

## Research



**Cite this article:** Hayakawa S, Oberhauser H, Lyons T. 2023 Hypercontractivity meets random convex hulls: analysis of randomized multivariate cubatures. *Proc. R. Soc. A* **479**: 20220725.  
<https://doi.org/10.1098/rspa.2022.0725>

Received: 1 November 2022

Accepted: 20 April 2023

**Subject Areas:**

statistics, computational mathematics, applied mathematics

**Keywords:**

random convex hull, cubature, hypercontractivity, kernel quadrature

**Author for correspondence:**

Satoshi Hayakawa  
 e-mail: [hayakawa@maths.ox.ac.uk](mailto:hayakawa@maths.ox.ac.uk)

# Hypercontractivity meets random convex hulls: analysis of randomized multivariate cubatures

Satoshi Hayakawa, Harald Oberhauser and Terry Lyons

Mathematical Institute, University of Oxford, Oxford, Oxfordshire, UK

SH, 0000-0003-1176-5837

Given a probability measure  $\mu$  on a set  $\mathcal{X}$  and a vector-valued function  $\phi$ , a common problem is to construct a discrete probability measure on  $\mathcal{X}$  such that the push-forward of these two probability measures under  $\phi$  is the same. This construction is at the heart of numerical integration methods that run under various names such as quadrature, cubature or recombination. A natural approach is to sample points from  $\mu$  until their convex hull of their image under  $\phi$  includes the mean of  $\phi$ . Here, we analyse the computational complexity of this approach when  $\phi$  exhibits a graded structure by using so-called hypercontractivity. The resulting theorem not only covers the classical cubature case of multivariate polynomials, but also integration on pathspace, as well as kernel quadrature for product measures.

## 1. Introduction

Let  $X$  be a random variable that takes values in a set  $\mathcal{X}$ , and  $\mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$  a linear, finite dimensional space of integrable functions from  $\mathcal{X}$  to  $\mathbb{R}$ . A cubature formula for  $(X, \mathcal{F})$  is a finite set of points  $(x_i) \subset \mathcal{X}$  and weights  $(w_i) \subset \mathbb{R}$  such that

$$\mathbb{E}[f(X)] = \sum_{i=1}^n w_i f(x_i) \quad \text{for all } f \in \mathcal{F}. \quad (1.1)$$

We also denote  $\mu = \text{Law}(X)$  and refer to  $\hat{\mu} = \sum_{i=1}^n w_i \delta_{x_i}$  as the cubature measure for  $(X, \mathcal{F})$ . The existence of such a cubature formula that further satisfies  $n \leq 1 + \dim \mathcal{F}$ ,  $w_i \geq 0$  and  $\sum_{i=1}^n w_i = 1$  is guaranteed by

© 2023 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

what is often referred to as Tchakaloff's theorem although a more accurate nomenclature would involve Wald, Richter, Rogosinski and Rosenbloom [1–5]; see [6] for a historical perspective. Arguably the most famous applications concerns the case when  $\mathcal{X}$  is a subset of  $\mathbb{R}^d$  and  $\mathcal{F}$  is the linear space of polynomials up to a certain degree, that is  $\mathcal{F}$  is spanned by monomials up to a certain degree. However, more recent applications include the case when  $\mathcal{X}$  is a space of paths and  $\mathcal{F}$  is spanned by iterated Ito–Stratonovich integrals [7], or kernel quadrature [8,9] where  $\mathcal{X}$  is a set that carries a positive definite kernel and  $\mathcal{F}$  is a subset of the associated reproducing kernel Hilbert space that is spanned by some ‘test functions’ including eigenfunctions of the integral operator induced by this kernel.

## (a) Convex hulls

If  $\mathcal{F}$  is spanned by  $m$  functions  $\varphi_1, \dots, \varphi_m: \mathcal{X} \rightarrow \mathbb{R}$ , then we can denote  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_m): \mathcal{X} \rightarrow \mathbb{R}^m$  and see that (1.1) is equivalent to

$$\mathbb{E}[\boldsymbol{\varphi}(X)] = \sum_{i=1}^n w_i \boldsymbol{\varphi}(x_i).$$

If we restrict attention to non-negative weights that sum up to one (equivalently,  $\hat{\mu}$  is a probability measure) this is equivalent to that statement that

$$\mathbb{E}[\boldsymbol{\varphi}(X)] \in \text{conv}\{\boldsymbol{\varphi}(x_1), \dots, \boldsymbol{\varphi}(x_n)\}, \quad (1.2)$$

where we denote for an  $A \subset \mathbb{R}^m$  its convex hull as

$$\text{conv } A = \left\{ c_1 a_1 + \dots + c_k a_k \mid k \geq 1, a_i \in A, c_i \geq 0, \sum_{i=1}^k c_i = 1 \right\}.$$

## (b) Random convex hulls

A natural and general approach to find points  $(x_i) \subset \mathcal{X}$  for which (1.2) holds was recently proposed in [10]: draw  $N \gg n$  independent random samples  $(X_i)_{i=1}^N$  from  $\mu$  and subsequently try to select a subset of  $n$  points  $(x_i)$ . The success of this approach amounts the event that

$$\mathbb{E}[\boldsymbol{\varphi}(X)] \in \text{conv}\{\boldsymbol{\varphi}(X_1), \dots, \boldsymbol{\varphi}(X_N)\}, \quad (1.3)$$

since then simple linear programming (LP) allows select the subset of  $x_i$ 's resp. compute the remaining weights that determine a cubature formula. The following guarantees that for large enough  $N$  this event occurs with high probability.

**Proposition 1.1 ([10]).** *If  $X_1, X_2, \dots$  are independent copies of  $X$ , then the probability of the event (1.3) tends to 1 as  $N \rightarrow \infty$ .*

Empirically, this approach turns out to work well already for ‘reasonable’ magnitudes of  $N$  [8,10]. The aim of this article is to fill this gap and provide theoretical guarantees for the number of samples  $N$  for which this approach leads with high probability to a successful cubature construction; that is to provide a quantitative version of Proposition 1.1 that applies to common cases.

## (c) Hypercontractivity

Our main tool is hypercontractivity. This allows to prove the existence of a constant  $C'_m$  satisfying (mainly for  $p = 4$ )

$$\mathbb{E}[|f(X)|^p] \leq C'_m \mathbb{E}[|f(X)|^2]^{p/2}$$

uniformly for a large class of functions  $f$ , and where  $X$  follows the product measure  $\mu^{\otimes d}$ . While hypercontractivity is classically studied for Gaussian, discrete, and uniform probability measures

on hypercubes or hyperspheres [11–14]. We generalize it to function classes that have a certain graded structure.

## (d) Contribution

Our main result is to provide an upper bound for the number of samples  $N$  such that an  $N$ -point i.i.d. sample of random vectors contains the expectation in its convex hull, i.e. the event (1.3) occurs, with a reasonable probability. Although the connection between the bound for  $N$  and the hypercontractivity of the given random vector/function class has implicitly been proven in a preceding study [15] in the form of theorem 2.3, generic conditions for having a good hypercontractivity constant and why the magnitude of required  $N$  becomes reasonably small in practice have not been established or understood.

In this paper, we address these questions by

- extending the hypercontractivity for the Wiener chaos to what we call generalized random polynomials (§3) and
- showing that this extension naturally applies to important examples in numerical analysis including classical cubature, cubature on Wiener space, and kernel quadrature (§4).

We explain the intuition behind these points by introducing theorem 1.2 and example 1.3:

**Theorem 1.2 (informal).** *Let  $\mu$  be a probability measure on  $\mathcal{X}$ . Suppose we have a ‘natural’ function class*

$$\mathcal{F} = \bigoplus_{d \geq 1} \bigcup_{m \geq 0} \mathcal{F}_{d,m},$$

where  $\mathcal{F}_{d,m}$  denotes a set of functions from  $\mathcal{X}^d$  to  $\mathbb{R}$  of ‘degree’ up to  $m$ . Then, under some integrability assumptions, there exists for every  $m$  a constant  $C_m = C_m(\mu, \mathcal{F}) > 0$  such that the following holds:

*Let  $d$  and  $D$  be two positive integers and  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_D) : \mathcal{X}^d \rightarrow \mathbb{R}^D$  with  $\varphi_1, \dots, \varphi_D \in \mathcal{F}_{d,m}$ . Then, for all integers  $N \geq C_m D$ , we have*

$$\mathbb{P}(\mathbb{E}[\boldsymbol{\varphi}(X)] \in \text{conv}\{\boldsymbol{\varphi}(X_1), \dots, \boldsymbol{\varphi}(X_N)\}) \geq \frac{1}{2},$$

where  $X, X_1, \dots, X_N$  are i.i.d. samples from the product measure  $\mu^{\otimes d}$  on  $\mathcal{X}^d$ .

**Example 1.3.** Although the above statement is somewhat abstract, the assumption of a ‘natural’ function class covers the following important examples:

- **Classical cubature** [16]:  $\mu$  is a probability measure with finite  $m$  moments and  $\mathcal{F}_{d,m}$  is the space of  $d$ -variate polynomials up to degree  $m$ .
- **Cubature on Wiener space** [7]:  $\mu$  is the Wiener measure and  $\mathcal{F}_{d,m}$  is spanned by up to  $m$ -times iterated Itô–Stratonovich integrals.
- **Kernel quadrature** [8,9]:  $\mu$  is a probability measure on set  $\mathcal{X}$  that carries a positive definite kernel  $k$  and  $\mathcal{F}_{d,m}$  is spanned by some test functions suitable to  $k^{\otimes d}$ , e.g. eigenfunctions down to some eigenvalue of the integral operator  $g \mapsto \int k^{\otimes d}(\cdot, x)g(x) d\mu^{\otimes d}(x)$ , where  $k^{\otimes d}$  is a tensor product kernel.

## (e) Related work

If the measure  $\mu$  has finite support, the problem (1.1) is also known as recombination. While in this case, the existence follows immediately from Caratheodory’s theorem, the design of efficient algorithms to compute the cubature measure is more recent; we mention efficient deterministic

algorithms [17–19] and randomized speedups [20]. For non-discrete measures, the majority of the cubature constructions are typically limited to algebraic approaches that cannot apply to general situations. Related to our convex hull approach but different, is a line of research aiming at constructing general cubature formulas with positive weights by using least-squares instead of the random convex hull approach [21,22]. As their theory is on the positivity of the resulting cubature formula given by solving a certain least squares problem, it requires more (or efficiently selected) points than are needed for simply obtaining a positively weighted cubature.

