Monte Carlo convergence rates for kth moments in Banach spaces

Kirchner, Kristin; Schwab, Christoph

DOI
10.1016/j.jfa.2023.110218

Publication date
2024

Document Version
Final published version

Published in
Journal of Functional Analysis

Journal of Functional Analysis 286 (2024) 110218

Contents lists available at ScienceDirect

Journal of Functional Analysis

journal homepage: www.elsevier.com/locate/jfa

Regular Article

Monte Carlo convergence rates for kth moments in Banach spaces

Kristin Kirchner [a,*], Christoph Schwab [b]

[a] *Delft Institute of Applied Mathematics, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands*
[b] *Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland*

ARTICLE INFO

Article history:
Received 15 December 2022
Accepted 17 October 2023
Available online 23 October 2023
Communicated by Sophie Grivaux

MSC:
primary 65C05
secondary 46A32, 60B11, 60H35

Keywords:
Injective tensor product
Multilevel Monte Carlo
Rademacher averages
Type of Banach space

ABSTRACT

We formulate standard and multilevel Monte Carlo methods for the kth moment $\mathbb{M}^k_\varepsilon[\xi]$ of a Banach space valued random variable $\xi\colon \Omega \to E$, interpreted as an element of the k-fold injective tensor product space $\otimes^k_\varepsilon E$. For the standard Monte Carlo estimator of $\mathbb{M}^k_\varepsilon[\xi]$, we prove the k-independent convergence rate $1-\frac{1}{p}$ in the $L_q(\Omega; \otimes^k_\varepsilon E)$-norm, provided that (i) $\xi \in L_{kq}(\Omega; E)$ and (ii) $q \in [p, \infty)$, where $p \in [1, 2]$ is the Rademacher type of E. By using the fact that Rademacher averages are dominated by Gaussian sums combined with a version of Slepian's inequality for Gaussian processes due to Fernique, we moreover derive corresponding results for multilevel Monte Carlo methods, including a rigorous error estimate in the $L_q(\Omega; \otimes^k_\varepsilon E)$-norm and the optimization of the computational cost for a given accuracy. Whenever the type of the Banach space E is $p = 2$, our findings coincide with known results for Hilbert space valued random variables.
We illustrate the abstract results by three model problems: second-order elliptic PDEs with random forcing or random coefficient, and stochastic evolution equations. In these cases, the solution processes naturally take values in non-Hilbertian Banach spaces. Further applications, where physical modeling

* Corresponding author.
E-mail address: k.kirchner@tudelft.nl (K. Kirchner).

https://doi.org/10.1016/j.jfa.2023.110218

constraints impose a setting in Banach spaces of type $p < 2$, are indicated.

1. Introduction

1.1. Background and motivation

Many applications in uncertainty quantification require the estimation of statistical moments. The first statistical moment, that is to say the mean, is often itself the quantity of interest, whereas higher-order moments are needed to infer certain characteristics about the probability distribution of the underlying real- or vector-valued random variable. In the case that this distribution is Gaussian, it is fully determined by the first two statistical moments. Third-order and fourth-order moments, which define the skewness and kurtosis of the probability distribution, play for instance an important role for tests if the distribution is Gaussian, see e.g. [38,49].

In order to estimate statistical moments one resorts to sampling strategies, i.e., Monte Carlo methods. It is well-known that for estimating the mean the convergence rate $1/2$ in the number of samples is achieved as long as the random variable ξ is square-integrable in Bochner sense with values in a Hilbert space H, i.e., $\xi \in L_2(\Omega; H)$. Moreover, this result extends to statistical moments of an arbitrary order $k \in \mathbb{N}$ when interpreted as elements of the Hilbert tensor product space $H^{(k)}$, provided that the random variable exhibits sufficient integrability in $L_{2k}(\Omega; H)$.

Vector-valued random variables occur, for instance, in the context of differential equations involving randomness. Here, numerical methods for generating samples of approximate solutions often allow for a hierarchical multilevel structure corresponding to different degrees of refinement of the discretization parameters. The idea of multilevel Monte Carlo (MLMC) methods is to reduce the computational cost for achieving a given target accuracy by optimizing the number of samples used on each level to compute the MLMC estimator. To the best of our knowledge this approach was first formulated by Giles [22] for stochastic ordinary differential equations (SDEs) after having previously been introduced by Heinrich [32] in the context of numerical integration. Since then MLMC methods have been used to approximate means of Hilbert space valued random variables for a variety of problems in uncertainty quantification, including but not limited to SDEs [13,21,23,25,58], partial differential equations (PDEs) with random coefficients [3,10,12,13,26,27,31,56], stochastic PDEs [2,24], and hyperbolic PDEs with random fluxes or uncertainties in the initial data [50–52].

With regard to higher-order moments, MLMC strategies have been applied to estimate *diagonals of central statistical moments* in [5,6], and combined with sparse tensor techniques to approximate (*full*) moments in [3,51,52]; the latter approach has been refined and generalized by the multi-index Monte Carlo method in [30].

All of the previously mentioned references have in common that the error analysis of the (multilevel) Monte Carlo estimators proceeds in Hilbert spaces. For the first statistical moment, it is known that estimating means of random variables taking values in a Banach space E via the standard Monte Carlo method *does in general not converge at the rate* $1/2$, even in the presence of high Bochner integrability. More specifically, the rate of convergence depends on geometric properties of the Banach space E: If E has *Rademacher type* $p \in [1, 2]$ and the random variable is an element of the Bochner space $L_q(\Omega; E)$, $q \geq 1$, the standard Monte Carlo method will, in general, converge only at the rate $1 - \frac{1}{\min\{p,q\}}$, see [47, Proposition 9.11]. This behavior necessitates tailoring of MLMC methods not only to the discretization of a particular problem, but also to the type of the Banach space. The only references known to the authors addressing this issue are [17] and [42], where conservation laws with random data are discretized by MLMC finite difference methods, and [14], where the authors perform a MLMC analysis in Hölder spaces for solutions to stochastic evolution equations.

Besides the aforementioned issue of type-dependent convergence rates, another difficulty occurs when considering higher-order moments of Banach space valued random variables: As opposed to the Hilbert space case, there is no canonical choice for the norm on the tensor product space $E \otimes E$. Two options which are widely used in the literature are the *projective* and *injective tensor product norms*, mostly caused by the fact that any reasonable cross norm is bounded from above, respectively from below, by these norms, see [54, Proposition 6.1(a)]. Janson and Kaijser [37] defined and analyzed statistical moments of order $k \in \mathbb{N}$ (in Bochner, Dunford or Pettis sense) as elements of projective and injective tensor product spaces. One of the findings [37, Theorem 3.8] shows that both the *projective* and *injective kth moment* of $\xi \colon \Omega \to E$ exist in Bochner sense (and coincide) whenever $\xi \in L_k(\Omega; E)$.

Clearly, the choice of the tensor product norm will play a crucial role in the error analysis of (multilevel) Monte Carlo estimation of higher-order statistical moments. A simple argumentation (see Example 3.21) shows that, no matter how "good" the (Rademacher) type of the Banach space E is, Monte Carlo methods for the second moment will in general not converge in the projective tensor product of E.

While this work is devoted to the numerical analysis of Monte Carlo (sampling) methods to approximate higher-order moments of vector-valued random variables, we remark that certain *linear* (or *linearized*) stochastic equations allow for alternative approaches to *deterministically compute* approximations to k-point correlations of random solutions. In this context, we mention [11,40,41,43,46] and the references therein. This methodology does not raise the mathematical issue of the type of a Banach space and its impact on the convergence of numerical approximations. In the present paper, we shall not pursue this direction further.

1.2. Contributions

We consider the injective kth moment $\mathbb{M}_\varepsilon^k[\xi]$ of a Banach space valued random variable $\xi \in L_k(\Omega; E)$ and, for the first time, formulate standard and multilevel Monte Carlo methods for this higher-order moment in the Banach space setting. We prove that the standard Monte Carlo estimator for $\mathbb{M}_\varepsilon^k[\xi]$ converges in the $L_q(\Omega; \otimes_\varepsilon^k E)$-norm at the rate $1 - \frac{1}{p}$ provided that (i) $\xi \in L_{kq}(\Omega; E)$ and (ii) $q \in [p, \infty)$, where $p \in [1, 2]$ is the Rademacher type of E, see Theorem 3.16. Here, $\otimes_\varepsilon^k E$ denotes the k-fold injective tensor product of E. Note, in particular, that this convergence rate is independent of the order k of the statistical moment $\mathbb{M}_\varepsilon^k[\xi]$. This result readily implies error estimates and convergence rates for abstract *single-level Monte Carlo* methods for $\mathbb{M}_\varepsilon^k[\xi]$, see Corollary 3.20.

By means of replacing Rademacher sums by Gaussian averages and exploiting a version of Slepian's inequality for Gaussian processes due to Fernique [18], we are furthermore able to formulate corresponding abstract results for *multilevel Monte Carlo* methods. This includes a rigorous error estimate in the $L_q(\Omega; \otimes_\varepsilon^k E)$-norm, see Theorem 3.24, and the optimization of the computational cost for a given accuracy (in the MLMC context also known as "$\alpha\beta\gamma$ theorem"), see Theorem 3.25.

We apply these abstract results to several classes of problems, where the stochastic solution processes naturally take values in (non-Hilbertian) Banach spaces: second-order elliptic PDEs (a) with random forcing taking values in L_p for some general $p \in (1, \infty)$, or (b) with log-Gaussian diffusion coefficient and right-hand side in L_p; and (c) stochastic evolution equations, where we extend the MLMC analysis in Hölder norms, performed in [14, Section 5] for mean values of solution processes, to their kth moments, being kth order (spatio-)temporal correlation functions.

1.3. Layout

In Section 2 we introduce the necessary notation, see Subsection 2.1, as well as the analytical preliminaries on tensor products of a Banach space E, with particular emphasis on the full and symmetric k-fold injective tensor product of E, see Subsection 2.2. We close this section with the definition of the injective kth moment of a Banach space valued random variable $\xi \in L_k(\Omega; E)$ in Subsection 2.3. Section 3 is dedicated to the analysis of Monte Carlo methods for the injective kth moment. For this purpose, we first need to formulate several auxiliary results for Rademacher and Gaussian averages in Subsection 3.1. We then perform the error analysis for the standard (and single-level) Monte Carlo method in Subsection 3.2 and for the multilevel Monte Carlo method in Subsection 3.3. In Section 4 we discuss several applications of our convergence results. Section 5 gives an outlook on extensions and further applications, where due to essential restrictions in the physical modeling a (non-Hilbertian) Banach space setting cannot be avoided.

2. Preliminaries

2.1. General notation and setting

Given parameter sets $\mathscr{P}$, $\mathscr{Q}$, and mappings $\mathscr{F}, \mathscr{G} \colon \mathscr{P} \times \mathscr{Q} \to \mathbb{R}$, we use the notation $\mathscr{F}(p,q) \lesssim_q \mathscr{G}(p,q)$ to indicate that for each $q \in \mathscr{Q}$ there exists a constant $C_q \in (0,\infty)$ such that $\mathscr{F}(p,q) \leq C_q \mathscr{G}(p,q)$ holds for every $p \in \mathscr{P}$. Whenever both relations, $\mathscr{F}(p,q) \lesssim_q \mathscr{G}(p,q)$ and $\mathscr{G}(p,q) \lesssim_q \mathscr{F}(p,q)$, hold simultaneously, we write $\mathscr{F}(p,q) \eqsim_q \mathscr{G}(p,q)$.

For a Banach space $(F, \|\cdot\|_F)$ over $\mathbb{R}$, we write $B_F := \{x \in F : \|x\|_F \leq 1\}$ for its closed unit ball, and $\mathcal{B}(F)$ for the Borel σ-algebra on $(F, \|\cdot\|_F)$, that is the σ-algebra generated by the open sets. The dual space of all continuous linear functionals $g \colon F \to \mathbb{R}$ is denoted by F'. We write $g(x)$ or $\langle g, x \rangle$ for the duality pairing between $g \in F'$ and $x \in F$, and $\|g\|_{F'} := \sup_{x \in B_F} |g(x)|$ for the norm on F'.

Throughout this article, we assume that $(\Omega, \mathcal{A}, \mathbb{P})$ is a complete probability space with expectation operator $\mathbb{E}$, and we mark statements which hold $\mathbb{P}$-almost surely with $\mathbb{P}$-a.s. For a second complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ with expectation operator $\widetilde{\mathbb{E}}$, $(\Omega \times \widetilde{\Omega}, \mathcal{A} \otimes \widetilde{\mathcal{A}}, \mathbb{P} \otimes \widetilde{\mathbb{P}})$ denotes the product probability space, i.e., $\Omega \times \widetilde{\Omega}$ is the set of all tuples $(\omega, \widetilde{\omega})$ with $\omega \in \Omega$, $\widetilde{\omega} \in \widetilde{\Omega}$, $\mathcal{A} \otimes \widetilde{\mathcal{A}}$ is the product σ-algebra generated by all sets of the form $A \times \widetilde{A}$ with $A \in \mathcal{A}$, $\widetilde{A} \in \widetilde{\mathcal{A}}$, and $\mathbb{P} \otimes \widetilde{\mathbb{P}}$ is the uniquely defined product measure satisfying $(\mathbb{P} \otimes \widetilde{\mathbb{P}})(A \times \widetilde{A}) = \mathbb{P}(A)\widetilde{\mathbb{P}}(\widetilde{A})$ for all $A \in \mathcal{A}$ and $\widetilde{A} \in \widetilde{\mathcal{A}}$. The expectation operator on $(\Omega \times \widetilde{\Omega}, \mathcal{A} \otimes \widetilde{\mathcal{A}}, \mathbb{P} \otimes \widetilde{\mathbb{P}})$ will be denoted by $\mathbb{E} \otimes \widetilde{\mathbb{E}}$.

In addition, we let $(E, \|\cdot\|_E)$ be a Banach space over $\mathbb{R}$, the set $\mathbb{N}$ contains all (strictly) positive integers and, unless otherwise stated, $k \in \mathbb{N}$ is a fixed positive integer which indicates the order of (statistical) moments.

2.2. k-fold tensor products of Banach spaces

In this subsection we define (full and symmetric) k-fold tensor products of the Banach space $(E, \|\cdot\|_E)$, with the aim of obtaining a new Banach space satisfying the following:

(i) It contains the set of all kth moments (in Bochner sense) of Bochner integrable random variables $\xi \colon \Omega \to E$ satisfying $\mathbb{E}\big[\|\xi\|_E^k\big] < \infty$.
(ii) The topology on this space (prescribed by its norm) allows to quantify the convergence of Monte Carlo estimation for statistical moments of order k.

For this purpose, symmetry of moments will be particularly important.

We start by defining the (full) *k-fold algebraic tensor product of E*,

$$\otimes^k E = \underbrace{E \otimes \cdots \otimes E}_{k \text{ times}},$$

that is the vector space consisting of all finite sums of the form

$$\sum_{j=1}^{M} x_{j,1} \otimes \cdots \otimes x_{j,k} = \sum_{j=1}^{M} \bigotimes_{n=1}^{k} x_{j,n}, \qquad x_{j,n} \in E, \quad 1 \le j \le M, \quad 1 \le n \le k,$$

equipped with the algebraic operations rendering the tensor product linear in each of its k components, see [19, Section 1.1]. The *injective tensor norm* $\|\cdot\|_\varepsilon$ for an element $U \in \otimes^k E$ with representation $U = \sum_{j=1}^{M} x_{j,1} \otimes \cdots \otimes x_{j,k}$ is defined by

$$\|U\|_\varepsilon := \sup\left\{ \left| \sum_{j=1}^{M} \prod_{n=1}^{k} f_n(x_{j,n}) \right| \,\middle|\, f_1, \ldots, f_k \in B_{E'} \right\}, \tag{2.1}$$

cf. [54, Section 3.1] and [37, Section 2.3.2]. Note that the value of the supremum in (2.1) is independent of the choice of the representation of $U \in \otimes^k E$, since

$$\left| \sum_{j=1}^{M} \prod_{n=1}^{k} f_n(x_{j,n}) \right| = \left| \sum_{j=1}^{M} \langle (f_1 \otimes \cdots \otimes f_k), (x_{j,1} \otimes \cdots \otimes x_{j,k}) \rangle \right| = |\langle (f_1 \otimes \cdots \otimes f_k), U \rangle|.$$

For $k = 1$, the Hahn–Banach theorem shows that $\|\cdot\|_\varepsilon = \|\cdot\|_E$. We call the completion of the k-fold algebraic tensor product space $\otimes^k E$ with respect to $\|\cdot\|_\varepsilon$ in (2.1) the (full) *k-fold injective tensor product of E* and denote it by $\otimes_\varepsilon^k E$.

In the context of moments of E-valued random variables, we will be interested in subspaces of $\otimes^k E$ and $\otimes_\varepsilon^k E$ containing only their symmetric elements. To this end, we first introduce for $x_1 \otimes \cdots \otimes x_k \in \otimes^k E$ its *symmetrization*

$$s(x_1 \otimes \cdots \otimes x_k) := \frac{1}{k!} \sum_{\sigma \in S_k} x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(k)}, \tag{2.2}$$

where S_k is the group of permutations of the set $\{1, ..., k\}$. The *k-fold symmetric algebraic tensor product of E*, denoted by $\otimes^{k,s} E$, is then defined as the linear span of the subset $\{s(x_1 \otimes \cdots \otimes x_k) : x_1, \ldots, x_k \in E\}$ in $\otimes^k E$, i.e.,

$$\begin{aligned}
\otimes^{k,s} E &:= \left\{ \sum_{j=1}^{M} s(x_{j,1} \otimes \cdots \otimes x_{j,k}) \,\middle|\, M \in \mathbb{N},\ x_{j,n} \in E,\ 1 \le j \le M,\ 1 \le n \le k \right\} \\
&= \left\{ \sum_{j=1}^{M} \lambda_j \otimes^k x_j \,\middle|\, M \in \mathbb{N},\ \lambda_j \in \mathbb{R},\ x_j \in E,\ 1 \le j \le M \right\} \\
&= \left\{ \sum_{j=1}^{M} \delta_j \otimes^k x_j \,\middle|\, M \in \mathbb{N},\ \delta_j \in \mathscr{E}(k),\ x_j \in E,\ 1 \le j \le M \right\},
\end{aligned}$$

see [19, Section 1.5] or [20, Section 1.1]. Here, we set

$$\otimes^k x := \underbrace{x \otimes \cdots \otimes x}_{k \text{ times}} \quad \forall x \in E, \quad \text{and} \quad \mathscr{E}(k) := \begin{cases} \{-1, 1\} & \text{if } k \text{ is even}, \\ \{1\} & \text{if } k \text{ is odd}. \end{cases}$$

The *symmetric injective tensor norm* $\|\cdot\|_{\varepsilon_s}$ on the k-fold symmetric algebraic tensor product space $\otimes^{k,s}E$ is given by (see [19, Section 3.1])

$$\|U\|_{\varepsilon_s} := \sup\left\{ \left| \sum_{j=1}^{M} \lambda_j f(x_j)^k \right| \,\middle|\, f \in B_{E'} \right\}, \tag{2.3}$$

if $U = \sum_{j=1}^{M} \lambda_j \otimes^k x_j$, where $\lambda_j \in \mathbb{R}$, $x_j \in E$ for $1 \leq j \leq M$. Note that, as for the injective tensor norm $\|\cdot\|_\varepsilon$, this definition does not depend on the choice of the representation of U. The *k-fold symmetric injective tensor product of E*, denoted by $\otimes_{\varepsilon_s}^{k,s}E$, is the completion of $\otimes^{k,s}E$ with respect to the norm $\|\cdot\|_{\varepsilon_s}$ in (2.3).

The symmetrization s in (2.2) extends linearly to a projection $s\colon \otimes^k E \to \otimes^{k,s}E$ and, since for $U = \sum_{j=1}^{M} x_{j,1} \otimes \cdots \otimes x_{j,k} \in \otimes^k E$ we moreover have that

$$\begin{aligned} \|s(U)\|_{\varepsilon_s} \leq \|s(U)\|_\varepsilon &= \left\| \sum_{j=1}^{M} \frac{1}{k!} \sum_{\sigma \in S_k} x_{j,\sigma(1)} \otimes \cdots \otimes x_{j,\sigma(k)} \right\|_\varepsilon \\ &\leq \frac{1}{k!} \sum_{\sigma \in S_k} \left\| \sum_{j=1}^{M} x_{j,\sigma(1)} \otimes \cdots \otimes x_{j,\sigma(k)} \right\|_\varepsilon = \left\| \sum_{j=1}^{M} x_{j,1} \otimes \cdots \otimes x_{j,k} \right\|_\varepsilon = \|U\|_\varepsilon, \end{aligned} \tag{2.4}$$

it also has a unique continuous extension to a linear projection $s_\varepsilon\colon \otimes_\varepsilon^k E \to \otimes_{\varepsilon_s}^{k,s}E$. Furthermore, on $\otimes_{\varepsilon_s}^{k,s}E$ the injective tensor norm $\|\cdot\|_\varepsilon$ and the symmetric injective tensor norm $\|\cdot\|_{\varepsilon_s}$ are equivalent, with k-dependent equivalence constants,

$$\|U\|_{\varepsilon_s} \leq \|U\|_\varepsilon \leq \frac{k^k}{k!} \|U\|_{\varepsilon_s} \quad \forall U \in \otimes_{\varepsilon_s}^{k,s}E, \tag{2.5}$$

see [20, Sections 2.3 and 2.7].

Remark 2.1. There are several meaningful options to define norms on the k-fold algebraic tensor product spaces $\otimes^k E$ and $\otimes^{k,s}E$. Besides the injective tensor norm, a common choice is the *projective tensor norm* $\|\cdot\|_\pi$, defined for $U \in \otimes^k E$ by

$$\|U\|_\pi := \inf\left\{ \sum_{j=1}^{M} \prod_{n=1}^{k} \|x_{j,n}\|_E \,\middle|\, M \in \mathbb{N},\, U = \sum_{j=1}^{M} x_{j,1} \otimes \cdots \otimes x_{j,k} \right\}. \tag{2.6}$$

The *symmetric projective tensor norm* on $\otimes^{k,s}E$ is given by (see [19, Section 2.2])

$$\|U\|_{\pi_s} := \inf\left\{ \sum_{j=1}^{M} |\lambda_j| \, \|x_j\|_E^k \,\middle|\, M \in \mathbb{N},\, U = \sum_{j=1}^{M} \lambda_j \otimes^k x_j \right\}. \tag{2.7}$$

Then, for every $U \in \otimes^k E$, $\|U\|_\pi \geq \|U\|_\varepsilon$ and $\|s(U)\|_{\pi_s} \geq \max\{\|s(U)\|_{\varepsilon_s}, \|s(U)\|_\pi\}$ hold. The closures of $\otimes^k E$ and of $\otimes^{k,s}E$ with respect to the norms $\|\cdot\|_\pi$ and $\|\cdot\|_{\pi_s}$, respectively, yield well-defined Banach spaces, the *full* and *symmetric k-fold projective tensor* *product of* E, denoted by $\otimes_\pi^k E$ and $\otimes_{\pi_s}^{k,s} E$. As shown in Example 3.21, the projective tensor norm is not suitable for the error analysis of Monte Carlo methods.

2.3. Moments of Banach space valued random variables

The purpose of this subsection is to generalize the notion of the kth moment, defined for a real-valued random variable $X\colon \Omega \to \mathbb{R}$ as

$$\mathbb{M}^k[X] := \mathbb{E}\left[X^k\right] = \int_\Omega X(\omega)^k \, \mathrm{d}\mathbb{P}(\omega),$$

to Banach space valued random variables $\xi\colon \Omega \to E$. To this end, we first specify the concept of vector-valued integration which we imply when taking expected values of E-valued random variables.

2.3.1. Vector-valued integration

An E-valued random variable ξ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a mapping $\xi\colon \Omega \to E$ which is measurable in a certain sense. Specifically, we consider the class of Bochner measurable random variables; these are all mappings $\xi\colon \Omega \to E$ which are (i) measurable with respect to the Borel σ-algebra $\mathcal{B}(E)$ on E, and (ii) almost surely separably valued, i.e., $\xi \in E_0$ $\mathbb{P}$-a.s. for some separable subspace $E_0 \subseteq E$. A Bochner measurable random variable ξ is often also called *strongly measurable*, emphasizing the contrast to the notion of weak measurability, which only requires the real-valued random variable $\langle f, \xi \rangle$ to be measurable for every $f \in E'$. Note that these notions are equivalent whenever ξ is almost surely separably valued (e.g., in the case of a separable Banach space E), see [37, Theorem 2.3].

Furthermore, it turns out that $\xi\colon \Omega \to E$ is Bochner measurable if and only if there exists a sequence of Borel measurable simple functions $\xi_n\colon \Omega \to E$, $n \in \mathbb{N}$, such that $\xi_n \to \xi$, $\mathbb{P}$-a.s. This characterization facilitates the definition of the Bochner integral

$$\int_\Omega \xi(\omega) \, \mathrm{d}\mathbb{P}(\omega) =: \mathbb{E}[\xi] \in E,$$

whenever ξ is Bochner measurable and $\mathbb{E}\left[\|\xi\|_E\right] < \infty$.

We close this subsection with introducing the corresponding Bochner L_q-spaces. For a real Banach space $(F, \|\cdot\|_F)$ and $q \in [1, \infty)$, $L_q(\Omega; F) := L_q(\Omega, \mathcal{A}, \mathbb{P}; F)$ is the space of all (equivalence classes of) F-valued Bochner measurable random variables $\eta\colon \Omega \to F$ such that $\mathbb{E}\left[\|\eta\|_F^q\right] < \infty$, with norm given by

$$\|\eta\|_{L_q(\Omega;F)} := \left(\mathbb{E}\left[\|\eta\|_F^q\right]\right)^{1/q} = \left(\int_\Omega \|\eta(\omega)\|_F^q \, \mathrm{d}\mathbb{P}(\omega)\right)^{1/q}.$$

2.3.2. Moments of order k

For an E-valued random variable ξ, its *injective kth moment* $\mathbb{M}_\varepsilon^k[\xi]$ is defined as the expectation (see e.g. [37, Section 3.1])

$$\mathbb{M}_\varepsilon^k[\xi] := \mathbb{E}\left[\otimes^k \xi\right] = \int_\Omega \otimes^k \xi(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_\Omega \underbrace{\xi(\omega) \otimes \cdots \otimes \xi(\omega)}_{k \text{ times}}\,\mathrm{d}\mathbb{P}(\omega), \tag{2.8}$$

whenever this integral exists (in Bochner sense) in the k-fold injective tensor product space $\otimes_\varepsilon^k E$. In what follows, we will always assume that $\xi \in L_q(\Omega; E)$ holds for some $q \in [k, \infty)$. This guarantees that $\mathbb{M}_\varepsilon^k[\xi]$ exists in Bochner sense: Firstly, Bochner measurability of ξ implies that also $\otimes^k \xi \colon \Omega \to \otimes_\varepsilon^k E$ is Bochner measurable since the non-linear mapping $E \ni x \mapsto \otimes^k x \in \otimes_\varepsilon^k E$ is continuous and, secondly,

$$\mathbb{E}\left[\|\otimes^k \xi\|_\varepsilon\right] = \mathbb{E}\left[\|\xi\|_E^k\right] \leq \mathbb{E}\left[\|\xi\|_E^q\right] = \|\xi\|_{L_q(\Omega;E)}^q < \infty.$$

Of particular relevance in the present context is that the injective kth moment is an element of the k-fold *symmetric* injective tensor product space $\otimes_{\varepsilon_s}^{k,s} E$, since

$$s_\varepsilon\left(\mathbb{M}_\varepsilon^k[\xi]\right) = s_\varepsilon\left(\mathbb{E}\left[\otimes^k \xi\right]\right) = \mathbb{E}\left[s_\varepsilon\left(\otimes^k \xi\right)\right] = \mathbb{E}\left[\otimes^k \xi\right] = \mathbb{M}_\varepsilon^k[\xi].$$

Here, we have used continuity of the symmetrization $s_\varepsilon \colon \otimes_\varepsilon^k E \to \otimes_{\varepsilon_s}^{k,s} E$, see (2.2) and (2.4), to exchange the order of $s_\varepsilon(\,\cdot\,)$ and $\mathbb{E}[\,\cdot\,]$.

Remark 2.2. The assumption $\xi \in L_k(\Omega; E)$ does not only guarantee the existence of the injective kth moment $\mathbb{M}_\varepsilon^k[\xi]$ but also that of the projective kth moment $\mathbb{M}_\pi^k[\xi]$, i.e., the integral in (2.8) converges in Bochner sense also in the stronger projective tensor norm defined in (2.6), see Remark 2.1. This observation follows from the chain of identities $\|\otimes^k \xi\|_\pi = \|\xi\|_E^k = \|\otimes^k \xi\|_\varepsilon$ showing that the above arguments may be translated verbatim, see also [37, Theorem 3.8].

3. Monte Carlo estimation of the kth moment

This section treats the analysis of abstract standard, single-level and multilevel Monte Carlo methods to estimate the injective kth moment $\mathbb{M}_\varepsilon^k[\xi]$ of a Banach space valued random variable $\xi \in L_k(\Omega; E)$. In Subsection 3.1 we first provide necessary definitions, including those of Rademacher and orthogaussian families as well as the type of a Banach space. Furthermore, we formulate auxiliary results based on comparison theorems for Rademacher and Gaussian averages. These observations facilitate the error analysis for standard and single-level Monte Carlo estimation in Subsection 3.2 and for the multilevel Monte Carlo method in Subsection 3.3.

