HAL
open science

Uncertainty Quantification for Stochastic Approximation Limits Using Chaos Expansion

Stéphane Crépey, Gersende Fort, Emmanuel Gobet, Uladzislau Stazhynski

To cite this version:

Stéphane Crépey, Gersende Fort, Emmanuel Gobet, Uladzislau Stazhynski. Uncertainty Quantification for Stochastic Approximation Limits Using Chaos Expansion. SIAM/ASA Journal on Uncertainty Quantification, In press. hal-01629952v3

HAL Id: hal-01629952

https://hal.science/hal-01629952v3

Submitted on 31 Jan 2019 (v3), last revised 28 May 2020 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Uncertainty Quantification for Stochastic Approximation Limits Using Chaos Expansion*

S. Crépey †, G. Fort ‡, E. Gobet§, and U. Stazhynski ¶

Abstract. The uncertainty quantification for the limit of a Stochastic Approximation (SA) algorithm is analyzed. In our setup, this limit $\zeta^\star$ is defined as a zero of an intractable function and is modeled as uncertain through a parameter θ. We aim at deriving the function $\zeta^\star$, as well as the probabilistic distribution of $\zeta^\star(\theta)$ given a probability distribution π for θ. We introduce the so-called Uncertainty Quantification for SA (UQSA) algorithm, an SA algorithm in increasing dimension for computing the basis coefficients of a chaos expansion of $\theta \mapsto \zeta^\star(\theta)$ on an orthogonal basis of a suitable Hilbert space. UQSA, run with a finite number of iterations K, returns a finite set of coefficients, providing an approximation $\widehat{\zeta^\star_K}(\cdot)$ of $\zeta^\star(\cdot)$. We establish the almost-sure and L^p-convergences in the Hilbert space of the sequence of functions $\widehat{\zeta^\star_K}(\cdot)$ when the number of iterations K tends to infinity. This is done under mild, tractable conditions, uncovered by the existing literature for convergence analysis of infinite dimensional SA algorithms. For a suitable choice of the Hilbert basis, the algorithm also provides an approximation of the expectation, of the variance-covariance matrix and of higher order moments of the quantity $\widehat{\zeta^\star_K}(\theta)$ when θ is random with distribution π. UQSA is illustrated and the role of its design parameters is discussed numerically.

Key words. Stochastic Approximation in Hilbert space, Uncertainty Quantification, Chaos expansion, Almost-sure convergence.

AMS subject classifications. 62L20, 41A10, 41A25, 60B12, 60F15, 90C15.

1. Introduction.

Since the seminal work of Robbins and Monro [26], the Stochastic Approximation (SA) method has become mainstream for various applications—see e.g. [19, 5] for a general introduction and examples of applications. SA is used to find a zero of an intractable function $h : \mathbb{R}^q \to \mathbb{R}^q$ of the form $z \mapsto \mathbb{E}[H(z, V)]$, where V is some random variable (r.v.) and $H(z, v)$ can be computed explicitly (as opposed to its expectation h).

Example 1. *A typical example comes from portfolio optimization (for related stochastic programming problems, see [29]): consider a manager who has to devise an investment portfolio made of $q+1$ stocks whose values at time t are denoted by $S_t = (S^i_t : 1 \le i \le q+1)$; we can assume without loss of generality that all the initial conditions are deterministic and the*

*This research benefited from the support of the Chair *Financial Risks* of the *Risk Foundation*, the Chair *Markets in Transition* of the *Fédération Bancaire Française*, the *Finance for Energy Market Research Centre*, the ANR project *CAESARS* (ANR-15-CE05-0024) and the ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02.

†Email: stephane.crepey@univ-evry.fr. LaMME, Univ. Evry, CNRS, Université Paris-Saclay, 91037, Evry, France.

‡Email: Gersende.Fort@math.univ-toulouse.fr. IMT, Université de Toulouse & CNRS, F-31062 Toulouse, France.

§Email: emmanuel.gobet@polytechnique.edu. Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique and CNRS, Université Paris-Saclay, Route de Saclay, 91128 Palaiseau Cedex, France. Corresponding author.

¶Email: uladzislau.stazhynski@polytechnique.edu. Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique and CNRS, Université Paris-Saclay, Route de Saclay, 91128 Palaiseau Cedex, France.

normalization $S_0^{q+1} = 1$ holds. Portfolio weights (real numbers) $w := (w_1, \dots, w_{q+1})$ in each stock have to be determined, such that some risk criterion (encoded by a convex function Λ) is minimized. Namely, if the investment horizon is $T > 0$, minimizing the portfolio risk writes

$$\inf_{w \in \mathbb{R}^{q+1}} \mathbb{E}\left[\Lambda\left(\sum_{i=1}^{q+1} w_i S_T^i\right)\right],$$

under a budget constraint

$$W_0 - \sum_{i=1}^{q+1} w_i S_0^i = 0$$

pertaining to some initial wealth W_0. The above problem is equivalent to

$$\inf_{z=(w_1,\dots,w_q)\in\mathbb{R}^q} \mathbb{E}\left[L(z, S_T)\right], \quad \text{where } L(z, S_T) := \Lambda\left(\sum_{i=1}^{q} w_i S_T^i + (W_0 - \sum_{i=1}^{q} w_i S_0^i) S_T^{q+1}\right). \tag{1.1}$$

This is a convex problem. However, except in simple cases such as a quadratic loss function Λ, minimizers $z^\star$ are not explicit when they exist. Actually, under smooth conditions on the function Λ, the optimal weights are the zeros of $z \mapsto \mathbb{E}[H(z, S_T)]$, where $H(z, S) = \nabla_z L(z, S)$. The optimal weights can be computed using the SA method [29, Section 5.9].

The reader acquainted with financial applications will notice that the above problem can be rephrased in terms of expected utility maximization, interpreting the function $(-\Lambda)$ as a (concave) utility function.

Uncertainty Quantification (UQ) applied to SA limits corresponds to the situation where the distribution of the r.v. V or/and the function H is not known exactly and the goal is to quantify in some sense how this uncertainty acts on the roots of h. In this paper, the UQ is studied in the form where *(i)* the distribution of V depends on an unknown parameter θ in Θ, i.e. $V \sim \mu(\theta, \mathrm{d}v)$; *(ii)* the function H may depend on the parameter θ; *(iii)* the value of θ is unknown, but some a priori knowledge given by a probability distribution π on Θ is available. Specifically, our mathematical framework is the following: for any $(z, \theta) \in \mathbb{R}^q \times \Theta$, set

$$h(z, \theta) := \int_{\mathcal{V}} H(z, v, \theta)\mu(\theta, \mathrm{d}v). \tag{1.2}$$

Define

$$\mathcal{S} := \{\zeta : \Theta \to \mathbb{R}^q \text{ measurable and s.t. for } \pi\text{-almost all } \theta,\ h(\zeta(\theta), \theta) = 0\}. \tag{1.3}$$

The full UQ problem consists in (i) computing some $\zeta^\star$ in $\mathcal{S}$ and (ii) characterizing the distribution of $\zeta^\star(\theta)$ when $\theta \sim \pi$. In this paper, (i) we propose an algorithm designed to produce $\zeta^\star \in \mathcal{S}$, and we partially address (ii) by providing quantities related to the moments of $\zeta^\star(\theta)$ when $\theta \sim \pi$.

Example 2. *In example 1, the uncertainty could bear on the volatility/correlation of the stocks S, or to the uncertainty on some parameters of the loss function Λ (which models the risk aversion). UQ has the purpose of quantifying how much the model uncertainty on S and Λ will affect the optimal allocation $w^\star$.*

Another well-known SA application is for the determination of a quantile $\zeta^\star$ of order α , for some $\alpha \in (0,1)$, of a distribution μ on the real line with continuous cumulative distribution function. This can be done by a SA procedure (see e.g. [10, Chapter 1]) with

$$H(z,v) := \alpha - \mathbf{1}_{v \le z} \qquad h(z) := \alpha - \int_{v \le z} \mu(\mathrm{d}v).$$

The quantile is not necessarily unique and a convergence analysis of the related SA algorithm can be found e.g. in [3]. In some situations, the distribution μ may depend on a parameter θ (see e.g. [4]) so that the challenge is to determine a function $\theta \mapsto \zeta^\star(\theta)$ satisfying $\alpha = \int_{v \le \zeta^\star(\theta)} \mu(\theta, \mathrm{d}v)$ for almost-all θ. In the setup of quantile regression, our contribution allows considering the class of functions $\zeta^\star$ such that $\zeta^\star(\theta)$ is a square-integrable r.v. when $\theta \sim \pi$, π being some a priori knowledge. This class is more general than what is usually addressed in the literature (see e.g. [17] for the linear quantile regression).

1.1. Review of the Literature. In UQ applications (see [20, 30]), the quantity of interest $\zeta^\star$ is typically related to the solution of a PDE. It is given as the output of a blackbox (for instance, a numerical solver), which requires heavy computations. In our case, $\zeta^\star$ is defined as the root of a non-explicit function defined as an expectation, so that it can be obtained as the output of an SA algorithm. In both cases, the quantity of interest may depend on a parameter θ (uncertainty on the inputs of the blackbox, uncertainty on the expectation). The question is then to characterize $\theta \mapsto \zeta^\star(\theta)$ through UQ. In the last two decades, UQ has become a huge concern for both research and industrial applications; our goal is to study the UQ problem for the SA limits, which, to the best of our knowledge, has not been investigated so far.

A crude Monte Carlo (MC) method for the above problem could consist in sampling M values $\{\theta_i, 1 \le i \le M\}$ under π, and then compute, for each sample θ_i, an approximation $\widehat{\zeta^\star(\theta_i)}$ of $\zeta^\star(\theta_i)$ where $\zeta^\star \in \mathcal{S}$. The distribution of the r.v. $\{\zeta^\star(\theta), \theta \sim \pi\}$ is then approximated by the empirical distribution of $\{\widehat{\zeta^\star(\theta_i)} : 1 \le i \le M\}$. When $\zeta^\star$ solves a PDE, a global error analysis is performed in [2], accounting for both the sampling error and the PDE discretization error, which are decoupled in some way. In our SA setting, a naive approach would be to compute $\widehat{\zeta^\star(\theta_m)}$ as the output of a standard SA algorithm for fixed θ_m. The approach we propose below will be different since it will couple in an efficient way the outer Monte Carlo sampling of θ_m and the inner Monte Carlo sampling used to feed the SA algorithm. The resulting algorithm computes the function $\theta \mapsto \zeta^\star(\theta)$ as a whole by a single Monte Carlo sampling, which is bound to work well if the function $\zeta^\star$ is regular (as made precise by our convergence result below)—a regularity which is not exploited in the naive approach.

A second method, developed in [21, 16], is a perturbative approach taking advantage of a stochastic expansion of $\{\zeta^\star(\theta), \theta \sim \pi\}$ that is available when θ has small variations (a restriction that we do not need or want to impose in our case).

A third strategy, which dates back to Wiener [33] and has been developed in the fields of engineering and UQ in the 2000s (see [12, 20] and references therein), is based on chaos expansions. This technique, also known as the spectral method, consists of projecting a point $\zeta^\star \in \mathcal{S}$ on an orthonormal basis $\{\theta \mapsto B_i(\theta), i \geq 0\}$ of real-valued and square-integrable (with respect to π) functions, and computing the $\mathbb{R}^q$-valued coefficients $\{u_i, i \geq 0\}$ of $\zeta = \zeta^\star$ in its decomposition

$$\zeta = \sum_{i \geq 0} u_i B_i, \qquad u_i = \langle \zeta; B_i \rangle_\pi := \int \zeta \; B_i \; \mathrm{d}\pi. \tag{1.4}$$

In the most common case where $B_0 \equiv 1$, the expectation and the variance-covariance matrix are related to the $\mathbb{R}^q$-valued projection coefficients $\{u_i, i \geq 0\}$ through

$$\mathbb{E}_{\theta \sim \pi}[\zeta(\theta)] = u_0 \text{ and } \mathbb{V}\mathrm{ar}_{\theta \sim \pi}(\zeta(\theta)) = \sum_{i \geq 1} u_i u_i^\top. \tag{1.5}$$

This provides a tool to approximate the first and second moments of $\{\zeta^\star(\theta), \theta \sim \pi\}$ from an approximation $\{\widehat{u}_i^\star, i \geq 0\}$ of the coefficients $\{u_i^\star, i \geq 0\}$. In the case of polynomial basis (see [20, Appendix C]), higher order moments are also explicitly related to the coefficients. For an approximation of more general statistics (confidence intervals for instance), one can define an approximation $\widehat{\zeta^\star}$ of the function $\zeta^\star$ by using $\{\widehat{u}_i^\star, i \geq 0\}$, sample independent and identically distributed (i.i.d.) variables θ_j's under the distribution π, compute the $\widehat{\zeta^\star}(\theta_j)$'s and obtain the empirical estimators of the desired quantities. Since such post-processing is quite obvious, from now on we focus on the computations of a function $\zeta^\star$ in $\mathcal{S}$ (or equivalently of its coefficients $\{u_i^\star, i \geq 0\}$).

