MOX–Report No. 36/2011

A stochastic collocation method for the second order wave equation with a discontinuous random speed

Motamed, M.; Nobile, F.; Tempone, R.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

A stochastic collocation method for the second order wave equation with a discontinuous random speed*

Mohammad Motamed[1], Fabio Nobile[2,3], Raúl Tempone[1]

November 7, 2011

[1] Applied Mathematics and Computational Science
4700 - King Abdullah University of Science and Technology
Thuwal 23955-6900, Kingdom of Saudi Arabia
mohammad.motamed@kaust.edu.sa, raul.tempone@kaust.edu.sa

[2] MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
fabio.nobile@polimi.it

[3] CSQI – MATHICSE
École Polytechnique Fédérale de Lausanne (EPFL)
Station 8, CH-1015 Lausanne, Switzerland
fabio.nobile@epfl.ch

Keywords: Stochastic partial differential equations, Wave equation, Collocation method, Finite differences, Finite elements, Uncertainty quantification, Error analysis

AMS Subject Classification: 65C20, 65M70, 65M60, 65M06, 65M15, 65Z05

Abstract

In this paper we propose and analyze a stochastic collocation method for solving the second order wave equation with a random wave speed and

*This work was supported by the King Abdullah University of Science and Technology (AEA project "Bayesian earthquake source validation for ground motion simulation"), the VR project "Effektiva numeriska metoder för stokastiska differentialekvationer med tillämpningar", and the PECOS center at ICES, University of Texas at Austin (Project Number 024550, Center for Predictive Computational Science). The second author was partially supported by the Italian grant FIRB-IDEAS (Project n. RBID08223Z) "Advanced numerical techniques for uncertainty quantification in engineering and life science problems".

subjected to deterministic boundary and initial conditions. The speed is piecewise smooth in the physical space and depends on a finite number of random variables. The numerical scheme consists of a finite difference or finite element method in the physical space and a collocation in the zeros of suitable tensor product orthogonal polynomials (Gauss points) in the probability space. This approach leads to the solution of uncoupled deterministic problems as in the Monte Carlo method. We consider both full and sparse tensor product spaces of orthogonal polynomials. We provide a rigorous convergence analysis and demonstrate different types of convergence of the probability error with respect to the number of collocation points for full and sparse tensor product spaces and under some regularity assumptions on the data. In particular, we show that, unlike in elliptic and parabolic problems, the solution to hyperbolic problems is not in general analytic with respect to the random variables. Therefore, the rate of convergence may only be algebraic. An exponential/fast rate of convergence is still possible for some quantities of interest and for the wave solution with particular types of data. We present numerical examples, which confirm the analysis and show that the collocation method is a valid alternative to the more traditional Monte Carlo method for this class of problems.

1 Introduction

Partial differential equations (PDEs) are important mathematical models for multidimensional physical systems. There is an increasing interest in including uncertainty in these models and quantifying its effects on the predicted solution or other quantities of physical interest. The uncertainty may be due to an intrinsic variability of the physical system. It may also reflect our ignorance or inability to accurately characterize all input data of the mathematical model. Examples include the variability of soil permeability in subsurface aquifers and heterogeneity of materials with microstructure.

Probability theory offers a natural framework to describe uncertainty by parametrizing the input data either in terms of a finite number of random variables or more generally by random fields. Random fields can in turn be accurately approximated by a finite number of random variables when the input data vary slowly in space, with a correlation length comparable to the size of the physical domain. A possible way to describe such random fields is to use the truncated Karhunen-Loéve [24, 25] or polynomial chaos expansion [39, 42].

There are different techniques for solving PDEs in probabilistic setting. The most popular one is the Monte Carlo sampling, see for instance [10]. It consists in generating independent realizations drawn from the input distribution and then computing sample statistics of the corresponding output val for ues. This allows one to reuse available deterministic solvers. While being very flexible and easy to implement, this technique features a very slow convergence rate.

In the last few years, other approaches have been proposed, which in certain situations feature a much faster convergence rate. They exploit the possible

regularity that the solution might have with respect to the input parameters, which opens up the possibility to use deterministic approximations of the response function (i.e. the solution of the problem as a function of the input parameters) based on global polynomials. Such approximations are expected to yield a very fast convergence. Stochastic Galerkin [11, 26, 41, 3, 35] and Stochastic Collocation [1, 29, 28, 40] are among these techniques.

Such new techniques have been successfully applied to stochastic elliptic and parabolic PDEs. In particular, we have shown in previous works [1, 27] that, under particular assumptions, the solution of these problems is analytic with respect to the input random variables. The convergence results are then derived from the regularity results. For stochastic hyperbolic problems, the analysis is not well developed. In the case of linear problems, there are a few works on the one-dimensional scalar advection equation with a time- and space-independent random wave speed [38, 12, 34]. Such problems also possess high regularity properties provided the data live in suitable spaces. The main difficulty arises when the coefficients vary in space or time. In this more general case, the solution of linear hyperbolic problems may have lower regularity than those of elliptic, parabolic and hyperbolic problems with constant random coefficients. There are also recent works on stochastic nonlinear conservation laws, see for instance [22, 23, 31, 36, 37].

In this paper, we consider the linear second order scalar wave equation with a piecewise smooth random wave speed. In many applications, such as seismology, acoustics, electromagnetism and general relativity, the underlying differential equations are systems of second order hyperbolic PDEs. In deterministic problems, these systems are often rewritten as first order systems and then discretized. This approach has the disadvantage of introducing auxiliary variables with their associated constraints and boundary conditions. This in turn reduces computational efficiency and accuracy [18, 17]. Here, we analyze the problem in the second order differential form, without reducing it to the first order form, and propose a numerical method that directly discretizes the second order PDE. The analysis of the first order and other types of second order hyperbolic systems with discontinuous random coefficients will be addressed elsewhere.

We propose a stochastic collocation method for solving the wave propagation problem in a medium consisting of non-overlapping sub-domains. In each sud-domain, the wave speed is smooth and is given in terms of one random variable. We assume that the interfaces of speed discontinuity are smooth. We derive a priori error estimates with respect to the number of collocation points. The main result is that unlike in elliptic and parabolic problems, the solution to hyperbolic problems is not in general analytic with respect to the random variables. Therefore, the convergence rate of error in the wave solution may only be algebraic. A fast spectral convergence is still possible for some linear quantities of interest with smooth mollifiers and for the wave solution with smooth data compactly supported within sub-domains. We also show that the semi-discrete solution is analytic with respect to the random variables with the radius of an-

alyticity proportional to the mesh size h. We therefore obtain an exponential rate of convergence which deteriorates as the quantity $h\,p$ gets smaller, with p representing the polynomial degree in the stochastic space.

The outline of the paper is as follows: in Sect. 2 we formulate the mathematical problem, prove its well-posedness, and provide regularity results on the solution and a quantity of interest. The collocation method for solving the underlying stochastic PDE is described in Sect. 3. In Sect. 4 we give a complete error analysis for the collocation method and obtain convergence results. In Sect. 5 we perform some numerical examples to illustrate the accuracy and efficiency of the method. Finally, we present our conclusions in Sect. 6.

2 Mathematical Setting

We consider the linear second order scalar wave equation with a discontinuous random wave speed and deterministic boundary and initial conditions. We study the well-posedness of the problem and regularity of the solution and a quantity of interest with respect to the input random parameters.

2.1 Problem statement

Let D be a convex bounded polygonal domain in $\mathbb{R}^d$ and $(\Omega, \mathcal{F}, P)$ be a complete probability space. Here, Ω is the set of outcomes, $\mathcal{F} \subset 2^\Omega$ is the σ-algebra of events and $P : \mathcal{F} \to [0,1]$ is a probability measure. Consider the stochastic initial boundary value problem (IBVP): find a random function $u : [0,T] \times \bar{D} \times \Omega \to \mathbb{R}$, such that P-almost everywhere in Ω, i.e. almost surely (a.s), the following holds

$$\begin{aligned} &u_{tt}(t,\mathbf{x},\omega) - \nabla \cdot \big(a^2(\mathbf{x},\omega)\nabla u(t,\mathbf{x},\omega)\big) = f(t,\mathbf{x}) && \text{in } [0,T]\times D\times\Omega \\ &u(0,\mathbf{x},\omega) = g_1(\mathbf{x}), \quad u_t(0,\mathbf{x},\omega) = g_2(\mathbf{x}) && \text{on } \{t=0\}\times D\times\Omega \\ &u(t,\mathbf{x},\omega) = 0 && \text{on } [0,T]\times\partial D\times\Omega \end{aligned} \tag{1}$$

where

$$f \in L^2([0,T]\times D), \qquad g_1 \in H_0^1(D), \qquad g_2 \in L^2(D). \tag{2}$$

We assume that the random wave speed a is bounded and uniformly coercive,

$$0 < a_{min} \le a(\mathbf{x},\omega) \le a_{max} < \infty, \quad \forall \mathbf{x} \in D, \quad \forall \omega \in \Omega. \tag{3}$$

In many wave propagation problems, the source of randomness can be described or approximated by using only a small number of uncorrelated random variables. For example, in seismic applications, a typical situation is the case of layered materials where the wave speeds in the layers are not perfectly known and therefore are described by uncorrelated random variables. The number of random variables corresponds therefore to the number of layers. Another example is the approximation of the random speed by a truncated Karhunen-Loéve expansion [2]. In this case the number of random variables is the number of

terms in the expansion. This motivates us to make the following *finite dimensional noise* assumption on the form of the wave speed,

$$a(\mathbf{x},\omega) = a(\mathbf{x}, Y_1(\omega), \ldots, Y_N(\omega)), \quad \forall \mathbf{x} \in D, \quad \forall \omega \in \Omega, \tag{4}$$

where $N \in \mathbb{N}_+$ and $Y = [Y_1, \ldots, Y_N] \in \mathbb{R}^N$ is a random vector. We denote by $\Gamma_n \equiv Y_n(\Omega)$ the image of Y_n and assume that Y_n is bounded. We let $\Gamma = \prod_{n=1}^{N} \Gamma_n$ and assume further that the random vector Y has a bounded joint probability density function $\rho : \Gamma \to \mathbb{R}_+$ with $\rho \in L^\infty(\Gamma)$. We note that by using a similar approach to [1] we can also treat unbounded random variables, such as Gaussian and exponential variables. Here, we consider only bounded random variables for simplicity.

In this paper, in particular, we consider a heterogeneous medium consisting of N sub-domains. In each sub-domain, the wave speed is smooth and represented by one random variable. The boundaries of sub-domains, which are interfaces of speed discontinuity, are assumed to be smooth and do not overlap. The random speed a can for instance be given by

$$a(\mathbf{x},\omega) = a_0(\mathbf{x}) + \sum_{n=1}^{N} a_n(\mathbf{x},\omega)\, \mathcal{X}_n(\mathbf{x}), \qquad a_n(\mathbf{x},\omega) = Y_n(\omega)\, \alpha_n(\mathbf{x}), \tag{5}$$

where $\mathcal{X}_n$ are indicator functions describing the geometry of each sub-domain, Y_n are independent and identically distributed random variables, and α_n are smooth functions defined on each sub-domain. Note that the representation of the coefficient a in (5) is exact, and there is no truncation error as in the Karhunen-Loéve expansion. The more general case where the wave speed in each sub-domain $a_n(\mathbf{x},\omega)$ is represented by a Karhunen-Loéve expansion can be treated in the same way. In this case the total number of random variables is $\sum_{n=1}^{N} M_n$, where M_n is the number of terms in the truncated Karhunen-Loéve expansion in each sub-domain. The case where the geometry of sub-domains is also random will be addressed elsewhere. For elliptic equations, random boundaries have been studied, e.g., in [43, 6, 13].

The finite dimensional noise assumption implies that the solution of the stochastic IBVP (1) can be described by only N random variables, i.e., $u(t,\mathbf{x},\omega) = u(t,\mathbf{x},Y_1(\omega),\ldots,Y_N(\omega))$. This turns the original stochastic problem into a deterministic IBVP for the wave equation with an N-dimensional parameter, which allows the use of standard finite difference and finite element methods to approximate the solution of the resulting deterministic problem $u = u(t,\mathbf{x},Y)$, where $t \in [0,T]$, $\mathbf{x} \in D$, and $Y \in \Gamma$. Note that the knowledge of $u = u(t,\mathbf{x},Y)$ fully determines the law of the random field $u = u(t,\mathbf{x},\omega)$. The ultimate goal is then the prediction of statistical moments of the solution u or statistics of some given quantities of physical interest.

Before studying the well-posedness and regularity in details, we start the discussion with two simple examples.

Example 1. A basic technique for studying the regularity of the solution of a PDE with respect to a parameter is based on analyzing the equation in the complex plane. In this approach, the parameter is first extended into the complex plane. Then, if the extended problem is well-posed and the first derivative of the resulting complex-valued solution with respect to the parameter satisfies the so called Cauchy-Riemann conditions, the solution can analytically be extended into the complex plane. This approach has been used in [27] to prove the analyticity of the solution of parabolic PDEs with stochastic parameters. As a first example, we therefore consider the Cauchy problem for the one-dimensional scalar wave equation with a complex-valued one-parameter wave speed,

$$\begin{aligned} &u_{tt}(t,x) - a^2\, u_{xx}(t,x) = 0, && \text{in } [0,T]\times\mathbb{R},\\ &u(0,x) = g(x), \quad u_t(0,x) = 0, && \text{on } \{t=0\}\times\mathbb{R}, \end{aligned}$$

with a constant, complex-valued coefficient

$$a = a_R + i\, a_I, \qquad a_R, a_I \in \mathbb{R}.$$

Assume that $g(x)$ is a smooth function that vanishes at infinity. We Fourier transform the problem with respect to x and get

$$\begin{aligned} &\hat{u}_{tt}(t,k) + a^2\, k^2\, \hat{u}(t,k) = 0,\\ &\hat{u}(0,k) = \hat{g}(k), \quad \hat{u}_t(0,k) = 0, \end{aligned}$$

where $\hat{u}(t,k) = \int_{\mathbb{R}} u(t,x)\, e^{-i\,k\,x} dx$ and $\hat{g}(k) = \int_{\mathbb{R}} g(x)\, e^{-i\,k\,x} dx$ are the Fourier transforms of $u(t,x)$ and $g(x)$ with respect to x, respectively. The solution of this linear, second order ordinary differential equation with parameter k is given by

$$\hat{u}(t,k) = \frac{\hat{g}(k)}{2}\left(e^{r_1\,t} + e^{r_2\,t}\right), \qquad r_{1,2} = \pm\, i\, a\, k.$$

When $a_I = 0$, then $r_{1,2} = \pm\, i\, a_R\, k$. Performing the inverse Fourier transform, we get the solution

$$u(t,x) = \frac{1}{2}\left(g(x + a_R\, t) + g(x - a_R\, t)\right),$$

and therefore the Cauchy problem is well-posed.

When $a_I \neq 0$, then $\mathrm{Re}(r_1) = -\mathrm{Re}(r_2) = -a_I\, k$, and

$$|\hat{u}(t,k)| \sim |\hat{g}(k)|\, e^{|a_I|\,|k|\,t}.$$

Therefore, regardless of the sign of $\mathrm{Re}(a^2) = a_R^2 - a_I^2$, the Fourier transform of the solution $\hat{u}(t,k)$ grows exponentially fast, i.e., $e^{|a_I|\,|k|\,t}$, unless the Fourier transform of the initial solution $\hat{g}(k)$ decays faster than $e^{-|a_I|\,|k|\,t}$. The Cauchy problem is therefore well-posed only if $g(x)$ is in a restricted class of Gevrey spaces.

Definition 1. *A function $g(x)$ is a Gevrey function of order $q > 0$, i.e., $g \in G^q(\mathbb{R})$, if $g \in C^\infty(\mathbb{R})$ and for every compact subset $D \subset \mathbb{R}$, there exists a positive constant C such that,*

$$\max_{x \in D} |\partial^n g(x)| \leq C^{n+1} (n!)^q.$$

In particular, $G^1(\mathbb{R})$ is the space of analytic functions [15]. For $0 < q < 1$, the class $G^q(\mathbb{R})$ is a subclass of the analytic functions, while for $1 < q < \infty$ it contains the analytic functions.

We now state a known result on the decay of the Fourire transform of Gevrey functions [32, 21].

Lemma 1. *A function $g(x)$ belongs to the Gevrey space $G^q(\mathbb{R})$ if and only if there exist positive constants C and ϵ such that $|\hat{g}(k)| \leq C\, e^{-\epsilon\, |k|^{1/q}}$.*

Therefore, for the Cauchy problem to be well-posed in the complex strip $\Sigma_r = \{(a_R + i\, a_I) \in \mathbb{C} : |a_I| \leq r\}$, we need $g \in G^q(\mathbb{R})$ with $q < 1$. Note that for $q = 1$, the problem is well-posed only for a finite time interval when $t \leq \epsilon/r$. This shows that even if the initial solution g is analytic, i.e, $g \in G^1(\mathbb{R})$, the solution is not analytic for all times in Σ_r. Reversing the argument, we can say that, starting from an analytic initial solution g, with $|\hat{g}(k)| \leq C\, e^{-\epsilon\, |k|}$, the solution at time t will be analytic only in the strip $\Sigma_{\epsilon/t}$, and the analyticity region becomes smaller and smaller as time increases.

Example 2. An important characteristic of waves in a heterogeneous medium in which the wave speed is piecewise smooth, is scattering by discontinuity interfaces. As a simple scattering problem, we consider the Cauchy problem for the second order scalar wave equation in a one-dimensional domain consisting of two homogeneous half-spaces separated by an interface at $x = 0$,

$$u_{tt}(t,x) - \big(a^2(x)\, u_x(t,x)\big)_x = 0, \quad \text{in } [0,T] \times \mathbb{R},$$

with a piecewise constant wave speed

$$a(x) = \begin{cases} a_-, & x < 0, \\ a_+, & x > 0. \end{cases}$$

In this setting, the wave speed contains two positive parameters, a_- and a_+. We choose the initial conditions such that the initial wave pulse is smooth, compactly supported, lies in the left half-space, and travels to the right. That is

$$u(0,x) = g(-x), \quad u_t(0,x) = a_-\, g'(-x), \quad g(x) \in C_0^\infty(0,\infty).$$

By d'Alembert's formula, the solution reads

$$u(t,x) = \begin{cases} g(a_-\, t - x) + \Phi_1(a_-\, t + x), & x < 0, \\ \Phi_2(a_+\, t - x), & x > 0. \end{cases}$$

Note that when $x < -a_-\, t$, the solution is purely right-going, $u = g(a_-\, t - x)$, and when $x > a_+\, t$, the solution is zero, $u = 0$.

