A Multilevel Stochastic Approximation Algorithm for Value-at-Risk and Expected Shortfall Estimation

Stéphane Crépey, Noufel Frikha, Azar Louzi

A Multilevel Stochastic Approximation Algorithm for Value-at-Risk and Expected Shortfall Estimation

Stéphane Crépey*, Noufel Frikha†, Azar Louzi‡

March 22, 2023

Abstract

We propose a multilevel stochastic approximation (MLSA) scheme for the computation of the Value-at-Risk (VaR) and the Expected Shortfall (ES) of a financial loss, which can only be computed via simulations conditional on the realization of future risk factors. Thus, the problem of estimating its VaR and ES is nested in nature and can be viewed as an instance of a stochastic approximation problem with biased innovation. In this framework, for a prescribed accuracy ε, the optimal complexity of a standard stochastic approximation algorithm is shown to be of order ε^{-3}. To estimate the VaR, our MLSA algorithm attains an optimal complexity of order $\varepsilon^{-2-\delta}$, where $\delta < 1$ is some parameter depending on the integrability degree of the loss, while to estimate the ES, it achieves an optimal complexity of order $\varepsilon^{-2} |\ln \varepsilon|^2$. Numerical studies of the joint evolution of the error rate and the execution time demonstrate how our MLSA algorithm regains a significant amount of the lost performance due to the nested nature of the problem.

Keywords: Value-at-Risk, Expected Shortfall, stochastic approximation algorithm, nested Monte Carlo, multilevel Monte Carlo.

MSC: 65C05, 62L20, 62G32, 91Gxx.

1 Introduction

The post-great recession era has witnessed the implementation of multiple risk measures with the goal of better controlling financial losses. With conservatism and coherence in mind, [7] triggered a shift from Value-at-Risk (VaR for short) to Expected Shortfall (ES for short, i.e. the average loss given this loss exceeds the VaR) as a reference regulatory risk measure. In addition, as far as financial derivatives are

*Université Paris Cité, Laboratoire de Probabilités, Statistique et Modélisation (LPSM), CNRS UMR 8001 (`stephane.crepey@lpsm.paris`). The research of S. Crépey has benefited from the support of the Chair *Capital Markets Tomorrow: Modeling and Computational Issues* under the aegis of the Institut Europlace de Finance, a joint initiative of Laboratoire de Probabilités, Statistique et Modélisation (LPSM) / Université Paris Cité and Crédit Agricole CIB.

†Université Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne (CES), 106 Boulevard de l'Hôpital, 75642 Paris Cedex 13, (`noufel.frikha@univ-paris1.fr`). The research of N. Frikha has benefited from the support of the Institut Europlace de Finance.

‡Université Paris Cité, Laboratoire de Probabilités, Statistique et Modélisation (LPSM), CNRS UMR 8001 (`azar.louzi@lpsm.paris`).

concerned, the future valuation of the position (which determines the corresponding loss) is defined aBardouFrikhaPagès2009:1s a conditional expectation that, in the case of exotic products or for nonlinear (such as CVA or FVA) computations at the portfolio level, can only be computed by numerical simulation. VaR and ES computations are then nested in nature and a brute force nested Monte Carlo computational approach à la Gordy and Juneja [18] is too heavy for being usable in practice. As a shortcut, an estimator of the inner conditional expectation by regression is implemented in [8] but the resulting regression error is difficult to control. Adopting the stochastic approximation (SA) viewpoint of [5] (see also [4], [3] and [11]), itself building on Rockafellar and Uryasev's [19] representation of the VaR and ES as the solution of a convex optimization problem, [6] revisits both [18] and [8] computational strategies under more realistic and easier to check assumptions. However, we clarify in the present paper that the complexity of the resulting stochastic approximation algorithm to reach an accuracy ε is of order ε^{-3}, while the nonasymptotic error bounds of the regression strategy involve large constants, so that these regression error bounds are mainly of theoretical interest.

In the present paper, we propose a multilevel stochastic approximation (MLSA for short) algorithm for the computation of the VaR and the ES of a loss that writes as a conditional expectation. MLSA algorithm was introduced in [12] as an extension of the multilevel Monte Carlo path simulation method [15] to the framework of stochastic optimization by means of stochastic approximation algorithm. It was then revisited in [10] from a different perspective. We also refer to [13] for the development of a Richardson-Romberg stochastic approximation method. However, let us importantly point out that the uniform mean-reverting assumption on the objective function as discussed in [12] is generally not satisfied in a VaR and ES setup, so that one cannot directly guarantee the theoretical performance of the MLSA scheme for the computation of both risk measures from the results therein. We also refer to [16] where the authors propose a multilevel Monte Carlo estimator for the computation of the nested expectation $\mathbb{P}(\mathbb{E}[X|Y] \geq 0)$, as well as a stochastic root-finding algorithm for the computation of the VaR and ES, but without a theoretical analysis. Our main contribution is to propose an optimized MLSA algorithm that achieves theoretical sharp complexities when focusing on either estimating the VaR or estimating the ES. The VaR focused estimation achieves a complexity of order $\varepsilon^{-2-\delta}$, where $\delta < 1$ is some specific parameter depending on the integrability degree of the loss. The ES focused estimation attains a complexity of order $\varepsilon^{-2}\,|\ln \varepsilon|^2$. The theoretical analysis is verified numerically on two financial case studies which show a considerable gain, exponentially increasing as ε goes to 0, in terms of computational time for the computation of both VaR and ES.

The paper is organized as follows. In Section 2, the problem of computing the VaR and ES by means of stochastic approximation scheme is presented. Section 3 analyzes the nested SA scheme for the computation of the VaR and ES. Section 4 introduces and analyzes the MLSA algorithm. The theoretical analysis and the gain in performance achieved by our optimized MLSA scheme are illustrated numerically in Sections 5 and 6. The proofs of some technical results are given in Appendix.

2 Problem Statement

2.1 Value-at-Risk and Expected Shortfall

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space on which a financial loss $X_0 \in \mathbb{R}$ is defined as

$$X_0 = \mathbb{E}[\varphi(Y, Z)|Y], \tag{2.1}$$

where $Y \in \mathbb{R}^d$ and $Z \in \mathbb{R}^d$ are two independent random variables, and $\varphi\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function such that $\varphi(y, Z) \in L^1(\mathbb{P})$ for all $y \in \mathbb{R}^d$. Hence, we can rewrite X_0 as follows:

$$X_0 = \mathbb{E}[\varphi(Y, Z)|Y] = \Phi(Y), \tag{2.2}$$

with $\Phi(y) = \mathbb{E}[\varphi(y, Z)]$, $y \in \mathbb{R}^d$. In practice, $\varphi(Y, Z)$ would model the future cash flows of a given portfolio, Y would model the dynamics of the portfolio up a certain point in time, and X_0 would model the value of the portfolio at that given point in time.

Assuming that $X_0 \in L^1(\mathbb{P})$, we are interested in computing the VaR and the ES of X_0 for a given confidence level $\alpha \in (0, 1)$. Via e.g. [14] and [1], these risk measures are defined respectively by

$$\mathrm{VaR} \coloneqq \mathbb{V}\mathrm{a}\mathbb{R}_\alpha(X_0) \coloneqq \inf\big\{\xi \in \mathbb{R}\colon \mathbb{P}(X_0 \leq \xi) \geq \alpha\big\}, \tag{2.3}$$

$$\mathrm{ES} \coloneqq \mathbb{ES}_\alpha(X_0) \coloneqq \frac{1}{1-\alpha}\int_\alpha^1 \mathbb{V}\mathrm{a}\mathbb{R}_a(X_0)\,\mathrm{d}a. \tag{2.4}$$

2.2 Stochastic Approximation Paradigm

It is well known that VaR and ES are linked via a convex optimization problem as established by [19]. In order to state precisely this connection, we introduce the map V_0 defined by

$$V_0(\xi) = \xi + \frac{1}{1-\alpha}\mathbb{E}[(X_0 - \xi)^+], \quad \xi \in \mathbb{R}. \tag{2.5}$$

The proof of the following result can be found in [3, Proposition 2.1].

Lemma 2.1. *Suppose that $X_0 \in L^1(\mathbb{P})$ and that the cdf of X_0 is continuous. Then the function V_0 is convex, continuously differentiable,* $\lim_{|\xi|\to\infty} V_0(\xi) = \infty$ *and*

$$\operatorname{arg\,min} V_0 = \{V_0' = 0\} = \{\xi \in \mathbb{R}\colon \mathbb{P}(X_0 \leq \xi) = \alpha\}$$

is a bounded non-empty interval of $\mathbb{R}$*, where V_0' is given by*

$$V_0'(\xi) = \mathbb{E}[H_1(\xi, X_0)], \quad H_1(\xi, x) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{x\geq\xi}. \tag{2.6}$$

Furthermore, it holds

$$\mathbb{ES}_\alpha(X_0) = \min V_0.$$

Observe that in general the set $\operatorname{arg\,min} V_0$ does not reduce to a singleton. If the cdf of X_0 is increasing then the $\mathbb{V}\mathrm{a}\mathbb{R}_\alpha(X_0)$ is unique so that $\operatorname{arg\,min} V_0 = \{\mathbb{V}\mathrm{a}\mathbb{R}_\alpha(X_0)\}$. Under the approach initiated by [3], in the case of an exactly computable function Φ defined by (2.2), one may compute the couple $(\mathbb{V}\mathrm{a}\mathbb{R}_\alpha(X_0), \mathbb{ES}_\alpha(X_0))$ using a two time scale SA algorithm $(\xi_n, C_n)_{n\geq0}$ with dynamics

$$\begin{cases} \xi_{n+1} &= \xi_n - \gamma_{n+1} H_1(\xi_n, X_0^{(n+1)}), \\ C_{n+1} &= C_n - \frac{1}{n+1} H_2(\xi_n, C_n, X_0^{(n+1)}), \end{cases} \tag{2.7}$$

where

$$H_2(\xi, C, x) = C - \Big(\xi + \frac{1}{1-\alpha}(x-\xi)_+\Big). \tag{2.8}$$

The sequence $(X_0^{(n)})_{n\geq 1}$ stands for i.i.d. copies of X_0 and (ξ_0, C_0) is a random vector independent of $(X_0^{(n)})_{n\geq 1}$ satisfying $\mathbb{E}[|\xi_0|^2] + \mathbb{E}[|C_0|^2] < \infty$. The learning rate sequence $(\gamma_n)_{n\geq 1}$ in (2.7) is deterministic, positive, non-increasing and satisfies

$$\sum_{n\geq 1} \gamma_n = \infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < \infty. \tag{2.9}$$

However, in our case, one does not have access to a simulator of the law of X_0 given by (2.1) inasmuch as the distribution of $\varphi(Y, Z)$ conditionally on Y is not known and an analytical expression of the function Φ in (2.2) is not available.