In [18], Kulkarni and Borkar provide a finite dimensional procedure to approximate the minimum of $\zeta \mapsto \int L(\zeta(v), v)\mu(\mathrm{d}v)$ on some subset of the real valued functions, for some explicit non-negative function L satisfying some strict convexity properties with respect to (w.r.t.) its first argument. They first introduce a truncation by considering functions of the form $\zeta = \sum_{i=1}^m x_i B_i$ for some adequate family of functions $\{B_1, \cdots, B_m\}$ such that computing ζ is equivalent to the computation of $x_1, \cdots, x_m$. The problem is then solved by a classical SA procedure in $\mathbb{R}^q$ where $H(z, v) = \partial_s L(s, v)|_{s=\sum_{i=1}^q z_i B_i(v)} \times \{B_k(v), 1 \leq k \leq q\}$. They analyze the error due to finite dimensional truncation. However, this analysis makes apparent that, in order to achieve actual convergence results (as opposed to error bounds in [18]), one really needs to let m increase during the algorithm. Moreover, even when the function ζ is explicitly known (which is not the case in our setup) and estimating individual coefficients u_i in (1.4) is straightforward by MC simulations, the global convergence of a method where more and more coefficients are computed by Monte Carlo is nontrivial, subject to a fine-tuning of the speeds at which the number of coefficients and the number of simulations go to infinity (see [13]).

Our methodological contribution is the derivation of an algorithm for the UQ analysis of SA limits. It is based on chaos expansion, it is able to overcome the finite dimensional truncation by increasing the dimension, and it uses Monte Carlo sampling only for feeding the SA algorithm.

Asymptotically, our method reaches the case with no truncation, so that it is among the family of infinite dimensional SA algorithms.

There exists a large number of works on infinite dimensional SA methods. Recently, numerous works have been devoted to statistical learning in Hilbert spaces, in particular, Reproducing Kernel Hilbert Spaces (RKHS, see e.g. [9] and references therein). However, statistical learning in Hilbert spaces reduces to finite dimensional SA: based on N input/output examples $\{(x_i, y_i), 1 \leq i \leq N\}$, it consists of solving

$$\text{argmin}_{\varphi \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} \left(L(\varphi(x_i), y_i) + \Omega(\|\varphi\|_{\mathcal{H}}) \right), \tag{1.6}$$

where $\mathcal{H}$ is a RKHS associated to a positive-definite real-valued kernel K, L is a non-negative loss function, and Ω is a strictly increasing function acting as a penalty term. By the Representer Theorem (see [28, Theorem 4.2]), each minimizer admits a representation of the form $\varphi^\star = \sum_{i=1}^{N} \omega_i^\star K(\cdot, x_i)$ so that, under regularity conditions on L and Ω, the solution of (1.6) can be solved by a SA algorithm in $\mathbb{R}^N$.

In [32], [6] and [35], the authors study SA in Hilbert spaces in the case $H(z, V) = \tilde{H}(z) + V$, where z here lives in a Hilbert space. The conditions of convergence are infinite dimensional formal analogues of those in the finite dimensional case (see Section 3.1). Unfortunately, although interesting from a theoretical point of view, these SA algorithms are defined directly in the infinite dimensional Hilbert space, so that they are not feasible in practice.

An alternative to the above is to iteratively increase the dimension, to infinity in the limit, while remaining finite at each iteration. There have been several papers in this direction, generally known as the sieve approach. Goldstein [14] proves almost-sure convergence in the norm topology for a modified Kiefer-Wolfowitz procedure (see [15]) in infinite dimensional Hilbert space using a sieve approach. Nixdorf [24] shows asymptotic normality for a modified sieve-type Robbins-Monro procedure. Yin [34] proves almost-sure convergence in the weak topology for a sieve-type Robbins-Monro procedure. The papers [14, 24, 34] treat specific expressions of $H : \mathcal{H} \times \mathcal{V} \to \mathcal{H}$ with sieve approaches. The paper [8] also proposes a sieve approach in a SA scheme, but each iteration necessitates the exact computation of a projection on a finite subspace of the Hilbert space (see [8, Algorithm TRMP]), which is not practical (exact computation of the coefficients u_i by the integral in (1.4)). We will see below that, when $\mathcal{H} = L^2_\pi$ for some probability distribution π, our approach allows replacing the computation of this expectation by a simulated sample mean.

None of the above literature in an infinite dimensional Hilbert space $\mathcal{H}$ is suitable for UQ. Indeed, all these previous papers solve problems of the form

$$\text{Finding } \zeta^\star \in \mathcal{H} : \qquad \int H(\zeta^\star, v) \mu(\mathrm{d}v) = 0, \tag{1.7}$$

so that, first, the distribution μ does not account for the uncertainty and, second, for any v, the computation of the quantity $H(\zeta, v)$ may have a prohibitive computational cost, as it entails the infinite dimensional input data (function) ζ.

1.2. Contributions and Outline of the Paper. Our SA algorithm combines *(i)* the sieve approach in the special case $\mathcal{H}$ is L^2_π for some probability distribution π, *(ii)* the UQ framework

by allowing μ in (1.7) to depend on $\theta \in \Theta$ with an a priori knowledge $\theta \sim \pi$ (see Eq. (1.2)), and *(iii)* a tractable computational cost for the function H. Moreover, while addressing a more general framework where the projection on a subspace of $\mathcal{H}$ is approximated, our convergence proof relies on weaker assumptions (see Section 3.1, where we show that, in our setup, the assumptions of [8] may fail to hold). As a result:

- We obtain a fully constructive, easy to implement algorithm for the UQ analysis of SA limits in a chaos expansion setup, dubbed UQSA (Uncertainty Quantification for Stochastic Approximation);
- We are able to derive a convergence proof under easy-to-check hypotheses, in terms of underlying problems corresponding to fixed values of θ, avoiding conditions involving Hilbert space notions that are often hard to check in practice (or may even be violated);
- We discuss complexity issues and provide extensive reports and discussion on numerical tests.

The paper is outlined as follows. The UQSA algorithm is introduced in Section 2. Section 3 states the almost-sure convergence of the algorithm and its L^p convergence w.r.t. the underlying Hilbert space norm. The proof is deferred to Section 4. Section 5 presents the results of numerical experiments, including a detailed discussion of the choice of the design parameters.

Note that, beyond model uncertainty, applications of our approach include sensitivity analysis w.r.t. θ, or quasi-regression of an unknown function (see [1]), for instance in the context of outer Monte Carlo computations involving some unknown inner function $\zeta^\star = \zeta^\star(\theta)$ (cf. [4]).

2. Problem Formulations and Algorithmic Designs.

Let $\mathcal{V}$ be a metric space endowed with its Borel σ-field, Θ be a measurable subset of $\mathbb{R}^d$, and $H : \mathbb{R}^q \times \mathcal{V} \times \Theta \to \mathbb{R}^q$ be given. Let π be a probability distribution on Θ and μ be a transition kernel from Θ to $\mathcal{V}$. For any measurable functions $f, g : \Theta \to \mathbb{R}$, we define the scalar product induced by π by

$$\langle f; g \rangle_\pi := \int_\Theta f(\theta) g(\theta) \pi(\mathrm{d}\theta). \tag{2.1}$$

By extension, for measurable functions $f = (f_1, \cdots, f_q) : \Theta \to \mathbb{R}^q$ and $g : \Theta \to \mathbb{R}$, we write in vector form

$$\langle f; g \rangle_\pi := \begin{bmatrix} \langle f_1; g \rangle_\pi \\ \cdots \\ \langle f_q; g \rangle_\pi \end{bmatrix}. \tag{2.2}$$

We denote by L^2_π the Hilbert space of functions $f : \Theta \to \mathbb{R}^q$ such that the norm $\|f\|_\pi := \sqrt{\sum_{i=1}^q \langle f_i; f_i \rangle_\pi}$ is finite. In the special case where $q = 1$, we write $L^{2,1}_\pi$.

Hereafter, all the r.v. are defined on the same probability space, with probability and expectation denoted by $\mathbb{P}$ and $\mathbb{E}$.

2.1. The naive SA algorithm. A naive approach for finding $\zeta^\star$ in $\mathcal{S}$ (see (1.3)) would be to calculate $\zeta^\star(\theta)$ for each value of θ separately, for example by the following standard SA scheme (see [5, 10, 19]): Given a deterministic sequence $\{\gamma_k, k \geq 1\}$ of positive step sizes and a sequence of i.i.d. r.v. $\{V_k, k \geq 1\}$ sampled from $\mu(\theta, \cdot)$, obtain $\zeta^\star(\theta)$ as the limit of an iterative scheme

$$\zeta^{k+1}(\theta) = \zeta^k(\theta) - \gamma_{k+1}\, H(\zeta^k(\theta), V_{k+1}, \theta), \qquad k \geq 0. \tag{2.3}$$

Explicit conditions can be formulated to the effect that $\zeta^\star(\theta) = \lim_k \zeta^k(\theta)$ holds $\mathbb{P}$-a.s. (see e.g. [10, Chapter 1]). When $\lim_k k\gamma_k = O(1)$, the error $\mathbb{E}\left[|\zeta^k(\theta) - \zeta^\star(\theta)|^2\right]$ after k iterations (and thus k Monte Carlo samples) is $O(1/k)$ (see [10, Chapter 2]). Here the expectation is w.r.t. the random variable $\zeta_k(\theta)$, for any fixed value θ. When $\lim_k k\gamma_k = +\infty$, the same rate of convergence can be reached by replacing $\zeta^k(\theta)$ with an averaged estimator (see the Polyak-Ruppert procedure [25]).

However, except in the case where Θ is finite with few elements, the estimation of $\zeta^\star(\theta)$, separately for each $\theta \in \Theta$, is unfeasible or too demanding computationally.

2.2. Chaos Expansion Setup. Let $\{\theta \mapsto B_i(\theta), i \geq 0\}$ be an orthonormal basis of $L^{2,1}_\pi$ (for the scalar product (2.1)). Orthonormal polynomials are natural candidates, but there are other possibilities. For examples of the most commonly used bases, we refer the reader to [7, Chapter 2]. For $x, y \in \mathbb{R}^q$ we denote by $x \cdot y$ and $|x|$ the scalar product and the Euclidean norm in $\mathbb{R}^q$. We denote by l^2 the normed vector space of the $\mathbb{R}^q$-valued sequences $\{u_i, i \geq 0\}$ with $\|u\|^2_{l^2} := \sum_{i\geq 0} |u_i|^2 < +\infty$. As is well known, given an orthonormal basis $\{B_i, i \geq 0\}$ in $L^{2,1}_\pi$, any function $\zeta \in L^2_\pi$ is characterized by a sequence $\{u_i, i \geq 0\}$ in l^2 such that

$$\zeta(\cdot) = \sum_{i\geq 0} u_i B_i(\cdot). \tag{2.4}$$

Throughout the paper, we use the natural isomorphism $\mathbf{Is} : l^2 \to L^2_\pi$ given by

$$\zeta = \mathbf{Is}(u) = \sum_{i\geq 0} u_i B_i, \text{ i.e. } u_i = \langle \zeta; B_i \rangle_\pi \text{ for each } i \in \mathbb{N}, \tag{2.5}$$

and the corresponding isometry $\|\zeta\|_\pi = \|u\|_{l^2}$ (see [22, Proposition 10.32]), hence

$$\langle u, u' \rangle_{l^2} = \langle \mathbf{Is}(u), \mathbf{Is}(u') \rangle_\pi, \; u, u' \in l^2, \tag{2.6}$$

where the scalar products $\langle \cdot, \cdot \rangle_{l^2}$ and $\langle \cdot, \cdot \rangle_\pi$ are the ones associated with the respective (Hilbert) norms.

An alternative strategy for finding $\zeta^\star$ in $\mathcal{S}$ under the additional constraint that $\zeta^\star \in L^2_\pi$ would consist in the combination of *(i)* a truncation at a fixed level m of the expansion (2.5), *(ii)* the estimation of the first $(m+1)$ $\mathbb{R}^q$-valued coefficients $\{u^\star_i, 0 \leq i \leq m\}$ of $\zeta^\star$, based on a Monte Carlo approximation

$$u^\star_i = \langle \zeta^\star; B_i \rangle_\pi \approx \hat{u^\star_i} := \frac{1}{M} \sum_{k=1}^{M} \widehat{\zeta^\star(\theta_{k,i})}\, B_i(\theta_{k,i}), \tag{2.7}$$

where $\{\theta_{k,i}, k \geq 1, i \leq m\}$ are i.i.d. with distribution π, and *(iii)* for each value $\theta_{k,i}$, the computation of an approximation $\widehat{\zeta^\star(\theta_{k,i})}$ of $\zeta^\star(\theta_{k,i})$ (which can be done e.g. through (2.3)). Let us discuss the computational cost of the step *(ii)* in the case where $q = 1$ (for ease of notation), $\theta \in \mathbb{R}^d$, and of a Jacobi polynomial basis for the B_i's. The following control on the truncation error of $\zeta^\star$ holds (see [11, Theorem 6.4.2] or [7, Chapter 5]):

$$\left\| \sum_{i>m} u_i^* B_i \right\|_\pi^2 = O\left(m^{-\frac{2(\eta-1)}{d}}\right), \tag{2.8}$$

where η is the order of continuous differentiability of $\zeta^\star$ (in some cases this bound may be strengthened to $O(m^{-\frac{2\eta}{d}})$). Neglecting the error associated with the approximation $\widehat{\zeta^\star(\theta_{k,i})}$, we have

$$\mathbb{E}\left[\left\| \sum_{i=0}^{m} (u_i^* - \widehat{u_i^*}) B_i \right\|_\pi^2 \right] = O\left(\frac{m}{M}\right). \tag{2.9}$$

For balancing the error components (2.8) and (2.9), we must set $M \sim m^{1+\frac{2(\eta-1)}{d}}$. To reach an accuracy ϵ, m has to increase as $\epsilon^{-d/(2(\eta-1))}$ and M has to increase as $\epsilon^{-(1+d/(2(\eta-1)))}$. The computational cost in terms of number of Monte Carlo samples to estimate m coefficients is therefore $m \times M = \epsilon^{-(1+d/(\eta-1))}$. This quantity suffers from the curse of dimensionality, which makes this approach inefficient for high d, especially when combined with a nested procedure for the computation of $\widehat{\zeta^\star(\theta_{k,i})}$.