The functions Φ_1 and Φ_2 are obtained by the interface jump conditions at $x = 0$,

$$u(t, 0^-) = u(t, 0^+), \quad a_-^2 \, u_x(t, 0^-) = a_+^2 \, u_x(t, 0^+). \tag{6}$$

After some manipulation, we get the solution

$$u(t, x) = \begin{cases} g(a_- \, t - x) + \frac{a_- - a_+}{a_- + a_+} \, g(a_- \, t + x), & x < 0, \\ \frac{2 \, a_-}{a_- + a_+} \, g(\frac{a_-}{a_+} \, (a_+ \, t - x)), & x > 0. \end{cases} \tag{7}$$

The interpretation of this solution is that the initial pulse $g(-x)$ inside the left half-space moves to the right with speed a_- until it reaches the interface. At the interface it is partially reflected (Φ_1) with speed a_- and partially transmitted (Φ_2) with speed a_+. The interface between two layers generates no reflections if the speeds are equal, $a_- = a_+$. From the closed form of the solution (7), we note that the solution $u(t, x)$ is infinitely differentiable with respect to both parameters a_- and a_+ in $(0, +\infty)$. Note that the smooth initial solution $u(0, x)$, which is contained in one layer with zero value at the interface, automatically satisfies the interface conditions (6) at time zero. Otherwise, if for instance the initial solution crosses the interface without satisfying (6), a singularity is introduced in the solution, and the high regularity result does no longer hold.

In the more general case of multi-dimensional heterogeneous media consisting of sub-domains, the interface jump conditions on a smooth interface Υ between two sub-domains D_I and D_{II} are given by

$$[u(t, .)]_\Upsilon = 0, \qquad [a^2(.) \, u_n(t, .)]_\Upsilon = 0. \tag{8}$$

Here, the subscript n represents the normal derivative, and $[v(.)]_\Upsilon$ is the jump in the function v across the interface Υ. In this general case, the high regularity with respect to parameters holds provided the smooth initial solution satisfies (8). The jump conditions are satisfied for instance when the initial data are contained within sub-domains. This result for Cauchy problems can easily be extended to IBVPs by splitting the problem to one pure Cauchy and two half-space problems. See Sect. 2.3.2 for more details.

Remark 1. *Immediate results of the above two examples are the following:*

1. *For the solution to be analytic with respect to the random wave speed at all times in a given complex strip Σ_r with $r > 0$, the initial datum needs to live in a space strictly contained in the space of analytic functions, which is the Gevrey space $G^q(\mathbb{R})$ with $0 < q < 1$. Moreover, if the problem is well-posed and the data are analytic, the solution may be analytic with respect to the parameter in Σ_r only for a short time interval.*

2. *In a heterogeneous medium with piecewise smooth wave speeds and smooth interfaces, if the data are smooth and the initial solution satisfies the interface jump conditions (8), the solution is smooth with respect to the wave speeds. If the initial solution does not satisfy (8), the solution is not smooth with respect to the wave speeds.*

We note that the above high regularity results with respect to parameters are valid only for particular types of smooth data. In real applications, the data are not smooth. We therefore study the well-posedness and regularity properties in the more general case when the data satisfy the minimal assumptions (2).

2.2 Well-posedness

We now show that the problem (1) with the data satisfying (2) and the assumption (3) is well-posed. For a function of the random vector Y, we introduce the space of square integrable functions:

$$L^2_\rho(\Gamma) = \{v : \Gamma \to \mathbb{R}, \int_\Gamma v(Y)^2\, \rho(Y)\, dY < \infty\},$$

with the inner product

$$(v_1, v_2)_{L^2_\rho(\Gamma)} = \mathbb{E}\,[v_1\, v_2] = \int_\Gamma v_1\, v_2\, \rho(Y)\, dY.$$

We also introduce the mapping $\mathbf{u} : [0,T] \to H^1_0(D) \otimes L^2_\rho(\Gamma)$, defined by

$$[\mathbf{u}(t)](\mathbf{x}, Y) := u(t, \mathbf{x}, Y), \quad \forall t \in [0,T],\, \mathbf{x} \in D,\, Y \in \Gamma.$$

Similarly, we introduce the function $\mathbf{f} : [0,T] \to L^2(D)$, defined by

$$[\mathbf{f}(t)](\mathbf{x}) := f(t, \mathbf{x}), \quad \forall t \in [0,T],\, \mathbf{x} \in D.$$

Finally, for a real Hilbert space X with norm $\|.\|_X$, we introduce the time-involving space

$$H_X \equiv L^2(0,T;X) \otimes L^2_\rho(\Gamma) \equiv L^2(0,T; X \otimes L^2_\rho(\Gamma)),$$

consisting of all measurable functions v with

$$\|v\|^2_{H_X} = \int_{[0,T]\times\Gamma} \|v\|^2_X\, \rho(Y)\, dt\, dY < \infty.$$

Examples of X include the $L^2(D)$ space and the Sobolev space $H^1_0(D)$ and its dual space $H^{-1}(D)$.

We now recall the notion of weak solutions for the IBVP (1).

Definition 2. *The function $\mathbf{u} \in H_{H^1_0(D)}$ with $\mathbf{u}' \in H_{L^2(D)}$ and $\mathbf{u}'' \in H_{H^{-1}(D)}$ is a weak solution to the IBVP (1) provided the following hold:*

(i) $\mathbf{u}(0) = g_1$ *and* $\mathbf{u}'(0) = g_2$,

(ii) for a.e. time $0 \le t \le T$ *and* $\forall v \in H_0^1(D) \otimes L_\rho^2(\Gamma)$*:*

$$\int_{D\times\Gamma} \mathbf{u}''(t)\, v\, \rho\, d\mathbf{x}\, dY + \int_\Gamma B(\mathbf{u}(t), v)\, \rho\, dY = \int_{D\times\Gamma} \mathbf{f}(t)\, v\, \rho\, d\mathbf{x}\, dY, \qquad (9)$$

where

$$B(v_1, v_2)(Y) = \int_D a^2(\mathbf{x}, Y)\, (\nabla v_1(\mathbf{x}, Y) \cdot \nabla v_2(\mathbf{x}, Y))\, d\mathbf{x}, \qquad \forall v_1, v_2 \in H_0^1(D) \otimes L_\rho^2(\Gamma).$$

Theorem 1. *There is a unique weak solution* $\mathbf{u} \in H_{H_0^1(D)}$ *to the IBVP (1). Moreover, it satisfies the energy estimate*

$$\max_{0\le t\le T} \left(\|\mathbf{u}(t)\|_{H_0^1(D)\otimes L_\rho^2(\Gamma)} + \|\mathbf{u}'(t)\|_{L^2(D)\otimes L_\rho^2(\Gamma)} \right) + \|\mathbf{u}''\|_{H_{H^{-1}(D)}} \le C\Big(\|\mathbf{f}\|_{L^2(0,T;L^2(D))} + \|g_1\|_{H_0^1(D)} + \|g_2\|_{L^2(D)} \Big). \qquad (10)$$

Proof. By the energy method, the assumptions (2) and (3) imply the existence and uniqueness of the weak solution. The proof is an easy extension of the proof for deterministic problems, see e.g. [9]. □

2.3 Regularity

In this section we study the regularity of the solution and of a quantity of interest with respect to the random input variable Y. The main result is that under the minimal assumptions (2) and (3) the solution, which is in $L_\rho^2(\Gamma)$, has in general only one bounded derivative with respect to Y, while the considered quantity of interest may have many bounded derivatives. The available regularity is then used to estimate the convergence rate of the error for the stochastic collocation method.

2.3.1 Regularity of the solution

We first investigate the regularity of the solution with respect to the random variable Y. For deterministic problems, for instance when Y is a fixed constant, it is well known that in the case of $\mathbf{x}$-discontinuous wave speed, with the data satisfying (2) and under the assumption (3), the solution of (1) is in general only $\mathbf{u} \in C^0(0, T; H_0^1(D))$, see for instance [30, 33]. In other words, in the presence of discontinuous wave speed, one should not expect higher spatial regularity than $H^1(D)$.

To investigate the Y-regularity of the solution in the stochastic space, we differentiate the IBVP (1) with respect to Y and obtain

$$\tilde{u}_{tt} - \nabla \cdot \left(a^2\, \nabla \tilde{u} \right) = \nabla \cdot \left(2\, a\, a_Y\, \nabla u \right), \qquad \tilde{u} = \partial_Y u, \qquad (11)$$

with homogeneous initial and boundary conditions. The force term in the above IBVP is $f_1 := \nabla \cdot \big(2\, a\, a_Y\, \nabla u\big) \in L^1(0,T;H^{-1}(D))$ for every $Y \in \Gamma$. In fact if $v \in L^1(0,T;H_0^1(D))$, then

$$\Big|\int_0^T \int_D f_1\, v\, d\mathbf{x}\, dt\Big| = |\langle \nabla \cdot \big(2\, a\, a_Y\, \nabla u\big), v\rangle| = |\langle 2\, a\, a_Y\, \nabla u, \nabla v\rangle| \leq$$
$$\leq 2\, |a\, a_Y|_\infty\, \|\nabla u\|_{L^1(0,T;L^2(D))}\, \|\nabla v\|_{L^1(0,T;L^2(D))} < \infty.$$

We now state an important result which is a generalization of a result given by Hörmander [14].

Lemma 2. *For arbitrary $f \in L^1(0,T;H^k(D))$, $g_1 \in H^{k+1}(D)$ and $g_2 \in H^k(D)$, with $k \in \mathbb{R}$, for every $Y \in \Gamma$, there is a unique weak solution $\mathbf{u} \in C^0(0,T;H^{k+1}(D)) \cap C^1(0,T;H^k(D))$ of the IBVP (1) with the $\mathbf{x}$-smooth wave speed (4) satisfying (3). Moreover, it satisfies the energy estimate*

$$\max_{0\leq t\leq T} \big(\|\mathbf{u}(t)\|_{H^{k+1}(D)} + \|\mathbf{u}'(t)\|_{H^k(D)}\big) \leq$$
$$C_{k,T}\Big(\|\mathbf{f}\|_{L^1(0,T;H^k(D))} + \|g_1\|_{H^{k+1}(D)} + \|g_2\|_{H^k(D)}\Big). \quad (12)$$

Proof. The proof is an easy extension of the proof of Lemma 23.2.1 and Theorem 23.2.2 in [14]. □

We note that Lemma 2 holds for $\mathbf{x}$-smooth wave speeds. When the wave speed is non-smooth, it holds only for $k = -1$ and $k = 0$ [33]. We apply Lemma 2 to (11) with $k = -1$ (which is valid also for non-smooth coefficients) and obtain

$$\tilde{u} \in C^0(0,T;L^2(D)), \qquad \forall\, Y \in \Gamma.$$

Moreover, the solution (7) of Example 2 with $g \in H_0^1(\mathbb{R})$ shows that the second and higher Y-derivatives do not exist. Therefore, under the minimal assumptions (2), the solution has at most one bounded Y-derivative in $L^2(D)$. We have proved the following result,

Theorem 2. *For the solution of the second order wave propagation problem (1) with data given by (2) and a random piecewise smooth wave speed satisfying (3) and (5), we have $\partial_Y u \in C^0(0,T;L^2(D))$ for every $Y \in \Gamma$.*

2.3.2 Regularity of quantities of interest

We now consider the quantity of interest

$$\mathcal{Q}(Y) = \int_0^T \int_D u(t,\mathbf{x},Y)\, \phi(\mathbf{x})\, d\mathbf{x}\, dt + \int_D u(T,\mathbf{x},Y)\, \psi(\mathbf{x})\, d\mathbf{x}, \quad (13)$$

where u solves (1) and the mollifiers ϕ and ψ are given functions of $\mathbf{x}$. As a corollary of Theorem 2, we can write,

Corollary 1. *With the assumptions of Theorem 2 and $\phi \in L^1(D)$ and $\psi \in L^1(D)$, we have $\frac{d}{dY}\mathcal{Q} \in L^2(\Gamma)$.*

We now assume that the mollifiers $\phi(\mathbf{x}) \in C_0^\infty(D)$ and $\psi(\mathbf{x}) \in C_0^\infty(D)$ are smooth functions and analytic in the interior of their supports. We further assume that their supports does not cross the speed discontinuity interfaces. We will show that the resulting quantity of interest (13) may have higher Y-regularity, without any higher regularity assumptions on the data than those in (2). For this purpose, we introduce the *influence function* (or dual solution) φ associated to the quantity of interest, $\mathcal{Q}$, as the solution of the dual problem

$$\begin{aligned} &\varphi_{tt}(t,\mathbf{x},Y) - \nabla\cdot\big(a^2(\mathbf{x},Y)\nabla\varphi(t,\mathbf{x},Y)\big) = \phi(\mathbf{x}) && \text{in } [0,T]\times D\times\Gamma \\ &\varphi(T,\mathbf{x},Y) = 0, \quad \varphi_t(T,\mathbf{x},Y) = -\psi(\mathbf{x}) && \text{on } \{t=T\}\times D\times\Gamma \\ &\varphi(t,\mathbf{x},Y) = 0 && \text{on } [0,T]\times\partial D\times\Gamma \end{aligned} \tag{14}$$

Note that this is a well-posed backward wave equation with smooth initial data at the final time T and a smooth force term.

We can write

$$\begin{aligned} \mathcal{Q}(Y) &= \int_0^T\int_D u\left(\varphi_{tt} - \nabla\cdot\big(a^2\nabla\varphi\big)\right)d\mathbf{x}\,dt + \int_D u(T,\mathbf{x},Y)\,\psi(\mathbf{x})\,d\mathbf{x} \\ &= \int_D\int_0^T u\,\varphi_{tt}\,d\mathbf{x}\,dt + \int_0^T\int_D a^2\,\nabla u\cdot\nabla\varphi\,d\mathbf{x}\,dt + \int_D u(T,\mathbf{x},Y)\,\psi(\mathbf{x})\,d\mathbf{x} \\ &= \int_0^T\int_D u_{tt}\,\varphi\,d\mathbf{x}\,dt - \int_0^T\int_D \varphi\,\nabla\cdot\big(a^2\,\nabla u\big)\,d\mathbf{x}\,dt \\ &\quad + \int_D\Big[\varphi_t\,u - \varphi\,u_t\Big]_0^T d\mathbf{x} + \int_D u(T,\mathbf{x},Y)\,\psi(\mathbf{x})\,d\mathbf{x} \\ &= \int_0^T\int_D \varphi(t,\mathbf{x},Y)\,f(t,\mathbf{x})\,d\mathbf{x}\,dt + \int_D\Big(g_2(\mathbf{x})\,\varphi(0,\mathbf{x},Y) - g_1(\mathbf{x})\,\varphi_t(0,\mathbf{x},Y)\Big)\,d\mathbf{x}. \end{aligned}$$

The last equality follows from the initial condition in (1) and in the dual problem (14). This shows that the regularity of the quantity of interest depends only on Y-regularity of the dual solution.

To investigate the Y-regularity of dual solution, we first note that the finite speed of wave propagation and the superposition principle due to the linearity of the dual problem (14) makes it possible to split the IBVP in $\mathbb{R}^d$ into two half-space problems and a pure Cauchy problem [16]. To clarify this, consider a one-dimensional strip problem ($d = 1$) for (14) on the physical domain $D = [0,1]$ with $N = 2$ layers with widths d_1 and d_2. Let $\vartheta_j \in C^\infty(D)$, $j = 1,2,3$, be monotone functions with

$$\vartheta_1(x) = \begin{cases} 1, & x \le \frac{d_1}{6}, \\ 0, & x \ge \frac{d_1}{3}. \end{cases} \qquad \vartheta_2(x) = \begin{cases} 1, & x \ge 1 - \frac{d_2}{6}, \\ 0, & x \le 1 - \frac{d_2}{3}. \end{cases}$$

and $\vartheta_3(x) = 1 - \vartheta_1(x) - \vartheta_2(x)$. Set $\varphi = \varphi_1 + \varphi_2 + \varphi_3$, where each φ_j solves

$$\begin{aligned} &\varphi_{jtt} - \nabla \cdot \left(a^2 \nabla \varphi_j\right) = \vartheta_j \, \phi && \text{in } [0,T] \times D \times \Gamma \\ &\varphi_j = 0, \quad \varphi_{jt} = -\vartheta_j \, \psi && \text{on } \{t = T\} \times D \times \Gamma \\ &\varphi_j = 0 && \text{on } [0,T] \times \partial D \times \Gamma \end{aligned}$$

The finite speed of propagation implies that there is a time $0 < T_1 \leq T$ where $\varphi_1 = 0$ for $x \in [d_1, 1]$ and $t \in [T - T_1, T]$. Therefore, we can consider φ_1 as the solution of the right half-space problem

$$\begin{aligned} &\varphi_{1tt} - \nabla \cdot \left(a^2 \nabla \varphi_1\right) = \vartheta_1 \, \phi, && t \in [T - T_1, T], \quad x \geq 0 \\ &\varphi_1 = 0, \quad \varphi_{1t} = -\vartheta_1 \, \psi, && t = T, \quad x \geq 0 \\ &\varphi_1 = 0, && t \in [T - T_1, T], \quad x = 0 \end{aligned}$$

Note that here we redefine the wave speed a by extending the speed corresponding to the left layer to the whole half space $0 \leq x < \infty$. Similarly, φ_2 and φ_3 locally solve a left half-space and a pure Cauchy problem, respectively. These considerations are valid in the time interval $[T - T_1, T]$. At time $t = T - T_1$, we obtain a new final dual solution and can restart.

