HAL
open science

Influence of sampling on the convergence rates of greedy algorithms for parameter-dependent random variables

Mohamed-Raed Blel, Virginie Ehrlacher, Tony Lelièvre

To cite this version:

Mohamed-Raed Blel, Virginie Ehrlacher, Tony Lelièvre. Influence of sampling on the convergence rates of greedy algorithms for parameter-dependent random variables. 2021. hal-03238244v2

HAL Id: hal-03238244

https://hal.science/hal-03238244v2

Preprint submitted on 23 Sep 2021

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.

Influence of sampling on the convergence rates of greedy algorithms for parameter-dependent random variables

Mohamed-Raed Blel, Virginie Ehrlacher, Tony Lelièvre

September 23, 2021

Abstract

The main focus of this article is to provide a mathematical study of the algorithm proposed in [6] where the authors proposed a variance reduction technique for the computation of parameter-dependent expectations using a reduced basis paradigm. We study the effect of Monte-Carlo sampling on the theoretical properties of greedy algorithms. In particular, using concentration inequalities for the empirical measure in Wasserstein distance proved in [14], we provide sufficient conditions on the number of samples used for the computation of empirical variances at each iteration of the greedy procedure to guarantee that the resulting method algorithm is a weak greedy algorithm with high probability. These theoretical results are not fully practical and we therefore propose a heuristic procedure to choose the number of Monte-Carlo samples at each iteration, inspired from this theoretical study, which provides satisfactory results on several numerical test cases.

1 Introduction

The aim of this article is to provide a mathematical study of the algorithm proposed in [6] where the authors proposed a variance reduction technique for the computation of parameter-dependent expectations using a reduced basis paradigm.

More precisely, the problematic we are considering here is the following: let us denote by $\mathcal{P} \subset \mathbb{R}^m$ a set of parameter values. In several applications, it is of significant interest to be able to rapidly compute the expectation of a random variable of the form $f_\mu(Z)$ for a large numbers of values of the parameter $\mu \in \mathcal{P}$, where Z is a random vector and where for all $\mu \in \mathcal{P}$, f_μ is a real-valued function. In practice, such expectations may not be computable analytically and are approximated using empirical means involving a large number of random samples of the random vector Z. Variance reduction methods are commonly used in such contexts in order to reduce the computational cost of approximating these expectations by means of standard Monte-Carlo algorithms. Among these, control variates, which are chosen as approximations of the random variable $f_\mu(Z)$ the expectation of which can be easily computed, can yield to interesting gains in terms of computational cost, provided that the variance of the difference between $f_\mu(Z)$ and its approximation is small. The construction of efficient control variates for a given application is thus fundamental for the variance reduction technique to yield significant computational gains.

In [6], the authors proposed a general algorithm in order to construct a control variate for $f_\mu(Z)$ using a reduced basis paradim. More precisely, the approximation of $f_\mu(Z)$ is constructed as a linear combination of $f_{\mu_1}(Z), \cdots, f_{\mu_n}(Z)$ for some small integer $n \in \mathbb{N}^*$ and well-chosen values $\mu_1, \cdots, \mu_n \in \mathcal{P}$ of the parameters. The choice of n and of the values of the parameters stems from an iterative procedure, called a greedy algorithm, which consists at iteration $n \in \mathbb{N}$ to compute

$$\mu_{n+1} \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \inf_{Z_n \in V_n} \mathrm{Var}\left[f_\mu(Z) - Z_n\right],$$

where $V_n := \mathrm{Span}\left\{f_{\mu_1}(Z), \cdots, f_{\mu_n}(Z)\right\}$. In the ideal (unpractical) case where variances can be exactly computed, the procedure boils down to a standard greedy algorithm in a Hilbert space [10]. It is now well-known [10] that such a greedy procedure provides a quasi-optimal set of parameters $\mu_1, \cdots, \mu_n$ in the sense that the error

$$\sup_{\mu \in \mathcal{P}} \inf_{Z_n \in V_n} \mathrm{Var}\left[f_\mu(Z) - Z_n\right] = \inf_{Z_n \in V_n} \mathrm{Var}\left[f_{\mu_{n+1}}(Z) - Z_n\right]$$

is comparable to the so-called Kolmogorov n-width of the set $\{f_\mu(Z), \mu \in \mathcal{P}\}$, defined by

$$\sup_{\mu \in \mathcal{P}} \inf_{\substack{W_n \text{ vectorial subspace} \\ \dim W_n = n}} \inf_{Z_n \in W_n} \mathrm{Var}\left[f_\mu(Z) - Z_n\right].$$

In other words, the subspace V_n is a quasi-optimal subspace of dimension n for the approximation of random variables $f_\mu(Z)$ for $\mu \in \mathcal{P}$ in an L^2 norm sense.

However, in practice, variances cannot be computed exactly and have to be approximated by empirical means involving a finite number of samples of the random vector Z, which may be different from one iteration of the greedy algorithm to another. The main result of this article is to give theoretical lower bounds on the number of samples which have to be taken at each iteration of the greedy algorithm in order to guarantee that the resulting Monte-Carlo greedy algorithm enjoys quasi-optimality properties close to those of an ideal greedy algorithm with high probability.

The mathematical analysis of algorithms which combine randomness and greedy procedures is a quite recent and active field of research among the model-order reduction community. Let us mention here a few works in this direction in which different settings than the one we focus on here are considered. In [9], the authors consider the effect of randomly sampling the set of parameters in order to define random trial sets at each iteration of the greedy algorithm and prove that the obtained procedure enjoys remarkable approximation properties which remain very close to the approximation properties of a greedy algorithm where minimization problems at each iteration are defined over the whole set of parameters. In [18, 17, 1, 2], the authors propose randomized residual-based error estimators for parametrized equations, with a view to using them for the acceleration of greedy algorithms for reduced basis techniques. Let us finally mention that significant research efforts are devoted by many different groups to the improvement of randomized algorithms for Singular Value Decompositions [8], which plays a fundamental role for model-order reduction.

The outline of the article is the following. In Section 2, we motivate the interest of greedy algorithms for the construction of control variates for variance reduction methods and recall some results of [7, 4, 10] on the mathematical analysis of greedy algorithms in Hilbert spaces. In Section 3, we present the Monte-Carlo greedy algorithm, which is the main focus of this article, our main theoretical result and its proof. This theoretical result does not yield a fully practical algorithm. To alleviate this difficulty, we propose in Section 4 a heuristic algorithm, inspired from the theoretical result, which provides satisfactory results on several test cases.

2 Motivation: greedy algorithms for reduced bases and variance reduction

2.1 Motivation: reduced basis control variate

The aim of this section is to present the motivation of our work, which aims at constructing control variates for reducing the variance of a Monte-Carlo estimator of the mean of parameter-dependent functions of random vectors.

Let us begin by introducing some notation. Let $d \in \mathbb{N}^*$, $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and Z a $\mathbb{R}^d$-valued random vector with associated probability measure ν. For all $q \in \mathbb{N}^*$, we denote by

$$L^q_\nu(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R}, \quad \int_{\mathbb{R}^d} |f(x)|^q \, d\nu(x) < +\infty \right\}.$$

Let $\mathcal{C}(\mathbb{R}^d)$ denote the set of continuous real-valued functions defined on $\mathbb{R}^d$. Let $p \in \mathbb{N}^*$, $\mathcal{P} \subset \mathbb{R}^p$ be a set of parameter values, and for all $\mu \in \mathcal{P}$, let f_μ be an element of $\mathcal{C}(\mathbb{R}^d) \cap L^2_\nu(\mathbb{R}^d)$.

For all $f, g \in \mathcal{C}(\mathbb{R}^d)$, any $M \in \mathbb{N}^*$ and any collection $\overline{Z} := (Z_k)_{1 \le k \le M}$ of random vectors of $\mathbb{R}^d$, we define the empirical averages:

$$\begin{aligned} \mathbb{E}_{\overline{Z}}(f) &:= \frac{1}{M} \sum_{k=1}^{M} f(Z_k), \\ \mathrm{Cov}_{\overline{Z}}(f, g) &:= \mathbb{E}_{\overline{Z}}(fg) - \mathbb{E}_{\overline{Z}}(f)\mathbb{E}_{\overline{Z}}(g), \\ \mathrm{Var}_{\overline{Z}}(f) &:= \mathrm{Cov}_{\overline{Z}}(f, f). \end{aligned}$$

The aim of our work is to propose and analyse from a mathematical point of view a numerical method in order to efficiently construct control variates to reduce the variance of a Monte-Carlo estimator of $\mathbb{E}[f_\mu(Z)]$ for all $\mu \in \mathcal{P}$ using a Reduced Basis paradigm [15, 3, 16, 12], which was originally proposed in [6].

More precisely, let $M_{\text{small}}, M_{\text{ref}} \in \mathbb{N}^*$ and assume that $M_{\text{ref}} \gg M_{\text{small}}$. Let $\overline{Z}^{\text{ref}} := \left(Z_k^{\text{ref}}\right)_{1\leq k\leq M_{\text{ref}}}$ and $\overline{Z}^{\text{small}} := \left(Z_k^{\text{small}}\right)_{1\leq k\leq M_{\text{small}}}$ be two independent collections of iid random vectors distributed according to the law of Z and **independent of** Z.

Let us assume that we have selected N values of parameters $(\mu_1, \mu_2, ..., \mu_N) \in \mathcal{P}^N$ for some $N \in \mathbb{N}^*$ and assume that the empirical means $(\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_i}))_{1\leq i\leq N}$ have been computed in an offline phase.

In an online phase, for all $\mu \in \mathcal{P}$, we can build an approximation of $\mathbb{E}\left[f_\mu(Z)\right]$, using a control variate which reads as $\overline{f}_\mu(Z)$ for some function $\overline{f}_\mu : \mathbb{R}^d \to \mathbb{R}$:

$$\mathbb{E}\left[f_\mu(Z)\right] \approx \mathbb{E}_{\overline{Z}^{\text{ref}}}(\overline{f}_\mu) + \mathbb{E}_{\overline{Z}^{\text{small}}}\left(f_\mu - \overline{f}_\mu\right). \tag{1}$$

Remark 2.1. *Let us point out that the statistical error between $\mathbb{E}_{\overline{Z}^{\text{ref}}}(\overline{f}_\mu)$ and $\mathbb{E}\left[\overline{f}_\mu(Z)\right]$ is close to*

$$\sqrt{\frac{\text{Var}\left[\overline{f}_\mu(Z)\right]}{M_{\text{ref}}}},$$

whereas the error between $\mathbb{E}_{\overline{Z}^{\text{small}}}\left(f_\mu - \overline{f}_\mu\right)$ and $\mathbb{E}\left[(f_\mu - \overline{f}_\mu)(Z)\right]$ is of the order of

$$\sqrt{\frac{\text{Var}\left[\left(f_\mu - \overline{f}_\mu\right)(Z)\right]}{M_{\text{small}}}}.$$

The aim of the Monte-Carlo greedy algorithm studied in this article is to give an approximation of $\mathbb{E}\left[f_\mu(Z)\right]$ with an error close to $\sqrt{\frac{\text{Var}[f_\mu(Z)]}{M_{\text{ref}}}}$ within a much smaller computational time than the one required by the computation of $\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_\mu)$.

In the method studied here, the control variate function $\overline{f}_\mu$ is constructed as follows:

$$\overline{f}_\mu = \sum_{i=1}^N \lambda_i^\mu f_{\mu_i}$$

where $\lambda^\mu := (\lambda_i^\mu)_{1\leq i\leq N} \in \mathbb{R}^N$ is a solution of the linear system

$$A\lambda^\mu = b^\mu \tag{2}$$

where $A := (A_{ij})_{1\leq i,j\leq N} \in \mathbb{R}^{N\times N}$ and $b^\mu := (b_i^\mu)_{1\leq i\leq N} \in \mathbb{R}^N$ are defined as follows: for all $1 \leq i, j \leq N$,

$$A_{ij} = \text{Cov}_{\overline{Z}^{\text{small}}}(f_{\mu_i}, f_{\mu_j}) \quad \text{and} \quad b_i^\mu = \text{Cov}_{\overline{Z}^{\text{small}}}(f_\mu, f_{\mu_i}). \tag{3}$$

Equivalently, the vector λ^μ is a solution of the minimization problem

$$\lambda^\mu \in \underset{\lambda:=(\lambda_i)_{1\leq i\leq N}\in\mathbb{R}^N}{\text{argmin}} \text{Var}_{\overline{Z}^{\text{small}}}\left(f_\mu - \sum_{i=1}^N \lambda_i f_{\mu_i}\right).$$

Let us point out that λ^μ is a random vector which can be written as a deterministic function of $\overline{Z}^{\text{small}}$. In other words, λ^μ is measurable with respect to $\overline{Z}^{\text{small}}$. Remarking that $\mathbb{E}_{\overline{Z}^{\text{ref}}}(\overline{f}_\mu) = \sum_{i=1}^N \lambda_i^\mu \mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_i})$, the computation of the approximation (1) of $\mathbb{E}\left[f_\mu(Z)\right]$ thus requires the following steps:

- **offline phase:** Compute $(\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_i}))_{1\leq i\leq N}$ (N empirical means with M_{ref} samples), $(\mathbb{E}_{\overline{Z}^{\text{small}}}(f_{\mu_i}))_{1\leq i\leq N}$ (N empirical means with M_{small} samples) and the matrix A (N^2 empirical covariances with M_{small} samples).
- **online phase:** For all $\mu \in \mathcal{P}$, compute b^μ (N empirical covariances with M_{small} samples) and solve the linear system (2) to obtain λ^μ. Then, compute the approximation (1) of $\mathbb{E}\left[f_\mu(Z)\right]$ as

$$\mathbb{E}\left[f_\mu(Z)\right] \approx \sum_{i=1}^N \lambda_i^\mu \mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_i}) + \mathbb{E}_{\overline{Z}^{\text{small}}}\left(f_\mu\right) - \sum_{i=1}^N \lambda_i^\mu \mathbb{E}_{\overline{Z}^{\text{small}}}\left(f_{\mu_i}\right), \tag{4}$$

which requires $\mathcal{O}(N)$ elementary operations and the computation of one empirical mean with M_{small} samples.

Naturally, the approximation of $\mathbb{E}\left[f_\mu(Z)\right]$ given by (1) can be interesting from a computational point of view in terms of variance reduction only if $\text{Var}\left[f_\mu(Z) - \overline{f}_\mu(Z)\right]$ is much smaller than $\text{Var}\left[f_\mu(Z)\right]$. The following question thus naturally arises: how can the set of parameters $(\mu_1, \mu_2, .., \mu_N) \in \mathcal{P}^N$ be chosen in the offline phase in order to ensure that $\text{Var}\left[f_\mu(Z) - \overline{f}_\mu(Z)\right]$ is as small as possible for any value of $\mu \in \mathcal{P}$?

Greedy algorithms stand as the state-of-the-art technique to construct such sets of snapshot parameters, enjoy very nice mathematical properties and are the backbone of the method proposed in [6] which we wish to analyze here. We present this family of algorithms and related existing theoretical convergence results in the next section.

2.2 Greedy algorithms for reduced basis

Let us recall here the results of [7, 4, 10] on the convergence rates of greedy algorithms for reduced bases, adapted to our context. Let us define

$$L^2_{\nu,0}(\mathbb{R}^d) := \left\{ g \in L^2_\nu(\mathbb{R}^d),\ \int_{\mathbb{R}^d} g \, d\nu = 0 \right\}.$$

It holds that $L^2_{\nu,0}(\mathbb{R}^d)$ is a Hilbert space, equipped with the scalar product $\langle \cdot, \cdot \rangle$ defined by

$$\forall g_1, g_2 \in L^2_{\nu,0}(\mathbb{R}^d), \quad \langle g_1, g_2 \rangle = \int_{\mathbb{R}^d} g_1 g_2 \, d\nu = \text{Cov}\left[g_1(Z), g_2(Z)\right].$$

The associated norm is denoted by $\| \cdot \|$ and is given by

$$\forall g \in L^2_{\nu,0}(\mathbb{R}^d), \quad \|g\| = \left(\int_{\mathbb{R}^d} |g|^2 \, d\nu \right)^{1/2} = \sqrt{\text{Var}\left[g(Z)\right]}.$$

For all $\mu \in \mathcal{P}$, let us define

$$g_\mu := f_\mu - \mathbb{E}\left[f_\mu(Z)\right] \tag{5}$$

and let us denote by

$$\mathcal{M} := \{g_\mu,\ \mu \in \mathcal{P}\} \tag{6}$$

so that $\mathcal{M} \subset L^2_{\nu,0}(\mathbb{R}^d)$. Let us assume that $\mathcal{M}$ is a compact subset of $L^2_{\nu,0}(\mathbb{R}^d)$. For all $n \in \mathbb{N}^*$, we introduce the Kolmogorov n-width of the set $\mathcal{M}$ in $L^2_{\nu,0}(\mathbb{R}^d)$, defined by

$$\begin{aligned} d_n(\mathcal{M}) := & \inf_{\substack{V_n \subset L^2_{\nu,0}(\mathbb{R}^d) \text{ subspace,} \\ \dim V_n = n}} \sup_{\mu \in \mathcal{P}} \inf_{g_n \in V_n} \sqrt{\text{Var}\left[g_\mu(Z) - g_n(Z)\right]} \\ = & \inf_{\substack{V_n \subset L^2_{\nu,0}(\mathbb{R}^d) \text{ subspace,} \\ \dim V_n = n}} \sup_{\mu \in \mathcal{P}} \inf_{g_n \in V_n} \|g_\mu - g_n\|. \end{aligned}$$

Let $0 < \gamma < 1$ and consider the following *weak greedy algorithm* with parameter γ.

