



Application of randomized quadrature formulas to the finite element method for elliptic equations

Raphael Kruse¹ · Nick Polydorides² · Yue Wu³

Received: 11 March 2025 / Accepted: 8 April 2026
© The Author(s) 2026

Abstract

The implementation of the finite element method for linear elliptic equations requires to assemble the stiffness matrix and the load vector. In general, the entries of this matrix-vector system are not known explicitly but need to be approximated by quadrature rules. If the coefficient functions of the differential operator or the forcing term are irregular, then standard quadrature formulas, such as the barycentric quadrature rule, may not be reliable. In this paper we investigate the application of two randomized quadrature formulas to the finite element method for such elliptic boundary value problems with irregular coefficient functions. We give a detailed error analysis of these methods, discuss their implementation, and demonstrate their capabilities in several numerical experiments.

Keywords Finite element method · Monte Carlo methods · Quadrature formulas · Elliptic partial differential equations

Mathematics Subject Classification 65C05 · 65D32 · 65N30

✉ Yue Wu
yue.wu@strath.ac.uk

Raphael Kruse
raphael.kruse@mathematik.uni-halle.de

Nick Polydorides
n.polydorides@ed.ac.uk

¹ Institut für Mathematik, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale) 06099, Germany

² School of Engineering, University of Edinburgh, Edinburgh EH9 3FB, UK

³ Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK

1 Introduction

Let $\mathcal{D} \subset \mathbb{R}^2$ be a convex, bounded, and polygonal domain. We consider a linear elliptic boundary value problem of the following form: Find a mapping $u : \mathcal{D} \rightarrow \mathbb{R}$ such that

$$\begin{cases} -\operatorname{div}(\sigma \nabla u) = f, & \text{in } \mathcal{D}, \\ u = 0, & \text{on } \partial \mathcal{D}, \end{cases} \quad (1.1)$$

where $\sigma, f : \mathcal{D} \rightarrow \mathbb{R}$ are given coefficient functions with $\sigma(x) \geq \sigma_0 > 0$ for all $x \in \mathcal{D}$. Provided σ is globally bounded and f is square-integrable, it is well-known that (1.1) admits a unique solution $u \in H_0^1(\mathcal{D})$ in the weak sense satisfying

$$\int_{\mathcal{D}} \sigma(x) \nabla u(x) \cdot \nabla v(x) \, dx = \int_{\mathcal{D}} f(x) v(x) \, dx \quad (1.2)$$

for all $v \in H_0^1(\mathcal{D})$. Here, we denote by $H_0^1(\mathcal{D})$ the Sobolev space of weakly differentiable and square-integrable functions which (in some sense) satisfy the homogeneous Dirichlet boundary condition. In Section 2 we provide more details on the function spaces used throughout this paper. We also refer, for instance, to [5, Chapters 8–9] or [11, Chapter 6] for an introduction to the variational formulation of elliptic boundary value problems of the form (1.1).

Elliptic equations such as (1.1) appear in many applications, e.g., in mechanical engineering and physics. It is also an intensively studied problem to introduce the Galerkin finite element method as found in many text books in numerical analysis, e.g. [4, 27, 28], [36]. In the same spirit, we use (1.1) as a model problem to demonstrate the applicability of randomized quadrature formulas to the finite element method.

To this end, we consider a family $(\mathcal{T}_h)_{h \in (0,1]}$ of finite subdivisions of the polygonal domain $\mathcal{D} \subset \mathbb{R}^2$ into triangles. Hereby, the parameter $h \in (0, 1]$ denotes the maximal edge length of the elements in \mathcal{T}_h . For every partition \mathcal{T}_h we define $S_h \subset H_0^1(\mathcal{D})$ as the associated finite element space consisting of piecewise linear functions.

Then, we obtain an approximation of the exact solution to the boundary value problem (1.1) by solving the following finite dimensional problem: For $h \in (0, 1]$ find $u_h \in S_h$ satisfying

$$\int_{\mathcal{D}} \sigma(x) \nabla u_h(x) \cdot \nabla v_h(x) \, dx = \int_{\mathcal{D}} f(x) v_h(x) \, dx \quad (1.3)$$

for all $v_h \in S_h$. For the practical computation of the approximation $u_h \in S_h$, it is then convenient to rewrite (1.3) as a system of linear equations. More precisely, let $(\varphi_j)_{j=1}^{N_h}$ be a basis of S_h , where $N_h = \dim(S_h)$ denotes the number of degrees of freedom. Then, we have the representation

$$u_h = \sum_{j=1}^{N_h} u_j \varphi_j,$$

where the entries of the vector $\mathbf{u} = [u_1, \dots, u_{N_h}]^\top \in \mathbb{R}^{N_h}$ are yet to be determined. After inserting this representation of u_h into the finite dimensional problem (1.3) and by testing with all basis functions $(\varphi_j)_{j=1}^{N_h}$ we arrive at a system of linear equations. In matrix-vector form this system is written as

$$A_h \mathbf{u} = f_h, \quad (1.4)$$

where the *stiffness matrix* $A_h \in \mathbb{R}^{N_h \times N_h}$ is given by

$$[A_h]_{i,j} = \int_{\mathcal{D}} \sigma(x) \nabla \varphi_i(x) \cdot \nabla \varphi_j(x) \, dx \quad (1.5)$$

for all $i, j \in \{1, \dots, N_h\}$. Moreover, the *load vector* $f_h \in \mathbb{R}^{N_h}$ has the entries

$$[f_h]_i = \int_{\mathcal{D}} f(x) \varphi_i(x) \, dx, \quad i \in \{1, \dots, N_h\}. \quad (1.6)$$

If, on the one hand, the entries of A_h and f_h are known explicitly, it is straight-forward to use standard solvers for the linear system (1.4) in order to determine $\mathbf{u} \in \mathbb{R}^{N_h}$ and, hence, $u_h \in S_h$ numerically. For instance, we refer to the monograph [17] for an overview of suitable solvers.

On the other hand, for general $\sigma \in L^\infty(\mathcal{D})$ and $f \in L^2(\mathcal{D})$, the entries of the stiffness matrix and the load vector are often not computable explicitly. Such irregular coefficients often appear in problems in uncertainty quantification to model incomplete knowledge of the problem parameters. See [2] and the references therein. In the literature, the reader is advised to approximate the entries by suitable quadrature formulas. For instance, we refer to [28, Section 5.6] and [36, Section 4.3].

However, standard methods for numerical integration, such as the trapezoidal sum, require point evaluations of the coefficient functions σ and f . Therefore, these quadrature formulas are, in general, only applicable if additional smoothness requirements, such as continuity, are imposed on σ and f . The purpose of this paper is to show that this problem can be circumvented if we approximate the entries of A_h and f_h by *randomized quadrature formulas*. As it will turn out, these quadrature formulas do not require the continuity of f and σ .

Before we give a more detailed outline of the content of this paper, let us mention that we consider randomized quadrature formulas of a form that has originally been introduced by S. Haber in [14–16]. His important observation was that the accuracy of the standard Monte Carlo method can be increased drastically, if the random sampling points are distributed more evenly over the integration domain. More precisely, he proposed to place the random sampling points in disjoint subdomains whose volumes decay asymptotically with the number of samples. If the integrand possesses more regularity than being merely square-integrable this approach reduces the variance of the randomized quadrature formula significantly. In particular, one often observes a higher order of convergence compared to standard Monte Carlo estimators or purely deterministic methods. For more details on this line of arguments we also refer to the proof of Lemma 1 further below. Moreover, related results are found in [6, 30].

More recently, it has been shown that such randomized quadrature formulas are also applicable to the numerical approximation of ordinary differential equations with time-irregular coefficient functions. We refer, for instance, to [8, 19, 21, 25, 34, 35] for results on randomized one-step methods. Further, these methods have also been applied for the *temporal discretization* of evolution equations in infinite dimensions, see [10, 20], and of stochastic differential equations, see [26, 32].

Besides [18], where the information based complexity of randomized algorithms for elliptic partial differential equations has been investigated, it appears that the application of randomized quadrature formulas to the *spatial discretization* of boundary value problems is not well-studied yet.

In this paper, we first consider a stratified Monte Carlo estimator in the spirit of [14]. More precisely, the estimator defined in (3.19) below, is based on an admissible triangulation \mathcal{T}_h of \mathcal{D} and exactly one uniformly distributed random point on each triangle of the triangulation. We show in Section 3 that this estimator gives approximations of the entries in the stiffness matrix and the load vector, which are convergent at least with order 1 with respect to the root-mean-square norm. Under slightly increased regularity assumptions, such as $f \in L^p(\mathcal{D})$ with $p \in (2, \infty]$ and $\sigma \in W^{s,q}(\mathcal{D})$ with $s \in (0, 1]$, $q \in (2, \infty]$, we also show that the resulting randomized finite element solution u_h^{MC} converges to the exact solution $u \in H_0^1(\mathcal{D})$. The precise error estimate is given in Theorem 2.

In Section 4, we propose an importance sampling estimator for the approximation of the load vector. Hereby, the random points are placed according to a non-uniform distribution, whose probability density function is proportional to the basis functions of the finite element space. The section also contains a detailed analysis of the error with respect to the norms in $L^2(\mathcal{D})$ and $H_0^1(\mathcal{D})$, where we purely focus on the associated finite element problem for the Poisson equation (4.41), i.e. Equation (1.1) with $\sigma \equiv 1$. These results are stated in Theorem 4 and Theorem 5.

In Section 5 we discuss the implementation of the randomized quadrature formulas. Essentially, this is achieved by a transformation to a reference triangle, typically the 2-simplex, and a general rejection algorithm. Finally, we report on some numerical experiments in Section 6.

2 Notation and preliminaries

In this section, we fix some notation and introduce several function spaces, which are used throughout this paper. We also revisit the variational formulation of the boundary value problem (1.1) and its approximation by the finite element method. The section also contains a brief overview of some terminology from probability.

By \mathbb{N} we denote the set of all positive integers, while $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. As usual, the set \mathbb{R} consists of all real numbers. By $|\cdot|$ we denote the Euclidean norm on the Euclidean space \mathbb{R}^d for any $d \in \mathbb{N}$. In particular, if $d = 1$ then $|\cdot|$ coincides with taking the absolute value.

Throughout this paper we often use C as a generic constant, which may vary from appearance to appearance. However, C is not allowed to depend on numerical parameters such as $h \in (0, 1]$.

Next, let us introduce some function spaces. Throughout this paper, we assume that $\mathcal{D} \subset \mathbb{R}^2$ is a bounded, convex and polygonal domain. By $L^p(\mathcal{D})$, $p \in [1, \infty]$, we denote the Banach space of (equivalence classes of) p -fold Lebesgue integrable functions, which is endowed with the norm

$$\|g\|_{L^p(\mathcal{D})} = \left(\int_{\mathcal{D}} |g(x)|^p \, dx \right)^{\frac{1}{p}} \quad \text{for } p \in [1, \infty),$$

$$\|g\|_{L^\infty(\mathcal{D})} = \operatorname{ess\,sup}_{x \in \mathcal{D}} |g(x)|.$$

As it is customary, we do not distinguish notationally between functions and their equivalence classes.

An important example of an element in $L^p(\mathcal{D})$ for any value of $p \in [1, \infty]$ is the indicator function of a measurable set $B \subseteq \mathcal{D}$ denoted by $\mathbb{1}_B$. This function fulfills $\mathbb{1}_B(x) = 1$ if $x \in B$, else $\mathbb{1}_B(x) = 0$.

Moreover, we denote by $W^{k,p}(\mathcal{D}) \subset L^p(\mathcal{D})$, $p \in [1, \infty]$, $k \in \mathbb{N}$, the Sobolev space with differentiation index k . To be more precise, $W^{k,p}(\mathcal{D})$ consists of all p -fold integrable functions that are k -times partially differentiable in the weak sense and whose derivatives are also p -fold integrable. If $W^{k,p}(\mathcal{D})$ is endowed with the norm

$$\|g\|_{W^{k,p}(\mathcal{D})} = \left(\sum_{\alpha \in \mathbb{N}_0^2, |\alpha| \leq k} \|\partial^\alpha g\|_{L^p(\mathcal{D})}^p \right)^{\frac{1}{p}} \quad \text{for } p \in [1, \infty),$$

$$\|g\|_{W^{k,\infty}(\mathcal{D})} = \sum_{\alpha \in \mathbb{N}_0^2, |\alpha| \leq k} \|\partial^\alpha g\|_{L^\infty(\mathcal{D})},$$

then it is also a Banach space. Here we make use of the standard multi-index notation for partial derivatives, that is, for $\alpha \in \mathbb{N}_0^2$ we define $|\alpha| = \alpha_1 + \alpha_2$ and

$$\partial^\alpha g := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}} g.$$

Further, if $p = 2$ then $L^2(\mathcal{D})$ and $H^k(\mathcal{D}) := W^{k,2}(\mathcal{D})$ are Hilbert spaces. The inner products are denoted by $(\cdot, \cdot)_{L^2(\mathcal{D})}$ and $(\cdot, \cdot)_{H^k(\mathcal{D})}$, respectively.

In order to incorporate homogeneous Dirichlet boundary conditions, we also introduce the space $H_0^1(\mathcal{D})$, which is defined as the closure of the set of all infinitely often differentiable functions with compact support in \mathcal{D} with respect to the norm in $H^1(\mathcal{D})$, that is

$$H_0^1(\mathcal{D}) := \overline{C_c^\infty(\mathcal{D})}^{\|\cdot\|_{H^1(\mathcal{D})}}.$$

It is well-known that the standard $H^1(\mathcal{D})$ -norm and the semi-norm

$$|g|_{H^1(\mathcal{D})} = \left(\sum_{i=1}^2 \left\| \frac{\partial}{\partial x_i} g \right\|_{L^2(\mathcal{D})}^2 \right)^{\frac{1}{2}} = \left(\int_{\mathcal{D}} |\nabla g|^2 \, dx \right)^{\frac{1}{2}}$$

are equivalent on $H_0^1(\mathcal{D})$. In particular, the space $(H_0^1(\mathcal{D}), |\cdot|_{H^1(\mathcal{D})}, (\cdot, \cdot)_{H_0^1(\mathcal{D})})$ is a separable Hilbert space. For a detailed introduction to Sobolev spaces we refer the reader, for instance, to [11, Chapter 5].

For a domain $\mathcal{D} \subset \mathbb{R}^2$, $p \in [1, \infty)$, and $s \in (0, 1)$ the Sobolev–Slobodeckij norm $\|\cdot\|_{W^{s,p}(\mathcal{D})}$ is given by

$$\|g\|_{W^{s,p}(\mathcal{D})} = \left(\|g\|_{L^p(\mathcal{D})}^p + \int_{\mathcal{D}} \int_{\mathcal{D}} \frac{|g(x_1) - g(x_2)|^p}{|x_1 - x_2|^{2+sp}} dx_2 dx_1 \right)^{\frac{1}{p}}. \quad (2.7)$$

Then, the fractional order Sobolev space $W^{s,p}(\mathcal{D})$ consists of all $g \in L^p(\mathcal{D})$ satisfying $\|g\|_{W^{s,p}(\mathcal{D})} < \infty$. By $|\cdot|_{W^{s,p}(\mathcal{D})}$ we denote the corresponding semi-norm, which only consists of the double integral part in (2.7). Further details on these spaces are found in [9].

Next, we revisit the variational formulation of the boundary value problem (1.1). If $\sigma \in L^\infty(\mathcal{D})$, $\sigma(x) \geq \sigma_0 > 0$ for almost every $x \in \mathcal{D}$, and $f \in L^2(\mathcal{D})$, then it is well-known that the bilinear form $a: H_0^1(\mathcal{D}) \times H_0^1(\mathcal{D}) \rightarrow \mathbb{R}$ and the linear functional $F: H_0^1(\mathcal{D}) \rightarrow \mathbb{R}$ given by

$$a(u, v) := \int_{\mathcal{D}} \sigma(x) \nabla u(x) \cdot \nabla v(x) dx, \quad (2.8)$$

$$F(v) := \int_{\mathcal{D}} f(x)v(x) dx \quad (2.9)$$

for all $u, v \in H_0^1(\mathcal{D})$ are well-defined. Moreover, a is strongly positive and bounded, that is, it holds

$$a(v, v) \geq \sigma_0 |v|_{H^1(\mathcal{D})}^2, \quad (2.10)$$

$$|a(u, v)| \leq \|\sigma\|_{L^\infty(\mathcal{D})} |u|_{H^1(\mathcal{D})} |v|_{H^1(\mathcal{D})} \quad (2.11)$$

for all $u, v \in H_0^1(\mathcal{D})$. Further, F is a bounded linear functional.

Therefore, the lemma of Lax–Milgram, cf. [11, Chapter 6], is applicable and ensures the existence of a unique weak solution $u \in H_0^1(\mathcal{D})$ satisfying

$$a(u, v) = F(v) \quad \text{for all } v \in H_0^1(\mathcal{D}). \quad (2.12)$$

Observe that (2.12) coincides with (1.2).

For the error analysis in Section 3 and Section 4, it will be necessary to impose the following additional regularity condition on the exact solution.

Assumption 1 The variational problem (2.12) admits a uniquely determined strong solution, i.e., the unique weak solution u to (2.12) is an element of $H_0^1(\mathcal{D}) \cap H^2(\mathcal{D})$.

We refer, for instance, to [13, Theorem 3.2.1.2], which gives sufficient conditions for the existence of a strong solution. For example, if \mathcal{D} is a convex, bounded and open subset of \mathbb{R}^2 and if $\sigma \in L^\infty(\mathcal{D})$ has a globally Lipschitz continuous extension on $\overline{\mathcal{D}}$, then Assumption 1 is satisfied for every $f \in L^2(\mathcal{D})$.

Next, we briefly review the finite element method for problem (1.1). To this end, let $(\mathcal{T}_h)_{h \in (0,1]}$ be a family of admissible triangulations of \mathcal{D} . More precisely, for every $h \in (0, 1]$ it holds that each triangle $T \in \mathcal{T}_h$ is an open subset of \mathcal{D} satisfying

$$\bigcup_{T \in \mathcal{T}_h} \bar{T} = \bar{\mathcal{D}} \quad \text{and} \quad T \cap T' = \emptyset, \quad \text{for all } T, T' \in \mathcal{T}_h, T \neq T'.$$

Further, it is assumed that no vertex of any triangle lies in the interior of an edge of any other triangle of the triangulation, cf. [4, Definition 3.3.11]. Typically, the parameter $h \in (0, 1]$ denotes the maximal edge length of all triangles in \mathcal{T}_h . Moreover, the area of a triangle T is denoted by $|T|$.

As usual, we define the finite element space S_h associated to a triangulation \mathcal{T}_h , $h \in (0, 1]$, by

$$S_h = \{v_h \in C(\bar{\mathcal{D}}) : v_h = 0 \text{ on } \partial\mathcal{D}, v_h|_T \in \Pi_1 \forall T \in \mathcal{T}_h\}.$$

Hereby, the set Π_1 consists of all polynomials up to degree 1. The finite element space S_h is finite dimensional and $N_h = \dim(S_h)$ is called the *number of degrees of freedom*. It coincides with the number of interior nodes $(z_i)_{i=1}^{N_h}$ of the triangulation. By $(\varphi_j)_{j=1}^{N_h} \subset S_h$ we denote the standard Lagrange basis of S_h determined by $\varphi_j(z_i) = \delta_{i,j}$ for all $i, j = 1, \dots, N_h$. Further details on the construction of finite element spaces are found, e.g., in [4, Chapter 3] or [28, Chapter 5].

For the error analysis in Section 3 and Section 4 we have to impose the following additional condition on the family of triangulations.

Assumption 2 We assume that $(\mathcal{T}_h)_{h \in (0,1]}$ is a family of admissible and quasi-uniform triangulations. In particular, the interior angles of each triangle in \mathcal{T}_h are bounded from below by a positive constant, independently of h . In addition, there exists $c \in (0, \infty)$ such that for every $h \in (0, 1]$ and $T \in \mathcal{T}_h$ it holds that $|T| \geq ch^2$.

The assumption enables us to make use of a maximum norm estimate for functions from the finite element space S_h , which we cite from [37, Lemma 6.4]: If Assumption 2 is satisfied then there exists $C \in (0, \infty)$, independently of $h \in (0, 1]$, such that

$$\|v_h\|_{L^\infty(\mathcal{D})} \leq C \ell_h^{\frac{1}{2}} |v_h|_{H^1(\mathcal{D})} \tag{2.13}$$

for every $v_h \in S_h$, where $\ell_h = \max(1, \log(1/h))$.