3.1. Auxiliary results on Rademacher and Gaussian averages

Definition 3.1 *(Rademacher family).* Let $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ be a complete probability space and $r_j \colon \widetilde{\Omega} \to \{-1, 1\}$, $j \in J \subseteq \mathbb{N}$, be a family of independent random variables such that $\widetilde{\mathbb{P}}(r_j = -1) = \widetilde{\mathbb{P}}(r_j = 1) = \frac{1}{2}$ for all $j \in J$. Then, the collection of random variables $(r_j)_{j \in J}$ is called a *Rademacher family.*

Definition 3.2 *(Orthogaussian family).* Let $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ be a complete probability space with expectation $\widetilde{\mathbb{E}}$. Suppose that $g_j \colon \widetilde{\Omega} \to \mathbb{R}$, $j \in J \subseteq \mathbb{N}$, are independent standard Gaussian random variables, i.e., $\widetilde{\mathbb{E}}[g_j] = 0$, $\widetilde{\mathbb{E}}\big[g_j^2\big] = 1$ for all $j \in J$, and $\widetilde{\mathbb{E}}[g_i g_j] = 0$ for $i \neq j$. Then, the collection $(g_j)_{j \in J}$ is called an *orthogaussian family.*

Assuming that $(z_j)_{j=1}^M$ is a Rademacher or orthogaussian family and $(x_j)_{j=1}^M$ are vectors in the Banach space E, the E-norm of the finite random sum $\sum_{j=1}^M z_j x_j$ is the supremum of a real-valued (Rademacher or Gaussian) stochastic process. More specifically, the Hahn–Banach theorem allows us to rewrite the norm as follows,

$$\Big\| \sum_{j=1}^M z_j x_j \Big\|_E = \sup_{f \in B_{E'}} \Big| \sum_{j=1}^M z_j f(x_j) \Big| = \sup_{(t_1, \ldots, t_M) \in T} \Big| \sum_{j=1}^M z_j t_j \Big|, \tag{3.1}$$

where T is the compact subset of $\mathbb{R}^M$ given by $T := \{(f(x_1), \ldots, f(x_M)) : f \in B_{E'}\}$.

The next lemma summarizes comparison theorems for Gaussian and Rademacher averages, see [47, Corollary 3.17 & Theorem 4.12]. It will facilitate generalizing the equality (3.1) to an upper bound for L_q-norms of random variables of the form

$$\sup_{f \in B_{E'}} \Big| \sum_{j=1}^M r_j f(x_j)^{k_j} \Big| \quad \text{and} \quad \sup_{f \in B_{E'}} \Big| \sum_{j=1}^M g_j f(x_j)^{k_j} \Big|, \qquad k_j \in \mathbb{N}, \quad 1 \leq j \leq M.$$

Lemma 3.3. *Let $M \in \mathbb{N}$, and let $z := (z_j)_{j=1}^M$ be a Rademacher or orthogaussian family on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$. Suppose that $(\varphi_j)_{j=1}^M$ are functions on $\mathbb{R}$ such that, for every $1 \leq j \leq M$,*

$$\varphi_j(0) = 0 \qquad \text{and} \qquad |\varphi_j(s) - \varphi_j(t)| \leq |s - t| \quad \forall s, t \in \mathbb{R}. \tag{3.2}$$

Assume further that $G \colon [0, \infty) \to [0, \infty)$ is convex and increasing. Then, we have for any bounded subset $T \subset \mathbb{R}^M$ the relation

$$\widetilde{\mathbb{E}} G\bigg(\frac{1}{2} \sup_{(t_1, \ldots, t_M) \in T} \Big| \sum_{j=1}^M z_j \varphi_j(t_j) \Big| \bigg) \leq \widetilde{\mathbb{E}} G\bigg(C_z \sup_{(t_1, \ldots, t_M) \in T} \Big| \sum_{j=1}^M z_j t_j \Big| \bigg), \tag{3.3}$$

where $\widetilde{\mathbb{E}}$ denotes the expectation operator on $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$, and C_z is given by

$$C_z := \begin{cases} 1 & \textit{if } z \textit{ is a Rademacher family}, \\ 2 & \textit{if } z \textit{ is an orthogaussian family}. \end{cases} \tag{3.4}$$

Proposition 3.4. *Let $M \in \mathbb{N}$, $x_1, \ldots, x_M \in E$, and $(z_j)_{j=1}^M$ be a Rademacher or orthogaussian family on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ (with expectation $\widetilde{\mathbb{E}}$). In addition, assume that $G\colon [0,\infty) \to [0,\infty)$ is a convex and increasing function.*

(i) *For integers $k_1, \ldots, k_M \in \mathbb{N}$, we have the relation*

$$\widetilde{\mathbb{E}} G\left(\sup_{f \in B_{E'}} \left| \sum_{j=1}^M z_j f(x_j)^{k_j} \right| \right) \leq \widetilde{\mathbb{E}} G\left(2C_z \left\| \sum_{j=1}^M z_j k_j \|x_j\|_E^{k_j - 1} x_j \right\|_E \right). \tag{3.5}$$

(ii) *For general exponents $q_1, \ldots, q_M \in [1,\infty)$, the following holds:*

$$\widetilde{\mathbb{E}} G\left(\sup_{f \in B_{E'}} \left| \sum_{j=1}^M z_j |f(x_j)|^{q_j} \right| \right) \leq \widetilde{\mathbb{E}} G\left(2C_z \left\| \sum_{j=1}^M z_j q_j \|x_j\|_E^{q_j - 1} x_j \right\|_E \right). \tag{3.6}$$

Here, the constant $C_z \in \{1,2\}$ is defined as in (3.4).

Remark 3.5. Part (i) of Proposition 3.4 is a generalization of the observation made by Ledoux and Talagrand in [47, Equation (4.19)]; there formulated for Rademacher averages, $G(x) := x$, and $k_1 = \ldots = k_M = 2$. That is, we generalize to higher-order polynomials on one hand, and to Gaussian averages on the other hand.

Proof of Proposition 3.4. We will prove (i) and (ii) using Lemma 3.3. Without loss of generality we may assume that $x_j \neq 0$ for all $1 \leq j \leq M$.

To derive (i), let $M \in \mathbb{N}$ and $k_j \in \mathbb{N}$, $\widetilde{x}_j \in E \setminus \{0\}$ for all $1 \leq j \leq M$. Furthermore, for $1 \leq j \leq M$, define $I_j := \big[-\|\widetilde{x}_j\|_E, \|\widetilde{x}_j\|_E\big]$ and

$$\widetilde{\varphi}_j \colon I_j \to \mathbb{R}, \qquad \widetilde{\varphi}_j(s) := k_j^{-1} \|\widetilde{x}_j\|_E^{1-k_j} s^{k_j}, \quad s \in I_j. \tag{3.7}$$

The function $\widetilde{\varphi}_j$ is continuously differentiable on the interior of I_j with $|\widetilde{\varphi}_j'(s)| \leq 1$. Let $P_j \colon \mathbb{R} \to I_j$ denote the projection $P_j(s) := \max\{-\|\widetilde{x}_j\|_E, \min\{s, \|\widetilde{x}_j\|_E\}\}$ onto the interval I_j. Then, for every $1 \leq j \leq M$, the function

$$\varphi_j \colon \mathbb{R} \to \mathbb{R}, \qquad \varphi_j(s) := \widetilde{\varphi}_j(P_j(s)),$$

satisfies the assumptions (3.2) of Lemma 3.3. Therefore, letting the bounded set $T \subset \mathbb{R}^M$ be given by $T := \{(f(\widetilde{x}_1), \ldots, f(\widetilde{x}_M)) : f \in B_{E'}\}$, we may combine the observation (3.1) with (3.3) and by also noting that $T \subseteq I_1 \times \ldots \times I_M$ we obtain that, for every convex and increasing function $G\colon [0,\infty) \to [0,\infty)$,

$$\widetilde{\mathbb{E}}G\left(\frac{1}{2}\sup_{f\in B_{E'}}\left|\sum_{j=1}^{M} z_j k_j^{-1}\|\widetilde{x}_j\|_E^{1-k_j} f(\widetilde{x}_j)^{k_j}\right|\right) = \widetilde{\mathbb{E}}G\left(\frac{1}{2}\sup_{(t_1,\ldots,t_M)\in T}\left|\sum_{j=1}^{M} z_j\varphi_j(t_j)\right|\right)$$
$$\leq \widetilde{\mathbb{E}}G\left(C_z \sup_{(t_1,\ldots,t_M)\in T}\left|\sum_{j=1}^{M} z_j t_j\right|\right) = \widetilde{\mathbb{E}}G\left(C_z\left\|\sum_{j=1}^{M} z_j\widetilde{x}_j\right\|_E\right). \tag{3.8}$$

Finally, for $x_1,\ldots,x_M \in E\setminus\{0\}$, we choose $\widetilde{x}_j := 2k_j\|x_j\|_E^{k_j-1}x_j \in E\setminus\{0\}$ for all $1\leq j\leq M$ and (3.8) shows (3.5).

For (ii) we modify the above arguments by replacing k_j and the interval I_j in the definition (3.7) of $\widetilde{\varphi}_j$ with q_j and $I_j := \big[0,\|\widetilde{x}_j\|_E\big]$, respectively, and the projection $P_j\colon \mathbb{R}\to I_j$ with $P_j(s) := \min\{|s|,\|\widetilde{x}_j\|_E\}$. Then, the function $\varphi_j(s) := \widetilde{\varphi}_j(P_j(s))$ satisfies the assumptions (3.2) of Lemma 3.3, since $\varphi_j(0)=0$, $|\widetilde{\varphi}_j'(s)|\leq 1$ holds for every $s\in(0,\|\widetilde{x}_j\|_E)$, and by the mean value theorem combined with the reverse triangle inequality we thus have that

$$|\varphi_j(s)-\varphi_j(t)| = |\widetilde{\varphi}_j(P_j(s))-\widetilde{\varphi}_j(P_j(t))| \leq |P_j(s)-P_j(t)| \leq \big||s|-|t|\big| \leq |s-t|.$$

Furthermore, for every $f\in B_{E'}$, we have that $P_j(f(\widetilde{x}_j)) = |f(\widetilde{x}_j)|$ and we obtain the analogue of (3.8),

$$\widetilde{\mathbb{E}}G\left(\frac{1}{2}\sup_{f\in B_{E'}}\left|\sum_{j=1}^{M} z_j q_j^{-1}\|\widetilde{x}_j\|_E^{1-q_j}|f(\widetilde{x}_j)|^{q_j}\right|\right) \leq \widetilde{\mathbb{E}}G\left(C_z\left\|\sum_{j=1}^{M} z_j\widetilde{x}_j\right\|_E\right).$$

The choice $\widetilde{x}_j := 2q_j\|x_j\|_E^{q_j-1}x_j$, $1\leq j\leq M$, completes the proof of (3.6). $\square$

The next lemma shows that we may symmetrize independent random variables with vanishing mean, when bounding L_q-norms of their sum. This result can be found, e.g., in [47, Lemma 6.3] or [14, Lemma 5.9].

Lemma 3.6 *(Symmetrization). Let $q\in[1,\infty)$, $M\in\mathbb{N}$, $(r_j)_{j=1}^M$ be a Rademacher family on a complete probability space $(\widetilde{\Omega},\widetilde{\mathcal{A}},\widetilde{\mathbb{P}})$, and let $\eta_1,\ldots,\eta_M\in L_q(\widetilde{\Omega};F)$ be centered random variables, i.e., $\widetilde{\mathbb{E}}[\eta_j]=0$ for $1\leq j\leq M$, taking values in a real Banach space $(F,\|\cdot\|_F)$ such that $\eta_1,\ldots,\eta_M,r_1,\ldots,r_M$ are independent. Then,*

$$\left\|\sum_{j=1}^{M}\eta_j\right\|_{L_q(\widetilde{\Omega};F)} \leq 2\left\|\sum_{j=1}^{M} r_j\eta_j\right\|_{L_q(\widetilde{\Omega};F)}.$$

Definition 3.7 *(Kahane–Khintchine constants).* Assume that $p,q\in[1,\infty)$. The (q,p) Kahane–Khintchine constant $K_{q,p}$ is the smallest constant $K\in(0,\infty)$ such that for any Rademacher family $(r_j)_{j\in\mathbb{N}}$ on a complete probability space $(\widetilde{\Omega},\widetilde{\mathcal{A}},\widetilde{\mathbb{P}})$, for any real Banach space $(F,\|\cdot\|_F)$, for all $n\in\mathbb{N}$, and every $x_1,\ldots,x_n\in F$, one has that

$$\left\| \sum_{j=1}^{n} r_j x_j \right\|_{L_q(\widetilde{\Omega};F)} \le K \left\| \sum_{j=1}^{n} r_j x_j \right\|_{L_p(\widetilde{\Omega};F)}. \tag{3.9}$$

Remark 3.8. For the case $q \le p$, Hölder's inequality shows that $K_{q,p} = 1$. Finiteness of the constant $K_{q,p}$ in the non-trivial case $q > p$ was derived by Kahane [39]; it implies that, for Rademacher sums, all L_q-norms with $q \in [1, \infty)$ are equivalent.

Remark 3.9. By invoking an argument based on the central limit theorem (see [47, p. 103]) the Kahane–Khintchine inequality for Rademacher sums (3.9) implies a corresponding result for Gaussian averages: For all $p, q \in [1, \infty)$, for any orthogaussian family $(g_j)_{j\in\mathbb{N}}$ on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$, for any real Banach space $(F, \|\cdot\|_F)$, for all $n \in \mathbb{N}$, and every $x_1, \dots, x_n \in F$, we have that

$$\left\| \sum_{j=1}^{n} g_j x_j \right\|_{L_q(\widetilde{\Omega};F)} \le K_{q,p} \left\| \sum_{j=1}^{n} g_j x_j \right\|_{L_p(\widetilde{\Omega};F)}. \tag{3.10}$$

Definition 3.10 *(Type p constant).* A real Banach space $(F, \|\cdot\|_F)$ is said to be of *(Rademacher) type* $p \in [1, 2]$ if there exists a constant $\tau \in (0, \infty)$ such that for any Rademacher family $(r_j)_{j\in\mathbb{N}}$ on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ (with expectation operator $\widetilde{\mathbb{E}}$), for every $n \in \mathbb{N}$, and all vectors $x_1, \dots, x_n \in F$,

$$\left\| \sum_{j=1}^{n} r_j x_j \right\|_{L_p(\widetilde{\Omega};F)} = \left(\widetilde{\mathbb{E}} \left[\left\| \sum_{j=1}^{n} r_j x_j \right\|_F^p \right] \right)^{1/p} \le \tau \left(\sum_{j=1}^{n} \|x_j\|_F^p \right)^{1/p}. \tag{3.11}$$

In this case, the smallest constant $\tau \in (0, \infty)$ in (3.11) is called the *type p constant of F* and will be denoted by $\tau_p(F)$.

Remark 3.11. The definition of the type of a Banach space $(F, \|\cdot\|_F)$ is often complemented with the notion of its cotype: F has *cotype* $q \in [2, \infty]$ if there exists a constant $c \in (0, \infty)$ such that for any Rademacher family $(r_j)_{j\in\mathbb{N}}$ on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$, for every $n \in \mathbb{N}$, and all vectors $x_1, \dots, x_n \in F$,

$$\left(\sum_{j=1}^{n} \|x_j\|_F^q \right)^{1/q} \le c \left\| \sum_{j=1}^{n} r_j x_j \right\|_{L_q(\widetilde{\Omega};F)} \qquad \text{if } q \in [2, \infty),$$

$$\sup_{1\le j\le n} \|x_j\|_F \le c \left\| \sum_{j=1}^{n} r_j x_j \right\|_{L_1(\widetilde{\Omega};F)} \qquad \text{if } q = \infty.$$

Remark 3.12. Every Banach space has type 1 and cotype ∞ by the triangle inequality and Lévy's inequality (see e.g. [47, Proposition 2.3]), respectively. Conversely, by the (classical) Khintchine inequalities (see e.g. [47, Lemma 4.1]) there exist constants

$A_q, B_q \in (0,\infty)$ depending only on $q \in [1,\infty)$ such that, for any Rademacher family $(r_j)_{j\in\mathbb{N}}$ on $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ and any finite sequence $(a_j)_{j=1}^n$ of real numbers,

$$A_q\left(\sum_{j=1}^n a_j^2\right)^{1/2} \leq \left(\widetilde{\mathbb{E}}\left[\left|\sum_{j=1}^n r_j a_j\right|^q\right]\right)^{1/q} = \left\|\sum_{j=1}^n r_j a_j\right\|_{L_q(\widetilde{\Omega};\mathbb{R})} \leq B_q\left(\sum_{j=1}^n a_j^2\right)^{1/2} \tag{3.12}$$

which implies that the type cannot be larger than 2 and the cotype cannot be smaller than 2. Kwapień [45] showed that a Banach space has type 2 and cotype 2 if and only if it is isomorphic to a Hilbert space.

Example 3.13. Let $(H, (\,\cdot\,,\,\cdot\,)_H)$ be a real separable Hilbert space. In this case, the *Hilbert tensor product space* $\otimes_{w_2}^2 H$ (see Appendix A for the definition) is again a Hilbert space and, consequently, has type $p = 2$. However, none of the tensor product spaces $\otimes_\pi^2 H$, $\otimes_{\pi_s}^{2,s} H$, $\otimes_\varepsilon^2 H$ or $\otimes_{\varepsilon_s}^{2,s} H$ has a non-trivial type $p > 1$.

This can be seen by the following counterexample: Let $(e_j)_{j\in\mathbb{N}}$ be an orthonormal basis for H. Then, for all $p \in [1,\infty)$ and every $n \in \mathbb{N}$, we have

$$\left(\sum_{j=1}^n \|e_j \otimes e_j\|_\pi^p\right)^{1/p} = \left(\sum_{j=1}^n \|e_j \otimes e_j\|_\varepsilon^p\right)^{1/p} = \left(\sum_{j=1}^n \|e_j\|_H^{2p}\right)^{1/p} = n^{1/p}.$$

Moreover, the calculations in Appendix A (see the identities (A.2) and (A.3) of Lemma A.1) imply that for any Rademacher family $(r_j)_{j\in\mathbb{N}}$ on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$, for all $p \in [1,\infty)$ and for every $n \in \mathbb{N}$,

$$\left\|\sum_{j=1}^n r_j\, e_j \otimes e_j\right\|_{L_p(\widetilde{\Omega};\otimes_\pi^2 H)} = n \qquad \text{and} \qquad \left\|\sum_{j=1}^n r_j\, e_j \otimes e_j\right\|_{L_p(\widetilde{\Omega};\otimes_\varepsilon^2 H)} = 1,$$

and the same statements hold when replacing $\otimes_\pi^2 H$ by the symmetric projective tensor product space $\otimes_{\pi_s}^{2,s} H$ and $\otimes_\varepsilon^2 H$ by the symmetric injective tensor product space $\otimes_{\varepsilon_s}^{2,s} H$, respectively. This shows that (i) neither $\otimes_\pi^2 H$ nor $\otimes_{\pi_s}^{2,s} H$ have a non-trivial type $p > 1$, and (ii) neither $\otimes_\varepsilon^2 H$ nor $\otimes_{\varepsilon_s}^{2,s} H$ have a non-trivial cotype $q < \infty$. Thus, $\otimes_\varepsilon^2 H$ and $\otimes_{\varepsilon_s}^{2,s} H$ do not have a non-trivial type either, cf. [36, Theorem 7.1.14].

In the next subsections we will see that the type $p \in [1,2]$ of a Banach space E determines the rate of convergence when approximating statistical moments of E-valued random variables by means of Monte Carlo methods and, moreover, that this convergence rate does not depend on the order k of the moment. However, as the above example illustrates, to derive this finding, it is not possible to argue by transferring the type of a Banach space to its k-fold tensor product.

3.2. Standard and single-level Monte Carlo estimation

The next proposition is the key result for proving convergence of Monte Carlo methods for means, i.e., statistical moments of order $k = 1$. It can be found, e.g., in [47, Proposition 9.11] for the case $q = p$ and in this generality in [14, Proposition 5.10].

Proposition 3.14. *Assume that $(E, \|\cdot\|_E)$ is of Rademacher type $p \in [1, 2]$. Let $q \in [p, \infty)$, $M \in \mathbb{N}$ and $\eta_1, \ldots, \eta_M \in L_q(\Omega; E)$ be independent E-valued random variables with vanishing mean, $\mathbb{E}[\eta_j] = 0$ for all $1 \le j \le M$. Then,*

$$\Big\| \sum_{j=1}^{M} \eta_j \Big\|_{L_q(\Omega;E)} \le 2K_{q,p}\tau_p(E) \left(\sum_{j=1}^{M} \|\eta_j\|_{L_q(\Omega;E)}^p \right)^{1/p}.$$

Corollary 3.15. *Assume that $(E, \|\cdot\|_E)$ is of Rademacher type $p \in [1, 2]$, and let $\eta \in L_1(\Omega; E)$. In addition, let $q \in [p, \infty)$, $M \in \mathbb{N}$ and $\xi_1, \ldots, \xi_M \in L_q(\Omega; E)$ be independent and identically distributed E-valued random variables. Then,*

$$\Big\| \mathbb{E}[\eta] - \frac{1}{M} \sum_{j=1}^{M} \xi_j \Big\|_{L_q(\Omega;E)} \le \big\| \mathbb{E}[\eta - \xi_1] \big\|_E + 2K_{q,p}\tau_p(E)\, M^{-\left(1-\frac{1}{p}\right)} \big\| \xi_1 - \mathbb{E}[\xi_1] \big\|_{L_q(\Omega;E)}.$$

Proof. By applying the triangle inequality on $L_q(\Omega; E)$ and Proposition 3.14 (noting that $\xi_j - \mathbb{E}[\xi_1]$, $1 \le j \le M$, are independent and centered), we find that

$$\begin{aligned} \Big\| \mathbb{E}[\eta] - \frac{1}{M} \sum_{j=1}^{M} \xi_j \Big\|_{L_q(\Omega;E)} &\le \big\| \mathbb{E}[\eta - \xi_1] \big\|_E + \frac{1}{M} \Big\| \sum_{j=1}^{M} (\xi_j - \mathbb{E}[\xi_1]) \Big\|_{L_q(\Omega;E)} \\ &\le \big\| \mathbb{E}[\eta - \xi_1] \big\|_E + 2K_{q,p}\tau_p(E)\, M^{-1} \left(\sum_{j=1}^{M} \big\| \xi_j - \mathbb{E}[\xi_1] \big\|_{L_q(\Omega;E)}^p \right)^{1/p}, \end{aligned}$$

and the claim follows by the identical distribution of $\xi_1, \ldots, \xi_M$. □

The remainder of this subsection is devoted to generalizing the approximation result of Monte Carlo estimation for the first statistical moment in Corollary 3.15 to (injective) statistical moments $\mathbb{M}_\varepsilon^k[\eta]$ of an arbitrary order $k \in \mathbb{N}$. Example 3.13 shows that it is in general not possible to argue via the type of the tensor product space. We therefore prove the convergence rates of Monte Carlo methods directly by means of the auxiliary results derived in Subsection 3.1.

Theorem 3.16. *Assume that $(E, \|\cdot\|_E)$ is of Rademacher type $p \in [1, 2]$. Let $q \in [p, \infty)$, $k, M \in \mathbb{N}$ and $\xi_1, \ldots, \xi_M \in L_{kq}(\Omega; E)$ be independent and identically distributed E-valued random variables. Then,*

$$\left\| \mathbb{M}^k_\varepsilon[\xi_1] - \frac{1}{M} \sum_{j=1}^{M} \otimes^k \xi_j \right\|_{L_q(\Omega; \otimes^{k,s}_{\varepsilon_s} E)} \leq C^{\mathsf{SL}}_{q,p,k}\, M^{-\left(1-\frac{1}{p}\right)} \|\xi_1\|^k_{L_{kq}(\Omega;E)}. \tag{3.13}$$

Here, we recall the constant $B_q \in (0,\infty)$ from the classical Khintchine inequalities (3.12), as well as the Kahane–Khintchine constant $K_{q,p}$ and type p constant $\tau_p(E)$ from Definitions 3.7 and 3.10, respectively, and set

$$C^{\mathsf{SL}}_{q,p,k} := 2(2kK_{q,p}\tau_p(E) + B_q). \tag{3.14}$$

Proof. Assume that $(r_j)_{j=1}^M$ is a Rademacher family on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ and, for $j \in \{1,\dots,M\}$, let $\boldsymbol{\xi}_j \colon \Omega \times \widetilde{\Omega} \to E$ and $\boldsymbol{r}_j \colon \Omega \times \widetilde{\Omega} \to \{-1,1\}$ denote the mappings that satisfy $\boldsymbol{\xi}_j(\omega,\widetilde{\omega}) = \xi_j(\omega)$ and $\boldsymbol{r}_j(\omega,\widetilde{\omega}) = r_j(\widetilde{\omega})$ for all $(\omega,\widetilde{\omega}) \in \Omega \times \widetilde{\Omega}$. Notice that on $(\Omega \times \widetilde{\Omega}, \mathcal{A} \otimes \widetilde{\mathcal{A}}, \mathbb{P} \otimes \widetilde{\mathbb{P}})$ the random variables $(\boldsymbol{r}_j)_{j=1}^M$ form a Rademacher family and $\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_M, \boldsymbol{r}_1,\dots,\boldsymbol{r}_M$ are independent. Moreover, $\xi_j \in L_{kq}(\Omega;E)$ implies that $\mathbb{M}^k_\varepsilon[\xi_1] \in \otimes^{k,s}_{\varepsilon_s} E$ and $\otimes^k \boldsymbol{\xi}_j \in L_q(\Omega \times \widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)$ are well-defined. We further note that by the identical distribution of $\xi_1,\dots,\xi_M$,

$$(\mathbb{E} \otimes \widetilde{\mathbb{E}})\big[\otimes^k \boldsymbol{\xi}_j - \mathbb{M}^k_\varepsilon[\xi_1]\big] = \mathbb{E}\big[\otimes^k \xi_j - \mathbb{M}^k_\varepsilon[\xi_1]\big] = \mathbb{E}\big[\otimes^k \xi_j\big] - \mathbb{M}^k_\varepsilon[\xi_j] = 0.$$

This shows that the independent random variables $\otimes^k \boldsymbol{\xi}_j - \mathbb{M}^k_\varepsilon[\xi_1] \colon \Omega \times \widetilde{\Omega} \to \otimes^{k,s}_{\varepsilon_s} E$, $1 \leq j \leq M$, are centered. Therefore, we can use Lemma 3.6 on the probability space $(\Omega \times \widetilde{\Omega}, \mathcal{A} \otimes \widetilde{\mathcal{A}}, \mathbb{P} \otimes \widetilde{\mathbb{P}})$ and for the Banach space $\otimes^{k,s}_{\varepsilon_s} E$ to deduce that

$$\left\| \sum_{j=1}^{M} \big(\otimes^k \boldsymbol{\xi}_j - \mathbb{M}^k_\varepsilon[\xi_1]\big) \right\|_{L_q(\Omega\times\widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)} \leq 2 \left\| \sum_{j=1}^{M} \boldsymbol{r}_j \big(\otimes^k \boldsymbol{\xi}_j - \mathbb{M}^k_\varepsilon[\xi_1]\big) \right\|_{L_q(\Omega\times\widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)}.$$

By the triangle inequality on $L_q(\Omega \times \widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)$ we then obtain that

$$\begin{aligned}
\left\| \sum_{j=1}^{M} \big(\otimes^k \xi_j - \mathbb{M}^k_\varepsilon[\xi_1]\big) \right\|_{L_q(\Omega; \otimes^{k,s}_{\varepsilon_s} E)} &= \left\| \sum_{j=1}^{M} \big(\otimes^k \boldsymbol{\xi}_j - \mathbb{M}^k_\varepsilon[\xi_1]\big) \right\|_{L_q(\Omega\times\widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)} \\
&\leq 2 \left\| \sum_{j=1}^{M} \boldsymbol{r}_j \big(\otimes^k \boldsymbol{\xi}_j - \mathbb{M}^k_\varepsilon[\xi_1]\big) \right\|_{L_q(\Omega\times\widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)} \\
&\leq 2 \left\| \sum_{j=1}^{M} \boldsymbol{r}_j \otimes^k \boldsymbol{\xi}_j \right\|_{L_q(\Omega\times\widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)} + 2 \left\| \sum_{j=1}^{M} \boldsymbol{r}_j \mathbb{M}^k_\varepsilon[\xi_1] \right\|_{L_q(\Omega\times\widetilde{\Omega}; \otimes^{k,s}_{\varepsilon_s} E)} \\
&=: 2(\mathrm{A}) + 2(\mathrm{B}).
\end{aligned} \tag{3.15}$$

To bound term (A) from above, we apply Fubini's theorem and the Kahane–Khintchine inequality (3.9) for the Banach space $F := \otimes^{k,s}_{\varepsilon_s} E$ and obtain that

$$(\mathrm{A}) = \left(\int_\Omega \Big\| \sum_{j=1}^M r_j(\,\cdot\,) \otimes^k \xi_j(\omega) \Big\|^q_{L_q(\widetilde{\Omega};\otimes^{k,s}_{\varepsilon_s} E)} \mathrm{d}\mathbb{P}(\omega) \right)^{1/q}$$

$$\leq K_{q,p} \left(\int_\Omega \Big\| \sum_{j=1}^M r_j(\,\cdot\,) \otimes^k \xi_j(\omega) \Big\|^q_{L_p(\widetilde{\Omega};\otimes^{k,s}_{\varepsilon_s} E)} \mathrm{d}\mathbb{P}(\omega) \right)^{1/q}.$$

Upon inserting the definition (2.3) of the symmetric injective tensor norm, we use Proposition 3.4(i) for the convex increasing function $G(t) := t^p$, $t \geq 0$, and the fact that the Banach space E has type $p \in [1, 2]$ with type constant $\tau_p(E) \in (0, \infty)$, to conclude that

$$(\mathrm{A}) \leq K_{q,p} \left(\int_\Omega \left(\widetilde{\mathbb{E}} \left[\left(\sup_{f \in B_{E'}} \Big| \sum_{j=1}^M r_j(\,\cdot\,) f\big(\xi_j(\omega)\big)^k \Big| \right)^p \right] \right)^{q/p} \mathrm{d}\mathbb{P}(\omega) \right)^{1/q}$$

$$\leq 2k K_{q,p} \left(\int_\Omega \left(\widetilde{\mathbb{E}} \left[\Big\| \sum_{j=1}^M r_j(\,\cdot\,) \|\xi_j(\omega)\|_E^{k-1} \xi_j(\omega) \Big\|_E^p \right] \right)^{q/p} \mathrm{d}\mathbb{P}(\omega) \right)^{1/q}$$

$$\leq 2k K_{q,p} \tau_p(E) \left(\int_\Omega \left(\sum_{j=1}^M \|\xi_j(\omega)\|_E^{kp} \right)^{q/p} \mathrm{d}\mathbb{P}(\omega) \right)^{1/q}$$

$$= 2k K_{q,p} \tau_p(E) \Big\| \sum_{j=1}^M \|\xi_j\|_E^{kp} \Big\|^{1/p}_{L_{q/p}(\Omega;\mathbb{R})}.$$

Since $q \geq p$, we can use the triangle inequality on $L_{q/p}(\Omega; \mathbb{R})$, yielding

$$(\mathrm{A}) \leq 2k K_{q,p} \tau_p(E) \left(\sum_{j=1}^M \|\xi_j\|^{kp}_{L_{kq}(\Omega;E)} \right)^{1/p} = 2k K_{q,p} \tau_p(E) M^{1/p} \|\xi_1\|^k_{L_{kq}(\Omega;E)}, \tag{3.16}$$

where we also used the identical distribution of $\xi_1, \ldots, \xi_M$.