Weak-Greedy Algorithm:

Initialization: Find $\mu_1 \in \mathcal{P}$ such that

$$\|g_{\mu_1}\|^2 \geq \gamma^2 \max_{\mu \in \mathcal{P}} \|g_\mu\|^2. \tag{7}$$

Set $V_1 := \text{Span}\{g_{\mu_1}\}$ and set $n = 2$.

Iteration $n \geq 2$: Find $\mu_n \in \mathcal{P}$ such that

$$\inf_{(\lambda_i)_{1 \leq i \leq n-1} \in \mathbb{R}^{n-1}} \left\| g_{\mu_n} - \sum_{i=1}^{n-1} \lambda_i g_{\mu_i} \right\|^2 \geq \gamma^2 \max_{\mu \in \mathcal{P}} \inf_{(\lambda_i)_{1 \leq i \leq n-1} \in \mathbb{R}^{n-1}} \left\| g_\mu - \sum_{i=1}^{n-1} \lambda_i g_{\mu_i} \right\|^2, \tag{8}$$

Set $V_n := V_{n-1} + \text{Span}\{g_{\mu_n}\} = \text{Span}\{g_{\mu_1}, \cdots, g_{\mu_n}\}$.

For all $n \in \mathbb{N}^*$, the error associated with the n-dimensional subspace V_n given by the weak greedy algorithm is defined by

$$\sigma_n(\mathcal{M}) := \max_{\mu \in \mathcal{P}} \inf_{(\lambda_i)_{1 \le i \le n} \in \mathbb{R}^n} \left\| g_\mu - \sum_{i=1}^{n} \lambda_i g_{\mu_i} \right\|.$$

The following result is then a direct corollary of the results proved in [10, Corollary 3.3].

Theorem 2.2. *For all $n \in \mathbb{N}^*$, $\sigma_n(\mathcal{M}) \le \sqrt{2}\gamma^{-1} \min_{0 \le m < n} (d_m(\mathcal{M}))^{\frac{n-m}{n}}$. In particular, for all $n \in \mathbb{N}^*$, $\sigma_{2n}(\mathcal{M}) \le \sqrt{2}\gamma^{-1}\sqrt{d_n(\mathcal{M})}$.*

This result indicates that the weak greedy algorithm provides a practical way to construct a quasi-optimal sequence $(V_n)_{n \in \mathbb{N}^*}$ of finite dimensional subspaces of $L^2_{\nu,0}(\mathbb{R}^d)$.

Of course, the weak greedy algorithm introduced above cannot be implemented in practice since it requires at the nth iteration of the algorithm the computation of the exact variances of $g_\mu(Z) - \sum_{i=1}^{n-1} \lambda_i g_{\mu_i}(Z)$ for $\mu, \mu_1, \cdots, \mu_{n-1} \in \mathcal{P}$ and $\lambda_1, \cdots, \lambda_{n-1} \in \mathbb{R}$, which is out of reach in our context. In practice, these quantities have to be approximated by Monte-Carlo estimators involving a finite number of samples of the random vector Z. The resulting greedy algorithm with Monte Carlo sampling is presented in Section 3. The mathematical analysis of this algorithm is the main purpose of the present article.

For the sake of simplicity, in the rest of the article, we assume that for all $n \in \mathbb{N}^*$, $d_n(\mathcal{M}) > 0$.

3 Greedy algorithm with Monte-Carlo sampling

3.1 Presentation of the algorithm

Let us begin by presenting the greedy algorithm with Monte Carlo sampling.

Let $(M_n)_{n \in \mathbb{N}^*}$ be a sequence of integers, which represents the number of samples used at iteration n. For all $n \in \mathbb{N}^*$, let $\overline{Z}^n := (Z^n_k)_{1 \le k \le M_n}$ be a collection of random vectors such that $(Z^n_k)_{n \ge 1,\, 1 \le k \le M_n}$ are independent and identically distributed according to the law of Z, and independent of Z. Let $\overline{Z}^{1:n} := \left(\overline{Z}^m\right)_{1 \le m \le n}$ and $\overline{Z}^{1:\infty} := \left(\overline{Z}^n\right)_{n \in \mathbb{N}^*}$.

For any random functions g_1, g_2 with values in $L^2_{\nu,0}(\mathbb{R}^d)$, we define

$$\langle g_1, g_2 \rangle_{\overline{Z}^{1:\infty}} := \mathrm{Cov}\left[g_1(Z), g_2(Z) \,\middle|\, \overline{Z}^{1:\infty} \right] \quad \text{and} \quad \|g_1\|_{\overline{Z}^{1:\infty}} := \sqrt{\mathrm{Var}\left[g_1(Z) \,\middle|\, \overline{Z}^{1:\infty} \right]}.$$

Let us make here an important remark. Since $\overline{Z}^{1:\infty}$ is a collection of random vectors which are all independent of Z, it holds that, for all $f, g \in L^2_{\nu,0}(\mathbb{R}^d)$, almost surely,

$$\begin{aligned} \langle f, g \rangle_{\overline{Z}^{1:\infty}} &= \mathrm{Cov}\left[f(Z), g(Z) \,\middle|\, \overline{Z}^{1:\infty} \right] = \mathrm{Cov}\left[f(Z), g(Z) \right] = \langle f, g \rangle, \\ \|g\|^2_{\overline{Z}^{1:\infty}} &= \mathrm{Var}\left[g(Z) \,\middle|\, \overline{Z}^{1:\infty} \right] = \mathrm{Var}[g(Z)] = \|g\|^2. \end{aligned}$$

Hence, almost surely, $\langle \cdot, \cdot \rangle_{\overline{Z}^{1:\infty}}$ defines a scalar product on $L^2_{\nu,0}(\mathbb{R}^d)$, which is a Hilbert space when equipped with this scalar product, and $\| \cdot \|_{\overline{Z}^{1:\infty}}$ is the associated norm.

The greedy algorithm with Monte-Carlo sampling reads as follows:

MC-Greedy Algorithm:

Initialization: Find $\overline{\mu}_1 \in \mathcal{P}$ such that, almost surely,

$$\overline{\mu}_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \mathrm{Var}_{\overline{Z}^1}(g_\mu) \quad \text{and} \quad g_{\overline{\mu}_1} \neq 0. \tag{9}$$

Set $\overline{V}_1 := \mathrm{Span}\{g_{\overline{\mu}_1}\}$ and set $n = 2$.

Iteration $n \geq 2$**:** Find $\overline{\mu}_n \in \mathcal{P}$ such that, almost surely,

$$\overline{\mu}_n \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \inf_{(\lambda_i)_{1 \leq i \leq n-1} \in \mathbb{R}^{n-1}} \mathrm{Var}_{\overline{Z}^n} \left(g_\mu - \sum_{i=1}^{n-1} \lambda_i g_{\overline{\mu}_i} \right) \quad \text{and} \quad g_{\overline{\mu}_n} \notin \overline{V}_{n-1}. \tag{10}$$

Set $\overline{V}_n := \overline{V}_{n-1} + \mathrm{Span}\{g_{\overline{\mu}_n}\} = \mathrm{Span}\{g_{\overline{\mu}_1}, \cdots, g_{\overline{\mu}_n}\}$.

Naturally, for all $n \in \mathbb{N}^*$, the parameter $\overline{\mu}_n$ and thus the finite-dimensional space $\overline{V}_n$ are $\overline{Z}^{1:n}$-measurable.

Let us first prove an auxiliary lemma.

Lemma 3.1. *Almost surely, all the iterations of the MC-Greedy Algorithm are well-defined, in the sense that, for all $n \in \mathbb{N}^*$, there always exists at least one element $\overline{\mu}_n \in \mathcal{P}$ such that (9) (when $n = 1$) or (10) (when $n \geq 2$) is satisfied.*

Proof of Lemma 3.1. Let us first consider the initialization step corresponding to $n = 1$. Two situations may a priori occur : either $\max_{\mu \in \mathcal{P}} \mathrm{Var}_{\overline{Z}^1}(g_\mu) > 0$ or $\max_{\mu \in \mathcal{P}} \mathrm{Var}_{\overline{Z}^1}(g_\mu) = 0$. In the first case, choosing $\overline{\mu}_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \mathrm{Var}_{\overline{Z}^1}(g_\mu)$ is sufficient to guarantee that $g_{\overline{\mu}_1} \neq 0$. Indeed, since $\mathcal{M} \subset \mathcal{C}(\mathbb{R}^d)$ (remember that f_μ is continuous for all $\mu \in \mathcal{P}$, and hence so is g_μ), the fact that $\mathrm{Var}_{\overline{Z}^1}\left(g_{\overline{\mu}_1}\right) > 0$ necessarily implies that $\mathrm{Var}\left[g_{\overline{\mu}_1}(Z) \middle| \overline{Z}^{1:\infty}\right] > 0$ almost surely. Since $\overline{Z}^{1:\infty}$ is independent of Z and $\overline{\mu}_1$ is a $\overline{Z}^{1:\infty}$ measurable random variable, this implies that almost surely $g_{\overline{\mu}_1} \neq 0$.

In the second case, it then holds that $\mathrm{Var}_{\overline{Z}^1}(g_\mu) = 0$ for all $\mu \in \mathcal{P}$. Then, the assumption $d_1(\mathcal{M}) > 0$ implies that, almost surely, there exists at least one element $\overline{\mu}_1 \in \mathcal{P}$ such that $g_{\overline{\mu}_1} \neq 0$. In addition, $\overline{\mu}_1 \in \operatorname{argmax}_{\mu \in \mathcal{P}} \mathrm{Var}_{\overline{Z}^1}(g_\mu)$.

Using similar arguments and the fact that $d_n(\mathcal{M}) > 0$ for all $n \in \mathbb{N}^*$, it is easy to see that, almost surely, all the iterations of the MC-Greedy algorithm are well-defined, in particular for $n \geq 2$. □

Remark 3.2. *We stress on the fact that the practical implementation of the MC-greedy algorithm does not require the knowledge of the value of $\mathbb{E}[f_\mu(Z)]$, even if $g_\mu = f_\mu - \mathbb{E}[f_\mu(Z)]$ for all $\mu \in \mathcal{P}$. Indeed, it holds that for all $g \in \mathcal{C}(\mathbb{R}^d)$, all $n \in \mathbb{N}^*$ and all $C \in \mathbb{R}$, $\mathrm{Var}_{\overline{Z}^n}(g) = \mathrm{Var}_{\overline{Z}^n}(g + C)$. Thus, for all $\mu \in \mathcal{P}$, $n \in \mathbb{N}^*$ and $\lambda := (\lambda_i)_{1 \leq i \leq n-1} \in \mathbb{R}^{n-1}$,*

$$\mathrm{Var}_{\overline{Z}^1}(g_\mu) = \mathrm{Var}_{\overline{Z}^1}(f_\mu) \quad \textit{and} \quad \mathrm{Var}_{\overline{Z}^n}\left(g_\mu - \sum_{i=1}^{n-1} \lambda_i g_{\overline{\mu}_i} \right) = \mathrm{Var}_{\overline{Z}^n}\left(f_\mu - \sum_{i=1}^{n-1} \lambda_i f_{\overline{\mu}_i} \right).$$

Thus, the MC-greedy algorithm naturally makes sense with a view to the construction of a reduced basis control variate for variance reduction as explained in Section 2.1.

Remark 3.3. *In practice, a discrete subset $\mathcal{P}_{\mathrm{trial}} \subset \mathcal{P}$ has to be introduced. The optimization problems (9) and (10) have to be replaced respectively by*

$$\overline{\mu}_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}_{\mathrm{trial}}} \mathrm{Var}_{\overline{Z}^1}(g_\mu) \quad \textit{and} \quad g_{\overline{\mu}_1} \neq 0,$$

and

$$\overline{\mu}_n \in \operatorname*{argmax}_{\mu \in \mathcal{P}_{\mathrm{trial}}} \inf_{(\lambda_i)_{1 \leq i \leq n-1} \in \mathbb{R}^{n-1}} \mathrm{Var}_{\overline{Z}^n} \left(g_\mu - \sum_{i=1}^{n-1} \lambda_i g_{\overline{\mu}_i} \right) \quad \textit{and} \quad g_{\overline{\mu}_n} \notin \overline{V}_{n-1}.$$

The influence of the choice of the set $\mathcal{P}_{\mathrm{trial}}$ on the mathematical properties of the MC-greedy algorithm is an important question which we do not address in our analysis for the sake of simplicity. For related discussion, we refer the reader to the work [9], where the authors study the mathematical properties of a greedy algorithm where the set $\mathcal{P}_{\mathrm{trial}}$ depends on the iteration n of the greedy algorithm and is randomly chosen according to appropriate probability distributions defined on the set of parameters $\mathcal{P}$.

For all $n \in \mathbb{N}^*$, we also define

$$\widehat{\sigma}_{n-1}(\mathcal{M}) := \max_{\mu \in \mathcal{P}} \inf_{(\lambda_i)_{1 \leq i \leq n-1} \in \mathbb{R}^{n-1}} \sqrt{\mathrm{Var}\left[g_\mu(Z) - \sum_{i=1}^{n-1} \lambda_i g_{\overline{\mu}_i}(Z) \middle| \overline{Z}^{1:\infty} \right]}, \tag{11}$$

$$\overline{\sigma}_{n-1}(\mathcal{M}) := \inf_{(\lambda_i)_{1 \leq i \leq n-1} \in \mathbb{R}^{n-1}} \sqrt{\mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1} \lambda_i g_{\overline{\mu}_i}(Z) \middle| \overline{Z}^{1:\infty} \right]} \tag{12}$$

Let us point out here that $\widehat{\sigma}_{n-1}(\mathcal{M})$ is a random variable which is measurable with respect to $\overline{Z}^{1:(n-1)}$ whereas $\overline{\mu}_n$ and $\overline{\sigma}_{n-1}(\mathcal{M})$ are measurable with respect to $\overline{Z}^{1:n}$.

3.2 Main theoretical result

The aim of this section is to study the effect of Monte-Carlo sampling on the convergence of such a greedy algorithm. We consider here the probability space $(\Omega, \mathcal{A}(\overline{Z}^{1:\infty}), \mathbb{P})$ the probabilty space where $\mathcal{A}(\overline{Z}^{1:\infty})$ denotes the set of events that are measurable with respect to $\overline{Z}^{1:\infty}$. We prove, under appropriate assumptions on the probability density ν and on the set of functions $\mathcal{M} = \{g_\mu, \mu \in \mathcal{P}\}$, that for all $0 < \gamma < 1$, there exist explicit conditions on the sequence $(M_n)_{n\in\mathbb{N}^*}$ so that, with high probability, the MC-greedy algorithm is actually a weak greedy algorithm with parameter γ. More precisely, under this set of assumptions, we prove that, with high probability, it holds that for all $n \in \mathbb{N}^*$,

$$\overline{\sigma}_{n-1}(\mathcal{M}) \geq \gamma \widehat{\sigma}_{n-1}(\mathcal{M}).$$

Let us now present the set of assumptions we make on ν and on the set $\mathcal{M} = \{g_\mu,\ \mu \in \mathcal{P}\}$ for our main result to hold.

From now on, we make the following assumption on the probability distribution ν.

Assumption (A): The probability law ν is such that there exist $\alpha > 1$ and $\beta > 0$ such that

$$\int_{\mathbb{R}} e^{\beta|x|^\alpha}\, d\nu(x) < +\infty.$$

Let us denote by $\mathcal{L}$ the set of Lipschitz functions of $\mathbb{R}^d$ and for all $f \in \mathcal{L}$, let us denote by $\|f\|_{\mathcal{L}}$ its Lipschitz constant. In the sequel, we denote by $\phi : \mathbb{R}_+^* \to \mathbb{R}_+^*$ the function defined by

$$\forall \kappa \in \mathbb{R}_+^*, \quad \phi(\kappa) := \begin{cases} \kappa^2 \mathbb{1}_{\kappa\leq 1} + \kappa^\alpha \mathbb{1}_{\kappa>1} & \text{if } d = 1, \\ (\kappa/\log(2+1/\kappa)^2)\mathbb{1}_{\kappa\leq 1} + \kappa^\alpha \mathbb{1}_{\kappa>1} & \text{if } d = 2, \\ \kappa^d \mathbb{1}_{\kappa\leq 1} + \kappa^\alpha \mathbb{1}_{\kappa>1} & \text{if } d \geq 3. \end{cases} \tag{13}$$

A key ingredient in our analysis is the use of concentration inequalities in the Wasserstein-1 distance between a probability distribution and its empirical measure proved in [5, 14]. Let us recall here a direct corollary of Theorem 2 of [14], which is the backbone of our analysis.