Further, we recall that for a quasi-uniform family of triangulations the following inverse estimate is satisfied

$$|v_h|_{H^1(\mathcal{D})} \leq Ch^{-1} \|v_h\|_{L^2(\mathcal{D})} \tag{2.14}$$

for every $v_h \in S_h$, where C is independent of the triangulation \mathcal{T}_h . For a proof of (2.14) we refer to [4, Section 4.5].

Next, we introduce the *Ritz projector* $R_h: H_0^1(\mathcal{D}) \rightarrow S_h$ as the orthogonal projector onto S_h with respect to the bilinear form a . To be more precise, as a consequence of the

lemma of Lax–Milgram, for each $v \in H_0^1(\mathcal{D})$ there exists a unique element $R_h v \in S_h$ fulfilling

$$a(R_h v, v_h) = a(v, v_h) \quad \text{for all } v_h \in S_h. \quad (2.15)$$

Note that $R_h : H_0^1(\mathcal{D}) \rightarrow S_h$ is a bounded linear operator. In addition, there exists $C \in (0, \infty)$ such that for every $h \in (0, 1]$ and $v \in H_0^1(\mathcal{D}) \cap H^2(\mathcal{D})$ it holds

$$|(R_h - I)v|_{H^1(\mathcal{D})} \leq Ch \|v\|_{H^2(\mathcal{D})}, \quad (2.16)$$

$$\|(R_h - I)v\|_{L^2(\mathcal{D})} \leq Ch^2 \|v\|_{H^2(\mathcal{D})}. \quad (2.17)$$

A proof is found, for instance, in [28, Theorem 5.5].

For the introduction and the error analysis of Monte Carlo methods, we also require some fundamental concepts from probability and stochastic analysis. For a general introduction readers are referred to standard monographs on this topic, for instance [23, 24]. For the measure theoretical background see also [3, 7].

First, let us recall that a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ consists of a measurable space (Ω, \mathcal{F}) endowed with a finite measure \mathbb{P} satisfying $\mathbb{P}(\Omega) = 1$. The value $\mathbb{P}(A) \in [0, 1]$ is interpreted as the *probability* of the *event* $A \in \mathcal{F}$. A mapping $X : \Omega \rightarrow \mathbb{R}^d$, $d \in \mathbb{N}$, is called a *random variable* if X is $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$ -measurable, where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel- σ -algebra generated by the set of all open subsets of \mathbb{R}^d . More precisely, it holds true that

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

for all $B \in \mathcal{B}(\mathbb{R}^d)$. Every random variable induces a probability measure on its image space. In fact, the measure $\mathbb{P}_X : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ given by $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$ for all $B \in \mathcal{B}(\mathbb{R}^d)$ is a probability measure on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Usually, \mathbb{P}_X is called the *distribution* of X .

If the distribution \mathbb{P}_X is absolutely continuous with respect to the Lebesgue measure, then there exists a measurable, non-negative mapping $g_X : \mathbb{R}^d \rightarrow \mathbb{R}$ with

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \int_B g_X(x) \, dx$$

for every $B \in \mathcal{B}(\mathbb{R}^d)$. The mapping g_X is called the *probability density function* of X and we write $X \sim g_X(x)$.

Next, let us recall that a random variable $X : \Omega \rightarrow \mathbb{R}^d$ is called *integrable* if $\int_{\Omega} |X(\omega)| \, d\mathbb{P}(\omega) < \infty$. Then, the *expectation* of X is defined as

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega) = \int_{\mathbb{R}^d} x \, d\mathbb{P}_X(x).$$

We say that X is centered if $\mathbb{E}[X] = 0$.

Moreover, we write $X \in L^p(\Omega; \mathbb{R}^d)$ with $p \in [1, \infty)$ if $\int_{\Omega} |X(\omega)|^p \, d\mathbb{P}(\omega) < \infty$. If $d = 1$, then we simply write $L^p(\Omega) := L^p(\Omega; \mathbb{R})$. In addition, the set $L^p(\Omega; \mathbb{R}^d)$ becomes a Banach space if we identify all random variables that only differ on a set of measure zero (i.e. probability zero) and if we endow $L^p(\Omega; \mathbb{R}^d)$ with the norm

$$\|X\|_{L^p(\Omega; \mathbb{R}^d)} = (\mathbb{E}[|X|^p])^{\frac{1}{p}} = \left(\int_{\Omega} |X(\omega)|^p \, d\mathbb{P}(\omega) \right)^{\frac{1}{p}}.$$

In Section 3, we frequently encounter a family of $\mathcal{U}(T)$ -distributed random variables $(Z_T)_{T \in \mathcal{T}_h}$. This means that for each $T \in \mathcal{T}$ the mapping $Z_T: \Omega \rightarrow \mathbb{R}^2$ is a random variable that is *uniformly distributed* on the triangle T . More precisely, the distribution \mathbb{P}_{Z_T} of Z_T is given by $\mathbb{P}_{Z_T}(A) = \frac{|A \cap T|}{|T|}$ for every $A \in \mathcal{B}(\mathbb{R}^2)$. Moreover, it follows from the transformation theorem that the expectation of $v \circ Z_T$ for an arbitrary function $v \in L^1(\mathcal{D})$ is given by

$$\mathbb{E}[v(Z_T)] = \int_T v(z) \frac{1}{|T|} \, dz = \int_{\mathcal{D}} v(z) \frac{1}{|T|} \mathbb{I}_T(z) \, dz,$$

where the mapping $g_{Z_T}(z) = \frac{1}{|T|} \mathbb{I}_T(z)$, $z \in \mathcal{D} \subset \mathbb{R}^2$, is the probability density function of Z_T .

Further, we say that a family of \mathbb{R}^d -valued random variables $(X_n)_{n \in \mathbb{N}}$ is *independent* if for any finite subset $M \subset \mathbb{N}$ and for arbitrary events $(A_m)_{m \in M} \subset \mathcal{B}(\mathbb{R}^d)$ we have the multiplication rule

$$\mathbb{P}\left(\bigcap_{m \in M} \{\omega \in \Omega : X_m(\omega) \in A_m\}\right) = \prod_{m \in M} \mathbb{P}(\{\omega \in \Omega : X_m(\omega) \in A_m\}).$$

On the level of distributions this basically means that the joint distribution of each finite subfamily $(X_m)_{m \in M}$ is equal to the product measure of the single distributions. This directly implies the multiplication rule for the expectation

$$\mathbb{E}\left[\prod_{m \in M} X_m\right] = \prod_{m \in M} \mathbb{E}[X_m], \tag{2.18}$$

provided X_m is integrable for each $m \in M$.

Finally, let us mention that we often encounter random variables taking values in a function space instead of \mathbb{R}^d . For instance, in Theorem 1 we construct a random variable with values in $S_h \subset H_0^1(\mathcal{D})$. Since S_h is finite dimensional all notions for \mathbb{R}^d -valued random variables carry over to this case in a straight-forward way. However, we often use the norm of the Bochner space $L^p(\Omega; V)$ with either $V = H_0^1(\mathcal{D})$ or $V = L^2(\mathcal{D})$, which is given by

$$\|X\|_{L^p(\Omega; V)} = (\mathbb{E}[\|X\|_V^p])^{\frac{1}{p}} = \left(\int_{\Omega} \|X(\omega)\|_V^p \, d\mathbb{P}(\omega) \right)^{\frac{1}{p}}$$

for $p \in [1, \infty)$. For an introduction to Bochner spaces we refer to [7, Appendix E].

3 A randomized quadrature formula on a triangulation

As already mentioned in the introduction, quadrature rules are often used for the assembly of the matrix-vector system (1.4) associated to the finite element method for (2.12). In this section, inspired by the stratified Monte Carlo method, we introduce the first randomized quadrature formula, which is linked to the underlying triangulation \mathcal{T}_h of the finite element space S_h . We discuss the well-posedness of the resulting method and derive error estimates in a similar way as for deterministic quadrature rules shown in [28, Section 5.6].

3.1 An unbiased randomized quadrature

Let \mathcal{T}_h , $h \in (0, 1]$, be an admissible triangulation of \mathcal{D} . For a given $v \in L^1(\mathcal{D})$, we consider the following Monte Carlo estimator

$$Q_{MC}[v] := \sum_{T \in \mathcal{T}_h} |T| v(Z_T), \quad (3.19)$$

where we sum over all triangles of the triangulation \mathcal{T}_h . Hereby, $(Z_T)_{T \in \mathcal{T}_h}$ denotes an independent family of random variables such that for each triangle $T \in \mathcal{T}_h$ the random variable Z_T is uniformly distributed on T , that is $Z_T \sim \mathcal{U}(T)$. We discuss the simulation of Z_T and the implementation of Q_{MC} in Subsection 5.4.

Observe that the randomized quadrature rule is independent of the considered equivalence class of $v \in L^1(\mathcal{D})$. If $v(x) = \tilde{v}(x)$ for almost every $x \in \mathcal{D}$, then it follows that $Q_{MC}[v] = Q_{MC}[\tilde{v}]$ with probability one.

Lemma 1 *Let \mathcal{T}_h be an admissible triangulation with maximal edge length $h \in (0, 1]$. Then, the randomized quadrature rule Q_{MC} is unbiased, i.e., for every $v \in L^1(\mathcal{D})$ it holds*

$$\mathbb{E}[Q_{MC}[v]] = \int_{\mathcal{D}} v(x) \, dx.$$

Moreover, if $v \in L^2(\mathcal{D})$ then it holds that

$$\mathbb{E}\left[\left|\int_{\mathcal{D}} v(x) \, dx - Q_{MC}[v]\right|^2\right] \leq \frac{\sqrt{3}}{2} h^2 \|v\|_{L^2(\mathcal{D})}^2.$$

In addition, if $v \in W^{s,2}(\mathcal{D})$ for some $s \in (0, 1)$ then it follows that

$$\mathbb{E}\left[\left|\int_{\mathcal{D}} v(x) \, dx - Q_{MC}[v]\right|^2\right] \leq h^{2+2s} |v|_{W^{s,2}(\mathcal{D})}^2.$$

Proof Due to $Z_T \sim \frac{1}{|T|} \mathbb{I}_T(z) \, dz$ for every $T \in \mathcal{T}_h$ we have

$$\mathbb{E}[|T| v(Z_T)] = |T| \int_{\mathcal{D}} v(z) \frac{1}{|T|} \mathbb{I}_T(z) \, dz = \int_T v(z) \, dz.$$

Then the first assertion follows by summing over all triangles of the triangulation.

Now, let $v \in L^2(\mathcal{D})$ be arbitrary. Then, the mean-square error is equal to

$$\begin{aligned} \mathbb{E}\left[\left|\int_{\mathcal{D}} v(x) \, dx - Q_{MC}[v]\right|^2\right] &= \mathbb{E}\left[\left|\sum_{T \in \mathcal{T}_h} \left(\int_T v(x) \, dx - |T|v(Z_T)\right)\right|^2\right] \\ &= \sum_{T \in \mathcal{T}_h} \mathbb{E}\left[\left|\int_T v(x) \, dx - |T|v(Z_T)\right|^2\right] \end{aligned}$$

since the summands are independent and centered random variables. Therefore, they are orthogonal with respect to the $L^2(\Omega)$ -inner product as can easily be deduced from (2.18).

Next, for every $T \in \mathcal{T}_h$ we make use of $Z_T \sim \frac{1}{|T|}\mathbb{1}_T(z) \, dz$ and the Cauchy–Schwarz inequality. This yields

$$\begin{aligned} \mathbb{E}\left[\left|\int_T v(x) \, dx - |T|v(Z_T)\right|^2\right] &= |T|^2 \mathbb{E}\left[\left|\frac{1}{|T|} \int_T v(x) \, dx - v(Z_T)\right|^2\right] \\ &= |T| \int_T \left|\frac{1}{|T|} \int_T v(x) \, dx - v(z)\right|^2 \, dz \quad (3.20) \\ &\leq \int_T \int_T |v(x) - v(z)|^2 \, dx \, dz. \end{aligned}$$

Then, since $v \in L^2(\mathcal{D})$ we get

$$\begin{aligned} \int_T \int_T |v(x) - v(z)|^2 \, dx \, dz &= \int_T \int_T (v(x)^2 - 2v(x)v(z) + v(z)^2) \, dx \, dz \\ &= 2|T| \int_T |v(x)|^2 \, dx - 2\left(\int_T v(x) \, dx\right)^2 \\ &\leq 2|T| \int_T |v(x)|^2 \, dx. \end{aligned}$$

Then, we recall Weitzenböck’s inequality [38], which yields an upper bound for the area $|T|$ of a triangle $T \in \mathcal{T}_h$ with maximal edge length h . More precisely, it holds

$$|T| \leq \frac{\sqrt{3}}{4} h^2. \quad (3.21)$$

Hence, after summing over all triangles we obtain

$$\left\|\int_{\mathcal{D}} v(x) \, dx - Q_{MC}[v]\right\|_{L^2(\Omega)}^2 \leq 2 \sum_{T \in \mathcal{T}_h} |T| \int_T |v(x)|^2 \, dx \leq \frac{\sqrt{3}}{2} h^2 \|v\|_{L^2(\mathcal{D})}^2.$$

This proves the second claim.

Finally, let $v \in W^{s,2}(\mathcal{D})$, $s \in (0, 1)$. The estimate in (3.20) is then continued by

$$\begin{aligned} \mathbb{E} \left[\left| \int_T v(x) \, dx - |T|v(Z_T) \right|^2 \right] &\leq \int_T \int_T |v(x) - v(z)|^2 \, dx \, dz \\ &\leq h^{2+2s} \int_T \int_T \frac{|v(x) - v(z)|^2}{|x - z|^{2+2s}} \, dx \, dz \\ &= h^{2+2s} |v|_{W^{s,2}(T)}^2 \end{aligned}$$

since $|x - z| \leq h$ for all $x, y \in T$. After summing over all triangles we directly obtain the third assertion. \square

3.2 Integrating the randomized quadrature into FEM

Next, we apply the randomized quadrature formula (3.19) for the approximation of the bilinear form a and the linear form F defined in (2.8) and (2.9). From this we obtain two randomized mappings $a_{MC}: S_h \times S_h \rightarrow L^\infty(\Omega)$ and $F_{MC}: S_h \rightarrow L^2(\Omega)$ which are given by

$$a_{MC}(v_h, w_h) := Q_{MC}[\sigma \nabla v_h \cdot \nabla w_h] = \sum_{T \in \mathcal{T}_h} |T| \sigma(Z_T) \nabla v_h(Z_T) \cdot \nabla w_h(Z_T) \tag{3.22}$$

and

$$F_{MC}(v_h) := Q_{MC}[f v_h] = \sum_{T \in \mathcal{T}_h} |T| f(Z_T) v_h(Z_T) \tag{3.23}$$

for all $v_h, w_h \in S_h$. In passing, we observe that $a_{MC}(v_h, w_h) = a(v_h, w_h)$ if $\sigma \equiv c \in (0, \infty)$ in \mathcal{D} . This holds true since the gradients of $v_h, w_h \in S_h$ are constant on each triangle.

The next lemma answers the question of well-posedness of a_{MC} and F_{MC} and contains some additional properties.

Lemma 2 *Let $(\mathcal{T}_h)_{h \in (0,1]}$ be a family of admissible triangulations of \mathcal{D} . Assume that $\sigma \in L^\infty(\mathcal{D})$ satisfies $\sigma(x) \geq \sigma_0 > 0$ for almost every $x \in \mathcal{D}$. Then, the mapping a_{MC} introduced in (3.22) is well-defined for every $h \in (0, 1]$. Moreover, it holds \mathbb{P} -almost surely that*

$$\begin{aligned} |a_{MC}(v_h, w_h)| &\leq \|\sigma\|_{L^\infty(\mathcal{D})} |v_h|_{H^1(\mathcal{D})} |w_h|_{H^1(\mathcal{D})}, \\ a_{MC}(v_h, v_h) &\geq \sigma_0 |v_h|_{H^1(\mathcal{D})}^2 \end{aligned}$$

for all $v_h, w_h \in S_h$.

In addition, if $f \in L^2(\mathcal{D})$ and the family of triangulations satisfies Assumption 2 then the mapping F_{MC} defined in (3.23) is also well-defined and there exists $C \in (0, \infty)$ independent of \mathcal{T}_h with

$$|F_{MC}(v_h)| \leq C \ell_h^{\frac{1}{2}} Q_{MC}[|f|] |v_h|_{H^1(\mathcal{D})} < \infty \quad \mathbb{P}\text{-a.s.},$$

$$\|F_{MC}(v_h)\|_{L^2(\Omega)} \leq C \|f\|_{L^2(\mathcal{D})} |v_h|_{H^1(\mathcal{D})}$$

for all $v_h \in S_h$, where $\ell_h = \max(1, \log(1/h))$.

Proof We first show that $a_{MC}(v_h, w_h) \in L^\infty(\Omega)$ for every $v_h, w_h \in S_h$. To see this, we recall that the functions in S_h are linear on each triangle T in \mathcal{T}_h . This implies that the gradient ∇v_h is piecewise constant for every $v_h \in S_h$. Hence, the random variables $\nabla v_h(Z_T), T \in \mathcal{T}_h$, are, in fact, constant with probability one. This implies that

$$|T| |\nabla v_h(Z_T)|^2 = \int_T |\nabla v_h(x)|^2 dx \quad \mathbb{P}\text{-almost surely.}$$

Together with the assumption $\sigma \in L^\infty(\mathcal{D})$ it therefore follows that the summands in (3.22) are essentially bounded random variables. More precisely, it holds \mathbb{P} -almost surely that

$$\begin{aligned} |a_{MC}(v_h, w_h)| &\leq \sum_{T \in \mathcal{T}_h} |T| |\sigma(Z_T)| |\nabla v_h(Z_T)| |\nabla w_h(Z_T)| \\ &\leq \|\sigma\|_{L^\infty(\mathcal{D})} \sum_{T \in \mathcal{T}_h} |T| |\nabla v_h(Z_T)| |\nabla w_h(Z_T)| \\ &\leq \|\sigma\|_{L^\infty(\mathcal{D})} \left(\sum_{T \in \mathcal{T}_h} |T| |\nabla v_h(Z_T)|^2 \right)^{\frac{1}{2}} \left(\sum_{T \in \mathcal{T}_h} |T| |\nabla w_h(Z_T)|^2 \right)^{\frac{1}{2}} \\ &= \|\sigma\|_{L^\infty(\mathcal{D})} |v_h|_{H^1(\mathcal{D})} |w_h|_{H^1(\mathcal{D})} \end{aligned}$$

for all $v_h, w_h \in S_h$.

Moreover, the same arguments yield for every $v_h \in S_h$

$$a_{MC}(v_h, v_h) = \sum_{T \in \mathcal{T}_h} |T| |\sigma(Z_T)| |\nabla v_h(Z_T)|^2 \geq \sigma_0 |v_h|_{H^1(\mathcal{D})}^2 \quad \mathbb{P}\text{-almost surely,}$$

since $\sigma(Z_T) \geq \sigma_0 > 0$ almost surely.