3 Nested Stochastic Approximation

The above discussion naturally suggests to replace the samples of the random variable X_0 by approximate samples in the dynamics (2.7). We first let $K_0 \in \mathbb{N}_0$ and consider a bias parameter

$$h = \frac{1}{K} \in \mathcal{H} \coloneqq \Big\{\frac{1}{K} \colon K \in K_0\mathbb{N}_0\Big\}. \tag{3.1}$$

We then approximate X_0 by the random variable X_h defined by

$$X_h = \frac{1}{K}\sum_{k=1}^{K} \varphi(Y, Z^{(k)}), \tag{3.2}$$

where the sequence $(Z^{(k)})_{1\leq k\leq K}$ consists of i.i.d. copies of Z and is independent of Y. In order to simulate X_h, it thus suffices to sample Y then to independently sample $Z^{(1)}, \dots, Z^{(K)}$ and eventually compute X_h as the sample mean (3.2).

In the spirit of the previous section, assuming that $X_h \in L^1(\mathbb{P})$ and that the cdf of X_h is continuous, we define the approximating optimization problem

$$\min_{\xi\in\mathbb{R}} V_h(\xi), \tag{3.3}$$

where

$$V_h(\xi) \coloneqq \xi + \frac{1}{1-\alpha}\mathbb{E}[(X_h - \xi)^+], \quad \xi \in \mathbb{R}. \tag{3.4}$$

By applying Lemma 2.1,

$$\arg\min V_h = \{V_h' = 0\} \neq \varnothing \quad \text{and} \quad C_\star^h \coloneqq \mathbb{ES}_\alpha(X_h) = \min V_h.$$

Moreover, any minimizer $\xi_\star^h$ of V_h satisfies

$$\mathbb{P}(X_h \leq \xi_\star^h) = \alpha. \tag{3.5}$$

In order to compute the couple $(\xi_\star^h, C_\star^h)$, we devise the following two time scale SA algorithm $(\xi_n^h, C_n^h)_{n\geq 0}$ with dynamics

$$\begin{cases} \xi_{n+1}^h & = \xi_n^h - \gamma_{n+1} H_1(\xi_n^h, X_h^{(n+1)}), \\ C_{n+1}^h & = C_n^h - \frac{1}{n+1} H_2(\xi_n^h, C_n^h, X_h^{(n+1)}), \end{cases} \tag{3.6}$$

where $(X_h^{(n)})_{n\geq 1}$ is a sequence of i.i.d. copies of X_h and (ξ_0^h, C_0^h) is an $\mathbb{R}^2$-valued random variable independent of $(X_h^{(n)})_{n\geq 1}$ satisfying $\mathbb{E}[|\xi_0^h|^2] + \mathbb{E}[|C_0^h|^2] < \infty$.

Observe that this scheme is nested in nature, in the sense that the update of the outer layer (ξ_{n+1}^h, C_{n+1}^h) at step $n+1$ entails simulating the inner layer which is the Monte Carlo estimator $X_h^{(n+1)}$ given by (3.2). Besides, this numerical scheme is biased inasmuch as the target of $(\xi_n^h, C_n^h)_{n\geq 0}$ is $(\xi_\star^h, C_\star^h)$, which hopefully should converge to $(\mathbb{V}\mathrm{a}\mathbb{R}_\alpha(X_0), \mathbb{E}\mathbb{S}_\alpha(X_0))$ as $\mathcal{H} \ni h \downarrow 0$.

Algorithm 1 below summarizes the above procedure for approximating the couple $(\mathrm{VaR}, \mathrm{ES})$.

Algorithm 1 Nested SA algorithm for estimating $(\mathrm{VaR}, \mathrm{ES})$

Require: $K, N \in \mathbb{N}_0$, a positive sequence $(\gamma_n)_{1\leq n\leq N}$ satisfying (2.9)
Choose (ξ_0^h, C_0^h) such that $\mathbb{E}[|\xi_0^h|^2] + \mathbb{E}[|C_0^h|^2] < \infty$
for $n = 0 \mathinner{..} N-1$ **do**
    Simulate $Y^{(n+1)} \sim Y$ and $Z^{(n+1,1)}, \dots, Z^{(n+1,K)} \overset{\text{iid}}{\sim} Z$ independently of $Y^{(n+1)}$
    $X_h^{(n+1)} \leftarrow \frac{1}{K}\sum_{k=1}^K \varphi(Y^{(n+1)}, Z^{(n+1,k)})$
    $\xi_{n+1}^h \leftarrow \xi_n^h - \gamma_{n+1} H_1(\xi_n^h, X_h^{(n+1)})$
    $C_{n+1}^h \leftarrow C_n^h - \frac{1}{n+1} H_2(C_n^h, \xi_n^h, X_h^{(n+1)})$
end for
return The couple (ξ_N^h, C_N^h)

3.1 Convergence Analysis

We here study the nested stochastic approximation scheme (3.6). We first analyze its bias and then prove an $L^2(\mathbb{P})$ error estimate for any fixed integer $N \in \mathbb{N}_0$. We eventually analyze the complexity of the algorithm and establish a way to tune the number of steps N with respect to the bias parameter $h \in \mathcal{H}$ in order to reach some prescribed error. For $h \in \mathcal{H} \cup \{0\}$, we denote by Θ_h the set of minimizers of V_h

$$\Theta_h \coloneqq \arg\min V_h, \tag{3.7}$$

assuming that $X_h \in L^1(\mathbb{P})$ and that the cdf of X_h is continuous.

Lemma 3.1. *Suppose that for any $h \in \mathcal{H} \cup \{0\}$ the random variable $X_h \in L^1(\mathbb{P})$, that its cdf F_{X_h} is continuous and that the sequence of random variables $(X_h)_{h\in\mathcal{H}}$ converges in distribution to X_0 as $\mathcal{H} \ni h \downarrow 0$. Then, for any sequence $(\xi_\star^h)_{h\in\mathcal{H}}$ such that $\xi_\star^h \in \arg\min V_h$, it holds*

$$\mathrm{dist}(\xi_\star^h, \Theta_0) \to 0 \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0.$$

Assume additionally that $(X_h)_{h\in\mathcal{H}}$ converges to X_0 in $L^1(\mathbb{P})$. Then

$$C_\star^h \to C_\star \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0.$$

Proof. See Appendix A. □

We now introduce the following set of assumptions on the sequence $(X_h)_{h\in\mathcal{H}}$.

Assumption 1. (i) *F_{X_h} admits the following first order Taylor expansion: for any $\xi \in \mathbb{R}$ and any $h \in \mathcal{H}$,*

$$F_{X_h}(\xi) - F_{X_0}(\xi) = v(\xi)h + \epsilon(\xi, h)h, \quad \xi \in \mathbb{R},$$

for some functions $v, \epsilon(., h)\colon \mathbb{R} \to \mathbb{R}$ satisfying $\int_{\xi_\star}^{\infty} v(\xi)\,\mathrm{d}\xi < \infty$ and $\lim_{h\downarrow 0} \epsilon(\xi_\star, h) = \lim_{h\downarrow 0} \int_{\xi_\star}^{\infty} \epsilon(\xi, h)\,\mathrm{d}\xi = 0$ for any $\xi_\star \in \Theta_0$.

(ii) *For any $h \in \mathcal{H} \cup \{0\}$, the random variable X_h admits a density function f_{X_h} with respect to the Lebesgue measure. Moreover, the sequence of density functions $(f_{X_h})_{h\in\mathcal{H}}$ converges locally uniformly towards f_{X_0}.*

The proposition below quantifies the implicit weak error implied by approximating the unbiased problem with the biased one, by stating a first order expansion of the error between $(\xi^h_\star, C^h_\star)$ and $(\xi_\star, C_\star)$ in terms of the bias parameter h. Its proof is postponed to Appendix B.

Proposition 3.1. *Suppose that for any $h \in \mathcal{H} \cup \{0\}$ the random variable $X_h \in L^1(\mathbb{P})$ with continuous cdf F_{X_h}, that the density function f_{X_0} is positive, and that Assumption 1 is satisfied. Then, for any $\xi^h_\star \in \arg\min V_h$, as $\mathcal{H} \ni h \downarrow 0$, it holds*

$$\xi^h_\star - \xi_\star = -\frac{v(\xi_\star)}{f_{X_0}(\xi_\star)} h + o(h), \tag{3.8}$$

and

$$C^h_\star - C_\star = -h \int_{\xi_\star}^{\infty} \frac{v(\xi)}{1-\alpha} \mathrm{d}\xi + o(h). \tag{3.9}$$

Before dealing with the statistical error, we need a technical lemma. We start with the following assumption.