For term (B) we use the estimate

$$\big\| \mathbb{M}^k_\varepsilon[\xi_1] \big\|_{\varepsilon_s} = \big\| \mathbb{E}\big[\otimes^k \xi_1\big] \big\|_{\varepsilon_s} \leq \mathbb{E}\big[\|\otimes^k \xi_1\|_{\varepsilon_s} \big] = \mathbb{E}\big[\|\xi_1\|_E^k \big] \leq \|\xi_1\|^k_{L_{kq}(\Omega;E)},$$

as well as the classical Khintchine inequalities (3.12) so that

$$\Big\| \sum_{j=1}^M r_j \Big\|_{L_q(\widetilde{\Omega};\mathbb{R})} \leq B_q\, M^{1/2} \leq B_q\, M^{1/p},$$

and conclude that

$$
\begin{aligned}
(\mathrm{B}) = \Big\| \mathbb{M}^k_\varepsilon[\xi_1] \sum_{j=1}^M \boldsymbol{r}_j \Big\|_{L_q(\Omega\times\widetilde{\Omega};\otimes^{k,s}_{\varepsilon_s}E)} &= \left(\int_{\widetilde{\Omega}} \Big\| \mathbb{M}^k_\varepsilon[\xi_1] \sum_{j=1}^M r_j(\widetilde{\omega}) \Big\|^q_{\varepsilon_s} \mathrm{d}\widetilde{\mathbb{P}}(\widetilde{\omega}) \right)^{1/q} \\
&= \big\| \mathbb{M}^k_\varepsilon[\xi_1] \big\|_{\varepsilon_s} \left(\int_{\widetilde{\Omega}} \Big| \sum_{j=1}^M r_j(\widetilde{\omega}) \Big|^q \mathrm{d}\widetilde{\mathbb{P}}(\widetilde{\omega}) \right)^{1/q} \leq B_q\, M^{1/p} \|\xi_1\|^k_{L_{kq}(\Omega;E)}.
\end{aligned} \tag{3.17}
$$

Finally, combining (3.15), (3.16) and (3.17) shows that

$$
\begin{aligned}
\Big\| \mathbb{M}^k_\varepsilon[\xi_1] - \frac{1}{M} \sum_{j=1}^M \otimes^k \xi_j \Big\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}E)} &= M^{-1} \Big\| \sum_{j=1}^M \big(\otimes^k \xi_j - \mathbb{M}^k_\varepsilon[\xi_1] \big) \Big\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}E)} \\
&\leq 2(2kK_{q,p}\tau_p(E) + B_q)\, M^{-\left(1-\frac{1}{p}\right)} \|\xi_1\|^k_{L_{kq}(\Omega;E)},
\end{aligned}
$$

which along with the definition (3.14) of $C^{\mathsf{SL}}_{q,p,k}$ completes the proof of (3.13). □

The estimate (3.16) of term (A) in the proof of Theorem 3.16 reveals the following analogue of Proposition 3.14 for independent (not necessarily identically distributed) random variables $\eta_1, \dots, \eta_M \in L_{kq}(\Omega; E)$ with vanishing kth moment.

Corollary 3.17. *Assume that $(E, \|\cdot\|_E)$ is of Rademacher type $p \in [1, 2]$. Let $q \in [p, \infty)$, $k, M \in \mathbb{N}$ and $\eta_1, \dots, \eta_M \in L_{kq}(\Omega; E)$ be independent E-valued random variables with vanishing kth moment, $\mathbb{M}^k_\varepsilon[\eta_j] = 0$ for all $1 \leq j \leq M$. Then,*

$$
\Big\| \sum_{j=1}^M \otimes^k \eta_j \Big\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}E)} \leq 4kK_{q,p}\tau_p(E) \left(\sum_{j=1}^M \|\eta_j\|^{kp}_{L_{kq}(\Omega;E)} \right)^{1/p}. \tag{3.18}
$$

Remark 3.18. Optimality of the convergence rate $1 - \frac{1}{p}$ in (3.13) is ultimately related to the question whether it is *necessary* that the Banach space E has Rademacher type $p \in [1, 2]$ for (3.18) to hold for all finite sequences $\eta_1, \dots, \eta_M \in L_{kq}(\Omega; E)$ of independent E-valued random variables with vanishing kth moment.

For the first moment, $k = 1$, it is evident that the choice $\eta_j := r_j x_j$ in (3.18), where $(r_j)_{j\in\mathbb{N}}$ is a Rademacher family on $(\Omega, \mathcal{A}, \mathbb{P})$ and $x_1, x_2, \ldots \in E$, implies that the Banach space E has Rademacher type p. However, for higher-order moments, this question is more complex due to the injective tensor norm on the left-hand side. For odd orders $k \in \mathbb{N}$ and the space $E := \ell_1$ of summable real-valued sequences (which only has Rademacher type $p = 1$), the choice $\eta_j := r_j e_j$ shows that (3.18) cannot hold for any $p > 1$. Here, $(e_j)_{j\in\mathbb{N}}$ denote the standard unit vectors in ℓ_1. In addition, the classical Khintchine inequalities (3.12) imply that, for any Banach space E, the convergence rate in (3.13) cannot be better that $1/2$ (one may take, e.g., $\xi_j := g_j\, x$, where $(g_j)_{j\in\mathbb{N}}$ is an orthogaussian family on $(\Omega, \mathcal{A}, \mathbb{P})$ and $x \neq 0$ is a non-zero vector in E). Sharpness of the rate $1 - \frac{1}{p}$ in (3.13) and necessity of the Rademacher type p for (3.18) for the case $p \in (1, 2)$ remains an open question.

The next lemma complements Theorem 3.16 when deriving convergence rates of single-level Monte Carlo methods for approximating injective kth moments.

Lemma 3.19. *Let $k \in \mathbb{N}$ and suppose that $\eta, \xi \in L_k(\Omega; E)$. Then,*

$$\left\|\mathbb{M}_\varepsilon^k[\eta] - \mathbb{M}_\varepsilon^k[\xi]\right\|_{\varepsilon_s} \le \left\|\mathbb{M}_\varepsilon^k[\eta] - \mathbb{M}_\varepsilon^k[\xi]\right\|_{\varepsilon} \le \|\eta - \xi\|_{L_k(\Omega;E)} \sum_{i=0}^{k-1} \Big[\|\eta\|_{L_k(\Omega;E)}^i \|\xi\|_{L_k(\Omega;E)}^{k-i-1} \Big].$$

Proof. The first inequality of the assertion is trivial. We next note that also the remaining relation is evident in the case $k = 1$, since

$$\left\|\mathbb{M}_\varepsilon^k[\eta] - \mathbb{M}_\varepsilon^k[\xi]\right\|_{\varepsilon} = \left\|\mathbb{E}[\eta - \xi]\right\|_E \le \mathbb{E}\left[\|\eta - \xi\|_E\right] = \|\eta - \xi\|_{L_k(\Omega;E)} \quad \text{if} \quad k = 1.$$

We now assume that $k \ge 2$ and observe that

$$\begin{aligned} \otimes^k \eta - \otimes^k \xi &= \sum_{i=0}^{k-1} \Big[\big(\otimes^{i+1}\eta\big) \otimes \big(\otimes^{k-(i+1)}\xi\big) - \big(\otimes^{i}\eta\big) \otimes \big(\otimes^{k-i}\xi\big) \Big] \\ &= \sum_{i=0}^{k-1} \Big[\big(\otimes^{i}\eta\big) \otimes (\eta - \xi) \otimes \big(\otimes^{k-(i+1)}\xi\big) \Big]. \end{aligned}$$

Therefore, we may estimate as follows,

$$\begin{aligned} &\left\|\mathbb{M}_\varepsilon^k[\eta] - \mathbb{M}_\varepsilon^k[\xi]\right\|_{\varepsilon} = \left\|\mathbb{E}\left[\otimes^k\eta - \otimes^k\xi\right]\right\|_{\varepsilon} \le \mathbb{E}\Big[\left\|\otimes^k\eta - \otimes^k\xi\right\|_{\varepsilon}\Big] \\ &\quad \le \mathbb{E}\left[\sum_{i=0}^{k-1} \left\| \big(\otimes^{i}\eta\big) \otimes (\eta - \xi) \otimes \big(\otimes^{k-i-1}\xi\big) \right\|_{\varepsilon} \right] = \sum_{i=0}^{k-1} \mathbb{E}\Big[\|\eta\|_E^i \|\eta - \xi\|_E \|\xi\|_E^{k-i-1} \Big]. \end{aligned}$$

Combined with the Hölder inequality this completes the proof, since

$$\begin{aligned} \mathbb{E}\Big[\|\eta - \xi\|_E \|\xi\|_E^{k-1}\Big] &\le \big(\mathbb{E}\big[\|\eta - \xi\|_E^k\big]\big)^{\frac{1}{k}} \big(\mathbb{E}\big[\|\xi\|_E^k\big]\big)^{\frac{k-1}{k}}, \\ \mathbb{E}\Big[\|\eta\|_E^{k-1} \|\eta - \xi\|_E\Big] &\le \big(\mathbb{E}\big[\|\eta - \xi\|_E^k\big]\big)^{\frac{1}{k}} \big(\mathbb{E}\big[\|\eta\|_E^k\big]\big)^{\frac{k-1}{k}}, \end{aligned}$$

and, whenever $k \ge 3$, we obtain for all $i \in \{1, \dots, k-2\}$

$$\mathbb{E}\Big[\|\eta\|_E^i \|\eta - \xi\|_E \|\xi\|_E^{k-i-1}\Big] \le \big(\mathbb{E}\big[\|\eta\|_E^k\big]\big)^{\frac{i}{k}} \big(\mathbb{E}\big[\|\eta - \xi\|_E^k\big]\big)^{\frac{1}{k}} \big(\mathbb{E}\big[\|\xi\|_E^k\big]\big)^{\frac{k-i-1}{k}}$$

by a triple Hölder inequality with $\big(\frac{k}{i}\big)^{-1} + k^{-1} + \big(\frac{k}{k-i-1}\big)^{-1} = 1$. □

We are now ready to state the main result of this subsection: an abstract convergence rate bound in $L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)$ for single-level Monte Carlo estimation of the injective kth moment $\mathbb{M}_\varepsilon^k[\eta]$, assuming at our disposal M independent samples of an approximation $\xi_1 \in L_{kq}(\Omega; E)$ to $\eta \in L_k(\Omega; E)$.

Corollary 3.20. *Assume that $(E, \|\cdot\|_E)$ is of Rademacher type $p \in [1,2]$. Let $q \in [p, \infty)$, $k, M \in \mathbb{N}$ and $\xi_1, \ldots, \xi_M \in L_{kq}(\Omega; E)$ be independent and identically distributed E-valued random variables. Then, for every $U \in \otimes_{\varepsilon_s}^{k,s} E$, we have*

$$\left\| U - \frac{1}{M} \sum_{j=1}^{M} \otimes^k \xi_j \right\|_{L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)} \le \left\| U - \mathbb{M}_\varepsilon^k[\xi_1] \right\|_{\varepsilon_s} + C_{q,p,k}^{\mathsf{SL}} M^{-\left(1-\frac{1}{p}\right)} \|\xi_1\|_{L_{kq}(\Omega;E)}^k,$$

where the constant $C_{q,p,k}^{\mathsf{SL}}$ is defined as in (3.14).

In particular, for all $\eta \in L_k(\Omega; E)$, we have

$$\left\| \mathbb{M}_\varepsilon^k[\eta] - \frac{1}{M} \sum_{j=1}^{M} \otimes^k \xi_j \right\|_{L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)} \le \|\eta - \xi_1\|_{L_k(\Omega;E)} \sum_{i=0}^{k-1} \left[\|\eta\|_{L_k(\Omega;E)}^i \|\xi_1\|_{L_k(\Omega;E)}^{k-i-1} \right] + C_{q,p,k}^{\mathsf{SL}} M^{-\left(1-\frac{1}{p}\right)} \|\xi_1\|_{L_{kq}(\Omega;E)}^k.$$

Proof. By the triangle inequality on $L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)$ we obtain, for every $U \in \otimes_{\varepsilon_s}^{k,s} E$,

$$\left\| U - \frac{1}{M} \sum_{j=1}^{M} \otimes^k \xi_j \right\|_{L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)} \le \left\| U - \mathbb{M}_\varepsilon^k[\xi_1] \right\|_{\varepsilon_s} + \left\| \mathbb{M}_\varepsilon^k[\xi_1] - \frac{1}{M} \sum_{j=1}^{M} \otimes^k \xi_j \right\|_{L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)},$$

and the first claim follows by applying the estimate (3.13) of Theorem 3.16. Subsequently, we derive the second assertion by combining this result with Lemma 3.19 which we use to bound the difference of the kth moments $\left\| \mathbb{M}_\varepsilon^k[\eta] - \mathbb{M}_\varepsilon^k[\xi_1] \right\|_{\varepsilon_s}$. □

We close this subsection with a counterexample which shows that the convergence results for the standard and single-level Monte Carlo estimators in Theorem 3.16 and Corollary 3.20 can, in general, not hold when measuring the error in the (symmetric or full) projective tensor norm. More specifically, we discuss this for second-order moments of Hilbert space valued random variables, i.e., the random variables take values in a Banach space of type $p = 2$.

Example 3.21. Let $(H, (\cdot\,,\cdot)_H)$ be a real separable Hilbert space, and let $(e_j)_{j\in\mathbb{N}}$ be an orthonormal basis for H. For $n \in \mathbb{N}$, consider the H-valued random variable $\xi_n \colon \Omega \to H$ on $(\Omega, \mathcal{A}, \mathbb{P})$, whose (discrete uniform) distribution is defined by

$$\forall i \in \{1, \ldots, n\}: \quad \mathbb{P}(\{\omega \in \Omega : \xi_n(\omega) = e_i\}) = n^{-1}.$$

Then, for all $n \in \mathbb{N}$, both the projective and injective second moments of ξ_n exist,

$$\mathbb{M}^2_\pi[\xi_n] = \mathbb{M}^2_\varepsilon[\xi_n] = \mathbb{E}\big[\xi_n \otimes \xi_n\big] = \frac{1}{n}\sum_{i=1}^n e_i \otimes e_i, \quad \big\|\mathbb{M}^2_\pi[\xi_n]\big\|_\pi = \big\|\mathbb{M}^2_\pi[\xi_n]\big\|_{\pi_s} = 1,$$

see (A.2) in Lemma A.1 of Appendix A for the norm identities.

In addition, for all $q \in [1,\infty)$ and every $n \in \mathbb{N}$, we have that

$$\|\xi_n\|^q_{L_q(\Omega;H)} = \mathbb{E}\big[\|\xi_n\|^q_H\big] = \frac{1}{n}\sum_{i=1}^n \|e_i\|^q_H = 1.$$

We let $q \in [1,\infty)$, $\xi_{n,1},\ldots,\xi_{n,M}$ be $M \in \mathbb{N}$ independent copies of ξ_n and estimate

$$\begin{aligned}
\operatorname{err}^{(n)}_{q,\pi} &:= \Bigg\|\mathbb{M}^2_\pi[\xi_n] - \frac{1}{M}\sum_{j=1}^M \otimes^2 \xi_{n,j}\Bigg\|^q_{L_q(\Omega;\otimes^2_\pi H)} = \mathbb{E}\left[\Bigg\|\mathbb{M}^2_\pi[\xi_n] - \frac{1}{M}\sum_{j=1}^M \otimes^2 \xi_{n,j}\Bigg\|^q_\pi\right] \\
&= \sum_{\nu_1=1}^n \cdots \sum_{\nu_M=1}^n \frac{1}{n^M}\Bigg\|\frac{1}{n}\sum_{i=1}^n (e_i \otimes e_i) - \frac{1}{M}\sum_{j=1}^M \big(e_{\nu_j} \otimes e_{\nu_j}\big)\Bigg\|^q_\pi \\
&\geq \sum_{\substack{1\leq \nu_1,\ldots,\nu_M \leq n \\ \text{pairwise distinct}}} \frac{1}{n^M}\Bigg\|\frac{1}{n}\sum_{i=1}^n (e_i \otimes e_i) - \frac{1}{M}\sum_{j=1}^M \big(e_{\nu_j} \otimes e_{\nu_j}\big)\Bigg\|^q_\pi.
\end{aligned}$$

Thus, assuming that $n \geq M$, again by (A.2) in Lemma A.1 we obtain that

$$\begin{aligned}
\operatorname{err}^{(n)}_{q,\pi} &\geq \sum_{\substack{1\leq \nu_1,\ldots,\nu_M \leq n \\ \text{pairwise distinct}}} n^{-M}\big[M\big(\tfrac{1}{M} - \tfrac{1}{n}\big) + (n-M)\tfrac{1}{n}\big]^q \\
&= 2^q\big(1 - \tfrac{M}{n}\big)^q n^{-M}[n\cdots(n-M+1)] \geq 2^q\big(1-\tfrac{M}{n}\big)^{q+M}.
\end{aligned}$$

Given $q \in [1,\infty)$ and $M \in \mathbb{N}$, we choose an integer $n_\star = n_\star(q,M) \in \mathbb{N}$ such that

$$n_\star \geq M\big(1 - 2^{-q/(q+M)}\big)^{-1} \qquad \Longrightarrow \qquad \big(1 - \tfrac{M}{n_\star}\big)^{q+M} \geq 2^{-q}.$$

This proves that, for all $q \in [1,\infty)$ and every $M \in \mathbb{N}$, there exists $n_\star = n_\star(q,M) \in \mathbb{N}$ such that

$$\Bigg\|\mathbb{M}^2_\pi[\xi_{n_\star}] - \frac{1}{M}\sum_{j=1}^M \otimes^2 \xi_{n_\star,j}\Bigg\|_{L_q(\Omega;\otimes^{2,s}_{\pi_s} H)} \geq \Bigg\|\mathbb{M}^2_\pi[\xi_{n_\star}] - \frac{1}{M}\sum_{j=1}^M \otimes^2 \xi_{n_\star,j}\Bigg\|_{L_q(\Omega;\otimes^2_\pi H)} \geq 1.$$

Since $\|\xi_{n_\star}\|_{L_q(\Omega;H)} = 1$ is also true for all $q \in [1,\infty)$ and since H has type $p = 2$, this shows that an analogue of (3.13) cannot hold with respect to the (full or symmetric) projective tensor norm.

3.3. Multilevel Monte Carlo estimation

Assuming that $(X_\ell)_{\ell=1}^L$ is a family of E-valued random variables corresponding to $L \in \mathbb{N}$ different refinement *levels* of underlying discretization parameters, translating the idea of multilevel Monte Carlo (MLMC) estimation, as formulated e.g. in [12, p. 5] for means of Hilbert space valued random variables, to higher-order moments of Banach space valued random variables results in exploiting the following telescopic sum (here: $X_0 := 0 \in E$)

$$\mathbb{E}\left[\otimes^k X_L\right] = \mathbb{M}_\varepsilon^k[X_L] = \sum_{\ell=1}^{L}\left(\mathbb{M}_\varepsilon^k[X_\ell] - \mathbb{M}_\varepsilon^k[X_{\ell-1}]\right) = \sum_{\ell=1}^{L}\mathbb{E}\left[\otimes^k X_\ell - \otimes^k X_{\ell-1}\right],$$

and estimating $\mathbb{E}[\otimes^k X_\ell - \otimes^k X_{\ell-1}]$ for each $1 \le \ell \le L$ via Monte Carlo sampling instead of $\mathbb{E}[\otimes^k X_L]$. As we will see in Theorem 3.25 and Remark 3.27, this approach considerably reduces the computational cost.

Evidently, the corresponding error analysis requires a Monte Carlo convergence result for estimating differences of injective kth moments, i.e., for expected values of the form $\mathbb{E}[\otimes^k\eta - \otimes^k\xi] = \mathbb{M}_\varepsilon^k[\eta] - \mathbb{M}_\varepsilon^k[\xi]$, via standard Monte Carlo methods. This auxiliary result is derived in Proposition 3.23 by means of the next lemma, Lemma 3.22, which acts as the analogue of the Rademacher type estimate (3.11) for Rademacher sums of differences $\otimes^k x_j - \otimes^k y_j$, $1 \le j \le M$.

Lemma 3.22. *Let $(r_j)_{j=1}^M$ be a Rademacher family on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ with expectation $\widetilde{\mathbb{E}}$. Assume that $(E, \|\cdot\|_E)$ has Rademacher type $p \in [1, 2]$, and let $q \in [p, \infty)$, $k, M \in \mathbb{N}$ and $x_1, \ldots, x_M, y_1, \ldots, y_M \in E$. Then,*

$$\left\| \sum_{j=1}^{M} r_j \left(\otimes^k x_j - \otimes^k y_j\right) \right\|_{L_q(\widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} \le C_{q,p,k}^{\mathsf{diff}} \sum_{i=1}^{k} \binom{k}{i} \left[\sum_{j=1}^{M} \|x_j - y_j\|_E^{ip} \|y_j\|_E^{(k-i)p} \right]^{1/p}. \tag{3.19}$$

Here, $\binom{k}{i} := \frac{k!}{i!(k-i)!}$ is the binomial coefficient and the constant $C_{q,p,k}^{\mathsf{diff}}$ is given by

$$C_{q,p,k}^{\mathsf{diff}} := 16k\sqrt{\pi} K_{q,p} \tau_p(E) K_{q,2}, \tag{3.20}$$

with the Kahane–Khintchine and type p constants from Definitions 3.7 and 3.10.

Proof. For the proof of (3.19), we assume that $(g_j)_{j=1}^M$, $(\widetilde{g}_j)_{j=1}^M$ are two independent orthogaussian families on $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$. We first note that by [47, Lemma 4.5 and (4.8)], applied for the convex function $t \mapsto t^q$ and the Banach space $\otimes_{\varepsilon_s}^{k,s} E$, and by the definition (2.3) of the symmetric injective tensor norm we have that

$$
\begin{aligned}
\Big\| \sum_{j=1}^{M} r_j \big(\otimes^k x_j - \otimes^k y_j \big) \Big\|_{L_q(\widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} &\le \sqrt{\frac{\pi}{2}} \Big\| \sum_{j=1}^{M} g_j \big(\otimes^k x_j - \otimes^k y_j \big) \Big\|_{L_q(\widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} \\
&= \sqrt{\frac{\pi}{2}} \Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} g_j \big(f(x_j)^k - f(y_j)^k \big) \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})} \\
&= \sqrt{\frac{\pi}{2}} \Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} g_j \sum_{i=1}^{k} \binom{k}{i} f(x_j - y_j)^i f(y_j)^{k-i} \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})} \\
&\le \sum_{i=1}^{k} \sqrt{\frac{\pi}{2}} \binom{k}{i} \Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} g_j f(x_j - y_j)^i f(y_j)^{k-i} \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})} .
\end{aligned}
\tag{3.21}
$$

Here, we also used the binomial expansion for $f(x_j)^k = [f(x_j - y_j) + f(y_j)]^k$ and the triangle inequality on $L_q(\widetilde{\Omega}; \mathbb{R})$. We now claim that, for every $i \in \{1, \dots, k-1\}$,

$$
\begin{aligned}
&\Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} g_j f(x_j - y_j)^i f(y_j)^{k-i} \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})} \\
&\quad \le 2\sqrt{2} \Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} \big(g_j f(x_j - y_j)^i \|y_j\|_E^{k-i} + \widetilde{g}_j \|x_j - y_j\|_E^i f(y_j)^{k-i} \big) \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})} .
\end{aligned}
\tag{3.22}
$$

To establish (3.22), we set $\delta_j := x_j - y_j \in E$ for all $1 \le j \le M$, and consider for a fixed $i \in \{1, \dots, k-1\}$ the following two real-valued centered Gaussian processes $\mathcal{G}_{i,1}, \mathcal{G}_{i,2} \colon B_{E'} \times \widetilde{\Omega} \to \mathbb{R}$, which are indexed by $f \in B_{E'}$,

$$
\mathcal{G}_{i,1}(f) := \sum_{j=1}^{M} g_j f(\delta_j)^i f(y_j)^{k-i}, \tag{3.23}
$$

$$
\mathcal{G}_{i,2}(f) := \sqrt{2} \sum_{j=1}^{M} \big(g_j f(\delta_j)^i \|y_j\|_E^{k-i} + \widetilde{g}_j \|\delta_j\|_E^i f(y_j)^{k-i} \big). \tag{3.24}
$$

For all $i \in \{1, \dots, k-1\}$ and every $f, h \in B_{E'}$, we then obtain by independence of the standard Gaussian random variables $g_1, \dots, g_M, \widetilde{g}_1, \dots, \widetilde{g}_M$ the following estimate,

$$
\begin{aligned}
\widetilde{\mathbb{E}} \big[|\mathcal{G}_{i,1}(f) - \mathcal{G}_{i,1}(h)|^2 \big] &= \sum_{j=1}^{M} \big(f(\delta_j)^i f(y_j)^{k-i} - h(\delta_j)^i h(y_j)^{k-i} \big)^2 \\
&= \sum_{j=1}^{M} \big([f(\delta_j)^i - h(\delta_j)^i] f(y_j)^{k-i} + h(\delta_j)^i [f(y_j)^{k-i} - h(y_j)^{k-i}] \big)^2 \\
&\le 2 \sum_{j=1}^{M} \Big([f(\delta_j)^i - h(\delta_j)^i]^2 f(y_j)^{2(k-i)} + h(\delta_j)^{2i} [f(y_j)^{k-i} - h(y_j)^{k-i}]^2 \Big)
\end{aligned}
$$

$$
\begin{aligned}
&\leq 2 \sum_{j=1}^{M} \Big(\big[f(\delta_j)^i - h(\delta_j)^i \big]^2 \|y_j\|_E^{2(k-i)} + \|\delta_j\|_E^{2i} \big[f(y_j)^{k-i} - h(y_j)^{k-i} \big]^2 \Big) \\
&= \widetilde{\mathbb{E}} \big[|\mathcal{G}_{i,2}(f) - \mathcal{G}_{i,2}(h)|^2 \big].
\end{aligned}
$$

Furthermore, for every $i \in \{1, \dots, k-1\}$, we have

$$
\mathcal{G}_{i,1}(f) = \sum_{j=1}^{M} g_j \Psi_i\big(f(\delta_j), f(y_j) \big), \quad f \in B_{E'}.
$$

Here, the function $(t_1, t_2) \mapsto \Psi_i(t_1, t_2) := t_1^i \, t_2^{k-i}$ is continuous on $\mathbb{R}^2$ and satisfies $\Psi_i(0,0) = 0$ for every $i \in \{1, \dots, k-1\}$.

We thus may apply the comparison result derived in Lemma B.2 (see Appendix B) for every $i \in \{1, \dots, k-1\}$, which shows that, for all $q \in [1, \infty)$,

$$
\widetilde{\mathbb{E}} \Big[\big(\sup\nolimits_{f \in B_{E'}} |\mathcal{G}_{i,1}(f)| \big)^q \Big] \leq 2^q \, \widetilde{\mathbb{E}} \Big[\big(\sup\nolimits_{f \in B_{E'}} |\mathcal{G}_{i,2}(f)| \big)^q \Big].
$$

Taking the qth root on both sides of this inequality and inserting the definitions of $\mathcal{G}_{i,1}$ and $\mathcal{G}_{i,2}$ from (3.23)–(3.24) completes the proof of (3.22).