Next, we turn to the mapping F_{MC} . From (2.13) it follows for $v_h \in S_h$ that

$$\begin{aligned} |F_{MC}(v_h)| &\leq \sum_{T \in \mathcal{T}_h} |T| |f(Z_T)| |\nabla v_h(Z_T)| \leq \|v_h\|_{L^\infty(\mathcal{D})} Q_{MC}[|f|] \\ &\leq C \ell_h^{\frac{1}{2}} Q_{MC}[|f|] |v_h|_{H^1(\mathcal{D})}. \end{aligned}$$

Observe that the bound on the right-hand side still contains a random quadrature formula and is, therefore, itself random. However, for $f \in L^2(\mathcal{D})$ it follows from applications of the Cauchy–Schwarz inequality and Lemma 1 that

$$\mathbb{E}[(Q_{MC}[|f|])^2] = \mathbb{E}\left[\left(\sum_{T \in \mathcal{T}_h} |T| |f(Z_T)|\right)^2\right]$$

$$\begin{aligned} &\leq |\mathcal{D}| \mathbb{E} \left[\sum_{T \in \mathcal{T}_h} |T| |f(Z_T)|^2 \right] \\ &= |\mathcal{D}| \int_{\mathcal{D}} |f(z)|^2 \, dz. \end{aligned}$$

In particular, we have that $Q_{MC}[|f|] < \infty$ with probability one. This also proves that $F_{MC}(v_h) \in L^2(\Omega)$. It remains to prove the asserted estimate of the $L^2(\Omega)$ -norm of $F_{MC}(v_h)$. For this we first observe that

$$\|F_{MC}(v_h)\|_{L^2(\Omega)}^2 = \|F_{MC}(v_h) - \mathbb{E}[F_{MC}(v_h)]\|_{L^2(\Omega)}^2 + (\mathbb{E}[F_{MC}(v_h)])^2$$

for every $v_h \in S_h$. From Lemma 1, the Cauchy–Schwarz inequality, and the Poincaré inequality on $H_0^1(\mathcal{D})$ it follows that

$$\begin{aligned} (\mathbb{E}[F_{MC}(v_h)])^2 &= \left(\int_{\mathcal{D}} f(x)v_h(x) \, dx \right)^2 \\ &\leq \int_{\mathcal{D}} |f(x)|^2 \, dx \int_{\mathcal{D}} |v_h(x)|^2 \, dx \leq C \|f\|_{L^2(\mathcal{D})}^2 |v_h|_{H^1(\mathcal{D})}^2, \end{aligned}$$

where the constant C only depends on \mathcal{D} . An application of Lemma 1 then yields

$$\begin{aligned} \mathbb{E}[|F_{MC}(v_h) - \mathbb{E}[F_{MC}(v_h)]|^2] &= \mathbb{E} \left[\left| Q_{MC}[fv_h] - \int_{\mathcal{D}} f(x)v_h(x) \, dx \right|^2 \right] \\ &\leq \frac{\sqrt{3}}{2} h^2 \|fv_h\|_{L^2(\mathcal{D})}^2 \\ &\leq \frac{\sqrt{3}}{2} h^2 \|f\|_{L^2(\mathcal{D})}^2 \|v_h\|_{L^\infty(\mathcal{D})}^2 \\ &\leq C \frac{\sqrt{3}}{2} h^2 \ell_h \|f\|_{L^2(\mathcal{D})}^2 |v_h|_{H^1(\Omega)}, \end{aligned}$$

where we also applied the maximum norm estimate (2.13). Hence, after taking note of $\sup_{h \in (0,1)} h^2 \ell_h = \sup_{h \in (0,1)} h^2 \max(1, \log(1/h)) < \infty$ the proof is completed. \square

Next, we introduce the finite element problem based on the randomized quadrature rule. In terms of a_{MC} and F_{MC} the problem is stated as follows:

$$\begin{cases} \text{Find } u_h^{MC} : \Omega \rightarrow S_h \text{ such that } \mathbb{P}\text{-almost surely} \\ a_{MC}(u_h^{MC}, v_h) = F_{MC}(v_h) \text{ for all } v_h \in S_h. \end{cases} \tag{3.24}$$

Theorem 1 *Suppose that $f \in L^2(\mathcal{D})$ and $\sigma \in L^\infty(\mathcal{D})$ with $\sigma(x) \geq \sigma_0 > 0$ for almost every $x \in \mathcal{D}$ are given. Then, for every admissible triangulation \mathcal{T}_h , $h \in (0, 1]$, there exists a uniquely determined solution $u_h^{MC} : \Omega \rightarrow S_h$ to the discrete problem (3.24). In addition, there exists $C \in (0, \infty)$ independent of \mathcal{T}_h such that*

$$|u_h^{MC}|_{H^1(\mathcal{D})} \leq C \ell_h^{\frac{1}{2}} Q_{MC}[|f|] \quad \mathbb{P}\text{-a.s.},$$

where $\ell_h = \max(1, \log(1/h))$.

Proof It follows from Lemma 2 that the bilinear form a_{MC} is \mathbb{P} -almost surely strictly positive and bounded. Moreover, an inspection of the proof reveals that the exceptional set $N_1 \subset \Omega$ of probability zero, where these properties might be violated, can be chosen independently of the arguments $v_h, w_h \in S_h$. This is true since only the gradients of v_h and w_h appear in $a_{MC}(v_h, w_h)$, which are piecewise constant on each triangle. Hence, on the set $\{Z_T \in T\} \in \mathcal{F}$, which has probability one, the randomness only occurs in the coefficient function σ . Therefore, for every $\omega \in \Omega \setminus N_1$ the mapping $S_h \times S_h \ni (v_h, w_h) \mapsto a_{MC}(v_h, w_h)(\omega) \in \mathbb{R}$ satisfies the conditions of the lemma of Lax–Milgram.

In the same way, there exists a measurable set $N_2 \subset \Omega$ of probability zero such that the mapping $S_h \ni v_h \mapsto F_{MC}(v_h)(\omega) \in \mathbb{R}$ is a bounded linear functional on $H_0^1(\mathcal{D})$ for all $\omega \in \Omega \setminus N_2$. In particular, we observe that the exceptional set N_2 can again be chosen independently of the mapping v_h due to the continuity of all elements in S_h . In addition, the following estimate, which was used in the proof of Lemma 2, is true for all $\omega \in \Omega$:

$$|v_h(Z_T(\omega))| \leq \|v_h\|_{L^\infty(\mathcal{D})}.$$

Consequently, for every fixed $\omega \in \Omega \setminus (N_1 \cup N_2)$ the lemma of Lax–Milgram uniquely determines an element $u_h^{MC}(\omega) \in S_h$ satisfying

$$a_{MC}(u_h^{MC}(\omega), v_h)(\omega) = F_{MC}(v_h)(\omega) \quad \text{for all } v_h \in S_h. \tag{3.25}$$

Let us define $u_h^{MC}(\omega) = 0 \in S_h$ for all $\omega \in N_1 \cup N_2$. Next, we have to prove that the mapping $\Omega \ni \omega \mapsto u_h^{MC}(\omega) \in S_h$ is measurable. However, this follows from an application of [10, Lemma 4.3] to the mapping $g: \Omega \times \mathbb{R}^{N_h} \rightarrow \mathbb{R}^{N_h}$ defined by $g(v, \omega) := [a_{MC}(\sum_{i=1}^{N_h} v_i \psi_i, \psi_j)(\omega) - F_{MC}(\psi_j)]_{j=1}^{N_h}$, where $v = [v_i]_{i=1}^{N_h} \in \mathbb{R}^{N_h}$ and $(\psi_j)_{j=1}^{N_h} \subset S_h$ is an arbitrary basis of the finite dimensional space S_h .

It remains to prove the stability estimate. Due to Lemma 2 and (3.25) it holds on $\Omega \setminus (N_1 \cup N_2)$ that

$$\sigma_0 |u_h^{MC}|_{H^1(\mathcal{D})}^2 \leq a_{MC}(u_h^{MC}, u_h^{MC}) = F_{MC}(u_h^{MC}) \leq C \ell_h^{\frac{1}{2}} Q_{MC}[|f|] |u_h^{MC}|_{H^1(\mathcal{D})}.$$

Hence, after canceling the norm of u_h^{MC} one time on both sides of the inequality we obtain the desired estimate. □

Let us emphasize that the solution to the discrete problem (3.24) is a random variable. In fact, it follows directly from Theorem 1 that $u_h^{MC} \in L^p(\Omega; H_0^1(\mathcal{D}))$ provided $f \in L^p(\mathcal{D})$ for $p \in [2, \infty]$.

As in the standard error analysis (cf. [28, Theorem 5.7]), we want to use u_h^{MC} as a test function in the discrete problem (3.24). However, in contrast to the situation in Lemma 1 we have, in general, that $|\mathbb{E}[F_{MC}(v_h)] - F(\mathbb{E}[v_h])| \neq 0$ for an arbitrary S_h -valued random function $v_h \in L^2(\Omega; H_0^1(\mathcal{D}))$. The following lemmas give an estimate of this difference when f lives in different spaces.

Lemma 3 *Let Assumption 2 be satisfied. Then, there exists $C \in (0, \infty)$ such that for every $h \in (0, 1]$, $f \in L^p(\mathcal{D})$, $p \in [2, \infty]$, and S_h -valued random variable $v_h \in L^2(\Omega; H_0^1(\mathcal{D}))$ it holds*

$$|\mathbb{E}[F_{MC}(v_h) - F(v_h)]| \leq \begin{cases} Ch^{1-\frac{2}{p}} \|f\|_{L^p(\mathcal{D})} \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}, & \text{if } p \in [2, \infty), \\ C\ell_h^{\frac{1}{2}} h \|f\|_{L^\infty(\mathcal{D})} \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}, & \text{if } p = \infty, \end{cases}$$

where $\ell_h = \max(1, \log(1/h))$.

Proof For the error analysis it is convenient to choose an $L^2(\mathcal{D})$ -orthonormal basis $(\psi_j)_{j=1}^{N_h}$ of S_h , which solves the discrete eigenvalue problem

$$a(\psi_j, w_h) = \lambda_{h,j} (\psi_j, w_h)_{L^2(\mathcal{D})} \tag{3.26}$$

for all $w_h \in S_h$. Hereby, $0 < \lambda_{h,1} \leq \lambda_{h,2} \leq \dots \leq \lambda_{h,N_h}$ denote the discrete eigenvalues of the bilinear form a on the finite element space $S_h \subset H_0^1(\mathcal{D})$. We refer to [28, Section 6.2] regarding the existence of $(\lambda_{h,j})_{j=1}^{N_h}$ and the associated orthonormal basis $(\psi_j)_{j=1}^{N_h}$.

Next, let $h \in (0, 1]$, $f \in L^p(\mathcal{D})$, $p \in [2, \infty]$, and an S_h -valued random variable $v_h \in L^2(\Omega; H_0^1(\mathcal{D}))$ be arbitrary. Then, we represent v_h in terms of the orthonormal basis $(\psi_j)_{j=1}^{N_h} \subset S_h$ by

$$v_h = \sum_{j=1}^{N_h} v_j \psi_j, \tag{3.27}$$

For this choice of the basis, the random coefficients $(v_j)_{j=1}^{N_h} \subset L^2(\Omega)$ are given by

$$v_j = (v_h, \psi_j)_{L^2(\mathcal{D})}.$$

In particular, it follows from the Cauchy–Schwarz inequality that v_j is indeed a real-valued and square-integrable random variable for every $j \in \{1, \dots, N_h\}$. Due to the linearity of F and F_{MC} we then arrive at the estimate

$$\begin{aligned} & |\mathbb{E}[F_{MC}(v_h) - F(v_h)]| \\ &= \left| \sum_{j=1}^{N_h} \mathbb{E}[v_j (F_{MC}(\psi_j) - F(\psi_j))] \right| \\ &\leq \sum_{j=1}^{N_h} (\mathbb{E}[|v_j|^2])^{\frac{1}{2}} (\mathbb{E}[|F_{MC}(\psi_j) - F(\psi_j)|^2])^{\frac{1}{2}} \\ &\leq \left(\sum_{j=1}^{N_h} \lambda_{h,j} \mathbb{E}[|v_j|^2] \right)^{\frac{1}{2}} \left(\sum_{j=1}^{N_h} \lambda_{h,j}^{-1} \mathbb{E}[|F_{MC}(\psi_j) - F(\psi_j)|^2] \right)^{\frac{1}{2}} \end{aligned}$$

by additional applications of the Cauchy–Schwarz inequality. From (3.27) and (3.26) we then get

$$a(v_h, v_h) = \sum_{i,j=1}^{N_h} v_j v_i a(\psi_j, \psi_i) = \sum_{i,j=1}^{N_h} \lambda_{h,j} v_i v_j (\psi_j, \psi_i)_{L^2(\mathcal{D})} = \sum_{j=1}^{N_h} \lambda_{h,j} v_j^2,$$

since $(\psi_j)_{j=1}^{N_h}$ is an orthonormal basis of S_h . From this it follows that

$$\left(\sum_{j=1}^{N_h} \lambda_{h,j} \mathbb{E}[|v_j|^2] \right)^{\frac{1}{2}} = \left(\mathbb{E}[a(v_h, v_h)] \right)^{\frac{1}{2}} \leq \|\sigma\|_{L^\infty(\mathcal{D})}^{\frac{1}{2}} \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}. \tag{3.28}$$

Moreover, an application of Lemma 1 yields

$$\mathbb{E}[|F_{MC}(\psi_j) - F(\psi_j)|^2] = \mathbb{E}\left[\left| Q_{MC}(f\psi_j) - \int_{\mathcal{D}} f\psi_j \, dx \right|^2 \right] \leq \frac{\sqrt{3}}{2} h^2 \|f\psi_j\|_{L^2(\mathcal{D})}^2$$

for every $j \in \{1, \dots, N_h\}$. Further, since $f \in L^p(\mathcal{D})$, $p \in [2, \infty]$, it follows from an application of Hölder’s inequality with conjugated exponent $p' \in [2, \infty]$ determined by $\frac{1}{p} + \frac{1}{p'} = \frac{1}{2}$ that

$$\|f\psi_j\|_{L^2(\mathcal{D})} \leq \|f\|_{L^p(\mathcal{D})} \|\psi_j\|_{L^{p'}(\mathcal{D})}.$$

An application of the Gagliardo–Nirenberg inequality, cf. [33, Theorem 1.24], yields

$$\|\psi_j\|_{L^{p'}(\mathcal{D})} \leq C \|\psi_j\|_{L^2(\mathcal{D})}^{\frac{2}{p'}} |\psi_j|_{H^1(\mathcal{D})}^{1-\frac{2}{p'}},$$

where the constant C is independent of $j \in \{1, \dots, N_h\}$. Since $\|\psi_j\|_{L^2(\mathcal{D})} = 1$ for every $j \in \{1, \dots, N_h\}$ and due to (2.10) and (3.26) we therefore obtain

$$\|\psi_j\|_{L^{p'}(\mathcal{D})} \leq C |\psi_j|_{H^1(\mathcal{D})}^{\frac{2}{p'}} \leq \frac{C}{\sigma_0^{\frac{1}{p}}} a(\psi_j, \psi_j)^{\frac{1}{p}} \leq \frac{C}{\sigma_0^{\frac{1}{p}}} \lambda_{h,j}^{\frac{1}{p}}$$

for every $p, p' \in [2, \infty]$ with $\frac{1}{p} + \frac{1}{p'} = \frac{1}{2}$. Altogether, we have the bound

$$\begin{aligned} \sum_{j=1}^{N_h} \lambda_{h,j}^{-1} \mathbb{E}[|F_{MC}(\psi_j) - F(\psi_j)|^2] &\leq \frac{\sqrt{3}}{2} h^2 \|f\|_{L^p(\mathcal{D})}^2 \sum_{j=1}^{N_h} \lambda_{h,j}^{-1} \|\psi_j\|_{L^{p'}(\mathcal{D})}^2 \\ &\leq Ch^2 \|f\|_{L^p(\mathcal{D})}^2 \sum_{j=1}^{N_h} \lambda_{h,j}^{-1+\frac{2}{p}}. \end{aligned} \tag{3.29}$$

Concerning the last sum we recall from [28, Theorem 6.7] that

$$\lambda_j \leq \lambda_{h,j}$$

for all $j \in \{1, \dots, N_h\}$, where $(\lambda_j)_{j \in \mathbb{N}}$ denotes the family of eigenvalues of the bilinear form a on the full space $H_0^1(\mathcal{D})$. Moreover, it is well-known, cf. [28, Section 6.1], that there exist constants $c_1, c_2 \in (0, \infty)$ only depending on σ and \mathcal{D} such that

$$c_1 j \leq \lambda_j \leq c_2 j.$$

From this it follows that

$$\sum_{j=1}^{N_h} \lambda_{h,j}^{-1+\frac{2}{p}} \leq \sum_{j=1}^{N_h} \lambda_j^{-1+\frac{2}{p}} \leq c_1^{-1+\frac{2}{p}} \sum_{j=1}^{N_h} j^{-1+\frac{2}{p}} \leq c_1^{-1+\frac{2}{p}} \left(1 + \int_1^{N_h} y^{-1+\frac{2}{p}} dy\right).$$

Hence, we obtain

$$\sum_{j=1}^{N_h} \lambda_{h,j}^{-1+\frac{2}{p}} \leq \begin{cases} \frac{p}{2} c_1^{-1+\frac{2}{p}} N_h^{\frac{2}{p}}, & \text{if } p \in [2, \infty), \\ c_1^{-1} (1 + \log(N_h)), & \text{if } p = \infty. \end{cases}$$

From (3.26), (2.11), and the inverse estimate (2.14) it then follows that

$$N_h \leq \frac{1}{c_1} \lambda_{h,N_h} = \frac{1}{c_1} a(\psi_{N_h}, \psi_{N_h}) \leq \frac{1}{c_1} \|\sigma\|_{L^\infty(\mathcal{D})} |\psi_{N_h}|_{H^1(\mathcal{D})}^2 \leq Ch^{-2}.$$

This implies that $\log(N_h) \leq C \max(1, \log(1/h)) = C\ell_h$. Altogether, this yields

$$\sum_{j=1}^{N_h} \lambda_{h,j}^{-1+\frac{2}{p}} \leq \begin{cases} Ch^{-\frac{4}{p}}, & \text{if } p \in [2, \infty), \\ C\ell_h, & \text{if } p = \infty. \end{cases} \tag{3.30}$$

Combining this with (3.28) and (3.29) then completes the proof. □

Lemma 4 *Let Assumption 2 be satisfied. Then there exists $C \in (0, \infty)$ such that for every $h \in (0, 1]$, $f \in W^{s',2}(\mathcal{D})$, $s' \in (0, 1)$, and S_h -valued random variable $v_h \in L^2(\Omega; H_0^1(\mathcal{D}))$ it holds*

$$\left| \mathbb{E}[F_{MC}(v_h) - F(v_h)] \right| \leq Ch^{s'} \|f\|_{W^{s',2}(\mathcal{D})} \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}.$$

Proof To show the quadrature estimate, we may need to introduce some auxiliary quadrature

$$F_{\text{sMC}}(v_h) := \sum_{T \in \mathcal{T}_h} |T| f(Z_T) v_h(C_T) \tag{3.31}$$

for all $v_h \in S_h$ and C_T is the centroid of the triangle of T . Note that

$$\int_T v_h(x) \, dx = |T|v_h(C_T) \tag{3.32}$$

because of v_h being piecewise linear on every T , see [28, Eq. (5.61)].