Corollary 3.4. *Let us assume that ν satisfies assumption (A). Then, there exist positive constants c, C depending only on ν, d, α and β, such that, for all $M \in \mathbb{N}^*$, all $\overline{Z} := (Z_k)_{1\leq k\leq M}$ iid random vectors distributed according to ν and all $\kappa > 0$, it holds that*

$$\mathbb{P}\left[\mathcal{T}_1\left(\overline{Z}\right) \geq \kappa\right] \leq Ce^{-cM\phi(\kappa)},$$

where

$$\mathcal{T}_1\left(\overline{Z}\right) := \sup_{f\in\mathcal{L};\|f\|_{\mathcal{L}}\leq 1} \left|\mathbb{E}[f(Z)] - \mathbb{E}_{\overline{Z}}(f)\right|.$$

Remark 3.5. *We would like to mention here that other concentration inequalities are stated in Theorem 2 of [14] under different sets of assumptions than (A) on the probability law ν. In particular, weaker concentration inequalities may be obtained when ν only has some finite polynomial moments. Our analysis can then be easily adapted to these different settings but we restrict ourselves here to a framework where ν satisfies Assumption (A) for the sake of clarity.*

We finally make the following set of assumptions on $\mathcal{M}$ defined in (6).

Assumption (B): The set $\mathcal{M}$ satisfies the four conditions:

(B1) $\mathcal{M}$ is a compact subset of $L^2_{\nu,0}(\mathbb{R}^d)$ and let $K_2 := \sup_{\mu\in\mathcal{P}} \|g_\mu\| < \infty$;

(B2) $\mathcal{M} \subset \mathcal{L}$ and $K_{\mathcal{L}} := \sup_{\mu\in\mathcal{P}} \|g_\mu\|_{\mathcal{L}} < +\infty$;

(B3) $\mathcal{M} \subset L^\infty(\mathbb{R}^d)$ and $K_\infty := \sup_{\mu\in\mathcal{P}} \|g_\mu\|_{L^\infty} < +\infty$;

(B4) for all $n \in \mathbb{N}^*$, $d_n(\mathcal{M}) > 0$.

Before presenting our main result, we need to introduce some additional notation. Using Lemma 3.1, we can almost surely define the sequence $(\overline{g}_n)_{n\in\mathbb{N}^*}$ as the orthonormal family of $L^2_{\nu,0}(\mathbb{R}^d)$ obtained by a Gram-Schmidt orthonormalization procedure (for $\|\cdot\|_{\overline{Z}^{1:\infty}}$) from the family $(g_{\overline{\mu}_n})_{n\in\mathbb{N}^*}$. More precisely, we define

$$\overline{g}_1 := \frac{g_{\overline{\mu}_1}}{\sqrt{\mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\,\middle|\,\overline{Z}^{1:\infty}\right]}}.$$

Moreover, for all $n \geq 2$, et $\overline{\lambda}^n := \left(\overline{\lambda}^n_i\right)_{1\leq i\leq n-1} \in \mathbb{R}^{n-1}$ be a solution to the minimization problem

$$\overline{\lambda}^n \in \underset{\lambda:=(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}{\mathrm{argmin}} \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i(Z)\,\middle|\,\overline{Z}^{1:\infty}\right].$$

Then it holds that

$$\overline{g}_n := \frac{g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}^n_i\overline{g}_i}{\sqrt{\mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}^n_i\overline{g}_i(Z)\,\middle|\,\overline{Z}^{1:\infty}\right]}}.$$

As a consequence, it always holds that $\overline{V}_n = \mathrm{Span}\left\{g_{\overline{\mu}_1},\cdots,g_{\overline{\mu}_n}\right\} = \mathrm{Span}\left\{\overline{g}_1,\cdots,\overline{g}_n\right\}$. Moreover, $\overline{g}_n$ is $\overline{Z}^{1:n}$-measurable.

We are now in position to state our main result, the proof of which is postponed to Section 3.3.

Theorem 3.6. *Let $0<\delta<1$ and $(\delta_n)_{n\in\mathbb{N}^*}\subset(0,1)^{\mathbb{N}^*}$ be a sequence of numbers satisfying $\Pi_{n\in\mathbb{N}^*}(1-\delta_n)\geq 1-\delta$. Let us assume that $\mathcal{M}$ satisfies assumption (B) and that ν satisfies assumption (A). Let $C,c>0$ be the constants defined in Corollary 3.4.*

For all $n\in\mathbb{N}^*$, let

$$K^n_\infty := \max\left(K_\infty, \|\overline{g}_1\|_{L^\infty},\cdots,\|\overline{g}_n\|_{L^\infty}\right) \quad \text{and} \quad K^n_{\mathcal{L}} := \max\left(K_{\mathcal{L}}, \|\overline{g}_1\|_{\mathcal{L}},\cdots,\|\overline{g}_n\|_{\mathcal{L}}\right). \tag{14}$$

Let us assume that there exists $0<\gamma<1$ such that for all $n\in\mathbb{N}^*$, $M_n\in\mathbb{N}^*$ is a $\overline{Z}^{1:(n-1)}$ measurable random variable which satisfies almost surely the following condition:

$$\forall n\geq 1,\quad M_n \geq -\ln\left(\frac{\delta_n}{C}\right)\frac{1}{c\phi\left(\kappa_{n-1}\right)}, \tag{15}$$

where κ_{n-1} is a deterministic function of $\overline{Z}^{1:(n-1)}$, defined by

$$\kappa_0 := \frac{\left(1-\gamma^2\right)\widehat{\sigma}_0(\mathcal{M})^2}{8K_\infty K_{\mathcal{L}}}; \tag{16}$$

and

$$\forall n\geq 2,\quad \kappa_{n-1} := \frac{\min\left(\frac{1}{2(n-1)},\ \frac{(1-\gamma^2)\widehat{\sigma}_{n-1}(\mathcal{M})^2}{n\left(9K_2^2+4\right)}\right)}{6K_\infty^{n-1}K_{\mathcal{L}}^{n-1}}. \tag{17}$$

Then, for all $n\in\mathbb{N}^*$, it holds that

$$\mathbb{P}\left[\overline{\sigma}_{n-1}(\mathcal{M})\geq\gamma\widehat{\sigma}_{n-1}(\mathcal{M})\,\middle|\,\overline{Z}^{1:(n-1)}\right]\geq 1-\delta_n. \tag{18}$$

As a consequence, denoting by $\mathcal{G}_n$ the event $\overline{\sigma}_{n-1}(\mathcal{M})\geq\gamma\widehat{\sigma}_{n-1}(\mathcal{M})$ for all $n\in\mathbb{N}^*$, it holds that

$$\mathbb{P}\left[\bigcap_{n\in\mathbb{N}^*}\mathcal{G}_n\right]\geq 1-\delta. \tag{19}$$

Thus, it then holds that the MC-greedy algorithm is a weak greedy algorithm with parameter γ and norm $\|\cdot\|_{\overline{Z}^{1:\infty}}$ with probability at least $1-\delta$.

We state here a direct corollary of Theorem 3.6, the proof of which is given below.

Corollary 3.7. *Under the assumptions of Theorem 3.6, with probability $1-\delta$, it holds that for all $n \in \mathbb{N}^*$,*

$$\widehat{\sigma}_n(\mathcal{M}) \leq \sqrt{2}\gamma^{-1} \min_{1 \leq m < n} \left(d_m(\mathcal{M})\right)^{\frac{n-m}{n}}. \tag{20}$$

In particular, with probability $1-\delta$, it holds that

$$\forall n \in \mathbb{N}^*, \quad \widehat{\sigma}_{2n}(\mathcal{M}) \leq \sqrt{2}\gamma^{-1}\sqrt{d_n(\mathcal{M})}. \tag{21}$$

Proof. With probability $1-\delta$, the MC-greedy algorithm is a weak greedy algorithm with parameter γ and norm $\|\cdot\|_{\overline{Z}^{1:\infty}}$. Thus, since for all $n \in \mathbb{N}^*$, $\overline{\mu}_n$ is a $\overline{Z}^{1:\infty}$ measurable random variable, if such an event is realized, using Theorem 2.2, it holds that for all $n \in \mathbb{N}^*$

$$\widehat{\sigma}_n(\mathcal{M}) \leq \sqrt{2}\gamma^{-1} \min_{1 \leq m < n} \left(d_m^{\overline{Z}^{1:\infty}}(\mathcal{M})\right)^{\frac{n-m}{n}},$$

where for all $n \in \mathbb{N}^*$,

$$\begin{aligned} d_n^{\overline{Z}^{1:\infty}}(\mathcal{M}) &:= \inf_{\substack{V_n \subset L^2_{\nu,0}(\mathbb{R}^d) \text{ subspace,} \\ \dim V_n = n}} \sup_{\mu \in \mathcal{P}} \inf_{g_n \in V_n} \sqrt{\mathrm{Var}\left[g_\mu(Z) - g_n(Z) \,\middle|\, \overline{Z}^{1:\infty}\right]} \\ &= \inf_{\substack{V_n \subset L^2_{\nu,0}(\mathbb{R}^d) \text{ subspace,} \\ \dim V_n = n}} \sup_{\mu \in \mathcal{P}} \inf_{g_n \in V_n} \sqrt{\mathrm{Var}\left[g_\mu(Z) - g_n(Z)\right]} \\ &= d_n(\mathcal{M}). \end{aligned}$$

Hence, we obtain (20), and (21) as a consequence. □

Some remarks are in order here.

Remark 3.8. *Note that, since the random variables K_∞^{n-1}, $K_\mathcal{L}^{n-1}$ and $\widehat{\sigma}_{n-1}(\mathcal{M})$ are measurable with respect to $\overline{Z}^{1:(n-1)}$, κ_{n-1} is also measurable with respect to $\overline{Z}^{1:(n-1)}$.*

Remark 3.9. *A natural question is then the following: can Theorem 3.6 be used (at least in principle) to design a* constructive strategy *to choose a number of samples M_n, so that the MC-greedy algorithm can be guaranteed to be a weak greedy algorithm with parameter γ? This can indeed be done in principle using the following remark: for all $n \in \mathbb{N}^*$, the quantity $\widehat{\sigma}_{n-1}(\mathcal{M})$ defined by (11) cannot be computed in practice since variances cannot be computed exactly for any parameter $\mu \in \mathcal{P}$. However, almost surely, it holds that $\overline{\sigma}_{n-1}(\mathcal{M})$ defined by (12) satisfies $\overline{\sigma}_{n-1}(\mathcal{M}) \leq \widehat{\sigma}_{n-1}(\mathcal{M})$. Let us recall that $\overline{\sigma}_{n-1}(\mathcal{M})$ depends on $\overline{Z}^{1:n}$, whereas $\widehat{\sigma}_{n-1}(\mathcal{M})$ only depends on $\overline{Z}^{1:(n-1)}$. Since ϕ is an increasing function, this implies that, if the sequence $(M_n)_{n\in\mathbb{N}^*}$ satisfies condition (15) where $\widehat{\sigma}_0(\mathcal{M})$ is replaced by $\overline{\sigma}_0(\mathcal{M})$ in (16) and $\widehat{\sigma}_{n-1}(\mathcal{M})$ is replaced by $\overline{\sigma}_{n-1}(\mathcal{M})$ in (17), the assumptions of Theorem 3.6 are satisfied. Besides, it is reasonable to expect in this case that $\overline{\sigma}_{n-1}(\mathcal{M})$ should provide a reasonable approximation of $\widehat{\sigma}_{n-1}(\mathcal{M})$ since, from Theorem 3.6, $\overline{\sigma}_{n-1}(\mathcal{M}) \geq \gamma\widehat{\sigma}_{n-1}(\mathcal{M})$ with high probability.*

Unfortunately, we will see that such an approach is not viable in practice, because it leads to much too large values of M_n for small values of n for the MC-greedy algorithm to be interesting with a view to the variance reduction method explained in Section 2.1. That is why in Section 4, we will present numerical results with heuristic ways to choose values of $(M_n)_{n\in\mathbb{N}^*}$ which are not theoretically guaranteeed, but which nevertheless yield satisfactory numerical results in several test cases.

3.3 Proof of Theorem 3.6

The aim of this section is to prove Theorem 3.6. For all $n \in \mathbb{N}^*$, we denote by $\mathcal{G}_n$ the event $\overline{\sigma}_n(\mathcal{M}) \geq \gamma\widehat{\sigma}_n(\mathcal{M})$.

Let us begin by proving some intermediate results which will be used later. We first need the following auxiliary lemma.

Lemma 3.10. *Let $n \in \mathbb{N}^*$. Then, almost surely,*

$$\sup_{\substack{f \ \overline{Z}^{1:\infty}\text{-measurable random function} \\ \text{such that } \|f\|_\mathcal{L} \leq 1 \text{ almost surely}}} \left|\mathbb{E}\left[f(Z)|\overline{Z}^{1:\infty}\right] - \mathbb{E}_{\overline{Z}^n}(f)\right| = \sup_{f\in\mathcal{L};\|f\|_\mathcal{L}\leq 1} \left|\mathbb{E}\left[f(Z)\right] - \mathbb{E}_{\overline{Z}^n}(f)\right|.$$

Proof. On the one hand, it is obvious to check that

$$\sup_{f\in\mathcal{L};\|f\|_{\mathcal{L}}\leq 1}\left|\mathbb{E}\left[f(Z)\right]-\mathbb{E}_{\overline{Z}^n}(f)\right|\leq \sup_{\substack{f\ \overline{Z}^{1:\infty}\text{-measurable random function}\\ \text{such that } \|f\|_{\mathcal{L}}\leq 1 \text{ almost surely}}}\left|\mathbb{E}\left[f(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^n}(f)\right|.$$

On the other hand, for any $\overline{Z}^{1:\infty}$-measurable random function f such that $\|f\|_{\mathcal{L}}\leq 1$ almost surely, it holds that, almost surely, since $\overline{Z}^{1:\infty}$ is independent of Z, $\mathbb{E}[f(Z)|\overline{Z}^{1:\infty}]=\mathbb{E}_Z[f(Z)]$ where the index Z in $\mathbb{E}_Z$ indicates that the expectation is only taken with respect to Z, and thus

$$\left|\mathbb{E}\left[f(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^n}(f)\right|\leq \sup_{f\in\mathcal{L};\|f\|_{\mathcal{L}}\leq 1}\left|\mathbb{E}\left[f(Z)\right]-\mathbb{E}_{\overline{Z}^n}(f)\right|.$$

Hence the result. □

We start by considering the case of the initialization of the MC-greedy algorithm.

Lemma 3.11. *Let $0<\gamma<1$. Then, it holds that almost surely,*

$$\mathbb{P}\left[\mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right]\geq\gamma^2\max_{\mu\in\mathcal{P}}\mathrm{Var}\left[g_{\mu}(Z)\left|\overline{Z}^{1:\infty}\right.\right]\right]\geq 1-\delta_1. \tag{22}$$

As a consequence, $\mathbb{P}\left[\mathcal{G}_1\right]\geq 1-\delta_1$ and (18) holds for $n=1$.

Proof. Let $\widehat{\mu}_1\in\mathcal{P}$ such that

$$\widehat{\sigma}_0(\mathcal{M})^2=\max_{\mu\in\mathcal{P}}\mathrm{Var}\left[g_{\mu}(Z)\left|\overline{Z}^{1:\infty}\right.\right]=\mathrm{Var}\left[g_{\widehat{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right].$$

Inequality (22) holds provided that

$$\mathbb{P}\left[\left(\mathrm{Var}\left[g_{\widehat{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right]-\mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right]\right)>\epsilon\widehat{\sigma}_0(\mathcal{M})^2\right]\leq\delta_1,$$

with $\epsilon:=\left(1-\gamma^2\right)$. Almost surely, since $\overline{\mu}_1\in\mathrm{argmax}_{\mu\in\mathcal{P}}\mathrm{Var}_{\overline{Z}^1}(g_\mu)$, it holds that

$$\begin{aligned}
&\mathrm{Var}\left[g_{\widehat{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right]-\mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right]\\
&=\mathrm{Var}\left[g_{\widehat{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right]-\mathrm{Var}_{\overline{Z}^1}\left(g_{\widehat{\mu}_1}\right)+\mathrm{Var}_{\overline{Z}^1}\left(g_{\widehat{\mu}_1}\right)-\mathrm{Var}_{\overline{Z}^1}\left(g_{\overline{\mu}_1}\right)+\mathrm{Var}_{\overline{Z}^1}\left(g_{\overline{\mu}_1}\right)-\mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right]\\
&\leq\mathrm{Var}\left[g_{\widehat{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right]-\mathrm{Var}_{\overline{Z}^1}\left(g_{\widehat{\mu}_1}\right)+\mathrm{Var}_{\overline{Z}^1}\left(g_{\overline{\mu}_1}\right)-\mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right]\\
&=\mathbb{E}\left[|g_{\widehat{\mu}_1}|^2(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^1}\left(|g_{\widehat{\mu}_1}|^2\right)+\mathbb{E}_{\overline{Z}^1}\left(g_{\widehat{\mu}_1}\right)^2-\mathbb{E}\left[g_{\widehat{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right]^2\\
&\quad-\mathbb{E}\left[|g_{\overline{\mu}_1}|^2(Z)|\overline{Z}^{1:\infty}\right]+\mathbb{E}_{\overline{Z}^1}\left(|g_{\overline{\mu}_1}|^2\right)-\mathbb{E}_{\overline{Z}^1}\left(g_{\overline{\mu}_1}\right)^2+\mathbb{E}\left[g_{\overline{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right]^2\\
&\leq\left|\mathbb{E}\left[|g_{\widehat{\mu}_1}|^2(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^1}\left(|g_{\widehat{\mu}_1}|^2\right)\right|+2K_\infty\left|\mathbb{E}_{\overline{Z}^1}\left(g_{\widehat{\mu}_1}\right)-\mathbb{E}\left[g_{\widehat{\mu}_1}(Z)\right]\right|\\
&\quad+\left|\mathbb{E}\left[|g_{\overline{\mu}_1}|^2(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^1}\left(|g_{\overline{\mu}_1}|^2\right)\right|+2K_\infty\left|\mathbb{E}_{\overline{Z}^1}\left(g_{\overline{\mu}_1}\right)-\mathbb{E}\left[g_{\overline{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right]\right|,\\
&\leq 2K_\infty K_{\mathcal{L}}\\
&\times\left(\left|\mathbb{E}\left[\frac{|g_{\widehat{\mu}_1}|^2}{2K_\infty K_{\mathcal{L}}}(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^1}\left(\frac{|g_{\widehat{\mu}_1}|^2}{2K_\infty K_{\mathcal{L}}}\right)\right|+\left|\mathbb{E}\left[\frac{|g_{\overline{\mu}_1}|^2}{2K_\infty K_{\mathcal{L}}}(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^1}\left(\frac{|g_{\overline{\mu}_1}|^2}{2K_\infty K_{\mathcal{L}}}\right)\right|\right.\\
&\quad\left.+\left|\mathbb{E}_{\overline{Z}^1}\left(\frac{g_{\overline{\mu}_1}}{K_{\mathcal{L}}}\right)-\mathbb{E}\left[\frac{g_{\overline{\mu}_1}}{K_{\mathcal{L}}}(Z)|\overline{Z}^{1:\infty}\right]\right|+\left|\mathbb{E}_{\overline{Z}^1}\left(\frac{g_{\widehat{\mu}_1}}{K_{\mathcal{L}}}\right)-\mathbb{E}\left[\frac{g_{\widehat{\mu}_1}}{K_{\mathcal{L}}}(Z)|\overline{Z}^{1:\infty}\right]\right|\right).
\end{aligned}$$