Next, combining (3.21) with (3.22) and the triangle inequality on $L_q(\widetilde{\Omega}; \mathbb{R})$ yields

$$
\begin{aligned}
\Big\| \sum_{j=1}^{M} r_j \big(\otimes^k x_j - \otimes^k y_j \big) \Big\|_{L_q(\widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} &\leq \sum_{i=1}^{k-1} \widetilde{C}_{k,i} \Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} g_j f(\delta_j)^i \|y_j\|_E^{k-i} \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})} \\
+ \sqrt{\frac{\pi}{2}} \Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} g_j f(\delta_j)^k \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})} &+ \sum_{i=1}^{k-1} \widetilde{C}_{k,i} \Big\| \sup_{f \in B_{E'}} \Big| \sum_{j=1}^{M} \widetilde{g}_j \|\delta_j\|_E^i f(y_j)^{k-i} \Big| \Big\|_{L_q(\widetilde{\Omega}; \mathbb{R})},
\end{aligned}
$$

where we set $\widetilde{C}_{k,i} := 2\sqrt{\pi} \binom{k}{i}$. By noting that $\sqrt{\pi}/\sqrt{2} \leq \widetilde{C}_{k,k}$ and estimating the $L_q(\widetilde{\Omega}; \mathbb{R})$-norms on the right-hand side for all $i \in \{1, \dots, k\}$ using Proposition 3.4(i), with $k_j = i$ and the vectors $\|y_j\|_E^{\frac{k-i}{i}} \delta_j$ (respectively, for every $i \in \{1, \dots, k-1\}$ with $k_j = k - i$ and $\|\delta_j\|_E^{\frac{i}{k-i}} y_j$) for all $1 \leq j \leq M$, we find that

$$
\begin{aligned}
\Big\| \sum_{j=1}^{M} r_j \big(\otimes^k x_j - \otimes^k y_j \big) \Big\|_{L_q(\widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} &\leq \sum_{i=1}^{k} \widetilde{C}_{k,i} \, 4i \Big\| \sum_{j=1}^{M} g_j \|\delta_j\|_E^{i-1} \|y_j\|_E^{k-i} \delta_j \Big\|_{L_q(\widetilde{\Omega}; E)} \\
&+ \sum_{i=1}^{k-1} \widetilde{C}_{k,i} \, 4(k-i) \Big\| \sum_{j=1}^{M} \widetilde{g}_j \|\delta_j\|_E^{i} \|y_j\|_E^{k-i-1} y_j \Big\|_{L_q(\widetilde{\Omega}; E)}.
\end{aligned}
$$

Finally, since $q \in [p, \infty)$ is assumed, we may use Proposition 3.14 for the independent, centered E-valued random variables

$$
\eta_j := g_j \|\delta_j\|_E^{i-1} \|y_j\|_E^{k-i} \delta_j \quad \text{resp.} \quad \widetilde{\eta}_j := \widetilde{g}_j \|\delta_j\|_E^{i} \|y_j\|_E^{k-i-1} y_j, \qquad 1 \leq j \leq M,
$$

to conclude that (recall the definitions $\widetilde{C}_{k,i} = 2\sqrt{\pi}\binom{k}{i}$ and $\delta_j = x_j - y_j$)

$$\left\| \sum_{j=1}^{M} r_j \left(\otimes^k x_j - \otimes^k y_j \right) \right\|_{L_q(\widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)}$$

$$\leq \sum_{i=1}^{k} \left[\widetilde{C}_{k,i} 4k\, 2K_{q,p}\tau_p(E) \left[\sum_{j=1}^{M} \|g_1\|^p_{L_q(\widetilde{\Omega};\mathbb{R})} \|x_j - y_j\|_E^{ip} \|y_j\|_E^{(k-i)p} \right]^{1/p} \right]$$

$$= 16k\sqrt{\pi} K_{q,p}\tau_p(E)\, \|g_1\|_{L_q(\widetilde{\Omega};\mathbb{R})} \sum_{i=1}^{k} \binom{k}{i} \left[\sum_{j=1}^{M} \|x_j - y_j\|_E^{ip} \|y_j\|_E^{(k-i)p} \right]^{1/p},$$

which completes the proof of (3.19), since $\|g_1\|_{L_q(\widetilde{\Omega};\mathbb{R})} \leq K_{q,2}\|g_1\|_{L_2(\widetilde{\Omega};\mathbb{R})} = K_{q,2}$ follows from (3.10), and $C^{\mathsf{diff}}_{q,p,k} = 16k\sqrt{\pi}K_{q,p}\tau_p(E)K_{q,2}$ by (3.20). □

Proposition 3.23. *Assume that $(E, \|\cdot\|_E)$ is of Rademacher type $p \in [1,2]$. Let $q \in [p,\infty)$, $k, M \in \mathbb{N}$ and $\eta_1, \dots, \eta_M, \xi_1, \dots, \xi_M \in L_{kq}(\Omega; E)$ be E-valued random variables such that the tuples $(\eta_1, \xi_1), \dots, (\eta_M, \xi_M)$ are independent and identically distributed. Then,*

$$\left\| \mathbb{M}^k_\varepsilon[\eta_1] - \mathbb{M}^k_\varepsilon[\xi_1] - \frac{1}{M}\sum_{j=1}^{M} \left(\otimes^k \eta_j - \otimes^k \xi_j \right) \right\|_{L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)}$$

$$\leq 2C^{\mathsf{diff}}_{q,p,k}\, M^{-\left(1-\frac{1}{p}\right)} \sum_{i=1}^{k} \left[\binom{k}{i} \|\eta_1 - \xi_1\|^i_{L_{kq}(\Omega;E)} \|\xi_1\|^{k-i}_{L_{kq}(\Omega;E)} \right]$$

$$+ 2B_q\, M^{-1/2} \|\eta_1 - \xi_1\|_{L_k(\Omega;E)} \sum_{i=0}^{k-1} \left[\|\eta_1\|^i_{L_k(\Omega;E)} \|\xi_1\|^{k-i-1}_{L_k(\Omega;E)} \right],$$

where $B_q, C^{\mathsf{diff}}_{q,p,k} \in (0,\infty)$ are defined as in (3.12) and (3.20), respectively.

Proof. We proceed similarly as in the proof of Theorem 3.16. We pick a Rademacher family $(r_j)_{j=1}^M$ on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$, and define the following random variables on the product probability space $(\Omega \times \widetilde{\Omega}, \mathcal{A} \otimes \widetilde{\mathcal{A}}, \mathbb{P} \otimes \widetilde{\mathbb{P}})$: For every $j \in \{1, \dots, M\}$, we set

$$\boldsymbol{\eta}_j(\omega, \widetilde{\omega}) := \eta_j(\omega), \qquad \boldsymbol{\xi}_j(\omega, \widetilde{\omega}) := \xi_j(\omega), \qquad \boldsymbol{r}_j(\omega, \widetilde{\omega}) := r_j(\widetilde{\omega}) \qquad \forall (\omega, \widetilde{\omega}) \in \Omega \times \widetilde{\Omega},$$

where we note that $(\boldsymbol{r}_j)_{j=1}^M$ is a Rademacher family on $(\Omega \times \widetilde{\Omega}, \mathcal{A} \otimes \widetilde{\mathcal{A}}, \mathbb{P} \otimes \widetilde{\mathbb{P}})$, and that $(\boldsymbol{\eta}_1, \boldsymbol{\xi}_1), \dots, (\boldsymbol{\eta}_M, \boldsymbol{\xi}_M), \boldsymbol{r}_1, \dots, \boldsymbol{r}_M$ are independent. Furthermore, the random variables $\otimes^k \boldsymbol{\eta}_j - \otimes^k \boldsymbol{\xi}_j - \mathbb{M}^k_\varepsilon[\eta_1] + \mathbb{M}^k_\varepsilon[\xi_1]$ are centered for all $1 \leq j \leq M$ so that by Lemma 3.6 and by the triangle inequality on $L_q(\Omega \times \widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)$ we find that

$$
\begin{aligned}
&\Big\| \sum_{j=1}^{M} \big(\otimes^k \eta_j - \otimes^k \xi_j - \mathbb{M}_\varepsilon^k[\eta_1] + \mathbb{M}_\varepsilon^k[\xi_1] \big) \Big\|_{L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E)} \\
&\qquad = \Big\| \sum_{j=1}^{M} \big(\otimes^k \boldsymbol{\eta}_j - \otimes^k \boldsymbol{\xi}_j - \mathbb{M}_\varepsilon^k[\eta_1] + \mathbb{M}_\varepsilon^k[\xi_1] \big) \Big\|_{L_q(\Omega \times \widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} \\
&\qquad \le 2 \Big\| \sum_{j=1}^{M} \boldsymbol{r}_j \big(\otimes^k \boldsymbol{\eta}_j - \otimes^k \boldsymbol{\xi}_j \big) \Big\|_{L_q(\Omega \times \widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} \\
&\qquad\quad + 2 \Big\| \sum_{j=1}^{M} \boldsymbol{r}_j \big(\mathbb{M}_\varepsilon^k[\eta_1] - \mathbb{M}_\varepsilon^k[\xi_1] \big) \Big\|_{L_q(\Omega \times \widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} =: 2(\mathrm{A}) + 2(\mathrm{B}).
\end{aligned}
\tag{3.25}
$$

For term (A) we use Fubini's theorem as well as Lemma 3.22 to find that

$$
\begin{aligned}
(\mathrm{A}) &= \Big\| \sum_{j=1}^{M} \boldsymbol{r}_j \big(\otimes^k \boldsymbol{\eta}_j - \otimes^k \boldsymbol{\xi}_j \big) \Big\|_{L_q(\Omega \times \widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)} \\
&= \Bigg(\int_\Omega \Big\| \sum_{j=1}^{M} r_j(\,\cdot\,) \big(\otimes^k \eta_j(\omega) - \otimes^k \xi_j(\omega) \big) \Big\|_{L_q(\widetilde{\Omega}; \otimes_{\varepsilon_s}^{k,s} E)}^q \,\mathrm{d}\mathbb{P}(\omega) \Bigg)^{1/q} \\
&\le C_{q,p,k}^{\mathsf{diff}} \Bigg(\int_\Omega \Bigg| \sum_{i=1}^{k} \binom{k}{i} \Bigg[\sum_{j=1}^{M} \| \eta_j(\omega) - \xi_j(\omega) \|_E^{ip} \| \xi_j(\omega) \|_E^{(k-i)p} \Bigg]^{1/p} \Bigg|^q \,\mathrm{d}\mathbb{P}(\omega) \Bigg)^{1/q}
\end{aligned}
$$

and, hence,

$$
(\mathrm{A}) \le C_{q,p,k}^{\mathsf{diff}} \Bigg\| \sum_{i=1}^{k} \binom{k}{i} \Bigg[\sum_{j=1}^{M} \| \eta_j - \xi_j \|_E^{ip} \| \xi_j \|_E^{(k-i)p} \Bigg]^{1/p} \Bigg\|_{L_q(\Omega; \mathbb{R})}.
$$

Next, we use the triangle inequality on $L_q(\Omega;\mathbb{R})$ as well as the fact that $q \in [p,\infty)$ so that also on $L_{q/p}(\Omega;\mathbb{R})$ we may apply the triangle inequality and conclude

$$
\begin{aligned}
(\mathrm{A}) &\le C_{q,p,k}^{\mathsf{diff}} \sum_{i=1}^{k} \binom{k}{i} \Big\| \sum_{j=1}^{M} \| \eta_j - \xi_j \|_E^{ip} \| \xi_j \|_E^{(k-i)p} \Big\|_{L_{q/p}(\Omega;\mathbb{R})}^{1/p} \\
&\le C_{q,p,k}^{\mathsf{diff}} \sum_{i=1}^{k} \binom{k}{i} \Bigg(\sum_{j=1}^{M} \Big\| \| \eta_j - \xi_j \|_E^{ip} \| \xi_j \|_E^{(k-i)p} \Big\|_{L_{q/p}(\Omega;\mathbb{R})} \Bigg)^{1/p} \\
&= C_{q,p,k}^{\mathsf{diff}} \, M^{1/p} \sum_{i=1}^{k} \binom{k}{i} \Big\| \| \eta_1 - \xi_1 \|_E^{ip} \| \xi_1 \|_E^{(k-i)p} \Big\|_{L_{q/p}(\Omega;\mathbb{R})}^{1/p},
\end{aligned}
$$

where the last step follows from the identical distribution of $(\eta_1, \xi_1), \ldots, (\eta_M, \xi_M)$. In addition, we observe that, for every $i \in \{1, \ldots, k-1\}$, by Hölder's inequality

$$\Big\| \|\eta_1 - \xi_1\|_E^{ip} \|\xi_1\|_E^{(k-i)p} \Big\|_{L_{q/p}(\Omega;\mathbb{R})}^{1/p} = \Big(\mathbb{E}\Big[\|\eta_1 - \xi_1\|_E^{iq} \|\xi_1\|_E^{(k-i)q} \Big] \Big)^{1/q}$$
$$\leq \Big(\big(\mathbb{E}\big[\|\eta_1 - \xi_1\|_E^{kq} \big] \big)^{\frac{i}{k}} \big(\mathbb{E}\big[\|\xi_1\|_E^{kq} \big] \big)^{\frac{k-i}{k}} \Big)^{1/q} = \|\eta_1 - \xi_1\|_{L_{kq}(\Omega;E)}^i \|\xi_1\|_{L_{kq}(\Omega;E)}^{k-i}$$

which completes the bound for term (A),

$$\text{(A)} \leq C_{q,p,k}^{\mathsf{diff}} \, M^{1/p} \sum_{i=1}^{k} \Big[\tbinom{k}{i} \|\eta_1 - \xi_1\|_{L_{kq}(\Omega;E)}^i \|\xi_1\|_{L_{kq}(\Omega;E)}^{k-i} \Big]. \tag{3.26}$$

For term (B) we obtain by the Khintchine inequalities (3.12) and by Lemma 3.19 the following estimate,

$$\begin{aligned} \text{(B)} &= \Big\| \sum_{j=1}^{M} r_j \Big\|_{L_q(\widetilde{\Omega};\mathbb{R})} \big\| \mathbb{M}_\varepsilon^k[\eta_1] - \mathbb{M}_\varepsilon^k[\xi_1] \big\|_{\varepsilon_s} \leq B_q \, M^{1/2} \big\| \mathbb{M}_\varepsilon^k[\eta_1] - \mathbb{M}_\varepsilon^k[\xi_1] \big\|_{\varepsilon_s} \\ &\leq B_q \, M^{1/2} \|\eta_1 - \xi_1\|_{L_k(\Omega;E)} \sum_{i=0}^{k-1} \Big[\|\eta_1\|_{L_k(\Omega;E)}^i \|\xi_1\|_{L_k(\Omega;E)}^{k-i-1} \Big]. \end{aligned} \tag{3.27}$$

The claim now follows by combining (3.25) with the estimates (3.26), (3.27) for the terms (A) and (B), upon dividing the resulting inequality by M. □

We are now ready to formulate our convergence result for abstract multilevel Monte Carlo methods to estimate higher-order statistical moments of Banach space valued random variables.

Theorem 3.24. *Let $(E, \|\cdot\|_E)$ be of Rademacher type $p \in [1,2]$, $q \in [p,\infty)$ and $k, L \in \mathbb{N}$. Suppose further that, for every $\ell \in \{1,\ldots,L\}$, $X_\ell \in L_{kq}(\Omega;E)$, $M_\ell \in \mathbb{N}$, and $\xi_{\ell,1},\ldots,\xi_{\ell,M_\ell}$ are independent copies of the $\otimes_{\varepsilon_s}^{k,s}E$-valued random variable*

$$\xi_\ell := \otimes^k X_\ell - \otimes^k X_{\ell-1} \in L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} E), \qquad X_0 := 0 \in E.$$

Then, for every $U \in \otimes_{\varepsilon_s}^{k,s} E$,

$$\begin{aligned} \Big\| U - \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{j=1}^{M_\ell} \xi_{\ell,j} \Big\|_{L_q(\Omega;\otimes_{\varepsilon_s}^{k,s} E)} &\leq \big\| U - \mathbb{M}_\varepsilon^k[X_L] \big\|_{\varepsilon_s} \\ &\quad + C_{q,p,k}^{\mathsf{ML}} \sum_{\ell=1}^{L} \Big[M_\ell^{-\left(1-\frac{1}{p}\right)} \|X_\ell - X_{\ell-1}\|_{L_{kq}(\Omega;E)} \\ &\qquad \cdot \sum_{i=0}^{k-1} \Big[\Big(\tbinom{k}{i+1} \|X_\ell - X_{\ell-1}\|_{L_{kq}(\Omega;E)}^i + \|X_\ell\|_{L_k(\Omega;E)}^i \Big) \|X_{\ell-1}\|_{L_{kq}(\Omega;E)}^{k-i-1} \Big] \Big], \end{aligned}$$

where $C^{\mathsf{ML}}_{q,p,k} := 2\max\{C^{\mathsf{diff}}_{q,p,k}, B_q\}$ *and the constants* $B_q, C^{\mathsf{diff}}_{q,p,k} \in (0,\infty)$ *are defined as in* (3.12) *and* (3.20).

Proof. First note that, for every $\ell \in \{1,\dots,L\}$ the random variables $\xi_{\ell,1},\dots,\xi_{\ell,M_\ell}$ are identically distributed and we have that

$$\mathbb{E}\left[\sum_{\ell=1}^{L}\frac{1}{M_\ell}\sum_{j=1}^{M_\ell}\xi_{\ell,j}\right] = \sum_{\ell=1}^{L}\mathbb{E}[\xi_\ell] = \sum_{\ell=1}^{L}\bigl(\mathbb{M}^k_\varepsilon[X_\ell] - \mathbb{M}^k_\varepsilon[X_{\ell-1}]\bigr) = \mathbb{M}^k_\varepsilon[X_L].$$

Thus, we find by the triangle inequality on $L_q(\Omega; \otimes^{k,s}_{\varepsilon_s} E)$ that, for every $U \in \otimes^{k,s}_{\varepsilon_s} E$,

$$\left\| U - \sum_{\ell=1}^{L}\frac{1}{M_\ell}\sum_{j=1}^{M_\ell}\xi_{\ell,j}\right\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}E)} \le \bigl\|U - \mathbb{M}^k_\varepsilon[X_L]\bigr\|_{\varepsilon_s} + \sum_{\ell=1}^{L}\operatorname{err}^{\mathsf{SL}}_{q,\varepsilon_s}(\xi_\ell),$$

where, for $\ell \in \{1,\dots,L\}$, we define

$$\operatorname{err}^{\mathsf{SL}}_{q,\varepsilon_s}(\xi_\ell) := \left\|\mathbb{E}[\xi_\ell] - \frac{1}{M_\ell}\sum_{j=1}^{M_\ell}\xi_{\ell,j}\right\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}E)}.$$

For every $\ell \in \{1,\dots,L\}$, we let the tuples $(X_{\ell-1,1}, X_{\ell,1}),\dots,(X_{\ell-1,M_\ell}, X_{\ell,M_\ell})$ be M_ℓ independent copies of $(X_{\ell-1}, X_\ell)$ and observe that

$$\operatorname{err}^{\mathsf{SL}}_{q,\varepsilon_s}(\xi_\ell) = \left\|\mathbb{M}^k_\varepsilon[X_\ell] - \mathbb{M}^k_\varepsilon[X_{\ell-1}] - \frac{1}{M_\ell}\sum_{j=1}^{M_\ell}\bigl(\otimes^k X_{\ell,j} - \otimes^k X_{\ell-1,j}\bigr)\right\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}E)}.$$

We are thus in the position to apply Proposition 3.23 on every level $\ell \in \{1,\dots,L\}$,

$$\begin{aligned}\operatorname{err}^{\mathsf{SL}}_{q,\varepsilon_s}(\xi_\ell) &\le 2C^{\mathsf{diff}}_{q,p,k}\, M_\ell^{-\left(1-\frac{1}{p}\right)}\sum_{i=1}^{k}\Bigl[\tbinom{k}{i}\|X_\ell - X_{\ell-1}\|^i_{L_{kq}(\Omega;E)}\|X_{\ell-1}\|^{k-i}_{L_{kq}(\Omega;E)}\Bigr] \\ &\quad + 2B_q\, M_\ell^{-1/2}\|X_\ell - X_{\ell-1}\|_{L_k(\Omega;E)}\sum_{i=0}^{k-1}\Bigl[\|X_\ell\|^i_{L_k(\Omega;E)}\|X_{\ell-1}\|^{k-i-1}_{L_k(\Omega;E)}\Bigr],\end{aligned}$$

which after recalling that $p \in [1,2]$ and $q \ge p \ge 1$ as well as combining the two sums completes the proof of the assertion. □

The error estimate of Theorem 3.24 facilitates optimizing the number of levels L as well as the number of samples on each level, $M_1,\dots,M_L$, to reduce the computational cost for achieving a target accuracy $\epsilon > 0$ of the MLMC estimator in the $L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}E)$-norm. This optimization is subject of the following "$\alpha\beta\gamma$ theorem".

Theorem 3.25. *Assume that $(E, \|\cdot\|_E)$ is of Rademacher type $p \in (1,2]$. Let $q \in [p,\infty)$, $k \in \mathbb{N}$, $X \in L_k(\Omega; E)$, $(X_\ell)_{\ell\in\mathbb{N}} \subset L_{kq}(\Omega; E)$ be a sequence of E-valued random variables and, for every $\ell \in \mathbb{N}$, define*

$$\xi_\ell := \otimes^k X_\ell - \otimes^k X_{\ell-1} \in L_q(\Omega; \otimes^{k,s}_{\varepsilon_s} E), \qquad X_0 := 0 \in E. \tag{3.28}$$

For $\ell \in \mathbb{N}$, let $\mathcal{C}_\ell$ denote the cost (number of floating point operations) to generate one sample of the random variable ξ_ℓ in (3.28), and suppose that there exist a sequence $(N_\ell)_{\ell\in\mathbb{N}}$ of positive integers and constants $\alpha, \beta, \gamma, C_\alpha, C_\beta, C_\gamma, C_{\mathsf{stab}} \in (0,\infty)$, $A \in (1,\infty)$ such that $N_\ell \eqsim A^\ell$ for all $\ell \in \mathbb{N}$ and, moreover,

$$\forall \ell \in \mathbb{N}: \qquad \left\| \mathbb{M}^k_\varepsilon[X] - \mathbb{M}^k_\varepsilon[X_\ell] \right\|_{\varepsilon_s} \leq C_\alpha N_\ell^{-\alpha}, \tag{α}$$

$$\forall \ell \in \mathbb{N}: \qquad \|X_\ell - X_{\ell-1}\|_{L_{kq}(\Omega;E)} \leq C_\beta N_\ell^{-\beta}, \tag{β}$$

$$\forall \ell \in \mathbb{N}: \qquad \mathcal{C}_\ell \leq C_\gamma N_\ell^{\gamma}, \tag{γ}$$

$$\forall \ell \in \mathbb{N}: \qquad \max\left\{ \|X\|_{L_k(\Omega;E)}, \|X_\ell\|_{L_{kq}(\Omega;E)} \right\} \leq C_{\mathsf{stab}}. \tag{stab}$$

For each $\ell \in \mathbb{N}$, let $(\xi_{\ell,j})_{j\in\mathbb{N}} \subset L_q(\Omega; \otimes^{k,s}_{\varepsilon_s} E)$ be a sequence of independent copies of the $\otimes^{k,s}_{\varepsilon_s} E$-valued random variable ξ_ℓ in (3.28).

Then, for every $\epsilon \in (0, 1/2]$, there exist integers $L \in \mathbb{N}$ and $M_1, \ldots, M_L \in \mathbb{N}$ such that the L_q-accuracy ϵ of the multilevel Monte Carlo estimator for $\mathbb{M}^k_\varepsilon[X]$,

$$\operatorname{err}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X) := \left\| \mathbb{M}^k_\varepsilon[X] - \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{j=1}^{M_\ell} \xi_{\ell,j} \right\|_{L_q(\Omega; \otimes^{k,s}_{\varepsilon_s} E)} < \epsilon, \tag{ϵ}$$

can be achieved at computational costs of the order

$$\mathcal{C}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X) \lesssim_{(\alpha,\beta,\gamma,A,p,q)} \begin{cases} \epsilon^{-\frac{\gamma}{\alpha}} + \epsilon^{-p'} & \text{if} \quad \beta p' > \gamma, \\ \epsilon^{-\frac{\gamma}{\alpha}} + \epsilon^{-p'} |\log_A \epsilon|^{p'+1} & \text{if} \quad \beta p' = \gamma, \\ \epsilon^{-\frac{\gamma}{\alpha}} + \epsilon^{-p' - \frac{\gamma - \beta p'}{\alpha}} & \text{if} \quad \beta p' < \gamma, \end{cases} \tag{$\mathcal{C}$}$$

where $p' \in [2,\infty)$ is such that $\frac{1}{p} + \frac{1}{p'} = 1$. The constant implied in $\lesssim$ may also depend on the constants $C_\alpha, C_\beta, C_\gamma$ and C_{stab} from the assumptions above.

Proof. We will show by explicit construction that, for every $\epsilon \in (0, 1/2]$, assumptions (α), (β), (γ) and (stab) allow to choose the algorithmic steering parameters $L \in \mathbb{N}$ and $M_1, \ldots, M_L \in \mathbb{N}$ so that (ϵ) holds with cost ($\mathcal{C}$).