Hypercontractivity is the key technical tool for our estimates. Although it is a classic tool in probability, its use seems to be novel in the context of cubature resp. random convex hull problems. Somewhat related to the special case of kernel quadrature, [23] proves a generalization error bound for kernel ridge regression with random features, however hypercontractivity appears there just as a technical assumption. Further, for low-degree polynomials of a sequence of random variables, Kim & Vu [24], Schudy & Sviridenko [25] give similar estimates on their higher order moments, but they mainly estimate the concentration of the moments, and do not generally analyse the kurtosis-type values appearing in the hypercontractivity.

## (f) Outline

In §2, we briefly explain recent results on random convex hulls, and give some assertions that additionally follow from them. In §3, we introduce the Gaussian hypercontractivity and show its generalization that is suitable for multivariate cubatures. Section 4 gives some applications of Gaussian/generalized hypercontractivity to random convex hulls with product structure, including cubature on Wiener space and kernel quadrature. Finally, we conclude the paper in §5. All the omitted proofs are given in appendix A.

## 2. Random convex hulls

Our main interest is the probability of the event (1.3) that the mean is contained in the random convex hull. To quantify this probability, it turns out to be more convenient to study a more general problem. Therefore, we define

**Definition 2.1.** Let  $X$  be a  $D$ -dimensional random vector and  $X_1, X_2, \dots$  be iid copies of  $X$ . For every integer  $N > 0$  and  $\theta \in \mathbb{R}^D$  define

$$p_{N,X}(\theta) := \mathbb{P}(\theta \in \text{conv}\{X_1, \dots, X_N\}) \quad \text{and} \quad N_X(\theta) := \inf\{N \mid p_{N,X}(\theta) \geq 1/2\}.$$

Both of these quantities are classically studied for symmetric  $X$  by Wendel [26], but more recently sharp inequalities for general  $X$  [15,27] as well and calculations on the Gaussian case [28] have been established. Note that, by considering  $m$  independent copies of  $(X_1, \dots, X_N)$ , we have  $1 - p_{mN,X}(\theta) \geq 1 - p_{N,X}(\theta)^m$ , and so knowing  $N_X(\theta)$  also yields some high-probability bounds.

Our main interest is the case  $\theta = \mathbb{E}[X]$ . We can bound  $N_X(\mathbb{E}[X])$  if the distribution of  $X$  satisfies some good properties including symmetry and log-concavity:

**Proposition 2.2.** For a  $D$ -dimensional random vector  $X$  for which  $\mathbb{E}[X]$  exists, we have

- (a) If the distribution of  $X$  is symmetric about  $\mathbb{E}[X]$ , then  $N_X(\mathbb{E}[X]) \leq 2D$ .
- (b) If the distribution of  $X$  is log-concave, then  $N_X(\mathbb{E}[X]) \leq \lceil 3eD \rceil$ .

Here, (a) is a well-known result [26]. The part (b) is novel but immediately follows from a combination of existing results [15,29]; see Appendix A(a) for details. What we use for our main results is however a bound based on moments given by Hayakawa *et al.* [15].

**Theorem 2.3 ([15]).** Let  $X$  be an arbitrary  $D$ -dimensional random vector with  $\mathbb{E}[\|X\|^3] < \infty$ . If a constant  $K > 0$  satisfies  $\|c^\top(X - \mathbb{E}[X])\|_{L^3} \leq K\|c^\top(X - \mathbb{E}[X])\|_{L^2}$  for all  $c \in \mathbb{R}^D$ , then we have

$$N_X(\mathbb{E}[X]) \leq 17(1 + 9K^6/4)D.$$

This result recovers a sharp bound  $N_X(\mathbb{E}[X]) = \mathcal{O}(D)$  up to constant for a Gaussian, where we have detailed information about the marginals. The sort of inequality assumed in the statement is also called Khintchin's inequality (see e.g. [30,31]) and there are known values of  $B$  for a certain class of  $X$  such as Rademacher vectors. Indeed, we can easily show the following estimate under a clear independence structure:

**Proposition 2.4.** *Let  $X = (X_1, \dots, X_D)^\top$  be a  $D$ -dimensional random vector whose coordinates are independent and identically distributed. If  $\mathbb{E}[X_1] = 0$  and  $\|X_1\|_{L^4} \leq K\|X_1\|_{L^2}$  holds for a constant  $K > 0$ , then we have  $\|c^\top X\|_{L^4} \leq K\|c^\top X\|_{L^2}$  for all  $c \in \mathbb{R}^D$ .*

While such an explicit independence yields  $N_X(\mathbb{E}[X]) = \mathcal{O}(D)$ , we can see that we can go much further by carefully looking at how one can prove hypercontractivity in Gaussian Wiener chaos. In the following section, we generalize the whole argument and provide natural conditions for  $X$  to achieve  $N_X(\mathbb{E}[X]) = \mathcal{O}(D)$ .

### 3. Hypercontractivity

The previous section provides bounds on  $N_X$  but the assumptions—log-concavity or coordinate-wise independence—are too strong for many applications. We now develop an approach via hypercontractivity; this results in bounds that apply under much less strict assumptions.

#### (a) Hypercontractivity: the Gaussian case

It is instructive to briefly review the classical results for Gaussian measures by following Janson [32] since we need several generalizations of this.

**Theorem 3.1 (Wiener Chaos decomposition).** *Let  $H$  be a Gaussian Hilbert space<sup>1</sup> and let  $\sigma(H)$  be the  $\sigma$ -algebra generated by  $H$ . Then*

$$L^2(\Omega, \sigma(H), \mathbb{P}) = \bigoplus_{n=0}^{\infty} H^{(n)},$$

where  $H^{(n)} := \overline{P_n(H)} \cap P_{n-1}(H)^\perp$  with

$$P_n(H) := \{f(Y_1, \dots, Y_m) \mid f \text{ is a polynomial of degree} \leq n, Y_1, \dots, Y_m \in H, m < \infty\}$$

with  $P_{-1}(H) := \{0\}$  and  $\overline{P_n(H)}$  denotes the completion in  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ .

Hence, for each  $X \in L^2(\Omega, \sigma(H), \mathbb{P})$ , we have a unique decomposition  $X = \sum_{n=0}^{\infty} X_n$  such that  $X_n \in H^{(n)}$ .

**Theorem 3.2 (Hypercontractivity, [32], theorem 5.8).** *For  $r \in [0, 1]$  denote*

$$T_r : L^2(\Omega, \sigma(H), \mathbb{P}) \rightarrow L^2(\Omega, \sigma(H), \mathbb{P}), \quad X \mapsto \sum_{n=0}^{\infty} r^n X_n.$$

If  $p > 2$  and  $0 < r \leq (p-1)^{-1/2}$ , then we have

$$\|T_r(X)\|_{L^p} \leq \|X\|_{L^2}.$$

From this, we have the following moment bound on  $\overline{P_n(H)}$ , which is also referred to as hypercontractivity, see for example [33].

**Theorem 3.3.** *Let  $n \geq 0$  be an integer. For each  $p > 2$ , we have*

$$\|X\|_{L^p} \leq (p-1)^{n/2} \|X\|_{L^2}, \quad X \in \overline{P_n(H)}.$$

<sup>1</sup>A Gaussian Hilbert space is a closed linear subspace of  $L^2(\Omega, \mathcal{G}, \mathbb{P})$  whose elements all follow Gaussian distributions.

*Proof.* Let  $X = \sum_{m=0}^n X_m$  with  $X_m \in H^{(m)}$ . For  $0 < r \leq (p-1)^{-1/2}$ , by theorem 3.2, we have

$$\begin{aligned} \|X\|_{L^p}^2 &= \left\| T_r \left( \sum_{m=0}^n r^{-m} X_m \right) \right\|_{L^p}^2 \leq \left\| \sum_{m=0}^n r^{-m} X_m \right\|_{L^2}^2 \\ &= \sum_{m=0}^n r^{-2m} \|X_m\|_{L^2}^2 \leq r^{-2n} \|X\|_{L^2}^2. \end{aligned}$$

We obtain the conclusion by letting  $r = (p-1)^{-1/2}$ . ■

We included the proof since we are going to generalize it in the following.

## (b) Hypercontractivity for generalized random polynomials

The phenomenon of hypercontractivity is not limited to the Gaussian setting. Indeed, the hypercontractivity of operators on the space of boolean functions (i.e.  $\{-1, 1\}^n \rightarrow \mathbb{R}$ ) associated with the uniform measure was established even before the Gaussian case [13,34]. Our focus is to obtain estimates analogous to theorem 3.3 when a graded class of test function is given; we refer to such a class as generalized random polynomials.

**Definition 3.4.** Under a probability space  $(\Omega, \mathcal{G}, \mathbb{P})$ , a triplet  $G = (Y, Q, \lambda)$  is called GRP if it satisfies the following conditions:

- $Y$  is a random variable taking values in a topological space  $\mathcal{X}$ .
- $Q = (Q_m)_{m=0}^\infty$  is a non-decreasing sequence of linear spaces of  $L^2(\mathbb{P}_Y)$ -integrable functions  $\mathcal{X} \rightarrow \mathbb{R}$ . Namely, if we let  $Q_m(Y) := \{f(Y) \mid f \in Q_m\}$ , then each  $Q_m$  is a linear subspace of  $L^2(\mathbb{P})$ , with  $Q_0 \subset Q_1 \subset \dots \subset L^2(\mathbb{P})$ . We additionally assume  $Q_0$  is the set of constant functions.
- $\lambda = (\lambda_m)_{m=0}^\infty$  satisfies  $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \dots \geq 0$ .

If  $G$  is a GRP, we also define  $\widetilde{\deg}_G X := \inf\{1/\lambda_m \mid m \geq 0, X \in \overline{Q_m(Y)}\}$ .

Intuitively, each  $Q_m$  is a generalization of degree- $m$  polynomials and  $\widetilde{\deg}_G$  indicates the ‘degree’ of such functions (though  $Y$  plays a role in the latter). In the setting of actual polynomials like Wiener chaos, we can define  $\lambda_m = b^{-m}$  for a certain  $b > 1$ , and then we have  $\deg X = \log_b \widetilde{\deg}_G X$  for the usual degree of  $X$  as a random polynomial.