Assumption 2. (i) *For any $R > 0$, it holds*

$$\inf_{\substack{h\in\mathcal{H}\\ \xi\in B(\xi_\star, R)}} f_{X_h}(\xi) > 0.$$

(ii) *The sequence of density functions $(f_{X_h})_{h\in\mathcal{H}}$ are uniformly bounded and uniformly Lipschitz, namely,*

$$\sup_{h\in\mathcal{H}} \big(\|f_{X_h}\|_\infty + [f_{X_h}]_{\mathrm{Lip}}\big) < \infty.$$

We follow a similar strategy to the one developed in [9, Section 2.1], and define, for $h \in \mathcal{H}$, $\upsilon \geq 0$ and integer q, the Lyapunov function $\mathcal{L}^{\upsilon}_{h,q}\colon \mathbb{R} \to \mathbb{R}_+$ given by

$$\mathcal{L}^{\upsilon}_{h,q}(\xi) = \big(V_h(\xi) - V_h(\xi^h_\star)\big)^q \exp\Big(\upsilon\big(V_h(\xi) - V_h(\xi^h_\star)\big)\Big). \tag{3.10}$$

The proof of the next result is postponed to Appendix C.

Lemma 3.2. *Under Assumption 2, for any $h \in \mathcal{H}$, $\upsilon \geq 0$ and $q \geq 1$:*

(i) *$\mathcal{L}^{\upsilon}_{h,q}$ is twice continuously differentiable on $\mathbb{R}$.*

(ii) *For any $\xi \in \mathbb{R}$, it holds*

$$V_h'(\xi)(\mathcal{L}_{h,q}^{\upsilon})'(\xi) \geq \lambda_{h,q}^{\upsilon}\mathcal{L}_{h,q}^{\upsilon}(\xi),$$

with

$$\lambda_{h,q}^{\upsilon} \coloneqq \frac{3}{8}qV_h''(\xi_\star^h) \wedge \upsilon\frac{V_h''(\xi_\star^h)^4}{4[V_h'']_{\mathrm{Lip}}^2}.$$

Let also $\lambda_{h,q} \coloneqq \lambda_{h,q}^{\bar{\upsilon}_h}$ where $\bar{\upsilon}_h \coloneqq q^2\|V_h''\|_\infty$. Then, for any $q \geq 1$

$$\inf_{h\in\mathcal{H}} \lambda_{h,q} > 0. \tag{3.11}$$

With the above notations, we also have

$$\mathcal{L}_{h,q}^{\bar{\upsilon}_h}(\xi) \leq k_\alpha^q\,|\xi - \xi_\star^h|^q \exp\Big(\frac{q^2}{1-\alpha}k_\alpha \sup_{h\in\mathcal{H}} \|f_{X_h}\|_\infty|\xi - \xi_\star^h|\Big), \tag{3.12}$$

with $k_\alpha \coloneqq 1 \vee \frac{\alpha}{1-\alpha}$.

(iii) *For any $\xi \in \mathbb{R}$, it holds*

$$|(\mathcal{L}_{h,q}^{\upsilon})''(\xi)| \leq \eta_{h,q}^{\upsilon}\big(\mathcal{L}_{h,q}^{\upsilon}(\xi) + \mathcal{L}_{h,q-1}^{\upsilon}(\xi)\big),$$

where

$$\eta_{h,q}^{\upsilon} \coloneqq (q\vee\upsilon)\|V_h''\|_\infty + k_\alpha^2\upsilon(\upsilon\vee 2) + q\big(2\upsilon\vee(q-1)\big)\bigg(\frac{3k_\alpha^2[V_h'']_{\mathrm{Lip}}^2}{V_h''(\xi_\star^h)^3} \vee \frac{3\|V_h''\|_\infty^2}{V_h''(\xi_\star^h)}\bigg),$$

In particular, introducing $\eta_{h,q} \coloneqq \eta_{h,q}^{\bar{\upsilon}_h}$, for any $q \geq 1$, it holds

$$|\lambda_{h,q}|^2 \leq \eta_{h,q}. \tag{3.13}$$

and $(\eta_{h,q})_{h\in\mathcal{H}}$ satisfies

$$\sup_{h\in\mathcal{H}} \eta_{h,q} < \infty. \tag{3.14}$$

(iv) *For any $\xi \in \mathbb{R}$, it holds*

$$(\xi - \xi_\star^h)^{2q} \leq \kappa_{h,q}\big(\mathcal{L}_{h,q}^{\upsilon}(\xi) + \mathcal{L}_{h,2q}^{\upsilon}(\xi)\big),$$

where

$$\kappa_{h,q} \coloneqq \frac{3^q}{V_h''(\xi_\star^h)^q} \vee \frac{3^{2q}[V_h'']_{\mathrm{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}}.$$

In addition, for all $q \geq 1$, the sequence $(\kappa_{h,q})_{h\in\mathcal{H}}$ satisfies

$$\sup_{h\in\mathcal{H}} \kappa_{h,q} < \infty. \tag{3.15}$$

For any integer q, we let $\lambda_q \coloneqq \inf_{h\in\mathcal{H}} \lambda_{h,q}$ recalling that $(\lambda_{h,q})_{h\in\mathcal{H}}$ is defined as in Lemma 3.2(ii).

Theorem 3.1. *Suppose that for any $h \in \mathcal{H} \cup \{0\}$ the random variable $X_h \in L^2(\mathbb{P})$ and that Assumptions 1 and 2 are satisfied.*

Assume that $\sup_{h\in\mathcal{H}} \mathbb{E}[|\xi_0^h|^2 \exp(\frac{4}{1-\alpha} k_\alpha \sup_{h\in\mathcal{H}} \|f_{X_h}\|_\infty |\xi_0^h|)] < \infty$ *and* $\sup_{h\in\mathcal{H}} \mathbb{E}[|C_0^h|] < \infty$. *If $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0,1]$, with $\lambda_1\gamma_1 > 1$ if $\beta = 1$, then, for any $h \in \mathcal{H}$ and any positive integer n, it holds*

$$\mathbb{E}[(\xi_n^h - \xi_\star^h)^2] \leq \bar{K}_{h,2}^\beta \gamma_n \tag{3.16}$$

for some constants $\bar{K}_{h,2}^\beta$ satisfying $\sup_{h\in\mathcal{H}} \bar{K}_{h,2}^\beta < \infty$. *Moreover, one has*

$$\mathbb{E}[|C_n^h - C_\star^h|] \leq C_h^\beta n^{-(\frac{1}{2}\wedge\beta)}$$

for some constants C_h^β such that $\sup_{h\in\mathcal{H}} C_h^\beta < \infty$.

Assume additionally that $\sup_{h\in\mathcal{H}} \mathbb{E}[|\xi_0^h|^4 \exp(\frac{16}{1-\alpha} k_\alpha \sup_{h\in\mathcal{H}} \|f_{X_h}\|_\infty |\xi_0^h|)] < \infty$ *and* $\sup_{h\in\mathcal{H}} \mathbb{E}[(C_0^h)^2] < \infty$. *If $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0,1]$, with $\lambda_2\gamma_1 > 2$ when $\beta = 1$, then, for any $h \in \mathcal{H}$ and any positive integer n, it holds*

$$\mathbb{E}[(\xi_n^h - \xi_\star^h)^4] \leq \bar{K}_{h,4}^\beta \gamma_n^2, \tag{3.17}$$

for some constants $\bar{K}_{h,4}^\beta$ satisfying $\sup_{h\in\mathcal{H}} \bar{K}_{h,4}^\beta < \infty$. *Besides, one has*

$$\mathbb{E}[(C_n^h - C_\star^h)^2] \leq \bar{C}_h^\beta n^{-(1\wedge 2\beta)}, \tag{3.18}$$

for some constants $\bar{C}_h^\beta$ such that $\sup_{h\in\mathcal{H}} \bar{C}_h^\beta < \infty$.

Proof. As we do in the proof of Lemma 3.2 in Appendix C, we drop the superscript υ from our notation and write $\mathcal{L}_{h,q}(\xi)$ for $\mathcal{L}_{h,q}^\upsilon(\xi)$.

Step 1. Inequality on $\mathbb{E}[\mathcal{L}_{h,q}(\xi_n^h)]$**.**

We first prove a general inequality on $\mathbb{E}[\mathcal{L}_{h,q}(\xi_n^h)]$, for any $n \geq 0$ and $q \geq 1$. For $h \in \mathcal{H}$ fixed, we introduce the filtration $\mathcal{F}^h = (\mathcal{F}_n^h)_{n\geq 0}$ with $\mathcal{F}_n^h = \sigma(\xi_0^h, C_0^h, X_h^{(1)}, \ldots, X_h^{(n)})$.

Observe now that the dynamics of $(\xi_n^h)_{n\geq 0}$ given by the first component of (3.6) satisfy

$$\xi_{n+1}^h = \xi_n^h - \gamma_{n+1} V_h'(\xi_n^h) - \gamma_{n+1} e_{n+1}^h, \tag{3.19}$$

where

$$e_{n+1}^h \coloneqq H_1(\xi_n^h, X_h^{(n+1)}) - V_h'(\xi_n^h), \quad n \geq 0, \tag{3.20}$$

is an $(\mathcal{F}^h, \mathbb{P})$-martingale increment. For $q \geq 1$, we test the Lyapunov function of Lemma 3.2, namely $\mathcal{L}_{h,q} = \mathcal{L}_{h,q}^{\bar{\upsilon}_h}$ with $\bar{\upsilon}_h = q^2 \|V_h''\|_\infty$, along dynamics (3.19). Using a second order Taylor expansion, we obtain

$$\begin{aligned}
\mathcal{L}_{h,q}(\xi_{n+1}^h) &= \mathcal{L}_{h,q}\big(\xi_n^h - \gamma_{n+1} V_h'(\xi_n^h) - \gamma_{n+1} e_{n+1}^h\big) \\
&= \mathcal{L}_{h,q}(\xi_n^h) - \gamma_{n+1} \mathcal{L}_{h,q}'(\xi_n^h)(V_h'(\xi_n^h) + e_{n+1}^h) \\
&\quad + \gamma_{n+1}^2 H_1(\xi_n^h, X_h^{(n+1)})^2 \int_0^1 (1-t) \mathcal{L}_{h,q}''(t\xi_{n+1}^h + (1-t)\xi_n^h)\, \mathrm{d}t.
\end{aligned} \tag{3.21}$$