By Theorem 3.24 and by assumptions (α), (β), (stab) we obtain the following error estimate,

$$
\begin{aligned}
\operatorname{err}_{q,\varepsilon_s}^{k,\mathsf{ML}}(X) &\le \big\| \mathbb{M}_\varepsilon^k[X] - \mathbb{M}_\varepsilon^k[X_L] \big\|_{\varepsilon_s} + C_{q,p,k}^{\mathsf{ML}} \sum_{\ell=1}^{L} \Big[M_\ell^{-\left(1-\frac{1}{p}\right)} \|X_\ell - X_{\ell-1}\|_{L_{kq}(\Omega;E)} \cdots \\
&\qquad \cdot \sum_{i=0}^{k-1} \Big[\Big(\tbinom{k}{i+1} \|X_\ell - X_{\ell-1}\|_{L_{kq}(\Omega;E)}^{i} + \|X_\ell\|_{L_k(\Omega;E)}^{i} \Big) \|X_{\ell-1}\|_{L_{kq}(\Omega;E)}^{k-i-1} \Big] \Big] \\
&\le C_\alpha N_L^{-\alpha} + C_{q,p,k}^{\mathsf{ML}} C_\beta \sum_{\ell=1}^{L} \Big[M_\ell^{-\left(1-\frac{1}{p}\right)} N_\ell^{-\beta} \sum_{i=0}^{k-1} \Big[\Big(\tbinom{k}{i+1} C_\beta^i + C_{\mathsf{stab}}^i \Big) C_{\mathsf{stab}}^{k-i-1} \Big] \Big] \\
&\le C_\alpha N_L^{-\alpha} + C_\star \sum_{\ell=1}^{L} \Big[M_\ell^{-1/p'} N_\ell^{-\beta} \Big],
\end{aligned}
$$

where $C_{q,p,k}^{\mathsf{ML}} \in (0,\infty)$ is as in Theorem 3.24, and $C_\star = C_\star(k,p,q,C_\beta,C_{\mathsf{stab}}) \in (0,\infty)$ is defined by

$$
C_\star := C_{q,p,k}^{\mathsf{ML}} C_\beta \big(k\, C_{\mathsf{stab}}^{k-1} + 2^k \max\{ C_\beta^{k-1}, C_{\mathsf{stab}}^{k-1} \} \big),
$$

since

$$
\begin{aligned}
\sum_{i=0}^{k-1} \Big[\Big(\tbinom{k}{i+1} C_\beta^i + C_{\mathsf{stab}}^i \Big) C_{\mathsf{stab}}^{k-i-1} \Big] &\le k\, C_{\mathsf{stab}}^{k-1} + \max\{ C_\beta^{k-1}, C_{\mathsf{stab}}^{k-1} \} \sum_{i=0}^{k-1} \tbinom{k}{i+1} \\
&\le k\, C_{\mathsf{stab}}^{k-1} + 2^k \max\{ C_\beta^{k-1}, C_{\mathsf{stab}}^{k-1} \}.
\end{aligned}
$$

Choose $L \in \mathbb{N}$ as the smallest integer such that $N_L^{-\alpha} < \min\{C_\alpha^{-1}, 1\} \frac{\epsilon}{2}$ holds and, for every $\ell \in \{1, \ldots, L\}$, let $M_\ell \in \mathbb{N}$ be defined as the smallest integer satisfying

$$
M_\ell \ge C_\star^{p'} N_L^{\alpha p'} S_L^{p'} N_\ell^{-\frac{(\beta+\gamma)p'}{p'+1}}, \qquad \text{where} \qquad S_L := \sum_{\ell=1}^{L} N_\ell^{\frac{\gamma-\beta p'}{p'+1}}.
$$

Note that the magnitude of S_L behaves asymptotically (for L large) as

$$
S_L = \sum_{\ell=1}^{L} N_\ell^{\frac{\gamma-\beta p'}{p'+1}} \eqsim_{(\beta,\gamma,A,p)} \begin{cases} 1 & \text{if} \quad \beta p' > \gamma, \\ L & \text{if} \quad \beta p' = \gamma, \\ N_L^{\frac{\gamma-\beta p'}{p'+1}} & \text{if} \quad \beta p' < \gamma. \end{cases} \tag{3.29}
$$

For this choice of L and $M_1, \ldots, M_L$, we can bound the error as follows,

$$
\begin{aligned}
\operatorname{err}_{q,\varepsilon_s}^{k,\mathsf{ML}}(X) &< C_\alpha C_\alpha^{-1} \frac{\epsilon}{2} + C_\star \sum_{\ell=1}^{L} \Big[C_\star^{-1} N_L^{-\alpha} S_L^{-1} N_\ell^{\frac{\beta+\gamma}{p'+1}} N_\ell^{-\beta} \Big] \\
&= \frac{\epsilon}{2} + N_L^{-\alpha} S_L^{-1} \sum_{\ell=1}^{L} N_\ell^{\frac{\gamma-\beta p'}{p'+1}} = \frac{\epsilon}{2} + N_L^{-\alpha} < \epsilon.
\end{aligned}
$$

For the total cost, we first compute

$$\mathcal{C}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X) \eqsim \sum_{\ell=1}^{L} \mathcal{C}_\ell M_\ell \leq C_\gamma \sum_{\ell=1}^{L} N_\ell^\gamma \Big(1 + C_\star^{p'} N_L^{\alpha p'} S_L^{p'} N_\ell^{-\frac{(\beta+\gamma)p'}{p'+1}} \Big)$$
$$\leq C_\gamma \sum_{\ell=1}^{L} N_\ell^\gamma + C_\gamma C_\star^{p'} N_L^{\alpha p'} S_L^{p'} \sum_{\ell=1}^{L} N_\ell^{\frac{\gamma-\beta p'}{p'+1}} = C_\gamma \sum_{\ell=1}^{L} N_\ell^\gamma + C_\gamma C_\star^{p'} N_L^{\alpha p'} S_L^{p'+1}.$$

By the choice of L we have $A^L \eqsim N_L \eqsim \epsilon^{-1/\alpha}$ and, since $\epsilon \in (0, 1/2]$, we find that $L \eqsim_\alpha |\log_A \epsilon|$. Thus, using (3.29) we conclude that the computational cost,

$$\mathcal{C}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X) \lesssim_{(\beta,\gamma,A,p,q)} \begin{cases} N_L^\gamma + N_L^{\alpha p'} & \text{if } \beta p' > \gamma, \\ N_L^\gamma + N_L^{\alpha p'} L^{p'+1} & \text{if } \beta p' = \gamma, \\ N_L^\gamma + N_L^{\alpha p' + \gamma - \beta p'} & \text{if } \beta p' < \gamma, \end{cases}$$

in terms of the accuracy ϵ behaves as follows,

$$\mathcal{C}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X) \lesssim_{(\alpha,\beta,\gamma,A,p,q)} \begin{cases} \epsilon^{-\frac{\gamma}{\alpha}} + \epsilon^{-p'} & \text{if } \beta p' > \gamma, \\ \epsilon^{-\frac{\gamma}{\alpha}} + \epsilon^{-p'} |\log_A \epsilon|^{p'+1} & \text{if } \beta p' = \gamma, \\ \epsilon^{-\frac{\gamma}{\alpha}} + \epsilon^{-p' - \frac{\gamma-\beta p'}{\alpha}} & \text{if } \beta p' < \gamma, \end{cases}$$

which completes the proof of the assertion. □

Remark 3.26 *(Strong convergence implies (α)).* Lemma 3.19 shows that under the stability condition (stab), assumption (α) is satisfied whenever there exists a constant $\widetilde{C}_\alpha \in (0,\infty)$ such that $\|X - X_\ell\|_{L_k(\Omega;E)} \leq \widetilde{C}_\alpha N_\ell^{-\alpha}$ holds for all $\ell \in \mathbb{N}$.

Remark 3.27 *(Comparison with single-level Monte Carlo).* Under the assumptions (α), (γ), (stab) the single-level Monte Carlo approach of Corollary 3.20 requires to choose the level L and the number of samples M_L such that $N_L \eqsim \epsilon^{-1/\alpha}$ and $M_L \eqsim \epsilon^{-p'}$, in order to achieve a target accuracy $\mathrm{err}^{k,\mathsf{SL}}_{q,\varepsilon_s}(X) = \epsilon \in (0,\infty)$. Thus, the single-level Monte Carlo method to estimate $\mathbb{M}^k_\varepsilon[X]$ causes computational cost of the order

$$\mathcal{C}^{k,\mathsf{SL}}_{q,\varepsilon_s}(X) \eqsim \mathcal{C}_L M_L \lesssim N_L^\gamma M_L \eqsim \epsilon^{-\frac{\gamma}{\alpha} - p'}.$$

Remark 3.28 *(Comparison with MLMC in Hilbert spaces).* In the case that E is a Hilbert space, we have that $p = p' = 2$ and the computational costs in ($\mathcal{C}$) coincide e.g. with those of [12, Theorem 1] for the two cases when $\beta p' \neq \gamma$. In the critical case $\beta p' = 2\beta = \gamma$, we obtain an additional log-factor $|\log_A \epsilon|$. This is due to the fact that we do not assume independence across the levels $\ell \in \{1, \dots, L\}$. Note that this independence can be exploited only if (i) E is a Hilbert space, and (ii) the error is measured in the L_2-norm with respect to $(\Omega, \mathcal{A}, \mathbb{P})$.

Remark 3.29 *(Full injective tensor norm).* On the symmetric injective tensor product space $\otimes_{\varepsilon_s}^{k,s} E$, the full and symmetric injective tensor norms, $\|\cdot\|_\varepsilon$ and $\|\cdot\|_{\varepsilon_s}$, are equivalent, see (2.5). Therefore, the results of Subsections 3.2 and 3.3 on convergence of Monte Carlo methods of standard (Theorem 3.16), single-level (Corollary 3.20) and multilevel type (Theorems 3.24 and 3.25) hold also with respect to the stronger norm on $L_q(\Omega; \otimes_\varepsilon^k E)$, with the additional constant $\frac{k^k}{k!}$.

4. Applications

In this section we illustrate the preceding, abstract theory by several examples of stochastic equations, where the need for the presently developed modifications of the standard Monte Carlo theory is entailed either by problem-specific constraints on the choices of non-Hilbertian function spaces for well-posedness or by the interest in error estimates in norms on Banach spaces (such as Hölder norms).

Specifically, Subsections 4.1 and 4.2 are concerned with the kth moment MLMC finite element convergence analysis for explicit, linear, second-order elliptic PDEs with random forcing (in dimensions $d \in \{2, 3\}$) or random diffusion coefficient (for $d = 1$), respectively. Here, the right-hand side is assumed to be an element of (or taking values in) $L_p(\mathrm{D})$ for some $p \in (1, \infty)$, where $\mathrm{D} \subset \mathbb{R}^d$ denotes the spatial domain. To obtain well-posed problems, the case $p \in (1, 2)$ necessitates variational formulations on Banach spaces, whereas for $p \in (2, \infty)$ such formulations may be advantageous to derive error estimates in Hölder norms via Sobolev embeddings.

In Subsection 4.3 we discuss the MLMC approximation of higher-order moments for vector-valued stochastic processes $X \colon [0, T] \times \Omega \to E$ in tensor norms of Hölder spaces $C^\delta([0, T]; E)$ for problem-specific Hölder exponents $\delta \in [0, 1)$. These results are applicable to many semi-discrete or fully discrete numerical schemes for SDEs and stochastic PDEs and we give some explicit examples.

4.1. Linear elliptic PDEs with random forcing

Let $\mathrm{D} \subset \mathbb{R}^d$ with $d \in \{2, 3\}$ be an open, bounded, polytopal Lipschitz domain (with closure $\overline{\mathrm{D}}$) and, for $p \in [1, \infty]$ and $m \in \mathbb{N}$, let $L_p(\mathrm{D})$ and $W_p^m(\mathrm{D})$ denote the standard Lebesgue and Sobolev spaces of real-valued functions on D.

We write $\mathring{W}_p^1(\mathrm{D})$ for the closure of $C_c^\infty(\mathrm{D})$ (the space of smooth functions with compact support inside D) with respect to the norm on $W_p^1(\mathrm{D})$, and $W_p^{-1}(\mathrm{D})$ for the dual space of $\mathring{W}_{p'}^1(\mathrm{D})$, where $\frac{1}{p} + \frac{1}{p'} = 1$.

4.1.1. Deterministic model problem

We assume given deterministic, continuous diffusion coefficients $a_{ij} \in C^0(\overline{\mathrm{D}})$, with $a_{ij} = a_{ji}$, $1 \le i, j \le d$, which are uniformly positive definite. Thus, there exist constants $0 < \underline{a} \le \overline{a} < \infty$ such that, for all $x \in \overline{\mathrm{D}}$,

$$\forall \phi, \psi \in \mathbb{R}^d : \quad \sum_{i,j=1}^{d} a_{ij}(x)\phi_i\phi_j \geq \underline{a} \, \|\phi\|_{\mathbb{R}^d}^2, \quad \sum_{i,j=1}^{d} a_{ij}(x)\phi_i\psi_j \leq \overline{a} \, \|\phi\|_{\mathbb{R}^d}\|\psi\|_{\mathbb{R}^d}, \tag{4.1}$$

where $\| \cdot \|_{\mathbb{R}^d}$ denotes the Euclidean norm on $\mathbb{R}^d$.

For $p \in (1, \infty)$ and a given source term f in $L_p(\mathrm{D})$ (which in the sequel shall be generalized to be random), we consider the following variational formulation of a homogeneous Dirichlet boundary value problem: Find

$$u \in \mathring{W}_p^1(\mathrm{D}) : \quad B(u, v) = \langle f, v \rangle \quad \forall v \in \mathring{W}_{p'}^1(\mathrm{D}). \tag{4.2}$$

Here, $\langle \cdot , \cdot \rangle$ denotes the $W_p^{-1}(\mathrm{D}) \times \mathring{W}_{p'}^1(\mathrm{D})$ duality pairing, and the bilinear form B is given by

$$B \colon \mathring{W}_p^1(\mathrm{D}) \times \mathring{W}_{p'}^1(\mathrm{D}) \to \mathbb{R}, \qquad B(w, v) := \int_{\mathrm{D}} \sum_{i,j=1}^{d} a_{ij}(x) \tfrac{\partial}{\partial x_i} w(x) \tfrac{\partial}{\partial x_j} v(x) \, \mathrm{d}x.$$

Evidently, the boundedness condition in (4.1) and Hölder's inequality imply continuity of B on $\mathring{W}_p^1(\mathrm{D}) \times \mathring{W}_{p'}^1(\mathrm{D})$. However, as opposed to the Hilbert space case $p = p' = 2$, the uniform strong ellipticity assumption in (4.1) is in general not sufficient to guarantee that the mapping $\mathring{W}_p^1(\mathrm{D}) \ni u \mapsto B(u, \cdot) \in W_p^{-1}(\mathrm{D})$ is an isomorphism. For the case of the Laplace operator (i.e., $a_{ij}(x) = \delta_{ij}$) and every $p \in (1, \infty)$, an inf-sup condition and hence well-posedness of (4.2) have been shown in [55, Theorem 6.1], see also [8, Equation (8.6.5)]. Following the arguments used in [8, Section 8.6] this can be generalized to diffusion coefficients $(a_{ij})_{i,j=1}^d$ satisfying (4.1), provided that p is sufficiently close to 2. In what follows, we will *require* for an appropriate range of integrability indices $p \in (1, \infty)$ that (4.2) has a unique solution and, moreover, that this solution is $W_p^2(\mathrm{D})$-regular. This is summarized in the next assumption.

Assumption 4.1. There exists $p_0 \in (d, \infty)$ such that, for each $p \in (1, p_0)$ and every $f \in L_p(\mathrm{D})$, the variational problem (4.2) admits a unique solution $u \in \mathring{W}_p^1(\mathrm{D})$, and

$$\forall p \in (1, p_0) \quad \exists C_p \in (0, \infty) \quad \forall f \in L_p(\mathrm{D}) : \quad \|u\|_{W_p^2(\mathrm{D})} \leq C_p \|f\|_{L_p(\mathrm{D})}. \tag{4.3}$$

Sufficient conditions for the $W_p^2(\mathrm{D})$-regularity (4.3) to hold for the Laplace problem in polygons (i.e., $d = 2$) can, for instance, be found in [28, Theorem 4.3.2.4].

Since $p_0 > d$ and $d \in \{2, 3\}$ are assumed, for a given $q \in [p_0, \infty)$, we may choose $p = \frac{qd}{d+q} \in (1, \min\{d, q\})$ in (4.3) and conclude by continuity of the Sobolev embedding $W_p^2(\mathrm{D}) \subseteq W_q^1(\mathrm{D})$ and Hölder's inequality that $\|u\|_{W_q^1(\mathrm{D})} \lesssim_{(q,\mathrm{D})} \|u\|_{W_p^2(\mathrm{D})} \lesssim_p \|f\|_{L_p(\mathrm{D})} \lesssim_{(q,\mathrm{D})} \|f\|_{L_q(\mathrm{D})}$. For $q \in (1, p_0)$ this estimate trivially holds by (4.3) so that we obtain the following stability estimate for all $p \in (1, \infty)$:

$$\forall p \in (1, \infty) : \quad \|u\|_{W_p^1(\mathrm{D})} \lesssim_{(p,\mathrm{D})} \|f\|_{L_p(\mathrm{D})}. \tag{4.4}$$

4.1.2. Finite element approximation

For the numerical approximation, we use a conforming finite element method (FEM) for (4.2) based on continuous, piecewise first-order Langrangean basis functions on $\overline{\mathrm{D}}$: on a regular, simplicial triangulation $\mathcal{T}_h$ of $\overline{\mathrm{D}}$ with mesh size $h \in (0,\infty)$, we consider the finite-dimensional space

$$S_0^1(\mathrm{D};\mathcal{T}_h) := \left\{ v \in C^0(\overline{\mathrm{D}}) : v|_{\partial \mathrm{D}} = 0,\ v|_T \in \mathbb{P}_1\ \forall T \in \mathcal{T}_h \right\},$$

where $\mathbb{P}_1$ denotes the space of polynomials of degree at most one. The corresponding Galerkin discretization of (4.2) reads: Find

$$u_h \in S_0^1(\mathrm{D};\mathcal{T}_h): \quad B(u_h,v_h) = \langle f, v_h\rangle \quad \forall v_h \in S_0^1(\mathrm{D};\mathcal{T}_h). \tag{4.5}$$

Evidently, this finite-dimensional problem is equivalent to a linear system of equations, with matrix that is symmetric and, by (4.1), positive definite, so that there exists a unique solution $u_h \in S_0^1(\mathrm{D};\mathcal{T}_h)$ of (4.5).

Under Assumption 4.1 and the additional condition that

$$\forall i,j \in \{1,\dots,d\}: \quad a_{ij} \in W_q^1(\mathrm{D}) \quad \text{for some} \quad \begin{cases} q > 2 & \text{if } d=2, \\ q \geq \frac{12}{5} & \text{if } d=3, \end{cases} \tag{4.6}$$

it is shown in [8, Theorem 8.5.3] that, for all $p \in (1,\infty)$, the Galerkin projection u_h in (4.5) is bounded in $\mathring{W}_p^1(\mathrm{D})$: There exists a mesh size $h_0 \in (0,\infty)$ such that

$$\forall h \in (0,h_0): \quad \|u_h\|_{W_p^1(\mathrm{D})} \lesssim_{(p,\mathrm{D})} \|u\|_{W_p^1(\mathrm{D})}. \tag{4.7}$$

Combining (4.4) and (4.7) implies stability of both the exact solution u and its approximation u_h: For all $p \in (1,\infty)$, there exist a mesh width $h_0 \in (0,\infty)$ and a constant $\widetilde{C}_{\mathsf{stab}} \in (0,\infty)$ such that

$$\forall h \in (0,h_0): \quad \max\left\{ \|u\|_{W_p^1(\mathrm{D})}, \|u_h\|_{W_p^1(\mathrm{D})} \right\} \leq \widetilde{C}_{\mathsf{stab}} \|f\|_{L_p(\mathrm{D})}. \tag{4.8}$$

Moreover, by [8, Equation (8.5.4)] u_h is quasi-optimal in $\mathring{W}_p^1(\mathrm{D})$ for all $p \in (1,\infty)$: There exist $h_0, C_{\mathsf{opt}} \in (0,\infty)$ (which may depend on D and p) such that

$$\forall h \in (0,h_0): \quad \|u-u_h\|_{W_p^1(\mathrm{D})} \leq C_{\mathsf{opt}} \inf_{v_h \in S_0^1(\mathrm{D};\mathcal{T}_h)} \|u-v_h\|_{W_p^1(\mathrm{D})}.$$

Therefore, under Assumption 4.1 and (4.6), for every quasi-uniform family of triangulations $(\mathcal{T}_h)_{h\in\mathcal{H}}$, standard approximation properties of the corresponding finite element spaces $S_0^1(\mathrm{D};\mathcal{T}_h)$, $h \in \mathcal{H} \subseteq (0,\infty)$, show that, for all $p \in (0,p_0)$,

$$\forall h \in \mathcal{H} \cap (0,h_0): \quad \|u-u_h\|_{W_p^1(\mathrm{D})} \lesssim_{(p,\mathrm{D})} h\,\|u\|_{W_p^2(\mathrm{D})} \lesssim_{(p,\mathrm{D})} h\,\|f\|_{L_p(\mathrm{D})}, \tag{4.9}$$

where we also used the assumed regularity (4.3).

4.1.3. Random forcing and MLMC-FEM

Suppose the setting of the previous subsections. In particular, the coefficients $(a_{ij})_{i,j=1}^d$ satisfy (4.1) and (4.6), and Assumption 4.1 holds for some $p_0 \in (d, \infty)$. In this subsection we fix $p \in (1, p_0)$, which determines the spatial integrability of a given random forcing. Random forcing in (4.2) amounts to assuming that the right-hand side f is an element of $L_r(\Omega; L_p(\mathrm{D}))$ for some suitable integrability index $r \in [1, \infty)$ with respect to the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, i.e., we seek a $\mathring{W}_p^1(\mathrm{D})$-valued random variable u such that

$$B(u(\omega), v) = \langle f(\omega), v\rangle \quad \forall v \in \mathring{W}_{p'}^1(\mathrm{D}), \quad \text{for almost all } \omega \in \Omega. \tag{4.10}$$

Under the above mentioned regularity requirements, see Assumption 4.1, we may argue for almost all $\omega \in \Omega$ to establish the existence and uniqueness of a (stochastic) solution $u \in L_r(\Omega; \mathring{W}_p^1(\mathrm{D}))$ satisfying (4.10), with

$$u \in L_r(\Omega; W_p^2(\mathrm{D})), \qquad \|u\|_{L_r(\Omega; W_p^2(\mathrm{D}))} \lesssim_{(p,\mathrm{D})} \|f\|_{L_r(\Omega; L_p(\mathrm{D}))}.$$

Multilevel finite element discretizations of (4.10) will be based on the discrete variational problem (4.5), considered $\mathbb{P}$-a.s. To this end, we denote by $\{\mathcal{T}_\ell\}_{\ell\in\mathbb{N}}$ a nested sequence of regular, simplicial triangulations $\mathcal{T}_\ell$ of $\overline{\mathrm{D}}$, with corresponding sequence of mesh sizes $\{h_\ell\}_{\ell\in\mathbb{N}}$. We assume that $\mathcal{T}_{\ell+1}$ is obtained from $\mathcal{T}_\ell$ via uniform red refinement. Then, $h_{\ell+1} \lesssim h_\ell/2$ and, without loss of generality, we may assume that $h_1 < h_0$, with h_0 as in (4.7)–(4.9). The corresponding sequence of Galerkin solutions $u_{h_\ell} \in S_0^1(\mathrm{D}; \mathcal{T}_\ell)$ shall be denoted by u_ℓ (with slight abuse of notation).

In the next corollary we verify that all assumptions of the "$\alpha\beta\gamma$ theorem", see Theorem 3.25, are satisfied to bound the computational costs of the MLMC-FEM estimator for $\mathbb{M}_\varepsilon^k[u]$ for a given accuracy and provide an upper bound for these costs.

Corollary 4.2. *Let (4.1) and (4.6) be satisfied and suppose that Assumption 4.1 holds for some $p_0 \in (d, \infty)$. Assume that $p \in (1, p_0)$, $q \in [\min\{p, 2\}, \infty)$, and $k \in \mathbb{N}$. For $f \in L_{kq}(\Omega; L_p(\mathrm{D}))$, let $u \in L_{kq}(\Omega; \mathring{W}_p^1(\mathrm{D}))$ be the solution to (4.10). Furthermore, let the FEM approximations $(u_\ell)_{\ell\in\mathbb{N}}$ be constructed as described above. Then, for $E := W_p^1(\mathrm{D})$ and $N_\ell := \dim\big(S_0^1(\mathrm{D}; \mathcal{T}_\ell)\big) \eqsim h_\ell^{-d} \eqsim A^\ell$, with $A := 2^d$, all conditions of Theorem 3.25 are fulfilled,*

$$(\alpha) \quad \forall \ell \in \mathbb{N}: \qquad \big\|\mathbb{M}_\varepsilon^k[u] - \mathbb{M}_\varepsilon^k[u_\ell]\big\|_{\varepsilon_s} \lesssim_{(k,p,\mathrm{D},f)} N_\ell^{-1/d}, \qquad \text{i.e., } \alpha = d^{-1},$$

$$(\beta) \quad \forall \ell \in \mathbb{N}: \qquad \|u_\ell - u_{\ell-1}\|_{L_{kq}(\Omega; W_p^1(\mathrm{D}))} \lesssim_{(k,p,q,\mathrm{D},f)} N_\ell^{-1/d}, \qquad \text{i.e., } \beta = d^{-1},$$

$$(\gamma) \quad \forall \ell \in \mathbb{N}: \qquad \mathcal{C}_\ell \lesssim N_\ell^k, \qquad \text{i.e., } \gamma = k,$$

$$(\mathsf{stab}) \quad \forall \ell \in \mathbb{N}: \qquad \max\big\{\|u\|_{L_k(\Omega; W_p^1(\mathrm{D}))}, \|u_\ell\|_{L_{kq}(\Omega; W_p^1(\mathrm{D}))}\big\} \le C_{\mathsf{stab}},$$

for some constant $C_{\mathsf{stab}} \in (0, \infty)$ depending only on p, D and $\|f\|_{L_{kq}(\Omega; L_p(\mathrm{D}))}$.

Furthermore, the L_q*-accuracy* $\operatorname{err}^{k,\mathsf{ML}}_{q,\varepsilon_s}(u) < \epsilon \in (0, 1/2]$ *of the multilevel Monte Carlo estimator for* $\mathbb{M}^k_\varepsilon[u]$ *can be achieved at computational costs of the order*

$$\mathcal{C}^{k,\mathsf{ML}}_{q,\varepsilon_s}(u) \lesssim_{(k,p,q,\mathrm{D},f)} \begin{cases} \epsilon^{-\bar{p}'} & \text{if } \bar{p}' > kd, \\ \epsilon^{-kd}\,|\log_2 \epsilon|^{kd+1} & \text{if } \bar{p}' = kd, \\ \epsilon^{-kd} & \text{if } \bar{p}' < kd, \end{cases} \tag{4.11}$$

where $\bar{p}' \in [2, \infty)$ *is such that* $\frac{1}{\min\{p,2\}} + \frac{1}{\bar{p}'} = 1$.

Proof. First note that we may apply the deterministic stability estimate (4.8) for almost all $\omega \in \Omega$, showing that, for all $\ell \in \mathbb{N}$,

$$\max\bigl\{ \|u\|_{L_k(\Omega;W^1_p(\mathrm{D}))}, \|u_\ell\|_{L_{kq}(\Omega;W^1_p(\mathrm{D}))} \bigr\} \leq \widetilde{C}_{\mathsf{stab}} \|f\|_{L_{kq}(\Omega;L_p(\mathrm{D}))} =: C_{\mathsf{stab}}.$$

The integrability of $f \in L_{kq}(\Omega; L_p(\mathrm{D}))$ combined with the deterministic FEM convergence result (4.9) implies strong convergence,

$$\forall \ell \in \mathbb{N}: \quad \|u - u_\ell\|_{L_{kq}(\Omega;W^1_p(\mathrm{D}))} \lesssim_{(p,\mathrm{D})} h_\ell \, \|f\|_{L_{kq}(\Omega;L_p(\mathrm{D}))}.$$

Since $h_\ell \eqsim N_\ell^{-1/d}$, we conclude that the conditions (α) and (β) of Theorem 3.25 are satisfied with $\alpha = \beta = d^{-1}$, where we also have used Remark 3.26 for (α) and the triangle inequality for (β).

Assuming a linear complexity solver (as, e.g., multigrid), the cost $\mathcal{C}_{\ell,1}$ for computing one sample of u_ℓ in (4.5) is bounded by $\mathcal{C}_{\ell,1} \leq C_{\gamma,1} N_\ell$ with some constant $C_{\gamma,1} \in (0, \infty)$ independent of ℓ. Since the computation of the kth Kronecker product of a vector of length N_ℓ causes computational cost of the magnitude N_ℓ^k, the total cost $\mathcal{C}_\ell$ for computing one sample of the random variable $\xi_\ell = \otimes^k u_\ell - \otimes^k u_{\ell-1}$ is of the order N_ℓ^k. Therefore, the condition (γ) holds for $\gamma = k$.

Thus, the assumptions (α), (β), (γ) and (stab) of Theorem 3.25 are satisfied, and the upper bound for the computational costs in (4.11) follows upon applying Theorem 3.25, since the Banach space $E = W^1_p(\mathrm{D})$ has type $\min\{p, 2\}$. □

Remark 4.3. In the Hilbert space case, it is in general not optimal to obtain a convergence rate bound for (α) by combining strong convergence with stability (stab), as outlined in Remark 3.26. For instance, the error analysis of Galerkin approximations for generalized Whittle–Matérn fields in [15, Proposition 4] reveals that the corresponding approximations of the covariance function converge more than twice as fast in the $L_2(\mathrm{D}\times\mathrm{D})$-norm as the corresponding Gaussian random field approximations in the strong $L_q(\Omega; L_2(\mathrm{D}))$-sense. However, it is not obvious if and how this behavior generalizes to random variables with values in Banach spaces.

4.2. Linear elliptic PDEs with log-Gaussian coefficient

We next consider a linear, second-order elliptic PDE with mixed Dirichlet–Neumann boundary conditions and right-hand side in $L_p(\mathrm{D})$. As opposed to Subsection 4.1, we now assume that the diffusion coefficient a is random. More specifically, as e.g. in [9,10,56], we suppose that a is log-Gaussian, i.e., $a(x) = \exp(g(x))$ for almost all $x \in \mathrm{D}$, $\mathbb{P}$-a.s., for some Gaussian random field $g\colon \mathrm{D} \times \Omega \to \mathbb{R}$. The spatial domain is assumed to be a bounded interval $\mathrm{D} = I = (0, b)$ of length $b \in (0, \infty)$. The restriction of the spatial dimension to $d = 1$ facilitates an explicit expression of the inf-sup constant of the bilinear form, appearing in the corresponding variational formulation, depending on $\underline{a} := \operatorname{ess\,inf}_{x \in I} a(x)$. In this setting, $\underline{a}$ is a random variable satisfying $\underline{a} \in L_q(\Omega; \mathbb{R})$ for all $q \in [1, \infty)$. Similarly as in [9], this, in turn, yields well-posedness of the variational problem and strong convergence of finite element approximations.