The Y-regularity of the dual solution φ is therefore obtained by the regularity of φ_1, φ_2 and φ_3. The first two functions, φ_1 and φ_2, which solve two half-space problems with smooth data and coefficients, are smooth [9] and have $s \geq 2$ bounded Y-derivatives. The third function φ_3, which solves a single interface Cauchy problem with smooth data whose support does not cross the interface and with a piecewise smooth wave speed, has again $s \geq 2$ bounded Y-derivatives. In one dimension ($d = 1$), when the wave speed is piecewise constant, we can solve the Cauchy wave equation by d'Alembert's formula and explicitly obtain the solution φ_3 which is smooth with respect to the wave speed and therefore is Y-smooth, see Example 2 as a simple illustration. When the wave speed is variable, we can employ the energy method to show Y-regularity, see Theorem A2 in the appendix. Note that the same result holds also for a multiple interface Cauchy problem. Therefore the dual solution φ and consequently the quantity of interest $\mathcal{Q}$ have $s \geq 2$ bounded Y-derivatives. We note that for $0 \leq t \leq T - T_1$, although the new final dual solution $\varphi(T - T_1, x, Y)$ may not be contained in one layer, but since it naturally satisfies the correct jump conditions at the interface, the Y-regularity of the dual solution holds in the time interval $[0, T]$. We have therefore proved the following result in a one-dimensional physical space,

Theorem 3. *Let $D \subset \mathbb{R}$. With the assumptions of Theorem 2 and if $\phi \in C_0^\infty(D)$ and $\psi \in C_0^\infty(D)$ and their supports do not cross the discontinuity interfaces, the quantity of interest (13) satisfies $\frac{d^s}{dY^s}\mathcal{Q}(Y) \in L^2(\Gamma)$ with $s \geq 2$.*

In a more general case of two-dimensional physical space ($d = 2$), φ_1 and φ_2 are again smooth [9] and have $s \geq 2$ bounded Y-derivatives. The proof of smoothness for φ_3 is more complicated. However, noting that the discontinuity

occurs in the normal direction to the interfaces, we can employ a localization argument and build a two-dimensional result by generalizing the one-dimensional ones. Based on this and numerical results, we therefore make the following conjecture,

Conjecture 1. *Let $D \subset \mathbb{R}^2$. With the assumptions of Theorem 2 and if $\phi \in C_0^\infty(D)$ and $\psi \in C_0^\infty(D)$ and their supports do not cross the discontinuity interfaces, the quantity of interest (13) satisfies $\frac{d^s}{dY^s}\mathcal{Q}(Y) \in L^2(\Gamma)$ with $s \geq 2$.*

Remark 2. *For quantities of interest which are nonlinear in u the high Y-regularity property does not hold in general. In fact, the corresponding dual problems have non-smooth forcing terms and data (assuming u is not smooth), and therefore the dual solutions are not smooth with respect to Y. In Sect. 5, we numerically study the Arias intensity which is a nonlinear quantity of interest and show that it is not regular with respect to Y.*

3 A stochastic collocation method

In this section, we review the stochastic collocation method for computing the statistical moments of the solution u to the problem (1), see for example [1, 40]. We first discretize the problem in space and time, using a deterministic numerical method, such as the finite element or the finite difference method, and obtain a semi-discrete problem. We next collocate the semi-discrete problem in a set of η collocation points $\{Y^{(k)}\}_{k=1}^{\eta} \in \Gamma$ and compute the approximate solutions $u_h(t, \mathbf{x}, Y^{(k)})$. Finally, we build a global polynomial approximation $u_{h,p}$ upon those evaluations

$$u_{h,p}(t, \mathbf{x}, Y) = \sum_{k=1}^{\eta} u_h(t, \mathbf{x}, Y^{(k)})\, L_k(Y),$$

for suitable multivariate polynomials $\{L_k\}_{k=1}^{\eta}$ such as Lagrange polynomials. Here, h and p represent the discretization mesh size and the polynomial degree, respectively.

In what follows, we address in more details the choice of collocation points. We seek a numerical approximation to u in a finite-dimensional subspace $H_{h,w}$ of the space $H_{H_0^1(D)} \equiv L^2(0, T; H_0^1(D)) \otimes L_\rho^2(\Gamma)$ in which the function u lives. We define the subspace based on a tensor product $H_{h,w} = H_h \otimes H_w$, where

- $H_h([0, T] \times D) \subset L^2(0, T; H_0^1(D))$ is the space of the semi-discrete solution in time and space for a constant Y. The subscript h denotes the spatial grid-lengths and the time step-size.
- $H_w(\Gamma) \subset L_\rho^2(\Gamma)$ is a tensor product space which is the span of the tensor product of orthogonal polynomials with degree at most $\mathbf{p} = [p_1(w), \ldots, p_N(w)]$. The positive integer w is called the *level*, and $p_n(w)$ is the maximum degree

of polynomials in the n-th direction, with $n = 1, \ldots, N$, given as a function of the level w. For each Y_n, $n = 1, \ldots, N$, with the density ρ_n, let $H_{p_n}(\Gamma_n)$ be the span of ρ_n-orthogonal polynomials $V_0^{(n)}, V_1^{(n)}, \ldots, V_{p_n}^{(n)}$. The tensor product space is then $H_w(\Gamma) = \bigotimes_{n=1}^N H_{p_n}(\Gamma_n)$. The dimension of H_w is $\dim(H_w) = \prod_{n=1}^N (p_n + 1)$. Without loss of generality, for bounded random variables, we assume $\Gamma = [-1, 1]^N$.

Having the finite-dimensional subspace $H_{h,w}$ constructed, we can use Lagrange interpolation to build an approximate solution u.

The ultimate goal of the computations is the prediction of statistical moments of the solution u (such as the mean value and variance) or statistics of some given quantities of interest $\mathcal{Q}(Y)$. For a linear bounded operator $\Psi(u)$,using the Gauss quadrature formula for approximating integrals, we write

$$\mathbb{E}\left[\Psi(u(.,Y))\right] \approx \mathbb{E}\left[\Psi(u_{h,p}(.,Y))\right] = \int_\Gamma \Psi(u_{h,p}(.,Y))\,\rho(Y)\,dY \approx \sum_{k=1}^{\eta} \theta_k\,\Psi(u_h(.,Y^{(k)})),$$

where the weights are

$$\theta_k = \prod_{n=1}^N \int_{\Gamma_n} L_{k_n}(Y_n)\,\rho_n(Y_n)\,dY_n, \qquad L_{k_n}(Y_n) = \prod_{i=0,\, i\neq k_n}^{\eta} \frac{Y_n - Y_n^{(i)}}{Y_n^{(k_n)} - Y_n^{(i)}},$$

and the collocation points $Y^{(k)} = [Y_1^{(k_1)}, \ldots, Y_N^{(k_N)}] \in \Gamma$ are tensorized Gauss points with $Y_n^{(k_n)}$, $k_n = 0, 1, \ldots, p_n$, being the zeros of the ρ_n-orthogonal polynomial of degree p_n+1. Here, for any vector of indices $[k_1, \ldots, k_N]$ with $0 \le k_n \le p_n$ the associated global index reads $k = 1 + k_1 + (p_1+1)\,k_2 + (p_1+1)\,(p_2+1)\,k_3 + \ldots$.

Remark 3. *The choice of orthogonal polynomials depends on the density function ρ. For instance, for uniform random variables $Y_n \sim \mathcal{U}(-1, 1)$, Legendre polynomials are used, i.e. $V_{-1}^{(n)} = 0$, $V_0^{(n)} = 1$, and*

$$V_{k+1}^{(n)}(Y_n) = \frac{2k+1}{k+1}\,Y_n\,V_k^{(n)}(Y_n) - \frac{1}{2(k+1/2)}\,V_{k-1}^{(n)}(Y_n), \quad k \ge 0.$$

Other well known orthogonal polynomials include Hermite polynomials for Gaussian random variables and Laguerre polynomials for exponential random variables [42].

Remark 4. *There are other choices for the approximation space H_w. For example, instead of orthogonal polynomials, we can choose a piecewise constant approximation using the Haar-wavelet basis. We can also choose a piecewise polynomial approximation. The choice of the approximating space may depend on the smoothness of the function with respect to Y. In general, for smooth functions, we choose a polynomial approximation, while for non-smooth functions, we choose a low-degree piecewise polynomial or wavelet-type approximation [19, 20].*

We now consider two possible approaches for constructing the tensor product space H_w and briefly review the Lagrange interpolation.

3.1 Full tensor product space and interpolation

For a given multi-index $\mathbf{j} = [j_1, \ldots, j_N] \in \mathbb{Z}_+^N$, containing N non-negative integers, we define

$$H_{\mathbf{j}}(Y) = V^{(1)}_{p(j_1)}(Y_1) \otimes \ldots \otimes V^{(N)}_{p(j_N)}(Y_N).$$

Given an index j, we calculate the polynomial degree $p(j)$ either by

$$p(j) = j, \tag{15}$$

or by

$$p(j) = 2^j \text{ for } j > 0, \qquad p(0) = 0. \tag{16}$$

The isotropic full tensor product space is then chosen as

$$H_w^T = \text{span}\{H_{\mathbf{j}},\ \forall \mathbf{j} := \max_n j_n \leq w\}.$$

In other words, in each direction we take all polynomials of degree at most $p(w)$, and therefore $\dim(H_w^T) = (p(w)+1)^N$. Since the dimension of the space grows exponentially fast with N (*curse of dimensionality*), the full tensor product approximation can be used only when the number of random variables N is small.

The multi-dimensional Lagrange interpolation corresponding to a multi-index $\mathbf{j}$ is

$$\mathcal{I}_{\mathbf{j}}^N[u](., Y) = \bigotimes_{n=1}^{N} \mathcal{U}^{j_n}(u)(Y) = \sum_{k_1=0}^{p(j_1)} \ldots \sum_{k_N=0}^{p(j_N)} u_h(., Y_{1,k_1}^{j_1}, \ldots, Y_{N,k_N}^{j_N}) \prod_{n=1}^{N} L_{n,k_n}^{j_n}(Y_n), \tag{17}$$

where, for each value of a non-negative index j_n in the multi-index $\mathbf{j}$, $\mathcal{U}^{j_n}$ is the one-dimensional Lagrange interpolation operator, the set $\{Y_{n,k_n}^{j_n}\}_{k_n=0}^{p(j_n)}$ is a sequence of abscissas for Lagrange interpolation on Γ_n, and $\{L_{n,k_n}^{j_n}(y)\}_{k_n=0}^{p(j_n)}$ are Lagrange polynomials of degree $p(j_n)$,

$$L_{n,k_n}^{j_n}(y) = \prod_{i=0, i\neq k_n}^{p(j_n)} \frac{y - Y_{n,i}^{j_n}}{Y_{n,k_n}^{j_n} - Y_{n,i}^{j_n}}.$$

The set of points where the function u_h is evaluated to construct (17) is the tensor grid

$$\mathcal{H}_{\mathbf{j}}^N = \{Y_{\mathbf{k}} = [Y_{1,k_1}^{j_1}, \ldots, Y_{N,k_N}^{j_N}],\ 0 \leq k_n \leq p(j_n)\}.$$

The isotropic full tensor interpolation is obtained when we take $\mathbf{j} = [w, w, \ldots, w]$ in (17), and the corresponding operator is denoted by $\mathcal{I}_{w,N}$,

$$\mathcal{I}_{w,N}[u](., Y) = \sum_{k_1=0}^{p(w)} \ldots \sum_{k_N=0}^{p(w)} u_h(., Y_{1,k_1}^{w}, \ldots, Y_{N,k_N}^{w}) \prod_{n=1}^{N} L_{n,k_n}^{w}(Y_n). \tag{18}$$

3.2 Sparse tensor product space and interpolation

Here, we briefly describe the isotropic Smolyak formulas [4]. The sparse tensor product space is chosen as

$$H^S_{w,N} = \text{span}\{H_{\mathbf{j}},\ \forall\, \mathbf{j} : |\mathbf{j}| \leq w\},$$

The dimension of the sparse space is much smaller than that of the full space for large N. For example, when $p(j) = j$, we have $\dim(H^S_{w,N}) = \sum_{|\mathbf{j}|\leq w} 1 = \frac{(N+w)!}{N!\,w!}$, which helps reducing the curse of dimensionality. This space corresponds to the space of polynomials of *total degree* at most $p(w)$.

The sparse interpolation formula can be written as a linear combination of Lagrange interpolations (17) on all tensor grids $\mathcal{H}^N_{\mathbf{j}}$. With $\mathcal{U}^{-1} = 0$, and for an index $j_n \geq 0$, define

$$\Delta^{j_n} := \mathcal{U}^{j_n} - \mathcal{U}^{j_n - 1}.$$

The isotropic Smolyak formula is then given by

$$\mathcal{A}_{w,N}\,[u](.,Y) = \sum_{|\mathbf{j}|\leq w} \left(\Delta^{j_1} \otimes \ldots \otimes \Delta^{j_N}\right) u(.,Y). \tag{19}$$

Equivalently, the formula (19) can be written as

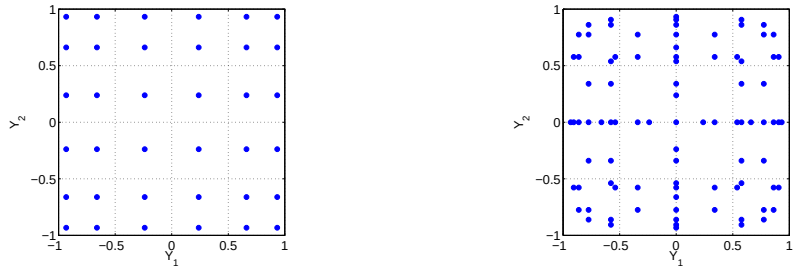
$$\mathcal{A}_{w,N}\,[u](.,Y) = \sum_{w-N+1\leq|\mathbf{j}|\leq w} (-1)^{w-|\mathbf{j}|} \binom{N-1}{w-|\mathbf{j}|} \mathcal{I}^N_{\mathbf{j}}\, u(.,Y). \tag{20}$$

The collection of all tensor grids used in the sparse interpolation formula is called the *sparse grid*,

$$\mathcal{H}^S_{w,N} = \bigcup_{w-N+1\leq|\mathbf{j}|\leq w} \mathcal{H}^N_{\mathbf{j}} \subset \Gamma.$$

Sparse interpolation implies evaluating $u_h(t,\mathbf{x},.)$ in all points of the sparse grid, known as collocation points. By construction, we have $\mathcal{A}_{w,N}\,[u](t,\mathbf{x},.) \in H^S_{w,N}$. Note that the number of collocation points is larger than the dimension of the approximating space $H^S_{w,N}$.

Example 3 Let N=2 and w=5, and consider $p(j) = j$. Moreover, let $Y = [Y_1, Y_2]$ be a random vector with independent and uniformly distributed random variables $Y_n \sim \mathcal{U}(-1,1)$. For a full tensor space, there are $(5+1)^2 = 36$ collocation points in the grid, shown in Fig. 1(a). For a sparse tensor space, there are $\frac{(2+5)!}{2!\,5!} = 21$ admissible sets of indices $\mathbf{j}$ and 89 collocation points in the grid shown in Fig. 1(b). Observe that the number of points in the full tensor grid grows much faster with the dimension N than the number of points in the sparse grid.



(a) Full grid

(b) Sparse grid

Figure 1: The full and sparse grids for a vector of two independent uniform random variables in $[-1, 1]$ with level $w = 5$.

3.3 Choice of interpolation abscissas

We propose two different abscissas in the construction of the Smolyak formula.

Gaussian abscissas. In this case, for a given index j_n, the sequence of abscissas $\{Y_{n,k_n}^{j_n}\}_{k_n=0}^{p(j_n)}$ are $p(j_n)+1$ zeros of the orthogonal polynomial $V_{p(j_n)+1}$. As the choice of the polynomial degree, we can use either the formula (15) or (16).

Clenshaw-Courtis abscissas. These abscissas are the extrema of Chebyshev polynomials and are given by

$$Y_{n,k_n}^{j_n} = -\cos\Big(\frac{\pi\, k_n}{p(j_n)}\Big), \qquad k_n = 0, \ldots, p(j_n).$$

It is recommended to use the formula (16) for the polynomial degree. In this case, one obtains nested sets of abscissas and thereby $\mathcal{H}_{w,N}^S \subset \mathcal{H}_{w+1,N}^S$.

We note that the structure of the stochastic collocation method, which involves solving η independent problems, allows us to use and perform parallel computations in a straight forward way.

4 Convergence analysis for stochastic collocation

In this section, we consider a linear bounded operator $\Psi(u)$ and give a priori estimates for the total error $\Psi(u) - \Psi(u_{h,w})$ in the norm $L^2(0,T;L^2(D)) \otimes L_\rho^2(\Gamma)$ when $\Psi(u) = u$, and in the norm $L_\rho^2(\Gamma)$ when $\Psi(u) = \mathcal{Q}(Y)$ with $\mathcal{Q}$ given in (13). We split the error into two parts and write

$$\varepsilon := ||\Psi(u) - \Psi(u_{h,w})|| \leq ||\Psi(u) - \Psi(u_h)|| + ||\Psi(u_h) - \Psi(u_{h,w})|| =: \varepsilon_I + \varepsilon_{II}. \quad (21)$$

The first term in the right hand side ε_I controls the convergence of the deterministic numerical scheme with respect to the mesh size h and is of order $\mathcal{O}(h^r)$,

where r is the minimum between the order of accuracy of the finite element or finite difference method used and the regularity of the solution. Notice that the constant in the term $\mathcal{O}(h^r)$ is uniform with respect to Y.

Here, we focus on the second term ε_{II} which is an interpolation error in the stochastic space. We first consider the case when $h \to 0$. We show that the error decays algebraically with respect to the number of collocation points η with an exponent proportional to $-s$, provided there are s bounded Y-derivatives (i.e., $\partial^s_{Y_n} \Psi < \infty$ with $n = 1, \ldots, N$) when the full tensor interpolation is used, and if the mixed Y-drivatives (i.e., $\partial^s_{Y_1} \partial^s_{Y_2} \ldots \partial^s_{Y_N} \Psi < \infty$) are bounded when the Smolyak interpolation is used. We next consider the case when $h^\beta\, w$, with $1 \leq \beta \leq 2$, is large. In this case, we show that the approximate solution u_h is Y-analytic with the radius of analyticity proportional to h^β. We therefore obtain a "fast" exponential rate of convergence which deteriorates as the quantity $h^\beta\, w$ gets smaller. The effective error ε_{II} will then be the minimum of the two errors corresponding to the case when $h \to 0$ and when $h^\beta\, w$ is large.