It follows from Lemma 3.2(ii,iii) that

$$\begin{aligned}
\mathcal{L}_{h,q}(\xi_{n+1}^h) \leq \mathcal{L}_{h,q}(\xi_n^h)(1 - \lambda_{h,q}^\upsilon \gamma_{n+1}) - \gamma_{n+1} \mathcal{L}_{h,q}'(\xi_n^h) e_{n+1}^h + \eta_{h,q}^\upsilon \gamma_{n+1}^2 H_1(\xi_n^h, X_h^{(n+1)})^2 \times \\
\int_0^1 (1-t)\big(\mathcal{L}_{h,q}(t\xi_{n+1}^h + (1-t)\xi_n^h) + \mathcal{L}_{h,q-1}(t\xi_{n+1}^h + (1-t)\xi_n^h)\big) \mathrm{d}t.
\end{aligned} \tag{3.22}$$

Let $t \in [0,1]$. The mean value theorem guarantees that there exists $\widetilde{\xi}_n^h(t) \in \mathbb{R}$ such that

$$V_h(t\xi_{n+1}^h + (1-t)\xi_n^h) = V_h(\xi_n^h) + tV_h'(\widetilde{\xi}_n^h(t))(\xi_{n+1}^h - \xi_n^h). \tag{3.23}$$

From the very definition of H_1, it holds (cf (3.6))

$$|\xi_{n+1}^h - \xi_n^h| = \gamma_{n+1}|H_1(\xi_n^h, X_h^{(n+1)})| \leq k_\alpha \gamma_{n+1}, \tag{3.24}$$

with $k_\alpha \coloneqq \frac{\alpha}{1-\alpha} \vee 1$. Applying the triangle inequality to (3.23), we get

$$V_h(t\xi_{n+1}^h + (1-t)\xi_n^h) - V_h(\xi_\star^h) \leq V_h(\xi_n^h) - V_h(\xi_\star^h) + k_\alpha^2 \gamma_{n+1}. \tag{3.25}$$

Using the inequality $\mathrm{e}^x \leq \mathrm{e}\mathbf{1}_{x\leq 1} + x^q \mathrm{e}^x \mathbf{1}_{x>1} \leq \mathrm{e}(1 + x^q \mathrm{e}^x)$ and the very definition of $\mathcal{L}_{h,q}$, for any $\xi \in \mathbb{R}$, we obtain

$$\mathcal{L}_{h,0}(\xi) \leq \mathrm{e}\big(1 + \upsilon^q \mathcal{L}_{h,q}(\xi)\big), \quad q \geq 1. \tag{3.26}$$

Henceforth, using (3.25) and then (3.26)

$$\begin{aligned} &\mathcal{L}_{h,q}(t\xi_{n+1}^h + (1-t)\xi_n^h) \\ &\qquad \leq \big(V_h(\xi_n^h) - V_h(\xi_\star^h) + k_\alpha^2 \gamma_{n+1}\big)^q \exp\big(\upsilon\big(V_h(\xi_n^h) - V_h(\xi_\star^h) + k_\alpha^2\gamma_{n+1}\big)\big) \\ &\qquad \leq 2^{q-1} \exp\big(\upsilon k_\alpha^2 \gamma_{n+1}\big)\big(\mathcal{L}_{h,q}(\xi_n^h) + k_\alpha^{2q}\gamma_{n+1}^q \mathcal{L}_{h,0}(\xi_n^h)\big) \\ &\qquad \leq \sigma_{\alpha,q}^\upsilon\big(\mathcal{L}_{h,q}(\xi_n^h) + \gamma_{n+1}^q\big), \end{aligned} \tag{3.27}$$

with $\sigma_{\alpha,q}^\upsilon \coloneqq 2^{q-1}\exp(\upsilon k_\alpha^2 \gamma_1)\big((1 + \mathrm{e}\upsilon^q k_\alpha^{2q}\gamma_1^q) \vee \mathrm{e}k_\alpha^{2q}\big)$. By (3.24),

$$\gamma_{n+1}^2 H_1(\xi_n^h, X_h^{(n+1)})^2 \leq k_\alpha^2 \gamma_{n+1}^2. \tag{3.28}$$

Plugging the upper bounds (3.27) and (3.28) into (3.22) we deduce

$$\begin{aligned} \mathcal{L}_{h,q}(\xi_{n+1}^h) \leq \mathcal{L}_{h,q}(\xi_n^h)(1 - \lambda_{h,q}^\upsilon \gamma_{n+1} + \zeta_{h,q}^\upsilon \gamma_{n+1}^2) - \gamma_{n+1}\mathcal{L}_{h,q}'(\xi_n^h)e_{n+1}^h \\ + \zeta_{h,q}^\upsilon \gamma_{n+1}^2 \mathcal{L}_{h,q-1}(\xi_n^h) + \zeta_{h,q}^\upsilon \gamma_{n+1}^{q+1}, \end{aligned} \tag{3.29}$$

with

$$\zeta_{h,q}^\upsilon \coloneqq \frac{1}{2}\eta_{h,q}^\upsilon k_\alpha^2\big((\gamma_1 \sigma_{\alpha,q}^\upsilon + \sigma_{\alpha,q-1}^\upsilon) \vee \sigma_{\alpha,q}^\upsilon\big), \tag{3.30}$$

which, according to the second part of Lemma 3.2(iii), satisfies

$$\zeta_q^\infty \coloneqq \sup_{h\in\mathcal{H}} \zeta_{h,q}^{\bar{\upsilon}_h} < \infty. \tag{3.31}$$

Note that, since $(e_n^h)_{n\geq 1}$ are $(\mathcal{F}^h, \mathbb{P})$-martingale increments, via the tower law,

$$\mathbb{E}[\mathcal{L}_{h,q}'(\xi_n^h)e_{n+1}^h] = \mathbb{E}\big[\mathcal{L}_{h,q}'(\xi_n^h)\mathbb{E}[e_{n+1}^h|\mathcal{F}_n^h]\big] = 0.$$

Hence, by taking the expectation in both sides of inequality (3.29),

$$\mathbb{E}[\mathcal{L}_{h,q}(\xi_{n+1}^h)] \leq \mathbb{E}[\mathcal{L}_{h,q}(\xi_n^h)]\big(1 - \lambda_{h,q}^\upsilon\gamma_{n+1} + \zeta_{h,q}^\upsilon\gamma_{n+1}^2\big) + \zeta_{h,q}^\upsilon\gamma_{n+1}^2\mathbb{E}[\mathcal{L}_{h,q-1}(\xi_n^h)] + \zeta_{h,q}^\upsilon\gamma_{n+1}^{q+1}. \tag{3.32}$$

Step 2. Inequality on $\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)]$***.***

We prove here a sharper upper estimate on $\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)]$. Taking $q = 1$ in (3.32) and (3.26), we get

$$\mathbb{E}[\mathcal{L}_{h,1}(\xi_{n+1}^h)] \leq \mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)]\big(1 - \lambda_{h,1}^\upsilon\gamma_{n+1} + (\mathrm{e}\upsilon + 1)\zeta_{h,1}^\upsilon\gamma_{n+1}^2\big) + (\mathrm{e}+1)\zeta_{h,1}^\upsilon\gamma_{n+1}^2. \tag{3.33}$$

Denote $\zeta_{h,q} := \zeta_{h,q}^{\bar{\upsilon}_h}$. Observe that, since $|\lambda_{h,q}|^2 \leq \eta_{h,q}$, recalling Lemmas 3.2(ii,iii) and (3.30), we obtain

$$\begin{aligned} \lambda_{h,q}\gamma_{n+1} - (\mathrm{e}\bar{\upsilon}_h^q + 1)\zeta_{h,q}\gamma_{n+1}^2 &\leq \sqrt{\eta_{h,q}}\gamma_{n+1} - \eta_{h,q}c_{\alpha,q}\gamma_{n+1}^2 \\ &= \sqrt{\eta_{h,q}}\gamma_{n+1}\big(1 - c_{\alpha,q}\sqrt{\eta_{h,q}}\gamma_{n+1}\big) \leq (4c_{\alpha,q})^{-1} \leq \frac{1}{2}, \end{aligned} \tag{3.34}$$

with $c_{\alpha,q}^{\upsilon} := (e\upsilon^q + 1)\zeta_{h,q}\eta_{h,q}^{-1} = 2^{-1}(\mathrm{e}\upsilon^q + 1)(\sigma_{\alpha,q}^{\upsilon} \vee (\gamma_1\sigma_{\alpha,q}^{\upsilon} + \sigma_{\alpha,q-1}^{\upsilon}))k_\alpha^2 > 1/2$, $c_{\alpha,q} := c_{\alpha,q}^{\bar{\upsilon}_h}$, and where we used the fact that the function $x(1 - c_{\alpha,q}x) \leq (4c_{\alpha,q})^{-1}$ for any $x \geq 0$. Hence, $1 - \lambda_{h,1}\gamma_{n+1} + (\mathrm{e}\bar{\upsilon}_h^q + 1)\zeta_{h,1}\gamma_{n+1}^2 \geq 7/8$ for any integer n. Evaluating (3.33) at $\upsilon = \bar{\upsilon}_h$, then iterating it n times inequality, we obtain

$$\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)] \leq \mathbb{E}[\mathcal{L}_{h,1}(\xi_0^h)]\Pi_{1:n}^{h,1} + (\mathrm{e}+1)\zeta_{h,1}\sum_{k=1}^{n}\gamma_k^2\,\Pi_{k+1:n}^{h,1}, \tag{3.35}$$

where

$$\Pi_{k:n}^{h,q} := \prod_{j=k}^{n}\big(1 - \lambda_{h,q}\gamma_j + (\mathrm{e}\bar{\upsilon}_h^q + 1)\zeta_{h,q}\gamma_j^2\big), \quad q \geq 1, \tag{3.36}$$

with the convention $\prod_\varnothing = 1$.