2.3. The UQSA Algorithm. Through the decomposition (2.4) and the isomorphism (2.5), finding $\zeta^\star$ in $L^2_\pi \cap \mathcal{S}$ (see (1.3)) can be restated as

$$\text{Finding } u^\star \text{ in } l^2: \quad \int_{\mathcal{V}} H\left(\sum_{i\geq 0} u_i^\star B_i(\theta), v, \theta\right) \mu(\theta, \mathrm{d}v) = 0, \ \pi\text{-a.s.} \tag{2.10}$$

If in addition, $h(\zeta^\star(\cdot), \cdot) \in L^2_\pi$ (see condition A4 below when deriving sufficient conditions for the convergence of UQSA), the problem is equivalent to finding $\zeta^\star \in L^2_\pi$ such that

$$\int_\Theta \left(\int_{\mathcal{V}} H\left(\zeta^\star(\theta), v, \theta\right) \ \mu(\theta, \mathrm{d}v) \right) B_i(\theta) \pi(\mathrm{d}\theta) = 0_{\mathbb{R}^q}, \qquad \forall i \geq 0. \tag{2.11}$$

This observation can be used for devising an original SA scheme for the $u_i^\star$ in (2.10). A first attempt in this direction is to restrict the problem to a set of functions $\zeta(\cdot)$ of the form $\sum_{i=0}^m u_i B_i(\cdot)$, together with $h(\zeta(\cdot), \cdot)$, for some fixed $m \geq 1$. In this case an SA algorithm for the computation of $\{u_i^\star, i \leq m\}$ is given by: for any $k \geq 0$,

$$u_i^{k+1} = u_i^k - \gamma_{k+1} \ H\left(\sum_{j=0}^m u_j^k B_j(\theta_{k+1}), V_{k+1}, \theta_{k+1}\right) B_i(\theta_{k+1}), \quad i = 0, \cdots, m, \tag{2.12}$$

where the $\{(\theta_k, V_k), k \geq 0\}$ are i.i.d. with distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$. However, in practice, neither $\zeta^\star$ nor $h(\zeta(\cdot), \cdot)$ are typically of the form $\sum_{i=0}^m u_i B_i$ and, even when they are, we may not know for which m. We emphasize that, in the general case $\zeta^\star \in L^2_\pi$, as the first argument of H in (2.12) is the current truncation $\sum_{i=0}^m u_i^k B_i(\theta_{k+1})$ and not $\zeta^\star(\theta_{k+1})$, this algorithm does not converge to the projection of $\zeta^\star$ onto the space spanned by $\{B_0, \cdots, B_m\}$. See the numerical evidence reported in Section 5.4.

In order to tackle the infinite dimensionality of the problem space L^2_π, our algorithm introduces further an $\mathbb{N}$-valued sequence $\{m_k, k \geq 0\}$ specifying the number of coefficients u_i that are updated at each iteration. The sequence $\{m_k, k \geq 0\}$ is non-decreasing and converges to ∞, which is the key property to avoid the truncation phenomenon.

The UQSA algorithm is given in Algorithm 2.1. The inputs of the algorithm are: a positive stepsize sequence $\{\gamma_k, k \geq 1\}$; two positive integer-valued sequences $\{m_k, k \geq 1\}$ and $\{M_k, k \geq 1\}$ corresponding to the number of non null coefficients in the approximation of $\zeta^\star$ and to the number of Monte Carlo draws of the pair (θ, V) at each iteration k; an initial value $u_0 \in \mathbb{R}^{q\, m_0}$; a total number of iterations K; a closed convex subset $\mathcal{A}$ of l^2 on which to project each newly updated sequence of coefficients, with related projection operator denoted by $\Pi_{\mathcal{A}}$.

Algorithm 2.1 The UQSA algorithm

Input: Sequences $\{\gamma_k, k \geq 1\}$, $\{m_k, k \geq 0\}$, $\{M_k, k \geq 1\}$, $K \geq 1$, $\{u_i^0, i = 0, \ldots, m_0\}$, a convex set $\mathcal{A} \subseteq l^2$.

for $k = 1, \ldots, K$ **do**

  for $s = 1, \ldots, M_k$ **do**

    Independently from the past, sample (θ_k^s, V_k^s) under the distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$

  end for

  For all $i > m_{k-1}$ define $u_i^k = 0$

  for $i = 0, \ldots, m_k$ **do**

    $\hat{u}_i^k = u_i^{k-1} - \gamma_k M_k^{-1} \sum_{s=1}^{M_k} H\left(\sum_{j=0}^{m_{k-1}} u_j^{k-1} B_j(\theta_k^s), V_k^s, \theta_k^s\right) B_i(\theta_k^s)$

  end for

  $u^k = \Pi_{\mathcal{A}}(\hat{u}^k)$

end for

return The vector $\{u_i^K, i = 0, \ldots, m_K\}$.

end

The motivations for the introduction of the projection set $\mathcal{A}$ at each iteration, and for the averaging over M_k Monte Carlo draws at step k, are discussed in Sections 3.2 and 5.5.3, respectively. The output of the algorithm is a sequence $\{u_i^K, i \leq m_K\}$ approximating a solution $u^\star$ to the problem (2.10). The associated function,

$$\zeta^K := \sum_{i=0}^{m_K} u_i^K B_i, \tag{2.13}$$

is an approximation of a solution $\zeta^\star$ to the problem (2.11).

3. The UQSA Algorithm Converges. The roots of h are unchanged if H is replaced with $(-H)$ (see (1.2)). In this section, we use the convention which allows h to be thought of as

the gradient of a convex function. This is also consistent with Example 1 (see also the last paragraph there), where $h(z) = \mathbb{E}\left[\nabla_z L(z, S_T)\right]$.

3.1. Assumptions. Let

$$\mathcal{T}^\star := \left\{ u^\star \in l^2; \qquad \int_{\mathcal{V}} H\left(\sum_{i\geq 0} u_i^\star B_i(\theta), v, \theta\right) \mu(\theta, \mathrm{d}v) = 0, \ \pi\text{-a.s.} \right\}. \tag{3.1}$$

Note that (2.10) is equivalent to $u^\star \in \mathcal{T}^\star$.

A1. *The set $\mathcal{T}^\star$ is compact and non-empty.*

$\mathcal{T}^\star$ may contain several elements. In the SA literature, allowing for multiple limits is quite standard. This encompasses for example, the important case of computing quantiles and average quantiles[1] (cf. [3, Section 2]): the SA approximation for the quantile component may converge to several limits (e.g. when, at these locations, the cumulative distribution function is flat), while, for the average quantile component, the limit is unique (see e.g. [4, Lemma A.1.]).

A2. *$\{M_k, k \geq 1\}$ and $\{m_k, k \geq 1\}$ are deterministic sequences of positive integers; $\{\gamma_k, k \geq 1\}$ is a deterministic sequence of positive real numbers such that, for some $\kappa \in (0, 1]$,*

$$\sum_{k\geq 1} \gamma_k = +\infty, \quad \sum_{k\geq 1} \gamma_k^{1+\kappa} < +\infty, \quad \sum_{k\geq 1} \gamma_k^2 \frac{Q_{m_k}}{M_k} < +\infty, \quad \sum_{k\geq 1} \gamma_k^{1-\kappa} q_{m_k} < +\infty, \tag{3.2}$$

where the sequences $\{q_m, m \geq 0\}$ and $\{Q_m, m \geq 0\}$ are defined by

$$q_m := \sup_{u^\star \in \mathcal{T}^\star} \sum_{i>m} |u_i^\star|^2, \qquad Q_m := \sup_{\theta\in\Theta} \sum_{i\leq m} |B_i(\theta)|^2. \tag{3.3}$$

Since $\mathcal{T}^\star$ is compact, we have $\lim_m q_m = 0$ (see e.g. Lemma A.1 in Section A). Assumption A2 requires, in particular, that $Q_m < +\infty$ for any m: if Θ is bounded, it holds for any continuous basis functions. In the case of polynomial basis, the inverse of the coefficient Q_m is the minimum of m-th Christoffel function, which is deeply studied in [23].

We now introduce two possible conditions on the behavior of H. In Section 3.2 we consider different cases illustrating the two sets of assumptions.

A3. *One of the following conditions:*

a) There exists a constant C_H such that, for any $z \in \mathbb{R}^q$,

$$\sup_{\theta\in\Theta} \int_{\mathcal{V}} |H(z, v, \theta)|^2 \mu(\theta, \mathrm{d}v) \leq C_H(1 + |z|^2).$$

b) The map from L^2_π into $\mathbb{R}$ defined by $\zeta \mapsto \int_{\mathcal{V}\times\Theta} |H(\zeta(\theta), v, \theta)|^2 \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta)$ is bounded, i.e. it maps bounded sets into bounded sets.

[1] also called Value-at-Risk and Expected Shortfall in financial applications.

The condition A3 implies that for any $z \in \mathbb{R}^q$, and $\theta \in \Theta$,

$$\int_{\mathcal{V}\times\Theta} |H(z,v,\theta)|\, \mu(\theta, \mathrm{d}v)\pi(\mathrm{d}\theta) < \infty, \qquad h(z,\theta) := \int_{\mathcal{V}} H(z,v,\theta)\, \mu(\theta,\mathrm{d}v) \text{ exists,} \tag{3.4}$$

and

$$\forall\, \zeta \in L^2_\pi, \qquad \theta \mapsto h(\zeta(\theta),\theta) \in L^2_\pi. \tag{3.5}$$

Note that the condition A3-a) implies A3-b), which in turn implies that $\zeta \mapsto h(\zeta(\cdot),\cdot)$ is a bounded map from L^2_π into itself. We also assume

A4. *The mapping $\zeta \mapsto h(\zeta(\cdot),\cdot)$ from L^2_π into itself is continuous.*

A5. *For π-almost every θ: for any $z, z^\star \in \mathbb{R}^q$ such that $h(z,\theta) \neq 0$ and $h(z^\star,\theta) = 0$,*

$$(z - z^\star)\cdot h(z,\theta) > 0,$$

where $a \cdot b$ denotes the usual scalar product on $\mathbb{R}^q$.

Note that A5 is the standard assumption for SA with fixed θ, and can be seen as a local separation condition. In the case of a single minimum ($\mathcal{T}^\star = \{u^\star\}$ and $\zeta^\star = \mathbf{Is}(\mathcal{T}^\star)$), previous works on SA in a Hilbert space $\mathcal{H}$ assume an *integrated* drift condition (see e.g. [8, Assumption A3P(2)]). For $\zeta \in L^2_\pi$, denoting $\zeta^m := \sum_{i=0}^m \langle \zeta; B_i\rangle_\pi B_i$, this condition is: there exists m^* such that, for any $m \geq m^*$ and $\eta > 0$,

$$\inf_{\{\zeta : \|\zeta^m - \zeta^\star\|^2_\pi > \eta\}} \int_\Theta (\zeta^m(\theta) - \zeta^\star(\theta)) \cdot (h(\zeta^m(\cdot),\cdot))^m(\theta)\ \pi(\mathrm{d}\theta) > 0. \tag{3.6}$$

The assumption (3.6) may not hold in some settings : for example, set $h(z,\theta) := z - \varphi(\theta)$ for some $\varphi \in L^2_\pi$; then $\zeta^\star = \varphi$, and $(h(\zeta^m(\cdot),\cdot))^m = \zeta^m - \zeta^{\star m}$. This implies that the integral in (3.6) is

$$\int_\Theta |\zeta^m(\theta) - \zeta^{\star m}(\theta)|^2\ \pi(\mathrm{d}\theta) = \|\zeta^m - \zeta^{\star m}\|^2_\pi$$

and, for any m, we can choose ζ and $\eta > 0$ such that this quantity is zero (choose $\zeta = \zeta^{\star m}$ and $\eta = 0.5\, \|\zeta^\star - \zeta^{\star m}\|^2_\pi$).

A6. *For any $B > 0$, there exists a constant $C_B > 0$ such that, for any $(\zeta, \zeta^\star) \in L^2_\pi \times \boldsymbol{Is}(\mathcal{T}^\star)$ with $\|\zeta - \zeta^\star\|_\pi \leq B$,*

$$\int_\Theta (\zeta(\theta) - \zeta^\star(\theta)) \cdot h(\zeta(\theta),\theta)\, \pi(\mathrm{d}\theta) \geq C_B \min_{\bar\zeta \in \boldsymbol{Is}(\mathcal{T}^\star)} \left\|\zeta - \bar\zeta\right\|^2_\pi.$$

Note that the above minimum exists since, by A1, $\mathbf{Is}(\mathcal{T}^\star)$ is compact.

3.2. Projection Set. Three possible choices regarding the projection set $\mathcal{A}$ in Algorithm 2.1 are considered for the convergence analysis. For all of them, it is assumed that $\mathcal{A}$ includes $\mathcal{T}^\star$.

Case 3.1. *$\mathcal{A}$ is l^2 (i.e. no projection).*

Case 3.2. *$\mathcal{A}$ is a closed ball of l^2 containing $\mathcal{T}^\star$.*

Case 3.3. *$\mathcal{A}$ is a closed convex set of l^2 containing $\mathcal{T}^\star$, with compact intersections with closed balls of l^2.*

Note that the projection set $\mathcal{A}$ is bounded in Case 3.2, unbounded in Case 3.1 and potentially unbounded in Case 3.3. Example 3 below is an illustration of Case 3.3.

