1, \dots, d\}$. We then split the set of contributing indices into two subsets,

$$\begin{aligned}
\mathbb{A}_N^* &= \Big\{ (\boldsymbol{n}, \boldsymbol{m}) \in \mathbb{A}_N : \; |n_j - m_j| \leqslant \min(\alpha_j, \beta_j) \Big\} \cap \Big\{ m_j \geqslant N^{1/2d} \quad \text{for every } j \Big\}, \\
\mathbb{A}_N^0 &= \Big\{ (\boldsymbol{n}, \boldsymbol{m}) \in \mathbb{A}_N : \; |n_j - m_j| \leqslant \min(\alpha_j, \beta_j) \Big\} \cap \Big\{ m_j < N^{1/2d} \quad \text{for at least one } j \Big\}.
\end{aligned}$$

It then follows from (5.44) that

$$\lim_{N\to\infty} \sup_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*} \big| E(\boldsymbol{n}, \boldsymbol{m}) \big| = 0 \tag{5.45}$$

and that there exists $C > 0$ independent on N satisfying

$$\sup_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^0} \big| E(\boldsymbol{n}, \boldsymbol{m}) \big| \leqslant \sup_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N} \big| E(\boldsymbol{n}, \boldsymbol{m}) \big| \leqslant C. \tag{5.46}$$

Next, we write

$$\begin{aligned}
&\frac{1}{N^{1-1/d}} \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N} E(\boldsymbol{n}, \boldsymbol{m}) \\
&= \frac{1}{N^{1-1/d}} \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*} E(\boldsymbol{n}, \boldsymbol{m}) + \frac{1}{N^{1-1/d}} \sum_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^0} E(\boldsymbol{n}, \boldsymbol{m}) \\
&\leqslant \frac{|\mathbb{A}_N^*|}{N^{1-1/d}} \sup_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*} \big| E(\boldsymbol{n}, \boldsymbol{m}) \big| + \frac{|\mathbb{A}_N^0|}{N^{1-1/d}} \sup_{(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^0} \big| E(\boldsymbol{n}, \boldsymbol{m}) \big|.
\end{aligned} \tag{5.47}$$

and claim that we have

$$\limsup_{N\to\infty} \frac{|\mathbb{A}_N^*|}{N^{1-1/d}} < \infty, \tag{5.48}$$

and, moreover,

$$\lim_{N\to\infty}\frac{\left|\mathbb{A}_N^0\right|}{N^{1-1/d}}=0. \tag{5.49}$$

Together with (5.45)–(5.46), this would prove (5.40) and thus the proposition.

We finally prove (5.48) and (5.49) in order to complete the proof of the proposition. Let us set $\kappa_j=\max(\alpha_j,\beta_j)$ for convenience. Clearly,

$$\begin{aligned}\left|\mathbb{A}_N^*\right| &= \Big|\bigcup_{\boldsymbol{n}\in\mathbb{N}^d}\Big\{\boldsymbol{m}\in\mathbb{N}^d:\ (\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*\Big\}\Big|\\ &\leqslant \max_{\boldsymbol{n}\in\mathbb{N}^d}\Big|\Big\{\boldsymbol{m}\in\mathbb{N}^d:\ (\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*\Big\}\Big|\times\Big|\Big\{\boldsymbol{n}\in\mathbb{N}^d:\ (\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^* \text{ for some } \boldsymbol{m}\in\mathbb{N}^d\Big\}\Big|.\end{aligned} \tag{5.50}$$

First, since $|n_j-m_j|\leqslant\kappa_j$ for every j as soon as $(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*$, we have the upper bound

$$\max_{\boldsymbol{n}\in\mathbb{N}^d}\Big|\Big\{\boldsymbol{m}\in\mathbb{N}^d:\ (\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*\Big\}\Big|\leqslant\prod_{j=1}^d(2\kappa_j+1). \tag{5.51}$$

Next, set $M=\lfloor N^{1/d}\rfloor$ so that $M^d\leqslant N<(M+1)^d$. If $(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^*$, then it satisfies $\boldsymbol{n}\in\mathcal{C}_{M+1}$ and $\boldsymbol{m}\in\mathbb{N}^d\setminus\mathcal{C}_M$, where $\mathcal{C}_M$ has been introduced in (5.19). Namely, it holds that $0\leqslant n_j\leqslant M$ for every j and there exists j_0 such that $m_{j_0}\geqslant M$. Together with $|n_{j_0}-m_{j_0}|\leqslant\kappa_{j_0}$, this yields $M-\kappa_{j_0}\leqslant n_{j_0}\leqslant M$ and thus provides the upper bound

$$\Big|\Big\{\boldsymbol{n}\in\mathbb{N}^d:\ (\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^* \text{ for some } \boldsymbol{m}\in\mathbb{N}^d\Big\}\Big|\leqslant(\max_{j_0=1}^d\kappa_{j_0}+1)(M+1)^{d-1}. \tag{5.52}$$