We now focus on our specific choice of learning sequence. Let us first assume that $\gamma_n = \gamma_1 n^{-\beta}$ with $\gamma_1 > 0$ and $\beta \in (0,1]$, and let

$$\varphi_\eta(t) := \begin{cases} \eta^{-1}(t^\eta - 1) & \eta \neq 0, \\ \ln t + \gamma^{\mathscr{E}} & \eta = 0, \end{cases} \tag{3.37}$$

recalling that $\gamma^{\mathscr{E}}$ is the Euler-Mascheroni constant.

To deal with the first term in the right hand side of (3.35), we use the simple bound $1 + x \leq \mathrm{e}^x$, $x \in \mathbb{R}$, to deduce

$$\Pi_{k+1:n}^{h,q} \leq \exp\Big(-\lambda_{h,q}\sum_{j=k+1}^{n}\gamma_j\Big)\exp\Big((\mathrm{e}\bar{\upsilon}_h^q + 1)\zeta_{h,q}\sum_{j=k+1}^{n}\gamma_j^2\Big),$$

with the convention $\sum_\varnothing = 0$.

Step 2.1.
We here assume that $\beta = 1$. Recall that $\sum_{j=k+1}^{n}\gamma_j = \gamma_1\big(\psi(n+1) - \psi(k+1)\big)$, where ψ is the digamma function that satisfies

$$\ln x - \frac{1}{x} \leq \psi(x) \leq \ln x - \frac{1}{2x}, \quad x > 0.$$

Hence, for $0 \leq k \leq n$,

$$\sum_{j=k+1}^{n}\gamma_j \geq \gamma_1\ln\Big(\frac{n+1}{k+1}\Big) - \frac{\gamma_1}{n+1} + \frac{\gamma_1}{2(k+1)} \geq \gamma_1\ln\Big(\frac{n+1}{k+1}\Big) - \frac{\gamma_1}{2},$$

which in turn yields

$$\Pi_{k+1:n}^{h,q} \leq \exp\Big((\mathrm{e}\bar{\upsilon}_h^q + 1)\zeta_{h,q}\frac{\pi^2}{6}\gamma_1^2 + \frac{\gamma_1\lambda_{h,q}}{2}\Big)\,\frac{(k+1)^{\lambda_{h,q}\gamma_1}}{(n+1)^{\lambda_{h,q}\gamma_1}}. \tag{3.38}$$

The above estimate together with a comparison between series and integrals gives, for $p \geq 0$,

$$\sum_{k=1}^{n} \gamma_k^p \, \Pi_{k+1:n}^{h,q} \leq \gamma_1^p \exp\Big((\mathrm{e}\bar{v}_h^q + 1)\zeta_{h,q} \frac{\pi^2}{6}\gamma_1^2 + \frac{\gamma_1 \lambda_{h,q}}{2} + \ln 2(\gamma_1 \lambda_{h,q} \vee p) \Big) \frac{\varphi_{\lambda_{h,q}\gamma_1 - p+1}(n+1)}{(n+1)^{\lambda_{h,q}\gamma_1}}. \tag{3.39}$$

Plugging the estimates (3.38) and (3.39) for $q = 1$ and $p = 2$ into (3.35) gives

$$\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)] \leq \widehat{K}_{h,1} \frac{\mathbb{E}[\mathcal{L}_{h,1}(\xi_0^h)]}{(n+1)^{\lambda_{h,1}\gamma_1}} + \bar{K}_{h,1} \frac{\varphi_{\lambda_{h,1}\gamma_1 - 1}(n+1)}{(n+1)^{\lambda_{h,1}\gamma_1}},$$

where $\widehat{K}_{h,1} \coloneqq \exp\big((\mathrm{e}\bar{v}_h + 1)\zeta_{h,1} \frac{\pi^2}{6}\gamma_1^2 + \frac{\gamma_1 \lambda_{h,1}}{2} \big)$ and $\bar{K}_{h,1} \coloneqq \widehat{K}_{h,1}\gamma_1^2(\mathrm{e}+1)\zeta_{h,1} 2^{\gamma_1 \lambda_{h,1} \vee 2}$.

Hence, if $\lambda_{h,1}\gamma_1 > 1$, recalling (3.37), it holds

$$\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)] \leq K_{h,1}^1 \gamma_n, \tag{3.40}$$

with

$$K_{h,1}^1 \coloneqq \gamma_1^{-1}(\widehat{K}_{h,1}\mathbb{E}[\mathcal{L}_{h,1}(\xi_0^h)] + \bar{K}_{h,1}).$$

Step 2.2.
We now assume that $0 < \beta < 1$. A comparison between series and integrals gives

$$\begin{aligned} \Pi_{k+1:n}^{h,q} \leq & \exp\Big(-\lambda_{h,q}\gamma_1 \big(\varphi_{1-\beta}(n+1) - \varphi_{1-\beta}(k+1) \big) \Big) \\ & \times \exp\Big(2^{2\beta} (\mathrm{e}\bar{v}_h^q + 1)\zeta_{h,q}\gamma_1^2 \big(\varphi_{1-2\beta}(n+1) - \varphi_{1-2\beta}(k+1) \big) \Big). \end{aligned} \tag{3.41}$$

We introduce $n_0 \coloneqq \inf\big\{ n \in \mathbb{N}_0 \colon \gamma_n \leq \frac{\lambda_{h,q}}{2(\mathrm{e}\bar{v}_h^q + 1)\zeta_{h,q}} \big\} - 1$, and remark that for $n \geq n_0 + 1$, $1 - \lambda_{h,q}\gamma_n + (\mathrm{e}\bar{v}_h^q + 1)\zeta_{h,q}\gamma_n^2 \leq 1 - \frac{\lambda_{h,q}}{2}\gamma_n$. Hence, for $p \geq 0$,

$$\begin{aligned} \sum_{k=1}^{n} \gamma_k^p \, \Pi_{k+1:n}^{h,q} &= \sum_{k=1}^{n_0 \wedge n} \gamma_k^p \, \Pi_{k+1:n_0 \wedge n}^{h,q} \Pi_{n_0 \wedge n+1:n}^{h,q} + \sum_{k=n_0 \wedge n+1}^{n} \gamma_k^p \, \Pi_{k+1:n}^{h,q} \\ &\leq \Big(\sum_{k=1}^{n_0 \wedge n} \gamma_k^p \prod_{j=k+1}^{n_0 \wedge n} \big(1 + (\mathrm{e}\bar{v}_h^q + 1)\zeta_{h,q}\gamma_j^2 \big) \Big) \prod_{j=n_0 \wedge n+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j \Big) \\ &\quad + \sum_{k=1}^{n} \gamma_k^p \prod_{j=k+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j \Big). \end{aligned} \tag{3.42}$$

In order to simplify the first term in the right hand side of the above inequality, we write, for $p \geq 2$,

$$\begin{aligned}
&\sum_{k=1}^{n_0\wedge n} \gamma_k^p \prod_{j=k+1}^{n_0\wedge n} \big(1+(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\gamma_j^2\big)\\
&\qquad \leq \gamma_1^{p-2} \sum_{k=1}^{n_0\wedge n} \gamma_k^2 \prod_{j=k+1}^{n_0\wedge n} \big(1+(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\gamma_j^2\big)\\
&\qquad = \frac{\gamma_1^{p-2}}{(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}} \sum_{k=1}^{n_0\wedge n} \bigg(\prod_{j=k}^{n_0\wedge n} \big(1+(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\gamma_j^2\big) - \prod_{j=k+1}^{n_0\wedge n} \big(1+(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\gamma_j^2\big)\bigg)\\
&\qquad \leq \frac{\gamma_1^{p-2}}{(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}} \prod_{j=1}^{n_0\wedge n} \big(1+(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\gamma_j^2\big)\\
&\qquad \leq \frac{\gamma_1^{p-2}}{(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}} \exp\bigg((\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q} \sum_{j=1}^{n_0\wedge n} \gamma_j^2\bigg),
\end{aligned}$$

and

$$\begin{aligned}
\prod_{j=n_0\wedge n+1}^{n} \Big(1-\frac{\lambda_{h,q}}{2}\gamma_j\Big) &\leq \exp\bigg(-\frac{\lambda_{h,q}}{2}\sum_{j=1}^{n}\gamma_j\bigg)\exp\bigg(\frac{\lambda_{h,q}}{2}\sum_{j=1}^{n_0\wedge n}\gamma_j\bigg)\\
&\leq \exp\bigg(-\frac{\lambda_{h,q}}{2}\sum_{j=1}^{n}\gamma_j\bigg)\exp\bigg((\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\sum_{j=1}^{n_0\wedge n}\gamma_j^2\bigg),
\end{aligned}$$

where we used the fact that $\gamma_j \geq \lambda_{h,q}/(2(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q})$ for $j \leq n_0$. Combining the two preceding estimates and using a comparison between series and integrals, we deduce, for $p \geq 2$,

$$\begin{aligned}
&\bigg(\sum_{k=1}^{n_0\wedge n} \gamma_k^p \prod_{j=k+1}^{n_0\wedge n} (1+(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\gamma_j^2)\bigg) \prod_{j=n_0\wedge n+1}^{n} \Big(1-\frac{\lambda_{h,q}}{2}\gamma_j\Big)\\
&\quad \leq \frac{\gamma_1^{p-2}}{(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}} \exp\bigg(2(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\sum_{j=1}^{n_0\wedge n}\gamma_j^2\bigg)\exp\bigg(-\frac{\lambda_{h,q}}{2}\sum_{j=1}^{n}\gamma_j\bigg)\\
&\quad \leq \frac{\gamma_1^{p-2}}{(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}} \exp\big(2^{2\beta+1}(\mathrm{e}\bar{v}_h^q+1)\zeta_{h,q}\gamma_1^2\varphi_{1-2\beta}(n+1)\big)\exp\Big(-\frac{\lambda_{h,q}}{2}\gamma_1\varphi_{1-\beta}(n+1)\Big).
\end{aligned} \tag{3.43}$$

We now deal with the second term appearing in the right hand side of (3.42). Since