4.1 The case when $h \to 0$

We only consider the operator $\Psi(u_h) = u_h$ and let $h \to 0$. The discrete solution u_h has then a Y-regularity of order $s = 1$ as the continuous solution u, i.e. $\partial_Y u_h \in C^0(\Gamma; W)$, where $W := L^2(0, T; L^2(D))$, see Sec. 2. The second term of the error ε_{II} will then be in the norm $L^2(0, T; L^2(D)) \otimes L^2_\rho(\Gamma)$. We notice that for the case $\Psi(u) = \mathcal{Q}(Y)$, where $\mathcal{Q}$ is the quantity of interest in (13) with compactly supported smooth mollifiers whose supports does not cross the interfaces, the corresponding error estimates are obtained by replacing $s = 1$ with $s \geq 2$.

The technique for obtaining error bounds for multivariate functions (when $N > 1$) is based on one-dimensional results. We first quote a useful result from Erdös and Turán [8] for univariate functions.

Lemma 3. *Let $N = 1$ and $\Gamma \subset \mathbb{R}$ be bounded. Let W be a Hilbert space. For every function $v \in C^0(\Gamma; W)$ the interpolation error with Lagrange polynomials based on Gaussian abscissas satisfies*

$$||v - \mathcal{U}^j(v)||_{L^2_\rho(\Gamma;W)} \leq 2 \inf_{v_0 \in W \otimes H_{p(j)}} ||v - v_0||_{L^\infty(\Gamma;W)}. \tag{22}$$

We then recall a Jackson-type theorem on the error of the best approximation of univariate functions with bounded derivatives by algebraic polynomials, see [4] for instance.

Lemma 4. *Let $N = 1$ and $\Gamma \subset \mathbb{R}$ be bounded. Set $W := L^2(0, T; L^2(D))$. Given a function $v \in C^0(\Gamma; W)$ with $s \geq 0$ bounded derivatives in Y, there holds*

$$E_p(v) := \min_{v_0 \in W \otimes H_p} ||v - v_0||_{L^\infty(\Gamma;W)} \leq C_s\, p^{-s} \max_{k=0,\ldots,s} ||D^k_Y v||_{L^\infty(\Gamma;W)}, \tag{23}$$

where the constant C_s depends only on s.

We consider one random variable $Y_n \in \Gamma_n$ with density ρ_n and denote by $\hat{Y}_n \in \hat{\Gamma}_n$ the remainder $N-1$ variables with density $\hat{\rho}_n = \Pi_{k=1,k\neq n}^N \rho_k$. We can now consider $u_h^{(n)} := u_h(.,Y_n,\hat{Y}_n) : \Gamma_n \to W_n$ as a univariate function of Y_n with values in the Hilbert space $W_n = W \otimes L^2_{\hat{\rho}_n}$. We are ready to prove the following result.

Theorem 4. *Consider the isotropic full tensor product interpolation formula (18), and let $u_{h,w} = \mathcal{I}_{w,N}[u_h]$. Then the interpolation error ε_{II} defined in (21) satisfies*

$$\varepsilon_{II} \leq C\, p(w)^{-s},$$

where the constant $C = C_s \sum_{n=1}^N \max_{k=0,\dots,s} ||D^k_{Y_n} u_{h,w}||_{L^\infty(\Gamma;W)}$ does not depend on w. Here, $p(w)$ is either w or 2^w depending on the choice of formula (15) or (16) for the polynomial degree, respectively.

Moreover, let η be the total number of collocation points, then

$$\varepsilon_{II} \leq \frac{C}{2}\, \eta^{-s/N}. \tag{24}$$

Proof. We consider the first random variable Y_1 and the corresponding one-dimensional Lagrange interpolation operator $\mathcal{I}_{w,1} = \mathcal{U}^w : C^0(\Gamma_1;W_1) \to L^2_{\rho_1}(\Gamma_1;W_1)$. The global interpolation $\mathcal{I}_{w,N}$ can be written as the composition of two interpolations operators, $\mathcal{I}_{w,N} = \mathcal{I}_{w,1} \circ \hat{\mathcal{I}}_{w,1}$, where $\hat{\mathcal{I}}_{w,1} : C^0(\hat{\Gamma}_1;W) \to L^2_{\rho_1}(\hat{\Gamma}_1;W)$ is the interpolation operator in all directions $Y_2,\dots,Y_N$ except Y_1. We have,

$$\varepsilon_{II} = ||u_h - \mathcal{I}_{w,N}[u_h]||_{L^2_\rho(\Gamma;W)} \leq \underbrace{||u_h - \mathcal{I}_{w,1}[u_h]||}_{\varepsilon_{II_1}} + \underbrace{||\mathcal{I}_{w,1}\big[u_h - \hat{\mathcal{I}}_{w,1}[u_h]\big]||}_{\varepsilon_{II_2}}.$$

By (22) and (23), we can bound the first term,

$$\varepsilon_{II_1} \leq C\, p(w)^{-s}, \qquad C = 2\, C_s \max_{k=0,\dots,s} ||D^k_{Y_1} u_{h,w}||_{L^\infty(\Gamma;W)}.$$

To bound the second term we use the inequality (see Lemma 4.2 in [1]), $||\mathcal{I}_{w,1}[v]||_{L^2_\rho(\Gamma;W)} \leq \tilde{C}\, ||v||_{L^\infty(\Gamma;W)}$, with $v \in C^0(\Gamma;W)$, for $v = u_h - \hat{\mathcal{I}}_{w,1}[u_h]$ and write

$$\varepsilon_{II_2} \leq \tilde{C}\, ||u_h - \hat{\mathcal{I}}_{w,1}[u_h]||_{L^\infty(\Gamma;W)}.$$

The right hand side is again an interpolation error in the remainder $N-1$ directions $Y_2,\dots,Y_N$. We can proceed iteratively and define an interpolation operator in direction Y_2 and so forth. Finally we arrive at

$$||u_h - \mathcal{I}_{w,N}[u_h]|| \leq C_s\, p(w)^{-s} \sum_{n=1}^N \max_{k=0,\dots,s} ||D^k_{Y_n} u_{h,w}||_{L^\infty(\Gamma;W)} =: C\, p(w)^{-s}.$$

Note that C_s denotes a positive constant depending on s whose value may change from one expression to another expression. This proves the first inequality. The second inequality follows noting that the total number of collocation points is $\eta = \big(p(w)+1\big)^N$. □

Remark 5. *If the anisotropic full tensor interpolation [28] is used, the number of collocation points is* $\eta = \prod_{n=1}^{N}(p(w_n)+1)$*, where* w_n *is the level in the* n*-th direction. In this case the error satisfies*

$$\varepsilon_{II} \leq C_s \sum_{n=1}^{N} D_n\, p(w_n)^{-s}, \quad D_n := \max_{k=0,\ldots,s} ||D_{Y_n}^k u_{h,w}||_{L^\infty(\Gamma;W)}.$$

In order to minimize the computational work η *subject to the constraint* $\varepsilon_{II} \leq TOL$*, we introduce the Lagrange function* $\mathcal{L} = \eta + \lambda\,(C_s \sum_{n=1}^{N} D_n\, p(w_n)^{-s} - TOL)$*, with the Lagrange multiplier* λ*. By equating the partial derivative of* $\mathcal{L}$ *with respect to* $p(w_n)$ *to zero, we obtain* $p(w_n) \propto D_n^{1/s}$*. Noting that* D_n *can be computed easily using just a few samples of* Y_n*, we can quickly build a fast way on how to choose polynomial degrees in different directions and build the anisotropic full tensor grid.*

To obtain error estimates using the isotropic Smolyak interpolation, we first recall another Jackson-type theorem on the error of the best approximation of univariate functions with bounded derivatives by algebraic polynomials, see [7] for instance.

Lemma 5. *Let* $N = 1$ *and* $\Gamma \subset \mathbb{R}$ *be bounded. Let* W *be a Hilbert space. For every function* $v \in L^2_\rho(\Gamma;W)$ *with* $s \geq 1$ *square integrable* Y*-derivatives, the interpolation error with Lagrange polynomials based on Gauss-Legendre abscissas satisfies*

$$||v - \mathcal{U}^j(v)||_{L^2_\rho(\Gamma;W)} \leq C_s\, ||\rho||_\infty^{1/2}\, p(j)^{-s} \max_{k=0,\ldots,s} ||D_Y^k v||_{L^2(\Gamma;W)}, \tag{25}$$

where the constant C_s *depends only on* s*.*

We also need the following lemma,

Lemma 6. *In the isotropic Smolyak formula (19), with* $p(j)$ *given by (16), if*

$$\begin{aligned} ||\Delta^{j_n} u_h^{(n)}||_{L^2_{\rho_n}(\Gamma_n;W_n)} &= ||(\mathcal{U}^{j_n} - \mathcal{U}^{j_n-1})u_h^{(n)}|| \\ &\leq 2\, C_s\, ||\rho_n||_\infty^{1/2}\, 2^{-s(j_n-1)} \max_{k_n=0,\ldots,s} ||D_{Y_n}^{k_n} u_h^{(n)}||_{L^2(\Gamma_n;W_n)}, \end{aligned}$$

then

$$||\bigotimes_{n=1}^{N} \Delta^{j_n} u_h||_{L^2_\rho(\Gamma;W)} \leq (2\, C_s)^N\, ||\rho||_\infty^{1/2}\, 2^{-s\sum_{n=1}^{N}(j_n-1)} \max_{0\leq k_1,\ldots,k_N\leq s} ||D_{Y_1}^{k_1} \ldots D_{Y_N}^{k_N} u_h||_{L^2(\Gamma;W)}. \tag{26}$$

Proof. We write

$$||\bigotimes_{n=1}^{N}\Delta^{j_n}u_h||^2_{L^2_\rho(\Gamma;W)} = \int_{\Gamma_1}\dots\int_{\Gamma_{N-1}}\Big[\int_{\Gamma_N}||\Delta^{j_N}\bigotimes_{n=1}^{N-1}\Delta^{j_n}u_h||^2_W\,\rho_N\,dY_N\Big]\rho_1\dots\rho_{N-1}\,dY_1\dots dY_{N-1}$$

$$\le (2C_s)^2\,2^{-2\,s\,(j_N-1)}\,||\rho_N||_\infty\int_{\Gamma_1}\dots\int_{\Gamma_{N-1}}\max_{k_N=0,\dots,s}\int_{\Gamma_N}||D^{k_N}_{Y_N}\bigotimes_{n=1}^{N-1}\Delta^{j_n}u_h||^2_W\,\rho_1\dots\rho_{N-1}\,dY$$

$$\le (2C_s)^2\,2^{-2\,s\,(j_N-1)}\,||\rho_N||_\infty\max_{k_N=0,\dots,s}\int_{\Gamma_N}\int_{\Gamma_1}\dots\int_{\Gamma_{N-1}}||\Delta^{j_{N-1}}\bigotimes_{n=1}^{N-2}\Delta^{j_n}D^{k_N}_{Y_N}u_h||^2_W\,\rho_1\dots\rho_{N-1}\,dY.$$

If we repeat the process, we finally arrive at (26). □

We can now prove the following result,

Theorem 5. *Consider the sparse tensor product interpolation formula (20) based on Gauss-Legendre abscissas when the formula (16) is used, and let* $u_{h,w} = \mathcal{A}_{w,N}[u_h]$. *Then for the discrete solution* u_h *with* $s \ge 1$ *bounded mixed derivatives in* Y, *the interpolation error* ε_{II} *defined in (21) satisfies*

$$\varepsilon_{II} \le \hat{C}\,(w+1)^{2\,N}\,2^{-s\,(w+1)},$$

with $\hat{C} = \frac{C_0}{2}\,\frac{1-C_0^N}{1-C_0}\,||\rho||^{1/2}_\infty\,\max_{d=1,\dots,N}D_d(u_h)$, *where* $C_0 = 2^{s+1}\,C_s$ *and*

$$D_d(u_h) := \max_{0\le k_1,\dots,k_d\le s}||D^{k_1}_{Y_1}\dots D^{k_d}_{Y_d}u_h||_{L^2(\Gamma;W)} \tag{27}$$

Here, the constant $\hat{C}$ *depends on the solution,* s *and* N, *but not on* w.

Moreover, let η *be the total number of collocation points, then*

$$\varepsilon_{II} \le \hat{C}\,\big(1+\log_2\frac{\eta}{N}\big)^{2\,N}\,\eta^{-s\,\frac{\log 2}{\xi+\log N}}, \tag{28}$$

with $\xi = 1+\log 2\,(1+\log_2 1.5) \approx 2.1$.

Proof. We follow [4] and start with rewriting the isotropic Smolyak formula (19) as

$$\mathcal{A}_{w,N} = \sum_{|\mathbf{j}|\le w}\bigotimes_{n=1}^{N}\Delta^{j_n} = \sum_{\sum_{n=1}^{N-1}j_n\le w}\Big[\big(\bigotimes_{n=1}^{N-1}\Delta^{j_n}\big)\otimes\big(\sum_{k=0}^{w-\sum_{n=1}^{N-1}j_n}\Delta^k\big)\Big]$$

$$= \sum_{\sum_{n=1}^{N-1}j_n\le w}\Big[\big(\bigotimes_{n=1}^{N-1}\Delta^{j_n}\big)\otimes\big(\mathcal{U}^{w-\sum_{n=1}^{N-1}j_n}\big)\Big].$$

Let $I_N : \Gamma\to\Gamma$ be the identity operator on an N-dimensional space and $I_1^{(n)}$: $\Gamma_n\to\Gamma_n$ be a one-dimensional identity operator for $n = 1,\dots,N$. We can

compute the error operator recursively,

$$\begin{aligned}\mathcal{E}_N &:= I_N - \mathcal{A}_{w,N} \\ &= I_N - \sum_{\sum_{n=1}^{N-1} j_n \le w} \Big[\big(\bigotimes_{n=1}^{N-1} \Delta^{j_n}\big) \otimes \big(\mathcal{U}^{w-\sum_{n=1}^{N-1} j_n} - I_1^{(N)}\big)\Big] - \sum_{\sum_{n=1}^{N-1} j_n \le w} \Big[\big(\bigotimes_{n=1}^{N-1} \Delta^{j_n}\big) \otimes I_1^{(N)}\Big].\end{aligned}$$

Noting that $\sum_{\sum_{n=1}^{N-1} j_n \le w} \bigotimes_{n=1}^{N-1} \Delta^{j_n} = \mathcal{A}_{w,N-1}$ and that $I_N = I_{N-1} \otimes I_1^{(N)}$, we can write

$$\mathcal{E}_N = \sum_{\sum_{n=1}^{N-1} j_n \le w} \Big[\big(\bigotimes_{n=1}^{N-1} \Delta^{j_n}\big) \otimes \big(I_1^{(N)} - \mathcal{U}^{w-\sum_{n=1}^{N-1} j_n}\big)\Big] + \mathcal{E}_{N-1} \otimes I_1^{(N)}.$$

If we repeat the process, we arrive at

$$\mathcal{E}_N = \sum_{d=2}^{N} \Big[\tilde{R}(w,d) \bigotimes_{n=d+1}^{N} I_1^{(n)}\Big] + \big(I_1^{(1)} - \mathcal{A}_{w,1}\big) \bigotimes_{n=2}^{N} I_1^{(n)},$$

where

$$\tilde{R}(w,d) = \sum_{\sum_{n=1}^{d-1} j_n \le w} \Big[\big(\bigotimes_{n=1}^{d-1} \Delta^{j_n}\big) \otimes \big(I_1^{(d)} - \mathcal{U}^{w-\sum_{n=1}^{d-1} j_n}\big)\Big]. \tag{29}$$

Then,

$$||(I_N - \mathcal{A}_{w,N})[u_h]|| \le \sum_{d=2}^{N} ||\big(\tilde{R}(w,d) \bigotimes_{n=d+1}^{N} I_1^{(n)}\big)[u_h]|| + ||\big((I_1^{(1)} - \mathcal{A}_{w,1}) \bigotimes_{n=2}^{N} I_1^{(n)}\big)[u_h]||, \tag{30}$$

where the norms are in $L^2_\rho(\Gamma; W)$. We first bound $\tilde{R}$. By (25), we have

$$||(I_1^{(n)} - \mathcal{U}^{j_n})(u_h^{(n)})||_{L^2_{\rho_n}(\Gamma_n;W_n)} \le C_s\, ||\rho_n||_\infty^{1/2}\, 2^{-s\, j_n} \max_{k_n=0,\ldots,s} ||D_{Y_n}^{k_n} u_h^{(n)}||_{L^2(\Gamma_n;W_n)}, \tag{31}$$

and therefore,

$$\begin{aligned}||\Delta^{j_n}(u_h^{(n)})||_{L^2_{\rho_n}(\Gamma_n;W_n)} &= ||(\mathcal{U}^{j_n} - \mathcal{U}^{j_n-1})(u_h^{(n)})||_{L^2_{\rho_n}(\Gamma_n;W_n)} \\ &\le 2\, C_s\, ||\rho_n||_\infty^{1/2}\, 2^{-s\,(j_n-1)} \max_{k_n=0,\ldots,s} ||D_{Y_n}^{k_n} u_h^{(n)}||_{L^2(\Gamma_n;W_n)}.\end{aligned}$$

By Lemma 6 and (29) and (31), we then have

$$\begin{aligned}||\tilde{R}(w,d)[u_h]||_{L^2_\rho(\Gamma;W)} &\le \sum_{\sum_{n=1}^{d-1} j_n \le w} \frac{(2\, C_s)^d}{2}\, ||\rho||_\infty^{1/2}\, 2^{-s\,(w-d+1)}\, D_d(u_h) \\ &= \binom{w+d-1}{w} \frac{(2\, C_s)^d}{2}\, ||\rho||_\infty^{1/2}\, 2^{-s\,(w-d+1)}\, D_d(u_h),\end{aligned}$$

with $D_d(u_h)$ given by (27). Moreover, since

$$||(I_1^{(1)} - \mathcal{A}_{w,1})[u_h]||_{L^2_\rho(\Gamma;W)} = ||(I_1^{(1)} - \mathcal{U}^w)[u_h^{(1)}]||_{L^2_{\rho_n}(\Gamma_n;W_n)}$$
$$\leq C_s\,||\rho_1||_\infty^{1/2}\,2^{-s\,w} \max_{k_1=0,\ldots,s} ||D_{Y_1}^{k_1} u_h||_{L^2(\Gamma_1;W_1)} \leq C_s\,||\rho||_\infty^{1/2}\,2^{-s\,w} \max_{k_1=0,\ldots,s} ||D_{Y_1}^{k_1} u_h||_{L^2(\Gamma;W)},$$

then by (30), we get

$$||(I_N - \mathcal{A}_{w,N})[u_h]||_{L^2_\rho(\Gamma;W)} \leq \frac{1}{2}\,||\rho||_\infty^{1/2} \sum_{d=1}^{N} \binom{w+d-1}{w} (2\,C_s)^d\,2^{-s\,(w-d+1)}\,D_d(u_h)$$
$$\leq \frac{1}{2}\,||\rho||_\infty^{1/2}\,2^{-s\,(w+1)} \max_{d=1,\ldots,N} D_d(u_h) \sum_{d=1}^{N} \binom{w+d-1}{w} (2^{s+1}\,C_s)^d.$$

The first inequality stated in Theorem 5 follows noting that $\binom{w+d-1}{w} \leq (w+1)^{2\,N}$ for $d = 1, \ldots, N$.