4.2.1. Deterministic model problem

For $p \in [1, \infty]$, $m \in \mathbb{N}$, we recall the Lebesgue and Sobolev spaces $L_p(I)$ and $W_p^m(I)$ from Subsection 4.1. We furthermore note that, since $d = 1$, for every $p \in [1, \infty]$, elements v in $W_p^1(I)$ coincide (upon a modification on a subset of $\overline{I}$ of zero Lebesgue measure) with a unique function which is continuous on $\overline{I} = [0, b]$, denoted by $\widetilde{v} \in C^0(\overline{I})$, and we define the subspace

$$\mathring{W}^1_{p,\{0\}}(I) := \bigl\{ v \in W_p^1(I) : \widetilde{v}(0) = 0 \bigr\}.$$

In virtue of the Poincaré inequality, on this subspace the map $v \mapsto |v|_{W_p^1(I)}$, with $|v|_{W_p^1(I)} := \|v'\|_{L_p(I)}$, defines a norm, where v' denotes the weak derivative of v. In addition, we write $p' \in [1, \infty]$ for the Hölder conjugate of $p \in [1, \infty]$, and we let $W^{-1}_{p,\{0\}}(I)$ be the dual space of $\mathring{W}^1_{p',\{0\}}(I)$, equipped with the norm

$$\|f\|_{W^{-1}_{p,\{0\}}(I)} := \sup_{0 \neq v \in \mathring{W}^1_{p',\{0\}}(I)} \frac{\langle f, v \rangle}{|v|_{W^1_{p'}(I)}}.$$

We assume given a finite partition $\mathcal{P} = \{J_i\}_{i=1}^{n_{\mathcal{P}}}$ of pairwise disjoint, open subintervals J_i of I such that $\overline{J_1} \cup \ldots \cup \overline{J_{n_{\mathcal{P}}}} = \overline{I} = [0, b]$. Furthermore, we suppose that the scalar diffusion coefficient satisfies

$$a \in W^1_\infty(I; \mathcal{P}), \quad \text{where} \quad W^1_\infty(I; \mathcal{P}) := \bigl\{ a \in L_\infty(I) \,|\, \forall J \in \mathcal{P} : \, a|_J \in W^1_\infty(J) \bigr\},$$

and it is positive in the sense that there exist constants $\underline{a}$, $\overline{a}$ such that

$$0 < \underline{a} \leq a(x) \leq \overline{a} < \infty \qquad \text{for almost all } x \in I.$$

For $p \in (1, \infty)$ and $f \in L_p(I)$ we then consider the following boundary value problem, with mixed (Dirichlet–Neumann) boundary conditions: One wishes to find

$$u\colon \overline{I} \to \mathbb{R}: \quad -(a(x)u'(x))' = f(x), \quad \text{a.a. } x \in I, \quad u(0) = a(b)u'(b) = 0. \tag{4.12}$$

The weak formulation of (4.12) reads: Find

$$u \in \mathring{W}^1_{p,\{0\}}(I): \quad B_a(u,v) = \langle f, v\rangle \quad \forall v \in \mathring{W}^1_{p',\{0\}}(I), \tag{4.13}$$

where the bilinear form B_a is defined by $B_a(w,v) := \int_0^b a(x)w'(x)v'(x)\,\mathrm{d}x$, for every $w \in \mathring{W}^1_{p,\{0\}}(I)$ and all $v \in \mathring{W}^1_{p',\{0\}}(I)$. For every $p \in (1,\infty)$, existence and uniqueness of a solution u to (4.13) follow from continuity of B_a,

$$\forall w \in \mathring{W}^1_{p,\{0\}}(I) \quad \forall v \in \mathring{W}^1_{p',\{0\}}(I): \quad |B_a(w,v)| \le \overline{a}\, |w|_{W^1_p(I)} |v|_{W^1_{p'}(I)}, \tag{4.14}$$

and the following *inf-sup condition*:

$$\inf_{0 \ne w \in \mathring{W}^1_{p,\{0\}}(I)} \ \sup_{0 \ne v \in \mathring{W}^1_{p',\{0\}}(I)} \frac{B_a(w,v)}{|w|_{W^1_p(I)} |v|_{W^1_{p'}(I)}} \ge \underline{a}. \tag{4.15}$$

For the homogeneous Dirichlet boundary value problem ($u(0) = u(b) = 0$), a constructive proof for the inf-sup condition on $\mathring{W}^1_p(I) \times \mathring{W}^1_{p'}(I)$ has been given in [1, Proof of Theorem 3.1]. We adjust the argument from [1], to derive (4.15) for the problem (4.12) with mixed boundary conditions. To this end, let $w \in \mathring{W}^1_{p,\{0\}}(I) \setminus \{0\}$ be arbitrary but fixed, and define

$$v_w(x) := \int_0^x \operatorname{sign}(w'(t)) |w'(t)|^{p-1}\,\mathrm{d}t, \quad x \in \overline{I} = [0,b]. \tag{4.16}$$

This function satisfies $v_w(0) = 0$, and it is weakly differentiable with weak derivative

$$v_w'(x) = \operatorname{sign}(w'(x)) |w'(x)|^{p-1} \qquad \text{for almost all } x \in I.$$

We furthermore obtain that, for almost all $x \in I$, $|v_w'(x)| = |w'(x)|^{p-1} = |w'(x)|^{p/p'}$, and conclude that $v_w \in \mathring{W}^1_{p',\{0\}}(I)$, with

$$|v_w|_{W^1_{p'}(I)} = \|v_w'\|_{L_{p'}(I)} = \|w'\|^{p/p'}_{L_p(I)} = |w|^{p/p'}_{W^1_p(I)} = |w|^{p-1}_{W^1_p(I)}.$$

The continuity (4.14) of B_a implies that $B_a(w, v_w)$ is finite, and we find that

$$B_a(w, v_w) = \int_0^b a(x) |w'(x)|^p\,\mathrm{d}x \ge \underline{a}\, |w|^p_{W^1_p(I)} = \underline{a}\, |w|_{W^1_p(I)} |v_w|_{W^1_{p'}(I)}.$$

Since $w \in \mathring{W}^1_{p,\{0\}}(I) \setminus \{0\}$ was arbitrary, (4.15) follows.

The inf-sup condition (4.15) (together with its symmetric counterpart which can be shown in the same fashion) implies that, for every $f \in W^{-1}_{p,\{0\}}(I)$, the variational problem (4.13) admits a unique solution $u \in \mathring{W}^1_{p,\{0\}}(I)$. Furthermore, the linear data-to-solution mapping $W^{-1}_{p,\{0\}}(I) \ni f \mapsto u \in \mathring{W}^1_{p,\{0\}}(I)$ is an isomorphism with

$$\|u'\|_{L_p(I)} = |u|_{W^1_p(I)} \le \underline{a}^{-1} \|f\|_{W^{-1}_{p,\{0\}}(I)}. \tag{4.17}$$

In the case that $f \in L_p(I)$, this solution is more regular: Considering the differential equation (4.12) in weak sense on $J \in \mathcal{P}$ implies, for $a \in W^1_\infty(I;\mathcal{P})$ and $f \in L_p(I)$, that the second weak derivative of $u|_J$ restricted to $J \subseteq I$ exists and

$$-u|''_J(x) = a(x)^{-1} \left[f(x) + a|'_J(x) u'(x) \right] \quad \text{for almost all } x \in J.$$

Taking here the $L_p(J)$-norm yields with elementary estimates that, for every $J \in \mathcal{P}$,

$$\begin{aligned} \|u|''_J\|_{L_p(J)} &\le \|a^{-1}\|_{L_\infty(J)} \left[\|f\|_{L_p(J)} + \|a|'_J\|_{L_\infty(J)} \|u'\|_{L_p(J)} \right] \\ &\le \underline{a}^{-1} \Big[\|f\|_{L_p(I)} + \underline{a}^{-1} \|a|'_J\|_{L_\infty(J)} \|f\|_{W^{-1}_{p,\{0\}}(I)} \Big] \le C^{\mathsf{reg}}_{a,p} \|f\|_{L_p(I)}, \end{aligned}$$

where the constant $C^{\mathsf{reg}}_{a,p} \in (0,\infty)$ is given by

$$C^{\mathsf{reg}}_{a,p} := \underline{a}^{-1} \Big[1 + \underline{a}^{-1} \max_{J \in \mathcal{P}} \|a|'_J\|_{L_\infty(J)} \, C_{L_p \to W^{-1}_{p,\{0\}}} \Big],$$

and $C_{L_p \to W^{-1}_{p,\{0\}}} := \sup_{f \in B_{L_p(I)}} \|f\|_{W^{-1}_{p,\{0\}}(I)}$ denotes the norm of the continuous embedding $L_p(I) \subset W^{-1}_{p,\{0\}}(I)$.

Hence, for every $f \in L_p(I)$, the unique weak solution to (4.13) satisfies

$$u \in W^2_p(I;\mathcal{P}) \cap \mathring{W}^1_{p,\{0\}}(I), \qquad \max_{J \in \mathcal{P}} \|u|''_J\|_{L_p(J)} \le C^{\mathsf{reg}}_{a,p} \|f\|_{L_p(I)}, \tag{4.18}$$

where $W^2_p(I;\mathcal{P}) := \left\{ v \in W^1_p(I) \,\middle|\, \forall J \in \mathcal{P} : v|_J \in W^2_p(J) \right\}$ is the space of functions in $W^1_p(I)$ which are piecewise in W^2_p on the partition $\mathcal{P}$ of I.

4.2.2. Finite element approximation

For the numerical approximation of the solution $u \in \mathring{W}^1_{p,\{0\}}(I)$ to (4.13) we use a similar conforming finite element discretization as in Subsection 4.1.2. That is, we use continuous, piecewise affine-linear functions on a partition $\mathcal{T}_h$ of $\overline{I}$ with mesh size $h \in (0,\infty)$,

$$S^1_{0,\{0\}}(I;\mathcal{T}_h) := \left\{ v \in C^0(\overline{I}) : v(0) = 0, \ v|_T \in \mathbb{P}_1 \ \forall T \in \mathcal{T}_h \right\}.$$

Evidently, $S^1_{0,\{0\}}(I;\mathcal{T}_h) \subset \mathring{W}^1_{p,\{0\}}(I) \cap \mathring{W}^1_{p',\{0\}}(I)$ and $\dim\big(S^1_{0,\{0\}}(I;\mathcal{T}_h)\big) = \#(\mathcal{T}_h)$. For given $f \in L_p(I)$, the Galerkin discretization of (4.13) reads: One wishes to find

$$u_h \in S^1_{0,\{0\}}(I;\mathcal{T}_h): \quad B_a(u_h, v_h) = \langle f, v_h \rangle \quad \forall v_h \in S^1_{0,\{0\}}(I;\mathcal{T}_h). \tag{4.19}$$

Unique solvability of (4.19) follows from the (h-uniform) discrete inf-sup condition:

$$\inf_{0\neq w_h \in S^1_{0,\{0\}}(I;\mathcal{T}_h)} \sup_{0\neq v_h \in S^1_{0,\{0\}}(I;\mathcal{T}_h)} \frac{B_a(w_h, v_h)}{|w_h|_{W^1_p(I)} |v_h|_{W^1_{p'}(I)}} \geq \underline{a}. \tag{4.20}$$

To verify (4.20), note that the proof of (4.15) carries over to the discrete case: Given a function $w_h \in S^1_{0,\{0\}}(I;\mathcal{T}_h)$, one checks that the expression (4.16) yields an element v_h in $S^1_{0,\{0\}}(I;\mathcal{T}_h)$, and that all steps in the proof of (4.15) may be repeated verbatim.

The discrete inf-sup condition (4.20) and the continuity (4.14) imply that (4.19) admits a unique solution $u_h \in S^1_{0,\{0\}}(I;\mathcal{T}_h)$ with

$$\|u_h'\|_{L_p(I)} = |u_h|_{W^1_p(I)} \leq \underline{a}^{-1} \|f\|_{W^{-1}_{p,\{0\}}(I)}, \tag{4.21}$$

which is, furthermore, quasi-optimal:

$$|u - u_h|_{W^1_p(I)} \leq \left(1 + \frac{\overline{a}}{\underline{a}}\right) \inf_{v_h \in S^1_{0,\{0\}}(I;\mathcal{T}_h)} |u - v_h|_{W^1_p(I)}. \tag{4.22}$$

Therefore, for every quasi-uniform family of grids $(\mathcal{T}_h)_{h\in\mathcal{H}}$ on $\overline{I}$ which is such that, for every $h \in \mathcal{H}$, the grid $\mathcal{T}_h$ is compatible with the partition $\mathcal{P}$, the quasi-optimality (4.22) and the regularity (4.18) imply the error bound

$$\begin{aligned} |u - u_h|_{W^1_p(I)} &\leq \left(1 + \frac{\overline{a}}{\underline{a}}\right) |u - \mathcal{I}_h u|_{W^1_p(I)} = \left(1 + \frac{\overline{a}}{\underline{a}}\right) \left[\sum_{J\in\mathcal{P}} |u - \mathcal{I}_h u|^p_{W^1_p(J)}\right]^{1/p} \\ &\leq C_{b,p} \left(1 + \frac{\overline{a}}{\underline{a}}\right) h \left[\sum_{J\in\mathcal{P}} \|u|''_J\|^p_{L_p(J)}\right]^{1/p} \leq C_{b,p}\, n_{\mathcal{P}}^{1/p} \left(1 + \frac{\overline{a}}{\underline{a}}\right) C^{\mathsf{reg}}_{a,p}\, h\, \|f\|_{L_p(I)}, \end{aligned} \tag{4.23}$$

upon choosing v_h in (4.22) as the nodal interpolant $\mathcal{I}_h u$ of u in $S^1_{0,\{0\}}(I;\mathcal{T}_h)$. Here, the constant $C_{b,p} \in (0,\infty)$ is independent of a, h and u.

4.2.3. Log-Gaussian random coefficient and MLMC-FEM

The a-priori stability and discretization error bounds (4.17), (4.21) and (4.23) are explicit in the dependence on the coefficient a. They allow to consider (4.12) with deterministic source term $f \in L_p(I)$ for some $p \in (1,\infty)$, and with random coefficient $a\colon I \times \Omega \to \mathbb{R}$ whose logarithm $g\colon I \times \Omega \to \mathbb{R}$ is a Gaussian random field.

More specifically, we assume that the mapping $\Omega \ni \omega \mapsto g(\,\cdot\,,\omega)$ is a vector-valued random variable taking values in $W^1_\infty(I;\mathcal{P})$, where we note that $W^1_\infty(I;\mathcal{P})$, equipped with the norm

$$\|v\|_{W^1_\infty(I;\mathcal{P})} := \|v\|_{L_\infty(I)} + \max_{J\in\mathcal{P}} \|v|'_J\|_{L_\infty(J)} = \operatorname*{ess\,sup}_{x\in I} |v(x)| + \max_{J\in\mathcal{P}} \operatorname*{ess\,sup}_{x\in J} |v|'_J(x)|,$$

is a Banach space. Furthermore, g is assumed to be centered Gaussian, i.e., for any finite collection $(f_1, \ldots, f_n)$ in the dual space $[W^1_\infty(I;\mathcal{P})]'$ the distribution of $(\langle f_1, g\rangle, \ldots, \langle f_n, g\rangle)$ is multivariate Gaussian with zero mean. In other words, the law μ of g, defined for every set B in the Borel σ-algebra $\mathcal{B}(W^1_\infty(I;\mathcal{P}))$ by

$$\mu(B) = \mathbb{P}(\{\omega \in \Omega : g(\,\cdot\,,\omega) \in B\}),$$

satisfies that $\mu \circ f^{-1}$ is a centered Gaussian measure on $\mathbb{R}$ for any $f \in [W^1_\infty(I;\mathcal{P})]'$, see e.g. [7, Definition 2.2.1].

Under these assumptions we have, for almost all $\omega \in \Omega$,

$$a(\,\cdot\,,\omega) = \exp(g(\,\cdot\,,\omega)), \quad \text{with} \quad g(\,\cdot\,,\omega) \in W^1_\infty(I;\mathcal{P}), \tag{4.24}$$

and the trajectories of g and a are $\mathbb{P}$-a.s. continuous on each subinterval $J_1, \ldots, J_{n_\mathcal{P}} \subseteq I$ of the partition $\mathcal{P}$ (upon modifications on subsets of I of zero Lebesgue measure, which we again denote by g and a). We thus may define, for almost all $\omega \in \Omega$,

$$\begin{aligned} \underline{a}(\omega) &:= \min\nolimits_{J\in\mathcal{P}} \inf\nolimits_{x\in J} a(x,\omega) = \min\nolimits_{J\in\mathcal{P}} \exp\bigl(\inf\nolimits_{x\in J} g(x,\omega)\bigr), \\ \overline{a}(\omega) &:= \max\nolimits_{J\in\mathcal{P}} \sup\nolimits_{x\in J} a(x,\omega) = \max\nolimits_{J\in\mathcal{P}} \exp\bigl(\sup\nolimits_{x\in J} g(x,\omega)\bigr), \end{aligned} \tag{4.25}$$

so that we obtain, for almost all $\omega \in \Omega$,

$$0 < \exp(-\|g(\,\cdot\,,\omega)\|_{L_\infty(I)}) \le \underline{a}(\omega) \le \overline{a}(\omega) \le \exp(\|g(\,\cdot\,,\omega)\|_{L_\infty(I)}) < \infty.$$

For stability and strong convergence of finite element approximations of the solution u to (4.13) with the log-Gaussian coefficient $a = \exp(g)$, integrability of $\underline{a}^{-1}$, $\overline{a}$ and of $\max_{J\in\mathcal{P}} \|a|'_J\|_{L_\infty(J)}$ with respect to the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ will be crucial. This is summarized in the next lemma.

Lemma 4.4. *The in* (4.25) $\mathbb{P}$*-a.s. defined mappings* $\omega \mapsto \underline{a}(\omega)$ *and* $\omega \mapsto \overline{a}(\omega)$ *yield random variables satisfying* $\underline{a}^{-1}, \overline{a} \in L_r(\Omega;\mathbb{R})$ *for all* $r \in [1,\infty)$.

In addition, the mapping $\overline{a}' : \omega \mapsto \max_{J\in\mathcal{P}} \|a|'_J(\,\cdot\,,\omega)\|_{L_\infty(J)}$ *is* $\mathbb{P}$*-a.s. well-defined and* $\overline{a}' \in L_r(\Omega;\mathbb{R})$ *for all* $r \in [1,\infty)$.

Proof. The centered Gaussian random field g takes values in the Banach space $W^1_\infty(I;\mathcal{P})$. Thus, for every $J \in \mathcal{P}$ and almost all $\omega \in \Omega$, g admits a representative which is continuous on $\overline{J}$, and the proof of [9, Proposition 2.3] using Fernique's theorem is applicable on each $J \in \mathcal{P}$, showing that $\underline{a}^{-1}, \overline{a} \in L_r(\Omega;\mathbb{R})$ for all $r \in [1,\infty)$.

We now consider $\overline{a}'$. To this end, we first note that by (4.24) $a \in W^1_\infty(I;\mathcal{P})$, $\mathbb{P}$-a.s., since $\exp(\,\cdot\,)$ is smooth and $g \in W^1_\infty(I;\mathcal{P})$, $\mathbb{P}$-a.s. In particular, $a = \exp(g)$ is also a $W^1_\infty(I;\mathcal{P})$-valued random variable. Therefore, for almost all $\omega \in \Omega$ and every $J \in \mathcal{P}$, $\|a|'_J(\,\cdot\,,\omega)\|_{L_\infty(J)} < \infty$. It follows that $\overline{a}'$ is $\mathbb{P}$-a.s. well-defined and measurable, since the mapping $W^1_\infty(I;\mathcal{P}) \ni v \mapsto \max_{J\in\mathcal{P}} \|v|'_J\|_{L_\infty(J)} \in \mathbb{R}$ is continuous. To prove the integrability of $\overline{a}'$, we observe that, for almost all $\omega \in \Omega$ and every $J \in \mathcal{P}$, the weak derivative of $a|_J$ is given by $a|'_J(\,\cdot\,,\omega) = g|'_J(\,\cdot\,,\omega)\exp(g|_J(\,\cdot\,,\omega))$. Thus, we obtain that

$$\max_{J\in\mathcal{P}} \|a|'_J(\,\cdot\,,\omega)\|_{L_\infty(J)} \le \max_{J\in\mathcal{P}} \|g|'_J(\,\cdot\,,\omega)\|_{L_\infty(J)} \exp\big(\|g(\,\cdot\,,\omega)\|_{L_\infty(I)}\big).$$

We have $\mathbb{E}\big[\max_{J\in\mathcal{P}} \|g|'_J\|^q_{L_\infty(J)}\big] < \infty$ for all $q \in (0,\infty)$, since the distribution of g is Gaussian, with values in $W^1_\infty(I;\mathcal{P})$. For every $r \in [1,\infty)$, $\overline{a}' \in L_r(\Omega;\mathbb{R})$ can then be derived along the lines of the proof of [9, Proposition 2.3], using Fernique's theorem which shows that also $\mathbb{E}\big[\exp\big(q\,\|g\|_{L_\infty(I)}\big)\big] < \infty$ holds for all $q \in (0,\infty)$. □

We now consider the model problem introduced in Subsection 4.2.1 with a log-Gaussian coefficient a as in (4.24). That is, given a deterministic source $f \in L_p(I)$ for some $p \in (1,\infty)$, we seek $u\colon I \times \Omega \to \mathbb{R}$ such that, for almost all $\omega \in \Omega$,

$$u(\,\cdot\,,\omega) \in \mathring{W}^1_{p,\{0\}}(I): \quad B_{a(\,\cdot\,,\omega)}(u(\,\cdot\,,\omega), v) = \langle f, v\rangle \quad \forall v \in \mathring{W}^1_{p',\{0\}}(I), \tag{4.26}$$

where the bilinear form is as in (4.13). The following proposition addresses well-posedness of (4.26) and regularity of its solution in $L_r(\Omega)$-sense.

Proposition 4.5. *The variational problem (4.26) admits a solution that is $\mathbb{P}$-a.s. unique, and belongs to $L_r\big(\Omega;\mathring{W}^1_{p,\{0\}}(I)\big) \cap L_r\big(\Omega;W^2_p(I;\mathcal{P})\big)$ for all $r \in [1,\infty)$, with*

$$\big(\mathbb{E}\big[|u|^r_{W^1_p(I)}\big]\big)^{1/r} \le \big\|\underline{a}^{-1}\big\|_{L_r(\Omega;\mathbb{R})} \|f\|_{W^{-1}_{p,\{0\}}(I)}, \tag{4.27}$$

$$\Big(\mathbb{E}\Big[\big(\max_{J\in\mathcal{P}} \|u|''_J\|_{L_p(J)}\big)^r\Big]\Big)^{1/r} \le C^{\mathsf{reg}}_{a,p,r} \|f\|_{L_p(I)}, \tag{4.28}$$

where $C^{\mathsf{reg}}_{a,p,r} := \|\underline{a}^{-1}\|_{L_r(\Omega;\mathbb{R})} + \|\underline{a}^{-1}\|^2_{L_{4r}(\Omega;\mathbb{R})} \|\overline{a}'\|_{L_{2r}(\Omega;\mathbb{R})} C_{L_p\to W^{-1}_{p,\{0\}}} \in (0,\infty)$.

Proof. Since $f \in L_p(I)$ is deterministic and $a \in W^1_\infty(I;\mathcal{P})$ holds $\mathbb{P}$-a.s., existence of a solution to (4.26), which is $\mathbb{P}$-a.s. unique, follows by arguing via the well-posedness in the deterministic case (see Subsection 4.2.1) for almost all $\omega \in \Omega$. Furthermore, for every $r \in [1,\infty)$, the deterministic stability bound (4.17) combined with the integrability $\underline{a}^{-1} \in L_r(\Omega;\mathbb{R})$, see Lemma 4.4, imply (4.27).

We now show the regularity estimate (4.28). Recalling again the random variables $\underline{a}, \overline{a}, \overline{a}'$ from Lemma 4.4, by (4.18) we find that, for almost all $\omega \in \Omega$:

$$\max_{J\in\mathcal{P}} \|u|''_J(\,\cdot\,,\omega)\|_{L_p(J)} \le \Big[\underline{a}(\omega)^{-1} + \underline{a}(\omega)^{-2}\,\overline{a}'(\omega)\, C_{L_p\to W^{-1}_{p,\{0\}}}\Big] \|f\|_{L_p(I)}.$$

Taking the $L_r(\Omega;\mathbb{R})$-norm, and using the Minkowski and Hölder inequalities completes the proof of (4.28), and the constant $C^{\mathsf{reg}}_{a,p,r} > 0$ is finite by Lemma 4.4. □

We are now ready to formulate the "$\alpha\beta\gamma$" theorem for multilevel approximations of moments of the random solution to (4.26). To this end, let $\{u_\ell\}_{\ell\in\mathbb{N}}$ be a sequence of Galerkin approximations $u_\ell := u_{h_\ell}$, see (4.19), on partitions $\mathcal{T}_\ell$ of $\overline{I}$ corresponding to mesh sizes $h_\ell \eqsim 2^{-\ell}$. For example, $\mathcal{T}_\ell$ may be obtained by ℓ-fold bisection of the initial partition $\mathcal{T}_1 := \mathcal{P}$. Note that then $N_\ell = \dim\bigl(S^1_{0,\{0\}}(I;\mathcal{T}_\ell)\bigr) \eqsim h_\ell^{-1} \eqsim 2^\ell$.

Corollary 4.6. *Let $p \in (1,\infty)$, $q \in [\min\{p,2\},\infty)$, and $k \in \mathbb{N}$. For $f \in L_p(I)$, let $u \in L_{kq}\bigl(\Omega;\mathring{W}^1_{p,\{0\}}(I)\bigr)$ be the solution to (4.26). Assume further that the Galerkin approximations $(u_\ell)_{\ell\in\mathbb{N}}$ to (4.26) are constructed as described above. Then, for the Banach space $(E,\|\cdot\|_E) := \bigl(\mathring{W}^1_{p,\{0\}}(I), |\cdot|_{W^1_p(I)}\bigr)$ and with $N_\ell \eqsim 2^\ell$, all conditions of Theorem 3.25 are fulfilled,*

$$
\begin{aligned}
&(\alpha) \quad \forall \ell\in\mathbb{N}: && \bigl\|\mathbb{M}^k_\varepsilon[u] - \mathbb{M}^k_\varepsilon[u_\ell]\bigr\|_{\varepsilon_s} \lesssim_{(k,b,p,\mathcal{P},a,f)} N_\ell^{-1}, && \textit{i.e., } \alpha = 1,\\
&(\beta) \quad \forall \ell\in\mathbb{N}: && \|u_\ell - u_{\ell-1}\|_{L_{kq}(\Omega;\mathring{W}^1_{p,\{0\}}(I))} \lesssim_{(k,b,p,q,\mathcal{P},a,f)} N_\ell^{-1}, && \textit{i.e., } \beta = 1,\\
&(\gamma) \quad \forall \ell\in\mathbb{N}: && \mathcal{C}_\ell \lesssim N_\ell^k, && \textit{i.e., } \gamma = k,\\
&(\mathsf{stab}) \quad \forall \ell\in\mathbb{N}: && \max\bigl\{\|u\|_{L_k(\Omega;\mathring{W}^1_{p,\{0\}}(I))}, \|u_\ell\|_{L_{kq}(\Omega;\mathring{W}^1_{p,\{0\}}(I))}\bigr\} \le C_{\mathsf{stab}}, &&
\end{aligned}
$$

for some constant $C_{\mathsf{stab}} \in (0,\infty)$ depending only on $\|\underline{a}^{-1}\|_{L_{kq}(\Omega;\mathbb{R})}$ and $\|f\|_{W^{-1}_{p,\{0\}}(I)}$.

The L_q-accuracy $\mathrm{err}^{k,\mathsf{ML}}_{q,\varepsilon_s}(u) < \epsilon \in (0,1/2]$ of the multilevel Monte Carlo estimator for $\mathbb{M}^k_\varepsilon[u]$ can be achieved at computational costs of the order (4.11) with $d = 1$.

Proof. We first note that (4.27) combined with the deterministic discrete stability estimate (4.21) and the fact that $\underline{a}^{-1} \in L_{kq}(\Omega;\mathbb{R})$ imply (stab):

$$
\forall \ell\in\mathbb{N}: \quad \max\Bigl\{\bigl(\mathbb{E}\bigl[|u|^k_{W^1_p(I)}\bigr]\bigr)^{1/k}, \bigl(\mathbb{E}\bigl[|u_\ell|^{kq}_{W^1_p(I)}\bigr]\bigr)^{1/kq}\Bigr\} \le \|\underline{a}^{-1}\|_{L_{kq}(\Omega;\mathbb{R})}\|f\|_{W^{-1}_{p,\{0\}}(I)}.
$$

Next, we observe strong convergence of the finite element approximations $(u_\ell)_{\ell\in\mathbb{N}}$: For all $r \in [1,\infty)$, we obtain by exploiting the deterministic error estimate (4.23) for almost all $\omega \in \Omega$ that

$$
\begin{aligned}
\bigl(\mathbb{E}\bigl[|u-u_\ell|^r_{W^1_p(I)}\bigr]\bigr)^{1/r} &\lesssim_{(b,p,\mathcal{P})} h_\ell \Bigl\|\bigl(1+\tfrac{\overline{a}}{\underline{a}}\bigr)\max_{J\in\mathcal{P}} \|u|''_J\|_{L_p(J)}\Bigr\|_{L_r(\Omega;\mathbb{R})}\\
&\le h_\ell\bigl(1+\|\overline{a}\|_{L_{4r}(\Omega;\mathbb{R})}\|\underline{a}^{-1}\|_{L_{4r}(\Omega;\mathbb{R})}\bigr)\Bigl\|\max_{J\in\mathcal{P}}\|u|''_J\|_{L_p(J)}\Bigr\|_{L_{2r}(\Omega;\mathbb{R})}\\
&\le h_\ell\bigl(1+\|\overline{a}\|_{L_{4r}(\Omega;\mathbb{R})}\|\underline{a}^{-1}\|_{L_{4r}(\Omega;\mathbb{R})}\bigr) C^{\mathsf{reg}}_{a,p,2r}\|f\|_{L_p(I)},
\end{aligned}
$$

where we also used (4.28) of Proposition 4.5. Thus, the conditions (α) and (β) are satisfied with $\alpha = \beta = 1$ by Remark 3.26 and the triangle inequality, respectively.