Now we may decompose our target term as

$$\begin{aligned} & |\mathbb{E}[F_{MC}(v_h) - F(v_h)]| \\ & \leq |\mathbb{E}[F_{MC}(v_h) - F_{sMC}(v_h)]| \\ & \quad + |\mathbb{E}[F_{sMC}(v_h) - \sum_{T \in \mathcal{T}_h} \int_T f(x) v_h(C_T) \, dx]| \\ & \quad + |\mathbb{E}[\sum_{T \in \mathcal{T}_h} \int_T f(x) v_h(C_T) \, dx - F(v_h)]| := \sum_{i=1}^3 T_i, \end{aligned}$$

For T_1 , we have

$$\begin{aligned} & |\mathbb{E}[F_{MC}(v_h) - F_{sMC}(v_h)]| \\ & = \left| \mathbb{E} \left[\sum_{T \in \mathcal{T}_h} |T| f(Z_T) (v_h(Z_T) - v_h(C_T)) \right] \right| \\ & \leq \sum_{T \in \mathcal{T}_h} (\mathbb{E}[|T| |f(Z_T)|^2])^{1/2} (\mathbb{E}[|T| |v_h(Z_T) - v_h(C_T)|^2])^{1/2} \\ & \leq \left(\sum_{T \in \mathcal{T}_h} \mathbb{E}[|T| |f(Z_T)|^2] \right)^{1/2} \left(\sum_{T \in \mathcal{T}_h} \mathbb{E}[|T| |v_h(Z_T) - v_h(C_T)|^2] \right)^{1/2} \\ & \leq \|f\|_{L^2(\mathcal{D})} \left(h^2 \sum_{T \in \mathcal{T}_h} \mathbb{E}[|T| |\nabla v_h(C_T)|^2] \right)^{1/2} \\ & = h \|f\|_{L^2(\mathcal{D})}^2 \left(\sum_{T \in \mathcal{T}_h} \mathbb{E} \left[\int_T |\nabla v_h(x)|^2 \, dx \right] \right)^{1/2} \\ & = h \|f\|_{L^2(\mathcal{D})} \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}, \end{aligned}$$

where we use the fact that $\nabla v_h(\cdot)$ is constant on each T to derive the last second line. The idea to bound T_3 is similar,

$$\begin{aligned} & |\mathbb{E}[\sum_{T \in \mathcal{T}_h} \int_T f(x) v_h(C_T) \, dx] - F(v_h)| \\ & = |\mathbb{E}[\sum_{T \in \mathcal{T}_h} \int_T f(x) (v_h(C_T) - v_h(x)) \, dx]| \\ & \leq \|f\|_{L^2(\mathcal{D})} \left(\sum_{T \in \mathcal{T}_h} \mathbb{E}[|T| |v_h(C_T) - v_h(x)|^2] \right)^{1/2} \end{aligned}$$

$$\leq h \|f\|_{L^2(\mathcal{D})} \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}.$$

For T_2 , we will use the fact (3.32) and the Cauchy–Schwarz inequality to derive

$$\begin{aligned} & \left| \mathbb{E} \left[F_{\text{sMC}}(v_h) - \sum_{T \in \mathcal{T}_h} \int_T f(x) v_h(C_T) \, dx \right] \right| \\ &= \left| \mathbb{E} \left[\sum_{T \in \mathcal{T}_h} \int_T (f(Z_T) - f(x)) v_h(C_T) \, dx \right] \right| \\ &\leq \left| \mathbb{E} \left[\sum_{T \in \mathcal{T}_h} \left(\int_T |f(Z_T) - f(x)|^2 \, dx \right)^{1/2} \left(\int_T |v_h(x)|^2 \, dx \right)^{1/2} \right] \right| \\ &\leq \left(\mathbb{E} \left[\sum_{T \in \mathcal{T}_h} \int_T |f(Z_T) - f(x)|^p \, dx \right] \right)^{1/2} \left(\mathbb{E} \left[\sum_{T \in \mathcal{T}_h} \int_T |v_h(x)|^2 \, dx \right] \right)^{1/2} \\ &= \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} \int_T \int_T |f(z) - f(x)|^p \, dx \, dz \right)^{1/2} \|v_h\|_{L^2(\Omega; L^2(\mathcal{D}))} \\ &\leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} h^{2+2s'} \int_T \int_T \frac{|f(z) - f(x)|^2}{|z - x|^{2+2s'}} \, dx \, dz \right)^{1/2} \|v_h\|_{L^2(\Omega; L^2(\mathcal{D}))} \\ &\leq h^s \|f\|_{W^{s',2}(\mathcal{D})} \|v_h\|_{H_0^1(\Omega; L^2(\mathcal{D}))}, \end{aligned}$$

where, to get the last second lines, we make use of fact that $|z - x| < h$ for all $z, x \in T$. □

Next, we state and prove the main result of this section.

Theorem 2 *Suppose that $\sigma \in L^\infty(\mathcal{D}) \cap W^{s,q}(\mathcal{D})$, $s \in (0, 1]$, $q \in (2, \infty)$, with $\sigma(x) \geq \sigma_0 > 0$ for almost every $x \in \mathcal{D}$. Let Assumptions 1 and 2 be satisfied.*

1. *If $f \in L^p(\mathcal{D})$, $p \in (2, \infty)$, then it holds*

$$\begin{aligned} \|u - u_h^{MC}\|_{L^2(\Omega; H_0^1(\mathcal{D}))} &\leq Ch \|u\|_{H^2(\mathcal{D})} + Ch^s \|u\|_{H^2(\mathcal{D})} |\sigma|_{W^{s,q}(\mathcal{D})} \\ &\quad + Ch^{1-\frac{2}{p}} \|f\|_{L^p(\mathcal{D})} \end{aligned}$$

for every $h \in (0, 1]$. Further, if $f \in L^\infty(\mathcal{D})$ then it holds

$$\begin{aligned} \|u - u_h^{MC}\|_{L^2(\Omega; H_0^1(\mathcal{D}))} &\leq Ch \|u\|_{H^2(\mathcal{D})} + Ch^s \|u\|_{H^2(\mathcal{D})} |\sigma|_{W^{s,q}(\mathcal{D})} \\ &\quad + C \ell_h^{\frac{1}{2}} h \|f\|_{L^\infty(\mathcal{D})}. \end{aligned}$$

2. *Moreover, if $f \in W^{s',2}(\mathcal{D})$, $s' \in (0, 1)$, then*

$$\begin{aligned} \|u - u_h^{MC}\|_{L^2(\Omega; H_0^1(\mathcal{D}))} &\leq Ch \|u\|_{H^2(\mathcal{D})} + Ch^s \|u\|_{H^2(\mathcal{D})} |\sigma|_{W^{s,q}(\mathcal{D})} \\ &\quad + Ch^{s'} \|f\|_{W^{s',2}(\mathcal{D})}. \end{aligned}$$

Proof Let us split the error into the following two parts

$$u_h^{MC} - u = u_h^{MC} - R_h u + R_h u - u =: \theta + \rho,$$

where $R_h: H_0^1(\mathcal{D}) \rightarrow S_h$ denotes the Ritz projector (see Section 2). Observe that θ and ρ are orthogonal with respect to the bilinear form a . Then, it follows from the positivity (2.10) and boundedness (2.11) of a that

$$\begin{aligned} \sigma_0 |u_h^{MC} - u|_{H^1(\mathcal{D})}^2 &\leq a(u_h^{MC} - u, u_h^{MC} - u) = a(\theta, \theta) + a(\rho, \rho) \\ &\leq \|\sigma\|_{L^\infty(\mathcal{D})} (|\theta|_{H^1(\mathcal{D})}^2 + |\rho|_{H^1(\mathcal{D})}^2). \end{aligned}$$

Standard error estimates for the conforming finite element method, cf. (2.16), yield

$$|\rho|_{H^1(\mathcal{D})} = |R_h u - u|_{H^1(\mathcal{D})} \leq Ch \|u\|_{H^2(\mathcal{D})}. \tag{3.33}$$

Moreover, from (2.12) and (3.24) we get \mathbb{P} -almost surely for every $v_h \in S_h$ that

$$\begin{aligned} a_{MC}(\theta, v_h) &= a_{MC}(u_h^{MC}, v_h) - a_{MC}(R_h u, v_h) \\ &= F_{MC}(v_h) - F(v_h) + a(R_h u, v_h) - a_{MC}(R_h u, v_h), \end{aligned}$$

since $F(v_h) = a(u, v_h) = a(R_h u, v_h)$ for every $v_h \in S_h$. In particular, for the choice $v_h = \theta(\omega) = u_h^{MC}(\omega) - R_h u \in S_h$ we obtain \mathbb{P} -almost surely that

$$\sigma_0 |\theta|_{H^1(\mathcal{D})}^2 \leq a_{MC}(\theta, \theta) = F_{MC}(\theta) - F(\theta) + a(R_h u, \theta) - a_{MC}(R_h u, \theta).$$

From Lemma 2 and Theorem 1 it follows directly that all terms on the right-hand side are integrable with respect to \mathbb{P} . Hence, after taking expectations it remains to give error estimates for the two terms

$$\begin{aligned} E_1 &= |\mathbb{E}[F_{MC}(\theta) - F(\theta)]|, \\ E_2 &= |\mathbb{E}[a(R_h u, \theta) - a_{MC}(R_h u, \theta)]|. \end{aligned}$$

If $f \in L^p(\mathcal{D})$, $p \in (2, \infty]$, an application of Lemma 3 directly yields

$$E_1 \leq \begin{cases} Ch^{1-\frac{2}{p}} \|f\|_{L^p(\mathcal{D})} \|\theta\|_{L^2(\Omega; H_0^1(\mathcal{D}))}, & \text{if } p \in [2, \infty), \\ C\ell_h^{\frac{1}{2}} \|f\|_{L^\infty(\mathcal{D})} \|\theta\|_{L^2(\Omega; H_0^1(\mathcal{D}))}, & \text{if } p = \infty. \end{cases}$$

If $f \in W^{s',2}(\mathcal{D})$, an application of Lemma 4 leads to

$$E_1 \leq Ch^{s'} \|f\|_{W^{s',2}(\mathcal{D})}.$$

Next, we turn to the term E_2 which is given by

$$E_2 = |\mathbb{E}[a(R_h u, \theta) - a_{MC}(R_h u, \theta)]|$$

$$= \left| \sum_{T \in \mathcal{T}_h} \mathbb{E} \left[\int_T \sigma(x) \nabla R_h u(x) \cdot \nabla \theta(x) \, dx - |T| \sigma(Z_T) \nabla R_h u(Z_T) \cdot \nabla \theta(Z_T) \right] \right|.$$

Since $R_h u \in S_h$ and $\theta: \Omega \rightarrow S_h$, the respective gradients are constant on each triangle.

Therefore, we have $\nabla R_h u(x) \cdot \nabla \theta(x) = \nabla R_h u(Z_T) \cdot \nabla \theta(Z_T)$ for every $x \in T$. Hence, we get

$$E_2 = \left| \sum_{T \in \mathcal{T}_h} \mathbb{E} \left[\left(\int_T \sigma(x) \, dx - |T| \sigma(Z_T) \right) \nabla R_h u(Z_T) \cdot \nabla \theta(Z_T) \right] \right|$$

$$\leq \sum_{T \in \mathcal{T}_h} \left(\mathbb{E} \left[\left| \left(\int_T \sigma(x) \, dx - |T| \sigma(Z_T) \right) \nabla R_h u(Z_T) \right|^2 \right] \right)^{\frac{1}{2}} \left(\mathbb{E} [|\nabla \theta(Z_T)|^2] \right)^{\frac{1}{2}}$$

$$\leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} \mathbb{E} \left[\left(\int_T \sigma(x) \, dx - |T| \sigma(Z_T) \right)^2 |\nabla R_h u(Z_T)|^2 \right] \right)^{\frac{1}{2}}$$

$$\times \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E} [|\nabla \theta(Z_T)|^2] \right)^{\frac{1}{2}}$$

by further applications of the Cauchy–Schwarz inequality. Moreover, by making again use of the fact that the gradient of θ is piecewise constant we obtain

$$\left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E} [|\nabla \theta(Z_T)|^2] \right)^{\frac{1}{2}} = \left(\mathbb{E} \left[\sum_{T \in \mathcal{T}_h} |T| |\nabla \theta(Z_T)|^2 \right] \right)^{\frac{1}{2}}$$

$$= \left(\mathbb{E} \left[\int_{\mathcal{D}} |\nabla \theta(x)|^2 \, dx \right] \right)^{\frac{1}{2}} = \|\theta\|_{L^2(\Omega; H_0^1(\mathcal{D}))}.$$

Further, due to $Z_T \sim |T|^{-1} \mathbb{1}_T(z) \, dz$ it holds

$$\left(\sum_{T \in \mathcal{T}_h} |T|^{-1} \mathbb{E} \left[\left(\int_T \sigma(x) \, dx - |T| \sigma(Z_T) \right)^2 |\nabla R_h u(Z_T)|^2 \right] \right)^{\frac{1}{2}}$$

$$= \left(\sum_{T \in \mathcal{T}_h} |T|^{-2} \int_T \left(\int_T (\sigma(x) - \sigma(z)) \, dx \right)^2 |\nabla R_h u(z)|^2 \, dz \right)^{\frac{1}{2}} \tag{3.34}$$

$$\leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-2} \int_T \left(\int_T (\sigma(x) - \sigma(z)) \, dx \right)^2 |\nabla (R_h - I)u(z)|^2 \, dz \right)^{\frac{1}{2}}$$

$$+ \left(\sum_{T \in \mathcal{T}_h} |T|^{-2} \int_T \left(\int_T (\sigma(x) - \sigma(z)) \, dx \right)^2 |\nabla u(z)|^2 \, dz \right)^{\frac{1}{2}},$$

where we applied Minkowski’s inequality in the last step. The first term is then estimated by

$$\begin{aligned}
 & \left(\sum_{T \in \mathcal{T}_h} |T|^{-2} \int_T \left(\int_T (\sigma(x) - \sigma(z)) \, dx \right)^2 |\nabla(R_h - I)u(z)|^2 \, dz \right)^{\frac{1}{2}} \\
 & \leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} \int_T \int_T (\sigma(x) - \sigma(z))^2 \, dx |\nabla(R_h - I)u(z)|^2 \, dz \right)^{\frac{1}{2}} \\
 & \leq C \|\sigma\|_{L^\infty(\mathcal{D})} \left(\sum_{T \in \mathcal{T}_h} \int_T |\nabla(R_h - I)u(z)|^2 \, dz \right)^{\frac{1}{2}} \\
 & \leq C \|\sigma\|_{L^\infty(\mathcal{D})} |(R_h - I)u|_{H^1(\mathcal{D})} \leq C \|\sigma\|_{L^\infty(\mathcal{D})} \|u\|_{H^2(\mathcal{D})} h
 \end{aligned}$$

by a further application of (3.33).

For the estimate of the last term in (3.34) we first consider $s \in (0, 1)$. Applying Hölder’s inequality with exponents $\rho = \frac{q}{2} \in (1, \infty)$ and $\rho' = \frac{q}{q-2} \in (1, \infty)$ yields

$$\begin{aligned}
 & \left(\sum_{T \in \mathcal{T}_h} |T|^{-2} \int_T \left(\int_T (\sigma(x) - \sigma(z)) \, dx \right)^2 |\nabla u(z)|^2 \, dz \right)^{\frac{1}{2}} \\
 & \leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-2} \left(\int_T \left(\int_T |\sigma(x) - \sigma(z)| \, dx \right)^{2\rho} \, dz \right)^{\frac{1}{\rho}} \left(\int_T |\nabla u(z)|^{2\rho'} \, dz \right)^{\frac{1}{\rho'}} \right)^{\frac{1}{2}} \\
 & \leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-\frac{1}{\rho}} \left(\int_T \int_T |\sigma(x) - \sigma(z)|^{2\rho} \, dx \, dz \right)^{\frac{1}{\rho}} \left(\int_T |\nabla u(z)|^{2\rho'} \, dz \right)^{\frac{1}{\rho'}} \right)^{\frac{1}{2}} \\
 & \leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} \int_T \int_T |\sigma(x) - \sigma(z)|^q \, dx \, dz \right)^{\frac{1}{q}} \left(\sum_{T \in \mathcal{T}_h} \int_T |\nabla u(z)|^{2\rho'} \, dz \right)^{\frac{1}{2\rho'}} \\
 & \leq \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} h^{2+qs} \int_T \int_T \frac{|\sigma(x) - \sigma(z)|^q}{|x - z|^{2+qs}} \, dx \, dz \right)^{\frac{1}{q}} \|u\|_{W^{1,2\rho'}(\mathcal{D})}
 \end{aligned}$$

since $|x - y| \leq h$ for all $x, y \in T$.

Next, recall that the Sobolev embedding theorem [1, Theorem 4.12] yields

$$\|u\|_{W^{1,2\rho'}(\mathcal{D})} \leq C \|u\|_{H^2(\mathcal{D})}.$$

In addition, we have $|T|^{-1} \leq c^{-1}h^{-2}$ due to Assumption 2. Altogether, this shows

$$\begin{aligned}
 & \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} h^{2+qs} \int_T \int_T \frac{|\sigma(x) - \sigma(z)|^q}{|x - z|^{2+qs}} \, dx \, dz \right)^{\frac{1}{q}} \|u\|_{W^{1,2\rho'}(\mathcal{D})} \\
 & \leq C \|u\|_{H^2(\mathcal{D})} |\sigma|_{W^{s,q}(\mathcal{D})} h^s.
 \end{aligned}$$

This completes the proof of the case $s \in (0, 1)$. The border case $s = 1$ follows by similar arguments and an additional application of the Poincaré–Wirtinger inequality. The details are left to the reader. \square

4 The modified randomized quadrature and its application to Poisson equation

The goal of this section is to increase the accuracy of the randomized quadrature formula Q_{MC} introduced in (3.19) by applying a standard variance reduction technique for Monte Carlo methods termed *importance sampling*. We briefly review the importance sampling in Section 4.1. Inspired by this technique, we propose a modified randomized quadrature for the load vector (denoted as F_{IS}) in Section 4.2, followed by its application to the Poisson equation with a detailed error analysis (see Section 4.3). Note that the randomized quadrature based on importance-sampling for the stiffness matrix coincides with a_{MC} , as discussed in Remark 1. Therefore, considering the analysis of a_{MC} in Section 3, integrating both F_{IS} and a_{MC} into the FEM framework for a general elliptic equation is a straightforward extension of the Poisson equation case. To avoid redundancy, we omit this discussion from the paper.

4.1 Importance sampling

An introduction to importance sampling and further variance reduction techniques is found, for instance, in [12, Chapter 6], [29, Chapter 3], and [31, Kapitel 5]. Let us briefly recall the main idea of importance sampling. Suppose one wants to approximate the integral

$$\int_{\mathcal{D}} v(x) \, dx,$$

where $v \in L^2(\mathcal{D})$ is given. Then, the standard Monte Carlo approach is to rewrite the integral as an expectation

$$\mathbb{E}[v(Z)] = |\mathcal{D}|^{-1} \int_{\mathcal{D}} v(x) \, dx,$$

where $Z: \Omega \rightarrow \mathcal{D}$ is a uniformly distributed random variable. In particular, the probability density function of Z is given by $p_Z(x) = \frac{1}{|\mathcal{D}|} \mathbb{1}_{\mathcal{D}}(x)$. Then, the standard Monte Carlo estimator of the integral is defined as

$$\frac{|\mathcal{D}|}{M} \sum_{i=1}^M v(Z_i),$$

where $(Z_i)_{i=1}^M$, $M \in \mathbb{N}$, is a family of independent and identically distributed copies of Z . This estimator is unbiased and its variance is equal to

$$\left\| \frac{|\mathcal{D}|}{M} \sum_{i=1}^M v(Z_i) - \int_{\mathcal{D}} v(x) \, dx \right\|_{L^2(\Omega)}^2 = \frac{1}{M} \text{var}(|\mathcal{D}|v(Z)).$$

Therefore, the accuracy of the Monte Carlo estimator is determined by the number of samples $M \in \mathbb{N}$ and the variance of the random variable $|\mathcal{D}|v(Z)$.

The main idea of importance sampling is then to increase the accuracy of the standard Monte Carlo estimator by replacing the uniformly distributed random variable Z with a random variable $Y : \Omega \rightarrow \mathcal{D}$ whose distribution is determined by a probability distribution function p_Y . If the density p_Y satisfies that $p_Y(x) = 0$ only if $v(x) = 0$, then it follows from the transformation theorem that

$$\int_{\mathcal{D}} v(x) \, dx = \int_{\mathcal{D}} \frac{v(x)}{p_Y(x)} p_Y(x) \, dx = \mathbb{E}_{p_Y} \left[\frac{v(Y)}{p_Y(Y)} \right],$$

where we write \mathbb{E}_{p_Y} to emphasise the expectation corresponds to distribution p_Y .

From this one derives the following *importance sampling estimator* given by

$$\frac{1}{M} \sum_{i=1}^M \frac{v(Y_i)}{p_Y(Y_i)},$$

where $(Y_i)_{i=1}^M$ denotes a family of independent and identically distributed copies of Y . The art of importance sampling is then to determine a suitable density p_Y such that the variance is reduced and, at the same time, the generation of random variates with density p_Y is computational feasible and affordable. It is known (cf. [12, Theorem 6.5]) that the optimal choice of the density p_Y is

$$p_Y^*(x) = \frac{|v(x)|}{\int_{\mathcal{D}} |v(y)| \, dy}, \quad \text{for } x \in \mathcal{D}.$$

Observe that p_Y^* suggests to avoid sampling in regions of $|\mathcal{D}|$, where $|v|$ is zero or very small. However, since the denominator is typically unknown it is, in general, not possible to use the density p_Y^* in practice.

Nevertheless, one can often still make use of the underlying idea to improve the accuracy of the randomized quadrature rule (3.19).