To show the second inequality (28), we note that the number of collocation points η at level w using the Smolyak formula with Gaussian abscissas and the polynomial degree (16) satisfies (see Lemma 3.17 in [29])

$$\frac{\log \eta}{\xi + \log N} \leq w + 1 \leq 1 + \log_2 \frac{\eta}{N}, \tag{32}$$

with $\xi = 1 + \log 2\,(1 + \log_2 1.5) \approx 2.1$. From the first inequality we have

$$||(I_N - \mathcal{A}_{w,N})[u_h]||_{L^2_\rho(\Gamma;W)} \leq \hat{C}\,\big(1 + \log_2 \frac{\eta}{N}\big)^{2\,N}\,2^{-s\,\frac{\log \eta}{\xi + \log N}}.$$

This completes the proof. □

Remark 6. *We note that the above estimates are uniform with respect to h in the case of smooth quantity of interest. For the solution, we have one Y-derivative uniformly bounded with respect to h in* $L^2(0,T;H_0^1(D))$.

Remark 7. *(algebraic rate of convergence) In full tensor interpolation, with the minimal assumptions (2) on the data, by (24) we have an upper error bound of order* $\mathcal{O}(\eta^{-s/N})$ *with* $s = 1$ *when* $\Psi(u) = u$ *and with* $s \geq 2$ *when* $\Psi(u) = \mathcal{Q}(Y)$. *In Smolyak interpolation, with the minimal assumptions (2), when* $\Psi(u) = u$, *then (28) implies an upper error bound of order* $\mathcal{O}(\eta^{-\delta\,s})$ *with* $s = 1$ *and some* $0 < \delta < 1$ *only when* $N = 1$ *for which* $D_d(u_h)$ *is bounded. As we showed in Sect. 2,* $D_d(u_h)$, *which involves mixed Y-derivatives of the solution for* $N \geq 2$, *is not bounded. This gives an algebraic error convergence for the solution when* $N = 1$. *When* $\Psi(u) = \mathcal{Q}(Y)$, *with the minimal assumptions (2),* $\mathcal{Q}(Y)$ *has* $s \geq 2$ *bounded mixed derivatives for* $N \geq 1$, *as shown in Sect. 2. We obtain an upper error bound of order* $\mathcal{O}(\eta^{-\delta\,s})$ *with* $s \geq 2$ *and some* $0 < \delta < 1$. *This gives a faster error convergence for the quantity of interest (13).*

Remark 8. *(full tensor versus sparse tensor) The slowdown effect that the dimension N has on the error convergence (24) when a full tensor product is employed is known as the curse of dimensionality. This is the main reason for not using isotropic full tensor interpolation when N is large. On the other hand, the isotropic Smolyak approximation has a larger exponent $\mathcal{O}(\frac{1}{\log N})$ in (28) compared to $\mathcal{O}(\frac{1}{N})$ in (24). This is a clear advantage of the isotropic Smolyak method over the full tensor method when bounded mixed Y-derivatives exisit.*

Remark 9. *(computational cost versus error) In order to find the optimal choice of the mesh size h, we need to minimize the computational complexity of the stochastic collocation method, η/h^{d+1}, subject to the total error constraint $\varepsilon_F \propto h^r + \eta^{-s/N} = TOL$ for the isotropic full tensor interpolation and $\varepsilon_S \propto h^r + \eta^{-s/\log N} = TOL$ for the isotropic Smolyak interpolation. We introduce the Lagrange functions $\mathcal{L}_F = \eta/h^{d+1} + \lambda\,(h^r + \eta^{-s/N} - TOL)$ and $\mathcal{L}_S = \eta/h^{d+1} + \lambda\,(h^r + \eta^{-s/\log N} - TOL)$, with the Lagrange multiplier λ. By equating the partial derivatives of the Lagrange functions with respect to η, h, and λ to zero, we obtain $h^r \approx TOL/(1 + \frac{r\,N}{s\,(d+1)})$ and $h^r \approx TOL/(1 + \frac{r\,\log N}{s\,(d+1)})$, making the computational works of order $TOL^{-N/s-(d+1)/r}$ and $TOL^{-\log N/s-(d+1)/r}$ for the full tensor and Smolyak interpolations, respectively.*

4.2 The case when $h^\beta\, w$ is large with $1 \le \beta \le 2$

We consider a finite element approximation of (1) using a quasi-uniform triangulation of the physical domain. Let h denote the size of the largest triangle in the triangulation and u_h be the semi-discrete solution. We leave $t \in [0, T]$ and $Y \in \Gamma$ continuous. The semi-discrete problem reads

$$\int_D \partial_{tt} u_h\, v\, d\mathbf{x} + \int_D a^2\, \nabla u_h \cdot \nabla v\, d\mathbf{x} = \int_D f\, v\, d\mathbf{x}. \tag{33}$$

We differentiate the semi-discrete equation (33) with respect to the random variable Y_n. We then set $\tilde{u} := \partial_{Y_n} u_h$ and let $v = \tilde{u}_t$ to obtain

$$(\tilde{u}_{tt}, \tilde{u}_t) + B[\tilde{u}, \tilde{u}_t] = -A_1[u_h, \tilde{u}_t], \tag{34}$$

where

$$(v_1, v_2) := \int_D v_1\, v_2\, d\mathbf{x}, \qquad B[v_1, v_2] := \int_D a^2\, \nabla v_1 \cdot \nabla v_2\, d\mathbf{x},$$

$$A_1[v_1, v_2] := \int_{D_n} 2\, a\, a_{Y_n}\, \nabla v_1 \cdot \nabla v_2\, d\mathbf{x}.$$

We observe that $(\tilde{u}_{tt}, \tilde{u}_t) = \frac{1}{2}\frac{d}{dt}\|\tilde{u}_t\|^2_{L^2(D)}$. Moreover, since a is time-independent, then $B[\tilde{u}, \tilde{u}_t] = \frac{1}{2}\frac{d}{dt}B[\tilde{u}, \tilde{u}] = \frac{1}{2}\frac{d}{dt}\|a\,\nabla\tilde{u}\|^2_{L^2(D)}$. Furthermore, by Hölder, inverse

and Cauchy inequalities [9], we have

$$\begin{aligned}|A_1[u_h,\tilde{u}_t]| &\leq C_n \,\|\nabla u_h\|_{L^2(D_n)}\,\|\nabla \tilde{u}_t\|_{L^2(D_n)}\\ &\leq C_n\, C_{inv}\, h^{-1}\,\|\nabla u_h\|_{L^2(D_n)}\,\|\tilde{u}_t\|_{L^2(D_n)}\\ &\leq \frac{T}{2}\, C_n^2\, C_{inv}^2\, h^{-2}\,\|\nabla u_h\|^2_{L^2(D)} + \frac{1}{2\,T}\,\|\tilde{u}_t\|^2_{L^2(D)},\end{aligned}$$

where $C_n := 2\,\|a\, a_{Y_n}\|_{L^\infty(D_n\times\Gamma_n)}$, and C_{inv} is the constant in the inverse inequality. From (34) we therefore get

$$\frac{d}{dt}\|\tilde{u}_t\|^2_{L^2(D)} + \frac{d}{dt}\|a\,\nabla\tilde{u}\|^2_{L^2(D)} \leq \frac{1}{T}\,\|\tilde{u}_t\|^2_{L^2(D)} + T\, C_n^2\, C_{inv}^2\, h^{-2}\,\|\nabla u_h\|^2_{L^2(D)}. \quad (35)$$

Now write

$$y_1 := \|\tilde{u}_t\|^2_{L^2(D)} + \|a\,\nabla\tilde{u}\|^2_{L^2(D)}, \quad y_2 := T\, C_n^2\, C_{inv}^2\, h^{-2}\,\|\nabla u_h\|^2_{L^2(D)}.$$

From the inequality (35), we have $y_1'(t) \leq \frac{1}{T}\, y_1(t) + y_2(t)$. By the Gronwall's inequality [9] and noting that $y_1(0) = 0$, we obtain

$$\|\tilde{u}_t\|^2_{L^2(D)} + \|a\,\nabla\tilde{u}\|^2_{L^2(D)} \leq e\, T\, C_n^2\, C_{inv}^2\, h^{-2} \int_0^T \|\nabla u_h\|^2_{L^2(D)}\, dt. \quad (36)$$

We now define the energy norm

$$\|u_h\|^2_E := \sup_{\substack{t\in(0,T)\\ Y\in\Gamma}} \big(\|\partial_t u_h(t,.)\|^2_{L^2(D)} + \|a\,\nabla u_h(t,.)\|^2_{L^2(D)}\big),$$

and the Sobolev norm

$$\|u_h\|^2_S := \sup_{\substack{t\in(0,T)\\ Y\in\Gamma}} \big(\|\partial_t u_h(t,.)\|^2_{L^2(D)} + \|\nabla u_h(t,.)\|^2_{L^2(D)}\big).$$

We consider two different cases. One case is when the uniform coercivity assumption (3) holds. The other case is when the wave speed $a(\mathbf{x}, Y)$ may be zero or negative due to possible negative values in the random vector Y, and therefore (3) does not hold.

4.2.1 The case of uniformly coercive wave speed

Under the uniform coercivity assumption (3), we have

$$\|u_h\|_S \leq \frac{1}{\tilde{a}_{min}}\,\|u_h\|_E, \qquad \tilde{a}_{min} := \min\{a_{min}, 1\} > 0.$$

Moreover, by (36), we obtain

$$\|\tilde{u}\|^2_E \leq e\, T^2\, C_n^2\, C_{inv}^2\, h^{-2}\,\|u_h\|^2_S.$$

Therefore,

$$\|\partial_{Y_n} u_h\|_S \leq \frac{e^{1/2}\, T\, C_n\, C_{inv}}{h\, \tilde{a}_{min}} \|u_h\|_S. \tag{37}$$

We now obtain the estimate on the growth of all mixed Y-derivatives of u_h. Let $\mathbf{k} \in \mathbb{Z}_+^N$ be a multi-index and $\partial_Y^{\mathbf{k}} u_h := \frac{\partial^{|\mathbf{k}|} u_h}{\partial_{Y_1}^{k_1} \ldots \partial_{Y_N}^{k_N}}$. In order to find an upper bound for the $|\mathbf{k}|$-th order mixed Y-derivative $\partial_Y^{\mathbf{k}} u_h$, we follow [5] and introduce a set $\mathcal{K}$ of indices with cardinality $n_{\mathcal{K}}$ such that $\partial_Y^{\mathcal{K}} u_h := \frac{\partial^{n_{\mathcal{K}}} u_h}{\Pi_{k \in \mathcal{K}} \partial_{Y_k}} = \partial_Y^{\mathbf{k}} u_h$. As an example, let $N = 5$ and consider the set $\mathcal{K} = \{1, 1, 2, 3, 5, 5, 5\}$ with $n_{\mathcal{K}} = 7$. Then the corresponding multi-index is $\mathbf{k} = [2\,1\,1\,0\,3]$ with $|\mathbf{k}| = 7$, and we have

$$\partial_Y^{\mathcal{K}} u_h = \frac{\partial^7 u_h}{\partial_{Y_1}\, \partial_{Y_1}\, \partial_{Y_2}\, \partial_{Y_3}\, \partial_{Y_5}\, \partial_{Y_5}\, \partial_{Y_5}} = \frac{\partial^7 u_h}{\partial_{Y_1}^2\, \partial_{Y_2}\, \partial_{Y_3}\, \partial_{Y_5}^3} = \partial_Y^{\mathbf{k}} u_h.$$

Before deriving the estimates, we need the following two lemmas.

Lemma 7. (generalized Leibniz rule) *Given a set of indices $\mathcal{K}$ with cardinality $n_{\mathcal{K}}$ and two functions $f, g \in \mathcal{C}^{\mathcal{K}}(\Gamma)$,*

$$\partial_Y^{\mathcal{K}}(f\, g) = \sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K})} \partial_Y^{\mathcal{S}} f\; \partial_Y^{\mathcal{K} \setminus \mathcal{S}} g,$$

where $\mathcal{P}(\mathcal{K})$ represents the power set of $\mathcal{K}$.

Lemma 8. *Let $C \in \mathbb{R}_+$ and $n \in \mathbb{Z}_+$. Then we have*

$$C \sum_{i=0}^{n-1} \frac{(C+1)^i}{(n-i)!} \leq (C+1)^n. \tag{38}$$

Proof. The left hand side of (38) can be written as

$$\sum_{i=0}^{n-1} \frac{1}{(n-i)!} \sum_{j=0}^{i} \binom{i}{j} C^{j+1} = \sum_{j=0}^{n-1} C^{j+1} \sum_{i=j}^{n-1} \frac{\binom{i}{j}}{(n-i)!}.$$

The right hand side of (38) can be written as

$$1 + \sum_{j=1}^{n} \binom{n}{j} C^j = 1 + \sum_{j=0}^{n-1} \binom{n}{j+1} C^{j+1}.$$

We now show that

$$\sum_{i=j}^{n-1} \frac{\binom{i}{j}}{(n-i)!} \leq \binom{n}{j+1}, \quad 0 \leq j \leq n-1, \tag{39}$$

from which the inequality (38) follows. We prove (39) by induction on $n \geq j+1$.

Case $n = j + 1$**.** In this case (39) reads $1 \leq 1$ which is true.
General case. We assume that (39) holds for $n \geq j + 1$ and show that

$$\sum_{i=j}^{n} \frac{\binom{i}{j}}{(n+1-i)!} \leq \binom{n+1}{j+1}.$$

We can use the induction hypothesis (39) and write

$$\begin{aligned}\sum_{i=j}^{n} \frac{\binom{i}{j}}{(n+1-i)!} &= \sum_{i=j}^{n-1} \frac{\binom{i}{j}}{(n-i)!\,(n+1-i)} + \frac{\binom{n}{j}}{1!} \leq \frac{1}{n+1-j} \sum_{i=j}^{n-1} \frac{\binom{i}{j}}{(n-i)!} + \binom{n}{j} \\ &\leq \frac{1}{n+1-j} \binom{n}{j+1} + \binom{n}{j} \leq \binom{n}{j+1} + \binom{n}{j} = \binom{n+1}{j+1},\end{aligned}$$

where the last equality is the Pascal's rule. Therefore, by induction the proof is complete. □

We are now ready to prove the following result,

Theorem 6. *The Y-derivatives of the semi-discrete solution u_h which solves (33) can be bounded as*

$$\|\partial_Y^{\mathbf{k}} u_h\|_S \leq |\mathbf{k}|!\,(C+1)^{|\mathbf{k}|}\,\|u_h\|_S, \qquad C = \frac{\hat{C}\,T}{h\,\tilde{a}_{min}}, \tag{40}$$

where $\mathbf{k} \in \mathbb{Z}_+^N$ is a multi-index, and $\hat{C}$ is independent of h.