By combining (5.50)–(5.52), we have proved (5.48). The proof of (5.49) is similar. More precisely, the only difference is that if $(\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^{(0)}$, then there exists j_1 such that $m_{j_1}<\sqrt{N^{\frac1d}}<\sqrt{M+1}$. Notice that necessarily $j_1\neq j_0$. Using moreover that $|n_{j_1}-m_{j_1}|\leqslant\kappa_{j_1}$, we obtain the upper bound

$$\Big|\Big\{\boldsymbol{n}\in\mathbb{N}^d:\ (\boldsymbol{n},\boldsymbol{m})\in\mathbb{A}_N^{(0)} \text{ for some } \boldsymbol{m}\in\mathbb{N}^d\Big\}\Big|\leqslant(\max_{\substack{j_0=1\\ j_0\neq j_1}}^d\kappa_{j_0}+1)(\kappa_{j_1}+\sqrt{M+1}\,)(M+1)^{d-2}$$

in place of (5.52), and (5.49) follows. □

5.4 Extension to $\mathscr{C}^1$ functions and conclusion

We consider a reference measure μ satisfying the assumptions of Theorem 2.1 and let $\mathbf{x}_1,\dots,\mathbf{x}_N$ be the associated multivariate OP Ensemble. For any d-multivariate polynomial P, we can write $P=\sum_{\boldsymbol{k}\in\mathbb{N}^d}\hat{P}(\boldsymbol{k})T_{\boldsymbol{k}}$, where the latter sum is finite. As a consequence of Propositions 5.2 and 5.5, we then obtain

$$\begin{aligned}&\lim_{N\to\infty}\frac{1}{N^{1-1/d}}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N P(\mathbf{x}_i)\right]\\ &=\sum_{\boldsymbol{k},\boldsymbol{\ell}\in\mathbb{N}^d}\hat{P}(\boldsymbol{k})\hat{P}(\boldsymbol{\ell})\lim_{N\to\infty}\frac{1}{N^{1-1/d}}\,\mathbb{C}\mathrm{ov}\left[\sum_{i=1}^N T_{\boldsymbol{k}}(\mathbf{x}_i),\sum_{i=1}^N T_{\boldsymbol{\ell}}(\mathbf{x}_i)\right]\\ &=\frac12\sum_{\boldsymbol{k}=(k_1,\dots,k_d)\in\mathbb{N}^d}(k_1+\cdots+k_d)\hat{P}(\boldsymbol{k})^2=\sigma_P^2.\end{aligned} \tag{5.53}$$

Therefore, we have proven Proposition 4.7 provided we restrict the test functions to polynomials. We finally extend this result to $\mathscr{C}^1$ test functions, and thus complete the proof of this proposition, by means of a density argument.

First, a standard computation yields

$$\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} f(\mathbf{x}_i)\right] = \frac{1}{2}\iint (f(x)-f(y))^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y). \tag{5.54}$$

This indeed follows from (2.1)–(2.2) with $k = 1,2$, and that $K_N(x,y)$ is a symmetric reproducing kernel.

Now, for any $f \in \mathscr{C}^1(I^d,\mathbb{R})$, we set

$$\|f\|_{\mathrm{Lip}} = \sup_{x\in I^d}\|\nabla f(x)\|, \tag{5.55}$$

so that $|f(x)-f(y)| \leqslant \|f\|_{\mathrm{Lip}}\|x-y\|$ for every $x \neq y$. If we consider the monomials defined by

$$e_j(x_1,\ldots,x_d) = x_j, \tag{5.56}$$

then formula (5.54) yields

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} f(\mathbf{x}_i)\right] &= \frac{1}{2}\iint (f(x)-f(y))^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y)\\
&\leqslant \|f\|_{\mathrm{Lip}}^2 \sum_{j=1}^{d} \frac{1}{2}\iint (e_j(x)-e_j(y))^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y)\\
&= \|f\|_{\mathrm{Lip}}^2 \sum_{j=1}^{d} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} e_j(\mathbf{x}_i)\right]
\end{aligned}$$

and, as a consequence of (5.53),

$$\limsup_{N\to\infty}\frac{1}{N^{1-1/d}}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} f(\mathbf{x}_i)\right] \leqslant C\|f\|_{\mathrm{Lip}}^2, \qquad C = \sum_{j=1}^{d}\sigma_{e_j}^2. \tag{5.57}$$