Finally, the complexity of computing the Galerkin approximation u_ℓ in (4.19) per one realization of the Gaussian random field $g(\cdot,\omega) = \log(a(\cdot,\omega))$ (assumed given) at discretization level $\ell \in \mathbb{N}$ scales linearly with $N_\ell = \dim\bigl(S^1_{0,\{0\}}(I;\mathcal{T}_\ell)\bigr)$: Observe that the linear system of N_ℓ equations for the N_ℓ unknowns corresponding to (4.19) for each sample of $a(\,\cdot\,,\omega)$ is tridiagonal and symmetric, positive definite when the standard Courant ("hat functions") basis is adopted in (4.19). The formation of the kth order (full, algebraic) tensor product $\otimes^k u_\ell$ then entails the cost bound $\mathcal{C}_\ell \lesssim \mathcal{C}_\gamma N_\ell^k$ for one sample of the random variable $\xi_\ell = \otimes^k u_\ell - \otimes^k u_{\ell-1}$. $\square$

4.3. Approximation of stochastic processes in Hölder spaces

In this subsection we let $T \in (0,\infty)$ be a time horizon and consider approximating higher-order moments of vector-valued stochastic processes $X\colon [0,T]\times\Omega \to E$ by means of multilevel Monte Carlo methods. In particular, we derive explicit convergence rates for the error of the corresponding approximation in injective tensor norms of Hölder spaces $C^\delta([0,T];E)$. We furthermore detail the implications of this general result for the Euler–Maruyama method for stochastic ordinary differential equations, and give an overview of further possible applications including approximations of stochastic partial differential equations.

In order to properly define the relevant Hölder spaces, we introduce for a Banach space $(F,\|\cdot\|_F)$ and $\delta\in(0,1)$ the mappings

$$|\cdot|_{C^\delta([0,T];F)},\ \|\cdot\|_{C^\delta([0,T];F)}\colon C([0,T];F)\to[0,\infty]$$

on the Banach space

$$\big(C([0,T];F), \|\cdot\|_{C([0,T];F)}\big), \qquad \|f\|_{C([0,T];F)} := \sup_{t\in[0,T]} \|f(t)\|_F,$$

of continuous functions from $[0,T]$ to $(F,\|\cdot\|_F)$ via

$$|f|_{C^\delta([0,T];F)} := \sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{\|f(s)-f(t)\|_F}{|s-t|^\delta},$$

$$\|f\|_{C^\delta([0,T];F)} := \sup_{t\in[0,T]} \|f(t)\|_F + |f|_{C^\delta([0,T];F)}.$$

We note that the norm $\|\cdot\|_{C^\delta([0,T];F)}$ renders the subspace

$$C^\delta([0,T];F) = \big\{ f\in C([0,T];F) : \|f\|_{C^\delta([0,T];F)} < \infty \big\} \subset C([0,T];F)$$

of F-valued, δ-Hölder continuous functions a Banach space. For brevity, we also use the notation $C^0([0,T];F) := C([0,T];F)$ to include the case $\delta = 0$.

We now consider the setting of [14, Section 5], that is, we are given a stochastic process $X\colon [0,T]\times\Omega\to E$ with continuous sample paths satisfying the following regularity assumption: There exists a constant $\bar\beta\in(0,1]$ such that

$$\forall \beta \in [0, \bar{\beta}) \quad \forall q \in [1, \infty): \quad X \in C^{\beta}([0,T]; L_q(\Omega; E)). \tag{4.29}$$

Recall from Subsection 2.1 that $(E, \|\cdot\|_E)$ is assumed to be a real Banach space. Additionally, we let $(Y^N)_{N\in\mathbb{N}}$ be a sequence of approximations $Y^N\colon [0,T] \times \Omega \to E$ to the process X with continuous sample paths, which is known to converge at the nodes of the temporal partitions

$$\Theta^N := \{t_0^N, t_1^N, \ldots, t_{\#(\Theta^N)-1}^N\}, \qquad \#(\Theta^N) < \infty, \qquad N \in \mathbb{N},$$
$$0 =: t_0^N < t_1^N < \ldots < t_{\#(\Theta^N)-2}^N < t_{\#(\Theta^N)-1}^N := T,$$

in the strong sense essentially at the rate $\bar{\beta} \in (0, 1]$, i.e.,

$$\forall \beta \in [0, \bar{\beta}) \quad \forall q \in [1, \infty): \quad \sup_{t \in \Theta^N} \left\| X(t) - Y^N(t) \right\|_{L_q(\Omega; E)} \lesssim_{(\bar{\beta}, q, T)} \left| \triangle t_{\max}^N \right|^{\beta}, \tag{4.30}$$

where $\triangle t_{\max}^N := \max_{j \in \{0,\ldots,\#(\Theta^N)-2\}} |t_{j+1}^N - t_j^N|$. These partitions do not necessarily have to be equidistant or nested. We only require the following quasi-uniformity:

$$\sup_{N\in\mathbb{N}} \frac{\triangle t_{\max}^N}{\triangle t_{\min}^N} < \infty, \qquad \text{and} \qquad \lim_{N\to\infty} \triangle t_{\max}^N = 0, \tag{4.31}$$

where $\triangle t_{\min}^N$ is defined as $\triangle t_{\max}^N$ with the maximum replaced by the minimum. Furthermore, we assume that, for every $N \in \mathbb{N}$, the approximation Y^N is linearly interpolated on the partition Θ^N, i.e., for all $j \in \{0, \ldots, \#(\Theta^N) - 2\}$,

$$Y^N(s) = \frac{\left(t_{j+1}^N - s\right) Y^N\left(t_j^N\right)}{t_{j+1}^N - t_j^N} + \frac{\left(s - t_j^N\right) Y^N\left(t_{j+1}^N\right)}{t_{j+1}^N - t_j^N}, \qquad s \in \left[t_j^N, t_{j+1}^N\right]. \tag{4.32}$$

This general setting facilitates combining the abstract multilevel Monte Carlo results of Subsection 3.3 with [14, Corollary 2.11] and, thus, quantifying the convergence of the MLMC estimator for $\mathbb{M}_\varepsilon^k[X]$ based on approximations $Y^{N_1}, \ldots, Y^{N_L}$ in the norm on $L_q(\Omega; \otimes_{\varepsilon_s}^{k,s} C^\delta([0,T]; E))$ for $q \in [p, \infty)$ and $\delta \in [0, \bar{\beta})$, where $p \in [1, 2]$ is the Rademacher type of E, see Theorem 4.8 below. To this end, the following proposition which readily follows from [14, Corollary 2.11] will be crucial.

Proposition 4.7. *Let $(\Theta^N)_{N\in\mathbb{N}} \subset [0,T]$ be a sequence of partitions fulfilling (4.31). Assume that $X, Y^N\colon [0,T] \times \Omega \to E$, $N \in \mathbb{N}$, are stochastic processes with continuous sample paths, such that, for all $N \in \mathbb{N}$ and $t \in [0,T]$, the random variables $X(t), Y^N(t)\colon \Omega \to E$ are Bochner measurable and there exists $\bar{\beta} \in (0,1]$ such that (4.29) and (4.30) hold. In addition, for every $N \in \mathbb{N}$, let Y^N be linearly interpolated on the partition Θ^N, see (4.32).*

Then, we have for every $q \in [1,\infty)$, $\delta \in [0,\bar{\beta})$, *and all* $\epsilon \in (0,\infty)$,

$$\|X\|_{L_q(\Omega;C^\delta([0,T];E))} + \sup_{N\in\mathbb{N}} \|Y^N\|_{L_q(\Omega;C^\delta([0,T];E))} < \infty,$$

$$\|X - Y^N\|_{L_q(\Omega;C^\delta([0,T];E))} \lesssim_{(\bar{\beta},\delta,\epsilon,q,T)} \left|\triangle t^N_{\max}\right|^{\bar{\beta}-\delta-\epsilon}.$$

Theorem 4.8. *Suppose that all assumptions of Proposition 4.7 are fulfilled. In addition, let* $(E, \|\cdot\|_E)$ *be of Rademacher type* $p \in [1,2]$, $k, L \in \mathbb{N}$ *and* $\{N_\ell\}_{\ell\in\mathbb{N}} \subseteq \mathbb{N}$ *be a strictly increasing sequence of integers. Assume further that, for all* $\ell \in \{1,\dots,L\}$, $M_\ell \in \mathbb{N}$ *and* $\xi_{\ell,1},\dots,\xi_{\ell,M_\ell}$ *are independent copies of the random variable*

$$\otimes^k Y^{N_\ell} - \otimes^k Y^{N_{\ell-1}} \colon \Omega \to \otimes^{k,s}_{\varepsilon_s} C([0,T];E), \qquad Y^{N_0} := 0 \in C([0,T];E).$$

Then, for every $q \in [p,\infty)$, $\delta \in [0,\bar{\beta})$, *and all* $\epsilon \in (0,\infty)$, *we have that*

$$\begin{aligned}\mathrm{err}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X;\delta) &:= \left\| \mathbb{M}^k_\varepsilon[X] - \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{j=1}^{M_\ell} \xi_{\ell,j} \right\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s} C^\delta([0,T];E))} \\ &\lesssim_{(\bar{\beta},\delta,\epsilon,k,p,q,T)} \left|\triangle t^{N_L}_{\max}\right|^{\bar{\beta}-\delta-\epsilon} + \sum_{\ell=1}^{L} M_\ell^{-\left(1-\frac{1}{p}\right)} \left|\triangle t^{N_{\ell-1}}_{\max}\right|^{\bar{\beta}-\delta-\epsilon}.\end{aligned}$$

Proof. We fix $q \in [p,\infty)$, $\delta \in [0,\bar{\beta})$, $\epsilon \in (0,\infty)$, and first observe that by the triangle inequality on $L_q(\Omega;\otimes^{k,s}_{\varepsilon_s} C^\delta([0,T];E))$,

$$\begin{aligned}\mathrm{err}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X;\delta) &\le \left\|\mathbb{M}^k_\varepsilon[X] - \mathbb{M}^k_\varepsilon\left[Y^{N_L}\right]\right\|_{\otimes^{k,s}_{\varepsilon_s} C^\delta([0,T];E)} \\ &\quad + \left\| \mathbb{M}^k_\varepsilon\left[Y^{N_L}\right] - \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{j=1}^{M_\ell} \xi_{\ell,j} \right\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s} C^\delta([0,T];E))} =: (\mathrm{A}) + (\mathrm{B}).\end{aligned}$$

Term (A) can be bounded by combining Lemma 3.19 with the stability and convergence results of Proposition 4.7, showing that

$$\begin{aligned}(\mathrm{A}) &\le \left\|X - Y^{N_L}\right\|_{L_k(\Omega;C^\delta([0,T];E))} \sum_{i=0}^{k-1} \left[\|X\|^i_{L_k(\Omega;C^\delta([0,T];E))} \left\|Y^{N_L}\right\|^{k-i-1}_{L_k(\Omega;C^\delta([0,T];E))} \right] \\ &\lesssim_{(\bar{\beta},\delta,\epsilon,k,T)} \left|\triangle t^{N_L}_{\max}\right|^{\bar{\beta}-\delta-\epsilon}.\end{aligned}$$

To bound term (B), we may without loss of generality assume that $\epsilon < \bar{\beta} - \delta$ and define $\beta := \delta + \frac{\epsilon}{2} \in (\delta,\bar{\beta})$. We then exploit continuous embeddings, similarly as in the proof of [14, Corollary 5.15]: There are constants $C_1, C_2 \in (0,\infty)$, depending only on β, δ, T, such that

$$\|f\|_{C^\delta([0,T];E)} \le C_1 \|f\|_{W^{\frac{s}{p}}_{\bar{p}}((0,T);E)} \le C_2 \|f\|_{C^\beta([0,T];E)} \quad \forall f \in C^\beta([0,T];E),$$

where $\bar{s} := \frac{\beta+\delta}{2} \in (\delta, \beta) \subset (0,1)$ and $\bar{p} := \frac{4}{\beta-\delta} \in (4, \infty)$. Here, for $s \in (0,1)$ and $q \in [1,\infty)$, the space $W^s_q((0,T);E)$ denotes the vector-valued fractional Sobolev space, see e.g. [35, Definition 2.5.16]. Continuous embeddings are preserved under (full or symmetric) injective tensor products and, thus, we also have that

$$\|u\|_{\otimes^{k,s}_{\varepsilon_s} C^\delta([0,T];E)} \lesssim_{(\beta,\delta,k,T)} \|u\|_{\otimes^{k,s}_{\varepsilon_s} W^{\bar{s}}_{\bar{p}}((0,T);E)} \quad \forall u \in \otimes^{k,s}_{\varepsilon_s} W^{\bar{s}}_{\bar{p}}((0,T);E).$$

In addition, we note that $\bar{E} := W^{\bar{s}}_{\bar{p}}((0,T);E)$ has type $\min\{p,\bar{p}\} = p$: This observation follows from the fact that both $L^{\bar{p}}((0,T);E)$ and $W^1_{\bar{p}}((0,T);E)$ have type $\min\{p,\bar{p}\}$ (see [36, Proposition 7.1.4]) combined with the property that

$$\bar{E} = W^{\bar{s}}_{\bar{p}}((0,T);E) = \big(L^{\bar{p}}((0,T);E), W^1_{\bar{p}}((0,T);E)\big)_{\bar{s},\bar{p}}$$

is the real interpolation space between $L^{\bar{p}}((0,T);E)$ and $W^1_{\bar{p}}((0,T);E)$ (cf. [35, Theorem 2.5.17]) and the specification of the type of interpolation spaces [36, Proposition 7.1.3]. Thus, we may conclude with Theorem 3.24 and Proposition 4.7 that

$$\begin{aligned}
\text{(B)} &\lesssim_{(\beta,\delta,k,T)} \Big\| \mathbb{M}^k_\varepsilon\big[Y^{N_L}\big] - \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{j=1}^{M_\ell} \xi_{\ell,j} \Big\|_{L_q(\Omega;\otimes^{k,s}_{\varepsilon_s}\bar{E})} \\
&\lesssim_{(k,p,q)} \sum_{\ell=1}^{L} \Big[M_\ell^{-\left(1-\frac{1}{p}\right)} \big\|Y^{N_\ell} - Y^{N_{\ell-1}}\big\|_{L_{kq}(\Omega;\bar{E})} \max_{1\le\ell\le L} \big\|Y^{N_\ell}\big\|^{k-1}_{L_{kq}(\Omega;\bar{E})} \Big] \\
&\lesssim_{(\beta,\delta,k,T)} \sup_{N\in\mathbb{N}} \big\|Y^N\big\|^{k-1}_{L_{kq}(\Omega;C^\beta([0,T];E))} \sum_{\ell=1}^{L} \Big[M_\ell^{-\left(1-\frac{1}{p}\right)} \big\|X - Y^{N_{\ell-1}}\big\|_{L_{kq}(\Omega;C^\beta([0,T];E))} \Big] \\
&\lesssim_{(\bar{\beta},\delta,\epsilon,\widetilde{\epsilon},k,q,T)} \sum_{\ell=1}^{L} M_\ell^{-\left(1-\frac{1}{p}\right)} \big|\triangle t^{N_{\ell-1}}_{\max}\big|^{\bar{\beta}-\beta-\widetilde{\epsilon}} = \sum_{\ell=1}^{L} M_\ell^{-\left(1-\frac{1}{p}\right)} \big|\triangle t^{N_{\ell-1}}_{\max}\big|^{\bar{\beta}-\delta-\frac{\epsilon}{2}-\widetilde{\epsilon}}
\end{aligned}$$

holds for all $\widetilde{\epsilon} \in (0,\infty)$, and the claim follows for the choice $\widetilde{\epsilon} := \frac{\epsilon}{2}$. $\square$

Example 4.9 *(Euler–Maruyama method for SDEs).* Let $(\mathcal{F}_t)_{t\in[0,T]}$ be a normal filtration on $(\Omega, \mathcal{A}, \mathbb{P})$ and let $B\colon [0,T]\times\Omega \to \mathbb{R}^m$ be an m-dimensional $(\mathcal{F}_t)_{t\in[0,T]}$-Brownian motion (with continuous sample paths). For Lipschitz continuous functions $\mu\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$, consider the $(\mathcal{F}_t)_{t\in[0,T]}$-adapted stochastic process $X\colon [0,T]\times\Omega \to \mathbb{R}^d$ with continuous sample paths that satisfies

$$X(t) = X(0) + \int_0^t \mu(X(s))\,\mathrm{d}s + \int_0^t \sigma(X(s))\,\mathrm{d}B(s), \quad \mathbb{P}\text{-a.s.},$$

as well as the linearly interpolated Euler–Maruyama approximations $(Y^N)_{N\in\mathbb{N}}$ to X, defined with respect to equidistant partitions of size $\triangle t^N_{\max} = \triangle t^N_{\min} = {}^T\!/_N$ as follows: $Y^N(0) := X(0)$, and for $j \in \{0,\dots,N-1\}$ and $s \in \big(\frac{jT}{N}, \frac{(j+1)T}{N}\big]$,

$$Y^N(s) := Y^N(\tfrac{jT}{N}) + \big(s - \tfrac{jT}{N}\big)\mu\big(Y^N(\tfrac{jT}{N})\big) + \big(\tfrac{sN}{T} - j\big)\sigma\big(Y^N(\tfrac{jT}{N})\big)\big(B(\tfrac{(j+1)T}{N}) - B(\tfrac{jT}{N})\big).$$

Then, using the notation of Theorem 4.8 with $E := \mathbb{R}^d$, we conclude that for every $q \in [2, \infty)$, $\delta \in [0, 1/2)$ and all $\epsilon \in (0, \infty)$,

$$\operatorname{err}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X;\delta) \lesssim_{(\delta,\epsilon,k,q,T)} N_L^{-(\frac{1}{2}-\delta-\epsilon)} + \sum_{\ell=1}^{L} M_\ell^{-\frac{1}{2}} N_{\ell-1}^{-(\frac{1}{2}-\delta-\epsilon)}.$$

For the choice $N_\ell := 2^\ell$ and $M_\ell := 2^{L-\ell}$, this yields the error bound

$$\operatorname{err}^{k,\mathsf{ML}}_{q,\varepsilon_s}(X;\delta) \lesssim_{(\delta,\epsilon,k,q,T)} 2^{-L(\frac{1}{2}-\delta-\epsilon)} + \sum_{\ell=1}^{L} 2^{-\frac{L-\ell}{2}} 2^{-\ell(\frac{1}{2}-\delta-\epsilon)} \lesssim 2^{-L(\frac{1}{2}-\delta-\epsilon)}.$$

Since $\otimes^k_\varepsilon C([0,T];\mathbb{R}) = C([0,T]^k;\mathbb{R})$, see [54, Section 3.2, p. 50], this error estimate holds in particular also on $C([0,T]^k;\mathbb{R})$.

Remark 4.10. We note that Theorem 4.8 is applicable to a variety of numerical schemes developed for stochastic evolution problems (such as SDEs and stochastic PDEs), for which the regularity (4.29) of the solution process is known, and strong convergence rates are available at the nodes of the temporal partitions in $L_q(\Omega; E)$-sense for any $q \in [1, \infty)$. This list includes for instance:

(a) SDEs with coefficients that are not globally Lipschitz continuous, see e.g. [29, Theorem 3.1], [34, Theorem 1.1] and [53, Theorem 4.5];
(b) fully discrete (in space and time) approximations for linear or semilinear parabolic stochastic PDEs, see e.g. [44, Theorem 3.14];
(c) fully discrete approximations for non-linear stochastic PDEs, such as the stochastic Allen–Cahn equation, see e.g. [4, Theorem 1.1].

5. Conclusions

We have analyzed the convergence of Monte Carlo sampling for higher-order moments of Banach space valued random variables. Specifically, for every $k \in \mathbb{N}$, we have derived explicit, k-independent strong convergence rates in the injective tensor norm for approximating the kth moment $\mathbb{M}^k_\varepsilon[X]$ of a random variable $X \colon \Omega \to E$, taking values in a Banach space E, by means of

I. *standard Monte Carlo* sampling, involving no further numerical approximation, see Theorem 3.16;
II. the *single-level Monte Carlo* method, combining Monte Carlo sampling with an approximation $X_1 \colon \Omega \to E_1$ of X to generate samples in a (usually finite-dimensional) subspace $E_1 \subseteq E$, see Corollary 3.20;

III. the *multilevel Monte Carlo* method, combining Monte Carlo sampling with a hierarchy of approximations $X_\ell \colon \Omega \to E_\ell$, $\ell \in \{1, \dots, L\}$, in (usually nested, finite-dimensional) subspaces $E_\ell \subseteq E$, see Theorems 3.24 and 3.25.

These findings extend the numerical analysis of Monte Carlo based algorithms in computational uncertainty quantification to a broad range of mathematical models beyond the classical theory in Hilbert spaces, which relies on assumed square-integrability and bias-variance decompositions. Several examples have illustrated the wide scope of the presently developed theory: linear, second-order elliptic PDEs with data affording well-posedness in W_p^1, and stochastic evolution equations with almost sure path regularity in Hölder spaces.

The results of Subsections 3.2 and 3.3 are essential for the error analysis of Monte Carlo approximations of k-point correlations for every operator equation with random input data which, due to modeling or physical constraints, does not admit a well-posed formulation in Hilbert spaces. We indicate some further applications, where this is of relevance: In [42] Monte Carlo finite difference discretizations for scalar, degenerate convection-diffusion equations with random initial data were considered. In that case, the particular structure of the degeneracy in the diffusion coefficient, imposed from physical properties of the underlying model, mandated a mathematical formulation in Banach spaces of type $p < 2$. Assuming random initial data, the corresponding Monte Carlo error analysis for mean values of the solution therefore required a setting in Banach spaces as in Corollary 3.15. With the abstract MLMC results of Theorems 3.24 and 3.25, the MLMC finite difference convergence analysis for first-order moments of [42] generalizes to spatiotemporal k-point correlations with $k \geq 2$.

Another application is related to fluid flows: For the compressible Navier–Stokes equations with spatially periodic solutions, the (isentropic) equation of state relates the pressure P to the fluid density ϱ via $P(\varrho) = a\varrho^\gamma$, where $a > 0$ and $\gamma > 1$ are *physical constants*. In well-posed variational formulations [48], the density $\varrho(t, \cdot)$ and the corresponding momentum $\boldsymbol{m}(t, \cdot)$ at time $t \in [0, T]$ take values in $L_\gamma(\mathbb{T}^d)$ and $L_{\frac{2\gamma}{\gamma+1}}(\mathbb{T}^d; \mathbb{R}^d)$, respectively, where $d \in \{2, 3\}$ and $\mathbb{T}^d$ denotes the d-dimensional torus. With random data, this entails a Banach space setting of type $p = \min\{\gamma, 2\}$ for the density and $p = \frac{2\gamma}{\gamma+1} \in (1, 2)$ for the momentum. The convergence of *single-level* Monte Carlo finite volume approximations for higher-order moments of $\varrho(t, \cdot)$ and $\boldsymbol{m}(t, \cdot)$ has been discussed in the recent work [17], using the isomorphic identification

$$\otimes_\gamma^k L_\gamma(\mathbb{T}^d) \cong L_\gamma(\mathbb{T}^{kd}) \tag{5.1}$$

(and similarly for the d components of the momentum). Here, $\otimes_\gamma^k$ indicates the appropriate Chevet–Saphar tensor product space, see e.g. [54, Chapter 6]. The present, abstract MLMC results apply directly to the setting of [17], implying corresponding convergence results for multilevel Monte Carlo approximations.

An interesting topic for future research is to investigate whether the Monte Carlo convergence results, derived for injective tensor product spaces in this work, hold also with respect to stronger cross norms. In particular, the identification of tensor products of L_p-spaces as in (5.1) raises the question if it is possible to use one of the pth Chevet–Saphar tensor norms d_p or g_p (see [54, p. 135]) if the Banach space E has type p. The Chevet–Saphar norms and the Hilbert tensor norm w_2 are unified by the tensor norms $\{\alpha_{p,q}\}_{1\leq p,q\leq\infty}$ due to Lapresté: $g_p = \alpha_{p,1}$, $d_p = \alpha_{1,p}$, and $w_2 = \alpha_{2,2}$, see [16, Sections 12.5–12.8] and the references there. However, in this generality, there do not seem to be symmetric versions of these tensor norms available in the literature. A corresponding (ML)MC convergence analysis would thus have to be based on considerably different arguments.

Furthermore, this work may be extended to *sparse tensor approximations* as considered in the Hilbert space setting in [3]. Specifically, we analyzed the MLMC approach for approximating the kth moment $\mathbb{M}^k_\varepsilon[X]$ using samples of the (exact, full) tensor product $\otimes^k X_\ell$ on levels $\ell \in \{1, \dots, L\}$. The formation of this k-fold tensor product $\otimes^k X_\ell$ on level ℓ typically entails costs in work and memory of the order $\mathcal{C}_\ell \lesssim N_\ell^\gamma$ with $\gamma = \max\{\gamma_1, k\}$, assuming that γ_1 is the exponent in the asymptotic cost bound for computing one sample of X_ℓ, and that one computed sample of X_ℓ requires storage of order N_ℓ. As it is well-known in the Hilbert space case, various consistent sparse tensor product approximations allow to reduce this complexity considerably. For example, for the applications discussed in Subsections 4.1 and 4.2, the sparse tensor product approach for the MLMC approximation of kth moments proposed in [3] for Hilbert spaces can be leveraged to reduce the parameter γ in Theorem 3.25 from $\max\{\gamma_1, k\}$ to $\gamma_1 + \delta$ for some (arbitrarily small) $\delta > 0$. Yet, in this setting the error analysis of Theorem 3.24 and, consequently, also of Theorem 3.25 does not readily apply.

Beyond the MLMC estimation of kth moments $\mathbb{M}^k_\varepsilon[\xi] = \mathbb{E}[\otimes^k \xi]$, one may consider *anisotropic* k-fold correlations of the form $\mathbb{E}[\xi_1 \otimes \cdots \otimes \xi_k]$. Here, the vector-valued random variables $\xi_1, \dots, \xi_k$ entering the anisotropic, injective tensor product formation may take values in Banach spaces $E_1, \dots, E_k$ of (possibly different) types $p_1, \dots, p_k \in [1, 2]$. This rather general setting has numerous applications, and can be analyzed with the techniques in the present paper, in conjunction with the *multi-index Monte Carlo* approach from [30]. Details shall be reported elsewhere.

Funding

KK acknowledges support of the research project *Efficient spatiotemporal statistical modelling with stochastic PDEs* (with project number VI.Veni.212.021) by the talent program *Veni* which is financed by the Dutch Research Council (NWO).

Declaration of competing interest

None.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was possible in part due to the visit of both authors at the Erwin Schrödinger Institute (ESI) in Vienna, Austria, during the ESI thematic program *Computational Uncertainty Quantification* in May and June 2022. ChS acknowledges a visit to Delft Institute of Applied Mathematics in July 2022, and KK a visit to the Forschungsinstitut für Mathematik (FIM) at ETH Zürich in October 2022.

KK acknowledges helpful comments on Slepian's inequality and Lemma B.2 by Jan van Neerven and Mark Veraar; and fruitful discussions on the counterexample mentioned in Example 3.21 as well as on the application considered in Subsection 4.3 with Sonja Cox. The authors furthermore thank an anonymous reviewer for valuable comments.

Appendix A. Tensor norms of symmetric elements of Hilbert spaces

In this section we consider a real separable Hilbert space $(H, (\,\cdot\,,\,\cdot\,)_H)$ and explicitly compute the projective and injective tensor norms of Subsection 2.2 for symmetric elements in $\otimes^{2,s} H$ of the form $\sum_{j=1}^n \lambda_j\, e_j \otimes e_j$, where $e_1, \dots, e_n$ are orthonormal in H, and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$.

For this purpose, we need the notion of the *k-fold Hilbert tensor product space* $\otimes_{w_2}^k H$, which is defined as the closure of the (full) k-fold algebraic tensor product space $\otimes^k H$ with respect to the norm which is induced by the inner product

$$\Bigl(\sum_{j=1}^{n} \bigotimes_{\nu=1}^{k} x_{j,\nu}, \sum_{i=1}^{\widetilde{n}} \bigotimes_{\nu'=1}^{k} y_{i,\nu'}\Bigr)_{w_2} := \sum_{j=1}^{n} \sum_{i=1}^{\widetilde{n}} \prod_{\nu=1}^{k} (x_{j,\nu}, y_{i,\nu})_H. \tag{A.1}$$

In particular, the tensor product space $(\otimes_{w_2}^k H, (\,\cdot\,,\,\cdot\,)_{w_2})$ is again a Hilbert space.