Proof. Let $\mathcal{K}$ be the index set corresponding to the multi-index $\mathbf{k}$. Then, according to Lemma 7, the $\partial_Y^{\mathcal{K}}$ derivative of the semi-discrete equation (33) is

$$\int_D \partial_Y^{\mathcal{K}} \partial_{tt} u_h\, v\, d\mathbf{x} + \int_D \sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K})} \partial_Y^{\mathcal{S}} \nabla u_h\, \partial_Y^{\mathcal{K}\setminus S} a^2 \cdot \nabla v\, d\mathbf{x} = 0.$$

Noting that $\mathcal{P}(\mathcal{K}) = \mathcal{K} \cup \big(\mathcal{P}(\mathcal{K}) \setminus \mathcal{K}\big)$, we write

$$\int_D \partial_Y^{\mathcal{K}} \partial_{tt} u_h\, v\, d\mathbf{x} + \int_D a^2\, \partial_Y^{\mathcal{K}} \nabla u_h \cdot \nabla v\, d\mathbf{x} = -\int_D \sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K})\setminus\mathcal{K}} \partial_Y^{\mathcal{S}} \nabla u_h\, \partial_Y^{\mathcal{K}\setminus S} a^2 \cdot \nabla v\, d\mathbf{x}.$$

Now let $v = \partial_Y^{\mathcal{K}} \partial_t u_h$ and obtain

$$\frac{1}{2}\frac{d}{dt}\|\partial_Y^{\mathcal{K}} \partial_t u_h\|_{L^2(D)}^2 + \frac{1}{2}\frac{d}{dt}\|a\,\partial_Y^{\mathcal{K}} \nabla u_h\|_{L^2(D)}^2 = -A_{\mathcal{K}}[u_h, \partial_Y^{\mathcal{K}} \partial_t u_h], \tag{41}$$

where

$$A_{\mathcal{K}}[v_1, v_2] := \sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K})\setminus\mathcal{K}} \int_D \partial_Y^{\mathcal{S}} \nabla v_1\, \partial_Y^{\mathcal{K}\setminus S} a^2 \cdot \nabla v_2\, d\mathbf{x}.$$

As before, by Hölder, inverse and Cauchy inequalities [9], we have

$$
\begin{aligned}
|A_{\mathcal{K}}[u_h, \partial_Y^{\mathcal{K}} \partial_t u_h]| &\leq \sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{K} \backslash S} a^2\|_{L^\infty(D)} \, \|\partial_Y^{\mathcal{S}} \nabla u_h\|_{L^2(D)} \, \|\partial_Y^{\mathcal{K}} \partial_t \nabla u_h\|_{L^2(D)} \\
&\leq \tilde{C} \, C_{inv} \, h^{-1} \, \|\partial_Y^{\mathcal{K}} \partial_t u_h\|_{L^2(D)} \sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{S}} \nabla u_h\|_{L^2(D)} \\
&\leq \frac{T}{2} \, \tilde{C}^2 \, C_{inv}^2 \, h^{-2} \Big(\sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{S}} \nabla u_h\|_{L^2(D)} \Big)^2 + \frac{1}{2\,T} \, \|\partial_Y^{\mathcal{K}} \partial_t u_h\|_{L^2(D)}^2,
\end{aligned}
$$

where $\tilde{C} := \max_{\mathcal{S} \in \mathcal{P}(\mathcal{K})} \|\partial_Y^{\mathcal{S}} a^2\|_{L^\infty(D \times \Gamma)}$, and C_{inv} is the constant in the inverse inequality. From (41) we therefore get

$$
\begin{aligned}
&\frac{d}{dt} \|\partial_Y^{\mathcal{K}} \partial_t u_h\|_{L^2(D)}^2 + \frac{d}{dt} \|a \, \partial_Y^{\mathcal{K}} \nabla u_h\|_{L^2(D)}^2 \leq \\
&\qquad \leq T \, \tilde{C}^2 \, C_{inv}^2 \, h^{-2} \Big(\sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{S}} \nabla u_h\|_{L^2(D)} \Big)^2 + \frac{1}{T} \, \|\partial_Y^{\mathcal{K}} \partial_t u_h\|_{L^2(D)}^2 \quad (42)
\end{aligned}
$$

Now we write

$$
\begin{aligned}
y_1 &:= \|\partial_Y^{\mathcal{K}} \partial_t u_h\|_{L^2(D)}^2 + \|a \, \partial_Y^{\mathcal{K}} \nabla u_h\|_{L^2(D)}^2, \\
y_2 &:= T \, \tilde{C}^2 \, C_{inv}^2 \, h^{-2} \Big(\sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{S}} \nabla u_h\|_{L^2(D)} \Big)^2.
\end{aligned}
$$

From the inequality (42), we have $y_1'(t) \leq \frac{1}{T} \, y_1(t) + y_2(t)$. By the Gronwall's inequality [9] and noting that $y_1(0) = 0$, we obtain

$$
\|\partial_Y^{\mathcal{K}} \partial_t u_h\|_{L^2(D)}^2 + \|a \, \partial_Y^{\mathcal{K}} \nabla u_h\|_{L^2(D)}^2 \leq e \, T \, \tilde{C}^2 \, C_{inv}^2 \, h^{-2} \int_0^T \Big(\sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{S}} \nabla u_h\|_{L^2(D)} \Big)^2 dt,
$$

and therefore,

$$
\|\partial_Y^{\mathcal{K}} u_h\|_E^2 \leq e \, T^2 \, \tilde{C}^2 \, C_{inv}^2 \, h^{-2} \sup_{t, Y} \Big(\sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{S}} \nabla u_h\|_{L^2(D)} \Big)^2.
$$

We finally obtain the formula

$$
\|\partial_Y^{\mathcal{K}} u_h\|_S \leq C \sum_{\mathcal{S} \in \mathcal{P}(\mathcal{K}) \backslash \mathcal{K}} \|\partial_Y^{\mathcal{S}} u_h\|_S, \qquad C = \frac{e^{1/2} \, T \, \tilde{C} \, C_{inv}}{h \, \tilde{a}_{min}}. \quad (43)
$$

We now by induction show that

$$
\|\partial_Y^{\mathcal{K}} u_h\|_S \leq n_{\mathcal{K}}! \, (C+1)^{n_{\mathcal{K}}} \, \|u_h\|_S, \quad (44)
$$

which is equivalent to the corresponding multi-index formulation (40).

Case $n_{\mathcal{K}} = 0$**.** In this case the set $\mathcal{K}$ is empty, and (44) reads $\|u_h\|_S \leq \|u_h\|_S$, which is true.
Case $n_{\mathcal{K}} = 1$**.** In this case $\mathcal{K} = \{k\}, 1 \leq k \leq N$, and (44) reads

$$\|\partial_{Y_k} u_h\|_S \leq (C+1)\,\|u_h\|_S,$$

which follows from (37).
General case. We now assume that (44) holds for all sets $\mathcal{S}$ with cardinality $0 \leq n_{\mathcal{S}} \leq n_{\mathcal{K}} - 1$. We have then the induction hypothesis,

$$\|\partial_Y^{\mathcal{S}} u_h\|_S \leq n_{\mathcal{S}}!\,(C+1)^{n_{\mathcal{S}}}\,\|u_h\|_S, \qquad 0 \leq n_{\mathcal{S}} \leq n_{\mathcal{K}} - 1. \tag{45}$$

From (43) we have

$$\begin{aligned}\|\partial_Y^{\mathcal{K}} u_h\|_S \leq C \sum_{\mathcal{S}\in\mathcal{P}(\mathcal{K})\setminus\mathcal{K}} \|\partial_Y^{\mathcal{S}} u_h\|_S &= C \sum_{i=0}^{n_{\mathcal{K}}-1} \sum_{\substack{\mathcal{S}\in\mathcal{P}(\mathcal{K})\\ n_{\mathcal{S}}=i}} \|\partial_Y^{\mathcal{S}} u_h\|_S \\ &\leq C \sum_{i=0}^{n_{\mathcal{K}}-1} \sum_{\substack{\mathcal{S}\in\mathcal{P}(\mathcal{K})\\ n_{\mathcal{S}}=i}} n_{\mathcal{S}}!\,(C+1)^{n_{\mathcal{S}}}\,\|u_h\|_S.\end{aligned}$$

Note that the number of subsets $\mathcal{S}$ of $\mathcal{P}(\mathcal{K})$ with cardinality i is $\binom{n_{\mathcal{K}}}{i}$. Then

$$\|\partial_Y^{\mathcal{K}} u_h\|_S \leq C \sum_{i=0}^{n_{\mathcal{K}}-1} i!\,(C+1)^i \binom{n_{\mathcal{K}}}{i} \|u_h\|_S \leq n_{\mathcal{K}}!\,(C+1)^{n_{\mathcal{K}}}\,\|u_h\|_S,$$

where the last inequality follows from Lemma 8. This completes the proof. □

Remark 10. *We note that the optimal choice of the mesh size h in Remark 9 in Section 4.1 is obtained by assuming that the Y-derivatives of the solution up to order s, which appear in the coefficients C and $\hat{C}$ in the error estimates (24) and (28), are uniformly bounded with respect to h. In the absence of such assumption, we can employ the estimate (40) and find the coefficients in the error bounds. For instance, for the full tensor interpolation, the coefficient C in the interpolation error (24) is $C \propto N\,h^{-s}$. The total error is then $\varepsilon_F \propto h^r + N\,h^{-s}\,\eta^{-s/N} = TOL$. By introducing the Lagrange function $\mathcal{L}_F = \eta/h^{d+1} + \lambda\,(h^r + N\,h^{-s}\,\eta^{-s/N} - TOL)$ and equating its partial derivatives with respect to η, h, and λ to zero, we obtain $h^r \approx TOL/(1 + \frac{r\,N}{s\,(N+d+1)})$, making the computational work of order $TOL^{-N/s-(N+d+1)/r}$.*

We now define for every $Y \in \Gamma$ the power series $u_h : \mathbb{C}^N \to L^\infty(0,T;H_0^1(D))$ as

$$u_h(t,\mathbf{x},Z) = \sum_{k=0}^{\infty} \sum_{|\mathbf{k}|=k} \frac{(Z-Y)^{\mathbf{k}}}{\mathbf{k}!}\,\partial_Y^{\mathbf{k}} u_h(t,\mathbf{x},Y), \tag{46}$$

where $\mathbf{k}! = \Pi_{n=1}^N (k_n!)$ and $Y^{\mathbf{k}} = \Pi_{n=1}^N Y_n^{k_n}$. By (40) we get

$$\|u_h(Z)\|_S \leq \sum_{k=0}^{\infty} \sum_{|\mathbf{k}|=k} \frac{(Z-Y)^{\mathbf{k}}}{\mathbf{k}!} \|\partial_Y^{\mathbf{k}} u_h(Y)\|_S \leq \sum_{k=0}^{\infty} \sum_{|\mathbf{k}|=k} \frac{|\mathbf{k}|!}{\mathbf{k}!} (C+1)^{|\mathbf{k}|} (Z-Y)^{\mathbf{k}} \|u_h(Y)\|_S.$$

We exploit the generalized Newton binomial formula for $\mathbf{v} = [v_1, \ldots, v_N] \in \mathbb{R}_+^N$ and $k \in \mathbb{Z}_+$,

$$\sum_{|\mathbf{k}|=k} \frac{|\mathbf{k}|!}{\mathbf{k}!} \mathbf{v}^{\mathbf{k}} = \big(\sum_{n=1}^N v_n\big)^k,$$

and obtain

$$\|u_h(Z)\|_S \leq \sum_{k=0}^{\infty} \Big(\sum_{n=1}^N (C+1)\,|Z_n - Y_n|\Big)^k \|u_h(Y)\|_S.$$

Therefore, the series (46) converges for all $Z \in \mathbb{C}^N$ such that $|Z_n - Y_n| \leq \tau < \frac{1}{N}(C+1)^{-1} = \mathcal{O}(h)$. By a continuation argument, the function u_h can analytically be extended on the whole region $\Sigma(\Gamma, \tau) = \{Z \in \mathbb{C}^N,\, dist(\Gamma_n, Z_n) \leq \tau,\, n = 1, \ldots, N\}$. We note that the radius of analyticity is proportional to h.

We now build an approximate solution $u_{h,w}$ to u_h based on Lagrange interpolation in Y. We investigate only the case of a tensor product interpolation on Gauss-Legendre points as described in Sect. 3. We recall a result on the error of the best approximation of univariate analytic functions by polynomials [1].

Lemma 9. *Let $N = 1$ and $\Gamma \subset \mathbb{R}$ be bounded. Set $W := L^2(0, T; L^2(D))$. Then, given a function $v(Y) \in L^\infty(\Gamma; W)$ which admits an analytic extension in the region of the complex plane $\Sigma(\Gamma, \tau) = \{Z \in \mathbb{C},\, dist(\Gamma, Z) \leq \tau\}$, for some $\tau > 0$, there holds*

$$E_p(v) := \min_{v_0 \in W \otimes H_p} ||v - v_0||_{L^\infty(\Gamma;W)} \leq \frac{2}{e^\sigma - 1}\, e^{-\sigma\, p} \max_{Z \in \Sigma} ||v(Z)||_W, \qquad (47)$$

where $0 < \sigma = \log\big(\frac{2\tau}{|\Gamma|} + \sqrt{1 + \frac{4\tau^2}{|\Gamma|^2}}\big)$.

In the above lemma, τ is smaller than the distance between Γ and the closest singularity of the extended function $v(z) : \mathbb{C} \to W$ in the complex plane.

In the multidimensional case when $N \geq 2$, we note that $\sigma_n = \log\big(\frac{2\tau}{|\Gamma_n|} + \sqrt{1 + \frac{4\tau^2}{|\Gamma_n|^2}}\big)$ depends on the direction n. We therefore set

$$\sigma^* = \min_{1 \leq n \leq N} \min_{\hat{Y}_n \in \hat{\Gamma}_n} \sigma_n, \qquad M^*(v) = \max_{1 \leq n \leq N} \max_{\hat{Y}_n \in \hat{\Gamma}_n} \max_{Z \in \Sigma(\Gamma, \tau)} ||v(Z)||_W.$$

Similar to the proof of Theorem 4, using (22) and (47), we can show that for the isotropic full tensor product interpolation formula (18), with $u_{h,w} = \mathcal{I}_{w,N}[u_h]$, the interpolation error ε_{II} defined in (21) satisfies

$$\varepsilon_{II} = \|u_h - u_{h,w}\|_{L^2_\rho(\Gamma;W)} \leq 2\, N\, M^*(u_h)\; e^{-\sigma^*\, p(w)}. \qquad (48)$$

We now consider the Smolyak interpolation formula (20) based on Gaussian abscissas when the formula (16) is used and let $u_{h,w} = \mathcal{A}_{w,N}[u_h]$. Similar to the proof of Lemma 3.16 in [29], we can show that the interpolation error ε_{II} defined in (21) satisfies

$$\varepsilon_{II} = \|u_h - u_{h,w}\|_{L^2_\rho(\Gamma;W)} \leq \hat{C}\, g(w), \qquad g(w) = \begin{cases} e^{-\sigma^*\, w\, e\, \log 2}, & 0 \leq w \leq \frac{N}{\log 2}, \\ e^{-\sigma^*\, N\, 2^{w/N}}, & otherwise, \end{cases} \tag{49}$$

with $\hat{C} = \frac{C_0}{2}\,\frac{1-C_0^N}{1-C_0}$ and $C_0 = \frac{16\, M^*(u_h)}{e^{4\,\sigma^*} - e^{2\,\sigma^*}}\,(1 + \frac{1}{\log 2}\,\sqrt{\frac{\pi}{2\,\sigma^*}})$.

From (48) and (49), we note that for both full tensor and Smolyak interpolations, since $\sigma^* = \mathcal{O}(h)$, we will have a fast exponential decay in the error when the product $h\,w$ is large. As a result, with a fixed h, the error convergence is slow (algebraic) for a small w and fast (exponential) for a large w. Moreover, the rate of convergence deteriorates as h gets smaller. These results are precisely what we observe in the numerical experiments presented in Sect. 5.

4.2.2 The case of non- coercive wave speed

We now relax the uniform coercivity assumption (3) and instead assume that

$$0 \leq a_{min} \leq a(\mathbf{x}, \omega) \leq a_{max} < \infty, \quad \forall \mathbf{x} \in D, \quad \forall \omega \in \Omega.$$

We apply the inverse inequality to (36) and write

$$\begin{aligned} \|\tilde{u}\|_E^2 &\leq e\, T\, C_n^2\, C_{inv}^4\, h^{-4} \int_0^T \|u_h\|_{L^2(D)}^2\, dt \\ &\leq e\, T^2\, C_n^2\, C_{inv}^4\, h^{-4} \sup_{t\in(0,T)} \|u_h(t)\|_{L^2(D)}^2 \\ &\leq e\, T^2\, C_n^2\, C_{inv}^4\, h^{-4}\, (T \sup_{t\in(0,T)} \|\partial_t u_h(t)\|_{L^2(D)} + \|u_h(0)\|_{L^2(D)})^2 \\ &\leq e\, T^4\, C_n^2\, C_{inv}^4\, h^{-4}\, \|u_h\|_E^2. \end{aligned}$$

In the last inequality, we assume for simplicity that $u_h(0) = 0$. Similar to Sect. 4.2.1, we obtain

$$\|\partial_Y^{\mathbf{k}} u_h\|_E \leq |\mathbf{k}|!\, (C_0 + 1)^{|\mathbf{k}|}\, \|u_h\|_E, \qquad C_0 := \frac{e^{1/2}\, T^2\, \tilde{C}\, C_{inv}^2}{h^2}. \tag{50}$$

Therefore, the series (46) converges for all $Z \in \mathbb{C}^N$ such that $|Z_n - Y_n| \leq \tau < \frac{1}{N}\,(C_0 + 1)^{-1} = \mathcal{O}(h^2)$. Comparing this with the case when the coercivity assumption (3) holds, we observe that as h decreases the radius of analyticity shrinks faster (proportional to h^2) for the non-coercive case than for the coercive case (proportional to h). We obtain the same estimates as (48) and (49) with $\sigma^* = \mathcal{O}(h^2)$. We will therefore have a fast exponential decay in the error when the product $h^2\, p(w)$ is large. Note that these estimates may not be sharp, as the numerical test 2 in Sect. 5 suggests that $\sigma^* \approx \mathcal{O}(h^{1.2})$. This may be related to the use of inverse inequality (which is not sharp) twice while obtaining (50).

5 Numerical examples

In this section, we consider the IBVP (1) in a two dimensional layered medium. We numerically simulate the problem by the stochastic collocation method and study the convergence of the statistical moments of the solution u, the linear quantity of interest (13) and a nonlinear quantity of interest called the Arias intensity

$$\mathcal{I}_A(Y) = \int_0^T \int_S |u_{tt}(t, \mathbf{x}, Y)|^2 \, d\mathbf{x} \, dt, \tag{51}$$

where, S is a sub-domain of the physical domain D, and T is a positive final time. We show that the computational results are in accordance with the convergence rates predicted by the theory.

We consider a rectangular physical domain $D = [-L_x, L_x] \times [-L_z, 0]$ and a random wave speed a of form (5) for a two-layered medium ($N = 2$). The computational domain containing two layers with widths d_1 and d_2 is shown in Fig. 2.

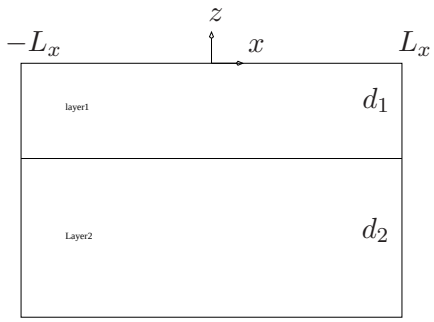


Figure 2: Two layered computational domain.