4.2 The modified randomized quadrature for the load vector

Recall that the entries of the load vector $f_h \in \mathbb{R}^{N_h}$ defined in (1.6) are given by

$$F(\varphi_j) = \int_{\mathcal{D}} f(x) \varphi_j(x) \, dx, \quad j \in \{1, \dots, N_h\},$$

where $(\varphi_j)_{j=1}^{N_h}$ denotes the standard Lagrange basis of the finite element space S_h . According to the results in the previous section, these entries are then approximated by an application of the randomized quadrature formula (3.19) given by

$$F_{MC}(\varphi_j) = Q_{MC}[f\varphi_j] = \sum_{T \in \mathcal{T}_h} |T|f(Z_T)\varphi_j(Z_T)$$

for every $j \in \{1, \dots, N_h\}$. Observe that for each triangle $T \in \mathcal{T}_h$ the term

$$|T|f(Z_T)\varphi_j(Z_T) \tag{4.35}$$

can be regarded as a standard Monte Carlo estimator with only $M = 1$ sample for the integral

$$\int_T f(x)\varphi_j(x) dx.$$

The idea of this section is to replace this term by a suitable importance sampling estimator.

Since we do not want to impose any additional assumption on f it is, as already mentioned above, not feasible to use the corresponding optimal density function p_Y^* with $v = f\varphi_j$. Instead, we recall that the piecewise linear basis function φ_j is equal to zero in two of the three vertices and equal to one in the remaining vertex of every triangle $T \in \mathcal{T}_h$ with $T \cap \text{supp}(\varphi_j) \neq \emptyset$. In particular, this implies $\varphi_j(x) \geq 0$ for every $x \in T$. Further, it holds

$$\int_T \varphi_j(x) dx = \frac{1}{3}|T|.$$

Therefore, the mapping $p_{T,j} : \mathcal{D} \rightarrow [0, \infty)$ defined by

$$p_{T,j}(x) = 3|T|^{-1}\varphi_j(x)\mathbb{I}_T(x), \quad x \in \mathcal{D}, \tag{4.36}$$

is a probability density function. By replacing Z_T in (4.35) with a random variable $Y_{T,j} \sim p_{T,j}(x) dx$ we arrive at the corresponding importance sampling estimator (again with only $M = 1$ sample)

$$\frac{f(Y_{T,j})\varphi_j(Y_{T,j})}{p_{T,j}(Y_{T,j})} = \frac{1}{3}|T|f(Y_{T,j})$$

for the integral $\int_T f(x)\varphi_j(x) dx$. Observe that the use of $Y_{T,j}$ significantly decreases the probability of the integrand $f\varphi_j$ being evaluated at a point $x \in T$ close to a vertex, where the basis function φ_j is equal to zero. We discuss the simulation of the random variable $Y_{T,j}$ in Section 5. Without causing ambiguity, throughout this section we will still adopt \mathbb{E} to denote the expectation with respect to $p_{T,j}$.

To sum up, this suggests to use the linear mapping $F_{IS}: S_h \rightarrow L^2(\Omega)$ given by

$$F_{IS}(v_h) = \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} v_j f(Y_{T,j}) \tag{4.37}$$

for every $v_h = \sum_{j=1}^{N_h} v_j \varphi_j \in S_h$. Hereby, $(Y_{T,j})_{T \in \mathcal{T}_h, j \in \{1, \dots, N_h\}}$ is a family of independent random variables with $Y_{T,j} \sim p_{T,j}(x) dx$. In particular, the entries of the load vector f_h are then approximated by

$$F_{IS}(\varphi_j) = \frac{1}{3} \sum_{\substack{T \in \mathcal{T}_h \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}} |T| f(Y_{T,j})$$

for every $j \in \{1, \dots, N_h\}$.

Remark 1 Note that the entries of the stiffness matrix A_h defined in (1.5) are given by

$$a(\varphi_i, \varphi_j) = \int_{\mathcal{D}} \sigma(x) \nabla \varphi_i(x) \cdot \nabla \varphi_j(x) dx = \sum_{T \in \mathcal{T}_h} \int_T \sigma(x) (\nabla \varphi_i(x) \cdot \nabla \varphi_j(x)) dx$$

for $i, j \in \{1, \dots, N_h\}$, where the function $(\nabla \varphi_i(\cdot) \cdot \nabla \varphi_j(\cdot))$ is constant on each T and is non-zero constant on each T with $T \cap \text{supp}(\varphi_i) \cap \text{supp}(\varphi_j) \neq \emptyset$. Recall that the (partially) importance sampling strategy described above achieves variance reduction by choosing an appropriate reference density function and assigning higher weights to the region of domain with higher density evaluation. Applying this strategy to approximate the summand of the right hand side through a reference density proportional to $(\nabla \varphi_i(\cdot) \cdot \nabla \varphi_j(\cdot))$ would leads to the estimator a_{MC} as the reference density function being constant over T .

Remark 1 demonstrates that the importance sampling strategy does not change the evaluation of stiffness matrix. Since this section focuses on assessing the performance of the modified quadrature (4.37) in comparison to (3.23), we will use the Poisson equation as an example, providing a detailed error analysis.

As the following lemma shows, the importance sampling estimator (4.37) is unbiased and convergent in the limit $h \rightarrow 0$.

Lemma 5 *Let \mathcal{T}_h be an admissible triangulation with maximal edge length $h \in (0, 1]$. Then, for every $f \in L^1(\mathcal{D})$ and $v_h \in S_h$ it holds that*

$$\mathbb{E}[F_{IS}(v_h)] = \int_{\mathcal{D}} f(x) v_h(x) dx.$$

Further, if $f \in L^p(\mathcal{D})$, $p \in [2, \infty]$, then it holds for every $v_h \in S_h$ that

$$\left\| \int_{\mathcal{D}} f(x) v_h(x) dx - F_{IS}(v_h) \right\|_{L^2(\Omega)}$$

$$\leq \frac{1}{\sqrt[4]{12}} h \|v_h\|_{L^\infty(\mathcal{D})}^{\frac{2}{p}} \|f\|_{L^p(\mathcal{D})} (2h \|v_h\|_{H^1(\mathcal{D})} + \|v_h\|_{L^2(\mathcal{D})})^{1-\frac{2}{p}}.$$

In addition, if $f \in W^{s,2}(\mathcal{D})$ for some $s \in (0, 1)$ then it holds for every $v_h \in S_h$ that

$$\left\| \int_{\mathcal{D}} f(x) v_h(x) \, dx - F_{IS}(v_h) \right\|_{L^2(\Omega)} \leq h^{1+s} \|v_h\|_{L^\infty(\mathcal{D})} |f|_{W^{s,2}(\mathcal{D})}.$$

Proof Let $v_h = \sum_{j=1}^{N_h} v_j \varphi_j \in S_h$ be arbitrary with coefficients $(v_j)_{j=1}^{N_h} \subset \mathbb{R}$. Due to $Y_{T,j} \sim \frac{3}{|T|} \varphi_j(z) \mathbb{I}_T(z) \, dz$ for every $T \in \mathcal{T}_h$ we have

$$\sum_{j=1}^{N_h} v_j \mathbb{E} \left[\frac{|T|}{3} f(Y_{T,j}) \right] = \sum_{j=1}^{N_h} v_j \frac{|T|}{3} \int_T f(z) \varphi_j(z) \frac{3}{|T|} \, dz = \int_T f(z) v_h(z) \, dz.$$

Then, the first assertion follows by summing over all triangles of the triangulation.

Now, let $f \in L^2(\mathcal{D})$ be arbitrary. In the same way as in the proof of Lemma 1, the mean-square error is shown to be equal to

$$\begin{aligned} & \left\| \int_{\mathcal{D}} f(x) v_h(x) \, dx - F_{IS}(v_h) \right\|_{L^2(\Omega)}^2 \\ &= \sum_{T \in \mathcal{T}_h} \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} v_j^2 \mathbb{E} \left[\left| \int_T f(x) \varphi_j(x) \, dx - \frac{|T|}{3} f(Y_{T,j}) \right|^2 \right], \end{aligned}$$

due to the independence of the random variables $(Y_{T,j})_{T \in \mathcal{T}_h, j \in \{1, \dots, N_h\}}$.

Then, for every $j \in \{1, \dots, N_h\}$ and $T \in \mathcal{T}_h$ with $T \cap \text{supp}(\varphi_j) \neq \emptyset$ we make use of $Y_{T,j} \sim \frac{3}{|T|} \mathbb{I}_T(z) \varphi_j(z) \, dz$ and the Cauchy–Schwarz inequality. This yields

$$\begin{aligned} & \mathbb{E} \left[\left| \int_T f(x) \varphi_j(x) \, dx - \frac{|T|}{3} f(Y_{T,j}) \right|^2 \right] \\ &= \frac{3}{|T|} \int_T \left| \int_T f(x) \varphi_j(x) \, dx - \frac{|T|}{3} f(z) \right|^2 \varphi_j(z) \, dz \\ &= \frac{3}{|T|} \int_T \left| \int_T (f(x) - f(z)) \varphi_j(x) \, dx \right|^2 \varphi_j(z) \, dz \\ &\leq \int_T \int_T (f(x) - f(z))^2 \varphi_j(x) \varphi_j(z) \, dx \, dz \\ &= \frac{2}{3} |T| \int_T |f(x)|^2 \varphi_j(x) \, dx - 2 \left(\int_T f(x) \varphi_j(x) \, dx \right)^2. \end{aligned}$$

We neglect the last term and insert this estimate into the mean-square error. An application of Weitzenböck’s inequality (3.21) then yields

$$\begin{aligned}
 & \left\| \int_{\mathcal{D}} f(x)v_h(x) \, dx - F_{IS}(v_h) \right\|_{L^2(\Omega)}^2 \\
 &= \sum_{T \in \mathcal{T}_h} \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} v_j^2 \mathbb{E} \left[\left| \int_T f(x)\varphi_j(x) \, dx - \frac{|T|}{3} f(Y_{T,j}) \right|^2 \right] \\
 &\leq \frac{2}{3} \sum_{T \in \mathcal{T}_h} \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} v_j^2 |T| \int_T |f(x)|^2 \varphi_j(x) \, dx \\
 &\leq \frac{1}{2\sqrt{3}} h^2 \sum_{T \in \mathcal{T}_h} \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} v_j^2 \int_T |f(x)|^2 \varphi_j(x) \, dx.
 \end{aligned} \tag{4.38}$$

Now, we assume that $f \in L^p(\mathcal{D})$ with $p \in [2, \infty]$. To every $v_h = \sum_{j=1}^{N_h} v_j \varphi_j \in S_h$ we then associate a mapping $v_h^\circ: \mathcal{D} \rightarrow \mathbb{R}$ defined by $v_h^\circ(x) = \sum_{T \in \mathcal{T}_h} v_T \mathbb{I}_T(x)$, where $v_T := v_h(z_T)$ and $z_T \in T$ denotes the barycenter of $T \in \mathcal{T}_h$. Observe that v_h° is piecewise constant on each triangle.

For every $T \in \mathcal{T}_h$ and $j \in \{1, \dots, N_h\}$ with $T \cap \text{supp}(\varphi_j) \neq \emptyset$ let $z_j \in \bar{T}$ be the uniquely determined node, which satisfies $\varphi_j(z_j) = 1$. Clearly, it holds $|z_j - z_T| \leq h$. Since v_h is affine linear we obtain that

$$|v_j - v_T| = |v_h(z_j) - v_h(z_T)| \leq |\nabla v_h(z_T)|h.$$

Then, we continue the estimate of the mean-square error in (4.38) by adding and subtracting the coefficients of v_h° as follows: For $\rho = \frac{p}{2} \in [1, \infty]$ let $\rho' = \frac{\rho}{\rho-2} \in [1, \infty]$ be the conjugated Hölder exponent determined by $\frac{1}{\rho} + \frac{1}{\rho'} = 1$, where we set $\frac{1}{\infty} = 0$. Then, we get

$$\begin{aligned}
 v_j^2 &= |v_j|^{\frac{2}{\rho}} |v_j|^{\frac{2}{\rho'}} \leq \max_i |v_i|^{\frac{2}{\rho}} (|v_j - v_T| + |v_T|)^{\frac{2}{\rho'}} \\
 &\leq \|v_h\|_{L^\infty(\mathcal{D})}^{\frac{2}{\rho}} (|\nabla v_h(z_T)|h + |v_T|)^{\frac{2}{\rho'}}.
 \end{aligned}$$

After inserting this into (4.38) we obtain

$$\begin{aligned}
 & \left\| \int_{\mathcal{D}} f(x)v_h(x) \, dx - F_{IS}(v_h) \right\|_{L^2(\Omega)}^2 \\
 &\leq \frac{1}{2\sqrt{3}} h^2 \|v_h\|_{L^\infty(\mathcal{D})}^{\frac{2}{\rho}} \sum_{T \in \mathcal{T}_h} \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} (|\nabla v_h(z_T)|h + |v_T|)^{\frac{2}{\rho'}} \int_T |f(x)|^2 \varphi_j(x) \, dx
 \end{aligned}$$

$$\leq \frac{1}{2\sqrt{3}} h^2 \|v_h\|_{L^\infty(\mathcal{D})}^{\frac{2}{\rho}} \int_{\mathcal{D}} (|\nabla v_h(x)|h + |v_h^\circ(x)|)^{\frac{2}{\rho'}} |f(x)|^2 dx,$$

since ∇v_h and v_h° are constant on each T . In addition, we also made use of

$$0 \leq \sum_{j=1}^{N_h} \varphi_j(x) \leq 1 \tag{4.39}$$

for every $x \in \overline{\mathcal{D}}$.

Therefore, applications of Hölder’s inequality and Minkowski’s inequality yield

$$\begin{aligned} & \left\| \int_{\mathcal{D}} f(x)v_h(x) dx - FIS(v_h) \right\|_{L^2(\Omega)}^2 \\ & \leq \frac{1}{2\sqrt{3}} h^2 \|v_h\|_{L^\infty(\mathcal{D})}^{\frac{2}{\rho}} \|f\|_{L^p(\mathcal{D})}^2 \left(\int_{\mathcal{D}} (|\nabla v_h(x)|h + |v_h^\circ(x)|)^2 dx \right)^{\frac{1}{\rho'}} \tag{4.40} \\ & \leq \frac{1}{2\sqrt{3}} h^2 \|v_h\|_{L^\infty(\mathcal{D})}^{\frac{2}{\rho}} \|f\|_{L^p(\mathcal{D})}^2 (h|v_h|_{H^1(\mathcal{D})} + \|v_h^\circ\|_{L^2(\mathcal{D})})^{\frac{2}{\rho'}}. \end{aligned}$$

Finally, we observe that

$$\begin{aligned} \|v_h - v_h^\circ\|_{L^2(\mathcal{D})}^2 &= \sum_{T \in \mathcal{T}_h} \int_T |v_h(x) - v_T|^2 dx \\ &= \sum_{T \in \mathcal{T}_h} \int_T |\nabla v_h(x) \cdot (x - z_T)|^2 dx \leq h^2 |v_h|_{H^1(\mathcal{D})}^2 \end{aligned}$$

since $|x - z_T| \leq h$ for every $x \in T$ and ∇v_h is piecewise constant on T . Consequently,

$$\|v_h^\circ\|_{L^2(\mathcal{D})} \leq \|v_h^\circ - v_h\|_{L^2(\mathcal{D})} + \|v_h\|_{L^2(\mathcal{D})} \leq h|v_h|_{H^1(\mathcal{D})} + \|v_h\|_{L^2(\mathcal{D})}.$$

Inserting this into (4.40) then completes the proof of the second assertion.

To prove the third assertion let $f \in W^{s,2}(\mathcal{D})$, $s \in (0, 1)$. As above we have

$$\begin{aligned} & \left\| \int_{\mathcal{D}} f(x)v_h(x) dx - FIS(v_h) \right\|_{L^2(\Omega)}^2 \\ & \leq \sum_{T \in \mathcal{T}_h} \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} v_j^2 \int_T \int_T (f(x) - f(z))^2 \varphi_j(x) \varphi_j(z) dx dz \\ & \leq \max_i |v_i|^2 \sum_{T \in \mathcal{T}_h} \int_T \int_T (f(x) - f(z))^2 dx dz, \end{aligned}$$

where we also used that $\varphi_j(z) \leq 1$ for all $z \in T$ and (4.39). Moreover, since $f \in W^{s,2}(\mathcal{D})$ we get

$$\begin{aligned} \sum_{T \in \mathcal{T}_h} \int_T \int_T (f(x) - f(z))^2 \, dx \, dz &\leq h^{2(1+s)} \sum_{T \in \mathcal{T}_h} \int_T \int_T \frac{|f(x) - f(z)|^2}{|x - z|^{2+2s}} \, dx \, dz \\ &\leq h^{2(1+s)} |f|_{W^{s,2}(\mathcal{D})}^2. \end{aligned}$$

Altogether, this completes the proof of the third assertion. □

The well-posedness of (4.37) is a consequence of Lemma 5. The following lemma contains some further estimates of F_{IS} provided the family of triangulations satisfies Assumption 2.

Corollary 1 *Suppose that $f \in L^2(\mathcal{D})$. Let $(\mathcal{T}_h)_{h \in (0,1]}$ be a family of triangulations satisfying Assumption 2. Then, there exists $C \in (0, \infty)$ independent of \mathcal{T}_h such that*

$$\begin{aligned} |F_{IS}(v_h)| &\leq C \ell_h^{\frac{1}{2}} \bar{F}_{IS,h} |v_h|_{H^1(\mathcal{D})} < \infty \quad \mathbb{P}\text{-a.s.}, \\ \|F_{IS}(v_h)\|_{L^2(\Omega)} &\leq C \|f\|_{L^2(\mathcal{D})} |v_h|_{H^1(\mathcal{D})}, \end{aligned}$$

for all $v_h \in S_h$, where $\ell_h = \max(1, \log(1/h))$ and $\bar{F}_{IS,h} : \Omega \rightarrow \mathbb{R}$ is defined as

$$\bar{F}_{IS,h} := \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} |f(Y_{T,j})|.$$

Proof We only verify the almost sure bound for $F_{IS}(v_h)$. The estimate of the $L^2(\Omega)$ -norm then follows from Lemma 5 and the same arguments as in the proof Lemma 2.

By the definition of F_{IS} and an application of (2.13) we have that

$$\begin{aligned} |F_{IS}(v_h)| &\leq \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} |v_j| |f(Y_{T,j})| \\ &\leq \frac{1}{3} \|v_h\|_{L^\infty(\mathcal{D})} \sum_{T \in \mathcal{T}_h} |T| \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} |f(Y_{T,j})| \\ &\leq C \ell_h^{\frac{1}{2}} |v_h|_{H^1(\mathcal{D})} \bar{F}_{IS,h}. \end{aligned}$$

It remains to show that $\bar{F}_{IS,h}$ is bounded \mathbb{P} -almost surely. But this follows immediately from

$$\mathbb{E}[\bar{F}_{IS,h}] = \frac{1}{3} \sum_{T \in \mathcal{T}_h} |T| \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} \mathbb{E}[|f(Y_{T,j})|]$$

$$\begin{aligned}
 &= \sum_{T \in \mathcal{T}_h} \sum_{\substack{j=1 \\ T \cap \text{supp}(\varphi_j) \neq \emptyset}}^{N_h} \int_T |f(y)| \varphi_j(y) \, dy \\
 &\leq \int_{\mathcal{D}} |f(y)| \, dy < \infty,
 \end{aligned}$$

where we used that $\sum_{j=1}^{N_h} \varphi_j(y) \leq 1$ for every $y \in \mathcal{D}$. In turn, this implies $\bar{F}_{IS,h} < \infty$ \mathbb{P} -almost surely. □

4.3 Its application to the Poisson equation

To demonstrate the performance of the modified quadrature in (4.37), we solely focus on the Poisson equation

$$\begin{cases} -\Delta u = f, & \text{in } \mathcal{D}, \\ u = 0, & \text{on } \partial\mathcal{D}, \end{cases} \tag{4.41}$$

where $\mathcal{D} \subset \mathbb{R}^2$ is a convex, bounded and polygonal domain and $f \in L^p(\mathcal{D})$ for some $p \in [2, \infty]$.

Observe that the Poisson equation is a particular case of the boundary value problem (1.1) with $\sigma \equiv 1$. In this case, the assembly of the stiffness matrix A_h in (1.4) does not require the application of a (randomized) quadrature rule. Therefore the term $a(\cdot, \cdot)$ remains when introducing the corresponding FEM in (4.42).

Next, we introduce the finite element problem based on the importance sampling estimator. In terms of F_{IS} the problem is stated as follows:

$$\begin{cases} \text{Find } u_h^{IS} : \Omega \rightarrow S_h \text{ such that } \mathbb{P}\text{-almost surely} \\ a(u_h^{IS}, v_h) = F_{IS}(v_h) \text{ for all } v_h \in S_h. \end{cases} \tag{4.42}$$

In the same way as in Theorem 1 one shows that the discrete problem (4.42) to Poisson equation (4.41) has a uniquely determined solution $u_h^{IS} : \Omega \rightarrow S_h$.