Proposition 2.4 also provides the upper bound

$$\sigma_f^2 \leqslant \frac{1}{2}\|f\|_{\mathrm{Lip}}^2. \tag{5.58}$$

Next, Theorem 5 of Peet [2009] yields the existence of a sequence of multivariate polynomials $(P_\varepsilon)_{\varepsilon>0}$ such that $\|P_\varepsilon - f\|_{\mathrm{Lip}} \leqslant \varepsilon$, and hence

$$\limsup_{N\to\infty}\frac{1}{N^{1-1/d}}\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} f(\mathbf{x}_i) - \sum_{i=1}^{N} P_\varepsilon(\mathbf{x}_i)\right] \leqslant C\varepsilon^2, \qquad \text{and} \qquad \sigma_{f-P_\varepsilon}^2 \leqslant \frac{\varepsilon^2}{2}. \tag{5.59}$$

Since $(X,Y) \mapsto \mathbb{C}\mathrm{ov}(X,Y)$ is a symmetric positive bilinear form, it satisfies the Cauchy-Schwartz inequality, and thus the triangle inequality $\mathbb{V}\mathrm{ar}(X+Y)^{1/2} \leqslant \mathbb{V}\mathrm{ar}(X)^{1/2} + \mathbb{V}\mathrm{ar}(Y)^{1/2}$ holds true, which in turn yields the inequality

$$\left|\mathbb{V}\mathrm{ar}(X)^{1/2} - \mathbb{V}\mathrm{ar}(Y)^{1/2}\right| \leqslant \mathbb{V}\mathrm{ar}(X-Y)^{1/2}. \tag{5.60}$$

For the same reason, the limiting variance satisfies $|\sigma_f - \sigma_g| \leqslant \sigma_{f-g}$. As a consequence, by taking $X = \sum f(\mathbf{x}_i)$ and $Y = \sum P_\varepsilon(\mathbf{x}_i)$ in (5.60), and using these two inequalities together with (5.53) and (5.59), we obtain by letting $N \to \infty$ and then $\varepsilon \to 0$ that

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N f(\mathbf{x}_i)\right] &= \lim_{\varepsilon\to 0}\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N P_\varepsilon(\mathbf{x}_i)\right] \\ &= \lim_{\varepsilon\to 0} \sigma_{P_\varepsilon}^2 \\ &= \sigma_f^2\end{aligned}$$

and the proof of Proposition 4.7 is therefore complete.

6 Monte Carlo with DPPs: proof of Theorem 2.7

The aim of this section is to prove the following variance decay.

Proposition 6.1. *Assume $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x$ with ω positive and $\mathscr{C}^1$ on $(-1,1)^d$. Assume further that μ satisfies Assumption 1. For every $f \in \mathscr{C}$, we have*

$$\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N \frac{Nf(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)} - \sum_{i=1}^N \frac{\omega(x)f(\mathbf{x}_i)}{\omega_{eq}^{\otimes d}(\mathbf{x}_i)}\right] = 0.$$

Before proving Proposition 6.1, we argue that it implies Theorem 2.7. Indeed, (5.60) then implies that

$$\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N \frac{Nf(\mathbf{x}_i)}{K_N(\mathbf{x}_i,\mathbf{x}_i)}\right] = \lim_{N\to\infty} \frac{1}{N^{1-1/d}} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N \frac{\omega(x)f(\mathbf{x}_i)}{\omega_{eq}^{\otimes d}(\mathbf{x}_i)}\right] = \Omega_{f,\omega}^2\,,$$

the last equality following from Theorem 2.1. Now Theorem 4.6 applies with $f_N(x) = Nf(x)/K_N(x,x)$ to yield Theorem 2.7.

From now on, we fix $f \in \mathscr{C}$. It is thus a $\mathscr{C}^1$ function and there exists $\varepsilon > 0$ so that $\mathrm{Supp}(f) \subset I_\varepsilon^d$. If we set for convenience

$$E_N(x) = \frac{N}{K_N(x,x)} - \sum_{i=1}^N \frac{\omega(x)}{\omega_{eq}^{\otimes d}(x)}, \qquad x \in I^d,$$

then Theorem 4.8 yields $\|fE_N\|_\infty = \sup_{I^d} |fE_N| \to 0$ as $N \to \infty$.

In order to prove Proposition 6.1, we start with the formula coming from (5.54),

$$\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^N f(\mathbf{x}_i)E_N(\mathbf{x}_i)\right] = \frac{1}{2}\iint \big(fE_N(x) - fE_N(y)\big)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y)$$

and split the integral in several terms that we shall analyse separately.