The deterministic solver employs a finite difference scheme based on second-order central difference approximation. Let $\Delta x = \frac{2\,L_x}{N_x}$ and $\Delta z = \frac{L_z}{N_z}$ denote the spatial grid-lengths, where N_x and N_z are natural numbers. For $i = 0, 1, \ldots, N_x$ and $j = 0, 1, \ldots, N_z$, let $(x_i, z_j) = (-L_x + i\,\Delta x, -L_z + j\,\Delta z)$ and $u_{i,j}(t)$ denote the corresponding grid point and the grid function approximating $u(t, x_i, z_j)$, respectively. On this spatial grid, we discretize the PDE in (1) and obtain the semi-discretization,

$$\frac{d^2 u_{i,j}(t)}{dt^2} = \frac{1}{\Delta x}\Big(a^2_{i+\frac{1}{2},j}\, D_{+i} u_{i,j}(t) - a^2_{i-\frac{1}{2},j}\, D_{-i} u_{i,j}(t)\Big) + \frac{1}{\Delta z}\Big(a^2_{i,j+\frac{1}{2}}\, D_{+j} u_{i,j}(t) - a^2_{i,j-\frac{1}{2}}\, D_{-j} u_{i,j}(t)\Big) + f_{i,j}(t).$$

Here, D_+ and D_- are forward and backward first-order difference operators, respectively. We then use the second-order central difference approximation in time to obtain the fully discrete deterministic scheme. In the stochastic space, we

use the isotropic Smolyak formula (20) based on Gaussian abscissas, described in Sect. 3.

We perform four numerical tests. In the first test, we consider a zero force term and smooth initial data and study the mean and standard deviation of the solution u. In the second test, we consider the same data as in the first test and select random variables so that the uniform coercivity assumption (3) is not satisfied, and we have $a_{min}^2 = 0$. We study the expected value of the solution u in this case and compare it with the case when $a_{min}^2 > 0$. In the third test, we consider zero initial data and a discontinuous time-independent forcing term and study the quantity of interest (13). Finally, in the fourth test, we study the Arias intensity (51) on the free surface due to a Ricker wavelet. In all computations, we use a time step-size $\Delta t = \Delta x/5$ which guarantees the stability of the deterministic numerical solver. We use homogeneous Neumann boundary conditions in all tests.

5.1 Numerical test 1

In the first test, we choose a computational domain $D = [-2, 2] \times [-3.5, 0]$ with two layers with widths $d_1 = 0.5$ and $d_2 = 3$. We consider a wave speed of form (5) with $a_0 = 0$, $\alpha_1 = 2$ and $\alpha_2 = 3$, and let $Y_n \sim \mathcal{U}(0.1, 0.5)$, $n = 1, 2$, be two independent and uniformly distributed random variables. We set $f = g_2 = 0$ and consider an initial Gaussian wave pulse,

$$g_1(x, z) = e^{-\frac{(x-x_c)^2}{2\,\sigma_x^2} - \frac{(z-z_c)^2}{2\,\sigma_z^2}}.$$

For computing the convergence rate of error, we consider a set of spatial grid-lenghts $\Delta x = \Delta z = 0.1, 0.05, 0.025, 0.0125$. For each grid-length $\Delta x = h$, we consider different levels $w \geq 1$ and compute the L^2-norm of error in the expected value of the solution at a fixed time $t = T$ by

$$\varepsilon_h(w) = \Big(\int_D \Big| \mathbb{E}\left[u_{h,w}\right](T, \mathbf{x}) - \mathbb{E}\left[u_{ref}\right](T, \mathbf{x}) \Big|^2 \, d\mathbf{x} \Big)^{1/2}.$$

Here, the reference solution u_{ref} is computed with a high level w_{ref} for a fixed $\Delta x = h$.

5.1.1 An irregular solution

We first put the center of the initial pulse at $(x_c, z_c) = (0, -1)$ and let $\sigma_x = \sigma_z = 0.2$. The initial solution is then in both layers and does not vanishes on the interface. In this case, since the smooth initial solution does not satisfy the interface jump conditions (8), the solution is not highly regular in Y. In fact, we only have $u_Y \in L^\infty(\Gamma; C^0(0, T; L^2(D)))$, and therefore, the solution has only one bounded Y-derivative and no bounded mixed derivatives in Y. Fig. 3 shows the initial solution and the expected value and standard deviation of the solution

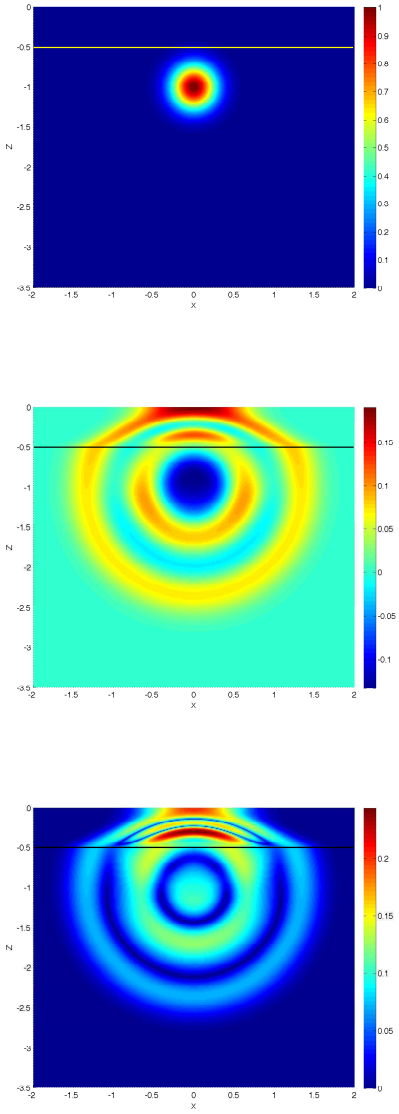


Figure 3: Test 1. The initial solution (top), the expected value of the solution (middle) and the standard deviation of the solution (bottom) at $t = 1$.

at time $t = 1$, computed with level $w = 5$ and $\Delta x = \Delta z = 0.0125$. Fig. 4 shows the L^2-norm of error in the expected value of the solution at $T = 1$ versus the number of collocation points $\eta(w)$. We observe a slow convergence of order $\mathcal{O}(\eta^{-\delta})$ with $0 < \delta < 1$, as expected due to low Y-regularity of the solution. We also note that for large values of $h\,\eta$, we observe exponential decay in the error, and as h decreases, more collocation points are needed to maintain a fixed

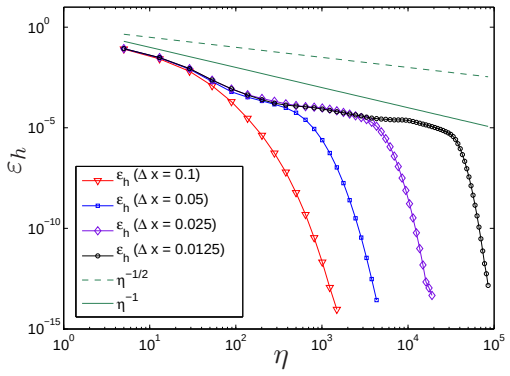


Figure 4: Test 1. The L^2-norm of error in the expected value of the solution, $\varepsilon_h(w)$, at time $T = 1$ versus the number of collocation points $\eta(w)$. The smooth initial wave pulse is in both layers and does not vanish on the interface. The solution has only one bounded Y-derivative and no mixed derivatives in Y.

accuracy (as predicted in Sect. 4.2).

5.1.2 A regular solution

We next put the center of the initial pulse at $(x_c, z_c) = (0, -1.5)$ and let $\sigma_x = \sigma_z = 0.11$. The initial solution is then essentially contained only in the bottom layer. In this case, since the smooth initial solution is zero at the interface, the interface conditions (8) are automatically satisfied. The solution remains smooth within each layer and satisfies the interface conditions. The solution is therefore highly regular in Y, see Sect. 2. Fig. 5 shows the L^2-norm of error in the expected value of the solution at $T = 1$ versus the number of collocation points $\eta(w)$. We observe a fast exponential rate of convergence in the error due to high regularity of the solution in Y.

5.2 Numerical test 2

In this test, we consider the same problem as the previous test in Sec. 5.1.1, except that we choose $Y_n \sim \mathcal{U}(-0.2, 0.5)$ so that the coercivity assumption (3) does not hold and we have $a^2_{min} = 0$. Fig. 6 shows the L^2-norm of error in the expected value of the solution at $T = 1$ versus the level w. For the sake of comparison, we also plot the error for the coercive wave speed in the numerical test 1.

In Tab. 1 we give the values of the spatial grid-lenghts $\Delta x = h$ and the level w at the *knee* point where the transition from slow to fast error convergence occurs. The values are given in both non-coercive and coercive cases, where $a^2_{min} \geq 0$ and $a^2_{min} > 0$, respectively. In the coercive case, when $h = 0.05$, the fast convergence starts at $h\, w = 0.05 \times 10 = 0.5$, and when $h = 0.025$, the fast

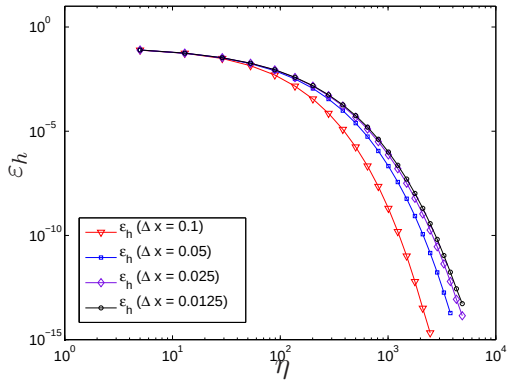


Figure 5: Test 1. The L^2-norm of error in the expected value of the solution, $\varepsilon_h(w)$, at time $T = 1$ versus the number of collocation points $\eta(w)$. The smooth initial wave pulse is contained only in one layer, and the solution remains smooth within that layer and has high Y-regularity.

Table 1: The values of h and w where the *knee* (transition from slow to fast convergence) occurs in both non-coercive and coercive cases.

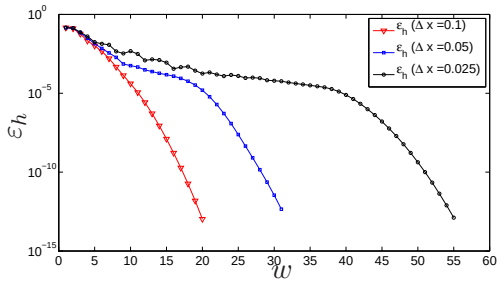
	w	
h	$a_{min}^2 \geq 0$	$a_{min}^2 > 0$
0.05	20	10
0.025	40	20

convergence starts at $h\,w = 0.025 \times 20 = 0.5$. In the non-coercive case, to obtain the same threshold 0.5, when $h = 0.05$ we need $h^\alpha \times w = 0.05^\alpha \times 20 = 0.5$, which gives $\alpha \approx 1.23$, and when $h = 0.025$ we need $h^\alpha \times w = 0.025^\alpha \times 40 = 0.5$, which gives $\alpha \approx 1.19$. This suggests that $\sigma^* \approx \mathcal{O}(h^{1.2})$ and shows that the estimates (48) and (49) with $\sigma^* = \mathcal{O}(h^2)$, derived in Sect. 4.2.2, may not be sharp.

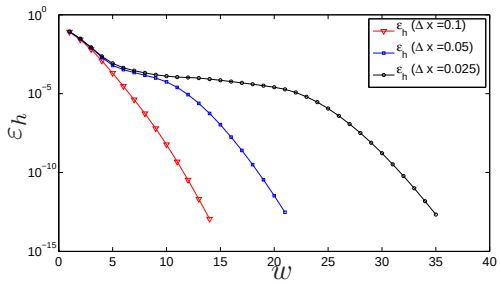
5.3 Numerical test 3

In the third test, we choose a computational domain $D = [-1.5, 1.5] \times [-3, 0]$ with two layers with equal widths $d_1 = d_2 = 1.5$. Let $a_0 = 0$, $\alpha_1 = 2$, $\alpha_2 = 3$ and $Y_n \sim \mathcal{U}(0.1, 0.5)$, $n = 1, 2$ in (5). We consider zero initial data $g_1 = g_2 = 0$ and a time-independent forcing term on $C = [-0.3, 0.3] \times [-1.8, -1.2]$ contained in D,

$$f(t, x, z) = \begin{cases} -10 \cos x \, \sin z, & \mathbf{x} \in C, \\ 0, & otherwise. \end{cases}$$



(a) $a^2_{min} = 0$



(b) $a^2_{min} > 0$

Figure 6: Test 2. The L^2-norm of error in the expected value of the solution, $\varepsilon_h(w)$, at time $T = 1$ versus the level w for non-coercive (top) and coercive (bottom) wave speeds. The smooth initial wave pulse is in both layers and does not vanish on the interface. The solution has only one bounded Y-derivative and no mixed derivatives in Y.

We note that since $f \in C^0(0,T;L^2(D))$, we will have $u_Y \in L^\infty(\Gamma; C^0(0,T;L^2(D)))$, and therefore, the solution has only one bounded Y-derivative and no bounded mixed derivatives in Y. Fig. 7 shows the convergence of the L^2-norm of error in the expected value of the solution $\varepsilon_h(w)$ at $T = 1$ versus the number of collocation points $\eta(w)$. We observe a slow convergence of order $\mathcal{O}(\eta^{-\delta})$ with $0 < \delta < 1$, as expected.

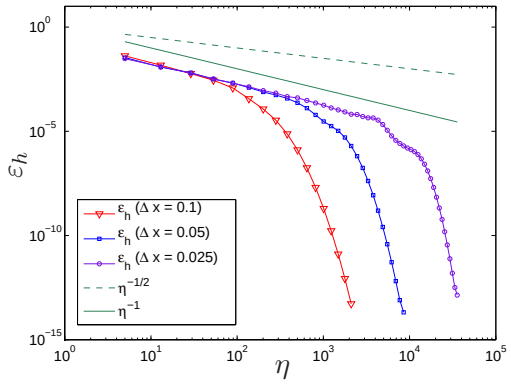


Figure 7: Test 3. The L^2-norm of error $\varepsilon_h(w)$ in the expected value of the solution at time $T = 1$ versus the number of collocation points $\eta(w)$. Due to a discontinuous force term, the solution has only one bounded Y-derivative and no mixed derivatives in Y.

Next we consider a quantity of interest of form (13) with $T = 1$, $\psi = 0$ and a smooth mollifier

$$\phi(x,z) = \begin{cases} 10\, e^{\frac{0.5}{x^2-0.5^2}+\frac{0.5}{(z+2.25)^2-0.5^2}}, & \mathbf{x} \in D_\phi \setminus \partial D_\phi, \\ 0, & otherwise, \end{cases}$$

with the support $D_\phi = [-0.5, 0.5] \times [-2.75, -1.75]$ contained in the bottom layer. Fig. 8 shows the error in the expected value of the quantity of interest, computed by

$$\varepsilon_{\mathcal{Q},h}(w) = \Big| \mathbb{E}\left[\mathcal{Q}[u_{h,w}]\right] - \mathbb{E}\left[\mathcal{Q}[u_{ref}]\right] \Big|.$$

We note that since the smooth mollifiers $\psi = 0$ and $\phi \in C_0^\infty(D)$ do no cross the interface, the quantity of interest (13) has high Y-regularity. We therefore expect a convergence rate faster than any polynomial rate. However, for the small values of w tested here, we observe an algebraic rate of order about $\mathcal{O}(\eta^{-3})$.

5.4 Numerical test 4

In this test, we study the Arias intensity (51) due to a Ricker wavelet. Arias Intensity is an important quantity of interest in seismology which describes earthquake shaking that triggers landslides. It determines the intensity of shaking

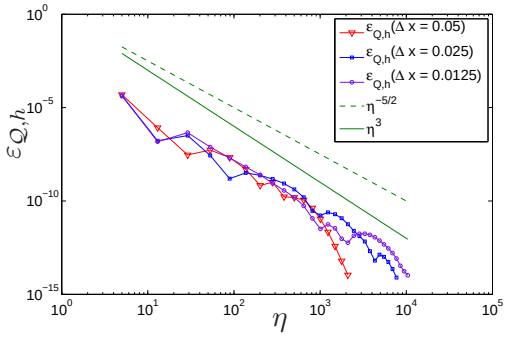


Figure 8: Test 3. Error in the expected value of the quantity of interest $\mathcal{Q}$ with $T = 1$ and smooth mollifiers compactly supported and contained in the bottom layer. Due to high Y-regularity of $\mathcal{Q}$, we expect a fast error convergence.

by measuring the acceleration of transient seismic waves. The Ricker wavelet, which is the negative normalized second derivative of a Gaussian function, is used to model the generation of seismic waves.

We choose a computational domain $D = [-10, 10] \times [-10, 0]$ with two layers with widths $d_1 = 1$ and $d_2 = 9$. Let $a_0 = 0$, $\alpha_1 = 2$, $\alpha_2 = 3$ and $Y_n \sim \mathcal{U}(0.1, 0.5)$, $n = 1, 2$ in (5). We consider zero initial data $g_1 = g_2 = 0$ and a forcing term consisting of a Ricker wavelet on a small region $R_c = [-0.1, 0.1] \times [-1.2, -1.1]$,

$$f(t, \mathbf{x}) = \psi(t)\, \mathcal{X}_{R_c}(\mathbf{x}), \quad \psi(t) = 100\,(1-\lambda\,(t-t_0)^2)\, e^{-0.5\,\lambda\,(t-t_0)^2}, \quad \lambda = 20, \quad t_0 = 0.1.$$

We compute the Arias intensity on a part of the free surface $S = \{(x, z) \mid x \in [0, 1],\ z = 0\}$. Fig. 9 shows the mean plus minus the standard deviation of the Arias intensity on S as a function of time, computed with the level $w = 15$ and the spatial grid-lenght $\Delta x = \Delta z = 0.0125$. Fig. 10 shows the response

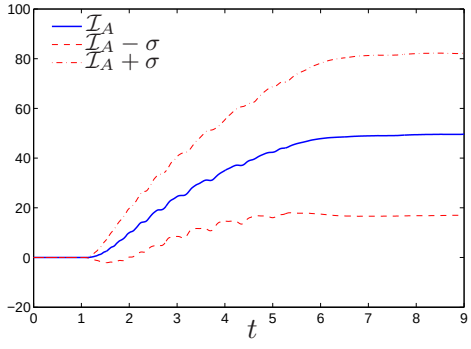


Figure 9: Test 4. Mean (solid line), plus and minus the standard deviation (dashed line) of the Arias intensity due to a Ricker wavelet on a small region in the bottom layer.

surface of the Arias intensity on S at the final time T=4 computed using sparse interpolation. We note that due to the nonlinearity of the Arias intensity in

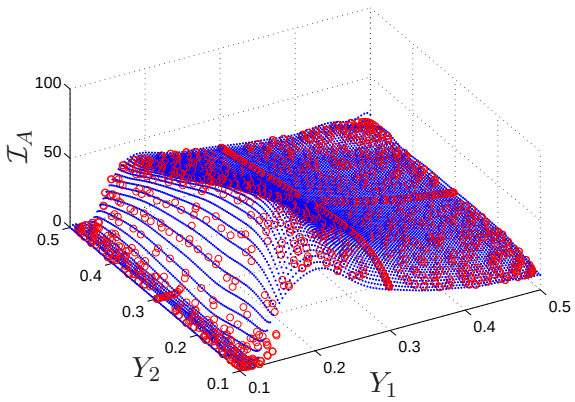


Figure 10: Test 4. Response surface of the Arias intensity with T=4 as a function of two random variables obtained by sparse interpolation. The red circles are the realizations of the sparse grid points, and the blue dots are interpolated values.

u_{tt}, we do not expect high Y-regularity. See Remark 3 in Sect. 2.3. This is also observable from the response surface of the Arias intensity in Fig. 10. Fig. 11 shows the error $\varepsilon_{\mathcal{I}_A,h}$ in the expected value of the Arias intensity at final time $T = 4$. We observe a slow rate of convergence $\mathcal{O}(\eta^{-\delta})$ with $0 < \delta < \frac{1}{2}$ as expected.