$(\gamma_n)_{n\geq 1}$ is a positive non increasing sequence, for any integer $1 \leq m \leq n$, it holds

$$
\begin{aligned}
&\sum_{k=1}^{n} \gamma_k^p \prod_{j=k+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j\Big) \\
&\quad = \sum_{k=1}^{m} \gamma_k^p \prod_{j=k+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j\Big) + \sum_{k=m+1}^{n} \gamma_k^p \prod_{j=k+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j\Big) \\
&\quad \leq \prod_{j=m+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j\Big) \sum_{k=1}^{m} \gamma_k^p + \gamma_m^{p-1} \sum_{k=m+1}^{n} \gamma_k \prod_{j=k+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j\Big) \\
&\quad \leq \exp\bigg(- \frac{\lambda_{h,q}}{2} \sum_{j=m+1}^{n} \gamma_j \bigg) \gamma_1^{p-2} \sum_{k=1}^{n} \gamma_k^2 + \frac{2\gamma_m^{p-1}}{\lambda_{h,q}} \bigg(1 - \prod_{j=m+1}^{n} \Big(1 - \frac{\lambda_{h,q}}{2}\gamma_j\Big)\bigg) \\
&\quad \leq 2^{2\beta}\gamma_1^p \exp\Big(- 2^{-(\beta+1)}\lambda_{h,q}\gamma_1\big(\varphi_{1-\beta}(n) - \varphi_{1-\beta}(m)\big)\Big)\varphi_{1-2\beta}(n+1) + \frac{2\gamma_1^{p-1}}{\lambda_{h,q} m^{(p-1)\beta}}.
\end{aligned}
\tag{3.44}
$$

We now select $m = \lceil n/2 \rceil$ and use the inequality $\varphi_{1-\beta}(n) - \varphi_{1-\beta}(n/2) \geq n^{1-\beta}/2$ for $n \geq 2$ which stems from the concavity of $\varphi_{1-\beta}$ together with the fact that $\varphi_{1-\beta}$ is non-decreasing. Hence, for any integer n and $p \geq 2$,

$$
\sum_{k=1}^{n} \gamma_k^p \prod_{j=k+1}^{n} \Big(1 - \frac{\lambda_{h,1}}{2}\gamma_j\Big) \leq 2^{2\beta}\gamma_1^p \exp\big(-2^{-(\beta+2)}\lambda_{h,1}\gamma_1 n^{1-\beta}\big)\varphi_{1-2\beta}(n+1) + \frac{2^{1+(p-1)\beta}\gamma_1^{p-1}}{\lambda_{h,1} n^{(p-1)\beta}}.
\tag{3.45}
$$

Going back to (3.35), by using on the one hand (3.41) and on the other hand (3.42) combined with (3.43) and (3.45) for $q = 1$ and $p = 2$, we obtain

$$
\begin{aligned}
\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)] \leq{} & \Big(\mathbb{E}[\mathcal{L}_{h,1}(\xi_0^h)] + \frac{\mathrm{e}+1}{\mathrm{e}\bar{\upsilon}_h + 1}\Big) \\
& \times \exp\Big(2^{2\beta+1}(\mathrm{e}\bar{\upsilon}_h + 1)\zeta_{h,1}\gamma_1^2\varphi_{1-2\beta}(n+1) - \frac{\lambda_{h,1}}{2}\gamma_1\varphi_{1-\beta}(n+1)\Big) \\
& + (\mathrm{e}+1)\zeta_{h,1}\Big(2^{2\beta}\gamma_1^2 \exp\big(- 2^{-(\beta+2)}\lambda_{h,1}\gamma_1 n^{1-\beta}\big)\varphi_{1-2\beta}(n+1) + \frac{2^{\beta+1}\gamma_1}{\lambda_{h,1} n^{\beta}}\Big).
\end{aligned}
\tag{3.46}
$$

Hence, for any $\beta \in (0,1)$ and any positive integer n,

$$
\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)] \leq K_{h,1}^{\beta}\gamma_n,
\tag{3.47}
$$

with

$$
\begin{aligned}
K_{h,1}^{\beta} \coloneqq{} & \gamma_1^{-1}\Big(\mathbb{E}[\mathcal{L}_{h,1}(\xi_0^h)] + \frac{\mathrm{e}+1}{\mathrm{e}\bar{\upsilon}_h + 1}\Big) \\
& \times \sup_{n\geq 1}\Big\{n^{\beta}\exp\Big(2^{2\beta+1}(\mathrm{e}\bar{\upsilon}_h + 1)\zeta_{h,1}\gamma_1^2\varphi_{1-2\beta}(n+1) - \frac{\lambda_{h,1}}{2}\gamma_1\varphi_{1-\beta}(n+1)\Big)\Big\} \\
& + (\mathrm{e}+1)\zeta_{h,1}\Big(2^{2\beta}\gamma_1 \sup_{n\geq 1}\Big\{n^{\beta}\exp\big(- 2^{-(\beta+2)}\lambda_{h,1}\gamma_1 n^{1-\beta}\big)\varphi_{1-2\beta}(n+1)\Big\} + \frac{2^{\beta+1}}{\lambda_{h,1}}\Big).
\end{aligned}
$$

Step 3. Inequality on $\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)]$.
We now take $q = 2$ and $\upsilon = \bar{\upsilon}_h$ in (3.32) and use (3.40) or (3.47). Hence

$$
\begin{aligned}
\mathbb{E}[\mathcal{L}_{h,2}(\xi_{n+1}^h)] &\leq \mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)]\big(1 - \lambda_{h,2}\gamma_{n+1} + \zeta_{h,2}\gamma_{n+1}^2\big) + \zeta_{h,2}\gamma_{n+1}^2\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)] + \zeta_{h,2}\gamma_{n+1}^3 \\
&\leq \mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)]\big(1 - \lambda_{h,2}\gamma_{n+1} + (\mathrm{e}\bar{\upsilon}_h^2 + 1)\zeta_{h,2}\gamma_{n+1}^2\big) + (K_{h,1}^{\beta} + 1)\zeta_{h,2}\gamma_{n+1}^3.
\end{aligned}
\tag{3.48}
$$

It follows from (3.34) that $1-\lambda_{h,2}\gamma_{n+1}-(\mathrm{e}\bar{v}_h^2+1)\zeta_{h,2}\gamma_{n+1}^2 \geq 1/2$. Iterating n times inequality (3.48) thus gives

$$\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)] \leq \mathbb{E}[\mathcal{L}_{h,2}(\xi_0^h)]\Pi_{1:n}^{h,2} + (2^\beta K_{h,1}^\beta + 1)\zeta_{h,2}\sum_{k=1}^n \gamma_k^3\,\Pi_{k+1:n}^{h,2}. \tag{3.49}$$

We skip here the computations that are similar to those performed in the previous step.

Step 3.1.
If $\gamma_n = \gamma_1 n^{-1}$, we obtain:

$$\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)] \leq \widehat{K}_{h,2}\frac{\mathbb{E}[\mathcal{L}_{h,2}(\xi_0^h)]}{(n+1)^{\lambda_{h,2}\gamma_1}} + \bar{K}_{h,2}\frac{\varphi_{\lambda_{h,2}\gamma_1-2}(n+1)}{(n+1)^{\lambda_{h,2}\gamma_1}}, \tag{3.50}$$

where $\widehat{K}_{h,2} := \exp\big((\mathrm{e}\bar{v}_h^2+1)\zeta_{h,2}\frac{\pi^2}{6}\gamma_1^2+\frac{\gamma_1\lambda_{h,2}}{2}\big)$ and $\bar{K}_{h,2} := \widehat{K}_{h,2}\gamma_1^3(2K_{h,1}^1+1)\zeta_{h,2}2^{\gamma_1\lambda_{h,2}\vee 3}$. Hence, since $\lambda_{h,2} > \lambda_{h,1} > 1$, one has $\gamma_1\lambda_{h,2} > 1$ so that for any positive integer n

$$\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)] \leq K_{h,2}^1\gamma_n, \tag{3.51}$$

with

$$K_{h,2}^1 := \gamma_1^{-1}\Big(\widehat{K}_{h,2}\mathbb{E}[\mathcal{L}_{h,2}(\xi_0^h)] + \bar{K}_{h,2}\sup_{n\geq 1}\frac{\varphi_{\lambda_{h,2}\gamma_1-2}(n+1)}{(n+1)^{\lambda_{h,2}\gamma_1-1}}\Big).$$

Step 3.2.
Otherwise, if $\gamma_n = \gamma_1 n^{-\beta}$ with $\beta \in (0,1)$, then it holds

$$\begin{aligned}\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)] \leq{}& \bigg(\mathbb{E}[\mathcal{L}_{h,2}(\xi_0^h)] + \frac{\gamma_1(2^\beta K_{h,1}^\beta+1)}{\mathrm{e}\bar{v}_h^2+1}\bigg)\\ &\times\exp\Big(2^{2\beta+1}\zeta_{h,2}(\mathrm{e}\bar{v}_h^2+1)\gamma_1^2\varphi_{1-2\beta}(n+1) - \frac{\lambda_{h,2}}{2}\gamma_1\varphi_{1-\beta}(n+1)\Big)\\ &+(2^\beta K_{h,1}^\beta+1)\zeta_{h,2}\Big(2^{2\beta}\gamma_1^3\exp\big(-2^{-(\beta+2)}\lambda_{h,2}\gamma_1 n^{1-\beta}\big)\varphi_{1-2\beta}(n+1) + \frac{2^{2\beta+1}\gamma_1^2}{\lambda_{h,2}n^{2\beta}}\Big).\end{aligned} \tag{3.52}$$