**Definition 3.5.** Let  $G = (Y, Q, \lambda)$  be a GRP. We define

$$H_m(Y) := \overline{Q_m(Y)} \cap Q_{m-1}(Y)^\perp$$

in terms of  $L^2(\mathbb{P})$  where  $Q_{-1}(Y) := \{0\}$  and

$$H_\infty := L^2(\Omega, \sigma(Y), \mathbb{P}) \cap \left( \bigcup_{m=0}^M Q_m(Y) \right)^\perp.$$

We refer to

$$L^2(\Omega, \sigma(Y), \mathbb{P}) = \left( \bigoplus_{m=0}^\infty H_m(Y) \right) \oplus H_\infty(Y)$$

as the orthogonal decomposition associated with  $G$ .

**Definition 3.6.** Let  $G = (Y, Q, \lambda)$  be a GRP. The operator  $T(G)$  is defined as

$$T(G) : L^2(\Omega, \sigma(Y), \mathbb{P}) \rightarrow L^2(\Omega, \sigma(Y), \mathbb{P}), \quad X \mapsto \sum_{m=0}^{\infty} \lambda_m X_m,$$

where  $(X_m)_{m \in \mathbb{N} \cup \infty}$  with  $X_m \in H_m(Y)$  is the orthogonal decomposition of  $X$  associated with the GRP  $G$ . We say that a GRP  $G = (Y, Q, \lambda)$  is  $(2, p; s)$ -hypercontractive if

$$\|T(G)^s X\|_{L^p} \leq \|X\|_{L^2}, \quad X \in L^2(\Omega, \sigma(Y), \mathbb{P}).$$

Thus,

$$T(G)^s X = \sum_{m=0}^{\infty} \lambda_m^s X_m \quad \text{for } s > 0$$

and if  $G$  is  $(2, p; s)$ -hypercontractive, it is  $(2, p; t)$ -hypercontractive for all  $t \geq s$  as  $T(G)^{t-s}$  is a contraction in  $L^2$ . The formulation of  $G$  associated with ‘degree’ concept given by  $\lambda$  then naturally extends to the multivariate case.

**Definition 3.7.** We call a set of  $d$  GRPs,  $G^{(i)} = (Y^{(i)}, Q^{(i)}, \lambda^{(i)})$  for  $i = 1, \dots, d$  independent, if the  $Y^{(i)}$ ’s are independent random variables taking values in  $\mathcal{X}^{(i)}$ ’s. For  $d$  independent GRPs, their product is a GRP  $G = (Y, Q, \lambda)$  that is defined as follows

- $Y = (Y^{(1)}, \dots, Y^{(d)}) \in \mathcal{X}^{(1)} \times \dots \times \mathcal{X}^{(d)}$ .
- $\lambda_m$  is the  $(m+1)$ -th largest value in the set  $\left\{ \prod_{i=1}^d \lambda_{m_i}^{(i)} \mid \lambda_{m_i}^{(i)} \in \lambda^{(i)}, i = 1, \dots, d \right\}$ .
- $Q_m = \text{span}\{f : (x_1, \dots, x_d) \mapsto \prod_{i=1}^d f_i(x_i) \mid f_i \in Q_{m_i}^{(i)}, \prod_{i=1}^d \lambda_{m_i}^{(i)} \leq \lambda_m\}$ .

As  $Q_m(Y) \subset L^2$  it follows from independence for each  $m$  that  $G = (Y, Q, \lambda)$  is indeed a GRP. We also denote it by  $G = G^{(1)} \otimes \dots \otimes G^{(d)}$ .

**Example 3.8.** Consider the case when  $Q_m^{(i)}$  are degree- $m$  polynomials of  $Y^{(i)}$  and  $\lambda_m^{(i)} = t^m$  for some  $t \in (0, 1)$  independent of  $i$ . This shows that the product GRP generalizes the multivariate random polynomials. Also, when  $Y^{(i)}$  are i.i.d. and  $(Q^{(i)}, \lambda^{(i)})$  are the same for all  $i = 1, \dots, d$ , then we say  $G^{(i)}$  are i.i.d. and we can in particular write  $G \simeq (G^{(1)})^{\otimes d}$ .

A straightforward generalization follows from the classical way of proving hypercontractivity. This turns out to be useful for treating multivariate hypercontractivity of our GRP setting.

**Theorem 3.9.** Let  $r \in (0, 1]$  and  $p > 2$ . If  $d$  independent GRPs  $G^{(1)}, \dots, G^{(d)}$  are all  $(2, p; s)$ -hypercontractive, then their product  $G = G^{(1)} \otimes \dots \otimes G^{(d)}$  is also  $(2, p; s)$ -hypercontractive.

**Remark 3.10.** We only use the  $(2, p; s)$ -hypercontractivity in this paper, but we can also deduce the same results for the general  $(q, p; s)$ -hypercontractivity with  $1 \leq q \leq p < \infty$  (for the operator norm of  $L^q \rightarrow L^p$ ), analogous to e.g. Janson [32].

The following is a result analogous to theorem 3.3 and the proof is almost identical.

**Proposition 3.11.** Let  $s > 0$  and  $p > 2$ . If  $G$  is a GRP that is  $(2, p; s)$ -hypercontractive, then we have  $\|X\|_{L^p} \leq (\deg_G X)^s \|X\|_{L^2}$  for all  $X \in L^2$ .

**Remark 3.12.** Although we have treated general GRPs  $G = (Y, Q, \lambda)$  in these propositions, we are basically only interested in the moment inequality for  $X$  up to some ‘degree’ fixed beforehand (in the case of Wiener chaos, it suffices to treat  $P_n(H)$  for some finite  $n$  to obtain theorem 3.3). Thus, our main interest is in ‘finite’ GRPs, satisfying  $Q_n = Q_{n+1} = \dots$  for some  $n$ , and their product in practice, which might be better for readers to have in mind when reading the next proposition.

We next show the following ‘converse’ result for the relation of the hypercontractivity and moment estimate for a (truncated) GRP when  $p = 4$ .



**Proposition 3.13.** Let  $G = (Y, Q, \lambda)$  be a GRP. Suppose there exists a  $s > 0$  such that

$$\|X_m\|_{L^4} \leq \lambda_m^{-s} \|X_m\|_{L^2}, \quad X_m \in H_m(Y)$$

holds for all  $m$ . If  $t > s$  satisfies

$$\sum_{m \geq 1} \lambda_m^{t-s} \leq 1/\sqrt{3}$$

and  $\lambda_1^t \leq 1/2$ , then  $G$  is  $(2, 4; t)$ -hypercontractive.

By using this, we can also prove the following as a non-quantitative result.

**Theorem 3.14.** Let  $K > 0$  and  $G$  be a GRP such that the space  $\{X \in L^2 \mid \widetilde{\deg}_G X \leq K\}$  is included in  $L^4(\Omega, \mathcal{F}, \mathbb{P})$  and finite-dimensional. Then, there exists a constant  $C = C(G, K)$  such that for an arbitrary  $d$ ,  $\|X\|_{L^4} \leq C\|X\|_{L^2}$  holds if we have  $\widetilde{\deg}_{G^{\otimes d}} X \leq K$ .

## 4. Applications

The generality of proposition 3.13 and theorem 3.14 allows us to quantify the number of samples resp. probability of success of the random convex hull approach to the problem of cubature. Concretely, we give formal statements of theorem 1.2 for various cubature constructions: (i) classical cubature, (ii) cubature on Wiener space, (iii) Kernel quadrature.

### (a) Classical polynomial cubatures

When the GRP  $G$  are actual random polynomials, we recover the following result.

**Corollary 4.1.** Let  $m$  be a positive integer and  $X^{(1)}, X^{(2)}, \dots$  be i.i.d. real-valued random variables with  $\mathbb{E}[|X^{(1)}|^{4m}] < \infty$ . Then, there exists a constant  $C_m > 0$  such that

$$\|f(X^{(1)}, \dots, X^{(d)})\|_{L^4} \leq C_m \|f(X^{(1)}, \dots, X^{(d)})\|_{L^2}$$

for any positive integer  $d$  and any polynomial  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  with degree up to  $m$ .

*Proof.* By introducing a truncated GRP given by a random variable  $X^{(1)}$ , function spaces  $Q_i$  of univariate polynomials up to degree  $i$ , and  $\lambda_i = 2^{-i}$  for  $0 \leq i \leq m$ , we can apply theorem 3.14 to obtain the desired result. ■

If we combine this with theorem 2.3, we obtain the following result for polynomial cubatures.

**Corollary 4.2.** Let  $m \geq 1$  be an integer and  $X^{(1)}, X^{(2)}, \dots$  be i.i.d. real-valued random variables with  $\mathbb{E}[|X^{(1)}|^{4m}] < \infty$ . Then, there exists a constant  $C_m > 0$  such that the following holds:

Let  $d \geq 1$  be an integer and  $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^D$  be a  $D$ -dimensional vector-valued function such that each coordinate is given by a polynomial up to degree  $m$ . If we let  $X_1^{(1:d)}, X_2^{(1:d)}, \dots$  be independent copies of  $X^{(1:d)} = (X^{(1)}, \dots, X^{(d)})$ , we have

$$\mathbb{P}(\mathbb{E}[\varphi(X^{(1:d)})] \in \text{conv}\{\varphi(X_1^{(1:d)}), \dots, \varphi(X_N^{(1:d)})\}) \geq \frac{1}{2}$$

for all integers  $N \geq C_m D$ .

### (b) Cubature on Wiener space

Cubature on Wiener space [7] is a weak approximation scheme for stochastic differential equations; at the heart of it is the construction of a finite measure on pathspace, such that the expectation of their first  $m$ -times iterated integrals matches those of Brownian motion. Some algebraic constructions are known for degree  $m = 3, 5$  [7] as well as  $m = 7$  [35]. The random convex hull approach applies in principle for any  $m$ , however, a caveat is that the discretization of paths becomes an issue in particular for high values of  $m$ ; some experimental results are available in



[36] by using random samples of piecewise linear approximations of Brownian motion. In this section, we use hypercontractivity to estimate the number of samples needed in this approach to cubature via sampling.