Lemma A.1. *Assume that $(H, (\,\cdot\,,\,\cdot\,)_H)$ is a real separable Hilbert space and $(e_j)_{j\in\mathbb{N}}$ is an orthonormal basis for H. Let the injective and projective tensor norms, $\|\cdot\|_\varepsilon$ and $\|\cdot\|_\pi$, be defined on $\otimes^2 H$ as in (2.1) and (2.6), and let the symmetric injective and projective tensor norms, $\|\cdot\|_{\varepsilon_s}$ and $\|\cdot\|_{\pi_s}$, be defined on $\otimes^{2,s} H$ as in (2.3) and (2.7), respectively. Let $n \in \mathbb{N}$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Then,*

$$\Bigl\|\sum_{j=1}^{n} \lambda_j\, e_j \otimes e_j\Bigr\|_{\pi} = \Bigl\|\sum_{j=1}^{n} \lambda_j\, e_j \otimes e_j\Bigr\|_{\pi_s} = \sum_{j=1}^{n} |\lambda_j|, \tag{A.2}$$

$$\Bigl\|\sum_{j=1}^{n} \lambda_j\, e_j \otimes e_j\Bigr\|_{\varepsilon} = \Bigl\|\sum_{j=1}^{n} \lambda_j\, e_j \otimes e_j\Bigr\|_{\varepsilon_s} = \max_{1\le j\le n} |\lambda_j|. \tag{A.3}$$

Proof. We set $U_n := \sum_{j=1}^n \lambda_j\, e_j \otimes e_j \in \otimes^{2,s} H$. Then, by the definitions of the projective norms in (2.6) and (2.7), see Remark 2.1, it follows that

$$\|U_n\|_\pi \leq \|U_n\|_{\pi_s} \leq \sum_{j=1}^n |\lambda_j|\, \|e_j\|_H^2 = \sum_{j=1}^n |\lambda_j|.$$

Furthermore, for any representation $U_n = \sum_{i=1}^{\widetilde{n}} x_i \otimes y_i \in \otimes^2 H$ of U_n we find that

$$\sum_{j=1}^n |\lambda_j| = \sum_{j\in\mathbb{N}} |(U_n, e_j \otimes e_j)_{w_2}| = \sum_{j\in\mathbb{N}} \left| \left(\sum_{i=1}^{\widetilde{n}} x_i \otimes y_i,\, e_j \otimes e_j \right)_{w_2} \right|,$$

where $(\,\cdot\,,\,\cdot\,)_{w_2}$ is the inner product on the Hilbert tensor product space $\otimes^2_{w_2} H$, defined as in (A.1) for $k = 2$. Thus, by the triangle and Cauchy–Schwarz inequalities,

$$\begin{aligned}\sum_{j=1}^n |\lambda_j| &\leq \sum_{i=1}^{\widetilde{n}} \sum_{j\in\mathbb{N}} \big|(x_i \otimes y_i, e_j \otimes e_j)_{w_2}\big| = \sum_{i=1}^{\widetilde{n}} \sum_{j\in\mathbb{N}} \big|(x_i, e_j)_H\big|\big|(y_i, e_j)_H\big| \\ &\leq \sum_{i=1}^{\widetilde{n}} \left(\sum_{j\in\mathbb{N}} (x_i, e_j)_H^2 \right)^{1/2} \left(\sum_{j\in\mathbb{N}} (y_i, e_j)_H^2 \right)^{1/2} = \sum_{i=1}^{\widetilde{n}} \|x_i\|_H \|y_i\|_H.\end{aligned}$$

By taking the infimum over all representations of $U_n \in \otimes^2 H$ we obtain the reverse inequality $\|U_n\|_\pi \geq \sum_{j=1}^n |\lambda_j|$ and, since also $\|U_n\|_{\pi_s} \geq \|U_n\|_\pi$, this proves (A.2).

To show (A.3), let $j_\star \in \{1, \dots, n\}$ be an index such that $|\lambda_{j_\star}| = \max_{1\leq j\leq n} |\lambda_j|$, and recall the definitions of the injective norms from (2.1) and (2.3). Then, we find

$$\|U_n\|_\varepsilon \geq \|U_n\|_{\varepsilon_s} = \sup_{f\in B_{H'}} \left| \sum_{j=1}^n \lambda_j \langle f, e_j\rangle^2 \right| \geq \left| \sum_{j=1}^n \lambda_j (e_{j_\star}, e_j)_H^2 \right| = |\lambda_{j_\star}| = \max_{1\leq j\leq n} |\lambda_j|.$$

The reverse estimates follow again by the Cauchy–Schwarz inequality combined with the Riesz representation theorem,

$$\begin{aligned}\|U_n\|_\varepsilon &= \sup_{f_1,f_2\in B_{H'}} \left| \sum_{j=1}^n \lambda_j \langle f_1, e_j\rangle \langle f_2, e_j\rangle \right| \leq |\lambda_{j_\star}| \sup_{f_1,f_2\in B_{H'}} \sum_{j=1}^n |\langle f_1, e_j\rangle|\, |\langle f_2, e_j\rangle| \\ &= \max_{1\leq j\leq n} |\lambda_j| \sup_{v_1,v_2\in B_H} \sum_{j=1}^n |(v_1, e_j)_H|\, |(v_2, e_j)_H| \leq \max_{1\leq j\leq n} |\lambda_j|.\end{aligned}$$

Thus, $\|U_n\|_{\varepsilon_s} \leq \|U_n\|_\varepsilon \leq \max_{1\leq j\leq n} |\lambda_j|$ completing the proof of (A.3). □

Remark A.2 *(Relation of $\otimes^{2,s} H$ to self-adjoint finite-rank linear operators).* In the setting of Lemma A.1, we may associate a self-adjoint linear operator on the Hilbert space H with the element $U_n := \sum_{j=1}^n \lambda_j\, e_j \otimes e_j$ in the symmetric algebraic tensor product space $\otimes^{2,s} H$. More specifically, we can define the self-adjoint finite-rank linear operator $T_{U_n} \colon H \to H$ associated with U_n by $T_{U_n} x := \sum_{j=1}^{n} \lambda_j (x, e_j)_H \, e_j$, for every $x \in H$. Using this definition, the norm identities in (A.2) and (A.3) can be reformulated in terms of the trace-class (or nuclear) norm, $\|T_{U_n}\|_{\mathcal{L}_1(H)} := \operatorname{tr}(|T_{U_n}|)$, and the operator norm, $\|T_{U_n}\|_{\mathcal{L}(H)} := \sup_{x \in B_H} \|T_{U_n} x\|_H$, of T_{U_n} as follows:

$$\|U_n\|_\pi = \|U_n\|_{\pi_s} = \sum_{j=1}^{n} |\lambda_j| = \|T_{U_n}\|_{\mathcal{L}_1(H)},$$

$$\|U_n\|_\varepsilon = \|U_n\|_{\varepsilon_s} = \max_{1 \le j \le n} |\lambda_j| = \|T_{U_n}\|_{\mathcal{L}(H)},$$

see e.g. [57, Theorem 14.15.(1) and Theorem 8.11] for the operator norm identities.

More generally, to every element $U \in \otimes^{2,s} H$, we can associate a self-adjoint linear operator $T_U \colon H \to H$, whose action on $x \in H$ is defined by

$$(T_U x, y)_H = (U, x \otimes y)_{w_2} \quad \forall y \in H.$$

Here, the Riesz representation theorem ensures that the linear operator T_U is well-defined. The implied linear mapping $\mathcal{I} : U \mapsto T_U$ extends continuously to an isometric isomorphism between the symmetric projective tensor product space $\otimes^{2,s}_{\pi_s} H$ and the space of self-adjoint trace-class linear operators on H (respectively, between the symmetric injective tensor product space $\otimes^{2,s}_{\varepsilon_s} H$ and the space of self-adjoint compact linear operators on H).

Appendix B. A consequence of Slepian's inequality

In this section we restate the version of Slepian's inequality for finite-dimensional Gaussian processes as formulated by Fernique [18]. We subsequently use it to derive a comparison result for real-valued Gaussian processes indexed by the closed unit ball $B_{E'}$ of the dual of a real Banach space E, see Lemma B.2. This result is needed in Subsection 3.3 to prove convergence of multilevel Monte Carlo methods.

The following theorem is taken from [18, Theorem 2.1.2]. We note that a more general version of Slepian's inequality, which includes Fernique's formulation as a special case, can be found in [33, Theorem 2.8].

Theorem B.1. *Let $N \in \mathbb{N}$ and $X = (X_1, \ldots, X_N)^\top$, $Y = (Y_1, \ldots, Y_N)^\top$ be two centered Gaussian random vectors in $\mathbb{R}^N$, defined on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ with expectation $\widetilde{\mathbb{E}}$. Assume further that*

$$\forall i, j \in \{1, \ldots, N\} : \quad \widetilde{\mathbb{E}}\big[|X_i - X_j|^2\big] \le \widetilde{\mathbb{E}}\big[|Y_i - Y_j|^2\big],$$

and let $G \colon [0, \infty) \to [0, \infty)$ be convex and increasing. Then,

$$\widetilde{\mathbb{E}} G\Big(\max_{1 \le i,j \le N} |X_i - X_j|\Big) \le \widetilde{\mathbb{E}} G\Big(\max_{1 \le i,j \le N} |Y_i - Y_j|\Big).$$

Lemma B.2. *Let $M \in \mathbb{N}$ and assume that $(g_j)_{j=1}^M$ is an orthogaussian family on a complete probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ (with expectation $\widetilde{\mathbb{E}}$), and that $\Psi\colon \mathbb{R}^2 \to \mathbb{R}$ is a continuous function such that $\Psi(0,0) = 0$. Let $(x_j)_{j=1}^M, (y_j)_{j=1}^M \subset E$, and define the centered Gaussian process $\mathcal{G}_1\colon B_{E'} \times \widetilde{\Omega} \to \mathbb{R}$ on $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ indexed by the closed unit ball $B_{E'}$ in the dual space E' by*

$$\mathcal{G}_1(f) := \sum_{j=1}^M g_j \Psi\big(f(x_j), f(y_j)\big), \quad f \in B_{E'}. \tag{B.1}$$

Let $\mathcal{G}_2\colon B_{E'} \times \widetilde{\Omega} \to \mathbb{R}$ be a second centered Gaussian process on $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ such that

$$\forall f, h \in B_{E'}: \quad \widetilde{\mathbb{E}}\big[|\mathcal{G}_1(f) - \mathcal{G}_1(h)|^2\big] \leq \widetilde{\mathbb{E}}\big[|\mathcal{G}_2(f) - \mathcal{G}_2(h)|^2\big]. \tag{B.2}$$

Then, for all $q \in [1, \infty)$ we have that

$$\widetilde{\mathbb{E}}\Big[\big(\sup\nolimits_{f \in B_{E'}} |\mathcal{G}_1(f)|\big)^q\Big] \leq 2^q \, \widetilde{\mathbb{E}}\Big[\big(\sup\nolimits_{f \in B_{E'}} |\mathcal{G}_2(f)|\big)^q\Big]. \tag{B.3}$$

Proof. Set $N_0 := 0$ and $f_0 := 0 \in E'$. Given $M \in \mathbb{N}$ and $(x_j)_{j=1}^M, (y_j)_{j=1}^M \subset E$, for $f \in E'$ and $\delta \in (0, \infty)$, define the subset

$$U_\delta(f) := \Big\{ h \in E' : \max_{1 \leq j \leq M} |f(x_j) - h(x_j)| < \delta, \ \max_{1 \leq j \leq M} |f(y_j) - h(y_j)| < \delta \Big\} \subseteq E'.$$

Then, for every $f \in E'$ and all $\delta \in (0, \infty)$, the set $U_\delta(f)$ is open (more precisely, an open neighborhood of f) in E' with respect to the weak*-topology on the dual space E'. By the Banach–Alaoglu theorem the closed unit ball $B_{E'}$ is weak*-compact. Hence, for every $n \in \mathbb{N}$, the open cover

$$\bigcup_{f \in B_{E'}} U_{\frac{1}{n}}(f) \supseteq B_{E'}$$

contains a finite subcover. Iteratively, for every $n \in \mathbb{N}$, one can find an integer $N_n \in \mathbb{N}$, satisfying $N_n > N_{n-1}$, and elements $f_{N_{n-1}+1}, \ldots, f_{N_n} \in B_{E'}$ such that

$$B_{E'} \subseteq \bigcup_{\nu=1}^{N_n} U_{\frac{1}{n}}(f_\nu).$$

Note, in particular, that this definition of $f_1, \ldots, f_{N_n}$, $n \in \mathbb{N}$, implies nestedness, $(f_1, \ldots, f_{N_n}) \subseteq (f_1, \ldots, f_{N_m})$ for $n < m$.

Next, we define for every non-negative integer $\nu \in \mathbb{N}_0$ the real-valued centered Gaussian random variables $X_\nu := \mathcal{G}_1(f_\nu)$ and $Y_\nu := \mathcal{G}_2(f_\nu)$. By assumption (B.2) we then have for all $n \in \mathbb{N}$ and every $\nu, \nu' \in \{0, \ldots, N_n\}$,

$$\widetilde{\mathbb{E}}\big[|X_\nu - X_{\nu'}|^2\big] = \widetilde{\mathbb{E}}\big[|\mathcal{G}_1(f_\nu) - \mathcal{G}_1(f_{\nu'})|^2\big] \le \widetilde{\mathbb{E}}\big[|\mathcal{G}_2(f_\nu) - \mathcal{G}_2(f_{\nu'})|^2\big]$$
$$= \widetilde{\mathbb{E}}\big[|Y_\nu - Y_{\nu'}|^2\big].$$

By Fernique's version of Slepian's inequality, see Theorem B.1, applied for the convex increasing function $G(t) := t^q$, $t \ge 0$, and by using the fact that $f_0 = 0 \in E'$ implies that $\mathcal{G}_1(f_0) = 0$ holds $\widetilde{\mathbb{P}}$-a.s., we find that, for all $n \in \mathbb{N}$, $q \in [1, \infty)$,

$$\begin{aligned} \widetilde{\mathbb{E}}\Big[\big(\max\nolimits_{0\le\nu\le N_n} |\mathcal{G}_1(f_\nu)|\big)^q\Big] &\le \widetilde{\mathbb{E}}\Big[\big(\max\nolimits_{0\le\nu,\nu'\le N_n} |\mathcal{G}_1(f_\nu) - \mathcal{G}_1(f_{\nu'})|\big)^q\Big] \\ &= \widetilde{\mathbb{E}}\Big[\big(\max\nolimits_{0\le\nu,\nu'\le N_n} |X_\nu - X_{\nu'}|\big)^q\Big] \le \widetilde{\mathbb{E}}\Big[\big(\max\nolimits_{0\le\nu,\nu'\le N_n} |Y_\nu - Y_{\nu'}|\big)^q\Big] \\ &\le 2^q\, \widetilde{\mathbb{E}}\Big[\big(\max\nolimits_{0\le\nu\le N_n} |Y_\nu|\big)^q\Big] \le 2^q\, \widetilde{\mathbb{E}}\Big[\big(\sup\nolimits_{f\in B_{E'}} |\mathcal{G}_2(f)|\big)^q\Big]. \end{aligned}$$

To derive (B.3), it remains to prove that $\lim_{n\to\infty} \|S_n\|_{L_q(\widetilde{\Omega};\mathbb{R})} = \|S_*\|_{L_q(\widetilde{\Omega};\mathbb{R})}$, where

$$S_n := \max_{0\le\nu\le N_n} |X_\nu| = \max_{0\le\nu\le N_n} |\mathcal{G}_1(f_\nu)| \qquad \text{and} \qquad S_* := \sup_{f\in B_{E'}} |\mathcal{G}_1(f)|.$$

By the assumptions on the process $\mathcal{G}_1$ in (B.1), there exists a set $\widetilde{\Omega}_0 \in \widetilde{\mathcal{A}}$ with $\widetilde{\mathbb{P}}(\widetilde{\Omega}_0) = 0$ such that $\overline{g}(\widetilde{\omega}) := 1 + \max_{1\le j\le M} |g_j(\widetilde{\omega})| < \infty$ for all $\widetilde{\omega} \in \widetilde{\Omega} \setminus \widetilde{\Omega}_0$. Fix $\widetilde{\omega} \in \widetilde{\Omega} \setminus \widetilde{\Omega}_0$ and $\epsilon \in (0, 1)$. Then, there exists $f^\epsilon = f^\epsilon(\widetilde{\omega}) \in B_{E'}$ such that

$$\sup_{f\in B_{E'}} |\mathcal{G}_1(f)(\widetilde{\omega})| \le |\mathcal{G}_1(f^\epsilon)(\widetilde{\omega})| + \frac{\epsilon}{2}.$$

In addition, there exists $\delta_\epsilon(\widetilde{\omega}) \in (0, \infty)$, such that the implication

$$h \in U_{\delta_\epsilon(\widetilde{\omega})}(f^\epsilon) \quad \Longrightarrow \quad |\mathcal{G}_1(f^\epsilon)(\widetilde{\omega}) - \mathcal{G}_1(h)(\widetilde{\omega})| < \frac{\epsilon}{2} \tag{B.4}$$

holds. Indeed, by continuity of $\Psi\colon \mathbb{R}^2 \to \mathbb{R}$ we may choose $\delta_\epsilon(\widetilde{\omega}) \in (0, \infty)$ such that

$$\begin{aligned} &\max_{1\le j\le M} \big\{\max\{|f^\epsilon(x_j) - h(x_j)|, |f^\epsilon(y_j) - h(y_j)|\}\big\} < \delta_\epsilon(\widetilde{\omega}) \\ &\qquad \Longrightarrow \quad \max_{1\le j\le M} \big|\Psi\big(f^\epsilon(x_j), f^\epsilon(y_j)\big) - \Psi\big(h(x_j), h(y_j)\big)\big| < \frac{\epsilon}{2}\, M^{-1}\overline{g}(\widetilde{\omega})^{-1}. \end{aligned}$$

Furthermore, by definition of the sequences $(N_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ and $(f_\nu)_{\nu\in\mathbb{N}} \subseteq B_{E'}$, there exist integers $n_\epsilon = n_\epsilon(\widetilde{\omega}) \in \mathbb{N}$ and $\nu_\star = \nu_\star(\widetilde{\omega}) \in \{1, \ldots, N_{n_\epsilon}\}$ such that

$$f^\epsilon \in U_{\delta_\epsilon(\widetilde{\omega})}(f_{\nu_\star}).$$

By combining this observation with (B.4) we conclude that

$$\big||\mathcal{G}_1(f^\epsilon)(\widetilde{\omega})| - |\mathcal{G}_1(f_{\nu_\star})(\widetilde{\omega})|\big| \le |\mathcal{G}_1(f^\epsilon)(\widetilde{\omega}) - \mathcal{G}_1(f_{\nu_\star})(\widetilde{\omega})| < \frac{\epsilon}{2},$$

and

$$S_*(\widetilde{\omega}) - S_{n_\epsilon}(\widetilde{\omega}) = \sup_{f \in B_{E'}} |\mathcal{G}_1(f)(\widetilde{\omega})| - \max_{0 \leq \nu \leq N_{n_\epsilon}} |\mathcal{G}_1(f_\nu)(\widetilde{\omega})|$$
$$\leq \sup_{f \in B_{E'}} |\mathcal{G}_1(f)(\widetilde{\omega})| - |\mathcal{G}_1(f_{\nu_\star})(\widetilde{\omega})| \leq |\mathcal{G}_1(f^\epsilon)(\widetilde{\omega})| - |\mathcal{G}_1(f_{\nu_\star})(\widetilde{\omega})| + \frac{\epsilon}{2} < \epsilon$$

follows. This shows that, for almost all $\widetilde{\omega} \in \widetilde{\Omega}$,

$$S_*(\widetilde{\omega}) = \sup_{f \in B_{E'}} |\mathcal{G}_1(f)(\widetilde{\omega})| = \sup_{n \in \mathbb{N}} \max_{0 \leq \nu \leq N_n} |\mathcal{G}_1(f_\nu)(\widetilde{\omega})| = \lim_{n \to \infty} S_n(\widetilde{\omega}).$$

Since the non-negative random variables $(S_n)_{n \in \mathbb{N}}$ are non-decreasing in $n \in \mathbb{N}$, $\widetilde{\mathbb{P}}$-a.s., the $L_q(\widetilde{\Omega}; \mathbb{R})$-convergence $\lim_{n \to \infty} \|S_n\|_{L_q(\widetilde{\Omega};\mathbb{R})} = \|S_*\|_{L_q(\widetilde{\Omega};\mathbb{R})}$ follows from the monotone convergence theorem. □

References

[1] I. Babuška, W.C. Rheinboldt, A posteriori error analysis of finite element solutions for one-dimensional problems, SIAM J. Numer. Anal. 18 (1981) 565–589.

[2] A. Barth, A. Lang, Ch. Schwab, Multilevel Monte Carlo method for parabolic stochastic partial differential equations, BIT 53 (2013) 3–27.

[3] A. Barth, Ch. Schwab, N. Zollinger, Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients, Numer. Math. 119 (2011) 123–161.

[4] S. Becker, B. Gess, A. Jentzen, P.E. Kloeden, Strong convergence rates for explicit space-time discrete numerical approximations of stochastic Allen-Cahn equations, Stoch. Partial Differ. Equ. Anal. Comput. 11 (2023) 211–268.

[5] C. Bierig, A. Chernov, Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems, Numer. Math. 130 (2015) 579–613.

[6] C. Bierig, A. Chernov, Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method, Stoch. Partial Differ. Equ. Anal. Comput. 4 (2016) 3–40.

[7] V.I. Bogachev, Gaussian Measures, Mathematical Surveys and Monographs, vol. 62, American Mathematical Society, Providence, RI, 1998.

[8] S.C. Brenner, L.R. Scott, The Mathematical Theory of Finite Element Methods, third ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008.

[9] J. Charrier, Strong and weak error estimates for elliptic partial differential equations with random coefficients, SIAM J. Numer. Anal. 50 (2012) 216–246.

[10] J. Charrier, R. Scheichl, A.L. Teckentrup, Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods, SIAM J. Numer. Anal. 51 (2013) 322–352.

[11] A. Chernov, Ch. Schwab, First order k-th moment finite element analysis of nonlinear operator equations with stochastic data, Math. Comput. 82 (2013) 1859–1888.

[12] K.A. Cliffe, M.B. Giles, R. Scheichl, A.L. Teckentrup, Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients, Comput. Vis. Sci. 14 (2011) 3–15.

[13] N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, R. Tempone, A continuation multilevel Monte Carlo algorithm, BIT 55 (2015) 399–432.

[14] S. Cox, M. Hutzenthaler, A. Jentzen, J. van Neerven, T. Welti, Convergence in Hölder norms with applications to Monte Carlo methods in infinite dimensions, IMA J. Numer. Anal. 41 (2021) 493–548.

[15] S.G. Cox, K. Kirchner, Regularity and convergence analysis in Sobolev and Hölder spaces for generalized Whittle-Matérn fields, Numer. Math. 146 (2020) 819–873.

[16] A. Defant, K. Floret, Tensor Norms and Operator Ideals, North-Holland Mathematics Studies, vol. 176, North-Holland Publishing Co., Amsterdam, 1993.

[17] E. Feireisl, M. Lukáčová-Medvid'ová, B. She, Y. Yuan, Convergence and error analysis of compressible fluid flows with random data: Monte Carlo method, Math. Models Methods Appl. Sci. 32 (2022) 2887–2925.

[18] X. Fernique, Regularité des trajectoires des fonctions aléatoires gaussiennes, in: École d'Été de Probabilités de Saint-Flour, IV-1974, in: Lecture Notes in Math., vol. 480, 1975, pp. 1–96.

[19] K. Floret, Natural norms on symmetric tensor products of normed spaces, in: Proceedings of the Second International Workshop on Functional Analysis (Trier, 1997), vol. 17, 1997, pp. 153–188, 1999.

[20] K. Floret, The extension theorem for norms on symmetric tensor products of normed spaces, in: Recent Progress in Functional Analysis, Valencia, 2000, in: North-Holland Math. Stud., vol. 189, North-Holland, Amsterdam, 2001, pp. 225–237.

[21] M.B. Giles, Improved multilevel Monte Carlo convergence using the Milstein scheme, in: Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer, Berlin, 2008, pp. 343–358.

[22] M.B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. 56 (2008) 607–617.

[23] M.B. Giles, K. Debrabant, A. Rössler, Analysis of multilevel Monte Carlo path simulation using the Milstein discretisation, Discrete Contin. Dyn. Syst., Ser. B 24 (2019) 3881–3903.

[24] M.B. Giles, C. Reisinger, Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance, SIAM J. Financ. Math. 3 (2012) 572–592.

[25] M.B. Giles, L. Szpruch, Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation, Ann. Appl. Probab. 24 (2014) 1585–1620.

[26] C.J. Gittelson, J. Könnö, Ch. Schwab, R. Stenberg, The multi-level Monte Carlo finite element method for a stochastic Brinkman problem, Numer. Math. 125 (2013) 347–386.

[27] I.G. Graham, R. Scheichl, E. Ullmann, Mixed finite element analysis of lognormal diffusion and multilevel Monte Carlo methods, Stoch. Partial Differ. Equ. Anal. Comput. 4 (2016) 41–75.

[28] P. Grisvard, Elliptic Problems in Nonsmooth Domains, Classics in Applied Mathematics, vol. 69, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[29] I. Gyöngy, M. Rásonyi, A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients, Stoch. Process. Appl. 121 (2011) 2189–2200.

[30] A.-L. Haji-Ali, F. Nobile, R. Tempone, Multi-index Monte Carlo: when sparsity meets sampling, Numer. Math. 132 (2016) 767–806.

[31] A.-L. Haji-Ali, F. Nobile, E. von Schwerin, R. Tempone, Optimization of mesh hierarchies in multilevel Monte Carlo samplers, Stoch. Partial Differ. Equ. Anal. Comput. 4 (2016) 76–112.

[32] S. Heinrich, Multilevel Monte Carlo methods, in: Large-Scale Scientific Computing, in: Lecture Notes in Comput. Sci., vol. 2179, Springer-Verlag, 2001, pp. 58–67.

[33] J. Hoffmann-Jørgensen, Slepian's inequality, modularity and integral orderings, in: High Dimensional Probability VI, in: Progr. Probab., vol. 66, Birkhäuser/Springer, Basel, 2013, pp. 19–53.

[34] M. Hutzenthaler, A. Jentzen, P.E. Kloeden, Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients, Ann. Appl. Probab. 22 (2012) 1611–1641.

[35] T. Hytönen, J. van Neerven, M. Veraar, L. Weis, Analysis in Banach Spaces. Vol. I. Martingales and Littlewood-Paley Theory, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics, vol. 63, Springer, Cham, 2016.

[36] T. Hytönen, J. van Neerven, M. Veraar, L. Weis, Analysis in Banach Spaces. Vol. II. Probabilistic Methods and Operator Theory, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics, vol. 67, Springer, Cham, 2017.

[37] S. Janson, S. Kaijser, Higher moments of Banach space valued random variables, Mem. Am. Math. Soc. 238 (2015), vii+110.

[38] C.M. Jarque, A.K. Bera, A test for normality of observations and regression residuals, Int. Stat. Rev. 55 (1987) 163–172.

[39] J.-P. Kahane, Some Random Series of Functions, second ed., Cambridge Studies in Advanced Mathematics, vol. 5, Cambridge University Press, Cambridge, 1985.

[40] K. Kirchner, Numerical methods for the deterministic second moment equation of parabolic stochastic PDEs, Math. Comput. 89 (2020) 2801–2845.

[41] K. Kirchner, A. Lang, S. Larsson, Covariance structure of parabolic stochastic partial differential equations with multiplicative Lévy noise, J. Differ. Equ. 262 (2017) 5896–5927.

[42] U. Koley, N.H. Risebro, Ch. Schwab, F. Weber, A multilevel Monte Carlo finite difference method for random scalar degenerate convection-diffusion equations, J. Hyperbolic Differ. Equ. 14 (2017) 415–454.

[43] M. Kovács, A. Lang, A. Petersson, Approximation of SPDE covariance operators by finite elements: a semigroup approach, IMA J. Numer. Anal. 43 (2023) 1324–1357.

[44] R. Kruse, Strong and Weak Approximation of Semilinear Stochastic Evolution Equations, Lecture Notes in Mathematics, vol. 2093, Springer, Cham, 2014.
[45] S. Kwapień, Isomorphic characterizations of inner product spaces by orthogonal series with vector valued coefficients, Stud. Math. 44 (1972) 583–595.
[46] A. Lang, S. Larsson, Ch. Schwab, Covariance structure of parabolic stochastic partial differential equations, Stoch. Partial Differ. Equ. Anal. Comput. 1 (2013) 351–364.
[47] M. Ledoux, M. Talagrand, Probability in Banach Spaces: Isoperimetry and Processes, Classics in Mathematics, Springer-Verlag, Berlin, 2011, Reprint of the 1991 edition.
[48] P.-L. Lions, Mathematical Topics in Fluid Mechanics. vol. 2, Compressible Models, Oxford Lecture Series in Mathematics and Its Applications, vol. 10, The Clarendon Press, Oxford University Press, New York, 1998.
[49] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.
[50] S. Mishra, N.H. Risebro, Ch. Schwab, S. Tokareva, Numerical solution of scalar conservation laws with random flux functions, SIAM/ASA J. Uncertain. Quantificat. 4 (2016) 552–591.
[51] S. Mishra, Ch. Schwab, Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data, Math. Comput. 81 (2012) 1979–2018.
[52] S. Mishra, Ch. Schwab, Monte-Carlo finite-volume methods in uncertainty quantification for hyperbolic conservation laws, in: Uncertainty Quantification for Hyperbolic and Kinetic Equations, in: SEMA SIMAI Springer Ser., vol. 14, Springer, Cham, 2017, pp. 231–277.
[53] T. Müller-Gronbach, L. Yaroslavtseva, A strong order 3/4 method for SDEs with discontinuous drift coefficient, IMA J. Numer. Anal. 42 (2022) 229–259.
[54] R.A. Ryan, Introduction to Tensor Products of Banach Spaces, Springer Monographs in Mathematics, Springer-Verlag London, Ltd., London, 2002.
[55] C.G. Simader, On Dirichlet's Boundary Value Problem: L^p-Theory Based on a Generalization of Gårding's Inequality, Lecture Notes in Mathematics, vol. 268, Springer-Verlag, Berlin-New York, 1972.
[56] A.L. Teckentrup, R. Scheichl, M.B. Giles, E. Ullmann, Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients, Numer. Math. 125 (2013) 569–600.
[57] J. van Neerven, Functional Analysis, Cambridge Studies in Advanced Mathematics, vol. 201, Cambridge University Press, Cambridge, 2022.
[58] Y. Xia, M.B. Giles, Multilevel path simulation for jump-diffusion SDEs, in: Monte Carlo and Quasi-Monte Carlo Methods 2010, in: Springer Proc. Math. Stat., vol. 23, Springer, Heidelberg, 2012, pp. 695–708.