We thus conclude that, for any $\beta \in (0,1)$ and any positive integer n,

$$\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)] \leq K_{h,2}^\beta\gamma_n, \tag{3.53}$$

with

$$\begin{aligned}K_{h,2}^\beta :={}& \bigg(\gamma_1^{-1}\mathbb{E}[\mathcal{L}_{h,2}(\xi_0^h)] + \frac{2^\beta K_{h,1}^\beta+1}{\mathrm{e}\bar{v}_h^2+1}\bigg)\\ &\times\sup_{n\geq1}\Big\{n^\beta\exp\Big(2^{2\beta+1}\zeta_{h,2}(\mathrm{e}\bar{v}_h^2+1)\gamma_1^2\varphi_{1-2\beta}(n+1) - \frac{\lambda_{h,2}}{2}\gamma_1\varphi_{1-\beta}(n+1)\Big)\Big\}\\ &+(2^\beta K_{h,1}^\beta+1)\zeta_{h,2}\Big(\gamma_1^2 2^{2\beta}\sup_{n\geq1}\Big\{n^\beta\exp\big(-2^{-(\beta+2)}\lambda_{h,2}\gamma_1 n^{1-\beta}\big)\varphi_{1-2\beta}(n+1)\Big\} + \gamma_1\frac{2^{2\beta+1}}{\lambda_{h,2}}\Big).\end{aligned}$$

Step 4. Inequality on $\mathbb{E}[(\xi_n^h-\xi_\star^h)^2]$***.***
Combining either estimate (3.40) with (3.51) or estimate (3.47) with (3.53), and using Lemma 3.2(iv) for $q=1$, we conclude

$$\mathbb{E}[(\xi_n^h-\xi_\star^h)^2] \leq \bar{K}_{h,2}^\beta\gamma_n, \tag{3.54}$$

with $\bar{K}^{\beta}_{h,2} := \kappa_{h,1}(K^{\beta}_{h,2} + K^{\beta}_{h,1})$ satisfying $\sup_{h\in\mathcal{H}} \bar{K}^{\beta}_{h,2} < \infty$ for any $\beta \in (0,1]$ since (3.12) implies that

$$\sup_{h\in\mathcal{H}} \mathbb{E}[\mathcal{L}_{h,q}(\xi^h_0)] \leq \sup_{h\in\mathcal{H}} \mathbb{E}\Big[(1+|\xi^h_0|^2)\exp\Big(\frac{4}{1-\alpha}k_\alpha \sup_{h\in\mathcal{H}} \|f_{X_h}\|_\infty |\xi^h_0|\Big)\Big] < \infty,\ q=1,2.$$

Step 5. Inequality on $\mathbb{E}[|C^h_n - C^h_\star|]$.

We now prove an $L^1(\mathbb{P})$ upper estimate for the difference $C^h_n - C^h_\star$, where the sequence $(C^h_n)_{n\geq 0}$ is given by (3.6). Without loss of generality, we here assume that $C^h_0 = 0$ inasmuch the general case $C^h_0 \neq 0$ is handled in a completely similar way. Observe that

$$\begin{aligned} C^h_n - C^h_\star &= \frac{1}{n}\sum_{k=1}^n \Big(\xi^h_{k-1} + \frac{1}{1-\alpha}(X^{(k)}_h - \xi^h_{k-1})^+\Big) - V_h(\xi^h_\star) \\ &= \frac{1}{n}\sum_{k=1}^n \varepsilon^h_k + \frac{1}{n}\sum_{k=1}^n V_h(\xi^h_{k-1}) - V_h(\xi^h_\star), \end{aligned} \tag{3.55}$$

where $\big\{\varepsilon^h_k := \xi^h_{k-1} + \frac{1}{1-\alpha}(X^{(k)}_h - \xi^h_{k-1})^+ - V_h(\xi^h_{k-1}), k\geq 1\big\}$ is a sequence of $(\mathcal{F}^h,\mathbb{P})$-martingale increments, that is, $\mathbb{E}[\varepsilon^h_k|\mathcal{F}^h_{k-1}] = 0$ and

$$\begin{aligned} \mathbb{E}[|\varepsilon^h_k|^2|\mathcal{F}^h_{k-1}] &\leq \frac{1}{(1-\alpha)^2}\mathbb{E}\Big[\big((X^{(k)}_h - \xi^h_{k-1})^+ - \mathbb{E}[(X^{(k)}_h - \xi^h_{k-1})^+|\mathcal{F}^h_{k-1}]\big)^2\Big|\mathcal{F}^h_{k-1}\Big] \\ &\leq \frac{1}{(1-\alpha)^2}\mathbb{E}\big[((X^{(k)}_h - \xi^h_{k-1})^+)^2\big|\mathcal{F}^h_{k-1}\big] \\ &\leq \frac{3}{(1-\alpha)^2}\big(\mathbb{E}[|X_h|^2] + (\xi^h_{k-1} - \xi^h_\star)^2 + |\xi^h_\star|^2\big). \end{aligned}$$

Using the above estimate together with the Cauchy-Schwarz inequality and (3.16), we obtain

$$\begin{aligned} \frac{1}{n}\mathbb{E}\Big[\Big|\sum_{k=1}^n \varepsilon^h_k\Big|\Big] &\leq \frac{1}{n}\Big(\sum_{k=1}^n \mathbb{E}[|\varepsilon^h_k|^2]\Big)^{\frac{1}{2}} \\ &\leq \frac{\sqrt{3}}{(1-\alpha)n}\Big(n\big(\sup_{h\in\mathcal{H}}\mathbb{E}[|X_h|^2] + \sup_{h\in\mathcal{H}}|\xi^h_\star|^2\big) + \bar{K}^\beta_h \sum_{k=2}^n \gamma_{k-1} + \mathbb{E}[(\xi^h_0 - \xi^h_\star)^2]\Big)^{\frac{1}{2}} \\ &\leq \frac{\sqrt{3}}{(1-\alpha)n^{\frac{1}{2}}}\Big(\big(\sup_{h\in\mathcal{H}}\mathbb{E}[|X_h|^2]^{\frac{1}{2}} + \sup_{h\in\mathcal{H}}|\xi^h_\star|\big) + (\gamma_1\bar{K}^\beta_h)^{\frac{1}{2}}(1-\beta)^{-\frac{1}{2}}n^{-\frac{\beta}{2}}\mathbf{1}_{\beta\in(0,1)} \\ &\qquad + (\gamma_1\bar{K}^\beta_h)^{\frac{1}{2}}\frac{(\ln n)^{\frac{1}{2}}}{n^{\frac{1}{2}}}\mathbf{1}_{\beta=1} + \frac{\mathbb{E}[(\xi^h_0 - \xi^h_\star)^2]^{\frac{1}{2}}}{n^{\frac{1}{2}}}\Big). \end{aligned}$$

For the second term appearing in the right hand side of (3.55), we use a second order Taylor expansion together with the fact that $\sup_{h\in\mathcal{H}}\|V''_h\|_\infty < \infty$ and then (3.16). We obtain

$$\begin{aligned} \frac{1}{n}\mathbb{E}\Big[\Big|\sum_{k=1}^n V_h(\xi^h_{k-1}) - V_h(\xi^h_\star)\Big|\Big] &\leq \frac{\|V''_h\|_\infty}{2n}\sum_{k=1}^n \mathbb{E}[(\xi^h_{k-1} - \xi^h_\star)^2] \\ &\leq \frac{\|V''_h\|_\infty}{2}\Big(\frac{\mathbb{E}[(\xi^h_0 - \xi^h_\star)^2]}{n} + \gamma_1 K^\beta_h\Big((1-\beta)^{-1}n^{-\beta}\mathbf{1}_{\beta\in(0,1)} + \frac{\ln n}{n}\mathbf{1}_{\beta=1}\Big)\Big). \end{aligned}$$

Coming back to (3.55) and combining the two previous estimates concludes the proof.

Step 6. Inequality on $\mathbb{E}[(\xi_n^h-\xi_\star^h)^4]$.
The proof of (3.17) relies on similar arguments as those used for (3.16). We will thus be brief and omit some technical details. We first note that using either (3.50) with $\gamma_1\lambda_{h,2}>2$ or (3.52), it holds

$$\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)]\leq \widetilde{K}_{h,2}^{\beta}\gamma_n^2, \tag{3.56}$$

for some constants $\widetilde{K}_{h,2}^{\beta}$ satisfying $\sup_{h\in\mathcal{H}}\widetilde{K}_{h,2}^{\beta}<\infty$. Then, we take $q=3$ and $\upsilon=\bar{\upsilon}_h$ in (3.32) and use the previous inequality. Hence,

$$\begin{aligned}\mathbb{E}[\mathcal{L}_{h,3}(\xi_{n+1}^h)]&\leq \mathbb{E}[\mathcal{L}_{h,3}(\xi_n^h)]\big(1-\lambda_{h,3}\gamma_{n+1}+\zeta_{h,3}\gamma_{n+1}^2\big)+\zeta_{h,3}\gamma_{n+1}^2\mathbb{E}[\mathcal{L}_{h,2}(\xi_n^h)]+\zeta_{h,3}\gamma_{n+1}^4\\&\leq \mathbb{E}[\mathcal{L}_{h,3}(\xi_n^h)]\big(1-\lambda_{h,3}\gamma_{n+1}+(\mathrm{e}\bar{\upsilon}_h^3+1)\zeta_{h,3}\gamma_{n+1}^2\big)+(2^{2\beta}\widetilde{K}_{h,2}^{\beta}+1)\zeta_{h,3}\gamma_{n+1}^4,\end{aligned} \tag{3.57}$$

so that

$$\mathbb{E}[\mathcal{L}_{h,3}(\xi_n^h)]\leq \mathbb{E}[\mathcal{L}_{h,3}(\xi_0^h)]\Pi_{1:n}^{h,3}+(2^{2\beta}\widetilde{K}_{h,2}^{\beta}+1)\zeta_{h,3}\sum_{k=1}^n\gamma_k^4\,\Pi_{k+1:n}^{h,3}. \tag{3.58}$$

Following similar lines of reasoning as those used in Step 2, we conclude that if $\gamma_n=\gamma_1 n^{-\beta}$, $\beta\in(0,1]$, with $\gamma_1\lambda_{h,3}>2$ if $\beta=1$, then it holds

$$\mathbb{E}[\mathcal{L}_{h,3}(\xi_n^h)]\leq K_{h,3}^{\beta}\gamma_n^2, \tag{3.59}$$

for some constants $K_{h,3}^{\beta}$ satisfying $\sup_{h\in\mathcal{H}}K_{h,3}^{\beta}<\infty$.