It holds that for all $\mu\in\mathcal{P}$, $\||g_\mu|^2\|_{\mathcal{L}}\leq 2K_\infty K_{\mathcal{L}}$. Indeed, for all $x,y\in\mathbb{R}^d$, we have

$$||g_\mu|^2(x)-|g_\mu|^2(y)|=|(g_\mu(x)+g_\mu(y))(g_\mu(x)-g_\mu(y))|\leq 2K_\infty K_{\mathcal{L}}|x-y|.$$

Thus, almost surely, it holds that

$$\mathrm{Var}\left[g_{\widehat{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right]-\mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right]\leq 8K_\infty K_{\mathcal{L}}\sup_{f\in\mathcal{L};\|f\|_{\mathcal{L}}\leq 1}\left|\mathbb{E}\left[f(Z)|\overline{Z}^{1:\infty}\right]-\mathbb{E}_{\overline{Z}^1}\left(f\right)\right|.$$

Then, using Lemma 3.10, we obtain that, almost surely,

$$\mathrm{Var}\left[g_{\widehat{\mu}_1}(Z)|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\overline{\mu}_1}(Z)\left|\overline{Z}^{1:\infty}\right.\right] \leq 8K_\infty K_{\mathcal{L}} \sup_{f\in\mathcal{L};\|f\|_{\mathcal{L}}\leq 1} \left|\mathbb{E}\left[f(Z)\right] - \mathbb{E}_{\overline{Z}^1}(f)\right|.$$

Thus, using Theorem 3.4, the assumption on M_1 and the definition of κ_0, we obtain that

$$\mathbb{P}\left[\sup_{f\in\mathcal{L};\|f\|_{\mathcal{L}}\leq 1} \left|\mathbb{E}\left[f(Z)\right] - \mathbb{E}_{\overline{Z}^1}(f)\right| \geq \kappa_0\right] \leq Ce^{-c\phi(\kappa_0)} \leq \delta_1.$$

Hence the desired result. □

We now turn to the case of the nth iteration of the algorithm, with $n \geq 2$, that we analyze in the next two lemmas.

Lemma 3.12. *Let $n \geq 2$. Let us denote by*

$$\mathcal{M}^{n-1} := \mathcal{M} \cup \{\overline{g}_1, \cdots, \overline{g}_{n-1}\}. \tag{23}$$

Then, for all $\epsilon > 0$, it holds that, almost surely,

$$\mathbb{P}\left[\sup_{g,h\in\mathcal{M}^{n-1}} \left|\mathrm{Cov}\left[g(Z),h(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right| \geq \epsilon \left|\overline{Z}^{1:(n-1)}\right.\right] \leq Ce^{-cM_n\phi\left(\frac{\epsilon}{6K_\infty^{n-1}K_{\mathcal{L}}^{n-1}}\right)},$$

where $K_{\mathcal{L}}^{n-1}$ and K_∞^{n-1} are defined by (14).

Proof. For all $g, h \in \mathcal{M}^{n-1}$, it holds that, almost surely,

$$\begin{aligned}
&\left|\mathrm{Cov}\left[g(Z),h(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right| \\
&\leq \left|\mathbb{E}\left[g(Z)h(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathbb{E}_{\overline{Z}^n}(gh)\right| \\
&\quad + K_\infty^{n-1}\left(\left|\mathbb{E}\left[g(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathbb{E}_{\overline{Z}^n}(g)\right| + \left|\mathbb{E}\left[h(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathbb{E}_{\overline{Z}^n}(h)\right|\right) \\
&\leq 2K_\infty^{n-1}K_{\mathcal{L}}^{n-1}\left(\left|\mathbb{E}\left[\frac{gh}{2K_\infty^{n-1}K_{\mathcal{L}}^{n-1}}(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathbb{E}_{\overline{Z}^n}\left(\frac{gh}{2K_\infty^{n-1}K_{\mathcal{L}}^{n-1}}\right)\right|\right. \\
&\quad \left. + \left|\mathbb{E}\left[\frac{g}{2K_{\mathcal{L}}^{n-1}}(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathbb{E}_{\overline{Z}^n}\left(\frac{g}{2K_{\mathcal{L}}^{n-1}}\right)\right| + \left|\mathbb{E}\left[\frac{h}{2K_{\mathcal{L}}^{n-1}}(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathbb{E}_{\overline{Z}^n}\left(\frac{h}{2K_{\mathcal{L}}^{n-1}}\right)\right|\right).
\end{aligned}$$

For all $g, h \in \mathcal{M}^{n-1}$, it holds that

$$\left\|\frac{gh}{2K_\infty^{n-1}K_{\mathcal{L}}^{n-1}}\right\|_{\mathcal{L}} \leq 1 \quad \text{and} \quad \left\|\frac{g}{2K_{\mathcal{L}}^{n-1}}\right\|_{\mathcal{L}} \leq 1.$$

This implies that, almost surely,

$$\sup_{g,h\in\mathcal{M}^{n-1}} \left|\mathrm{Cov}\left[g(Z),h(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right| \leq 6K_\infty^{n-1}K_{\mathcal{L}}^{n-1} \sup_{f\in\mathcal{L},\ \|f\|_{\mathcal{L}}\leq 1} \left|\mathbb{E}\left[f(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathbb{E}_{\overline{Z}^n}(f)\right|.$$

Using Lemma 3.10, this yields that, almost surely,

$$\sup_{g,h\in\mathcal{M}^{n-1}} \left|\mathrm{Cov}\left[g(Z),h(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right| \leq 6K_\infty^{n-1}K_{\mathcal{L}}^{n-1} \sup_{f\in\mathcal{L},\ \|f\|_{\mathcal{L}}\leq 1} \left|\mathbb{E}\left[f(Z)\right] - \mathbb{E}_{\overline{Z}^n}(f)\right|.$$

We finally obtain, using Corollary 3.4, that

$$\begin{aligned}
&\mathbb{P}\left[\sup_{g,h\in\mathcal{M}^{n-1}} \left|\mathrm{Cov}\left[g(Z),h(Z)\left|\overline{Z}^{1:\infty}\right.\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right| > \epsilon \left|\overline{Z}^{1:(n-1)}\right.\right] \\
&\leq \mathbb{P}\left[\sup_{f\in\mathcal{L},\ \|f\|_{\mathcal{L}}\leq 1} \left|\mathbb{E}\left[f(Z)\right] - \mathbb{E}_{\overline{Z}^n}(f)\right| > \frac{\epsilon}{6K_\infty^{n-1}K_{\mathcal{L}}^{n-1}} \left|\overline{Z}^{1:(n-1)}\right.\right] \\
&\leq \mathbb{P}\left[\sup_{f\in\mathcal{L},\ \|f\|_{\mathcal{L}}\leq 1} \left|\mathbb{E}\left[f(Z)\right] - \mathbb{E}_{\overline{Z}^n}(f)\right| > \frac{\epsilon}{6K_\infty^{n-1}K_{\mathcal{L}}^{n-1}} \left|\overline{Z}^{1:(n-1)}\right.\right] \\
&\leq Ce^{-cM_n\phi\left(\frac{\epsilon}{6K_\infty^{n-1}K_{\mathcal{L}}^{n-1}}\right)}.
\end{aligned}$$

Hence the result. □

Lemma 3.13. *Let $0 < \gamma < 1$ and $n \geq 2$. Then, it holds that almost surely*

$$\mathbb{P}\left[\mathcal{G}_n | \overline{Z}^{1:(n-1)}\right] \geq 1 - \delta_n.$$

Proof. Since $(\overline{g}_1, \cdots, \overline{g}_n)$ forms a basis of $\overline{V}_{n-1}$, for all $\mu \in \mathcal{P}$, there exists one unique minimizer to

$$\min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \mathrm{Var}\left[g_\mu(Z) - \sum_{i=1}^{n-1}\lambda_i \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right].$$

Let $\overline{\lambda}^n := \left(\overline{\lambda}_i^n\right)_{1\leq i\leq n-1} \in \mathbb{R}^{n-1}$ be the unique minimizer of

$$\overline{\lambda}^n := \operatorname*{argmin}_{\lambda:=(\lambda_i)_{1\leq i\leq n-1}} \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\lambda_i \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]. \tag{24}$$

As a consequence, it holds that

$$\overline{\sigma}_{n-1}(\mathcal{M}) = \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]$$

where $\overline{\sigma}_{n-1}(\mathcal{M})$ is defined by (12).

Let $\widehat{\mu}_n \in \mathcal{P}$ such that

$$\widehat{\mu}_n \in \operatorname*{argmax}_{\mu\in\mathcal{P}} \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \mathrm{Var}\left[g_\mu(Z) - \sum_{i=1}^{n-1}\lambda_i \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right],$$

so that

$$\widehat{\sigma}_{n-1}(\mathcal{M}) = \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\lambda_i \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right],$$

where $\widehat{\sigma}_{n-1}(\mathcal{M})$ is defined in (11).

Let $\widehat{\lambda}^n := \left(\widehat{\lambda}_i^n\right)_{1\leq i\leq n-1} \in \mathbb{R}^{n-1}$ the unique minimizer of

$$\widehat{\lambda}^n := \operatorname*{argmin}_{\lambda:=(\lambda_i)_{1\leq i\leq n-1}} \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\lambda_i \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right], \tag{25}$$

so that

$$\widehat{\sigma}_{n-1}(\mathcal{M}) = \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right].$$

The event $\mathcal{G}_n$ holds if and only if

$$\mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \geq \gamma^2 \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]. \tag{26}$$

Let us begin by pointing out that, since

$$\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \geq \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right],$$

if the inequality

$$\begin{aligned}&\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]\\ &\leq (1-\gamma^2)\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]\end{aligned} \tag{27}$$

holds, then (26) is necessarily statisfied. The rest of the proof consists in estimating the probability that (27) is realized.

To this aim, as a first step, we are going to prove an upper bound on

$$\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z)-\sum_{i=1}^{n-1}\widehat{\lambda}_i^n\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]-\mathrm{Var}\left[g_{\overline{\mu}_n}(Z)-\sum_{i=1}^{n-1}\overline{\lambda}_i^n\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]$$

as a function of

$$\eta:=\sup_{g,h\in\mathcal{M}_{n-1}}\left|\mathrm{Cov}\left[g(Z),h(Z)|\,\overline{Z}^{1:\infty}\right]-\mathrm{Cov}_{\overline{Z}^n}\left(g,h\right)\right|, \tag{28}$$

which is the quantity estimated in Lemma 3.12. More precisely, let us now prove that

$$\begin{aligned}&\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z)-\sum_{i=1}^{n-1}\widehat{\lambda}_i^n\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]-\mathrm{Var}\left[g_{\overline{\mu}_n}(Z)-\sum_{i=1}^{n-1}\overline{\lambda}_i^n\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]\\&\leq n\left(2+\left(\frac{K_2+\sqrt{n-1}\eta}{1-(n-1)\eta}\right)^2+K_2^2\right)\eta.\end{aligned} \tag{29}$$

It holds that for all $1\leq i\leq n-1$, from (24) and (25),

$$\widehat{\lambda}_i^n=\mathrm{Cov}\left[g_{\widehat{\mu}_n}(Z),\overline{g}_i(Z)|\overline{Z}^{1:\infty}\right]\quad\text{ and }\overline{\lambda}_i^n=\mathrm{Cov}\left[g_{\overline{\mu}_n}(Z),\overline{g}_i(Z)|\overline{Z}^{1:\infty}\right],$$

and it then holds that, almost surely,

$$\max\left(\|\widehat{\lambda}^n\|_{\ell^2},\|\overline{\lambda}^n\|_{\ell^2}\right)\leq\max\left(\|g_{\widehat{\mu}_n}\|,\|g_{\overline{\mu}_n}\|\right)\leq K_2, \tag{30}$$

where $\|\cdot\|_{\ell^2}$ denotes the Euclidean norm of $\mathbb{R}^{n-1}$. Let now $\widehat{\lambda}^{n,n}:=\left(\widehat{\lambda}_i^{n,n}\right)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}$ be a minimizer of

$$\widehat{\lambda}^{n,n}:=\underset{\lambda:=(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}{\mathrm{argmin}}\ \mathrm{Var}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n}-\sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right),$$

and $\overline{\lambda}^{n,n}:=\left(\overline{\lambda}_i^{n,n}\right)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}$ be a minimizer of

$$\overline{\lambda}^{n,n}:=\underset{\lambda:=(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}{\mathrm{argmin}}\ \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n}-\sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)$$

It then holds that for all $1\leq i\leq n-1$, $\widehat{\lambda}^{n,n}$ and $\overline{\lambda}^{n,n}$ are solution to the linear systems

$$A^n\widehat{\lambda}^{n,n}=\widehat{b}^n\quad\text{ and }A^n\overline{\lambda}^{n,n}=\overline{b}^n,$$

where $A^n:=\left(A_{ij}^n\right)_{1\leq i,j\leq n-1}\in\mathbb{R}^{(n-1)\times(n-1)}$, $\widehat{b}^n:=\left(\widehat{b}_i^n\right)_{1\leq i\leq n-1}$, $\overline{b}^n:=\left(\overline{b}_i^n\right)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}$ are defined as follows: for all $1\leq i,j\leq n-1$,

$$A_{ij}^n=\mathrm{Cov}_{\overline{Z}^n}\left(\overline{g}_i,\overline{g}_j\right),\ \widehat{b}_i^n=\mathrm{Cov}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n},\overline{g}_i\right)\ \text{ and }\overline{b}_i^n=\mathrm{Cov}_{\overline{Z}^n}\left(g_{\overline{\mu}_n},\overline{g}_i\right).$$

Then, it holds that, almost surely,

$$\max_{1\leq i\leq n-1}\left(\left|\widehat{b}_i^n-\widehat{\lambda}_i^n\right|,\left|\overline{b}_i^n-\overline{\lambda}_i^n\right|\right)\leq\eta,$$

which implies that

$$\max\left(\|\widehat{b}_n\|_{\ell^2},\|\overline{b}_n\|_{\ell^2}\right)\leq K_2+\sqrt{n-1}\eta.$$

Moreover, we have

$$\max_{1\leq i,j\leq n-1}\left|A_{ij}^n-\delta_{ij}\right|\leq\eta,$$

which yields that for all $\xi\in\mathbb{R}^{n-1}$,

$$(1-(n-1)\eta)\|\xi\|_{\ell^2}^2\leq\xi^TA^n\xi\leq(1+(n-1)\eta)\|\xi\|_{\ell^2}^2.$$

Assume for now that $\eta(n-1) < 1$, this implies that, for all $\xi \in \mathbb{R}^{n-1}$,

$$\|(A^n)^{-1}\xi\|_{\ell^2} \leq \frac{1}{1-(n-1)\eta}\|\xi\|_{\ell^2}. \tag{31}$$

Using (31), we obtain that

$$\max\left(\|\overline{\lambda}^{n,n}\|_{\ell^2}, \|\widehat{\lambda}^{n,n}\|_{\ell^2}\right) \leq \frac{K_2 + \sqrt{n-1}\eta}{1-(n-1)\eta}. \tag{32}$$

We then have,

$$\begin{aligned}
&\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \\
&= \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n} \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \\
&\quad + \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n} \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n} - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n}\overline{g}_i\right) \\
&\quad + \mathrm{Var}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n} - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n}\overline{g}_i\right) - \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n,n}\overline{g}_i\right) \\
&\quad + \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n,n}\overline{g}_i\right) - \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n}\overline{g}_i\right) \\
&\quad + \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n}\overline{g}_i\right) - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n}\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right].
\end{aligned}$$