### (c) Random convex hulls of iterated integrals

For a continuous bounded variation (BV) path  $x = (x^0, \dots, x^d): [0, 1] \rightarrow \mathbb{R}^{d+1}$  and a  $d$ -dimensional standard Brownian motion  $B = (B^1, \dots, B^d)$  with  $B_t^0 := t$ , we define the iterated integrals as

$$I^\alpha(x) := \int_{0 < t_1 < \dots < t_k < 1} dx_{t_1}^{\alpha_1} \dots dx_{t_k}^{\alpha_k} \quad \text{and} \quad I^\alpha(B) := \int_{0 < t_1 < \dots < t_k < 1} \circ dB_{t_1}^{\alpha_1} \dots \circ dB_{t_k}^{\alpha_k},$$

where the latter is given by the Stratonovich stochastic integral. Then, a degree  $m$  cubature formula on Wiener space for  $d$ -dimensional Brownian motion is a set of BV paths  $x_1, \dots, x_n: [0, 1] \rightarrow \mathbb{R}^{d+1}$  and convex weights  $w_1, \dots, w_n$  such that  $\sum_{i=1}^n w_i I^\alpha(x_i) = \mathbb{E}[I^\alpha(B)]$  for all multi-indices  $\alpha = (\alpha_1, \dots, \alpha_k) \in \bigcup_{\ell \geq 1} \{0, 1, \dots, d\}^\ell$  with  $\|\alpha\| := k + |\{j \mid \alpha_j = 0\}| \leq m$ .

Indeed, if we consider the Gaussian Hilbert space given by

$$H := \left\{ \sum_{i=1}^d \int_0^1 f_i(t) dB_t^i \mid f_1, \dots, f_d \in L^2([0, 1]) \right\},$$

the iterated integral  $I^\alpha(B)$  with  $\|\alpha\| \leq m$  is in the  $m$ -th Wiener chaos  $\overline{P_m(H)}$  (see §3) as it can be expressed as a limit of polynomials of increments of  $B$ . We thus have the hyperconstructivity given in theorem 3.3 and the following assertion:

**Corollary 4.3.** *Let  $d, m \geq 1$  be integers and  $B$  be a  $d$ -dimensional Brownian motion. Then, for an arbitrary linear combination  $X = \sum_{\|\alpha\| \leq m} c_\alpha I^\alpha(B)$  with  $c_\alpha \in \mathbb{R}$ , we have  $\|X\|_{L^3} \leq 2^{m/2} \|X\|_{L^2}$ .*

As the bound is independent of the dimension  $d$  of the underlying Brownian motion, we have the following version of theorem 1.2 by combining it with theorem 2.3 as follows:

**Corollary 4.4.** *Let  $d, m \geq 1$  be integers and  $B, B_1, B_2, \dots$  be independent standard  $d$ -dimensional Brownian motions. Then, if  $\varphi(B)$  is a  $D$ -dimensional random vector such that each coordinate is given by a linear combination of  $(I^\alpha(B))_{\|\alpha\| \leq m}$ , then we have*

$$\mathbb{P}(\mathbb{E}[\varphi(B)] \in \text{conv}\{\varphi(B_1), \dots, \varphi(B_N)\}) \geq \frac{1}{2}$$

for all integers  $N \geq 17(1 + 18 \cdot 8^{m-1})D$ .

The above allows us to choose the number of candidate paths that need to be sampled. However, as mentioned above, one challenge that is specific to cubature on pathspace is that one cannot sample a true Brownian trajectory which leads to an additional discretization error. However, we conjecture that the number of random samples divided by  $D$  and the number of time partitions for piecewise linear approximation in constructing cubature on Wiener space can be independent of the underlying dimension  $d$ .

**Remark 4.5.** One can also apply these estimates to fractional Brownian motion [37], though we also need to obtain the exact expectations of iterated integrals of fractional Brownian motion (we can find some results on the Ito-type iterated integrals without the time integral by  $B_t^0 = t$  in the literature [38, Theorem 31]).

### (d) Kernel quadrature for product measures

Let  $\mathcal{X}$  be a topological space and  $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  be a positive definite kernel with the reproducing kernel Hilbert space (RKHS)  $\mathcal{H}_k$  [39]. A *kernel quadrature* for a random variable  $X$  or equivalently

a Borel probability measure  $\mu$  (i.e.  $X \sim \mu$ ) on  $\mathcal{X}$  is a set of points  $x_i \in \mathcal{X}$  and weights  $w_i \in \mathbb{R}$  such that  $\tilde{\mu}_n = \sum w_i \delta_{x_i}$ , where our aim is to make the following worst-case error as small as possible:

$$\text{wce}(\tilde{\mu}_n; \mathcal{H}_k, \mu) := \sup_{\|f\|_{\mathcal{H}_k} \leq 1} \left| \mathbb{E}[f(X)] - \sum_{i=1}^n w_i f(x_i) \right|, \quad (4.1)$$

which we might just denote by  $\text{wce}(\tilde{\mu}_n)$ . We call a kernel quadrature convex if  $\tilde{\mu}_n$  is a probability measure, i.e.  $w_i \geq 0$  and  $\sum_{i=1}^n w_i = 1$ . Algorithms for better kernel quadrature rules have been pursued from the viewpoint of optimization [40–42] as well as sampling [8,43,44].

In this section, we explain the algorithm of Hayakawa *et al.* [8], which obtains good kernel quadrature formulas based on random convex hulls, and see how our GRP-based argument provides bounds for its computational complexity in a certain setting.

## (e) Tensor product kernels

When there are  $d$  pairs of space and kernel  $(\mathcal{X}_1, k_1), \dots, (\mathcal{X}_d, k_d)$ , the *tensor product kernel* on the product space  $\mathcal{X}_1 \times \dots \times \mathcal{X}_d$  is defined as

$$(k_1 \otimes \dots \otimes k_d)(x, y) := \prod_{i=1}^d k_i(x_i, y_i), \quad x = (x_1, \dots, x_d), y = (y_1, \dots, y_d) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_d.$$

This is indeed the reproducing kernel of the tensor product  $\mathcal{H}_{k_1} \otimes \dots \otimes \mathcal{H}_{k_d}$  in terms of RKHS [39]. The most important example of this construction is when the underlying  $d$  kernels are the same,  $k^{\otimes d} = k \otimes \dots \otimes k$ . Given a probability measure  $\mu$  in the (conceptually univariate) space  $\mathcal{X}$ , constructing a kernel quadrature for  $\mu^{\otimes d}$  with respect to  $k^{\otimes d}$  is a natural multivariate extension of kernel quadrature that is widely studied in the literature [9,43,45,46], and corresponds to high-dimensional QMCs [47]. While we will ultimately consider kernel quadrature for  $(k^{\otimes d}, \mu^{\otimes d})$ , let us start from the ‘univariate’  $(k, \mu)$  in the following.

## (f) Mercer-like expansions and quadrature

Let us consider the Mercer-type expansion

$$k(x, y) = \sum_{\ell=1}^{\infty} \sigma_{\ell} e_{\ell}(x) e_{\ell}(y), \quad (4.2)$$

where we suppose  $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$  and  $e_{\ell} \in L^2(\mu)$  (not necessarily normalized). Although  $(\sigma_{\ell}, e_{\ell})_{\ell=1}^{\infty}$  are given by normalized eigenpairs of the integral operator  $\mathcal{K}: f \mapsto \int_{\mathcal{X}} k(\cdot, y) f(y) d\mu(y)$  in the case of Mercer expansion [48], we can also use other expansions such as power-series expansion [49] and the Nyström-based (truncated) expansion [8,50]. The following shows how a finite-dimensional cubature in the sense of (1.1) yields some meaningful kernel quadrature.

**Proposition 4.6 ([8]).** *Let  $\tilde{\mu}_n = (w_i, x_i)_{i=1}^n$  be a convex kernel quadrature satisfying  $\int_{\mathcal{X}} e_{\ell}(x) d\mu(x) = \sum_{i=1}^n w_i e_{\ell}(x_i)$  for each  $\ell = 1, \dots, n-1$ . Then, by letting  $r_n(x) := \sum_{m=n}^{\infty} \sigma_m e_m(x)^2$ , we have  $\text{wce}(\tilde{\mu}_n)^2 \leq 4 \sup_{x \in \mathcal{X}} r_n(x)$ .*

We can have more favourable bounds on  $\text{wce}(\tilde{\mu}_n)$  by assuming more, but the important fact here is that the event (1.3) for a vector-valued  $\varphi$  given by ‘basis’ functions  $e_1, \dots, e_{n-1}$  enables us to construct an interesting numerical scheme. A similar approach, specialized to a Gaussian kernel over a Gaussian measure, can be found in [9]. Given a Mercer-like expansion (4.2), we can also consider the multivariate version

$$k^{\otimes d}(x, y) = \sum_{\ell_1, \dots, \ell_d=1}^{\infty} \sigma_{\ell_1} \dots \sigma_{\ell_d} (e_{\ell_1} \otimes \dots \otimes e_{\ell_d})(x) (e_{\ell_1} \otimes \dots \otimes e_{\ell_d})(y), \quad (4.3)$$

and the same result as proposition 4.6 holds for the properly reordered expansion. For the interaction between the convergence rate and the dimension  $d$  in the case of Mercer expansion, Bach [43, Section 3.4] provides some informative examples.

As the construction of such  $\tilde{\mu}_n$  in proposition 4.6 for general  $k$  and  $\mu$  relies on random sampling, we want to estimate  $N_{\varphi(X)}(\mathbb{E}[\varphi(X)])$  for  $X \sim \mu$  and  $\varphi = (e_1, \dots, e_{n-1})$ , where our main interest lies in the multivariate case despite using the univariate notation for simplicity.

## (g) From reproducing kernel Hilbert space to GRP

To make it compatible with the framework of GRPs introduced in the previous section, we further assume the following condition, which ensures that the kernel is in an appropriate scaling.

**Assumption 4.7.** *The expansion (4.2) converges pointwise,  $\sum_{\ell=1}^{\infty} \sigma_{\ell} < \infty$ ,  $\sigma_1 \leq 1$ , and the strict inequality  $\sigma_{\ell} < 1$  holds if  $e_{\ell} \in L^2(\mu)$  is not constant.*

Under assumption 4.7, we can naturally define a GRP  $G = (Y, Q, \lambda)$  with  $Y$  following  $\mu$ ,  $Q_m = \text{span}\{1, e_1, \dots, e_m\}$  and  $\lambda_m = \sigma_m$  for  $m \geq 1$ . Note that it violates the condition  $\lambda_1 < 1$  if  $\sigma_1 = 1$  and  $e_1$  is constant, but in that case we can simply decrement all the indices of  $(Q_m, \lambda_m)$  by one. We call it the *natural GRP* for  $k$  (with the expansion) and  $\mu$ , and we denote it by  $G = G_{k,\mu}$ .