Case 3.1 is the most convenient from the algorithmic viewpoint since no actual projection is required. However, it requires the stronger condition A3-a) to ensure the stability, combined with A6 to ensure the convergence.

The projection in the Case 3.2 is explicit: given the ball $\{u \in l^2 : \|u\|_{l^2} \leq B\}$, it is

$$u \mapsto \min\left(1, \frac{B}{\|u\|_{l^2}}\right) u. \tag{3.7}$$

The milder assumption A3-b) is required for the stability, again combined with A6 for the convergence.

Case 3.3 requires a potentially nontrivial projection on a closed convex set: see e.g. Example 3 below. The stronger condition A3-a) is required for both the stability and the convergence, but A6 is not needed.

Example 3. *Given a positive sequence $\{a_n, n \geq 0\}$ such that $\sum_{i\geq 0} a_i^2 < \infty$ and a strictly increasing sequence of non-negative integers $\{d_n, n \geq 0\}$, define*

$$\mathcal{A} := \left\{ u \in l^2 : \sum_{d_n \leq i < d_{n+1}} |u_i|^2 \leq a_n^2 \quad \forall n \geq 0 \right\}. \tag{3.8}$$

If $\mathcal{A}$ is bounded (which holds when $d_0 = 0$ for example), then it is a compact convex subset of l^2. Similarly, the set $\mathcal{A} \cap \{u \in l^2 : \sum_{i\geq 0} u_i^2 \leq B\}$ is a compact subset for any $B > 0$ (see Lemma B.1 for both cases). The orthogonal projection on $\mathcal{A}$ consists of projecting $(u_{d_n}, \ldots, u_{d_{n+1}-1})$ on the ball of radius a_n for all $n \geq 0$. The final choice of the sequence $(a_n)_n$ is left to the user expertise, typically based on a preliminary mathematical analysis giving bounds on $(u_i)_i$.

3.3. Main Result.

Theorem 1. *Assume A1, A2, A4 and A5; also assume A3-a) if $\mathcal{A}$ is unbounded and A3-b) if $\mathcal{A}$ is bounded. Let there be given i.i.d. random variables $\{(\theta_k^s, V_k^s), 1 \leq s \leq M_k, k \geq 1\}$ with distribution $\pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v)$. Let u^K be the output of the UQSA Algorithm, and let ζ^K be the associated function through (2.13).*

Stability. *For any $\zeta^\star \in \boldsymbol{Is}(\mathcal{T}^\star)$, $\lim_{k\to+\infty} \left\|\zeta^k - \zeta^\star\right\|_\pi$ exists, is finite a.s., and we have*

$$\sup_{k\geq 0} \mathbb{E}\left[\left\|\zeta^k - \zeta^\star\right\|_\pi^2\right] < +\infty. \tag{3.9}$$

Convergence. *In addition, in case 3.3, and in cases 3.1 and 3.2 under the additional assumption A6, there exists a random variable ζ^∞ taking values in $\boldsymbol{Is}(\mathcal{T}^\star)$ such that*

$$\lim_{k\to\infty} \left\|\zeta^k - \zeta^\infty\right\|_\pi = 0 \text{ a.s. and, for any } p \in (0,2), \lim_{k\to\infty} \mathbb{E}\left[\left\|\zeta^k - \zeta^\infty\right\|_\pi^p\right] = 0. \tag{3.10}$$

Remark 3.1. *The standard assumption ensuring a Central Limit Theorem (CLT) for SA algorithms in a Hilbert space (cf. [8, Assumption B3(1)] or [24, Section 3, equation 3.3]) is not satisfied in our setup: as a counter-example, one can take a function $h(z,\theta)$ such that $\partial_z h(\zeta^\star(\theta),\theta) = \theta$ and any polynomial basis, due to the recurrence relations of order two that are intrinsic to such bases. The study of convergence rates and CLT for the UQSA algorithm is therefore a problem per se, which we leave for future research (see [31, Chapter 6] for partial results).*

4. Proof of Theorem 1.

For any $z = (z_1, \cdots, z_q) \in \mathbb{R}^q$ and any real-valued sequence $p := \{p_i, i \geq 0\}$ such that $\sum_{i\geq 0} p_i^2 < \infty$ we write

$$z \otimes p := ((z_1 p_0, \cdots, z_q p_0), (z_1 p_1, \cdots, z_q p_1), \cdots),$$

which is in l^2. Set $\mathbf{B}^m(\theta) := (B_0(\theta), \ldots, B_m(\theta), 0, 0, \ldots)$. Define the filtration

$$\mathcal{F}_k := \sigma\left(\theta_\ell^s, V_\ell^s, 1 \leq s \leq M_\ell, 1 \leq \ell \leq k\right), \; k \geq 1.$$

We fix $u^\star \in \mathcal{T}^\star$, which exists by A1, and we set $\zeta^\star := \mathbf{Is}(u^\star)$.

4.1. Stability.

The first step is to prove that the algorithm is stable in the sense that

$$\lim_k \left\| u^k - u^\star \right\|_{l^2} \text{ exists a.s. }, \tag{4.1}$$

$$\sup_k \mathbb{E}\left[\left\| u^k - u^\star \right\|_{l^2}^2 \right] < +\infty, \tag{4.2}$$

$$\liminf_{k\to\infty} \int_\Theta (\zeta^k(\theta) - \zeta^\star(\theta)) \cdot h(\zeta^k(\theta),\theta)\pi(\mathrm{d}\theta) = 0, \quad \text{a.s.} \tag{4.3}$$

where ζ^k is associated to u^k through (2.13). By definition of u^{k+1} (see Algorithm 2.1) and the property $\Pi_{\mathcal{A}}(u^\star) = u^\star$ (which holds since $\mathcal{T}^\star \subseteq \mathcal{A}$), we obtain

$$\begin{aligned}\left\| \zeta^{k+1} - \zeta^\star \right\|_\pi^2 = \left\| u^{k+1} - u^\star \right\|_{l^2}^2 = \left\| \Pi_{\mathcal{A}}(\hat{u}^{k+1}) - \Pi_{\mathcal{A}}(u^\star) \right\|_{l^2}^2 \leq \left\| \hat{u}^{k+1} - u^\star \right\|_{l^2}^2 \\ \leq \left\| u^k - u^\star - \gamma_{k+1}\mathcal{H}^k - \gamma_{k+1}\eta^{k+1} \right\|_{l^2}^2,\end{aligned}$$

where

$$\begin{aligned}\mathcal{H}^k :=& \mathbb{E}\left[\frac{1}{M_{k+1}} \sum_{s=1}^{M_{k+1}} H\left(\zeta^k(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) \otimes \mathbf{B}^{m_{k+1}}(\theta_{k+1}^s) \middle| \mathcal{F}_k \right] \\ =& \int_{\Theta\times\mathcal{V}} H\left(\zeta^k(\theta), v, \theta\right) \otimes \mathbf{B}^{m_{k+1}}(\theta)\; \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v) = \int_\Theta h(\zeta^k(\theta),\theta) \otimes \mathbf{B}^{m_{k+1}}(\theta)\; \pi(\mathrm{d}\theta), \\ \eta^{k+1} :=& \frac{1}{M_{k+1}} \sum_{s=1}^{M_{k+1}} H\left(\zeta^k(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) \otimes \mathbf{B}^{m_{k+1}}(\theta_{k+1}^s) - \mathcal{H}^k.\end{aligned}$$

For the equivalent definitions of $\mathcal{H}^k$, we used the Fubini theorem (using A2 and A3). Observe that, by definition of $\mathbf{B}^{m_k}$, $\mathcal{H}^k$ and η^{k+1} are sequences in l^2 such that $\mathcal{H}_i^k = 0_{\mathbb{R}^q}$ and $\eta_i^{k+1} = 0_{\mathbb{R}^q}$ for all $i > m_{k+1}$. Define

$$\overline{\mathcal{H}}_i^k := \begin{cases} \mathcal{H}_i^k & i \leq m_{k+1}, \\ \int_\Theta h(\zeta^k(\theta), \theta) B_i(\theta)\ \pi(\mathrm{d}\theta) & i > m_{k+1}. \end{cases}$$

Recalling that $u_i^k = 0_{\mathbb{R}^q}$ for $i > m_{k+1}$, we obtain

$$\begin{aligned} \left\|u^{k+1} - u^\star\right\|_{l^2}^2 &= \left\|u^k - u^\star\right\|_{l^2}^2 - 2\gamma_{k+1} \sum_{i=0}^{m_{k+1}} (u_i^k - u_i^\star) \cdot \mathcal{H}_i^k - 2\gamma_{k+1} \sum_{i=0}^{m_{k+1}} (u_i^k - u_i^\star) \cdot \eta_i^{k+1} \\ &\quad + 2\gamma_{k+1}^2 \sum_{i=0}^{m_{k+1}} \eta_i^{k+1} \cdot \mathcal{H}_i^k + \gamma_{k+1}^2 \left\|\eta^{k+1}\right\|_{l^2}^2 + \gamma_{k+1}^2 \left\|\mathcal{H}^k\right\|_{l^2}^2 \\ &= \left\|u^k - u^\star\right\|_{l^2}^2 - 2\gamma_{k+1} \sum_{i \geq 0} (u_i^k - u_i^\star) \cdot \overline{\mathcal{H}}_i^k - 2\gamma_{k+1} \sum_{i=0}^{m_{k+1}} (u_i^k - u_i^\star) \cdot \eta_i^{k+1} \\ &\quad + 2\gamma_{k+1}^2 \sum_{i=0}^{m_{k+1}} \eta_i^{k+1} \cdot \mathcal{H}_i^k + \gamma_{k+1}^2 \left\|\eta^{k+1}\right\|_{l^2}^2 + \gamma_{k+1}^2 \left\|\mathcal{H}^k\right\|_{l^2}^2 - 2\gamma_{k+1} \sum_{i > m_{k+1}} u_i^\star \cdot \overline{\mathcal{H}}_i^k. \end{aligned} \tag{4.4}$$

A5 implies that, for each θ,

$$\sum_{i \geq 0} \left((u_i^k - u_i^\star) \cdot h\left(\zeta^k(\theta), \theta\right) \right) B_i(\theta) = (\zeta^k(\theta) - \zeta^\star(\theta)) \cdot h\left(\zeta^k(\theta), \theta\right) \geq 0.$$

Taking expectation w.r.t. $\theta \sim \pi$, we obtain by the isometry (2.6), for all $k \geq 0$

$$R^k := \int_\Theta (\zeta^k(\theta) - \zeta^\star(\theta)) \cdot h\left(\zeta^k(\theta), \theta\right) \pi(\mathrm{d}\theta) = \sum_{i \geq 0} (u_i^k - u_i^\star) \cdot \overline{\mathcal{H}}_i^k, \tag{4.5}$$

which is non negative. Also note that $\sum_{i=0}^{+\infty} (u_i^k - u_i^\star) \cdot \mathcal{H}_i^k \in \mathcal{F}_k$. By definition, $\mathbb{E}\left[\eta_i^{k+1} | \mathcal{F}_k\right] = 0$, $\mathcal{H}^k \in \mathcal{F}_k$ and $u^k \in \mathcal{F}_k$ so that

$$\mathbb{E}\left[\sum_{i=0}^{m_{k+1}} \eta_i^{k+1} \cdot \mathcal{H}_i^k \middle| \mathcal{F}_k\right] = 0, \qquad \mathbb{E}\left[\sum_{i=0}^{m_{k+1}} (u_i^k - u_i^\star) \cdot \eta_i^{k+1} \middle| \mathcal{F}_k\right] = 0. \tag{4.6}$$

Let us consider the term $\|\eta^{k+1}\|_{l^2}^2$. We write

$$\begin{aligned}
\mathbb{E}\left[\left\|\eta^{k+1}\right\|_{l^2}^2 \Big| \mathcal{F}_k\right] &\leq \mathbb{E}\left[\left\|\frac{1}{M_{k+1}}\sum_{s=1}^{M_{k+1}} H\left(\zeta^k(\theta_{k+1}^s), V_{k+1}^s, \theta_{k+1}^s\right) \otimes \mathbf{B}^{m_{k+1}}(\theta_{k+1}^s) - \mathcal{H}^k\right\|_{l^2}^2 \Big| \mathcal{F}_k\right] \\
&\leq \frac{1}{M_{k+1}}\int_{\Theta\times\mathcal{V}} \left\|H\left(\zeta^k(\theta), v, \theta\right) \otimes \mathbf{B}^{m_{k+1}}(\theta)\right\|_{l^2}^2 \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v) \\
&= \frac{1}{M_{k+1}}\int_{\Theta\times\mathcal{V}} \left|H\left(\zeta^k(\theta), v, \theta\right)\right|^2 \left(\sum_{i=0}^{m_{k+1}} B_i(\theta)^2\right) \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v) \\
&\leq \frac{Q_{m_{k+1}}}{M_{k+1}}\int_{\Theta\times\mathcal{V}} \left|H\left(\zeta^k(\theta), v, \theta\right)\right|^2 \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v). \qquad (4.7)
\end{aligned}$$