Using the fact that

$$\begin{aligned}
&\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n} \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \leq 0, \\
&\mathrm{Var}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n} - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n}\overline{g}_i\right) - \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n,n}\overline{g}_i\right) \leq 0, \\
&\mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n,n}\overline{g}_i\right) - \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^{n}\overline{g}_i\right) \leq 0,
\end{aligned}$$

from the definition of $\widehat{\lambda}^n$, $\widehat{\lambda}^{n,n}$, $\overline{\lambda}^{n,n}$, $\overline{\mu}_n$, we obtain that

$$\begin{aligned}
&\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \\
&\leq \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n} \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n} - \sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n}\overline{g}_i\right) \\
&\quad + \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n} - \sum_{i=1}^{n-1}\overline{\lambda}_i^n\overline{g}_i\right) - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right], \\
&= \mathrm{Var}\left[g_{\widehat{\mu}_n}(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n}\right) - 2\sum_{i=1}^{n-1}\widehat{\lambda}_i^{n,n}\left(\mathrm{Cov}\left[g_{\widehat{\mu}_n}(Z), \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Cov}_{\overline{Z}^n}\left(g_{\widehat{\mu}_n}, \overline{g}_i\right)\right) \\
&\quad + \sum_{i,j=1}^{n-1}\widehat{\lambda}_i^{n,n}\widehat{\lambda}_j^{n,n}\left(\mathrm{Cov}\left[\overline{g}_i(Z), \overline{g}_j(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Cov}_{\overline{Z}^n}\left(\overline{g}_i, \overline{g}_j\right)\right) \\
&\quad + \mathrm{Var}_{\overline{Z}^n}\left(g_{\overline{\mu}_n}\right) - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z)\middle|\overline{Z}^{1:\infty}\right] - 2\sum_{i=1}^{n-1}\overline{\lambda}_i^n\left(\mathrm{Cov}_{\overline{Z}^n}\left(g_{\overline{\mu}_n}, \overline{g}_i\right) - \mathrm{Cov}\left[\overline{g}_i(Z), \overline{g}_j(Z)\middle|\overline{Z}^{1:\infty}\right]\right) \\
&\quad + \sum_{i,j=1}^{n-1}\overline{\lambda}_i^n\overline{\lambda}_j^n\left(\mathrm{Cov}_{\overline{Z}^n}\left(\overline{g}_i, \overline{g}_j\right) - \mathrm{Cov}\left[g_{\overline{\mu}_n}(Z), \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]\right).
\end{aligned}$$

Now, using the definition of $\mathcal{M}^{n-1}$ given in (23), we obtain that

$$\begin{aligned}
&\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \\
&\leq \left(1 + 2\sum_{i=1}^{n-1}|\widehat{\lambda}_i^{n,n}| + \sum_{i,j=1}^{n-1}|\widehat{\lambda}_i^{n,n}||\widehat{\lambda}_j^{n,n}| + 1 + 2\sum_{i=1}^{n-1}|\overline{\lambda}_i^n| + \sum_{i,j=1}^{n-1}|\overline{\lambda}_i^n||\overline{\lambda}_j^n|\right) \\
&\quad \times \sup_{g,h\in\mathcal{M}_{n-1}}\left|\mathrm{Cov}\left[g(Z), h(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right|.
\end{aligned}$$

Since $\sup_{g,h\in\mathcal{M}_{n-1}}\left|\mathrm{Cov}\left[g(Z), h(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right| = \eta$, we then have, almost surely,

$$\begin{aligned}
&\mathrm{Var}\left[g_{\widehat{\mu}_n}(Z) - \sum_{i=1}^{n-1}\widehat{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Var}\left[g_{\overline{\mu}_n}(Z) - \sum_{i=1}^{n-1}\overline{\lambda}_i^n \overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right] \\
&\leq \left[\left(1 + \sum_{i=1}^{n-1}|\widehat{\lambda}_i^{n,n}|\right)^2 + \left(1 + \sum_{i=1}^{n-1}|\overline{\lambda}_i^n|\right)^2\right]\sup_{g,h\in\mathcal{M}_{n-1}}\left|\mathrm{Cov}\left[g(Z), h(Z)\middle|\overline{Z}^{1:\infty}\right] - \mathrm{Cov}_{\overline{Z}^n}(g,h)\right| \\
&\leq n\left(2 + \sum_{i=1}^{n-1}|\widehat{\lambda}_i^{n,n}|^2 + \sum_{i=1}^{n-1}|\overline{\lambda}_i^n|^2\right)\eta \\
&\leq n\left(2 + \|\widehat{\lambda}^{n,n}\|_{\ell^2}^2 + \|\overline{\lambda}^n\|_{\ell^2}^2\right)\eta.
\end{aligned}$$

Finally, using (30) and (32), we obtain (29), i.e.

$$\widehat{\sigma}_{n-1}(\mathcal{M}) - \overline{\sigma}_{n-1}(\mathcal{M}) \leq n\left(2 + \left(\frac{K_2 + \sqrt{n-1}\eta}{1-(n-1)\eta}\right)^2 + K_2^2\right)\eta.$$

Let us now evaluate the probability, conditioned to the knowledge of $\overline{Z}^{1:\infty}$, that

$$n\left(2 + \left(\frac{K_2 + \sqrt{n-1}\eta}{1-(n-1)\eta}\right)^2 + K_2^2\right)\eta \leq (1-\gamma^2)\widehat{\sigma}_{n-1}(\mathcal{M}). \tag{33}$$

If η is chosen to be smaller that $\frac{1}{2(n-1)}$, then it holds that

$$2+\left(\frac{K_2+\sqrt{n-1}\eta}{1-(n-1)\eta}\right)^2+K_2^2\leq 2+(2K_2+1)^2+K_2^2\leq 9K_2^2+4.$$

A sufficient condition for (33) to hold is then to ensure that $\eta\leq\epsilon$ with

$$\epsilon:=\min\left(\frac{1}{2(n-1)},\quad\frac{(1-\gamma^2)\widehat{\sigma}_{n-1}^2(\mathcal{M})}{n\left(9K_2^2+4\right)}\right),$$

Then, it holds that

$$\begin{aligned}\mathbb{P}\left[\mathcal{G}_n\left|\overline{Z}^{1:\infty}\right.\right]&=\mathbb{P}\left[\overline{\sigma}_{n-1}(\mathcal{M})^2\geq\gamma^2\widehat{\sigma}_{n-1}(\mathcal{M})^2\left|\overline{Z}^{1:\infty}\right.\right]\\&=\mathbb{P}\left[\widehat{\sigma}_{n-1}(\mathcal{M})^2-\overline{\sigma}_{n-1}(\mathcal{M})^2\leq(1-\gamma^2)\widehat{\sigma}_{n-1}(\mathcal{M})^2\left|\overline{Z}^{1:\infty}\right.\right]\\&\geq\mathbb{P}\left[n\left(2+\left(\frac{K_2+\sqrt{n-1}\eta}{1-(n-1)\eta}\right)^2+K_2^2\right)\eta\leq(1-\gamma^2)\widehat{\sigma}_{n-1}(\mathcal{M})\left|\overline{Z}^{1:\infty}\right.\right]\\&\geq\mathbb{P}\left[\eta\leq\epsilon\left|\overline{Z}^{1:\infty}\right.\right].\end{aligned}$$

Thus, using the definition of η given by (28) and applying Lemma 3.12, we then obtain that

$$\begin{aligned}\mathbb{P}\left[\mathcal{G}_n|\overline{Z}^{1:(n-1)}\right]&\geq\mathbb{P}\left[\eta\leq\epsilon\left|\overline{Z}^{1:\infty}\right.\right]\\&=\mathbb{P}\left[\sup_{g,h\in\mathcal{M}^{n-1}}\left|\mathrm{Cov}\left[g(Z),h(Z)\left|\overline{Z}^{1:\infty}\right.\right]-\mathrm{Cov}_{\overline{Z}^n}(g,h)\right|\leq\epsilon\left|\overline{Z}^{1:(n-1)}\right.\right]\\&\geq 1-\delta_n,\end{aligned}$$

since

$$Ce^{-cM_n\phi(\kappa_{n-1})}\leq\delta_n,$$

with

$$\kappa_{n-1}:=\frac{\min\left(\frac{1}{2(n-1)},\quad\frac{(1-\gamma^2)\widehat{\sigma}_{n-1}^2(\mathcal{M})}{n\left(9K_2^2+4\right)}\right)}{6K_\infty^{n-1}K_{\mathcal{L}}^{n-1}},$$

which yields the desired result. □

We are now in position to end the proof of Theorem 3.6.

Proof of Theorem 3.6. Collecting Lemma 3.11 and Lemma 3.13, we obtain (18) for all $n\in\mathbb{N}^*$. Let us now prove (19).

Let us first prove by recursion that for all $n\in\mathbb{N}^*$,

$$\mathbb{P}\left[\bigcap_{k=1}^n\mathcal{G}_k\right]\geq\Pi_{k=1}^n(1-\delta_k).\tag{34}$$

Using Lemma 3.11, it holds that (34) is true for $n=1$. Now we turn to the proof of the recursion. Let $n\in\mathbb{N}^*$. For any event $\mathcal{Z}$, we denote by $\mathbb{1}_{\mathcal{Z}}$ the random variable which is equal to 1 if $\mathcal{Z}$ is realized and 0 if not. Using the fact that $\bigcap_{k=1}^n\mathcal{G}_k$ is measurable with respect to $\overline{Z}^{1:n}$, it holds that

$$\begin{aligned}\mathbb{P}\left[\bigcap_{k=1}^{n+1}\mathcal{G}_k\right]&=\mathbb{E}\left[\mathbb{1}_{\bigcap_{k=1}^{n+1}\mathcal{G}_k}\right]\\&=\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\mathcal{G}_{n+1}}\mathbb{1}_{\bigcap_{k=1}^n\mathcal{G}_k}\left|\overline{Z}^{1:n}\right.\right]\right]\\&=\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\mathcal{G}_{n+1}}\left|\overline{Z}^{1:n}\right.\right]\mathbb{1}_{\bigcap_{k=1}^n\mathcal{G}_k}\right]\\&=\mathbb{E}\left[\mathbb{P}\left[\mathcal{G}_{n+1}\left|\overline{Z}^{1:n}\right.\right]\mathbb{1}_{\bigcap_{k=1}^n\mathcal{G}_k}\right].\end{aligned}$$

Now using Lemma 3.13, it holds that almost surely $\mathbb{P}\left[\mathcal{G}_{n+1}\left|\overline{Z}^{1:n}\right.\right] \geq 1-\delta_{n+1}$. Hence, it holds that

$$\mathbb{P}\left[\bigcap_{k=1}^{n+1}\mathcal{G}_k\right] \geq (1-\delta_{n+1})\mathbb{E}\left[\mathbb{1}_{\bigcap_{k=1}^{n}\mathcal{G}_k}\right] = (1-\delta_{n+1})\mathbb{P}\left[\bigcap_{k=1}^{n}\mathcal{G}_k\right].$$

The recursion is thus proved, together with (34), which implies (19).

If $\bigcap_{n\in\mathbb{N}}\mathcal{G}_n$ is realised, it then holds that the MC-greedy algorithm is a weak greedy algorithm with parameter γ and norm $\|\cdot\|_{\overline{Z}^{1:\infty}} = \sqrt{\mathrm{Var}\left[\cdot|\overline{Z}^{1:\infty}\right]}$.

□

4 Numerical results

The aim of this section is to compare several procedures to choose the value of the sequence $(M_n)_{n\in\mathbb{N}^*}$ in the MC-greedy algorithm presented in Section 3.1.

4.1 Three numerical procedures

As mentioned in Remark 3.9, it is possible to design a constructive way to define a sequence of numbers of samples $(M_n)_{n\in\mathbb{N}^*}$ which satisfies assumptions of Theorem 3.6, and thus which guarantees that the corresponding MC-greedy algorithm is a weak-greedy algorithm with high probability. Unfortunately, it appears that such a procedure cannot be used in practice to design a variance reduction method since the values of the sequence $(M_n)_{n\in\mathbb{N}^*}$ increases too sharply. The objective of this section is to propose a *heuristic* procedure to choose a sequence $(M_n)_{n\in\mathbb{N}^*}$ for an MC-greedy algorithm. This heuristic procedure appears to yield a reduced basis approximation $\overline{f}_\mu$ of f_μ which provides very satisfactory results in terms of variance reduction, at least on the different test cases presented below.

We use here the same notation as in Section 2.1 and consider $M_{\mathrm{ref}} \in \mathbb{N}^*$ such that $M_{\mathrm{ref}} \gg 1$. The idea of this heuristic method is the following: assume that the sequence $(M_n)_{n\in\mathbb{N}^*}$ can be chosen so that for all $n \in \mathbb{N}^*$, the inequality

$$\begin{aligned}&\left|\inf_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}\mathrm{Var}\left[g_\mu(Z)-\sum_{i=1}^{n-1}\lambda_i\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]-\inf_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}\mathrm{Var}_{\overline{Z}^n}\left(g_\mu-\sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)\right|\\ &\leq(1-\gamma^2)\inf_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}\mathrm{Var}\left[g_\mu(Z)-\sum_{i=1}^{n-1}\lambda_i\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]\end{aligned} \tag{35}$$

holds for all $\mu \in \mathcal{P}$. Then, it can easily be checked that such an MC-greedy algorithm is a weak greedy algorithm with parameter γ. Of course, such an algorithm could not be of any use for variance reduction since it would imply the computation of $\inf_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}\mathrm{Var}\left[g_\mu(Z)-\sum_{i=1}^{n-1}\lambda_i\overline{g}_i(Z)\middle|\overline{Z}^{1:\infty}\right]$ (or an approximation of this quantity of the form $\inf_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}\mathrm{Var}_{\overline{Z}^{\mathrm{ref}}}\left(g_\mu-\sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)$ with $\overline{Z}^{\mathrm{ref}} = \left(Z_k^{\mathrm{ref}}\right)_{1\leq k\leq M^{\mathrm{ref}}}$ a collection of iid random variables with the same law as Z and independent of Z) for all $\mu \in \mathcal{P}$.

The idea of the heuristic procedure is then to check if the inequality (35) holds, *only for the value* $\mu = \overline{\mu}_n$. If the inequality holds, the value of M_{n+1} is chosen to be equal to M_n for the next iteration. Otherwise, the value of M_n is increased and the nth iteration of the MC-greedy algorithm is performed again.

This heuristic procedure leads to the Heuristic MC-greedy algorithm (or HMC-greedy algorithm), see Algorithm 1. Notice that we introduce here $\mathcal{P}_{\mathrm{trial}}$ a finite subset of $\mathcal{P}$, which is classically called the trial set of parameters in reduced basis methods.

For the sake of comparison, we introduce two other algorithms, which cannot be implemented in practice, but which will allow us to compare the performance of the HMC-greedy algorithm with ideal procedures. The first method, called SHMC-greedy algorithm and also presented in Algorithm 1 as a variant, consists in designing the sequence $(M_n)_{n\in\mathbb{N}^*}$ in order to ensure that the inequality (35) is satisfied for all $\mu \in \mathcal{P}_{\mathrm{trial}}$ (and not only for $\overline{\mu}_n$). The second one consists in performing an *ideal* MC-greedy algorithm, called herefater

IMC-greedy algorithm, see Algorithm 2, where all the variances and expectations are evaluated using M_{ref} number of samples of the random variable Z at each iteration of the MC-greedy algorithm.

Let us comment on the termination criterion

$$\frac{\mathrm{Var}_{\overline{Z}^{\text{ref}}}\left(\overline{f}_{\mu_{n-1}^{(S)H}}\right)}{M_{\text{ref}}} > \frac{\mathrm{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_{n-1}^{(S)H}} - \overline{f}_{\mu_{n-1}^{(S)H}}\right)}{M_{n-1}}$$

introduced in line 11 of the (S)HMC-greedy algorithm. Recall that, for $\mu = \mu_{n-1}^{(S)H}$, the expectation $\mathbb{E}\left[f_{\mu_{n-1}^{(S)H}}(Z)\right]$ is approximated after $n-1$ iterations of the greedy algorithm by the control variate formula (see (1))

$$\mathbb{E}_{\overline{Z}^{\text{ref}}}(\overline{f}_{\mu_{n-1}^{(S)H}}) + \mathbb{E}_{\overline{Z}^{n-1}}\left(f_{\mu_{n-1}^{(S)H}} - \overline{f}_{\mu_{n-1}^{(S)H}}\right). \tag{36}$$

This criterion ensures that the iterative scheme ends when the statistical error associated with the second term in (36) becomes smaller than the statistical error of the first term (see Remark 2.1).

4.2 Definitions of quantities of interest

For each of the test cases presented below, we plot different quantities of interest which we define here.

For all $n \in \mathbb{N}^*$, we denote by $\mu_1^H, \cdots, \mu_n^H$ (respectively $\mu_1^{SH}, \cdots, \mu_n^{SH}$ and $\mu_1^I, \cdots, \mu_n^I$) the set of parameter values selected after n iterations of the HMC-greedy (respectively SHMC-greedy and IMC-greedy) algorithm. We also denote by $\overline{V}_n^H := \mathrm{Span}\left\{g_{\mu_1^H}, \cdots, g_{\mu_n^H}\right\}$, $\overline{V}_n^{SH} := \mathrm{Span}\left\{g_{\mu_1^{SH}}, \cdots, g_{\mu_n^{SH}}\right\}$ and $\overline{V}_n^I := \mathrm{Span}\left\{g_{\mu_1^I}, \cdots, g_{\mu_n^I}\right\}$.