**Remark 4.8.** The scaling given in assumption 4.7 is essential to the hypercontractivity under the framework of tensor product kernels when considering ‘eigenspace down to some eigenvalue’. Indeed, if  $\sigma_{\ell} \geq 1$  for some nonconstant  $e_{\ell}$ , we have, for  $p > 2$ ,

$$\frac{\|e_{\ell}^{\otimes d}\|_{L^p(\mu^{\otimes d})}}{\|e_{\ell}^{\otimes d}\|_{L^2(\mu^{\otimes d})}} = \left( \frac{\|e_{\ell}\|_{L^p(\mu)}}{\|e_{\ell}\|_{L^2(\mu)}} \right)^d$$

which increases exponentially as  $d$  grows, whereas the corresponding  $(\sigma_{\ell})^d$  is lower bounded by 1. So the hypercontractivity in our sense never gets satisfied if  $\sigma_{\ell} \geq 1$  for a nonconstant  $e_{\ell}$ .

By introducing GRPs as above, we can prove the following statement, written without GRPs.

**Proposition 4.9.** *Let  $k$  satisfy assumption 4.7 and  $Y_1, Y_2, \dots$  independently follow  $\mu$ . For each  $\varepsilon > 0$ , define a set of random variables as*

$$S(\varepsilon) := \text{span}(\{1\} \cup \{e_{\ell_1}(Y_{m_1}) \cdots e_{\ell_d}(Y_{m_d}) \mid d \geq 1, m_1 < \cdots < m_d, \sigma_{\ell_1} \cdots \sigma_{\ell_d} \geq \varepsilon\}).$$

*Then, if  $\|e_{\ell}(Y_1)\|_{L^4} < \infty$  holds for all  $\ell$  with  $\sigma_{\ell} \geq \varepsilon$ , then there is a constant  $C_{\varepsilon} > 0$  such that  $\|X\|_{L^4} \leq C_{\varepsilon} \|X\|_{L^2}$  for all  $X \in S(\varepsilon)$ .*

*Proof.* The finiteness of the dimension of ‘eigenspace’ for  $Y_1$ , i.e. the finiteness of the number of  $\ell$  satisfying  $\sigma_{\ell} \geq \varepsilon$  follows from  $\sum_{\ell=1}^{\infty} \sigma_{\ell} < \infty$  in assumption 4.7. Thus, theorem 3.14 gives the conclusion. ■

If we only had  $Y_1, \dots, Y_d$ , then  $S(\varepsilon)$  would correspond to the truncation of the  $d$ -variate expansion (4.3). So this assertion includes a hypercontractivity statement for an ‘eigenspace’ of  $k^{\otimes d}$  and  $\mu^{\otimes d}$  given the expansion (4.3). However, we can go further to a quantitative statement by imposing another assumption in the case of actual Mercer expansion.

## (h) Quantitative bounds for Mercer expansion

We first set up two additional assumptions for obtaining a quantitative statement. The following assumption says that (4.2) is actually the Mercer expansion.

**Assumption 4.10.**  *$(e_{\ell})_{\ell=1}^{\infty}$  and  $(\sqrt{\sigma_{\ell}} e_{\ell})_{\ell=1}^{\infty}$  are orthonormal sets in  $L^2(\mu)$  and  $\mathcal{H}_k$ , respectively.*

Mild conditions already imply that assumption 4.10 holds, e.g.  $\text{supp } \mu = \mathcal{X}$ ,  $k$  is continuous, and  $x \mapsto k(x, x)$  is in  $L^1(\mu)$  is sufficient, see [48]. Another assumption requires further orthogonality of these test functions against a constant function.

**Assumption 4.11.** The kernel  $k$  can be written as  $k = 1 + k_0$ , where  $k_0 : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a positive definite kernel satisfying  $\int_{\mathcal{X}} k_0(x, y) d\mu(y) = 0$  for  $(\mu$ -almost) all  $x \in \mathcal{X}$ .

Under assumptions 4.7 and 4.10, this is simply equivalent to  $e_1$  being constant. This might seem artificial, but naturally arises in the following situations:

- (a)  $\mathcal{X}$  is a compact group and  $\mu$  is its Haar measure.  $k$  is a positive definite kernel given as  $k(x, y) = g(x^{-1}y)$ , where  $g : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$  and  $\int_{\mathcal{X}} g(x) d\mu(x) = 1$ .
- (b)  $k_0$  is a kernel called Stein kernel [51,52] with appropriate scaling.

One theoretically sufficient condition for these assumptions can be described as follows:

**Proposition 4.12.** Let  $\mathcal{X}$  be compact metrizable and path-connected,  $\text{supp } \mu = \mathcal{X}$ , and  $k$  be continuous and nonnegative. If  $\int_{\mathcal{X}} k(x, y) d\mu(y) = 1$  holds for all  $x \in \mathcal{X}$ , assumptions 4.7–4.11 hold.

From this proposition, for instance, an appropriately scaled exponential/Gaussian kernel over the  $n$ -sphere with the uniform measure satisfies assumptions 4.7–4.11.

Under these two assumptions, the operator  $T(G_{k,\mu})$  in terms of GRPs corresponds to the integral operator  $\mathcal{K} : f \mapsto \int_{\mathcal{X}} k(\cdot, y) f(y) d\mu(y)$ , so the situation becomes simpler. We can directly apply proposition 3.13 by replacing  $\lambda$ 's with  $\sigma$ 's, but we also have the following sufficient conditions for the hypercontractivity without explicitly using the eigenvalue sequence. In the following,  $\|\mathcal{K}_0\| := \sigma_2 < 1$  is the operator norm of  $\mathcal{K}_0 : f \mapsto \int_{\mathcal{X}} k_0(\cdot, y) f(y) d\mu(y)$  on  $L^2(\mu)$ , and  $\text{tr}(\mathcal{K}_0) := \int_{\mathcal{X}} k_0(x, x) d\mu(x)$ . We have the following quantitative condition for hypercontractivity.

**Proposition 4.13.** Let  $k = 1 + k_0$  satisfy assumptions 4.7–4.11. When  $\|\mathcal{K}_0\| > 0$ , if  $r, s \geq 1$  satisfy

$$\|\mathcal{K}_0\|^{-(r+s)} \geq 2, \quad \|\mathcal{K}_0\|^{-(r-1)} \geq \sqrt{3} \text{tr}(\mathcal{K}_0) \quad \text{and} \quad \|\mathcal{K}_0\|^{-(s-1)} \geq \|k_0\|_{L^4(\mu \otimes \mu)},$$

then  $G_{k,\mu}$  is  $(2, 4; r + s)$ -hypercontractive. In particular, if we have  $\sup_{x \in \mathcal{X}} |k_0(x, x)| \leq 1/\sqrt{3}$ , then  $G_{k,\mu}$  is  $(2, 4; 2)$ -hypercontractive.

**Example 4.14 (Periodic Sobolev spaces over the torus).** Following Bach [43], we consider periodic kernels over  $[0, 1]$ . Therefore, let  $\mathcal{X} = [0, 1]$ ,  $\mu$  be the uniform distribution on  $\mathcal{X}$ , and define

$$k_{r,\delta}(x, y) = 1 + \delta \cdot \frac{(-1)^{r-1} (2\pi)^{2r}}{(2r)!} B_{2r}(|x - y|) \quad (4.4)$$

for each positive integer  $s$  and  $\delta \in (0, 1)$ , where  $B_{2r}$  is the  $2r$ -th Bernoulli polynomial [53].  $\delta = 1$  is assumed in the original definition, but it violates assumption 4.7 (see also remark 4.8). Albeit this slight modification, the kernel  $k_{r,\delta}$  gives an equivalent norm to the periodic Sobolev space in the literature. For  $\delta \in (0, 1)$ ,  $k_{r,\delta}$  satisfies assumptions 4.7–4.11. The eigenvalues and eigenfunctions with respect to the uniform measure are known [43]; the eigenvalues are: 1 for the constant function, and  $\delta m^{-2r}$  for  $c_m(\cdot) := \sqrt{2} \cos(2\pi m \cdot)$  and  $s_m(\cdot) := \sqrt{2} \sin(2\pi m \cdot)$  for  $m \geq 1, 2, \dots$ . We now apply proposition 3.13 with (for the sake of concreteness)  $\delta = 1/3$ . This gives  $\|c_m\|_{L^4(\mu)} = \|s_m\|_{L^4(\mu)} = (3/2)^{1/4}$ . Thus, to satisfy the condition of proposition 3.13, it suffices for  $s < t$  to satisfy  $3^s \geq (3/2)^{1/4} \cdot \delta^{t-s} \zeta(2r(t-s))$ ,  $3^t \geq 2$ , where  $\zeta$  is Riemann's zeta function. Hence a simple numerical sufficient condition for this is  $s = 0.1$  and  $t = 1.1$  for  $r = 1$ , and  $t = \log_3 2 \leq 0.631$  for  $r \geq 2$ , which can be derived by letting  $2r(t-s) \geq 2$ . To sum up, in the case  $r \geq 2$ , we only need  $\mathcal{O}(\lambda^{-0.631D})$  times of sampling for meeting (1.3) with probability over a half, if  $X \sim \mu^{\otimes d}$  and each coordinate of  $\varphi : \mathcal{X}^d \rightarrow \mathbb{R}^D$  is in the eigenspace of the eigenvalue  $\lambda$ .

## 5. Concluding remarks

We investigated the number of samples needed for the expectation vector to be contained in their convex hull from the viewpoint of product/graded structure. We showed that the fact that we empirically only need  $\mathcal{O}(D)$  times of sampling for the  $D$ -dimensional random vector in practical examples can partially be explained by the hypercontractivity in the Gaussian case as well as

the generalized situation including random polynomials and product kernels. There are also interesting questions for further research; for example, although in the asymptotic  $d \rightarrow \infty$  we established that the required number of sampling divided by  $D$  is independent of  $d$ , the constants are larger than the purely empirical estimates given in [8,10] (where  $10D$  is sufficient in practice). Another direction is the case of cubature of Wiener space, as one cannot actually sample from Brownian motion and discretization errors propagate to higher order  $m$ ; a promising research direction could be to study ‘approximate sampling’ or consider unbiased simulations [54] for the iterated integrals.