6.1 The off-diagonal contribution

Given any $\delta > 0$, we first consider the contribution

$$\frac{1}{2} \iint_{\|x-y\|>\delta} \big(fE_N(x) - fE_N(y)\big)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y). \tag{6.1}$$

By rough estimates, we obtain

$$\begin{aligned}
(6.1) &\leqslant \|fE_N\|_\infty^2 \iint_{\|x-y\|>\delta} K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y) \\
&\leqslant \frac{1}{\delta^2} \|fE_N\|_\infty^2 \sum_{j=1}^{d} \iint (x_j - y_j)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y) \\
&\leqslant \frac{2}{\delta^2} \|fE_N\|_\infty^2 \sum_{j=1}^{d} \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{N} e_j(\mathbf{x}_i)\right],
\end{aligned}$$

where the monomials e_j were defined in (5.56). As a consequence, using Proposition 4.7 and that $\|fE_N\|_\infty \to 0$ as $N \to \infty$, we get

$$\lim_{N\to\infty} \frac{1}{N^{1-1/d}} \iint_{\|x-y\|>\delta} \big(fE_N(x) - fE_N(y)\big)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y) = 0 \tag{6.2}$$

for every $\delta > 0$.

6.2 The diagonal contribution

By (6.2), it is sufficient to show

$$\limsup_{\delta\to 0} \limsup_{N\to\infty} \frac{1}{N^{1-1/d}} \iint_{\|x-y\|\leqslant\delta} \big(fE_N(x) - fE_N(y)\big)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y) = 0 \tag{6.3}$$

in order to complete the proof of Proposition 6.1.

Set for convenience

$$\mathscr{D}_g(x,y) = \frac{g(x) - g(y)}{\|x-y\|}, \qquad x, y \in I^d, \tag{6.4}$$

so that, $|\mathscr{D}_g(x,y)| \leqslant \|g\|_{\mathrm{Lip}}$. For every $\delta > 0$ small enough, we have for any x, y satisfying $\|x - y\| \leqslant \delta$,

$$\begin{aligned}
\mathscr{D}_{fE_N}(x,y)^2 &= (\mathscr{D}_f(x,y)E_N(x) + \mathscr{D}_{E_N}(x,y)f(y))^2 \\
&\leqslant 2\mathscr{D}_f(x,y)^2 E_N(x)^2 + 2\mathscr{D}_{E_N}(x,y)^2 f(y)^2 \\
&\leqslant 2\|f\|_{\mathrm{Lip}}^2 \|\mathbf{1}_{I^d_{\varepsilon/2}} E_N\|_\infty^2 + 2\|f\|_\infty^2 \mathscr{D}_{E_N}(x,y)^2 \; \mathbf{1}_{I^d_{\varepsilon/2} \times I^d_{\varepsilon/2}}(x,y).
\end{aligned}$$

Indeed, notice that if $x \in I^d_\varepsilon$ or $y \in I^d_\varepsilon$, then $x, y \in I^d_{\varepsilon/2}$ for every $\delta > 0$ small enough. Since f is by assumption supported on I^d_ε, we know that $\mathscr{D}_{fE_N}(x,y)$ vanishes outside of $I^d_{\varepsilon/2} \times I^d_{\varepsilon/2}$. This is the reason for the presence of $\mathbf{1}_{I^d_{\varepsilon/2}}$ in the last inequality.

With the notation Q_N introduced in (4.31), we thus obtain

$$\begin{aligned}
&\frac{1}{2N^{1-1/d}} \iint_{\|x-y\|\leqslant\delta} \big(fE_N(x) - fE_N(y)\big)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y) \\
&= \frac{1}{2} \iint_{\|x-y\|\leqslant\delta} \mathscr{D}_{fE_N}(x,y)^2 Q_N(\mathrm{d}x,\mathrm{d}y) \\
&\leqslant \|f\|_{\mathrm{Lip}}^2 \|\mathbf{1}_{I^d_{\varepsilon/2}} E_N\|_\infty^2 \iint_{\|x-y\|\leqslant\delta} Q_N(\mathrm{d}x,\mathrm{d}y) \\
&\qquad + \|f\|_\infty^2 \iint_{I^d_{\varepsilon/2}\times I^d_{\varepsilon/2},\ \|x-y\|\leqslant\delta} \mathscr{D}_{E_N}(x,y)^2 Q_N(\mathrm{d}x,\mathrm{d}y).
\end{aligned} \tag{6.5}$$

Moreover, because ω is $\mathscr{C}^1$ on $I^d_{\varepsilon/2}$ by assumption, and because $\omega_{eq}^{\otimes d}$ is also $\mathscr{C}^1$ and positive there, one similarly has, for every $x, y \in I^d_{\varepsilon/2}$,

$$\mathscr{D}_{E_N}(x,y)^2 \leqslant 2\mathscr{D}_N(x,y)^2 + 2\mathscr{D}_{\omega/\omega_{eq}^{\otimes d}}(x,y)^2 \leqslant 2\mathscr{D}_N(x,y)^2 + 2\|\omega/\omega_{eq}^{\otimes d}\|_{\mathrm{Lip}}, \tag{6.6}$$

where $\mathscr{D}_N$ is defined in (4.30).