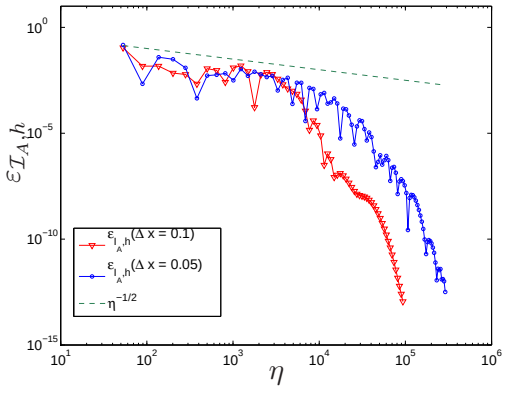


Figure 11: Test 4. Error in the expected value of the Arias intensity $\mathcal{I}_A$ with $T = 4$. The slow rate of convergence shows that the Arias intensity is not Y-regular.

6 Conclusion

We have proposed a stochastic collocation method for solving the second order wave equation in a heterogeneous random medium with a piecewise smooth random wave speed. The medium consists of non-overlapping sub-domains. In each sud-domain, the wave speed is smooth and is given in terms of one random variable. We assume that the interfaces of speed discontinuity are smooth. One important example is wave propagation in multi-layered media with smooth interfaces. We have derived a priori error estimates with respect to the number of collocation points for the stochastic collocation method based on full and sparse tensor interpolations.

The main result is that unlike in elliptic and parabolic problems, the solution to hyperbolic problems is not in general analytic with respect to the random variables. Therefore, the convergence rate of error in the wave solution is only algebraic. A fast spectral convergence is still possible for some linear quantities of interest with smooth mollifiers and for the wave solution with smooth data compactly supported within sub-domains. We also show that the semi-discrete solution is analytic with respect to the random variables with the radius of analyticity proportional to the mesh size h. We therefore obtain an exponential rate of convergence which deteriorates as the quantity $h\,p$ gets smaller, with p representing the polynomial degree in the stochastic space. We have shown that analytical results and numerical examples are consistent and that the stochastic collocation method may be a valid alternative to the more traditional Monte Carlo method.

Future directions will include the analysis of other types of second order hyperbolic problems such as elastic wave equation and the case where the position of discontinuity interfaces is also stochastic.

Appendix

Lemma A1. Consider the 1D Cauchy problem for the scalar wave equation in the conservative form

$$u_{tt} - \partial_x\big(c(x)\,\partial_x u\big) = f(t,x), \qquad (t,x) \in (0,\infty)\times\mathbb{R}, \tag{52}$$

and in the non-conservative form

$$u_{tt} - c(x)\,\partial_{xx} u = f(t,x), \qquad (t,x) \in (0,\infty)\times\mathbb{R}, \tag{53}$$

subjected to the initial conditions

$$u(0,x) = g(x), \qquad u_t(0,x) = h(x).$$

Suppose that

- $c(x)$ is positive bounded away from zero and smooth everywhere except at $x = 0$ where it has a discontinuity,
- $g(x)$ and $h(x)$ are smooth, compactly supported functions and $0 \notin \operatorname{supp} g \cup \operatorname{supp} h$,
- $\partial_t^k f \in L^2(\mathbb{R})$ for each fixed t, and $\partial_t^k f = 0$ at $t = 0$ for all $k \geq 0$.

Then, for each fixed t, for solutions u to any of the two wave equations (52) and (53),

$$\partial_t^k u_t \in L^2(\mathbb{R}), \qquad \partial_t^k u_x \in L^2(\mathbb{R}), \qquad \forall k \geq 0.$$

Proof. Let $v := \partial_t^k u$. Then, for the conservative form, v solves the Cauchy problem

$$v_{tt} - \partial_x\big(c(x)\,\partial_x v\big) = \partial_t^k f, \qquad (t, x) \in (0, \infty) \times \mathbb{R}, \tag{54}$$

with the initial conditions

$$v(0, x) = \partial_t^k u(0, x) = \begin{cases} (\partial_x c\,\partial_x)^{k/2} g, & k \text{ even}, \\ (\partial_x c\,\partial_x)^{(k-1)/2} h, & k \text{ odd}, \end{cases}$$

and

$$v_t(0, x) = \partial_t^{k+1} u(0, x) = \begin{cases} (\partial_x c\,\partial_x)^{k/2} h, & k \text{ even}, \\ (\partial_x c\,\partial_x)^{(k+1)/2} g, & k \text{ odd}. \end{cases}$$

For the non-conservative form, v solves the Cauchy problem

$$v_{tt} - c(x)\,\partial_{xx} v = \partial_t^k f, \qquad (t, x) \in (0, \infty) \times \mathbb{R}, \tag{55}$$

with the initial conditions

$$v(0, x) = \partial_t^k u(0, x) = \begin{cases} (c\,\partial_{xx})^{k/2} g, & k \text{ even}, \\ (c\,\partial_{xx})^{(k-1)/2} h, & k \text{ odd}, \end{cases}$$

and

$$v_t(0, x) = \partial_t^{k+1} u(0, x) = \begin{cases} (c\,\partial_{xx})^{k/2} h, & k \text{ even}, \\ (c\,\partial_{xx})^{(k+1)/2} g, & k \text{ odd}. \end{cases}$$

Since the functions g and h are smooth and their support does not include the discontinuity point of $c(x)$, the initial data for v in both problems are smooth for all k. It is well known that for the wave equations (54) and (55) with smooth initial data and L^2 forcing term [9],

$$v_t,\, v_x \in L^2(\mathbb{R}).$$

This completes the proof.

Theorem A2. Consider the 1D Cauchy problem for the scalar wave equation in the conservative form

$$u_{tt} - \partial_x\big(c(x,y)\,\partial_x u\big) = f(x), \qquad (t,x,y) \in (0,\infty)\times\mathbb{R}\times\mathbb{R}, \tag{56}$$

subjected to the initial conditions

$$u(0,x,y) = g(x), \qquad u_t(0,x,y) = h(x). \tag{57}$$

Let $z_{k,l} := \partial_y^k \partial_t^l c u_x$, and assume that the assumptions of Lemma A1 hold. If $\partial_y^k c \in L^\infty(\mathbb{R}), \forall k \geq 0$, then for each fixed t and y,

$$\partial_t z_{k,l} \in L^2(\mathbb{R}), \qquad \partial_{xx} z_{k,l} \in L^2(\mathbb{R}), \qquad \forall k,l \geq 0. \tag{58}$$

Proof. We show the result by induction on k.
Case $k = 0$**.** We have $\partial_t z_{0,l} = c\,\partial_t^{l+1} u_x$ which belongs to $L^2(\mathbb{R})$ by Lemma A2 for all $l \geq 0$. Moreover, differentiating (56) l times with respect to t and once with respect to x and multiplying by c, we obtain

$$\partial_{tt} z_{0,l} = c\,\partial_{xx} z_{0,l}. \tag{59}$$

Therefore, $\partial_{xx} z_{0,l} \in L^2(\mathbb{R})$ for all $l \geq 0$, because $\partial_t z_{0,l+1} \in L^2(\mathbb{R})$.
General case. We assume that (58) holds with $k < K$. Differentiating (59) K times with respect to y gives us

$$\partial_{tt} z_{K,l} - c\,\partial_{xx} z_{K,l} = \sum_{k=0}^{K-1} \binom{K}{k} \partial_y^{K-k} c\ \partial_{xx} z_{k,l}.$$

Since the right hand side belongs to $L^2(\mathbb{R})$ by the induction hypothesis, Lemma A1 tells us that $\partial_t z_{K,l} \in L^2(\mathbb{R})$ for all $l \geq 0$. Moreover,

$$\partial_{xx} z_{K,l} = \frac{1}{c}\left(\partial_t z_{K,l+1} - \sum_{k=0}^{K-1} \binom{K}{k} \partial_y^{K-k} c\ \partial_{xx} z_{k,l}\right),$$

where the right hand side is in $L^2(\mathbb{R})$. This completes the proof.

Acknowledgements

The authors would like to thank Olof Runborg and Georgios Zouraris for stimulating discussions and Lorenzo Tamellini for providing parts of source codes for implementing the stochastic collocation algorithm. The first author would also like to thank the MOX center in Politecnico di Milano for a two month visit funded by the Italian grant FIRB-IDEAS (Project n. RBID08223Z) "Advanced numerical techniques for uncertainty quantification in engineering and life science problems".

References

[1] I. M. Babuska and F. Nobile and R. Tempone, (2007) *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal. 45, 1005–1034.

[2] I. Babuska and R. Tempone and G. E. Zouraris, (2004) *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal. 42, 800–825.

[3] I. Babuska and R. Tempone and G. E. Zouraris (2005), *Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation*, Computer Methods in Applied Mechanics and Engineering 194(1216), 1251–1294.

[4] V. Barthelmann and E. Novak and K. Ritter, (2000) *High dimensional polynomial interpolation on sparse grids*, Adv. Comput. Math. 12, 273–288.

[5] J. Beck and R. Tempone and F. Nobile and L. Tamellini, (2011) *On the optimal approximation of stochastic PDEs by Galerkin and collocation methods*, To appear in Math. Model. Meth. Appl. Sci.

[6] C. Canuto and T. Kozubek, (2007) *A fictitious domain approach to the numerical solution of PDEs in stochastic domains*, Mumer. Math. 107 (2), 257–293.

[7] C. Canuto and M. Y. Hussaini and A. Quarteroni and T. A. Zang, (2006) *Spectral methods: fundamentals in single domains*, Springer-Verlag, Berlin.

[8] P. Erdös and P. Turán, (1937) *On interpolation I. Quadrature- and mean-convergence in the Lagrange-interpolation*, Ann. of Math. (2) 38, 142–155.

[9] L. C. Evans, (1998) *Partial differential equations*, Graduate Studies in Mathematics, Vol. 19, AMS.

[10] G. S. Fishman, (1996) *Monte Carlo: Concepts, Algorithms, and Applications*, Springer Ser. Oper. Res., Springer-Verlag, New York.

[11] R. G. Ghanem and P. D. Spanos (1991), *Stochastic finite elements: A spectral approach*, Springer, New York.

[12] D. Gottlieb and D. Xiu, (2008) *Galerkin method for wave equations with uncertain coefficients*, Commun. Comput. Phys. 3(2), 505–518.

[13] H. Harbrecht and R. Schneider and C. Schwab, (2008) *Sparse second moment analysis for elliptic problems in stochastic domains*, Numer. Math. 109, 385–414.

[14] L. Hörmander, (1994) *The analysis of linear partial differential operators III, pseudo-differential operators*, Classics in Mathematics, Springer, Berlin.

[15] F. John, (1982) *Partial Differential Equations*, Springer, New York.

[16] H.-O. Kreiss and J. Lorenz, (2004) *Initial-boundary value problems and the Navier-Stokes equations*, Classics in Applied Mathematics 47, SIAM, Philadelphia.

[17] H.-O. Kreiss and O. E. Ortiz, (2002) *Some mathematical and numerical questions connected with first and second order time-dependent systems of partial differential equations*, Lecture Notes in Physics, Vol. 604, 359–370.

[18] H.-O. Kreiss and N. A. Petersson and J. Yström, (2002) *Difference approximations for the second order wave equation*, SIAM J. Numer. Anal. 40, 1940–1967.

[19] O. P. Le Maitre and O. M. Knio and H. N. Najm and R. G. Ghanem, (2004) *Uncertainty propagation using Wiener-Haar expansions*, J. Comput. Phys. 197(1), 28–57.

[20] O. P. Le Maitre and H. N. Najm and R. G. Ghanem and O. M. Knio, (2004) *Multi-resolution analysis of Wiener-type uncertainty propagation schemes*, J. Comput. Phys. 197(2), 502–531.

[21] C. D. Levermore and M. Oliver, (1997) *Analyticity of solutions for a generalized Euler equation*, J. Differential Equations 133, 321–339.

[22] G. Lin and C.-H. Su and G. E. Karniadakis, (2006) *Predicting shock dynamics in the presence of uncertainties*, J. Comput. Phys. 217(1), 260–276.

[23] G. Lin and C.-H. Su and G. E. Karniadakis, (2006) *Stochastic modeling of random roughness in shock scattering problems: theory and simulations*, Comput. Methods Appl. Mech. Engrg. 197(43-44), 3420–3434.

[24] M. Loéve, (1977) *Probability theory I*, Grad. Texts in Math. 45, Springer-Verlag, New York.

[25] M. Loéve, (1978) *Probability theory II*, Grad. Texts in Math. 46, Springer-Verlag, New York.

[26] H. G. Matthies and A. Keese (2005), *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Computer Methods in Applied Mechanics and Engineering 194(1216), 1295–1331.

[27] F. Nobile and R. Tempone, (2009) *Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients*, Int. J. Numer. Meth. Engng 80, 979–1006.

[28] F. Nobile and R. Tempone and C. G. Webster, (2008) *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal. 46, 2411–2442.

[29] F. Nobile and R. Tempone and C. G. Webster, (2008) *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal. 46, 2309–2345.

[30] H. Owhadi and L. Zhang, (2008) *Numerical homogenization of the acoustic wave equations with a continuum of scales*, Comput. Methods Appl. Mech. Engrg. 198, 397–406.

[31] G. Poette and B. Després and D. Lucor, (2009) *Uncertainty quantification for systems of conservation laws*, J. Comput. Phys. 228(7), 2443–2467.

[32] L. Rodino, (1993) *Linear partial differential operators in Gevrey spaces*, World Scientific Publishin Co., Singapore.

[33] C. C. Stolk, (2000) *On the modeling and inversion of seismic data*, PhD Thesis, Utrecht University.

[34] T. Tang and T. Zhou, (2010) *Convergence analysis for stochastic collocation methods to scalar hyperbolic equations with a random wave speed*, Commun. Comput. Phys. 8(1), 226–248.

[35] R. A. Todor and C. Schwab (2007), *Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients*, IMA Journal of Numerical Analysis 27(2), 232–261.

[36] J. Tryoen and O. Le Maitre and M. Ndjinga and A. Ern, (2010) *Intrusive projection methods with upwinding for uncertain nonlinear hyperbolic systems*, J. Comput. Phys. 229, 6485–6511.

[37] J. Tryoen and O. Le Maitre and M. Ndjinga and A. Ern, (2010) *Roe solver with entropy corrector for uncertain hyperbolic systems*, J. Comput. Appl. Math. 235(2), 491–506.

[38] X. Wang and G. E. Karniadakis, (2006) *Long-term behavior of ploynomial chaos in stochastic flow simulations*, Comput. Methods Appl. Mech. Engrg 195, 5582–5596.

[39] N. Wiener, (1938) *The homogeneous chaos*, Amer. J. Math. 60, 897–936.

[40] D. Xiu andJ. S. Hesthaven, (2005) *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput. 27(3), 1118–1139.

[41] D. Xiu and G. E. Karniadakis (2002), *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Computer Methods in Applied Mechanics and Engineering 191(43), 4927–4948.

[42] D. Xiu and G. E. Karniadakis, (2002) *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput. 24, 619–644.

[43] D. Xiu and D. M. Tartakovsky, (2006) *Numerical methods for differential equations in random domains*, SIAM J. Sci. Comput. 28(3), 1167–1185.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

36/2011 Motamed, M.; Nobile, F.; Tempone, R.
A stochastic collocation method for the second order wave equation with a discontinuous random speed

35/2011 Iapichino, L.; Quarteroni, A.; Rozza, G.
A Reduced Basis Hybrid Method for the coupling of parametrized domains represented by fluidic networks

34/2011 Benacchio, T.; Bonaventura, L.
A spectral collocation method for the one dimensional shallow water equations on semi-infinite domains

33/2011 Antonietti, P.F.; Beirao da Veiga, L.; Lovadina, C.; Verani, M.
Hierarchical a posteriori error estimators for the mimetic discretization of elliptic problems

32/2011 Aletti, G; Ghiglietti, A; Paganoni, A.
A modified randomly reinforced urn design

31/2011 Astorino, M.; Becerra Sagredo, J.; Quarteroni, A.
A modular lattice Boltzmann solver for GPU computing processors

30/2011 Nobile, F.; Pozzoli, M.; Vergara, C.
Time accurate partitioned algorithms for the solution of fluid-structure interaction problems in haemodynamics

29/2011 Morin, P.; Nochetto, R.H.; Pauletti, S.; Verani, M.
AFEM for Shape Optimization

28/2011 Pischiutta, M.; Formaggia, L.; Nobile, F.
Mathematical modelling for the evolution of aeolian dunes formed by a mixture of sands: entrainment-deposition formulation

27/2011 Antonietti, P.F.; Bigoni, N.; Verani, M.
A Mimetic Discretization of Elliptic Control Problems