**Data accessibility.** This article has no additional data.

**Authors’ contributions.** S.H.: conceptualization, formal analysis, investigation, methodology, writing—original draft; H.O.: supervision, validation, writing—review and editing; T.L.: supervision.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** This work was supported in part by the EPSRC (grant number EP/S026347/1), in part by The Alan Turing Institute under the EPSRC grant EP/N510129/1, the Data Centric Engineering Programme (under the Lloyd’s Register Foundation grant G0095), the Defence and Security Programme (funded by the UK Government) and the Office for National Statistics & The Alan Turing Institute (strategic partnership) and in part by the Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA).

## Appendix A. Proofs

### (a) Proof of proposition 2.2(b)

For a  $D$ -dimensional random vector  $X$ , define the following quantity

$$\alpha_X(\theta) := \inf_{c \in \mathbb{R}^D \setminus \{0\}} \mathbb{P}(c^\top (X - \theta) \leq 0). \quad (\text{A } 1)$$

It is the so-called Tukey depth [55,56], and recently found to be closely related to  $N_X(\theta)$ :

**Theorem A.1 ([15]).** For  $\theta \in \mathbb{R}^D$ , we have

$$\frac{1}{2\alpha_X(\theta)} \leq N_X(\theta) \leq \left\lceil \frac{3D}{\alpha_X(\theta)} \right\rceil.$$

The above can be used to provide a novel bound on  $N_X(\mathbb{E}[X])$  for a general class of distributions called *log-concave* as in the statement of proposition 2.2. A function  $f: \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$  is called *log-concave* if it satisfies

$$f(tx + (1-t)y) \geq f(x)^t f(y)^{1-t}$$

for all  $x, y \in \mathbb{R}^d$  and  $t \in [0, 1]$ . A probability distribution with a log-concave density is also called log-concave, and this class includes the multivariate Gaussian/exponential/Wishart distributions, the uniform distribution over a convex domain, and many more univariate common distributions [57,58]. For the log-concave random vectors, the following result is known:

**Theorem A.2 ([29]).** If  $X$  is a  $d$ -dimensional random vector with log-concave density, then we have  $\alpha_X(\mathbb{E}[X]) \geq 1/e$ .

The case when  $X$  is uniform over a convex set is proven in Grünbaum [59], and Lovász & Vempala [60, Section 5] gives simpler proofs than the original result in Caplin & Nalebuff [29].

Simply combining theorem A.1 and theorem A.2 yields the desired result.

## (b) Proof of proposition 2.4

*Proof.* It suffices to consider the case  $\|X_1\|_{L^4} < \infty$ . If we write  $c = (c_1, \dots, c_D)^\top$ , then by using independence we have

$$\begin{aligned}\|c^\top X\|_{L^4}^4 &= \mathbb{E}[(c^\top X)^4] = \sum_{i=1}^D c_i^4 \mathbb{E}[X_i^4] + \sum_{1 \leq i < j \leq D} c_i^2 c_j^2 \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] \\ &\leq K^4 \sum_{i=1}^D c_i^4 \mathbb{E}[X_i^2]^2 + \sum_{1 \leq i < j \leq D} c_i^2 c_j^2 \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] \\ &\leq K^4 \left( \sum_{i=1}^D c_i^2 \mathbb{E}[X_i^2] \right)^2 \leq K^4 \mathbb{E}[(c^\top X)^2]^2,\end{aligned}$$

as we clearly have  $K \geq 1$  (or  $X = 0$  almost surely). ■

## (c) Proof of theorem 3.9

*Proof.* We give the proof by generalizing the proof of lemma 5.3 in Janson [32].

It suffices to prove the statement for  $d = 2$ , as the product of GRPs is associative. Let  $G^{(i)} = (Y^{(i)}, Q^{(i)}, \lambda^{(i)})$  for  $i = 1, 2$  be independent GRPs. Let  $H_m^{(i)}(Y^{(i)}) := \overline{Q_m^{(i)}(Y^{(i)})} \cap Q_{m-1}^{(i)}(Y^{(i)})^\perp$  for  $i = 1, 2$ . If we denote the product by  $G = G^{(1)} \otimes G^{(2)}$ . Then, for a random variable  $X = \sum_{\ell, m} X_{\ell, m}$  with  $X_{\ell, m} \in H_\ell^{(1)} \otimes H_m^{(2)}$ , the operator  $T(G)$  acts as

$$T(G)X = \sum_{\ell, m} \lambda_\ell^{(1)} \lambda_m^{(2)} X_{\ell, m}.$$

If each  $X_{\ell, m}$  can be written as a finite sum  $X_{\ell, m} = \sum_k X_{k, \ell, m}^{(1)} X_{k, \ell, m}^{(2)}$  with  $X_{k, \ell, m}^{(1)} \in H_\ell^{(1)}(Y^{(1)})$  and  $X_{k, \ell, m}^{(2)} \in H_m^{(2)}(Y^{(2)})$ , then by using Minkowski's integral inequality [61] and the  $(2, p; s)$ -hypercontractivity of  $G^{(1)}$  and  $G^{(2)}$ , we have

$$\begin{aligned}\|T(G)^s X\|_{L^p} &= \mathbb{E}_{Y^{(1)}} \left[ \mathbb{E}_{Y^{(2)}} \left[ \left| \sum_{\ell, m} (\lambda_\ell^{(1)} \lambda_m^{(2)})^s X_{\ell, m} \right|^p \right] \right]^{1/p} \\ &= \mathbb{E}_{Y^{(1)}} \left[ \mathbb{E}_{Y^{(2)}} \left[ \left| \sum_{k, \ell, m} (\lambda_\ell^{(1)})^s X_{k, \ell, m}^{(1)} (\lambda_m^{(2)})^s X_{k, \ell, m}^{(2)} \right|^p \right] \right]^{1/p} \\ &\leq \mathbb{E}_{Y^{(1)}} \left[ \mathbb{E}_{Y^{(2)}} \left[ \left| \sum_{k, \ell, m} (\lambda_\ell^{(1)})^s X_{k, \ell, m}^{(1)} X_{k, \ell, m}^{(2)} \right|^2 \right]^{p/2} \right]^{1/p} && \text{(by } G^{(2)}) \\ &\leq \mathbb{E}_{Y^{(2)}} \left[ \mathbb{E}_{Y^{(1)}} \left[ \left| \sum_{k, \ell, m} (\lambda_\ell^{(1)})^s X_{k, \ell, m}^{(1)} X_{k, \ell, m}^{(2)} \right|^p \right]^{2/p} \right]^{1/2} && \text{(by Minkowski)} \\ &\leq \mathbb{E}_{Y^{(2)}} \left[ \mathbb{E}_{Y^{(1)}} \left[ \left| \sum_{k, \ell, m} X_{k, \ell, m}^{(1)} X_{k, \ell, m}^{(2)} \right|^2 \right] \right]^{1/2} = \|X\|_{L^2}. && \text{(by } G^{(1)})\end{aligned}$$

The general case follows from the limit argument. ■

### (d) Proof of proposition 3.11

*Proof.* Let  $G = (Y, Q, \lambda)$ . Suppose  $\widetilde{\deg}_G X < \infty$  and let  $n$  be the minimum integer satisfying  $X \in Q_n(Y)$ . Then, by decomposing  $X = \sum_{m=0}^n X_m$  with  $X_m \in H_m(Y)$ , we obtain

$$\|X\|_{L^p} = \left\| T(G)^s \sum_{m=0}^n \lambda_m^{-s} X_m \right\|_{L^p} \leq \left\| \sum_{m=0}^n \lambda_m^{-s} X_m \right\|_{L^2} \leq \lambda_m^{-s} \|X\|_{L^2},$$

where we have used the  $(2, p; s)$ -hypercontractivity in the second inequality. ■

### (e) Proof of proposition 3.13

*Proof.* It suffices to consider  $X$  having the decomposition  $X = \sum_m X_m$  with  $X_m \in H_m(Y)$ . Recall that we have assumed that  $Q_0$  is the space of constant functions, so  $X_0$  is a constant. First, we consider the case  $X_0 = 0$ . In this case, for  $t > s$ , we have

$$\begin{aligned} \|T(G)^t X\|_{L^4}^2 &= \left\| \sum_{m \geq 1} \lambda_m^t X_m \right\|_{L^4}^2 \leq \left( \sum_{m \geq 1} \lambda_m^{t-s} \lambda_m^s \|X_m\|_{L^4} \right)^2 \\ &\leq \left( \sum_{m \geq 1} \lambda_m^{t-s} \|X_m\|_{L^2} \right)^2 \leq \left( \sum_{m \geq 1} \lambda_m^{2(t-s)} \right) \|X\|_{L^2}^2. \end{aligned} \quad (\text{Cauchy-Schwarz})$$

Therefore, when  $\sum_{m \geq 1} \lambda_m^{2(t-s)} \leq 1/\sqrt{3}$  we have

$$\|T(G)^t X\|_{L^4} \leq 3^{-1/4} \|X\|_{L^2} \quad (\text{A } 2)$$

for all  $X$  satisfying  $X_0 = 0$ .