Next, we have for every $C > 0$,

$$\begin{aligned}
&\iint_{I^d_{\varepsilon/2}\times I^d_{\varepsilon/2},\ \|x-y\|\leqslant\delta} \mathscr{D}_N(x,y)^2 Q_N(\mathrm{d}x,\mathrm{d}y) \\
&\leqslant C^2 \iint_{\|x-y\|\leqslant\delta} Q_N(\mathrm{d}x,\mathrm{d}y) \\
&\qquad + \iint_{I^d_{\varepsilon/2}\times I^d_{\varepsilon/2},\ \|x-y\|\leqslant\delta} \mathbf{1}_{|\mathscr{D}_N(x,y)|>C}\ \mathscr{D}_N(x,y)^2 Q_N(\mathrm{d}x,\mathrm{d}y).
\end{aligned} \tag{6.7}$$

We now make use of the following lemma, the proof of which is deferred to Section 6.3.

Lemma 6.2.

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \iint_{\|x-y\|\leqslant\delta} Q_N(\mathrm{d}x,\mathrm{d}y) = 0. \tag{6.8}$$

As a consequence, (6.5), (6.6), and (6.7) together yield, for every $C > 0$,

$$\begin{aligned}
&\limsup_{\delta\to 0} \limsup_{N\to\infty} \frac{1}{2N^{1-1/d}} \iint_{\|x-y\|\leqslant\delta} \big(fE_N(x) - fE_N(y)\big)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y) \\
&\leqslant 2\|f\|_\infty^2 \limsup_{\delta\to 0} \limsup_{N\to\infty} \iint_{I^d_{\varepsilon/2}\times I^d_{\varepsilon/2},\ \|x-y\|\leqslant\delta} \mathscr{D}_N(x,y)^2 \mathbf{1}_{|\mathscr{D}_N(x,y)|>C}\, Q_N(\mathrm{d}x,\mathrm{d}y).
\end{aligned} \tag{6.9}$$

Assumption 1 allows us to conclude the proof of Proposition 6.1, up to the proof of Lemma 6.2.

6.3 Proof of Lemma 6.2

Proof. First,

$$\iint_{\|x-y\|\leqslant\delta} Q_N(\mathrm{d}x,\mathrm{d}y) \leqslant \frac{1}{N^{1-1/d}} \sum_{j=1}^{d} \iint_{|x_j-y_j|\leqslant\delta} (x_j - y_j)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y). \tag{6.10}$$

We fix $j \in \{1, \ldots, d\}$ and use the notation of the proof of Lemma 4.12. It comes

$$K_N(x,y) = \sum_{[\boldsymbol{k}] \in \Gamma_N/\sim} K^{(j)}_{N_j([\boldsymbol{k}])+1}(x_j, y_j) \prod_{\alpha \neq j} \varphi_{k_\alpha}(x_\alpha)\varphi_{k_\alpha}(y_\alpha).$$

Squaring, integrating and using the orthonormality relations,

$$\begin{aligned}
&\iint_{|x_j - y_j| \leqslant \delta} (x_j - y_j)^2 K_N(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y) \\
&= \sum_{[\boldsymbol{k}],[\boldsymbol{\ell}] \in \Gamma_N/\sim} \mathbf{1}_{\sigma(\boldsymbol{k})=\sigma(\boldsymbol{\ell})} \\
&\qquad \times \iint_{|x_j - y_j| \leqslant \delta} (x_j - y_j)^2 K^{(j)}_{N_j([\boldsymbol{k}])+1}(x_j, y_j) K^{(j)}_{N_j([\boldsymbol{\ell}])+1}(x_j, y_j) \mu_j(\mathrm{d}x_j)\mu_j(\mathrm{d}y_j) \\
&= \sum_{[\boldsymbol{k}] \in \Gamma_N/\sim} \iint_{|x_j - y_j| \leqslant \delta} (x_j - y_j)^2 K^{(j)}_{N_j([\boldsymbol{k}])+1}(x_j, y_j)^2 \mu_j(\mathrm{d}x_j)\mu_j(\mathrm{d}y_j). \qquad (6.11)
\end{aligned}$$

Recall $M = \lfloor N^{1/d} \rfloor$ and $\mathcal{C}_M \subset \Gamma_N \subset \mathcal{C}_{M+1}$. By definition of $\mathfrak{b}$ we have, for every $1 \leqslant m \leqslant M-2$,

$$|\{[\boldsymbol{k}] \in \Gamma_N/\sim :\ N_j([\boldsymbol{k}]) = m\}| \leqslant d M^{d-2}.$$