Next we consider the term $2\gamma_{k+1}\sum_{i>m_{k+1}} u_i^\star \cdot \overline{\mathcal{H}}_i^k$. By using $2ab \leq a^2 + b^2$ with $a \leftarrow (\gamma_{k+1}^{1-\kappa})^{1/2}|u_i^\star|$ and $b \leftarrow (\gamma_{k+1}^{1+\kappa})^{1/2}|\overline{\mathcal{H}}_i^k|$, we have

$$\begin{aligned}
\left|2\gamma_{k+1}\sum_{i>m_{k+1}} u_i^\star \cdot \overline{\mathcal{H}}_i^k\right| &\leq \gamma_{k+1}^{1-\kappa}\left(\sum_{i>m_{k+1}}^{+\infty} |u_i^\star|^2\right) + \gamma_{k+1}^{1+\kappa}\left(\sum_{i>m_{k+1}}^{+\infty} \left|\overline{\mathcal{H}}_i^k\right|^2\right) \\
&\leq \gamma_{k+1}^{1-\kappa} q_{m_{k+1}} + \gamma_{k+1}^{1+\kappa}\left\|\overline{\mathcal{H}}^k\right\|_{l^2}^2, \qquad (4.8)
\end{aligned}$$

where we used A2 in the last inequality. Note that

$$\begin{aligned}
\left\|\overline{\mathcal{H}}^k\right\|_{l^2}^2 &= \sum_{i=0}^{+\infty}\left|\int_\Theta h(\zeta^k(\theta), \theta)B_i(\theta)\ \pi(\mathrm{d}\theta)\right|^2 = \int_\Theta \left|h(\zeta^k(\theta), \theta)\right|^2\ \pi(\mathrm{d}\theta) \\
&\leq \int_{\Theta\times\mathcal{V}} \left|H\left(\zeta^k(\theta), v, \theta\right)\right|^2 \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v). \qquad (4.9)
\end{aligned}$$

Combining (4.4) to (4.9), we obtain

$$\begin{aligned}
\mathbb{E}\Big[\left\|u^{k+1} - u^\star\right\|_{l^2}^2 | \mathcal{F}_k\Big] &\leq \left\|u^k - u^\star\right\|_{l^2}^2 - 2\gamma_{k+1}R^k + \gamma_{k+1}^{1-\kappa}q_{m_{k+1}} \\
&+ \left(\gamma_{k+1}^2 + \gamma_{k+1}^{1+\kappa} + \gamma_{k+1}^2\frac{Q_{m_{k+1}}}{M_{k+1}}\right)\int_{\Theta\times\mathcal{V}} \left|H\left(\zeta^k(\theta), v, \theta\right)\right|^2 \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v). \qquad (4.10)
\end{aligned}$$

To control the integral in (4.10), we distinguish two cases.

First case: $\mathcal{A}$ is unbounded. Using A3-a) we write

$$\begin{aligned}
\int_{\Theta\times\mathcal{V}} \left|H\left(\zeta^k(\theta), v, \theta\right)\right|^2 \pi(\mathrm{d}\theta)\mu(\theta, \mathrm{d}v) &\leq C_H\int_\Theta \left(1 + \left|\zeta^k(\theta)\right|^2\right)\pi(\mathrm{d}\theta) \\
&\leq C_1\left(1 + \left\|u^k - u^\star\right\|_{l^2}^2\right),
\end{aligned}$$

where $C_1 := 2C_H(1 + \sup_{u^\star\in\mathcal{T}^\star}\|u^\star\|_{l^2}^2)$. Note that C_1 is finite by A1.

Second case: $\mathcal{A}$ is bounded. Note that, by definition of u^k, there exists a constant B such that a.s. $\sup_{k\geq 0}\left\|u^k\right\|_{l^2}\leq B$. Assumption A3-b) implies that, for some deterministic $C_2>0$,

$$\sup_{k\geq 0}\int_{\Theta\times\mathcal{V}}|H(\zeta^k(\theta),v,\theta)|^2\pi(\mathrm{d}\theta)\mu(\theta,\mathrm{d}v)\leq C_2.$$

In either case, we deduce from (4.10) that

$$\mathbb{E}\left[\left\|u^{k+1}-u^\star\right\|_{l^2}^2\Big|\mathcal{F}_k\right]\leq\left\|u^k-u^\star\right\|_{l^2}^2-2\gamma_{k+1}R^k+\gamma_{k+1}^{1-\kappa}q_{m_{k+1}}$$
$$+\left(\gamma_{k+1}^2+\gamma_{k+1}^{1+\kappa}+\gamma_{k+1}^2\frac{Q_{m_{k+1}}}{M_{k+1}}\right)(C_1\vee C_2)\left(1+\left\|u^k-u^\star\right\|_{l^2}^2\right). \tag{4.11}$$

Conclusion. In view of the above controls and A2, the assumptions of the Robbins-Siegmund lemma are verified (see [27]). An application of this lemma yields that $\lim_k\left\|u^k-u^\star\right\|_{l^2}^2$ exists and $\sum_{k\geq 0}\gamma_{k+1}R^k<+\infty$ a.s.. This concludes the proof of (4.1). Taking expectations in (4.11) and applying the Robbins-Siegmund lemma to the sequence $\mathbb{E}\left[\left\|u^k-u^\star\right\|_{l^2}^2\right]$ yields (4.2). Also note that

$$R:=\liminf_{k\to+\infty}R^k=0,\quad\text{a.s.} \tag{4.12}$$

Indeed, on the event $\{R>0\}$, there exists a random index K, finite a.s. such that $R^k>R/2$ holds for any $k\geq K$, which implies that $\sum_{k\geq 0}\gamma_{k+1}R^k=+\infty$ (as, by assumption, $\sum_{k\geq 1}\gamma_k=+\infty$). Therefore $\{R>0\}\subseteq\{\sum_{k\geq 0}\gamma_{k+1}R^k=+\infty\}$, where we saw above that $\{\sum_{k\geq 0}\gamma_{k+1}R^k=+\infty\}$ is a zero probability event. Hence so is $\{R>0\}$, which proves (4.3).

We know from (4.1) that for any $\bar\zeta\in\mathbf{Is}(\mathcal{T}^\star)$, $\lim_k\left\|\zeta^k-\bar\zeta\right\|_\pi$ exists a.s. For later use, we need the existence of this limit simultaneously for all $\bar\zeta\in\mathbf{Is}(\mathcal{T}^\star)$ with probability one. Note that $\lim_k\left\|\zeta^k-\bar\zeta\right\|_\pi$ is continuous in $\bar\zeta$ (by triangle inequality). Using that $\mathbf{Is}(\mathcal{T}^\star)$ is separable as a subset of the separable Hilbert space L^2_π, we deduce that

$$\lim_k\left\|\zeta^k-\bar\zeta\right\|_\pi\text{ exists for all }\bar\zeta\in\mathbf{Is}(\mathcal{T}^\star),\quad\text{a.s.} \tag{4.13}$$

4.2. Proof of the Almost Sure Convergence in (3.10).

Proof for Case 3.1 or Case 3.2. Under the assumption A1, $\mathbf{Is}(\mathcal{T}^\star)$ is bounded so that, by (4.1), the r.v. $B:=\sup_{\zeta^\star\in\mathcal{T}^\star}\sup_k\left\|\zeta^k-\zeta^\star\right\|_\pi$ is finite with probability one. Since by (4.12) $\liminf_k R^k=0$, with probability one, there exists a subsequence k' such that $\lim_{k'}R^{k'}=0$.

From (4.5) and by A6 applied with $\zeta\leftarrow\zeta^{k'}$ and $\zeta^\star\leftarrow\mathbf{Is}(u^\star)$, there exists a positive random variable C_B (finite a.s. and independent of k by definition of the r.v. B) such that

$$R^{k'}\geq C_B\min_{\bar\zeta\in\mathbf{Is}(\mathcal{T}^\star)}\left\|\zeta^{k'}-\bar\zeta\right\|_\pi^2.$$

Let $\{\bar\zeta^k,k\geq 0\}$ be an $\mathbf{Is}(\mathcal{T}^\star)$-valued sequence such that, for all k,

$$\min_{\bar\zeta\in\mathbf{Is}(\mathcal{T}^\star)}\left\|\zeta^{k'}-\bar\zeta\right\|_\pi^2=\left\|\zeta^{k'}-\bar\zeta^k\right\|_\pi^2.$$

Such a sequence exists since $\mathcal{T}^\star$ is compact by A1. Using that $\lim_k R^{k'} = 0$ we obtain $\lim_k \left\|\zeta^{k'} - \bar{\zeta}^k\right\|_\pi = 0$ a.s. Since the sequence $\{\bar{\zeta}^k, k \geq 0\}$ is in a compact set $\mathbf{Is}(\mathcal{T}^\star)$ (see A1), up to extraction of a subsequence it converges to a random limit $\zeta^\infty \in \mathbf{Is}(\mathcal{T}^\star)$. Hence

$$\lim_k \left\|\zeta^{k'} - \zeta^\infty\right\|_\pi = 0 \text{ a.s.}$$

In view of (4.13), we deduce

$$\lim_k \left\|\zeta^k - \zeta^\infty\right\|_\pi = \lim_k \left\|\zeta^{k'} - \zeta^\infty\right\|_\pi = 0 \quad \text{a.s.}$$

This concludes the proof of (3.10).

Proof for Case 3.3. Since by (4.12) $\liminf_k R^k = 0$ with probability one, there exists a (random) subsequence $\{k', k \geq 1\}$ such that $\lim_k R^{k'} = 0$ a.s. Since the sequence $\{u^{k'}, k \geq 0\}$ is bounded in l^2 a.s. (as $\lim_k \left\|u^k - u^\star\right\|_{l^2}$ exists a.s.) and belongs to the convex set $\mathcal{A}$ by construction, it belongs to a compact set (see Lemma B.1). Therefore we can assume (up to extraction of another subsequence) the existence of $u^\infty \in L^2_\pi$ such that $\lim_k \left\|u^{k'} - u^\infty\right\|_{l^2} = 0$ a.s. We now prove that u^∞ is a $\mathcal{T}^\star$-valued r.v. (possibly depending on the choice of $u^\star \in \mathcal{T}^\star$). Set $\zeta^\infty := \mathbf{Is}(u^\infty)$ and define

$$R^\infty := \int_\Theta \left(\zeta^\infty(\theta) - \zeta^\star(\theta)\right) \cdot h(\zeta^\infty(\theta), \theta)\, \pi(\mathrm{d}\theta).$$

Then for any $j \geq 1$,

$$\begin{aligned} R^j - R^\infty = &\int_\Theta \left(\zeta^j(\theta) - \zeta^\infty(\theta)\right) \cdot h\left(\zeta^j(\theta), \theta\right) \pi(\mathrm{d}\theta) \\ &+ \int_\Theta \left(\zeta^\infty(\theta) - \zeta^\star(\theta)\right) \cdot \left(h\left(\zeta^j(\theta), \theta\right) - h\left(\zeta^\infty(\theta), \theta\right)\right) \pi(\mathrm{d}\theta). \end{aligned}$$

By either A3-b) or A3-a) (depending on whether $\mathcal{A}$ is bounded or not) and since $\sup_k \left\|u^k\right\|_{l^2} < \infty$ a.s., we have $\sup_k \left\|h\left(\zeta^{k'}(\cdot), \cdot\right)\right\|_\pi^2 < \infty$ a.s. Since

$$\lim_k \left\|\zeta^{k'} - \zeta^\infty\right\|_\pi = \lim_k \left\|u^{k'} - u^\infty\right\|_{l^2} = 0, \quad \text{a.s.},$$

it follows that

$$\lim_k \int_\Theta \left(\zeta^{k'}(\theta) - \zeta^\infty(\theta)\right) \cdot h\left(\zeta^{k'}(\theta), \theta\right) \pi(\mathrm{d}\theta) = 0, \quad \text{a.s.}$$

Furthermore, since, by A4, $\zeta \mapsto h(\zeta(\cdot), \cdot)$ is continuous in L^2_π, we have

$$\lim_k \int_\Theta \left(\zeta^\infty(\theta) - \zeta^\star(\theta)\right) \cdot \left(h\left(\zeta^{k'}(\theta), \theta\right) - h\left(\zeta^\infty(\theta), \theta\right)\right) \pi(\mathrm{d}\theta) = 0 \text{ a.s.}$$

Hence $0 = \lim_k R^{k'} = R^\infty$ a.s. In view of the definition of R^∞ and of A5, we deduce that $u^\infty \in \mathcal{T}^\star$ a.s. In view of (4.13), this implies that $\lim_k \left\|\zeta^k - \zeta^\infty\right\|_\pi = \lim_k \left\|\zeta^{k'} - \zeta^\infty\right\|_\pi = 0$.

4.3. Proof of the L^2-Control (3.9) and of the L^p-Convergence in (3.10). The L^2-control

$$\sup_{k\geq 0}\mathbb{E}\left[\left\|\zeta^k-\zeta^\infty\right\|_\pi^2\right]<+\infty$$

follows directly from (4.2) and the boundedness of $\mathcal{T}^\star$ (see A1). This proves (3.9). Let $C>0$ and $p\in(0,2)$. We write

$$\mathbb{E}\left[\left\|\zeta^k-\zeta^\infty\right\|_\pi^p\right]=\mathbb{E}\left[\left\|\zeta^k-\zeta^\infty\right\|_\pi^p\ 1_{\{\|\zeta^k-\zeta^\infty\|_\pi>C\}}\right]+\mathbb{E}\left[\left\|\zeta^k-\zeta^\infty\right\|_\pi^p\ 1_{\{\|\zeta^k-\zeta^\infty\|_\pi\leq C\}}\right].$$

The first term on the right hand side converges to 0 as $C\to+\infty$, uniformly in k: indeed, we have

$$\mathbb{E}\left[\left\|\zeta^k-\zeta^\infty\right\|_\pi^p 1_{\{\|\zeta^k-\zeta^\infty\|_\pi>C\}}\right]\leq\frac{\sup_{l\geq 0}\mathbb{E}\left[\left\|\zeta^l-\zeta^\infty\right\|_\pi^2\right]}{C^{2-p}}.$$

For any fixed $C>0$, the second term converges to zero by the dominated convergence theorem. This concludes the proof of Theorem 1.