Theorem 3 *For every admissible triangulation \mathcal{T}_h , $h \in (0, 1]$, there exists a uniquely determined measurable mapping $u_h^{IS} : \Omega \rightarrow S_h$ which solves the discrete problem (4.42). In addition, there exists $C \in (0, \infty)$ independent of \mathcal{T}_h such that*

$$|u_h^{IS}|_{H^1(\mathcal{D})} \leq C \ell_h^{\frac{1}{2}} \bar{F}_{IS,h} \quad \mathbb{P}\text{-a.s.},$$

where $\ell_h = \max(1, \log(1/h))$.

The following theorem contains an estimate of the total error of the approximation u_h^{IS} with respect to the $L^2(\Omega; H_0^1(\mathcal{D}))$ -norm.

Theorem 4 *Let Assumptions 1 and 2 be satisfied. If $f \in L^p(\mathcal{D})$, $p \in (2, \infty]$, then there exists $C \in (0, \infty)$ such that for every $h \in (0, 1]$*

$$\|u_h^{IS} - u\|_{L^2(\Omega; H_0^1(\mathcal{D}))} \leq Ch\|u\|_{H^2(\mathcal{D})} + C\ell_h^{\frac{1}{2} + \frac{1}{p}} h^{1 - \frac{2}{p}} \|f\|_{L^p(\mathcal{D})},$$

where $\ell_h = \max(1, \log(1/h))$.

Proof As in the proof of Theorem 2 we split the error into the two parts

$$u_h^{IS} - u = u_h^{IS} - R_h u + R_h u - u =: \theta + \rho,$$

where we recall the definition of the Ritz projector $R_h: H_0^1(\mathcal{D}) \rightarrow S_h$ from Section 2. Since the associated bilinear form a for (4.41) coincides with the inner product in $H_0^1(\mathcal{D})$ it follows that

$$\begin{aligned} |u_h^{IS} - u|_{H^1(\mathcal{D})}^2 &= a(u_h^{IS} - u, u_h^{IS} - u) = a(\theta, \theta) + a(\rho, \rho) \\ &= |\theta|_{H^1(\mathcal{D})}^2 + |\rho|_{H^1(\mathcal{D})}^2. \end{aligned}$$

Then, due to (2.16) it holds

$$|\rho|_{H^1(\mathcal{D})} = |R_h u - u|_{H^1(\mathcal{D})} \leq Ch\|u\|_{H^2(\mathcal{D})}.$$

Further, from the variational formulation of (4.41) and (4.42) we get \mathbb{P} -almost surely for every $v_h \in S_h$ that

$$\begin{aligned} a(\theta, v_h) &= a(u_h^{IS}, v_h) - a(R_h u, v_h) \\ &= F_{IS}(v_h) - F(v_h), \end{aligned}$$

since $a(R_h u, v_h) = a(u, v_h) = F(v_h)$ for every $v_h \in S_h$. In particular, for the choice $v_h = \theta(\omega) = u_h^{IS}(\omega) - R_h u \in S_h$ we obtain \mathbb{P} -almost surely that

$$|\theta|_{H^1(\mathcal{D})}^2 = a(\theta, \theta) = F_{IS}(\theta) - F(\theta).$$

From Corollary 1 and Theorem 3 it follows directly that all terms on the right-hand side are integrable with respect to \mathbb{P} . Hence, after taking expectations it remains to prove an estimate for the term

$$E_{IS} = |\mathbb{E}[F_{IS}(\theta) - F(\theta)]|.$$

This is accomplished by the same arguments as in the proof of Lemma 3. More precisely, we represent θ in terms of an orthonormal basis $(\psi_j)_{j=1}^{N_h} \subset S_h$ by

$$\theta = \sum_{j=1}^{N_h} \theta_j \psi_j,$$

where $\theta_j = (\theta, \psi_j)_{L^2(\mathcal{D})}$, $j = 1, \dots, N_h$, are real-valued and square-integrable random variables. Hereby, we assume again that $(\psi_j)_{j=1}^{N_h}$ is a solution to the discrete eigenvalue problem (3.26). Then, by the linearity of F and F_{IS} and the Cauchy–Schwarz inequality we obtain the estimate

$$E_{IS} = \left| \mathbb{E} \left[\sum_{j=1}^{N_h} \theta_j (F_{IS}(\psi_j) - F(\psi_j)) \right] \right| \leq \left(\sum_{j=1}^{N_h} \lambda_{h,j} \mathbb{E}[|\theta_j|^2] \right)^{\frac{1}{2}} \left(\sum_{j=1}^{N_h} \lambda_{h,j}^{-1} \mathbb{E}[|F_{IS}(\psi_j) - F(\psi_j)|^2] \right)^{\frac{1}{2}},$$

where $(\lambda_{h,j})_{j=1}^{N_h} \subset (0, \infty)$ denote the discrete eigenvalues in (3.26). Then, as in (3.28) one computes

$$\left(\sum_{j=1}^{N_h} \lambda_{h,j} \mathbb{E}[|\theta_j|^2] \right)^{\frac{1}{2}} = \|\theta\|_{L^2(\Omega; H_0^1(\mathcal{D}))}.$$

Moreover, since $f \in L^p(\mathcal{D})$ and $\|\psi_j\|_{L^2(\mathcal{D})} = 1$ it follows from Lemma 5 that

$$\mathbb{E}[|F_{IS}(\psi_j) - F(\psi_j)|^2] \leq \frac{1}{\sqrt{12}} h^2 \|f\|_{L^p(\mathcal{D})}^2 \|\psi_j\|_{L^\infty(\mathcal{D})}^{\frac{4}{p}} (2h|\psi_j|_{H^1(\mathcal{D})} + 1)^{2-\frac{4}{p}}.$$

Next, we recall from (2.13) and (2.14) that $\|\psi_j\|_{L^\infty(\mathcal{D})} \leq C\ell_h^{\frac{1}{2}}|\psi_j|_{H^1(\mathcal{D})} \leq C\ell_h^{\frac{1}{2}}h^{-1}$ for every $j \in \{1, \dots, N_h\}$, since $\|\psi_j\|_{L^2(\mathcal{D})} = 1$. Therefore,

$$\mathbb{E}[|F_{IS}(\psi_j) - F(\psi_j)|^2] \leq C\ell_h^{\frac{2}{p}}h^{2-\frac{4}{p}}\|f\|_{L^p(\mathcal{D})}^2$$

for some constant $C \in (0, \infty)$ independent of $h \in (0, 1]$ and $j \in \{1, \dots, N_h\}$.

Altogether, we have shown that

$$E_{IS} \leq C\ell_h^{\frac{1}{p}}h^{1-\frac{2}{p}}\|f\|_{L^p(\mathcal{D})}\|\theta\|_{L^2(\Omega; H_0^1(\mathcal{D}))} \left(\sum_{j=1}^{N_h} \lambda_{h,j}^{-1} \right)^{\frac{1}{2}}.$$

Together with (3.30) this completes the proof. □

Finally, we show an error estimate with respect to the norm in $L^2(\Omega; L^2(\mathcal{D}))$.

Theorem 5 *Let Assumptions 1 and 2 be satisfied. If $f \in W^{s,2}(\mathcal{D})$, $s \in [0, 1)$, then there exists $C \in (0, \infty)$ such that for every $h \in (0, 1]$*

$$\|u_h^{IS} - u\|_{L^2(\Omega; L^2(\mathcal{D}))} \leq Ch^2\|u\|_{H^2(\mathcal{D})} + C\ell_h h^{1+s}|f|_{W^{s,2}(\mathcal{D})},$$

where $\ell_h = \max(1, \log(1/h))$.

Proof As in the proof of Theorem 4 we again split the error into the two parts

$$u_h^{IS} - u = u_h^{IS} - R_h u + R_h u - u =: \theta + \rho.$$

Then, it follows from (2.17) that

$$\|\rho\|_{L^2(\mathcal{D})} = \|(R_h - I)u\|_{L^2(\mathcal{D})} \leq Ch^2 \|u\|_{H^2(\mathcal{D})}$$

for every $h \in (0, 1]$.

In order to give an estimate of the $L^2(\Omega; L^2(\mathcal{D}))$ -norm of θ we apply Nitsche’s duality trick. More precisely, we consider the auxiliary problem of finding a random mapping $w_h: \Omega \rightarrow S_h$ satisfying \mathbb{P} -almost surely

$$a(v_h, w_h) = (\theta, v_h)_{L^2(\mathcal{D})}, \quad \text{for all } v_h \in S_h. \tag{4.43}$$

Observe that (4.43) is a linear variational problem with a random right-hand side. The existence of a uniquely determined solution $w_h: \Omega \rightarrow S_h$ can be shown in the same way as in the proof of Theorem 1.

Testing (4.43) with $v_h = \theta(\omega) \in S_h$ then gives for \mathbb{P} -almost every $\omega \in \Omega$ that

$$\begin{aligned} \|\theta(\omega)\|_{L^2(\mathcal{D})}^2 &= a(\theta(\omega), w_h(\omega)) = a(u_h^{IS}(\omega), w_h(\omega)) - a(R_h u, w_h(\omega)) \\ &= F_{IS}(w_h(\omega)) - F(w_h(\omega)), \end{aligned}$$

where we also applied (4.42), (2.12), and (2.15). Therefore, we have

$$\|\theta\|_{L^2(\Omega; L^2(\mathcal{D}))}^2 = |\mathbb{E}[F_{IS}(w_h) - F(w_h)]|.$$

Then, as in the proof of Lemma 3 we represent w_h in terms of the orthonormal basis $(\psi_j)_{j=1}^{N_h}$ consisting of discrete eigenfunctions to the eigenvalue problem (3.26). After inserting this into the $L^2(\Omega; L^2(\mathcal{D}))$ -norm of θ , an application of the Cauchy–Schwarz inequality yields

$$\begin{aligned} \|\theta\|_{L^2(\Omega; L^2(\mathcal{D}))}^2 &= \left| \mathbb{E} \left[\sum_{j=1}^{N_h} w_j (F_{IS}(\psi_j) - F(\psi_j)) \right] \right| \\ &\leq \left(\sum_{j=1}^{N_h} \lambda_{h,j}^2 \mathbb{E}[|w_j|^2] \right)^{\frac{1}{2}} \left(\sum_{j=1}^{N_h} \lambda_{h,j}^{-2} \mathbb{E}[|F_{IS}(\psi_j) - F(\psi_j)|^2] \right)^{\frac{1}{2}}, \end{aligned}$$

where $w_j = (\psi_j, w_h)_{L^2(\mathcal{D})}$, $j \in \{1, \dots, N_h\}$, and $(\lambda_{h,j})_{j=1}^{N_h} \subset (0, \infty)$ are the discrete eigenvalues in (3.26).

Then, it follows from (3.26), (4.43) and Parseval’s identity that

$$\left(\sum_{j=1}^{N_h} \lambda_{h,j}^2 \mathbb{E}[|w_j|^2] \right)^{\frac{1}{2}} = \left(\sum_{j=1}^{N_h} \mathbb{E}[|\lambda_{h,j}(\psi_j, w_h)_{L^2(\mathcal{D})}|^2] \right)^{\frac{1}{2}}$$

$$\begin{aligned}
 &= \left(\sum_{j=1}^{N_h} \mathbb{E}[|a(\psi_j, w_h)|^2] \right)^{\frac{1}{2}} \\
 &= \left(\sum_{j=1}^{N_h} \mathbb{E}[|(\theta, \psi_j)_{L^2(\mathcal{D})}|^2] \right)^{\frac{1}{2}} = \|\theta\|_{L^2(\Omega; L^2(\mathcal{D}))}.
 \end{aligned}$$

Hence, this term can be cancelled from both sides of the inequality.

Furthermore, an application of Lemma 5 shows that

$$\mathbb{E}[|F_{IS}(\psi_j) - F(\psi_j)|^2] \leq h^{2(1+s)} \|\psi_j\|_{L^\infty(\mathcal{D})}^2 |f|_{W^{s,2}(\mathcal{D})}^2. \tag{4.44}$$

After recalling from (2.13) and (3.26) that

$$\|\psi_j\|_{L^\infty(\mathcal{D})}^2 \leq C \ell_h |\psi_j|_{H^1(\mathcal{D})}^2 = C \ell_h a(\psi_j, \psi_j) = C \ell_h \lambda_{h,j}$$

for every $j \in \{1, \dots, N_h\}$, we finally arrive at

$$\begin{aligned}
 \left(\sum_{j=1}^{N_h} \lambda_{h,j}^{-2} \mathbb{E}[|F_{IS}(\psi_j) - F(\psi_j)|^2] \right)^{\frac{1}{2}} &\leq C \ell_h^{\frac{1}{2}} h^{1+s} |f|_{W^{s,2}(\mathcal{D})} \left(\sum_{j=1}^{N_h} \lambda_{h,j}^{-1} \right)^{\frac{1}{2}} \\
 &\leq C \ell_h h^{1+s} |f|_{W^{s,2}(\mathcal{D})},
 \end{aligned}$$

where we also inserted (3.30) in the last step. Altogether, this completes the proof for $s \in (0, 1)$. The border case $s = 0$ is proven analogously. \square

Remark 2 Note that even if one applies F_{MC} defined in (3.23) instead of F_{IS} , the same rate of convergence in Theorem 5 can be obtained for solving the Poisson equation (4.41), where one may need to use Lemma 1 in (4.44). When $s = 0$, i.e., $f \in L^2(\mathcal{D})$, the coefficient of the approximation error bound of u_h^{IS} is smaller than the Monte-Carlo counterpart.

5 Implementation of the randomized quadrature formulas

This section is devoted to a brief instruction on how to implement the randomized quadrature formulas (3.19) and (4.37).

To be more precise, we apply the *general rejection algorithm* to sample the random variables $Y_{T,j} \sim p_{T,j}(x) dx$ introduced in (4.36) for each element $T \in \mathcal{T}_h$ and $j \in \{1, \dots, N_h\}$. We briefly review the rejection algorithm in Section 5.1. To simplify its implementation it is convenient to use a change of coordinates such that the sampling can be done on a fixed reference triangle. This will be discussed in detail in Section 5.2. In Section 5.3 we then show how the required samples are generated on the reference triangle using the rejection algorithm. Moreover, Section 5.4 briefly considers the uniform sampling of $Z_T \sim \mathcal{U}(T)$ on an arbitrary triangle $T \in \mathcal{T}_h$. Finally, in Section 5.5 we sketch how the randomized quadrature formula (3.19) can be embedded into the finite element method.

5.1 General rejection algorithm

In this subsection we briefly recall the general rejection algorithm for the simulation of a non-uniformly distributed random variable whose distribution is given by a probability density function. For more details on this method we refer to [29, Chapter 2.3.2].

For $d \in \mathbb{N}$ let $p: \mathbb{R}^d \rightarrow \mathbb{R}$ be a given probability density function. The goal is to generate samples of a random variable $X: \Omega \rightarrow \mathbb{R}^d$ whose distribution is given by $p(x) dx$. To this end, we assume that we already know how to generate samples of a random variable $Z: \Omega \rightarrow \mathbb{R}^d$ which is distributed according to a further probability density function $g: \mathbb{R}^d \rightarrow \mathbb{R}$. Suppose that there exists $c \in (0, \infty)$ such that

$$p(x) \leq cg(x), \quad \text{for all } x \in \mathbb{R}^d. \tag{5.45}$$

Then, the *general rejection algorithm* is given by:

1. Generate a sample $Z \sim g(x) dx$.
2. Generate a sample $Y \sim \mathcal{U}(0, c)$ independently from Z .
3. Return the value of Z if $Y \cdot g(Z) \leq p(Z)$, otherwise go back to Step 1.

It can be shown that the output of the algorithm is distributed according to the density p . Moreover, the expected number of samples of (Z, Y) needed until a value of Z is accepted is equal to c . It is therefore desirable to choose c in (5.45) as small as possible. For a proof we refer to [29, Theorem 2.15].

5.2 Transformation to a reference triangle

In this subsection we describe how to generate a sample of a random variable whose distribution depends on a specific triangle T of a given triangulation \mathcal{T}_h by making use of a transformation to a reference triangle. The same approach is widely used in practice for the assembly of the stiffness matrix (1.4) and can therefore easily be added to existing code.

We purely focus on generating samples of the random variables $Y_{T,j}$, $T \in \mathcal{T}_h$, $j \in \{1, \dots, N_h\}$, introduced in Section 4. Recall that the probability density function associated to $Y_{T,j}$ is given by

$$p_{T,j}(x) = 3|T|^{-1}\varphi_j(x)\mathbb{I}_T(x), \quad x \in \mathcal{D} \subset \mathbb{R}^2.$$

Let us fix a triangle $T \in \mathcal{T}_h$ with vertices (x_1, y_1) , (x_2, y_2) and (x_3, y_3) , such that $T \cap \text{supp}(\varphi_j) \neq \emptyset$. Without loss of generality we assume that $\varphi_j(x_1, y_1) = 1$.

We want to use the general rejection algorithm in order to generate samples of $Y_{T,j}$. However, the probability density function $p_{T,j}$ depends on the specific triangle and the basis function φ_j . Since it is inconvenient to set up the rejection method for each element and basis function separately, we will now describe in detail, how to simplify this problem by using a so called isoparametric transformation denoted by $\Gamma: T \rightarrow S_2$. Hereby, $S_2 \subset \mathbb{R}^2$ denotes the standard 2-simplex.

As illustrated in Figure 1 we denote the coordinates of a point in the given triangle T by (x, y) , while the ones in the standard 2-simplex S_2 are written as (α, β) . Then,

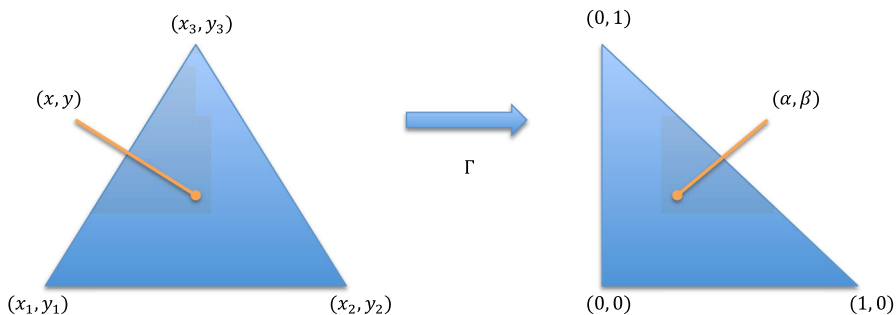


Fig. 1 Triangle transformation to the standard 2-simplex, where (x, y) and $(\alpha, \beta) = \Gamma(x, y)$ represent interior points of the respective triangles

the coordinate transformation $\Gamma : T \rightarrow S_2$ is given by

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \Gamma(x, y) := \begin{bmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{bmatrix}^{-1} \begin{bmatrix} x - x_1 \\ y - y_1 \end{bmatrix},$$

while the inverse $\Gamma^{-1} : S_2 \rightarrow T$ is explicitly determined by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \Gamma^{-1}(\alpha, \beta) := \begin{bmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}. \tag{5.46}$$

Observe that $\Gamma^{-1}(0, 0) = (x_1, y_1)$.