Notice also that (4.13) yields

$$\iint (x_j - y_j)^2 K^{(j)}_m(x_j, y_j)^2 \mu_j(\mathrm{d}x_j)\mu_j(\mathrm{d}y_j) = 2a_m^2,$$

which is bounded for every m since $a_m \to 1/2$ by assumption. Now

$$\begin{aligned}
(6.11) \;&=\; \left[\sum_{\substack{[\boldsymbol{k}] \in \Gamma_N/\sim \\ N_j([\boldsymbol{k}]) < \sqrt{M}}} + \sum_{\substack{[\boldsymbol{k}] \in \Gamma_N/\sim \\ N_j([\boldsymbol{k}]) \geqslant \sqrt{M}}} \right] \iint_{|x_j - y_j| \leqslant \delta} (x_j - y_j)^2 K^{(j)}_{N_j([\boldsymbol{k}])+1}(x_j, y_j)^2 \mu_j(\mathrm{d}x_j)\mu_j(\mathrm{d}y_j) \\
&\leqslant\; \mathcal{O}(M^{d-2+1/2}) \\
&\qquad + M^{d-1} \max_{\sqrt{M} \leqslant m \leqslant M} \iint_{|x_j - y_j| \leqslant \delta} (x_j - y_j)^2 K^{(j)}_{m+1}(x_j, y_j)^2 \mu_j(\mathrm{d}x_j)\mu_j(\mathrm{d}y_j). \; (6.12)
\end{aligned}$$

Moreover, Lemma 4.5 yields

$$\max_{\sqrt{M} \leqslant m \leqslant M} \iint_{|x_j - y_j| \leqslant \delta} (x_j - y_j)^2 K^{(j)}_{m+1}(x_j, y_j)^2 \mu_j(\mathrm{d}x_j)\mu_j(\mathrm{d}y_j) \to \iint_{|x-y| \leqslant \delta} L(x,y)\mathrm{d}x\mathrm{d}y$$

as $M \to \infty$. Combined with (6.10)–(6.12), we obtain

$$\limsup_{N \to \infty} \iint_{\|x-y\| \leqslant \delta} Q_N(\mathrm{d}x, \mathrm{d}y) \leqslant d \iint_{|x-y| \leqslant \delta} L(x,y)\mathrm{d}x\mathrm{d}y.$$

Finally, since L is integrable, the lemma follows by letting $\delta \to 0$. $\square$

The proof of Proposition 6.1 is therefore complete.

7 Discussion and perspectives

As detailed in Remark 2.6, Monte Carlo with DPPs is a stochastic counterpart to Gaussian quadrature, introduced in Section 1.1. Compared to the Monte Carlo methods introduced in Section 1.2, and 1.3, Theorem 2.9 is an importance sampling procedure, with negatively correlated importance samples. This negative correlation results in a variance reduction that impacts the decay rate of the variance. Loosely speaking, this is reminiscent of the interesting kernel density approach to importance sampling of Delyon and Portier [2016] described in Section 1.2. Our rates are better for equivalent smoothness in $d = 1$, but for $d > 1$, the comparison is less clear. Further investigation is thus needed to properly compare our methods. In terms of sampling cost, parallelizable approaches such as [Delyon and Portier, 2016] scale very well with N, while tackling the cost of sampling DPPs is a priority for future research, see Section 2.4.

Monte Carlo with DPPs is also reminiscent of randomized quasi-Monte Carlo methods such as scrambled nets, introduced in Section 1.3. The important difference is that randomness and discrepancy are tied in our DPP proposal. The similarities with QMC are an interesting lead for future research. In particular, fast constructions of nets in QMC [Dick et al., 2013] could yield fast sampling algorithms for DPPs.

Although not traditionally known as QMC methods, Bayesian quadrature is also familiar with our contribution. O'Hagan [1991] remarked that if we put a Gaussian process prior [Rasmussen and Williams, 2006] over the integrand, then the conditional of its integral given N evaluations is a univariate Gaussian, with a closed-form mean and variance. Sequentially minimizing this posterior variance by picking up nodes yields kernel herding [Chen et al., 2010] or Bayesian quadrature [Huszár and Duvenaud, 2012]. The biggest downside of such algorithms is the requirement for some key closed-form integrals [Bach et al., 2012, Section 4.3], and the absence of a general proof that the convergence rate is better than Monte Carlo [Bach et al., 2012, Briol et al., 2015]. As pointed out in Section 2.4, sampling DPPs is related to sequentially maximizing the variance of a Gaussian process. If a detailed connection is made with kernel herding, our results could potentially lead to better understanding of the behaviour of kernel herding. Conversely, the efficient Frank-Wolfe optimization procedures given for herding by Bach et al. [2012], Briol et al. [2015] could influence fast sampling algorithms for DPPs.

Acknowledgments: AH would like to thank Bernhard Beckermann, Jeff Geronimo and Kurt Johansson for useful and stimulating discussions. This work started while RB and AH were respectively at University of Oxford and KTH Royal Institute of Technology, funded by EPSRC grant number EP/I017909/1 and grant KAW 2010.0063 from the Knut and Alice Wallenberg Foundation. RB also acknowledges support from ANR Bnpsi and both authors acknowledge support from CNRS through PEPS JCJC DppMc.