In the case  $X_0 \neq 0$ , we can assume  $X_0 = 1$  without loss of generality. Let  $W = X - 1$  and  $Z = T(G)^t W = T(G)^t X - 1$ . Note that  $\mathbb{E}[W] = \mathbb{E}[Z] = 0$  holds by the orthogonality. We can explicitly expand the  $L^4$  norm as follows:

$$\begin{aligned} \|T(G)^t X\|_{L^4}^4 &= 1 + 6\mathbb{E}[Z^2] + 4\mathbb{E}[Z^3] + \mathbb{E}[Z^4] \\ &\leq 1 + 8\mathbb{E}[Z^2] + 3\mathbb{E}[Z^4]. \end{aligned} \quad (\text{A M-GM})$$

We also have

$$\|X\|_{L^2}^4 = \mathbb{E}[(1 + W)^2]^2 = (1 + \mathbb{E}[W^2])^2 = 1 + 2\mathbb{E}[W^2] + \mathbb{E}[W^2]^2.$$

So it suffices to show  $4\mathbb{E}[Z^2] \leq \mathbb{E}[W^2]$  and  $3\mathbb{E}[Z^4] \leq \mathbb{E}[W^2]^2$ , but the latter immediately follows from (A 2). The former holds when  $\lambda_1^t \leq 1/2$ :

$$\mathbb{E}[Z^2] = \sum_{m \geq 1} \lambda_m^{2t} \mathbb{E}[X_m^2] \leq \lambda_1^{2t} \mathbb{E}[W^2].$$

Therefore, we have completed the proof. ■

### (f) Proof of theorem 3.14

*Proof.* Let  $G = (Y, Q, \lambda)$  and  $\mathcal{X}$  be the space in which  $Y$  takes values. By truncating  $Q$  and  $\lambda$  (i.e. ignoring  $Q_m$  with  $1/\lambda_m > K$ ), we can assume that  $Q(Y) = \{X \in L^2 \mid \widetilde{\deg}_G X \leq K\}$ . Then, as  $\dim Q <$



$\infty$ , we can take a vector-valued measurable function

$$\varphi = (\varphi_1, \dots, \varphi_N)^\top : \mathcal{X} \rightarrow \mathbb{R}^N$$

such that  $(\varphi_i(Y))_{i=1}^N$  is an orthonormal basis of  $Q(Y)$ . Then, we have

$$\sup_{X \in Q(Y) \setminus \{0\}} \frac{\|X\|_{L^4}}{\|X\|_{L^2}} = \sup_{c \in \mathbb{R}^N \setminus \{0\}} \frac{\|c^\top \varphi(Y)\|_{L^4}}{\|c^\top \varphi(Y)\|_{L^2}} = \sup_{c \in \mathbb{R}^N, \|c\|=1} \|c^\top \varphi(Y)\|_{L^4} < \infty,$$

where the right-hand side is the supremum of a continuous function over a compact domain, and so is indeed finite. Hence, we can apply proposition 3.13, and there exists a constant  $s > 0$  such that

$$\|T(G)^t X\|_{L^4} \leq \|X\|_{L^2}, \quad X \in Q(Y),$$

because  $\lambda_1 < 1$  and  $(\lambda_m)_m$  is of finite length now. So  $G = (Y, Q, \lambda)$  (with truncation by  $K$ ) is actually  $(2, p; t)$ -hypercontractive and it extends to  $G^{\otimes d}$  for any  $d$  by theorem 3.9 (note that the truncation does not affect the random variables with  $\deg_{G^{\otimes d}} X \leq K$ ). Then, we finally use proposition 3.11 to obtain the desired result with  $C = K^t$ . ■

### (g) Proof of proposition 4.12

*Proof.* Let  $f \in L^2(\mu)$  be an eigenfunction with eigenvalue  $\lambda \geq 0$  of the integral operator, i.e. it satisfies  $\int_{\mathcal{X}} k(x, y)f(y) d\mu(y) = \lambda f(x)$  (assume this equality holds for all  $x$ , not just  $\mu$ -almost all). As  $\sum_{\ell=1}^{\infty} \sigma_{\ell} < \infty$  and assumption 4.10 is met from the general theory [48], it suffices to show  $\lambda \geq 1$  if and only if  $f$  is constant. Note that  $f = 1$  is an eigenfunction for  $\lambda = 1$  by assumption.

Assume  $\lambda \geq 1$ . Since  $k$  is bounded from the assumption, for an  $(x_n)_{n=1}^{\infty}$  converging to  $x$ , we have  $f(x_n) = (1/\lambda) \int_{\mathcal{X}} k(x_n, y)f(y) d\mu(y) \rightarrow (1/\lambda) \int_{\mathcal{X}} k(x, y)f(y) d\mu(y) = f(x)$  by the dominated convergence theorem. Thus,  $f$  is continuous. Let  $F = \max_{x \in \mathcal{X}} f(x)$ . If  $x^* \in f^{-1}(\{F\})$ , then

$$0 = F - f(x^*) = \int_{\mathcal{X}} k(x^*, y) \left( F - \frac{1}{\lambda} f(y) \right) d\mu(y).$$

As  $k(x^*, \cdot)$  is a probability density (recall  $k \geq 0$  from the assumption) with respect to  $\mu$  and  $\text{supp } \mu = \mathcal{X}$ , we must have  $\lambda \leq 1$  and  $k(x^*, y) = 0$  for all  $y \notin f^{-1}(\{F\})$ . Now, it suffices to prove  $f^{-1}(\{F\}) = \mathcal{X}$  actually holds when  $\lambda = 1$ . Let  $K = \max_{x, y \in \mathcal{X}} k(x, y)$ . By taking an  $\varepsilon > 0$  such that  $\mu(f^{-1}([F - \varepsilon, F])) \leq 1/(2K)$ , we have, for  $x \notin f^{-1}(\{F\})$ ,

$$\begin{aligned} f(x) &= \int_{\mathcal{X}} k(x, y)f(y) d\mu(y) \\ &\leq \int_{f^{-1}((-\infty, F-\varepsilon))} k(x, y)f(y) d\mu(y) + \int_{f^{-1}([F-\varepsilon, F])} k(x, y)f(y) d\mu(y) \\ &\leq (F - \varepsilon) \int_{f^{-1}((-\infty, F-\varepsilon))} k(x, y) d\mu(y) + F \int_{f^{-1}([F-\varepsilon, F])} k(x, y) d\mu(y) \\ &\leq (F - \varepsilon) + \varepsilon \int_{f^{-1}([F-\varepsilon, F])} k(x, y) d\mu(y) \leq (F - \varepsilon) + \frac{\varepsilon}{2} = F - \frac{\varepsilon}{2}. \end{aligned}$$

Therefore, if  $f^{-1}(\{F\}) = \mathcal{X}$ ,  $f$  is disconnected (because  $\mathcal{X}$  is path-connected), and it is contradiction. This completes the proof. ■

### (h) Proof of proposition 4.13

We first prove the following lemma.

**Lemma A.3.** For  $p > 2$ , we have  $\|\mathcal{K}_0 f\|_{L^p} \leq \|k_0\|_{L^p(\mu \otimes \mu)} \|f\|_{L^2}$  for all  $f \in L^2(\mu)$ .

*Proof.* By Minkowski's integral inequality, we have

$$\begin{aligned}
 \|\mathcal{K}_0 f\|_{L^p} &= \left( \int_{\mathcal{X}} \left| \int_{\mathcal{X}} k_0(x, y) f(y) \, d\mu(y) \right|^p \, d\mu(x) \right)^{1/p} \\
 &\leq \int_{\mathcal{X}} \left( \int_{\mathcal{X}} |k_0(x, y) f(y)|^p \, d\mu(x) \right)^{1/p} \, d\mu(y) \\
 &\leq \int_{\mathcal{X}} \left( \int_{\mathcal{X}} |k_0(x, y)|^p \, d\mu(x) \right)^{1/p} |f(y)| \, d\mu(y) \\
 &\leq \left( \int_{\mathcal{X}} \left( \int_{\mathcal{X}} |k_0(x, y)|^p \, d\mu(x) \right)^{2/p} \, d\mu(y) \right)^{1/2} \|f\|_{L^2} \quad (\text{Cauchy-Schwarz}) \\
 &\leq \|k_0\|_{L^p(\mu \otimes \mu)} \|f\|_{L^2}.
 \end{aligned}$$

From this lemma, we have

$$\|e_m\|_{L^p} = \frac{1}{\sigma_m} \|\mathcal{K}_0 e_m\|_{L^p} \leq \frac{\|k_0\|_{L^p(\mu \otimes \mu)}}{\sigma_m} \|e_m\|_{L^2} \quad (\text{A } 3)$$

for each  $m \geq 2$ .

*Proof of proposition 4.13.* It suffices to consider the case  $\|k_0\|_{L^4(\mu \otimes \mu)} < \infty$ . Note that  $\lambda_{\ell-1} = \sigma_{\ell}$  for  $\ell = 1, 2, \dots$  for the GRP  $G_{k, \mu}$ , so  $\lambda_1 = \sigma_2 = \|\mathcal{K}_0\|$ .

Let  $r_0$  be the minimum nonnegative number satisfying  $\|\mathcal{K}_0\|^{-r_0} \geq \sqrt{3} \operatorname{tr}(\mathcal{K}_0)$ . Then, for  $r := 1 + r_0$ , we have

$$\sum_{\ell=2}^{\infty} \sigma_{\ell}^r \leq \sigma_2^{r_0} \sum_{\ell=2}^{\infty} \sigma_{\ell} = \|\mathcal{K}\|^{r_0} \operatorname{tr}(\mathcal{K}_0) \leq \frac{1}{\sqrt{3}} \quad (\text{A } 4)$$

Let  $s_0$  be the minimum non-negative number satisfying  $\|\mathcal{K}_0\|^{-s_0} \geq \|k_0\|_{L^4(\mu \otimes \mu)}$ . As  $\|\mathcal{K}_0\| \in (0, 1)$  from assumption 4.7,  $s_0$  is well-defined. Then, for  $s := 1 + s_0$  and  $m \geq 2$ , from (A 3), we have

$$\|e_m\|_{L^4} \leq \frac{\|k_0\|_{L^4(\mu \otimes \mu)}}{\sigma_m} \|e_m\|_{L^2} \leq \frac{1}{\sigma_m \|\mathcal{K}_0\|^{s_0}} \|e_m\|_{L^2} \leq \sigma_m^{-1-s_0} \|e_m\|_{L^2}. \quad (\text{A } 5)$$

Thus, the condition for  $s$  and  $t := r + s$  of proposition 3.13 is satisfied, and so we have the desired conclusion. ■

## References

1. Richter H. 1957 Parameterfreie abschätzung und realisierung von erwartungswerten. *Bl. DGVFM* **3**, 147–162. (doi:10.1007/BF02808864)
2. Rogosinski WW. 1958 Moments of non-negative mass. *Proc. R. Soc. Lond. A* **245**, 1–27. (doi:10.1098/rspa.1958.0062)
3. Rosenbloom PC. 1952 Quelques classes de problèmes extrémaux. II. *Bull. Soc. Math. France* **80**, 183–215. (doi:10.24033/bsmf.1429)
4. Tchakaloff V. 1957 Formules de cubature mécanique à coefficients non négatifs. *Bull. Sci. Math.* **81**, 123–134.
5. Wald A. 1939 Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments. *Trans. Am. Math. Soc.* **46**, 280–306. (doi:10.1090/S0002-9947-1939-0000076-8)
6. di Dio PJ, Schmudgen K. 2018 The multidimensional truncated moment problem: the moment cone. (<https://arxiv.org/abs/1809.00584>)
7. Lyons T, Victoir N. 2004 Cubature on Wiener space. *Proc. R. Soc. A* **460**, 169–198. (doi:10.1098/rspa.2003.1239)
8. Hayakawa S, Oberhauser H, Lyons T. 2022 Positively weighted kernel quadrature via subsampling. *Adv. Neural Inf. Process. Syst.* **35**, 6886–6900.