Next, we consider the mapping $\hat{\varphi} : S_2 \rightarrow \mathbb{R}$ defined by

$$\hat{\varphi}(\alpha, \beta) = 1 - \alpha - \beta, \quad \text{for all } (\alpha, \beta) \in S_2. \tag{5.47}$$

Since $\hat{\varphi}$ is affine linear one easily verifies that

$$\hat{\varphi}(\alpha, \beta) = \varphi_j(\Gamma^{-1}(\alpha, \beta)), \quad \text{for all } (\alpha, \beta) \in S_2.$$

Moreover, it holds

$$\int_{S_2} \hat{\varphi}(\alpha, \beta) \, d(\alpha, \beta) = \frac{1}{3} |S_2| = \frac{1}{6}.$$

Therefore, the mapping $\hat{p} : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$\hat{p}(\alpha, \beta) = 6\hat{\varphi}(\alpha, \beta)\mathbb{I}_{S_2}(\alpha, \beta), \quad \text{for } (\alpha, \beta) \in \mathbb{R}^2, \tag{5.48}$$

is a probability density function. Suppose that $\hat{Y} : \Omega \rightarrow \mathbb{R}^2$ is a random variable with distribution $\hat{p}(\alpha, \beta) \, d(\alpha, \beta)$. Then, it follows that

$$Y_{T,j} \sim \Gamma^{-1}(\hat{Y}),$$

i.e. both random variables are identically distributed with the probability density function $p_{T,j}$. In fact, for every $B \in \mathcal{B}(\mathbb{R}^2)$ it holds

$$\mathbb{P}(\{\Gamma^{-1}(\hat{Y}) \in B\}) = \mathbb{P}(\{\hat{Y} \in \Gamma(B)\}) = \int_{\Gamma(B)} \hat{p}(\alpha, \beta) \, d(\alpha, \beta).$$

After inserting \hat{p} and since $\Gamma(B) \cap S_2 = \Gamma(B \cap T)$ we arrive at

$$\begin{aligned} \mathbb{P}(\{\Gamma^{-1}(\hat{Y}) \in B\}) &= 6 \int_{\Gamma(B)} \hat{\varphi}(\alpha, \beta) \mathbb{I}_{S_2}(\alpha, \beta) \, d(\alpha, \beta) = 6 \int_{\Gamma(B \cap T)} \hat{\varphi}(\alpha, \beta) \, d(\alpha, \beta) \\ &= 6 \int_{B \cap T} \hat{\varphi}(\Gamma(x, y)) |\det(D\Gamma)(x, y)| \, d(x, y) \\ &= 6 \int_B \varphi_j(x, y) \mathbb{I}_T(x, y) |\det(D\Gamma)(x, y)| \, d(x, y) \end{aligned}$$

by a change of coordinates. Since Γ is affine linear, the Jacobian $D\Gamma \in \mathbb{R}^{2,2}$ is constant and the determinant is easily computed as

$$|\det(D\Gamma)| = |\det(D\Gamma^{-1})|^{-1} = \frac{1}{2|T|}.$$

Therefore,

$$\mathbb{P}(\{\Gamma^{-1}(\hat{Y}) \in B\}) = \frac{3}{|T|} \int_B \varphi_j(x, y) \mathbb{I}_T(x, y) \, d(x, y) = \int_B p_{T,j}(x, y) \, d(x, y).$$

Consequently, in order to generate a sample of the random variable $Y_{T,j} \sim p_{T,j}$ it is sufficient to generate a sample of $\hat{Y} \sim \hat{p}$ and to apply the transformation Γ^{-1} .

In addition, for the cases of $\varphi_j(x_2, y_2) = 1$ or $\varphi_j(x_3, y_3) = 1$, if using the same triangle transform as illustrated in Figure 1, the only step that differs from the above description is in (5.47). It needs to be changed accordingly to

$$\hat{\varphi}(\alpha, \beta) = \alpha, \quad \text{for all } (\alpha, \beta) \in S_2,$$

if $\varphi_j(x_2, y_2) = 1$, or

$$\hat{\varphi}(\alpha, \beta) = \beta, \quad \text{for all } (\alpha, \beta) \in S_2,$$

in the case of $\varphi_j(x_3, y_3) = 1$.

5.3 Generating samples of \hat{Y} on the reference triangle

Next, we discuss how to generate samples of the random variable \hat{Y} introduced in (5.48). To this end, we recall that $\hat{Y} \sim \hat{p}(\alpha, \beta) \, d(\alpha, \beta)$. We apply the general rejection

algorithm from Section 5.1 with

$$g(\alpha, \beta) = 2\mathbb{I}_{S_2}(\alpha, \beta), \quad \text{for } (\alpha, \beta) \in \mathbb{R}^2,$$

as the probability density function of the random variable Z , i.e. $Z \sim \mathcal{U}(S_2)$. As above $S_2 \subset \mathbb{R}^2$ denotes the standard 2-simplex. We also define

$$c := \sup \left\{ \frac{\hat{p}(\alpha, \beta)}{g(\alpha, \beta)} \mid (\alpha, \beta) \in S_2 \right\} = \sup \left\{ \frac{1}{2} \hat{p}(\alpha, \beta) \mid (\alpha, \beta) \in S_2 \right\} = 3.$$

Then, (5.45) is satisfied. Therefore, the general rejection algorithm is applicable and generates samples of $\hat{Y} \sim \hat{p}(\alpha, \beta) d(\alpha, \beta)$ as follows:

1. Generate $Z = (Z_1, Z_2) \sim \mathcal{U}(S_2)$ as follows:
 - (a) Generate $U_1, U_2 \sim \mathcal{U}(0, 1)$ independently.
 - (b) If $U_1 + U_2 \leq 1$ then set $Z := (U_1, U_2)$, else set $Z := (1 - U_1, 1 - U_2)$.
2. Generate $Y \sim \mathcal{U}(0, c)$ independently of Z .
3. Output $Z = (Z_1, Z_2)$ if $Yg(Z_1, Z_2) \leq \hat{p}(Z_1, Z_2)$, else go back to Step 1.

Remark 3 As an alternative to the rejection method one could generate samples of $\hat{Y} = (\hat{Y}_1, \hat{Y}_2)$ by first applying the inversion method, cf. [29, Chapter 2], for the simulation of the marginal distribution of the first variable \hat{Y}_1 . Thereafter, a further application of the inversion method can be used to generate a sample of \hat{Y}_2 conditional on the already generated sample of \hat{Y}_1 . Depending on the actual implementation, this could be more efficient. However, this approach is much harder to generalize to other probability density functions or to higher dimensional domains.

5.4 Generating uniformly distributed samples on arbitrary elements

In this subsection, we briefly discuss the generation of uniformly distributed random variables $Z_T \sim \mathcal{U}(T)$ for an arbitrary triangle $T \in \mathcal{T}_h$. These random variables are required for the randomized quadrature formula (3.19). This is easily accomplished by making use of the results from the previous two subsections. Indeed, we just have to generate a sample of a uniformly distributed random variable $Z \sim \mathcal{U}(S_2)$, where S_2 again denotes the 2-simplex. Then, we apply the corresponding inverse transformation Γ^{-1} from (5.46) associated to the given triangle $T \in \mathcal{T}_h$. As a result, we obtain $Z_T = \Gamma^{-1}(Z) \sim \mathcal{U}(T)$ for $T \in \mathcal{T}_h$.

The sampling procedure is summarized in the following two steps.

1. Generate $Z = (Z_1, Z_2) \sim \mathcal{U}(S_2)$ as follows:
 - (a) Generate $U_1, U_2 \sim \mathcal{U}(0, 1)$ independently.
 - (b) If $U_1 + U_2 \leq 1$ then set $Z := (U_1, U_2)$, else set $Z := (1 - U_1, 1 - U_2)$.
2. Output: $Z_T = \Gamma^{-1}(Z_1, Z_2)$, where Γ^{-1} in (5.46) uses the coordinates of the vertices of T .

5.5 Implementation of the FEM with randomized quadrature formulas

In this part, we illustrate the implementation of the finite element method with the randomized quadrature formula (3.19) for the elliptic equation (1.1). The implementation of (4.42), which is based on the importance sampling estimator, can be done in a similar way.

Algorithm 1 lists one possibility to compute a realization of the numerical approximation of the solution to (1.1) based on the Monte Carlo estimator (3.19).

Algorithm 1 FEM with MC estimator (3.19) for the elliptic equation (1.1)

- 1: **Input:** \mathcal{T}_h triangulation of domain \mathcal{D} , functions f and σ ;
 - 2: Get the set of interior nodes $(z_j)_{j=1}^{N_h}$ of \mathcal{T}_h with associated Lagrange basis functions $(\varphi_j)_{j=1}^{N_h}$;
 - 3: Generate $Z_T^1, Z_T^2 \sim \mathcal{U}(T)$ independently for every $T \in \mathcal{T}_h$ (see Section 5.4);
 - 4: Compute the function values $(\sigma(Z_T^1))_{T \in \mathcal{T}_h}$ and $(f(Z_T^2))_{T \in \mathcal{T}_h}$;
 - 5: Assemble the stiffness matrix A_{MC} with entries $(a_{MC}(\varphi_{k_1}, \varphi_{k_2}))_{k_1, k_2=1}^{N_h}$ based on the values $(Z_T^1)_{T \in \mathcal{T}_h}$ and $(\sigma(Z_T^1))_{T \in \mathcal{T}_h}$ as in (3.22);
 - 6: Assemble the load vector F_{MC} with entries $(F_{MC}(\varphi_k))_{k=1}^{N_h}$ based on the values $(Z_T^2)_{T \in \mathcal{T}_h}$ and $(f(Z_T^2))_{T \in \mathcal{T}_h}$ as in (3.23);
 - 7: Solve the linear equation $A_{MC}u_h^{MC} = F_{MC}$ to obtain u_h^{MC} ;
 - 8: **Output:** One realization of u_h^{MC} .
-

Observe in Step 5 that one only has to sum over those triangles in (3.22), which are contained in the joint support of the basis functions $\varphi_{k_1}, \varphi_{k_2}$. Hence, the sum in (3.22) consists of at most two non-zero terms if $k_1 \neq k_2$. In particular, the stiffness matrix A_{MC} remains sparse and the complexity of assembling A_{MC} grows only linearly with N_h . In addition, the matrix A_h remains positive definite and allows the application of linear solvers for large sparse systems as described in, e.g., [17].

6 Numerical experiments

This section is devoted to numerical experiments, illustrating the performance of the randomized quadrature formulas based on the MC estimator (3.19) and/or the IS estimator (4.37).

To this end, on the domain $\mathcal{D} = (0, 1)^2 \subset \mathbb{R}^2$, we consider in the first experiment example the Poisson equation (4.41) with homogeneous Dirichlet boundary conditions, focusing on the performance of both estimators in approximating the load vectors; and in the second one the general elliptic equations with homogeneous Dirichlet boundary conditions, focusing on the performance of the MC estimator in approximating the stiffness matrix. In each example, we also briefly compare the performance of the randomized quadrature formula with the deterministic *barycentric quadrature rule* (BQR), which is also known as a one-point Gaussian quadrature formula. We refer to [22] and [28, Section 5.6].

6.1 The general set-up

For the finite element method we choose a family of structured uniform meshes. To be more precise, the domain \mathcal{D} is first subdivided into squares with uniform mesh size $h = 2^{-n}$, $n \in \{2, \dots, 8\}$. Then, we obtain the triangulation \mathcal{T}_h by bisecting each square along the diagonal from the upper left to the lower right vertex. As in the previous sections, the shape functions are chosen to be piecewise linear. For each fixed triangulation \mathcal{T}_h we then solve the discrete problems (3.24) and/or (4.42) as sketched in Algorithm 1. As above, we denote the corresponding discrete solutions by u_h^{MC} and u_h^{IS} , respectively.

To justify the performance of the two randomized quadrature formulas, we focus on the distances (in $L^2(\Omega; H^1(\mathcal{D}))$ -semi-norm and $L^2(\Omega; L^2(\mathcal{D}))$ -norm) between the discrete solutions u_h^{MC} and/or u_h^{IS} and the standard finite element solution $u_h = R_h u$. Recall that the computation of the $H^1(\mathcal{D})$ -semi-norm is easily accomplished in practice by making use of the relationship

$$|v_h|_{H^1(\mathcal{D})}^2 = a(v_h, v_h) = \sum_{i,j=1}^{N_h} v_i v_j a(\varphi_i, \varphi_j) = \mathbf{v}^\top A_h \mathbf{v}, \quad (6.49)$$

for every $v_h = \sum_{j=1}^{N_h} v_j \varphi_j \in S_h$ with $\mathbf{v} = [v_1, \dots, v_{N_h}]^\top \in \mathbb{R}^{N_h}$. Similarly, $|v_h|_{L^2(\mathcal{D})}^2$ is obtained if the stiffness matrix A_h is replaced by the mass matrix $M_h = [(\varphi_i, \varphi_j)_{L^2(\mathcal{D})}]_{i,j=1}^{N_h}$ in (6.49). Based on those observations, one can get Monte Carlo approximations of the errors $\|u_h^{\text{MC}} - u_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}$ and $\|u_h^{\text{IS}} - u_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}$, achieved by generating $M = 10^4$ independent realizations of the random variables u_h^{MC} and u_h^{IS} and taking suitable averages.

6.2 Example: the Poisson equation with different forcing terms

In this example, we consider the Poisson equation which satisfies (1.3) with $\sigma \equiv 1$ and is subject to different forcing terms f_1 or f_2 . The first forcing term f_1 is singular but still square-integrable, defined by

$$f_1(x, y) := |x - y|^{-q} + 10 \sin(2^3 \pi x) \operatorname{sgn}(2y - x), \quad \text{for } (x, y) \in \mathcal{D}, \quad (6.50)$$

with $q = 0.49$ and $\operatorname{sgn}: \mathbb{R} \rightarrow \mathbb{R}$ given by

$$\operatorname{sgn}(x) := \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases}$$

The second forcing term $f_2: \mathcal{D} \rightarrow \mathbb{R}$ is taken more regular by setting

$$f_2(x, y) := 8x(1-x)y(1-y), \quad \text{for } (x, y) \in \mathcal{D}. \quad (6.51)$$

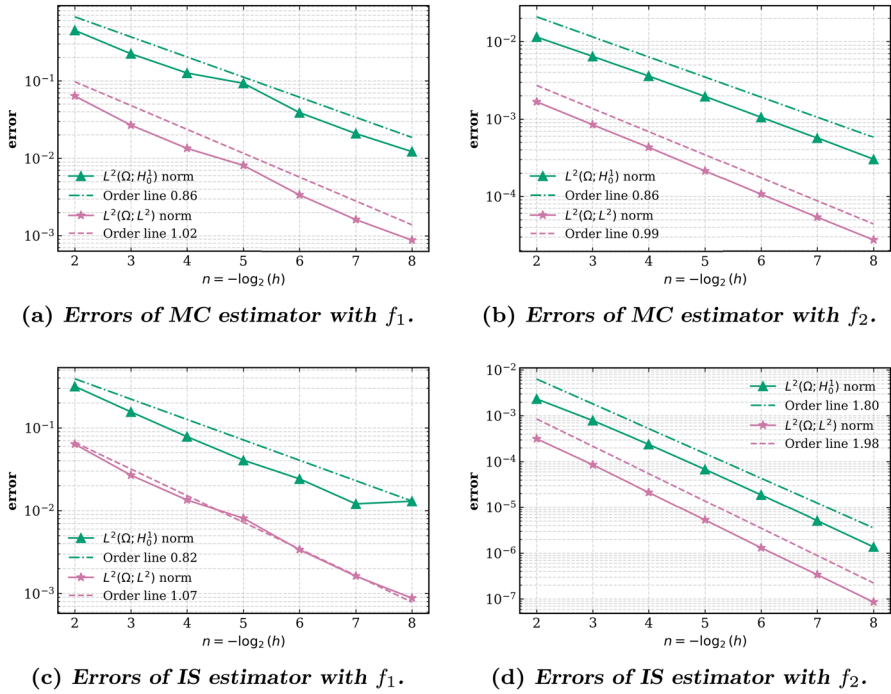


Fig. 2 Error plots of the MC estimator (3.19) and IS estimator (4.37) for the Poisson equation (4.41) with singular forcing term f_1 and smooth forcing term f_2

In fact, it can be easily verified that $f_2 \in H_0^1(\mathcal{D}) \cap H^2(\mathcal{D})$.

Considering Poisson equation allows us to neglect the approximation error $u_h - u$ stemming from the finite element method itself. More precisely, we first take note of the fact that $\mathbb{E}[u_h^{MC}] = \mathbb{E}[u_h^{IS}] = u_h$. In fact, since $\sigma \equiv 1$ we have that $a_{MC} = a$ in (3.24). Hence, after taking expectation in (3.24) and since Q_{MC} is unbiased we obtain that

$$a(\mathbb{E}[u_h^{MC}], v_h) = \mathbb{E}[a_{MC}(u_h^{MC}, v_h)] = \mathbb{E}[F_{MC}(v_h)] = F(v_h)$$

for every $v_h \in S_h$. Therefore, the function $\mathbb{E}[u_h^{MC}] \in S_h$ is a solution to (1.3), i.e. $\mathbb{E}[u_h^{MC}] = u_h$ for every $h \in (0, 1]$. The same arguments apply to $\mathbb{E}[u_h^{IS}]$. This motivates to replace u_h in the error computation by the Monte Carlo means

$$u_h \approx \frac{1}{M} \sum_{i=1}^M u_{h,i}^{MC}, \quad \text{and} \quad u_h \approx \frac{1}{M} \sum_{i=1}^M u_{h,i}^{IS},$$

where $(u_{h,i}^{MC})_{i=1}^M$ and $(u_{h,i}^{IS})_{i=1}^M$ denote families of independent and identically distributed copies of u_h^{MC} and u_h^{IS} , respectively.

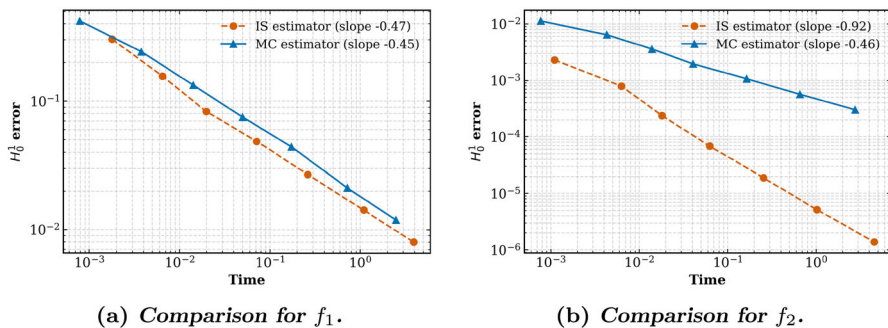


Fig. 3 Computational time versus errors in $L^2(\Omega; H_0^1(\mathcal{D}))$ -norm of the MC estimator (3.19) and IS estimator (4.37) with singular forcing term f_1 and smooth forcing term f_2

Figure 2 shows the results of experiments using MC estimator and IS estimator. In each of the four subfigures the Monte Carlo approximations of the $L^2(\Omega; H_0^1(\mathcal{D}))$ -norm and the $L^2(\Omega; L^2(\mathcal{D}))$ -norm of the errors $u_h^{MC} - u_h$ and $u_h^{IS} - u_h$ are plotted versus the mesh size $h = 2^{-n}$, $n \in \{2, \dots, 8\}$. Hereby, the first two subfigures show the corresponding errors for the MC estimator (3.19) applied to the Poisson equation with the forcing terms f_1 and f_2 defined in (6.50) and (6.51), respectively. As it can be seen from the order lines, the errors decay approximately with orders roughly 0.86 and 1. Given that f_1 is singular and only square-integrable, the experimental order of convergence is therefore larger than it is predicted by Theorem 2.

In Figures 2 (c) and (d) we see the corresponding results for the IS estimator (4.37). While the values in Figure 2 (c) are comparable to those in Figure 2 (a), it can be seen from Figure 2 (d) that the IS estimator benefits considerably from the additional smoothness of f_2 . In fact, the experimental order of convergence is close to 2 in Figure 2 (d), which is in line with the results in Theorem 5.

In Figure 3, we plot the estimated values of the errors in the $L^2(\Omega; H_0^1(\mathcal{D}))$ -norm versus the computational time. This allows a better comparison of the performance of the two randomized quadrature rules since the IS estimator is computationally more expensive due to the application of the general rejection method. Hereby, the computational time is taken as the average time needed to assemble the load vector $f_h \in \mathbb{R}^{N_h}$ for f_1 or f_2 with either (3.19) or (4.37). More precisely, we only measured the time of Step 6 in Algorithm 1. The other steps are neglected, since they are essentially independent of the choice of the randomized quadrature formula.

As it can be seen in both subfigures, the importance sampling estimator (4.37) is superior to the MC estimator. For both forcing terms the higher computational cost is offset by the better accuracy of the IS estimator (4.37). In particular, this is true for the smooth forcing term f_2 due to the better experimental order of convergence of (4.37). On the other hand, it is not very pronounced for the singular forcing term f_1 as can be seen in Figure 3 (a).

Finally, let us also briefly compare the performance of the randomized quadrature formula with BQR. Table 1 lists the corresponding estimates of the errors stemming from the application of the deterministic quadrature rule. Hereby, the errors are mea-

Table 1 Discretization errors of the (deterministic) barycentric quadrature rule applied to (4.41) with f_1

mesh size h^{-n}	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
error in $H_0^1(\mathcal{D})$ -norm	1.4e+6	7.7e+5	4.0e+5	2.1e+5	1.0e+5	5.2e+4

sured with respect to the semi-norm in $H^1(\mathcal{D})$. Apparently, BQR is not useful for approximating the load vector involving the singular forcing term f_1 .