References

Y. Ameur, H. Hedenmalm, and N. Makarov. Fluctuations of eigenvalues of random normal matrices. *Duke Math. J.*, 159(1):31–81, 2011.

Y. Ameur, H. Hedenmalm, and N. Makarov. Random normal matrices and Ward identities. *Ann. Probab.*, 43(3):1157–1201, 2015.

F. Bach, S. Lacoste-Julien, and G. Obozinski. On the equivalence between herding and conditional gradient algorithms. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2012.

R. Berman. Sharp asymptotics for Toeplitz determinants and convergence towards the Gaussian free field on Riemann surfaces. *Int. Math. Res. Not. IMRN*, (22):5031–5062, 2012.

R. J. Berman. Bergman kernels for weighted polynomials and weighted equilibrium measures of $\mathbb{C}^n$. *Indiana Univ. Math. J.*, 58(4):1921–1946, 2009a.

R. J. Berman. Bergman kernels and equilibrium measures for line bundles over projective manifolds. *Amer. J. Math.*, 131(5):1485–1524, 2009b.

R. J. Berman. Kähler-Einstein metrics, canonical random point processes and birational geometry. *Preprint arXiv: 1307.3634*, 2013.

R. J. Berman. Determinantal point processes and fermions on complex manifolds: large deviations and bosonization. *Comm. Math. Phys.*, 327(1):1–47, 2014.

I. Bogaert. Iteration-free computation of Gauss–Legendre quadrature nodes and weights. *SIAM J. Sci. Comput.*, 36(3):A1008–1026, 2014.

H. Brass and K. Petras. *Quadrature theory.* Amer. Math. Soc., Providence, RI, 2011.

J. Breuer and M. Duits. Central limit theorems for biorthogonal ensembles and asymptotics of recurrence coefficients. *J. Amer. Math. Soc., to appear. Preprint ArXiv:1309.6224*, 2013.

J. Breuer and M. Duits. Universality of mesocopic fluctuation in orthogonal polynomial ensembles. *to appear in Comm. Math. Phys. Preprint arXiv:1411.5205*, 2014.

F.-X. Briol, C. Oates, M. Girolami, and M. A. Osborne. Frank-Wolfe Bayesian quadrature: Probabilistic integration with theoretical guarantees. In *Advances In Neural Information Processing Systems (NIPS)*, pages 1162–1170, 2015.

Y. Chen, M. Welling, and A. Smola. Super-samples from kernel herding. In *Proceedings of the conference on Uncertainty in Artificial Intelligence (UAI)*, 2010.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes.* Springer, 2nd edition, 2003.

P. J. Davis and P. Rabinowitz. *Methods of numerical integration.* Academic Press, New York, 2nd edition, 1984.

B. Delyon and F. Portier. Integral approximation by kernel smoothing. *to appear in Bernoulli*, 2016.

P. Diaconis and S. N. Evans. Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.*, 353(7):2615–2633, 2001.

J. Dick and F. Pilichshammer. *Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo Integration.* Cambridge University Press, 2010.

J. Dick, F. Y. Kuo, and I. H. Sloan. High-dimensional integration: the quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 2013.

R. Douc, É. Moulines, and D. Stoffer. *Nonlinear time series*. Chapman-Hall, 2014.

C. F. Gauss. *Methodus nova integralium valores per approximationem inveniendi*. Heinrich Dietrich, Göttingen, 1815.

W. Gautschi. *Orthogonal polynomials: computation and approximation*. Oxford University Press, USA, 2004.

W. Gautschi. How sharp is Bernstein's inequality for Jacobi polynomials. *Electronic Transactions on Numerical Analysis*, 36:1–8, 2009.

W. Gautschi and R. S. Varga. Error bounds for Gaussian quadrature of analytic functions. *Siam J. Numer. Anal.*, 20(6):1170–1186, 1983.

A. Glaser, X. Liu, and V. Rokhlin. A fast algorithm for the calculation of the roots of special functions. *SIAM Journal on Scientific Computing*, 29(4):1420–1438, 2007.

G. H. Golub and C. F. Van Loan. *Matrix computations*. JHU Press, 2012.

N. Hale and A. Townsend. Fast and accurate computation of Gauss-Legendre and Gauss-Jacobi quadrature nodes and weights. *SIAM J. Sci. Comput.*, 35(2):A652–A674, 2013.

A. Hardy. Average characteristic polynomials of determinantal point processes. *Ann. Inst. H. Poincare Probab. Statist.*, 51(1):283–303, 2015.

J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág. Determinantal processes and independence. *Probab. Surv.*, 3:206–229, 2006.