For all $\mu \in \mathcal{P}$ and $n \in \mathbb{N}^*$, we define for the three algorithms presented in Section 4.1,

$$\theta_n(\mu) := \inf_{(\lambda_i)_{1\le i\le n-1}\in\mathbb{R}^{n-1}} \sqrt{\mathrm{Var}_{\overline{Z}^{\text{ref}}}\left(g_\mu - \sum_{i=1}^{n-1}\lambda_i \overline{g}_i\right)} \tag{37}$$

and

$$\theta_n := \sup_{\mu\in\mathcal{P}_{\text{trial}}} \theta_n(\mu). \tag{38}$$

In what follows, we denote by $\theta_n^H(\mu)$ and θ_n^H (respectively by $\theta_n^{SH}(\mu)$, θ_n^{SH}, $\theta_n^I(\mu)$ and θ_n^I) the quantities defined by (37) and (38) obtained with the HMC-greedy (respectively the SHMC-greedy and IMC-greedy) algorithm. Note that, by definition of the IMC-greedy algorithm, $\theta_n^I = \theta_n^I(\mu_n^I)$.

A second quantity of interest for the HMC-greedy and the SHMC-greedy algorithms is given, for all $n \in \mathbb{N}^*$ and $\mu \in \mathcal{P}$, by

$$\beta_n(\mu) := \inf_{(\lambda_i)_{1\le i\le n-1}\in\mathbb{R}^{n-1}} \sqrt{\mathrm{Var}_{\overline{Z}^{n}}\left(g_\mu - \sum_{i=1}^{n-1}\lambda_i \overline{g}_i\right)}. \tag{39}$$

and

$$\beta_n := \sup_{\mu\in\mathcal{P}_{\text{trial}}} \beta_n(\mu). \tag{40}$$

In the sequel, we denote by β_n^H (respectively by β_n^{SH}) the quantity defined by (40) obtained with the HMC-greedy (respectively the SHMC-greedy) algorithm.

Let us point out that when $\mathcal{P}_{\text{trial}} = \mathcal{P}$ and when $M_{\text{ref}} = \infty$, it holds that $\theta_n = \widehat{\sigma}_{n-1}(\mathcal{M})$ and $\theta_n(\overline{\mu}_n) = \overline{\sigma}_{n-1}(\mathcal{M})$ where $\widehat{\sigma}_{n-1}(\mathcal{M})$ and $\overline{\sigma}_{n-1}(\mathcal{M})$ are defined respectively in (11) and (12).

We finally wish to evaluate the error made on the approximation of $\mathbb{E}[f_\mu(Z)]$ obtained by using the variance reduction method based on these MC-greedy algorithm. More precisely, this approximation is computed as

$$\sum_{i=1}^{n-1}\lambda_i^{n,\mu}\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\overline{\mu}_i}) + \mathbb{E}_{\overline{Z}^{n}}\left(f_\mu - \sum_{i=1}^{n-1}\lambda_i^{n,\mu} f_{\overline{\mu}_i}\right)$$

where

$$(\lambda_i^{n,\mu})_{1\le i\le n-1} = \underset{(\lambda_i)_{1\le i\le n-1}\in\mathbb{R}^{n-1}}{\mathrm{argmin}} \mathrm{Var}_{\overline{Z}^{n}}\left(f_\mu - \sum_{i=1}^{n-1}\lambda_i f_{\overline{\mu}_i}\right)$$

Algorithm 1: (S)HMC-greedy algorithm

input : $\gamma > 0$, $\epsilon > 0$, $M_1 \in \mathbb{N}^*$, $\mathcal{P}_{\text{trial}}$ trial set of parameters (finite subset of $\mathcal{P}$), $M_{\text{ref}} \in \mathbb{N}^*$ (high fidelity sampling number, which has a vocation to satisfy $M_{\text{ref}} \gg M_1$.

output: $N \in \mathbb{N}^*$ size of the reduced basis, $\mu_1^{(S)H}, \mu_2^{(S)H}, \cdots, \mu_N^{(S)H} \in \mathcal{P}_{\text{trial}}$, $\left(\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_n^{(S)H}})\right)_{1\leq n\leq N}$.

Let $\overline{Z}^{\text{ref}} := \left(Z_k^{\text{ref}}\right)_{1\leq k\leq M_{\text{ref}}}$ be a collection of M_{ref} iid random variables with the same law as Z and independent of Z.

Set $R^1 = 1$.

while $R^1 \geq 1-\gamma^2$ **do**

  Let $\overline{Z}^1 := (Z_k^1)_{1\leq k\leq M_1}$ be a collection of M_1 iid random variables with the same law as Z and independent of Z and $\overline{Z}^{\text{ref}}$.

  Compute $\mu_1^{(S)H} \in \underset{\mu\in\mathcal{P}_{\text{trial}}}{\text{argmax}}\, \text{Var}_{\overline{Z}^1}(f_\mu)$.

  Compute $\overline{f}_{\mu_1^{(S)H}} = 0$.

  Compute $\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_1^{(S)H}})$.

  - **HMC case:** Set $\theta_1^H(\mu_1^H) = \sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_1^H}\right)}$ and $\beta_1^H(\mu_1^H) = \sqrt{\text{Var}_{\overline{Z}^1}\left(f_{\mu_1^H}\right)}$. Set $R^1 = \frac{\left|\theta_1^H(\mu_1^H)^2 - \beta_1^H(\mu_1^H)^2\right|}{\theta_1^H(\mu_1^H)^2}$.
  - **SHMC case:** Set $\theta_1^{SH}(\mu) = \sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}(f_\mu)}$ and $\beta_1^{SH}(\mu) = \sqrt{\text{Var}_{\overline{Z}^1}(f_\mu)}$ for all $\mu \in \mathcal{P}_{\text{trial}}$. Set $R^1 = \sup_{\mu\in\mathcal{P}_{\text{trial}}} \frac{\left|\theta_1^{SH}(\mu)^2 - \beta_1^{SH}(\mu)^2\right|}{\theta_1^{SH}(\mu)^2}$.

  if $R^1 \geq 1-\gamma^2$ **then**

    Set $b_1 := 1.1$ and $M_1 = \lceil b_1 M_1 + 1\rceil$.

  end

end

Compute $\overline{g}^1 = \frac{f_{\mu_1^{(S)H}} - \mathbb{E}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_1^{(S)H}}\right)}{\theta_1^{(S)H}(\mu_1^{(S)H})}$. Set $n = 2$ and $M_n = M_1$.

while $\frac{\text{Var}_{\overline{Z}^{\text{ref}}}\left(\overline{f}_{\mu_{n-1}^{(S)H}}\right)}{M_{\text{ref}}} \leq \frac{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_{n-1}^{(S)H}} - \overline{f}_{\mu_{n-1}^{(S)H}}\right)}{M_{n-1}}$ **do**

  Set $R^n = 1$.

  while $R^n \geq 1-\gamma^2$ **do**

    Let $\overline{Z}^n := (Z_k^n)_{1\leq k\leq M_n}$ be a collection of M_n iid random variables with the same law as Z and independent of Z and $\overline{Z}^{\text{ref}}$.

    Compute $\mu_n^{(S)H} \in \underset{\mu\in\mathcal{P}_{\text{trial}}}{\text{argmax}} \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \text{Var}_{\overline{Z}^n}\left(f_\mu - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)$

    Compute $(\overline{\lambda}_i^n)_{1\leq i\leq n-1} = \text{argmin}_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_n^{(S)H}} - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)}$.

    Compute $\overline{f}_{\mu_n^{(S)H}} = \sum_{i=1}^{n-1}\overline{\lambda}_i^n\overline{g}_i$.

    - **HMC case:** Compute $\theta_n^H(\mu_n^H) = \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_n^{(S)H}} - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)} = \sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_n^{(S)H}} - \overline{f}_{\mu_n^{(S)H}}\right)}$ and $\beta_n^H(\mu_n^H) = \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \sqrt{\text{Var}_{\overline{Z}^n}\left(f_{\mu_n^H} - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)}$. Set $R^n = \left|\frac{\theta_n^H(\mu_n^H)^2 - \beta_n^H(\mu_n^H)^2}{\theta_n^H(\mu_n^H)^2}\right|$.
    - **SHMC case:** For all $\mu \in \mathcal{P}_{\text{trial}}$, compute $\theta_n^{SH}(\mu) = \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_\mu - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)}$ and $\beta_n^{SH}(\mu) = \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \sqrt{\text{Var}_{\overline{Z}^n}\left(f_\mu - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)}$. Set $R^n = \sup_{\mu\in\mathcal{P}_{\text{trial}}} \left|\frac{\theta_n^{SH}(\mu)^2 - \beta_n^{SH}(\mu)^2}{\theta_n^{SH}(\mu)^2}\right|$.

    if $R^n \geq 1-\gamma^2$ **then**

      Compute $r_n := \frac{\phi(\theta_{n-1}^{(S)H}(\mu_{n-1}^{(S)H})^2)}{\phi\left(\theta_n^{(S)H}(\mu_n^{(S)H})^2\right)}$, $b_n := \max(1.1, r_n)$ and set $M_n = \lceil b_n M_n + 1\rceil$.

    end

  end

  Compute $\overline{g}_n = \frac{f_{\mu_n^{(S)H}} - \sum_{i=1}^{n-1}\overline{\lambda}_i^n\overline{g}_i}{\theta_n^{(S)H}(\mu_n^{(S)H})} = \frac{f_{\mu_n^{(S)H}} - \overline{f}_{\mu_n^{(S)H}}}{\theta_n^{(S)H}(\mu_n^{(S)H})}$ and $\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_n^{(S)H}})$.

  Set $M_{n+1} = M_n$ and $n = n+1$.

end

Set $N = n$, $M_N = M_n$.

Algorithm 2: IMC-greedy algorithm

input : $\epsilon > 0$, $\mathcal{P}_{\text{trial}}$ trial set of parameters (finite subset of $\mathcal{P}$), $M_{\text{ref}} \in \mathbb{N}^*$ (high fidelity sampling number).

output: $N \in \mathbb{N}^*$ size of the reduced basis, $\mu_1^I, \mu_2^I, \cdots, \mu_N^I \in \mathcal{P}_{\text{trial}}$, $\left(\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_n^I})\right)_{1\leq n\leq N}$.

Let $\overline{Z}^{\text{ref}} := \left(Z_k^{\text{ref}}\right)_{1\leq k\leq M_{\text{ref}}}$ be a collection of M_{ref} iid random variables with the same law as Z and independent of Z.

Compute $\mu_1^I \in \underset{\mu\in\mathcal{P}_{\text{trial}}}{\text{argmax}}\, \text{Var}_{\overline{Z}^{\text{ref}}}(f_\mu)$.

Set $\theta_1^I(\mu_1^I) := \text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_1^I}\right)$ and compute $\mathbb{E}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_1^I}\right)$.

Set $n = 2$, $M_n = M_1$, $\overline{g}^1 = \dfrac{f_{\mu_1^I}}{\sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_1^I}\right)}}$

while $\theta_{n-1}^I(\mu_{n-1}^I) \geq \epsilon$ **do**

  Compute $\mu_n^I \in \underset{\mu\in\mathcal{P}_{\text{trial}}}{\text{argmax}} \underset{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}{\min} \text{Var}_{\overline{Z}^{\text{ref}}}\left(f_\mu - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)$

  Compute $\theta_n^I(\mu_n^I) = \min_{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}} \sqrt{\text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_n^I} - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)}$

  Compute $(\overline{\lambda}_i^n)_{1\leq i\leq n-1} = \underset{(\lambda_i)_{1\leq i\leq n-1}\in\mathbb{R}^{n-1}}{\text{argmin}} \text{Var}_{\overline{Z}^{\text{ref}}}\left(f_{\mu_n^I} - \sum_{i=1}^{n-1}\lambda_i\overline{g}_i\right)$

  Compute $\overline{g}_n = \dfrac{f_{\mu_n^I} - \sum_{i=1}^{n-1}\overline{\lambda}_i^n\overline{g}_i}{\theta_n^I(\mu_n^I)}$ and $\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\mu_n^I})$. Set $n = n+1$.

end

Set $N = n$.

and $\overline{\mu}_i$ is equal to μ_i^H, μ_i^{SH} or μ_i^I depending on the chosen algorithm (remember formula (1)). This quantity has to be compared with the approximation obtained with a standard Monte-Carlo with M_{ref} samples, i.e. $\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_\mu)$. To this aim, for all $n \in \mathbb{N}^*$ and $\mu \in \mathcal{P}$, we define

$$e_n(\mu) := \frac{\left|\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_\mu) - \left[\sum_{i=1}^{n-1}\lambda_i^{n,\mu}\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_{\overline{\mu}_i}) + \mathbb{E}_{\overline{Z}^n}\left(f_\mu - \sum_{i=1}^{n-1}\lambda_i^{n,\mu}f_{\overline{\mu}_i}\right)\right]\right|}{\left|\mathbb{E}_{\overline{Z}^{\text{ref}}}(f_\mu)\right|}. \tag{41}$$

In what follows, we denote by $e_n^H(\mu)$ (respectively $e_n^{SH}(\mu)$) the quantity defined by (41) obtained by the HMC-greedy (respectively the SHMC-greedy) algorithm.

4.3 Explicit one-dimensional functions

We consider in this section two sets of one-dimensional explicit functions. The motivation for considering these two simple test cases is that the decays of the Kolmogorov n-widths of the associated sets $\mathcal{M}$ are known.

4.3.1 First test case

Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined such that

$$\forall x \in \mathbb{R}, \quad f(x) := \begin{cases} 2\,x & \text{if} \quad 0 \leq x \leq 0.5, \\ 1 & \text{if} \quad 0.5 \leq x \leq 1.5, \\ 4 - 2\,x & \text{if} \quad 1.5 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

Let $\mathcal{P} = [0, 3]$ be the set of parameter values. We consider in this first test case the family of functions $(f_\mu)_{\mu\in\mathcal{P}}$ such that $f_\mu(x) = f(x - \mu)$ for all μ in $\mathcal{P}$ and $x \in \mathbb{R}$. Let Z be a real-valued random variable with probability measure $\nu = \mathcal{U}(0, 5)$. In this case, it is known [11] that there exists a constant $c > 0$ such that $d_n(\mathcal{M}) \geq cn^{-1/2}$ for all $n \in \mathbb{N}^*$.

In this example, $M_1 = 10$, $M_{\text{ref}} = 10^5$, $\gamma = 0.9$ and the trial set $\mathcal{P}_{\text{trial}}$ is chosen to be a set of 300 random parameter values which were uniformly sampled in $\mathcal{P}$.

Figure 1 illustrates the evolution of the values of M_n as a function of n for the HMC and SHMC algorithms.

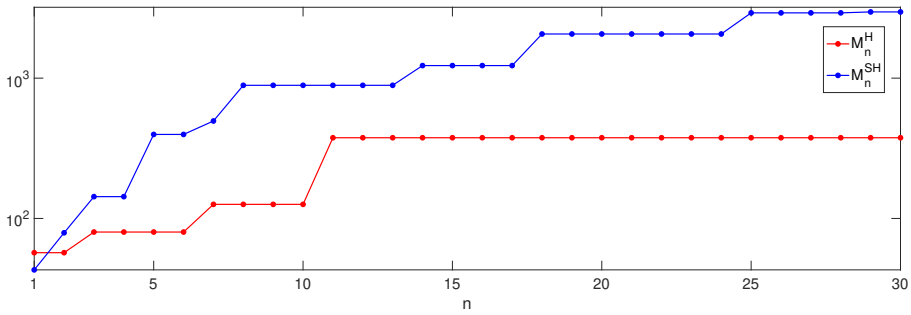

Figure 1: Evolution of M_n as a function of n for the HMC and SHMC-greedy algorithms in test case 1.

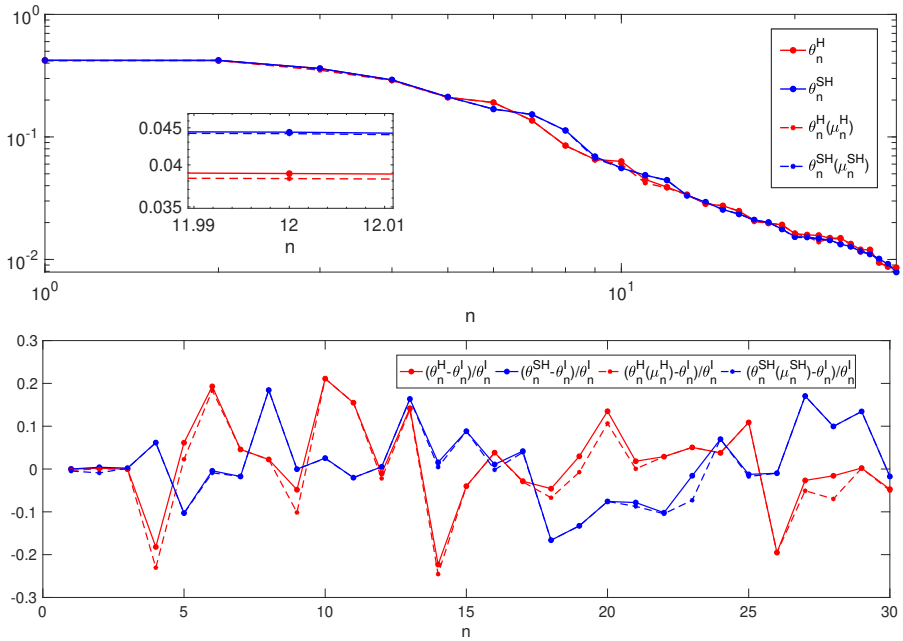

Figure 2: Evolution of $\theta_n^H(\mu_n^H)$, $\theta_n^{SH}(\mu_n^{SH})$, θ_n^H, θ_n^{SH} as a function of n in test case 1.