5. Numerical Investigations. This section is devoted to the numerical analysis of the convergence of the UQSA algorithm. The parameterization of the algorithm is discussed and the sensitivity of its performance w.r.t. the design parameters is tested empirically.

Notably, the possibility of letting the number m_k of estimated coefficients tend to infinity appears not only as a necessary ingredient for proving the theoretical convergence of the algorithm (see Theorem 1), but also as an important feature for its numerical performance, regarding, in particular, the estimation of the lower order coefficients $u_i^\star$ and the mitigation of the burn-in phase. We illustrate this assertion numerically, by testing both the genuine UQSA algorithm with increasing m_k and the fixed dimension version with $m_k=m$ (for different values of m), respectively referred to as the "increasing m_k" and the "fixed m" algorithms henceforth. The speed of the dimension growth turns out to be a determining factor of the practical convergence rate of the algorithm. A correct tuning of this speed allows achieving the right balance between the truncation error, i.e. the error due to the non-estimation of the coefficients beyond the m_k^{th} one, and the estimation error on the "active" coefficients up to m_k. Balancing these two contributions of the error seems to be the way to reach an optimal performance of the algorithm.

5.1. Design Parameterization of the UQSA Algorithm. When running the UQSA algorithm, the user has to choose some design parameters: given a problem of the form (2.10) and the corresponding sequence $\{q_m, m\geq 0\}$ via (3.2), the user has to choose the orthogonal basis $\{B_i(\theta), i\geq 0\}$, which fixes in turn the sequence $\{Q_m, m\geq 0\}$. It remains to choose $\{\gamma_k, k\geq 1\}$, $\{m_k, \geq 0\}$ and $\{M_k, k\geq 1\}$. In this section, we consider sequences of the form

$$\gamma_k=k^{-a},\quad m_k=\lfloor k^b\rfloor+1,\quad M_k=\lfloor k^p\rfloor+1, \tag{5.1}$$

for $a,p\geq 0$ and $b>0$, and we discuss how to choose these constants assuming that

$$q_m=O\left(m^{-\delta}\right),\qquad Q_m=O\left(m^{\Delta}\right), \tag{5.2}$$

for some $\delta > 0$ and $\Delta \geq 0$.

An easy calculation shows that A2 is satisfied if and only if

$$\frac{1}{2} < a \leq 1, \qquad 2 - \delta b < 2a, \qquad b\Delta + 1 < 2a + p \tag{5.3}$$

(for $\kappa = \frac{1}{a} - 1$ in A2, if $a < 1$, or $\kappa > 0$ small enough, if $a = 1$). Given $\delta > 0$ and $\Delta \geq 0$, there always exist a, b, p satisfying these conditions. Figure 1 displays the lines $x \mapsto 1$, $x \mapsto 2(1-x)/\delta$ and $x \mapsto (2x-1)/\Delta$ for different values of the pair (δ, Δ) with $\Delta > 0$. The colored area corresponds to the points (a, b) satisfying the conditions (5.3) in the case $p = 0$, i.e. in the case where the numbers of Monte Carlo draws is constant over iterations. Note that this set becomes all the more restrictive as $\delta \to 0$ and $\Delta \to \infty$. Choosing $p > 0$ gives more flexibility, but it also leads to higher computational cost (since the number of Monte Carlo simulations increases along iterations, see the discussion in Section 5.5.3).

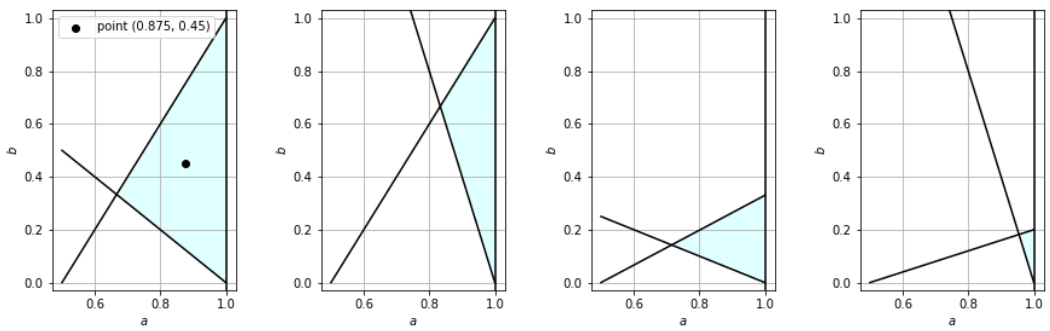

Figure 1. *For different values of (δ, Δ), in the case $p = 0$, the colored area is the admissible set of points (a, b) satisfying (5.3). From left to right: $(\delta, \Delta) = (2, 1), (0.5, 1), (4, 3)$, and $(0.5, 5)$.*

5.2. Benchmark Problem. We consider the problem (2.10) in the case where

$$\Theta = [-\pi, \pi], \qquad \pi(\mathrm{d}\theta) = \frac{1}{2\pi}\mathbf{1}_{[-\pi,\pi]}\,\mathrm{d}\theta. \tag{5.4}$$

We perform tests for two different models of function H:

1. $H_1(z, v, \theta) := (z - \zeta^\star(\theta))\left(1 + \frac{\cos(v)}{2}\cos(z - \zeta^\star(\theta))\right) + v,$
2. $H_2(z, v, \theta) := (z - \zeta^\star(\theta))\left(1 + \frac{\cos(v)}{2}\sin(z - \zeta^\star(\theta))\right),$

for a common function $\zeta^\star : \Theta \to \mathbb{R}$ given by

$$\zeta^\star(\theta) := \left|\frac{4}{5} + \frac{1}{4}\exp(\sin(\theta)) - \cosh(\sin(\theta)^2)\right|(1 + \sin(2\theta)), \tag{5.5}$$

and where, for any $\theta \in \Theta$, the conditional distribution $\mu(\theta, \mathrm{d}v)$ is a centered Gaussian law with variance θ^2. In both cases we have $q = 1$ and $\mathbf{Is}(\mathcal{T}^\star) = \{\zeta^\star\}$. It is easily checked that for any $z \in \mathbb{R}$, $\theta \in \Theta$ and $i = 1, 2$ we have

$$\int_{\mathcal{V}} |H_i(z, v, \theta)|^2 \mu(\theta, \mathrm{d}v) \leq 8|z - \zeta^\star(\theta)|^2 + 2\theta^2, \qquad (z - \zeta^\star(\theta)) \cdot h_i(z, \theta) \geq \frac{1}{2}(z - \zeta^\star(\theta))^2,$$

where h_i is derived from H_i following (1.2). Hence, the assumptions A3, A4, A5 and A6 are satisfied for both models.

The two models for H above correspond to two possible behaviors of the martingale increment sequence $\{\eta^k, k \geq 1\}$ (cf. Section 4.1): $\mathbb{E}\left[\left\|\eta^k\right\|_{l^2}\right]$ bounded away from 0 (case of Model 1) or $\mathbb{E}\left[\left\|\eta^k\right\|_{l^2}\right] \to 0$ (case of Model 2). While the first case is more general, the second one may also appear in practice and leads to quite different behavior of the UQSA algorithm, requiring a different tuning of the parameters.

The choice of the function $\zeta^\star$ (see the plot on Figure 3) is motivated by its relatively slow truncation error decay. Since $\zeta^\star$ extended by periodicity outside $[-\pi, \pi]$ is piecewise continuously differentiable, its truncation error satisfies (5.2) with $\delta = 2$ (see Lemma A.1). Numerically, one can check that the practical rate of convergence lies somewhere between 2 and 3, i.e. the theoretical value $\delta = 2$ above is reasonably sharp (meaning that our example $\zeta^\star$ is close to a "real" $\delta = 2$ example and not much "easier", which also motivated our choice of this particular function $\zeta^\star$). In real-life applications, the target function $\zeta^\star$ is bound to be less challenging than the present one, e.g. monotone and/or convex/concave w.r.t. θ or some of its components (for instance in the context of financial applications, see e.g. [4]). Moreover, the user may be interested with a few coefficients $u_i^\star$ only, whereas we show numerical results up to $m_K = 250$ below.

The choice of $\mathcal{N}(0, \theta^2)$ for the kernel $\mu(\theta, \mathrm{d}v)$ is purely illustrative. This distribution could be replaced by any other one (simulatable i.i.d.) without expectable impact regarding the qualitative conclusions drawn from the numerical experiments below.

Finally, for the orthonormal basis $\{B_i, i \geq 0\}$, we choose the normalized trigonometric basis on $\Theta = [-\pi, \pi]$ (cf. [7, Section 2.1]). Therefore, we have $\sup_{i\geq 0} \sup_\Theta |B_i(\theta)| < +\infty$, so that

$$Q_m = O(m),$$

i.e. $\Delta = 1$ in (5.1).

5.3. Performance Criteria. In the numerical experiments that follow, we compare the performances of the algorithms with increasing m_k and fixed m, for different choices of (a, b, p). The comparison relies on root-mean-square errors, approximated by the corresponding sample means over 50 independent runs of the algorithms. After K iterations, the square of the total error $\mathcal{E}^2$ is decomposed into the mean squared SA error $\mathcal{E}_{sa}^2$, which is the error restricted to the $(m_K + 1)$ estimated coefficients of $\zeta^\star$, and the squared truncation error $\mathcal{E}_{tr}^2$, i.e. $\mathcal{E}^2 = \mathcal{E}_{sa}^2 + \mathcal{E}_{tr}^2$ where

$$\mathcal{E}^2 = \mathbb{E}\left[\left\|u^K - u^\star\right\|_{l^2}^2\right], \quad \mathcal{E}_{sa}^2 = \mathbb{E}\left[\sum_{i=0}^{m_K} (u_i^K - u_i^\star)^2\right], \quad \text{and} \quad \mathcal{E}_{tr}^2 = \sum_{i=m_K+1}^{+\infty} (u_i^\star)^2$$

(recalling $u_i^K = 0$ for $i > m_K$). The benchmark values for the coefficients $u_i^\star$ are pre-calculated by high-accuracy numerical integration. With the exception of Figures 3 and 5[left], all our graphs are error plots in log-log scale.

5.4. Impact of the Increasing Dimension. In this section, we discuss the role of the sequence $\{m_k, k \geq 0\}$. Since $(\delta, \Delta) = (2, 1)$, the set of admissible pairs (a, b) for our benchmark problem is given by the leftmost graph of Figure 1.

We take $a = 0.875$, which is in the middle of the corresponding admissible interval. For Model 1 (i.e. $H = H_1$), a heuristic may be applied for a clever choice of b. In finite dimensional SA schemes, the mean squared error after k iterations is typically of the order of $\gamma_k = k^{-a}$ (see [10, Chapter 2]). For the UQSA algorithm, we may expect (in the case $p = 0$) a growth of the variance proportional to the dimension $m_k \approx k^b$. This suggests a heuristic guess for the SA-error order given by $k^{-a} \times k^b = 1/k^{a-b}$. Now, by (5.2) (with in our case $\delta = 2$), the truncation error is of order $k^{-b\delta}$. Hence, to optimize the convergence rate, we take b such that $b\delta = a - b$. This approximately corresponds to $b = 0.3$, which is the value that we use in our tests for Model 1. In the case of Model 2 (i.e. $H = H_2$), this rule of thumb does not apply, because the variance of the martingale increments goes to 0. In this case we simply take $b = 0.45$, so that $(a, b) = (0.875, 0.45)$ lies in the middle of the admissible set (see Figure 1[left]). Also note that the range $[0.25, 0.5]$ for b is reasonable in view of the total number K of iterations that we commonly use in the algorithm and of the number m_K of coefficients of interest.

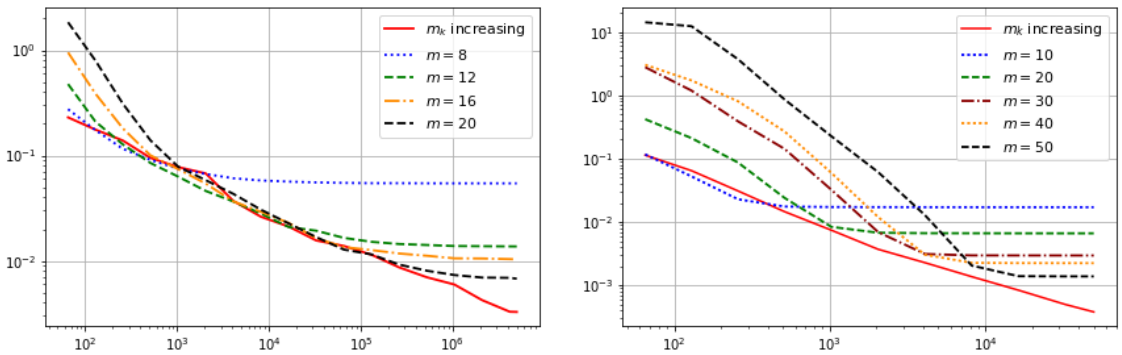

Figure 2. *The total error $\mathcal{E}$ as a function of the number of iterations, for different choices of the sequence $\{m_k, k \geq 0\}$: m_k increasing (solid line) and: [left] for Model 1 with $m_k = m = 8, 12, 16, 20$ (other lines); [right] for Model 2 with $m_k = m = 10, 20, 30, 40, 50$ (other lines).*

Figure 2 displays the total error $\mathcal{E}$ for different strategies on the sequence $\{m_k, k \geq 0\}$: the solid line is the case $m_k = \lfloor k^b \rfloor + 1$ (with $b = 0.3$ and 0.45 for Models 1 and 2 respectively), while the other lines correspond to the cases $m_k = m = 8, 12, 16, 20$ for Model 1 and $m_k = m = 10, 20, 30, 40, 50$ for Model 2 (larger values of m are used here since the convergence is expected to be faster for Model 2). The performance of the algorithm with increasing m_k is similar or better throughout the whole iteration path. This holds true in the burn-in phase, which is typically related to disproportion between the first values of γ_k and the magnitude of the solution (estimated coefficients). In fact, with increasing m_k, the dimension gradually grows with k, with larger values of γ_k naturally associated with the estimation of the first (and larger, for classical basis) coefficients, whereas, when $m_k = m$ is constant, the higher order "small" coefficients are involved from the very beginning along with the larger values of γ_k, leading to a longer burn-in phase. It is also true on the convergence part, where the algorithms with fixed m only converge up to a certain accuracy depending on the value of m. The better performance of the increasing m_k version is more explicit for Model 2; while for

Model 1, fixed m versions have similar performance within certain ranges of values of k.