9. Karvonen T, Särkkä S. 2019 Gaussian kernel quadrature at scaled Gauss–Hermite nodes. *BIT Numer. Math.* **59**, 877–902. (doi:10.1007/s10543-019-00758-3)
10. Hayakawa S. 2021 Monte Carlo cubature construction. *Jpn. J. Ind. Appl. Math.* **38**, 561–577. (doi:10.1007/s13160-020-00451-x)
11. Beckner W. 1975 Inequalities in fourier analysis. *Ann. Math.* **102**, 159–182. (doi:10.2307/1970980)
12. Beckner W. 1992 Sobolev inequalities, the Poisson semigroup, and analysis on the sphere  $s^n$ . *Proc. Natl Acad. Sci. USA* **89**, 4816–4819. (doi:10.1073/pnas.89.11.4816)
13. Bonami A. 1970 Étude des coefficients de Fourier des fonctions de  $L^p(G)$ . *Ann. Inst. Fourier* **20**, 335–402. (doi:10.5802/aif.357)
14. Nelson E. 1973 The free Markoff field. *J. Funct. Anal.* **12**, 211–227. (doi:10.1016/0022-1236(73)90025-6)
15. Hayakawa S, Lyons T, Oberhauser H. 2023 Estimating the probability that a given vector is in the convex hull of a random sample. *Probab. Theory Relat. Fields* **185**, 705–746. (doi:10.1007/s00440-022-01186-1)
16. Stroud AH. 1971 *Approximate calculation of multiple integrals*. Englewood Cliffs, NJ: Prentice-Hall.
17. Litterer C, Lyons T. 2012 High order recombination and an application to cubature on Wiener space. *Ann. Appl. Probab.* **22**, 1301–1327. (doi:10.1214/11-AAP786)
18. Maalouf A, Jubran I, Feldman D. 2019 Fast and accurate least-mean-squares solvers. In *Advances in neural information processing systems* (eds H Wallach, H Larochelle, A Beygelzimer, F d'Alché Buc, E Fox, R Garnett), vol. 32, pp. 8305–8316. Vancouver: Curran Associates, Inc.
19. Tchernychova M. 2015 Carathéodory cubature measures. PhD thesis, University of Oxford.
20. Cosentino F, Oberhauser H, Abate A. 2020 A randomized algorithm to reduce the support of discrete measures. *Adv. Neural Inf. Process. Syst.* **33**, 15 100–15 110.
21. Glaubitz J. 2021 Stable high-order cubature formulas for experimental data. *J. Comput. Phys.* **447**, 110693. (doi:10.1016/j.jcp.2021.110693)
22. Migliorati G, Nobile F. 2022 Stable high-order randomized cubature formulae in arbitrary dimension. *J. Approx. Theory* **275**, 105706. (doi:10.1016/j.jat.2022.105706)
23. Mei S, Misiakiewicz T, Montanari A. 2021 Generalization error of random feature and kernel methods: hypercontractivity and kernel matrix concentration. *Appl. Comput. Harmon. Anal.* **59**, 3–84. (doi:10.1016/j.acha.2021.12.003)
24. Kim JH, Vu VH. 2000 Concentration of multivariate polynomials and its applications. *Combinatorica* **20**, 417–434. (doi:10.1007/s004930070014)
25. Schudy W, Sviridenko M. 2012 Concentration and moment inequalities for polynomials of independent random variables. In *Proc. of the 23rd Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 437–446. Philadelphia, PA: SIAM.
26. Wendel JG. 1963 A problem in geometric probability. *Math. Scand.* **11**, 109–111. (doi:10.7146/math.scand.a-10655)
27. Wagner U, Welzl E. 2001 A continuous analogue of the upper bound theorem. *Discrete Comput. Geom.* **26**, 205–219. (doi:10.1007/s00454-001-0028-9)
28. Kabluchko Z, Zaporozhets D. 2020 Absorption probabilities for Gaussian polytopes and regular spherical simplices. *Adv. Appl. Probab.* **52**, 588–616. (doi:10.1017/apr.2020.7)
29. Caplin A, Nalebuff B. 1991 Aggregation and social choice: a mean voter theorem. *Econometrica* **59**, 1–23. (doi:10.2307/2938238)
30. Eskenazis A, Nayar P, Tkocz T. 2018 Sharp comparison of moments and the log-concave moment problem. *Adv. Math.* **334**, 389–416. (doi:10.1016/j.aim.2018.06.014)
31. König H. 2014 On the best constants in the Khintchine inequality for Steinhaus variables. *Isr. J. Math.* **203**, 23–57. (doi:10.1007/s11856-013-0006-y)
32. Janson S. 1997 *Gaussian Hilbert spaces*. Cambridge, UK: Cambridge University Press.
33. Nourdin I, Peccati G, Reinert G. 2010 Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. *Ann. Probab.* **38**, 1947–1985. (doi:10.1214/10-AOP531)
34. Simon B, Høegh-Krohn R. 1972 Hypercontractive semigroups and two dimensional self-coupled Bose fields. *J. Funct. Anal.* **9**, 121–180. (doi:10.1016/0022-1236(72)90008-0)
35. Ninomiya S, Shinozaki Y. 2021 On implementation of high-order recombination and its application to weak approximations of stochastic differential equations. In *Proc. of the NFA 29th Annual Conf.* Tokyo: Nippon Finance Association.
36. Hayakawa S, Tanaka K. 2022 Monte Carlo construction of cubature on Wiener space. *Jpn. J. Ind. Appl. Math.* **39**, 543–571. (doi:10.1007/s13160-021-00496-6)

37. Passeggeri R. 2016 Some results on the signature and cubature of the fractional Brownian motion for  $H > \frac{1}{2}$ . Preprint. (<https://arxiv.org/abs/1609.07352>)
38. Baudoin F, Coutin L. 2007 Operators associated with a stochastic differential equation driven by fractional Brownian motions. *Stoch. Process. Appl.* **117**, 550–574. (doi:10.1016/j.spa.2006.09.004)
39. Berlinet A, Thomas-Agnan C. 2004 *Reproducing kernel Hilbert spaces in probability and statistics*. New York, NY: Springer.
40. Bach F, Lacoste-Julien S, Obozinski G. 2012 On the equivalence between herding and conditional gradient algorithms. In *Int. Conf. on Machine Learning*, pp. 1355–1362. Madison, WI, USA: Edinburgh/Omnipress.
41. Chen Y, Welling M, Smola A. 2010 Super-samples from kernel herding. In *Proc. of the 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010)* (eds P Grünwald, P Spirtes), pp. 109–116. Corvallis, Oregon: AUAI Press.
42. Huszár F, Duvenaud D. 2012 Optimally-weighted herding is Bayesian quadrature. In *Proc. of the 28th Conference on Uncertainty in Artificial Intelligence (UAI 2012)* (eds K Murphy, N de Freitas), pp. 377–386. Corvallis, Oregon: AUAI Press.
43. Bach F. 2017 On the equivalence between kernel quadrature rules and random feature expansions. *J. Mach. Learn. Res.* **18**, 714–751.
44. Belhadji A, Bardenet R, Chainais P. 2019 Kernel quadrature with DPPs. *Adv. Neural Inf. Process. Syst.* **32**, 12 907–12 917.
45. Kanagawa M, Sriperumbudur BK, Fukumizu K. 2016 Convergence guarantees for kernel-based quadrature rules in misspecified settings. *Adv. Neural Inf. Process. Syst.* **29**, 3296–3304.
46. O'Hagan A. 1991 Bayes–Hermite quadrature. *J. Stat. Plann. Inference* **29**, 245–260. (doi:10.1016/0378-3758(91)90002-V)
47. Dick J, Kuo FY, Sloan IH. 2013 High-dimensional integration: the quasi-Monte Carlo way. *Acta Numer.* **22**, 133–288. (doi:10.1017/S0962492913000044)
48. Steinwart I, Scovel C. 2012 Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSs. *Constr. Approx.* **35**, 363–417. (doi:10.1007/s00365-012-9153-3)
49. Zwicknagl B. 2009 Power series kernels. *Constr. Approx.* **29**, 61–84. (doi:10.1007/s00365-008-9012-4)
50. Hayakawa S, Oberhauser H, Lyons T. 2023 Sampling-based Nyström approximation and kernel quadrature. Preprint. (<https://arxiv.org/abs/2301.09517>)
51. Anastasiou A *et al.* 2021 Stein's method meets statistics: a review of some recent developments. Preprint. (<https://arxiv.org/abs/2105.03481>)
52. Oates C, Girolami M, Chopin N. 2017 Control functionals for Monte Carlo integration. *J. R. Stat. Soc. B* **79**, 695–718. (doi:10.1111/rssb.12185)
53. Wahba G. 1990 *Spline models for observational data*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
54. Henry-Labordere P, Tan X, Touzi N. 2017 Unbiased simulation of stochastic differential equations. *Ann. Appl. Probab.* **27**, 3305–3341. (doi:10.1214/17-AAP1281)
55. Rousseeuw PJ, Ruts I. 1999 The depth function of a population distribution. *Metrika* **49**, 213–244. (doi:10.1007/PL00020903)
56. Tukey JW. 1974 Mathematics and the picturing of data. In *Proc. of the Int. Congress of Mathematicians, Vancouver, 1974*, vol. 2, pp. 523–531. Vancouver: Canadian Mathematical Congress.
57. An MY. 1997 Log-concave probability distributions: theory and statistical testing. *Duke University Dept of Economics Working Paper*.
58. Boyd S, Boyd SP, Vandenberghe L. 2004 *Convex optimization*. Cambridge, UK: Cambridge University Press.
59. Grünbaum B. 1960 Partitions of mass-distributions and of convex bodies by hyperplanes. *Pac. J. Math.* **10**, 1257–1261. (doi:10.2140/pjm.1960.10.1257)
60. Lovász L, Vempala S. 2007 The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms* **30**, 307–358. (doi:10.1002/rsa.20135)
61. Hardy GH, Littlewood JE, Pólya G. 1952 *Inequalities*. Cambridge, UK: Cambridge University Press.