Figure 2 illustrates the fact that at each iteration $n \in \mathbb{N}^*$, for the (S)HMC-algorithm, the value of the selected parameter $\mu_n^{(S)H}$ is relevant since we observe numerically that $\theta^{(S)H}(\mu_n^{(S)H})$ is very close to $\theta_n^{(S)H} = \sup_{\mu \in \mathcal{P}_{\text{trial}}} \theta_n^{(S)H}(\mu)$. In addition, we observe that the resulting reduced spaces $\overline{V}_n^{(S)H}$ have very good approximability properties with respect to the set $\mathcal{M}$, in the sense that the values of $\theta_n^{(S)H}$ and $\theta_n^{(S)H}(\mu_n^{(S)H})$ are very close to θ_n^I, which is computed with the reference IMC algorithm.

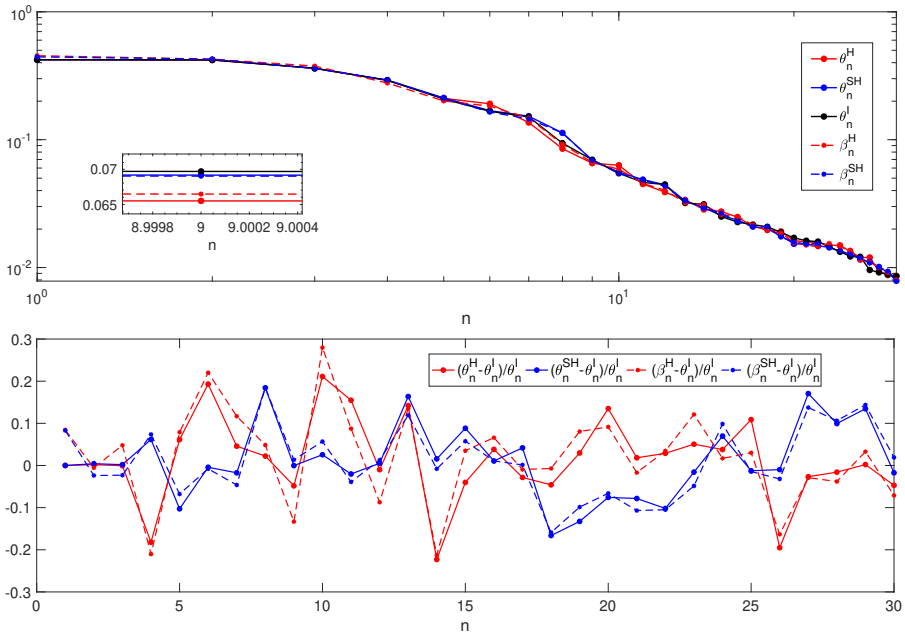

Figure 3: Evolution of β_n^H, β_n^{SH}, θ_n^H, θ_n^{SH}, θ_n^I as a function of n in test case 1.

Figure 3 illustrates the fact that the value of the number of samples M_n chosen at each iteration $n \in \mathbb{N}^*$ enables to compute empirical variances that are close to exact variances since the values of $\beta_n^{(S)H}$ are very close to the $\theta_n^{(S)H}$ for the (S)HMC-algorithm.

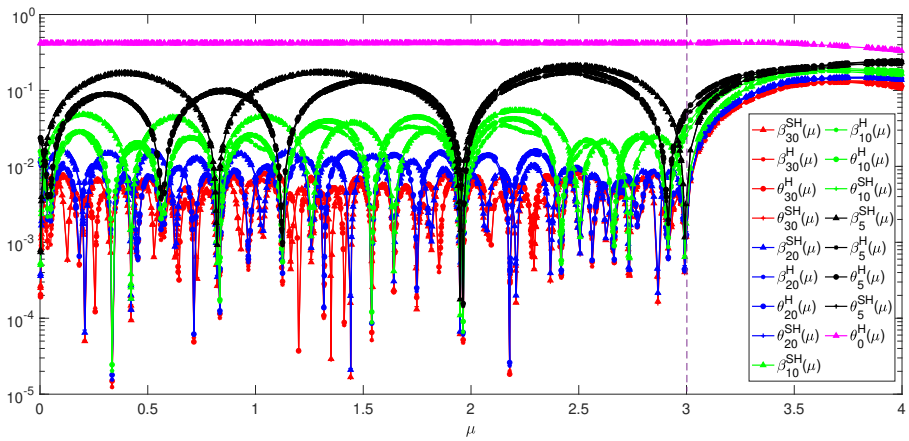

Figure 4: $\theta_n^H(\mu)$, $\theta_n^{SH}(\mu)$, $\beta_n^H(\mu)$, $\beta_n^{SH}(\mu)$ as a function of μ for $n = 0, 5, 10, 20, 30$ on $\mathcal{P}_{test} = [0, 4]$.

In Figure 4, the values of $\theta_n^{(S)H}(\mu)$ and $\beta_n^{(S)H}(\mu)$ are plotted as a function of $\mu \in \mathcal{P}_{test} = [0, 4]$ for different values of n ($n = 0, 5, 10, 20, 30$).

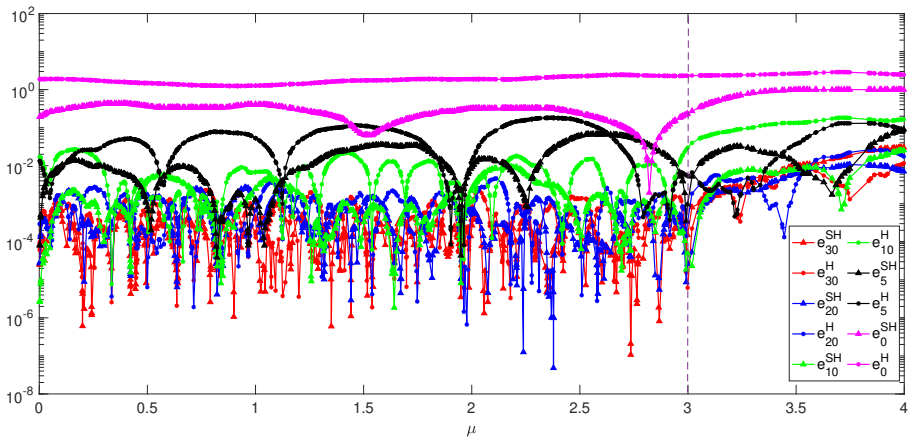

Figure 5: $e_n^H(\mu)$ and $e_n^{SH}(\mu)$ as a function of μ for $n = 0, 5, 10, 20, 30$ on $\mathcal{P}_{test} = [0, 4]$.

In comparison, in Figure 5, the relative error $e_n^{(S)H}(\mu)$ is plotted as a function of μ for $n = 0, 5, 10, 20, 30$. In particular, we observe that this error remains lower than 1% as soon as $n \geq 10$ on $\mathcal{P}$. Naturally, this error is larger for $\mu \in \mathcal{P}_{test} \setminus \mathcal{P}$.

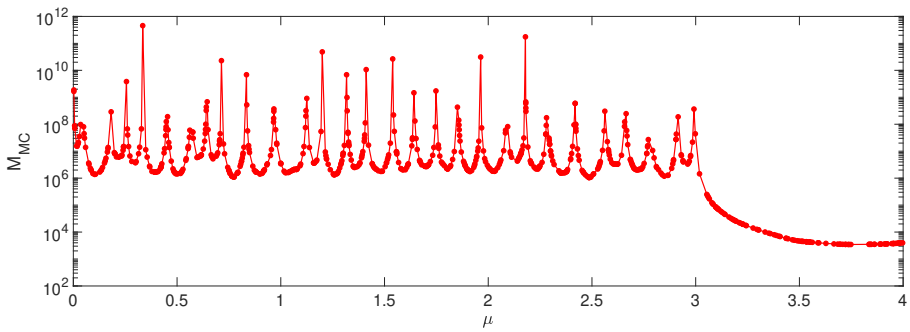

Figure 6: $M_{MC}(\mu)$ as a function of $\mu \in \mathcal{P}_{test} = [0, 4]$.

Finally, to illustrate the gain of our proposed method in terms of variance reduction, we plot on Figure 6 the value of the number of random Monte-Carlo samples $M_{MC}(\mu)$ that would have been necessary to compute an approximation of the mean of $f_\mu(Z)$ with a standard Monte-Carlo method with the same level of accuracy than the one given by the HMC-algorithm after $n = 30$ iterations. In this case, let us point out that $M_n = 349$. More precisely, we compute $M_{MC}(\mu)$ by the follwoing formula:

$$M_{MC}(\mu) = \frac{\mathrm{Var}_{\overline{Z}^{ref}}\left(f_\mu\right) \times M_n}{\mathrm{Var}_{\overline{Z}^n}\left(f_\mu - \sum_{i=1}^n \lambda_i^\mu f_{\overline{\mu}_i}\right)}. \tag{43}$$

Figure 6 illustrates that, for all $\mu \in \mathcal{P}$, the classical Monte Carlo method would have required a number of samples $M_{MC}(\mu)$ in the range $10^6 \leq M_{MC}(\mu) \leq 10^{12}$ in order to obtain the same level of statistical error. Thus, we see that the HMC-algorithm significantly improves the efficiency of the computation of the expectation of $f_\mu(Z)$ with respect to a standard Monte-Carlo algorithm.

4.3.2 Second test case

In this example, we consider a second family of one-dimensional functions where $\mathcal{P} = [0, 1]$ is the set of parameter values. More precisely, we consider the family of functions $(f_\mu)_{\mu \in \mathcal{P}}$ such that, for all μ in $\mathcal{P}$:

$$\forall x \in [0,1], \quad f_\mu(x) := \begin{cases} \sqrt{x + 0.1} & \text{if} \quad x \in [0, \mu] \\ \frac{1}{2}(\mu + 0.1)^{-\frac{1}{2}} x - \frac{1}{2}(\mu + 0.1)^{-\frac{1}{2}} \mu + (\mu + 0.1)^{\frac{1}{2}} & \text{if} \quad x \in [\mu, 1] \end{cases} \tag{44}$$

Let us point out that for all $\mu \in \mathcal{P}$, f_μ is a $\mathcal{C}^1$ function on $[0, 1]$. In this case, it is known [13] that there exists a constant $c > 0$ such that $d_n(\mathcal{M}) \leq cn^{-2}$ for all $n \in \mathbb{N}^*$.

Let Z be a random variable with probability measure $\nu = \mathcal{U}(0, 1)$.

In this example, $M_1 = 10$, $M_{\text{ref}} = 10^5$, $\gamma = 0.9$ and the trial set $\mathcal{P}_{\text{trial}}$ is chosen to be a set of 300 random parameter values which were uniformly sampled in $\mathcal{P}$. In this test case, we osbserve a similar behaviour of the (S)HMC-algorithm as in the first test case.

Figure 11 illustrates the computational gain brought by the HMC algorithm after $n = 70$ iterations (so that $M_n = 3109$) with respect to the classical Monte Carlo method. Indeed, the quantity $M_{MC}(\mu)$ defined in (43) is observed to vary in this case between 10^{12} and 10^{18}.

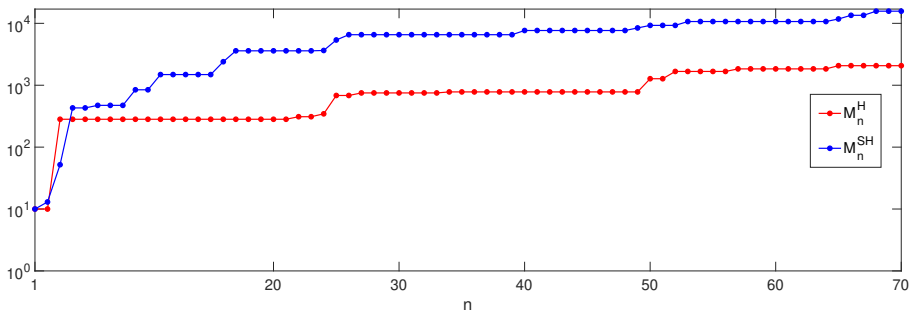

Figure 7: Evolution of M_n as a function of n for the HMC and SHMC-greedy algorithms.

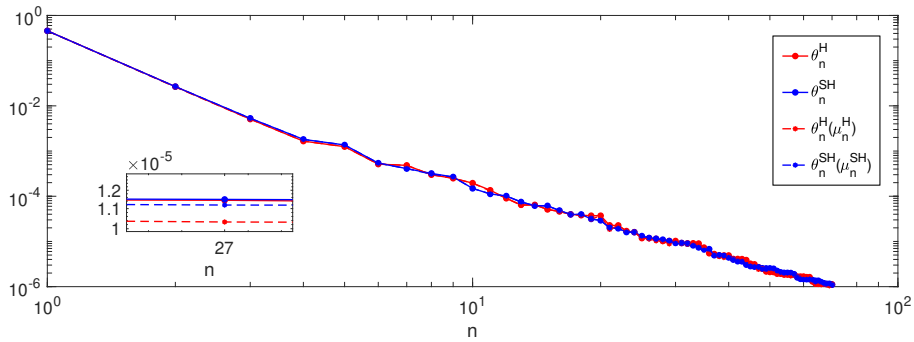

Figure 8: Evolution of $\theta_n^H(\mu_n^H)$, $\theta_n^{SH}(\mu_n^{SH})$, θ_n^H, θ_n^{SH} as a function of n in test case 2.

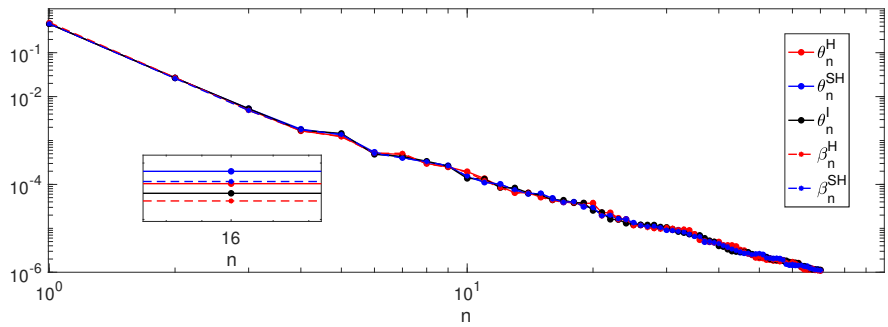

Figure 9: Evolution of β_n^H, β_n^{SH}, θ_n^H, θ_n^{SH}, θ_n^I as a function of n in test case 2.

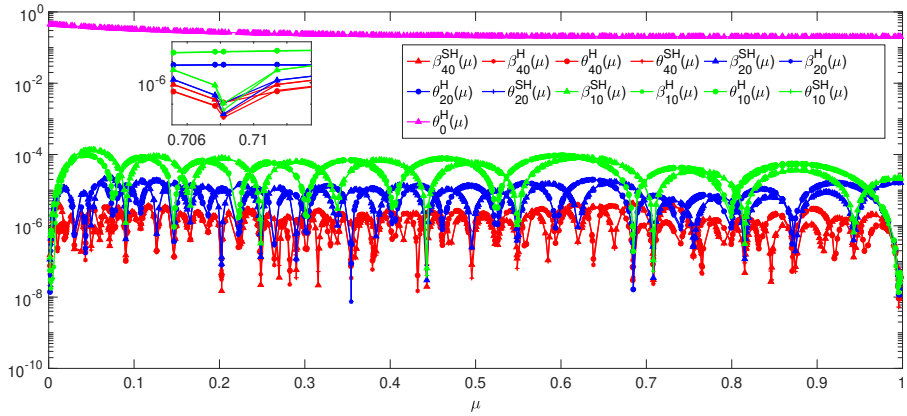

Figure 10: $\theta_n^H(\mu)$, $\theta_n^{SH}(\mu)$, $\beta_n^H(\mu)$, $\beta_n^{SH}(\mu)$ as a function of μ for $N = 0, 10, 20, 40$ on $\mathcal{P}_{test} = [0, 1]$.

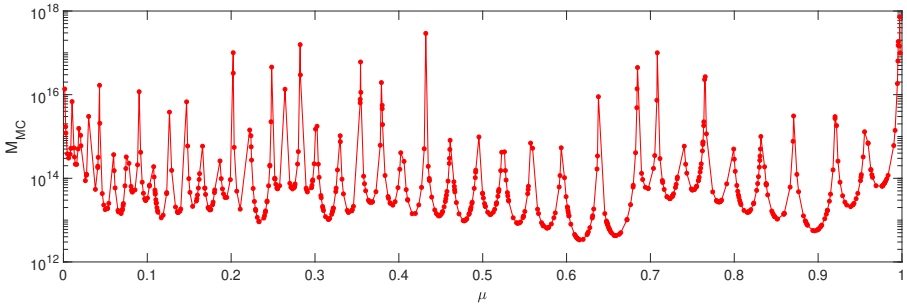

Figure 11: $M_{MC}(\mu)$ as a function of $\mu \in \mathcal{P}_{test} = [0, 1]$.