F. Huszár and D. Duvenaud. Optimally-weighted herding is Bayesian quadrature. In *Uncertainty in Artificial Intelligence (UAI)*, 2012.

K. Johansson. On random matrices from the compact classical groups. *Ann. Math.*, 145 (3):519–545, 1997.

K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

K. Johansson. *Random matrices and determinantal processes, Mathematical Statistical Physics*. Elsevier B.V. Amsterdam, 2006.

K. Johansson and G. Lambert. Gaussian and non-Gaussian fluctuations for mesoscopic linear statistics in determinantal processes. *Preprint arXiv:1504.06455*, 2015.

W. Köning. Orthogonal polynomial ensembles in probability theory. *Probab. Surv.*, 2 (385–447), 2005.

T. Kriecherbauer and M. Shcherbina. Fluctuations of eigenvalues of matrix models and their applications. *Preprint arXiv:1003.6121*, 2010.

A. B. J. Kuijlaars, K. T.-R. McLaughlin, W. Van Assche, and M. Vanlessen. The Riemann-Hilbert approach to strong asymptotics for orthogonal polynomials on [1,1]. *Adv. Math.*, 188(2):337–398, 2004.

G. Lambert. Mesoscopic fluctuations for unitary invariant ensembles. *Preprint arXiv:1510.03641*, 2015a.

G. Lambert. CLT for biorthogonal ensembles and related combinatorial identities. *Preprint arXiv:1511.06121*, 2015b.

F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference: Extended version. *Preprint arXiv: 1205.4818*, 2014.

F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 77: 853–877, 2015.

R. Lyons. Determinantal probability measures. *Publ. Math. Inst. Hautes Etudes Sci.*, 98: 167–212, 2003.

J. Møller, M. Nielsen, E. Porcu, and E. Rubak. Determinantal point process models on the sphere. *Research Report 13, Centre for Stochastic Geometry and Advanced Bioimaging. Submitted for journal publication.*, 2015.

A. O'Hagan. Bayes-Hermite quadrature. *Journal of Statistical Planning and Inference*, 29:245–260, 1991.

A. B. Owen. Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, 25(4):1541–1562, 1997.

A. B. Owen. Local antithetic sampling with scrambled nets. *The Annals of Statistics*, 36 (5):2319–2343, 2008.

L. Pastur. Limiting laws of linear eigenvalue statistics for Hermitian matrix models. *J. Math. Phys.*, 47, 2006.

M. M. Peet. Exponentially stable nonlinear systems have polynomial Lyapunov functions on bounded regions. *IEEE Trans. Automat. Control*, 54(5):979–987, 2009.

I. Popescu. General tridiagonal random matrix models, limiting distributions and fluctuations. *Probabability Theory Related Fields*, 144(1-2):179–220, 2009.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning.* MIT Press, 2006.

B. Rider and B. Virág. The noise in the circular law and the Gaussian free field. *Int. Math. Res. Not.*, Art. ID rnm006(2):33 pp., 2007.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer-Verlag, New York, 2004.

E. B. Saff and V. Totik. *Logarithmic potentials with external fields.* Number 316 in Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, Berlin, 1997.

A. Scardicchio, C. E. Zachary, and S. Torquato. Statistical properties of determinantal point processes in high-dimensional euclidean spaces. *Phys. Rev. E*, (3) 79(4, 041108): 19 pp., 2009.

M. Shcherbina. Fluctuations of linear eigenvalue statistics of matrix models in the multicut regime. *J. Stat. Phys.*, 151:1004–1034, 2013.

B. Simon. *Szegő's Theorem and its Descendants: Spectral Theory for L^2 Perturbations of Orthogonal Polynomials.* M. B. Porter Lecture Series, Princeton Univ. Press, Princeton, NJ, 2011.

A. Soshnikov. Determinantal random point fields. *Russian Math. Surveys*, 55:923–975, 2000.

A. Soshnikov. Gaussian limit for determinantal random point fields. *Ann. Probab.*, 30(1): 171–187, 2002.

H. Stahl and V. Totik. *General orthogonal polynomials.* Number 43 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1992.

G. Szegő. *Orthogonal polynomials*, volume Vol. XXIII. American Mathematical Society, Providence, R.I., 4th edition edition, 1974.

S. Torquato, A. Scardicchio, and C. E. Zachary. Point processes in arbitrary dimension from fermionic gases, random matrix theory, and number theory. *J. Stat. Mech. Theory Exp.*, (11, P11019):39 pp., 2008.

S. Xiang. An improved error bound on Gauss quadrature. *Applied Mathematics Letters*, 58:42–48, 2016.

S. Xiang and F. Bornemann. On the convergence rates of Gauss and Clenshaw–Curtis quadrature for functions of limited regularity. *SIAM J. Numer. Anal.*, 50(5):2581–2587, 2012.