In practice, we do not know in advance the length of the burn-in phase or the magnitude of the truncation error for various m. Hence, the genuine UQSA algorithm with increasing m_k, which ensures convergence without the need for additional knowledge, is preferable.

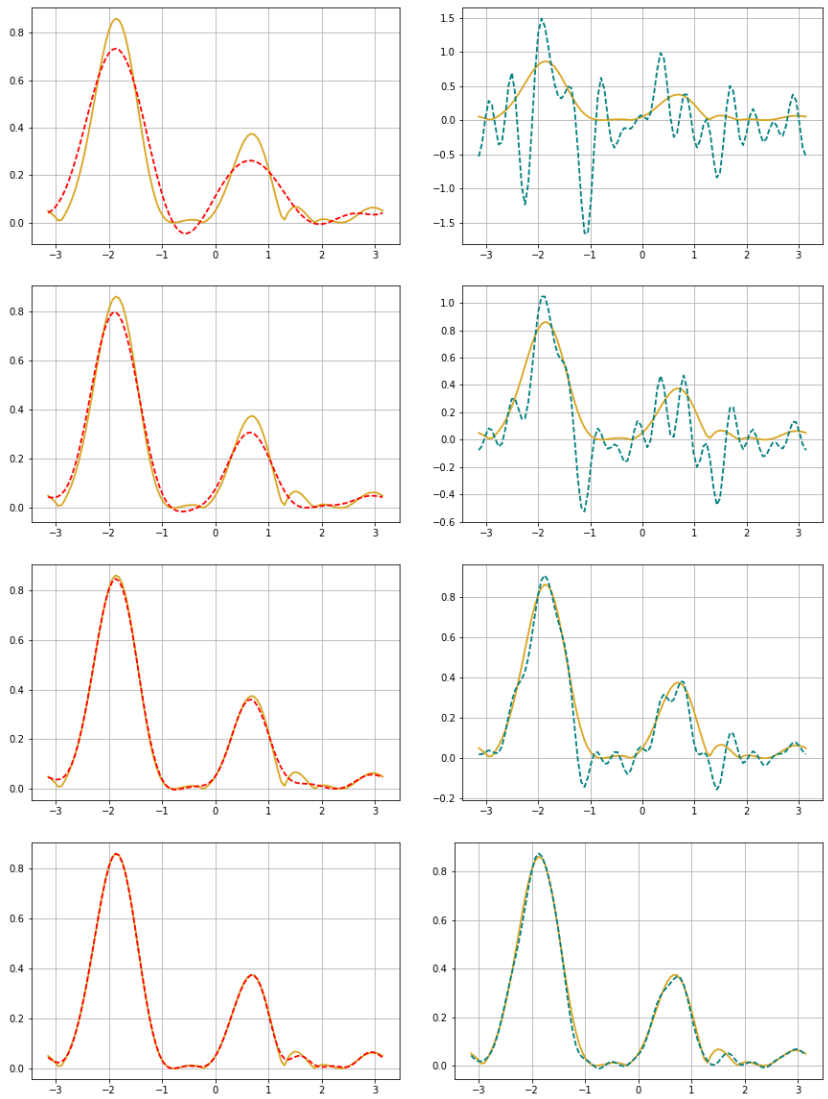

Figure 3. *The functions $\zeta^\star$ and ζ^K are displayed in respective solid line and dashed lines, as a function of $\theta \in [-\pi, \pi]$. On the left, $\{m_k, k \geq 0\}$ is increasing and on the right, it is constant and equal to $m = 30$. From top to bottom, $K \in \{128, 256, 512, 1024\}$.*

Let us now analyze the weak burn-in phase performance of the fixed m version for Model 2 (cf. Figure 2[right]). Figure 3 displays the result of a single run of the UQSA algorithm in this case. In dashed line, the function $\theta \mapsto \zeta^K(\theta)$ is displayed for $\theta \in [-\pi, \pi]$. For comparison, the function $\theta \mapsto \zeta^\star(\theta)$ is displayed in solid line. To illustrate the advantage of the UQSA

algorithm with increasing m_k in the burn-in phase, we show the estimated function ζ^K for different values of K (from top to bottom, $K \in \{128, 256, 512, 1024\}$) and for m_k increasing (left panels) versus $m_k = m = 30$ for any k (right panels). The increasing dimension m_k leads to a smoother convergence, with intermediate iterations looking closer to a projection of $\zeta^\star$ on the subspace spanned by a smaller number of basis functions. We conclude that, in the case of Model 2 where the variance of the martingale increment is small or goes to 0, the increasing m_k feature plays a key role in the UQSA algorithm performance.

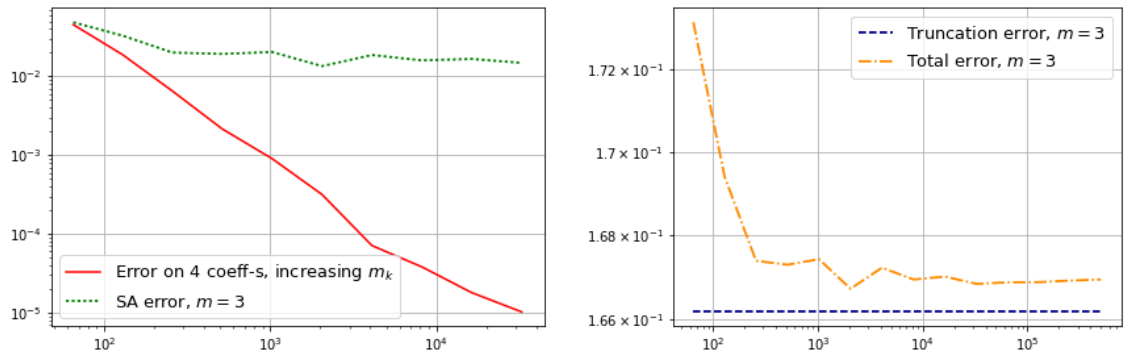

Figure 4. *Model 2: [left] in the case $m_k \to \infty$ (solid line) and $m_k = m = 3$ (dotted line), the SA error $\left(\mathbb{E}\sum_{i=0}^{3}(u_i^K - u_i^\star)^2\right)^{1/2}$ as a function of the number of iterations K; [right] in the case $m_k = m = 3$, the truncation error $\mathcal{E}_{tr}$ (dashed line) and the total error $\mathcal{E}$ (dash-dot line) displayed as a function of K.*

In Figure 4, we show that increasing m_k is also key for an accurate determination of the lower order coefficients (e.g. in the case where only the first few coefficients of the expansion of $\zeta^\star$ are of interest to the user). In fact, as already mentioned in Section 2.3, the algorithm with fixed m does typically not converge to the first $(m+1)$ coefficients of the decomposition of $\zeta^\star$. In Figure 4[left], the L^2-error on the first 4 coefficients is displayed as a function of the number of iterations K, for two strategies on m_k (case of Model 2): the solid line is the case $m_k = O(k^b)$ with $b = 0.45$ and the dotted line is the case $m = 3$. In Figure 4[right], the total error $\mathcal{E}$ and the truncation error $\mathcal{E}_{tr}$ are displayed, resp. in dash-dot line and dashed line in the case m_k is the constant sequence equal to $m = 3$. These figures show that, when $m_k \to +\infty$, UQSA converges (which is the claim of Theorem 1), whereas, when $m_k = m$ for any k, it does not: the total error does not reach the truncation error since there is a non vanishing bias on the estimation of the first $(m+1)$ coefficients (the SA-error $\mathcal{E}_{sa}$ does not vanish when $K \to +\infty$).

For Model 1, similar effects (not reported here) are visible, even if a bit less obvious due to the slower convergence of both versions (fixed and increasing dimension) of the algorithm in this case.

5.5. Impact of the Design Parameters for the Increasing m_k UQSA Algorithm. In this section, we discuss the impact of the choice of a, b, and p on the performance on the UQSA algorithm.

5.5.1. Impact of the increase rate b of $m_k \to +\infty$. In this paragraph, we set $a = 0.875$, $p = 0$, and we test different values for b. The range of admissible values of b is $(0.125, 0.75)$. We take $b \in \{0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$ for Model 1 and $b \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$ for Model 2.

Figure 5 displays the evolution of the total error $\mathcal{E}$ as a function of the number of iterations K for different values of b, and for both models of H. For Model 1, the variance increases with the dimension, which makes the SA error larger as b increases. At the same time, the truncation error decreases at the rate $b\delta$, so that for too small values of b, the truncation error dominates the SA error. Hence there is a trade-off between the two errors, with optimal values of b somewhere in the range. This phenomenon is observed on Figure 5[left, center]. For $b = 0.2, 0.25, 0.3$, the total error is dominated by the truncation error, while from $b = 0.35$ the error is dominated by the SA error; as b increases further, the convergence becomes slower due to additional variance which increases the SA error. For Model 2, since the variance of the martingale increments goes to 0, the effect of additional variance due to a larger dimension is not visible. We observe that larger values of b implies a better convergence, up to $b = 0.70$. However, as we may see, the gain in the rate of convergence from taking larger b decreases as it tends to the boundary of the admissible interval. In addition, this analysis in terms of the number of iterations K does not take into account the higher computational cost due to a dimension growing faster and each iteration becoming longer when b is larger. For example, for $b = 0.70$, we made only $K = 2500$ iterations, because the computational effort becomes too large beyond this. To conclude, we suggest that in this case optimal values of b (for given a) in terms of both convergence and computational cost lie near the middle of the admissible interval.

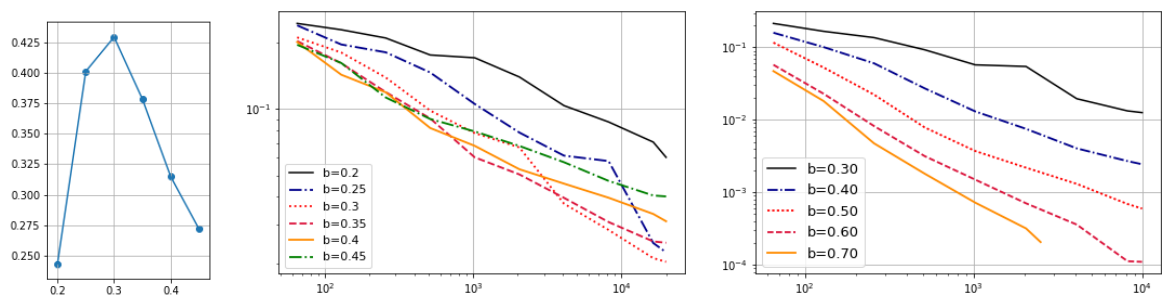

Figure 5. *[left] Empirical L^2 convergence rate for $b \in \{0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$, Model 1. Total error $\mathcal{E}$ as a function of the number of iterations K, for different values of b; [middle] Model 1, $b \in \{0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}$; [right] Model 2, $b \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$.*

5.5.2. the decrease rate a of the learning rate γ_k. In this paragraph, again for $p = 0$, taking $b = 0.3$ for Model 1 and $b = 0.45$ for Model 2 as in Section 5.4, we compare different values of a.

Figure 6 displays the total error $\mathcal{E}$ as a function of the number of iterations K for different values of a for both models. For Model 1, we see that the convergence rate is better for larger values of a. This is in line with classical results for finite dimensional stochastic approximation, whereby the mean square error is of order $\gamma_k^{1/2} \sim k^{-a/2}$ (see [10, Chapter 2]). For Model 2,

varying a does not produce much effect (except for a slight decline as a approaches 1), because the step-size controls the variance of the corresponding martingale noise, but in the case of Model 2, the variance of the martingale increments goes to 0 anyway.