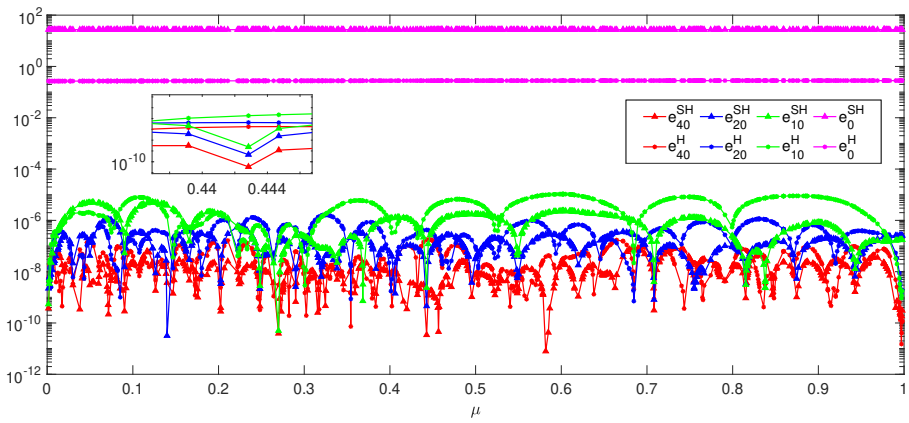

Figure 12: $e_N^H(\mu)$ and $e_N^{SH}(\mu)$ as a function of μ for $n = 0, 10, 20, 40$ on $\mathcal{P}_{test} = [0, 1]$.

4.4 Two-dimensional heat equation

Let Z_1 and Z_2 be two independent real-valued random varibales with probability laws respectively $\mathcal{U}(0.5, 2)$ and $\mathcal{N}(0, 1)$) and let $Z = (Z_1, Z_2)$. Let $\mathcal{D} = (0, 2)^2$, $\mathcal{P} := [0, 10]$. The trial set $\mathcal{P}_{\text{trial}}$ is constructed by selecting 50 random values uniformly distributed in $\mathcal{P}$.

For all $\mu \in \mathcal{P}$ and $z := (z_1, z_2) \in (0, 5, 2) \times \mathbb{R}$, we introduce

$$D^{\mu,z} : \begin{cases} \mathcal{D} & \to & \mathbb{R}^{2\times 2} \\ (x, y) & \mapsto & \begin{bmatrix} D_{11}^{\mu,z}(x, y) & 0 \\ 0 & D_{22}^{\mu,z}(x, y) \end{bmatrix} \end{cases}$$

where

$$\forall (x, y) \in \mathcal{D}, \quad D_{11}^{\mu,z}(x, y) = 13 + \mu \sin(2\pi x/z_1) + 0.5 z_2 \quad \text{and } D_{22}^{\mu,z}(x, y) = 13 + \mu \sin(2\pi y/z_1) + 0.5 z_2.$$

We introduce a conform triangular mesh $\mathcal{T}$ of the domain $\mathcal{D}$ as represented on the left-hand side plot of Figure 13 and denote by

$$V_h := \{u \in \mathcal{C}(\mathcal{D}), \quad u|_T \in \mathbb{P}_1 \ \forall T \in \mathcal{T}, \quad u|_{\partial\mathcal{D}} = 0\},$$

the standard $\mathbb{P}_1$ finite element space associated to this mesh.

For $\mu \in \mathcal{P}$ and $z \in (0.5, 2) \times \mathbb{R}$, we define $u_h^{\mu,z} \in V_h$ the unique solution to

$$a_{\mu,z}(u_h^{\mu,z}, v) = b(v), \quad \forall v \in V_h, \tag{45}$$

where

$$\forall v, w \in H_0^1(\mathcal{D}), \quad a_{\mu,z} = \int_{\mathcal{D}} \nabla v \cdot D^{\mu,z} \nabla w, \quad b(v) = \int_{\mathcal{D}} rv,$$

and where $r \in L^2(\mathcal{D})$ is defined by

$$r(x,y) = \exp\left(-(x-1)^2 - (y-1)^2\right), \quad \forall (x,y) \in \mathcal{D}.$$

The function $u_h^{\mu,z}$ is thus the standard $\mathbb{P}_1$ finite element approximation of the unique solution $u^{\mu,z} \in H_0^1(\mathcal{D})$ to

$$\begin{cases} -\mathrm{div}\left(D^{\mu,z} \nabla u^{\mu,z}\right) = r, & \text{in } \mathcal{D}, \\ u^{\mu,z} = 0 & \text{on } \partial\mathcal{D}. \end{cases} \tag{46}$$

Let $T_1 \in \mathcal{T}$ be the triangle colored in red in the left-hand side plot of Figure 13. For all $\mu \in \mathcal{P}$ and $z \in (0,5,2) \times \mathbb{R}$, we define by

$$f_\mu(z) := \frac{1}{|T_1|} \int_{T_1} u_h^{\mu,z}.$$

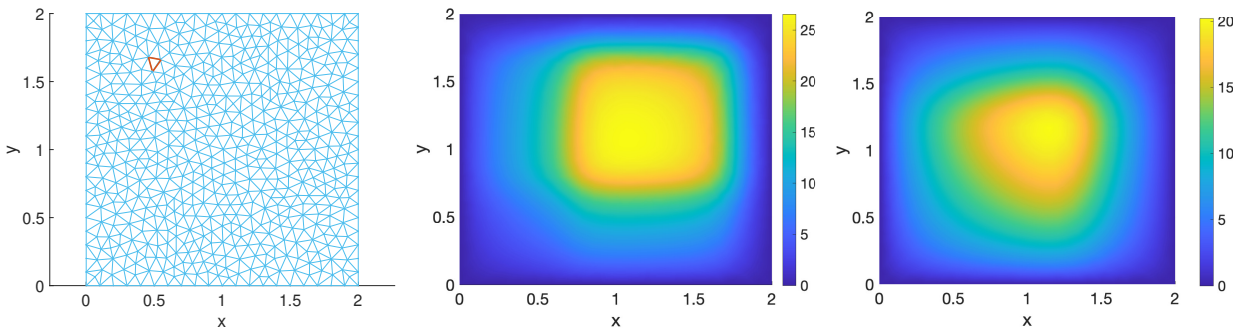

Figure 13: Left: mesh $\mathcal{T}$ (the triangle T_1 is highlighted in red color); Center: $u_h^{\mu,z}$ for $\mu = 9$ and $z = (1, 0)$; Right: $u_h^{\mu,z}$ for $\mu = 9$ and $z = (1.777, 0.2062)$.

In this example, $M_{\text{ref}} = 10^5$, $M_1 = 800$ and $\gamma = 0.9$. Figure 14 illustrates the evolution of the values of M_n as a function of n for the HMC and SHMC algorithms.

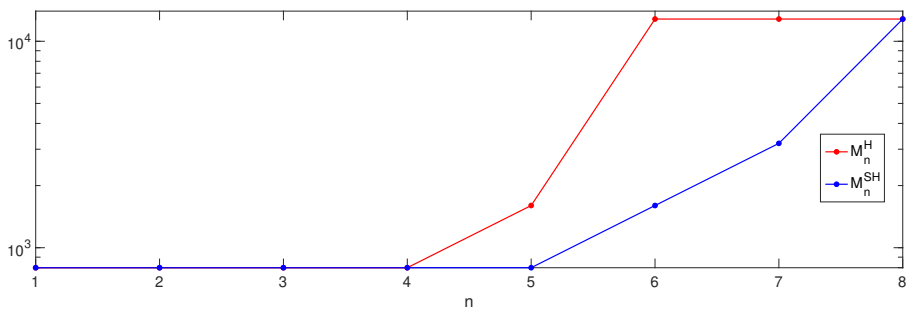

Figure 14: Evolution of M_n as a function of n for the HMC and SHMC-greedy algorithms in test case 3.

It is to be noted here that the quantities θ_n^H, θ_n^{SH} and θ_n^I are very close: the quality of approximation of the reduced spaces V_n^H or V_n^{SH} is very close to the quality of approximation of the reduced space V_n^I given by an ideal greedy algorithm.

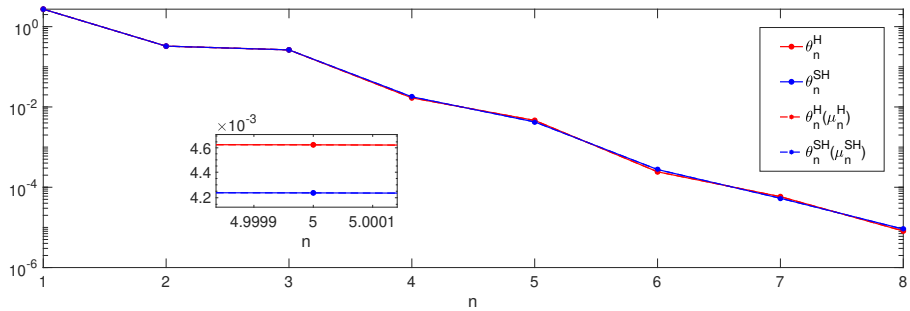

Figure 15: Evolution of $\theta_n^H(\mu_n^H)$, $\theta_n^{SH}(\mu_n^{SH})$, θ_n^H, θ_n^{SH} as a function of n in test case 3.

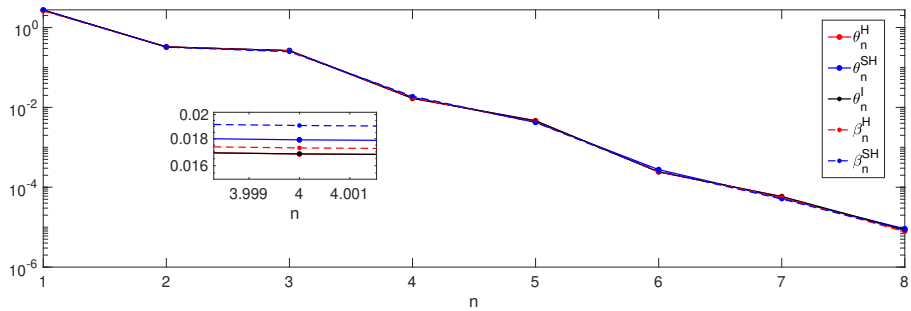

Figure 16: Evolution of β_n^H, β_n^{SH}, θ_n^H, θ_n^{SH}, θ_n^I as a function of n in test case 3.

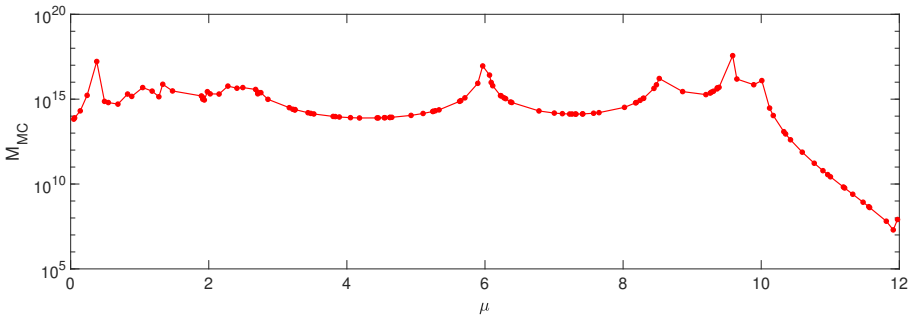

Figure 17: $M_{MC}(\mu)$ as a function of $\mu \in \mathcal{P}_{test} = [0, 12]$.

Figure 17 shows the value of $M_{MC}(\mu)$ given by (43), knowing that $M_n = 12800$ after $n = 7$ iterations of the HMC algorithm. We observe that in this case $10^{14} \leq M_{MC}(\mu) \leq 10^{20}$, which shows the huge computational gain brought by the HMC algorithm with respect to a standard Monte-Carlo method in this test case.

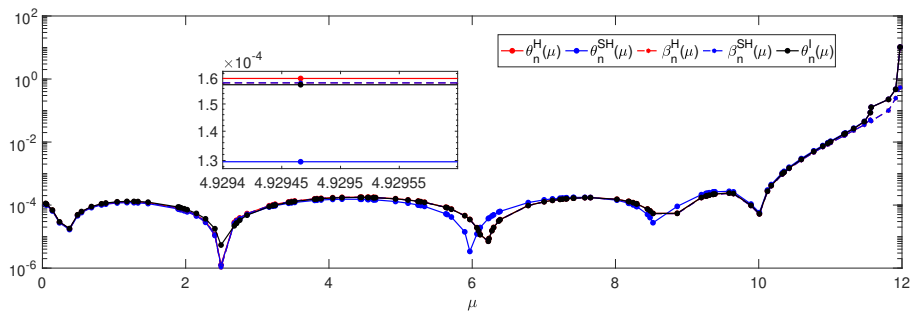

Figure 18: $\theta_n^H(\mu)$, $\theta_n^{SH}(\mu)$, $\theta_n^I(\mu)$, $\beta_n^H(\mu)$, $\beta_n^{SH}(\mu)$ as a function of μ for $n = 5$ in test case 3.

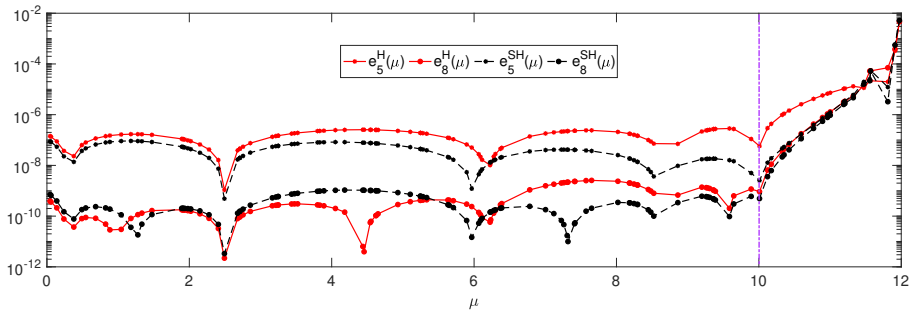

Figure 19: $e_n^H(\mu)$ and $e_n^{SH}(\mu)$ as a function of μ for $n = 5$ and $n = 8$ in test case 3.

Acknowledgements

The authors acknowledges financial support from the Mohammed VI Polytechnic University for funding Raed Blel's PhD thesis. Virginie Ehrlacher acknowledges support from project COMODO (ANR-19-CE46-0002). This work is partially supported by a funding from the European ResearchCouncil (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 810367), project EMC2.

References

[1] Oleg Balabanov and Anthony Nouy. Randomized linear algebra for model reduction. part i: Galerkin methods and error estimation. *Advances in Computational Mathematics*, 45(5):2969–3019, 2019.

[2] Oleg Balabanov and Anthony Nouy. Randomized linear algebra for model reduction. part ii: minimal residual methods and dictionary-based approximation. *arXiv preprint arXiv:1910.14378*, 2019.

[3] Maxime Barrault, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T Patera. An 'empirical interpolation'method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.

[4] Peter Binev, Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM journal on mathematical analysis*, 43(3):1457–1472, 2011.

[5] François Bolley, Arnaud Guillin, and Cédric Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2007.

[6] Sébastien Boyaval, Tony Lelièvre, et al. A variance reduction method for parametrized stochastic differential equations using the reduced basis paradigm. *Communications in Mathematical Sciences*, 8(3):735–762, 2010.

[7] Annalisa Buffa, Yvon Maday, Anthony T Patera, Christophe Prud'homme, and Gabriel Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical modelling and numerical analysis*, 46(3):595–603, 2012.

[8] Sayan Chakraborty, Souvik Chatterjee, Nilanjan Dey, Amira S Ashour, and Aboul Ella Hassanien. Comparative approach between singular value decomposition and randomized singular value decomposition-based watermarking. In *Intelligent techniques in signal processing for multimedia security*, pages 133–149. Springer, 2017.

[9] Albert Cohen, Wolfgang Dahmen, Ronald Devore, and James Nichols. Reduced basis greedy selection using random training sets. *ESAIM: Mathematical Modelling and Numerical Analysis*, 54(5):1509–1524, 2020.

[10] Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Greedy algorithms for reduced bases in banach spaces. *Constructive Approximation*, 37(3):455–466, 2013.

[11] Ronald A DeVore. Nonlinear approximation. *Acta numerica*, 7:51–150, 1998.

[12] Ronald A DeVore. The theoretical foundation of reduced basis methods. *Model Reduction and approximation: Theory and Algorithms*, pages 137–168, 2014.

[13] Virginie Ehrlacher, Damiano Lombardi, Olga Mula, and François-Xavier Vialard. Nonlinear model reduction on metric spaces. application to one-dimensional conservative pdes in wasserstein spaces. *ESAIM. Mathematical Modelling and Numerical Analysis*, 54, 2020.

[14] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

[15] Jan S Hesthaven, Gianluigi Rozza, Benjamin Stamm, et al. *Certified reduced basis methods for parametrized partial differential equations*, volume 590. Springer, 2016.

[16] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced basis methods for partial differential equations: an introduction*, volume 92. Springer, 2015.

[17] Kathrin Smetana and Olivier Zahm. Randomized residual-based error estimators for the proper generalized decomposition approximation of parametrized problems. *International Journal for Numerical Methods in Engineering*, 2019.

[18] Kathrin Smetana, Olivier Zahm, and Anthony T Patera. Randomized residual-based error estimators for parametrized equations. *SIAM journal on scientific computing*, 41(2):A900–A926, 2019.