This is easily explained by the geometry of the triangulation \mathcal{T}_h . For every mesh size $h = 2^{-n}$ there always exist triangles in \mathcal{T}_h whose barycenters lie on the diagonal in \mathcal{D} , where f_1 is singular. To avoid NaN entries in the load vector we replaced f_1 by the modification

$$\tilde{f}_1(x, y) := (\text{eps} + |x - y|)^{-q} + 10 \sin(2^3 \pi x) \text{sgn}(2y - x), \quad \text{for } (x, y) \in \mathcal{D},$$

where eps is equal to the machine precision (in MATLAB[©] $\text{eps} \approx 2.2204 \times 10^{-16}$). Nevertheless, the discretization errors indicate that BQR is not reliable for applications with singular forcing terms. This can only be circumvented by adapting the mesh to avoid point evaluations close to singularities of the given forcing term. However, this requires a priori knowledge of the position of the singularities or adaptive methods for their automatic detection when generating the mesh. The randomized quadrature formulas, on the other hand, lead to a robustification of the finite element method based on rudimentary uniform meshes without using any preknowledge of the forcing term.

6.3 Example: elliptic equations with different σ

In this example, we introduce two σ functions with different regularity, both of which are in the form of

$$\sigma(x, y; \theta, c) := \sigma_1(x, y; \theta_1) \sigma_2(x, y; \theta_2, \theta_3, \theta_4) + c,$$

and consider a smooth forcing term f_2 defined in (6.51). Here the sigmoid function $0 \leq \sigma_1(x, y; \theta_1) \leq 50$ is to introduce a smooth or sharp gradient at the interface $x + y - 1$, defined as

$$\sigma_1(x, y; \theta_1) = \frac{50}{1 + e^{-\theta_1(x+y-1)}},$$

for a parameter $\theta_1 > 0$ that controls the magnitude of the gradient of σ_1 at the interface, and a decaying ripple feature is captured in

$$\sigma_2(x, y; \theta_2, \theta_3, \theta_4) = \left| \cos[2\pi(\theta_2 x - \theta_3 x^2 - \theta_4 y^2)] \right|,$$

for some positive parameters $\theta_2, \theta_3, \theta_4$ controlling the initial frequency of the ripple, its decaying rate, and rotation respectively. Let $c = 1$ be an offset to impose positivity

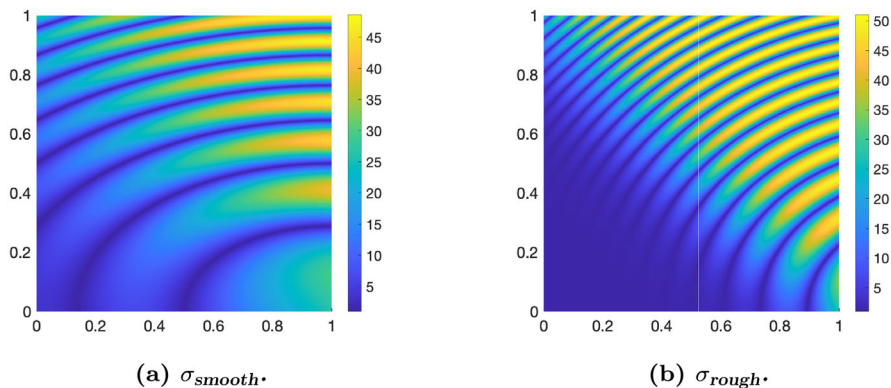


Fig. 4 The profiles of two $\sigma(x, y)$ functions used in the simulation of the elliptic equation at resolution $n = 8$

in the function, then a smooth sigma can be obtained by

$$\sigma_{\text{smooth}}(x, y) := \sigma(x, y; \theta = (3, 2, 1, 3))$$

while another realisation with a more rough profile - higher gradient magnitudes profile can be obtained by

$$\sigma_{\text{rough}}(x, y) := \sigma(x, y; \theta = (9, 8, 3, 5))$$

as depicted in figure 4.

For each combination of σ (from σ_{rough} and σ_{smooth}) and $f = f_2$ we apply the Monte Carlo estimator (3.19) and solve the finite element problem defined in (3.24), using the classical FEM with the BQR as a benchmark. As the gradient of basis φ_i being constant on triangle elements, the difference in approximating the stiffness matrix in (1.5) is highlighted by the estimator’s ability to approximate integrals of σ functions. That is,

$$\begin{aligned} \int_{\mathcal{D}} \sigma(x) \nabla \varphi_i(x) \cdot \nabla \varphi_j(x) \, dx &= \sum_{T \in \mathcal{T}} \int_T \sigma(x) \nabla \varphi_i(x) \cdot \nabla \varphi_j(x) \, dx \\ &= \sum_{T \in \mathcal{T}} (\nabla \varphi_i(C_T) \cdot \nabla \varphi_j(C_T)) \int_T \sigma(x) \, dx, \end{aligned}$$

where C_T is the centroid of T . To approximate $\int_T \sigma(x) \, dx$, both estimators use single value evaluation $|T|\sigma(x_T)$: the MC estimator (3.19) uses a random sample on T as x_T while the BQR uses $x_T = C_T$.

Figure 5 demonstrate the results of experiments using MC estimator (3.19) with varying regularity of σ . In each of subfigures the Monte Carlo approximations of the $L^2(\Omega; H_0^1(\mathcal{D}))$ -norm and the $L^2(\Omega; L^2(\mathcal{D}))$ -norm of the errors $u_h^{\text{MC}} - u_h$ are plotted versus the mesh size $h = 2^{-n}$, $n \in \{2, \dots, 8\}$. The convergence plots illustrate that the

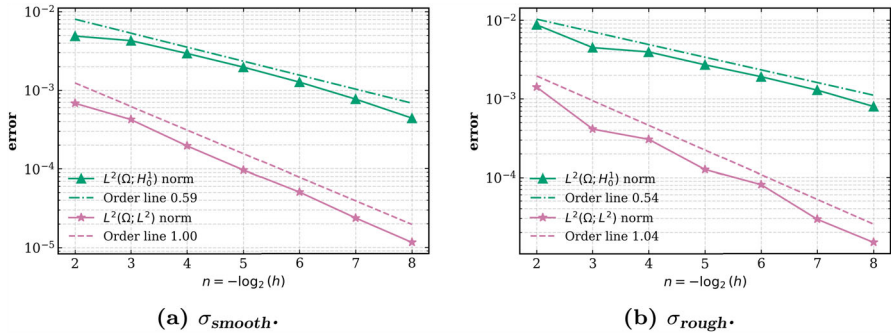


Fig. 5 Error plots of the MC estimator (3.19) for the Elliptic equation (1.1) with f_2 and different choices for σ

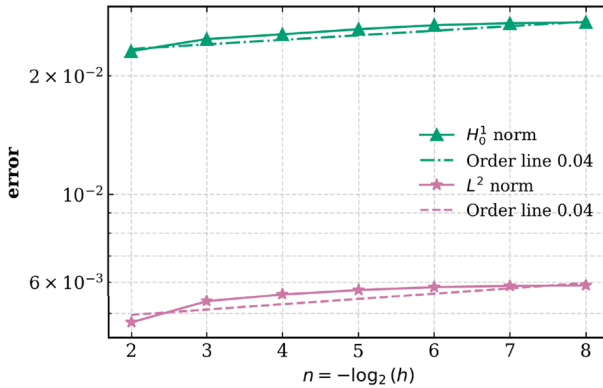


Fig. 6 Error plot of the BQR for the Elliptic equation (1.1) with f_2 and σ_{smooth}

numerical scheme achieves robust convergence rates for both coefficients, exhibiting an order-one convergence in the $L^2(\Omega; L^2(\mathcal{D}))$ norm and an order-half convergence in the $L^2(\Omega; H_0^1(\mathcal{D}))$ norm, where the first one is in general an open question for the future research. While the convergence order is preserved, the error magnitude clearly indicates that the rough coefficient, σ_{rough} , results in a significantly higher magnitude of the error across both norms compared to the smooth coefficient, σ_{smooth} , for any fixed mesh resolution. For σ_{rough} , a crucial observation is the imperfect alignment of the measured error points with the theoretical order line, indicating a greater influence of pre-asymptotic behavior and higher error constants C compared to the σ_{smooth} case.

For both cases, BQR shows divergence. Figure 6 demonstrates that employing BQR severely degrades the performance of the numerical scheme, even with smooth coefficients. This degradation occurs because the low precision of BQR introduces a quadrature error that dominates the total discretization error, preventing the method from achieving its full theoretical convergence orders in both the solution and the gradient norms.

An alternative estimate for Lemma 3

Lemma 6 *Let Assumption 2 be satisfied. Then there exists $C \in (0, \infty)$ such that for every $h \in (0, 1]$, $f \in L^2(\mathcal{D})$ and S_h -valued random variable $v_h \in L^2(\Omega; H_0^1(\mathcal{D}))$ it holds*

$$|\mathbb{E}[F_{MC}(v_h) - F(v_h)]| \leq \|f\|_{L^2(\mathcal{D})} (2h \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))} + \|v_h - \mathbb{E}[v_h]\|_{L^2(\Omega; L^2(\mathcal{D}))}).$$

Proof First, we collect some observations: Due to Lemma 1 we have

$$\mathbb{E}[F(v_h)] = F(\mathbb{E}[v_h]) = \mathbb{E}[F_{MC}(\mathbb{E}[v_h])].$$

Hence, we have

$$\begin{aligned} |\mathbb{E}[F_{MC}(v_h) - F(v_h)]| &= |\mathbb{E}[F_{MC}(v_h) - F_{MC}(\mathbb{E}[v_h])]| \\ &= \left| \sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[f(Z_T)(v_h(Z_T) - \mathbb{E}[v_h(Z_T)])] \right| \\ &\leq \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|f(Z_T)|^2] \right)^{\frac{1}{2}} \\ &\quad \times \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(Z_T) - \mathbb{E}[v_h(Z_T)]|^2] \right)^{\frac{1}{2}}, \end{aligned}$$

where we also applied the Cauchy–Schwarz inequality. Next, let us observe that

$$\left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|f(Z_T)|^2] \right)^{\frac{1}{2}} = \|f\|_{L^2(\mathcal{D})}.$$

By $C_T \in T$ we denote the barycenter of the triangle T . Then, an application of the triangle inequality yields

$$\begin{aligned} &\left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(Z_T) - \mathbb{E}[v_h(Z_T)]|^2] \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(Z_T) - v_h(C_T)|^2] \right)^{\frac{1}{2}} \\ &\quad + \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(C_T) - \mathbb{E}[v_h(C_T)]|^2] \right)^{\frac{1}{2}} \\ &\quad + \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(C_T) - v_h(Z_T)|^2] \right)^{\frac{1}{2}} \\ &\leq 2 \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(Z_T) - v_h(C_T)|^2] \right)^{\frac{1}{2}} \end{aligned}$$

$$\begin{aligned}
 & + \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(C_T) - \mathbb{E}[v_h(C_T)]|^2] \right)^{\frac{1}{2}} \\
 & =: \Gamma_1 + \Gamma_2.
 \end{aligned}$$

For Γ_1 we recall that $v_h \in L^2(\Omega; S_h)$. Thus, $x \mapsto \nabla v_h(x)$ is constant on each triangle almost surely. In addition, we have $|Z_T - C_T| \leq h$. Together, this yields

$$\begin{aligned}
 \Gamma_1 & = 2 \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(Z_T) - v_h(C_T)|^2] \right)^{\frac{1}{2}} \\
 & = 2 \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|\nabla v_h(C_T) \cdot (Z_T - C_T)|^2] \right)^{\frac{1}{2}} \\
 & \leq 2h \|v_h\|_{L^2(\Omega; H_0^1(\mathcal{D}))}.
 \end{aligned}$$

For Γ_2 we make use of the fact that

$$\int_T w_h(x) \, dx = |T| w_h(C_T) \tag{1.52}$$

which holds for every affine-linear mapping w_h on T . From this it follows that

$$v_h(C_T) - \mathbb{E}[v_h(C_T)] = |T|^{-1} \int_T (v_h(x) - \mathbb{E}[v_h(x)]) \, dx.$$

Therefore, it holds

$$\begin{aligned}
 \Gamma_2 & = \left(\sum_{T \in \mathcal{T}_h} |T| \mathbb{E}[|v_h(C_T) - \mathbb{E}[v_h(C_T)]|^2] \right)^{\frac{1}{2}} \\
 & = \left(\sum_{T \in \mathcal{T}_h} |T|^{-1} \mathbb{E} \left[\left| \int_T (v_h(x) - \mathbb{E}[v_h(x)]) \, dx \right|^2 \right] \right)^{\frac{1}{2}} \\
 & \leq \left(\sum_{T \in \mathcal{T}_h} \mathbb{E} \left[\int_T |v_h(x) - \mathbb{E}[v_h(x)]|^2 \, dx \right] \right)^{\frac{1}{2}} \\
 & = \|v_h - \mathbb{E}[v_h]\|_{L^2(\Omega; L^2(\mathcal{D}))}.
 \end{aligned}$$

□

Remark 4 The new version of Lemma 6 is only useful, if we can prove an estimate for

$$\|u_h^{MC} - \mathbb{E}[u_h^{MC}]\|_{L^2(\Omega; L^2(\mathcal{D}))}.$$

In Theorem 2 we apply Lemma 3 with $v_h = \theta = u_h^{MC} - R_h u$.

Acknowledgements The authors like to thank Monika Eisenmann for helpful comments. NP and YW are grateful to EPSRC for funding this work through the project EP/R041431/1, titled ‘Randomness: a resource for real-time analytics’. NP acknowledges funding support from UKRI EPSRC under grant EP/V028618/1. YW would acknowledge EPSRC project EP/S026347/1, titled ‘Unparameterised multi-modal data, high order signatures, and the mathematics of data science’, as well as Alan Turing Institute, for travel support.

Declarations

Conflicts of Interest None of the authors have a conflict of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces. Pure and Applied Mathematics, vol. 140, 2nd edn. Elsevier/Academic Press, Amsterdam (2003)
2. Barth, A., Stein, A.: A study of elliptic partial differential equations with jump diffusion coefficients. *SIAM/ASA J. Uncertain. Quantif.* **6**(4), 1707–1743 (2018)
3. Bauer, H.: Measure and Integration Theory. de Gruyter Studies in Mathematics, vol. 26. Walter de Gruyter & Co., Berlin (2001). Translated from the German by Robert B. Burckel
4. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods. Texts in Applied Mathematics, vol. 15, 3rd edn. Springer, New York (2008)
5. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext. Springer, New York (2011)
6. Cambanis, S., Masry, E.: Trapezoidal stratified Monte Carlo integration. *SIAM J. Numer. Anal.* **29**(1), 284–301 (1992)
7. Cohn, D.L.: Measure Theory. Birkhäuser Advanced Texts: Basler Lehrbücher, 2nd edn. Birkhäuser/Springer, New York (2013)
8. Daun, T.: On the randomized solution of initial value problems. *J. Complex.* **27**(3–4), 300–311 (2011)
9. Di Nezza, E., Palatucci, G., Valdinoci, E.: Hitchhiker’s guide to the fractional Sobolev spaces. *Bull. Sci. Math.* **136**(5), 521–573 (2012)
10. Eisenmann, M., Kovács, M., Kruse, R., Larsson, S.: On a randomized backward Euler method for nonlinear evolution equations with time-irregular coefficients. *Found. Comput. Math.* **19**(6), 1387–1430 (2019). (**Online first**)
11. Evans, L.C.: Partial Differential Equations. Graduate Studies in Mathematics, vol. 19, 2nd edn. American Mathematical Society, Providence, RI (2010)
12. Evans, M., Swartz, T.: Approximating Integrals via Monte Carlo and Deterministic Methods. Oxford Statistical Science Series. Oxford University Press, Oxford (2000)
13. Grisvard, P.: Elliptic Problems in Nonsmooth Domains. Classics in Applied Mathematics, vol. 69. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2011). Reprint of the 1985 original, With a foreword by S. C. Brenner
14. Haber, S.: A modified Monte-Carlo quadrature. *Math. Comp.* **20**, 361–368 (1966)
15. Haber, S.: A modified Monte-Carlo quadrature. II. *Math. Comp.* **21**, 388–397 (1967)
16. Haber, S.: Stochastic quadrature formulas. *Math. Comp.* **23**, 751–764 (1969)
17. Hackbusch, W.: Iterative Solution of Large Sparse Systems of Equations, vol. 95, 2nd edn. Springer-Verlag, Cham (2016)
18. Heinrich, S.: The randomized information complexity of elliptic PDE. *J. Complex.* **22**(2), 220–249 (2006)

19. Heinrich, S., Milla, B.: The randomized complexity of initial value problems. *J. Complex.* **24**(2), 77–88 (2008)
20. Hofmanová, M., Knöller, M., Schratz, K.: Stratified exponential integrator for modulated nonlinear Schrödinger equations. Preprint at [arXiv:1711.01091](https://arxiv.org/abs/1711.01091) (2017)
21. Jentzen, A., Neuenkirch, A.: A random Euler scheme for Carathéodory differential equations. *J. Comput. Appl. Math.* **224**(1), 346–359 (2009)
22. Jin, J.-M.: *The Finite Element Method in Electromagnetics*, 3rd edn. John Wiley & Sons, New York (2015)
23. Kallenberg, O.: *Foundations of Modern Probability. Probability and its Applications*, 2nd edn. Springer-Verlag, New York (2002)
24. Klenke, A.: *Probability Theory: A Comprehensive Course*, 2nd edn. Universitext. Springer, London (2014)
25. Kruse, R., Wu, Y.: Error analysis of randomized Runge-Kutta methods for differential equations with time-irregular coefficients. *Comput. Methods Appl. Math.* **17**(3), 479–498 (2017)
26. Kruse, R., Wu, Y.: A randomized Milstein method for stochastic differential equations with non-differentiable drift coefficients. *Discrete Contin. Dyn. Syst. Ser. B* **24**(8), 3475–3502 (2019)
27. Larson, M.G., Bengzon, F.: *The Finite Element Method: Theory, Implementation, and Applications. Texts in Computational Science and Engineering*, vol. 10. Springer, Heidelberg (2013)
28. Larsson, S., Thomée, V.: *Partial Differential Equations with Numerical Methods, Texts in Applied Mathematics*, vol. 45. Springer-Verlag, Berlin (2009). Paperback reprint of the 2003 edition
29. Madras, N.: *Lectures on Monte Carlo Methods. Fields Institute Monographs*, vol. 16. American Mathematical Society, Providence, RI (2002)
30. Masry, E., Cambanis, S.: Trapezoidal Monte Carlo integration. *SIAM J. Numer. Anal.* **27**(1), 225–246 (1990)
31. Müller-Gronbach, T., Novak, E., Ritter, K.: *Monte Carlo-Algorithmen. Springer-Lehrbuch. Springer-Verlag, Heidelberg* (2012)
32. Przybyłowicz, P., Morkisz, P.: Strong approximation of solutions of stochastic differential equations with time-irregular coefficients via randomized Euler algorithm. *Appl. Numer. Math.* **78**, 80–94 (2014)
33. Roubíček, T.: *Nonlinear Partial Differential Equations with Applications. International Series of Numerical Mathematics*, vol. 153, 2nd edn. Birkhäuser/Springer Basel AG, Basel (2013)
34. Stengle, G.: Numerical methods for systems with measurable coefficients. *Appl. Math. Lett.* **3**(4), 25–29 (1990)
35. Stengle, G.: Error analysis of a randomized numerical method. *Numer. Math.* **70**(1), 119–128 (1995)
36. Strang, G., Fix, G.J.: *An Analysis of the Finite Element Method. Prentice-Hall Series in Automatic Computation. Prentice-Hall Inc, Englewood Cliffs, N. J.* (1973)
37. Thomée, V.: *Galerkin Finite Element Methods for Parabolic Problems. Springer Series in Computational Mathematics*, vol. 25, 2nd edn. Springer-Verlag, Berlin (2006)
38. Weitzenböck, R.: Über eine Ungleichung in der Dreiecksgeometrie. *Math. Z.* **5**(1–2), 137–146 (1919)