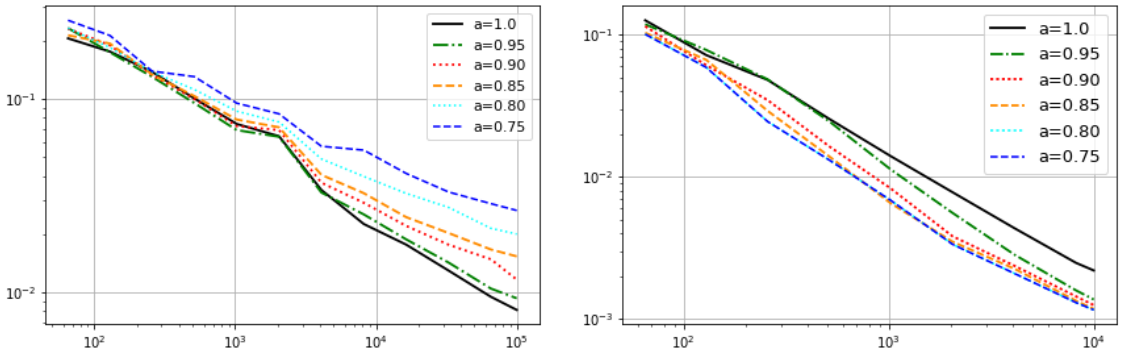

Figure 6. *The total error $\mathcal{E}$ as a function of the number of iterations, for different values of a in $\{0.75, 0.80, 0.85, 0.9, 0.95, 1.0\}$: [left] Model 1 and [right] Model 2.*

5.5.3. Role of p. In this section we consider the case $p > 0$, i.e. the number of Monte Carlo samples at each iteration increases along the UQSA iterations. One may check that all the triples of the parameters (a, b, p) used below lie in the admissible set (cf. (5.3)).

In the analysis that follows, we want to keep track of the dependence of the error w.r.t. a computational cost proxied by the total number of Monte Carlo draws of the pair (θ, v), i.e., after K iterations, $\sum_{k=0}^{K-1} M_k \approx O(K^{p+1})$. As we want to estimate the same number of coefficients for a given computational budget, we take $b = \bar{b}(p+1)$, with $\bar{b} = 0.3$ for Model 1 and $\bar{b} = 0.45$ for Model 2.

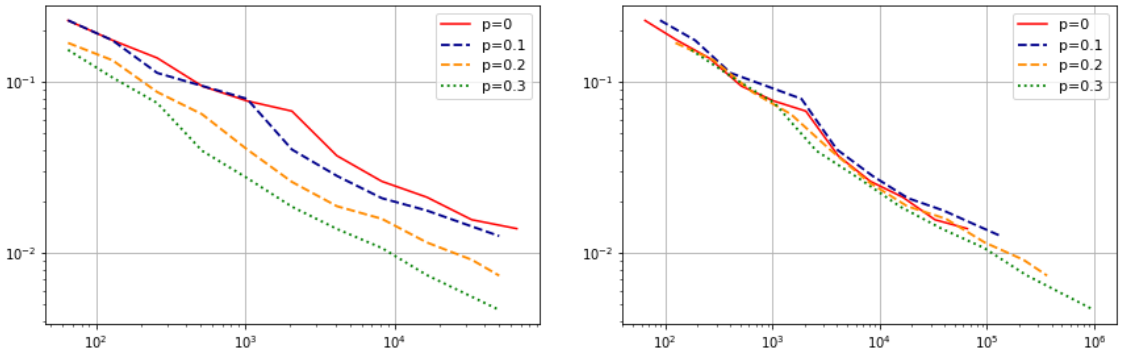

Figure 7. *Model 1, total error $\mathcal{E}$ of the UQSA algorithm for different values of $p \in \{0, 0.1, 0.2, 0.3\}$ as a function of the number of iterations [left] and of the total number of Monte Carlo draws [right]. Here $a = 0.875$ and $b = 0.3(p+1)$.*

We set $a = 0.875$ (as in Sections 5.4 and 5.5.1). Figure 7 displays for Model 1 the total error $\mathcal{E}$ as a function of the number of iterations (left) and as a function of the total number

of Monte Carlo draws (right) for triples of the form $(a, \bar{b}(p+1), p)$ with various p. The results show that, even though larger p yield a better convergence in terms of the number of iterations K, there is no much difference when the computational cost is taken into account (i.e. in terms of the number of Monte Carlo draws). The results are very similar for Model 2 and therefore not reported here.

5.6. Conclusion. To summarize, we observed from our experiments, that the increasing dimension feature of the UQSA algorithm is essential for the asymptotic convergence, as well as for the lower order coefficients estimation; it also yields a milder burn-in phase. As expected, the convergence is generally faster for Model 2 due to reduced variance effect. In the more general setting of Model 1, the parameter b plays a crucial role in the performance of the algorithm and a good rule of thumb (in the case $p = 0$) is $b = (a - b)/\delta$. However, in the special case where the variance of the martingale increments goes to 0, one should take larger values of b. Convergence rates should be naturally bounded by $a/2$ for Model 1 (which is the corresponding result for finite dimensional SA), while for Model 2 the convergence should be faster. Taking $p > 0$ may potentially be useful for verifying the assumptions of Theorem 1 in some cases, but it makes no real difference in terms of convergence speed when the latter is assessed w.r.t. the total number of simulations.

Appendix A. Truncation error for trigonometric basis.

Lemma A.1. *Let $\zeta : \mathbb{R} \to \mathbb{R}$ be 2π-periodic and piecewise continuously differentiable. Let $\{u_i, i \geq 0\}$ be the coefficients of its decomposition w.r.t. the normalized trigonometric basis (cf. [7, Section 2.1]). Then, for some $C > 0$,*

$$\sum_{i=m+1}^{+\infty} |u_i|^2 \leq Cm^{-2}.$$

Proof. Consider the Fourier decomposition of the function ζ' on $[-\pi, \pi]$:

$$\zeta'(x) = v_0 + \sum_{m\geq 1} (v_{2m-1} \sin(mx) + v_{2m} \cos(mx)).$$

Using that ζ is 2π-periodic (i.e. $\zeta(-\pi) = \zeta(\pi)$), it is not hard to deduce via integration by parts that, for any $m \geq 1$, $u_{2m-1} = v_{2m}/m$, and $u_{2m} = -v_{2m-1}/m$. Hence,

$$\sum_{i=2m-1}^{+\infty} |u_i|^2 \leq \frac{1}{m^2} \sum_{i=2m-1}^{+\infty} |v_i|^2 \leq \frac{\|\zeta'\|_\pi^2}{m^2},$$

which implies the result. ∎

Appendix B. Compact sets in l^2.

Let $\{a_n, n \geq 0\}$ be a positive sequence such that $\sum_{i\geq 0} a_i^2 < \infty$ and $\{d_n, n \geq 0\}$ be a strictly increasing sequence of non-negative integers. Define

$$\mathcal{A} := \left\{ u \in l^2 : \sum_{d_n \leq i < d_{n+1}} |u_i|^2 \leq a_n^2 \quad \forall n \geq 0 \right\} \tag{B.1}$$

Note that $\mathcal{A}$ is closed and convex.

Lemma B.1. 1. *Assume $d_0 = 0$. Then $\mathcal{A}$ is compact.*

2. *For any constant $B > 0$, the set $\{u \in \mathcal{A} : \|u\|_{l^2} < B\}$ is convex and compact.*

Proof. Proof of (1). First note that since $d_0 = 0$ and $\sum_n a_n^2 < \infty$, then $\mathcal{A}$ is bounded. Since $\mathcal{A}$ is closed, we prove that $\mathcal{A}$ is relatively compact, and to that goal, we check the sufficient condition given in [18, Theorem 3]: since l^2 is spanned by a complete orthonormal basis and $\mathcal{A}$ is bounded, $\mathcal{A}$ is relatively compact if and only if for any $\epsilon > 0$, there exists N_ϵ such that for any $u \in \mathcal{A}$, $\sum_{i \geq N_\epsilon} |u_i|^2 \leq \epsilon$. For any N and $d_{\ell(N)}$ such that $d_{\ell(N)} \leq N$, and for any $u \in \mathcal{A}$, it holds

$$\sum_{i \geq N} |u_i|^2 \leq \sum_{i \geq d_{\ell(N)}} |u_i|^2 \leq \sum_{j \geq \ell(N)} a_j^2. \tag{B.2}$$

When $N \to \infty$, we can choose $\ell(N)$ so that $\ell(N) \to \infty$ (since the sequence $\{d_n, n \geq 0\}$ is strictly increasing). Therefore, since $\sum_n a_n^2 < \infty$, the RHS in (B.2) tends to zero as $N \to \infty$.

Proof of (2). The result follows directly from the previous item, using that $\sum_{i<d_0} |u_i|^2 \leq B^2$ for any $u \in \mathcal{A}$ which implies (without assuming that $d_0 = 0$) that the set is bounded. ■

REFERENCES

[1] J. An and A. Owen, *Quasi-regression*, Journal of Complexity, 17 (2001), pp. 588–607.

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

[3] O. Bardou, N. Frikha, and G. Pages, *Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling*, Monte Carlo Methods and Applications, 15 (2009), pp. 173–210.

[4] D. Barrera, S. Crépey, B. Diallo, G. Fort, E. Gobet, and U. Stazhynski, *Stochastic approximation schemes for economic capital and risk margin computations*, ESAIM: Proceedings and Surveys, (2018). Forthcoming (preprint available at https://hal-univ-tlse2.archives-ouvertes.fr/IMT/hal-01710394v1).

[5] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*, vol. 22 of Applications of Mathematics, Springer-Verlag, 1990.

[6] N. Berman and A. Shwartz, *Abstract stochastic approximations and applications*, Stochastic Processes and their Applications, 31 (1989), pp. 133–149.

[7] C. Canuto, M. Hussaini, A. Quarteroni, and T. Zang, *Spectral Methods: Fundamentals in Single Domains*, Springer Verlag, 2006.

[8] X. Chen and H. White, *Asymptotic properties of some projection-based Robbins-Monro procedures in a Hilbert space*, Studies in Nonlinear Dynamics and Econometrics, 6 (2002), pp. 1–55.

[9] A. Dieuleveut and F. Bach, *Nonparametric stochastic approximation with large step-sizes*, The Annals of Statistics, 44 (2016), pp. 1363–1399.

[10] M. Duflo, *Random Iterative Models*, vol. 34 of Applications of Mathematics, Springer, 1997.

[11] D. Funaro, *Polynomial Approximation of Differential Equations*, vol. 8 of Lecture Notes in Physics Monographs, Springer-Verlag Heidelberg, 1992.

[12] R. Ghanem and P. Spanos, *Stochastic finite elements: a spectral approach*, Courier Corporation, 2003.

[13] E. Gobet and K. Surana, *A new sequential algorithm for L_2-approximation and application to Monte-Carlo integration*, Preprint available on https://hal.archives-ouvertes.fr/hal-00972016, (2014).

[14] L. Goldstein, *Minimizing noisy functionals in Hilbert space: An extension of the Kiefer-Wolfowitz procedure*, Journal of Theoretical Probability, 1 (1988), pp. 189–204.

[15] J. Kiefer and J. Wolfowitz, *Stochastic estimation of the maximum of a regression function*, Annals of Mathematical Statistics, 23 (1952), pp. 462–466.

[16] M. Kleiber and T. Hien, *The Stochastic Finite Element Method: basic perturbation technique and computer implementation*, John Wiley & Sons Ltd, 1992.
[17] R. Koenker, *Quantile Regression*, Econometric Society Monographs, Cambridge University Press, 2005.
[18] A. Kulkarni and V. Borkar, *Finite dimensional approximation and Newton-based algorithm for stochastic approximation in Hilbert space*, Automatica, 45 (2009), pp. 2815–2822.
[19] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, vol. 35 of Application of Mathematics, Springer, 1997.
[20] O. Le Maitre and O. Knio, *Spectral Methods for Uncertainty Quantification. With Applications to Computational Fluid Dynamics, Scientific Computation*, Springer Science & Business Media, 2010.
[21] W. Liu, T. Belytschko, and A. Mani, *Random field finite elements*, International journal for numerical methods in engineering, 23 (1986), pp. 1831–1845.
[22] J. Muscat, *Functional Analysis. An Introduction to Metric Spaces, Hilbert Spaces, and Banach Algebras*, Springer, 2014.
[23] P. Nevai, *Géza Freud, orthogonal polynomials and Christoffel functions. a case study*, Journal of Approximation Theory, 48 (1986), pp. 3–167.
[24] R. Nixdorf, *An invariance principle for a finite dimensional stochastic approximation method in a Hilbert space*, Journal of Multivariate Analysis, 15 (1984), pp. 252–260.
[25] B. Polyak and A. Juditsky, *Acceleration of stochastic approximation by averaging*, SIAM J. Control Optim., 30 (1992), pp. 838–855.
[26] H. Robbins and S. Monro, *A stochastic approximation method*, Annals of Mathematical Statistics, 22 (1951), pp. 400–407.
[27] H. Robbins and D. Siegmund, *A convergence theorem for nonnegative almost supermartingales and some applications*, in Optimizing Methods in Statistics, J. Rustagi, ed., Academic Press, New-York, 1971, pp. 233–257.
[28] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, USA, 2001.
[29] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on stochastic programming: modeling and theory*, SIAM, Philadelphia, 2009.
[30] R. Smith, *Uncertainty quantification. Theory, implementation, and applications*, vol. 12 of Computational Science & Engineering, Springer, Philadelphia, PA, 2014.
[31] U. Stazhynski, *Discretization of processes at stopping times and Uncertainty quantification for stochastic approximation limits*, PhD thesis, Ecole Polytechnique, 2018.
[32] H. Walk, *An invariance principle for the Robbins-Monro process in a Hilbert space*, Zeitschrift fur Warscheinlichkeitstheorie und Verwandte Gebiete, 39 (1977), pp. 135–150.
[33] N. Wiener, *The homogeneous chaos*, American Journal of Mathematics, 60 (1938), pp. 897–936.
[34] G. Yin, *On H-valued stochastic approximation: Finite dimensional projections*, Stochastic Analysis and Applications, 10 (1992), pp. 363–377.
[35] G. Yin and Y. Zhu, *On H-valued Robbins-Monro processes*, Journal of Multivariate Analysis, 34 (1990), pp. 116–140.