Finally, we take $q=4$ and $\upsilon=\bar{\upsilon}_h$ in (3.32) and use (3.59). We thus obtain

$$\begin{aligned}\mathbb{E}[\mathcal{L}_{h,4}(\xi_{n+1}^h)]&\leq \mathbb{E}[\mathcal{L}_{h,4}(\xi_n^h)]\big(1-\lambda_{h,4}\gamma_{n+1}+\zeta_{h,4}\gamma_{n+1}^2\big)+\zeta_{h,4}\gamma_{n+1}^2\mathbb{E}[\mathcal{L}_{h,3}(\xi_n^h)]+\zeta_{h,3}\gamma_{n+1}^5\\&\leq \mathbb{E}[\mathcal{L}_{h,4}(\xi_n^h)]\big(1-\lambda_{h,4}\gamma_{n+1}+(\mathrm{e}\bar{\upsilon}_h^4+1)\zeta_{h,4}\gamma_{n+1}^2\big)+(2^{2\beta}K_{h,3}^{\beta}+\gamma_1)\zeta_{h,4}\gamma_{n+1}^4,\end{aligned}$$

so that

$$\mathbb{E}[\mathcal{L}_{h,4}(\xi_n^h)]\leq \mathbb{E}[\mathcal{L}_{h,4}(\xi_0^h)]\Pi_{1:n}^{h,4}+(2^{2\beta}K_{h,3}^{\beta}+\gamma_1)\zeta_{h,4}\sum_{k=1}^n\gamma_k^4\,\Pi_{k+1:n}^{h,4}. \tag{3.60}$$

In a completely analogous manner, we deduce from the previous inequality that if $\gamma_n=\gamma_1 n^{-\beta}$, $\beta\in(0,1]$, with $\gamma_1\lambda_{h,4}>2$ if $\beta=1$, then for any $h\in\mathcal{H}$ and any positive integer n

$$\mathbb{E}[\mathcal{L}_{h,4}(\xi_n^h)]\leq K_{h,4}^{\beta}\gamma_n^2, \tag{3.61}$$

where $K_{h,4}^{\beta}$ is a constant satisfying $\sup_{h\in\mathcal{H}}K_{h,4}^{\beta}<\infty$.

Combining (3.56) with (3.61) as well as Lemma 3.2(iv) with $q=2$ and recalling that $\lambda_{h,4}>\lambda_{h,2}$, we eventually conclude that if $\gamma_n=\gamma_1 n^{-\beta}$, $\beta\in(0,1]$, with $\gamma_1\lambda_{h,2}>2$ if $\beta=1$, then for any $h\in\mathcal{H}$ and any positive integer n,

$$\mathbb{E}[(\xi_n^h-\xi_\star^h)^4]\leq \bar{K}_{h,4}^{\beta}\gamma_n^2 \tag{3.62}$$

holds, where $\bar{K}_{h,4}^{\beta}\coloneqq\kappa_{h,2}(K_{h,4}^{\beta}+\widetilde{K}_{h,2}^{\beta})$ satisfies $\sup_{h\in\mathcal{H}}\bar{K}_{h,4}^{\beta}<\infty$ for any $\beta\in(0,1]$ inasmuch (3.12) implies that

$$\sup_{h\in\mathcal{H}}\mathbb{E}[\mathcal{L}_{h,q}(\xi_0^h)]\leq\sup_{h\in\mathcal{H}}\mathbb{E}\Big[(1+|\xi_0^h|^4)\exp\Big(\frac{16}{1-\alpha}k_\alpha\sup_{h\in\mathcal{H}}\|f_{X_h}\|_\infty|\xi_0^h|\Big)\Big]<\infty,\ q=3,4.$$

The proof of (3.17) is now complete.

Step 7. Inequality on $\mathbb{E}[(C_n^h - C_\star^h)^2]$.
Here again assuming that $C_0^h = 0$, from the decomposition (3.55) and similar computations to the ones performed in Step 5, we get, for some constants $\bar{C}_h^\beta$ that may change from line to line,

$$\begin{aligned}\mathbb{E}[(C_n^h - C_\star^h)^2] &\le \bar{C}_h^\beta\bigg\{\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}[|\varepsilon_k^h|^2] + \frac{1}{n^2}\bigg(\sum_{k=1}^n \mathbb{E}\big[\big(V_h(\xi_{k-1}^h) - V_h(\xi_\star^h)\big)^2\big]^{\frac{1}{2}}\bigg)^2\bigg\}\\
&\le \bar{C}_h^\beta\bigg\{\frac{1}{n} + \frac{1}{n^2}\bigg(\sum_{k=1}^n \mathbb{E}[(\xi_{k-1}^h - \xi_\star^h)^4]^{\frac{1}{2}}\bigg)^2\bigg\}\\
&\le \bar{C}_h^\beta\bigg\{\frac{1}{n} + \frac{1}{n^2}\bigg(\sum_{k=1}^n \gamma_k\bigg)^2\bigg\}\\
&\le \frac{\bar{C}_h^\beta}{n^{1\wedge 2\beta}},
\end{aligned} \tag{3.63}$$

where the last inequality follows by a comparison between series and integrals for $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0,1]$. This concludes the proof. $\square$

3.2 Complexity Analysis

As a consequence of Proposition 3.1 and Theorem 3.1, if $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0,1]$, with $\lambda_1\gamma_1 > 1$ if $\beta = 1$, then there exists $C > 0$ such that, for any $h \in \mathcal{H}$ and any positive integer n,

$$\mathbb{E}[|\xi_n^h - \xi_\star|] \le C\big(h + n^{-\frac{\beta}{2}}\big) \quad \text{and} \quad \mathbb{E}[|C_n^h - C_\star|] \le C\big(h + n^{-\frac{1}{2}\wedge\beta}\big).$$

The above inequality gives us a heuristic method to balance the bias parameter $h \in \mathcal{H}$ with respect to the number of steps n in the stochastic approximation scheme (3.6) to achieve a prescribed error of order $\varepsilon \in (0,1)$. Indeed, letting

$$h = \varepsilon \quad \text{and} \quad n = \big\lceil \varepsilon^{-\frac{2}{\beta}} \big\rceil,$$

gives a global $L^1(\mathbb{P})$ error of order ε. The corresponding computational cost of the algorithm is thus given by

$$\mathrm{Cost} = Cnh^{-1} \sim C\varepsilon^{-\frac{2}{\beta}-1} \quad \text{as} \quad \varepsilon \downarrow 0,$$

for some constant C independent of ε. Observe that the minimal computational cost is of order ε^{-3} and is achieved if one selects $\gamma_n = \gamma_1 n^{-\beta}$ with $\beta = 1$ under the restriction $\lambda_1\gamma_1 > 1$.

The next section is devoted to the development of the MLSA algorithm which is a different approach of constructing estimators of VaR and ES by combining multiple SA estimates obtained through Algorithm 1 for a geometric sequence of bias parameters, all the while reducing the complexity of the entailed simulations.

4 Multilevel Stochastic Approximation

The first part explains the construction of the multilevel SA algorithm to approximate VaR and ES, and the second part provides an analysis of the proposed algorithm on both convergence and complexity aspects.

Let $h \in \mathcal{H}$ be a fixed bias parameter and M and L (the number of levels) be two positive integers such that $M \geq 2$. We consider the geometric sequence of bias parameters

$$h_\ell = \frac{h}{M^\ell}, \quad \ell = 0, \ldots, L.$$

Under Assumptions 1 and 2, letting $(\xi_\star^{h_\ell}, C_\star^{h_\ell}) = (\arg\min V_{h_\ell}, \min V_{h_\ell})$ for $\ell \in \{0, \ldots, L\}$, we use the following decomposition in telescopic sum

$$\xi_\star^{h_L} = \xi_\star^{h_0} + \sum_{\ell=1}^{L} \xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}, \tag{4.1}$$

$$C_\star^{h_L} = C_\star^{h_0} + \sum_{\ell=1}^{L} C_\star^{h_\ell} - C_\star^{h_{\ell-1}}. \tag{4.2}$$

In the previous section, we saw how to approximate each pair $(\xi_\star^{h_\ell}, C_\star^{h_\ell})$ through Algorithm 1. We let $\mathbf{N} = (N_0, \ldots, N_L) \in \mathbb{N}_0^{L+1}$. Following [12], we can define the following multilevel SA estimators $\xi_{\mathbf{N}}^{\mathrm{ML}}$ and $\mathrm{C}_{\mathbf{N}}^{\mathrm{ML}}$ of $\xi_\star^{h_L}$ and $C_\star^{h_L}$:

$$\begin{cases} \xi_{\mathbf{N}}^{\mathrm{ML}} &= \xi_{N_0}^{h_0} + \sum_{\ell=1}^{L} \xi_{N_\ell}^{h_\ell} - \xi_{N_\ell}^{h_{\ell-1}}, \\ \mathrm{C}_{\mathbf{N}}^{\mathrm{ML}} &= C_{N_0}^{h_0} + \sum_{\ell=1}^{L} C_{N_\ell}^{h_\ell} - C_{N_\ell}^{h_{\ell-1}}, \end{cases} \tag{4.3}$$

where the initializations (ξ_0^h, C_0^h) are generated such that $\mathbb{E}[|\xi_0^h|^2] + \mathbb{E}[|C_0^h|^2] < \infty$, and the iterates $(\xi_n^h, C_n^h)_{n\geq 1}$ are computed using the SA scheme (3.6). However, we importantly point out that for any fixed level $\ell \in \{1, \ldots, L\}$, the random variables $(X_{h_{\ell-1}}^{(n)}, X_{h_\ell}^{(n)})_{1\leq n\leq N_\ell}$ used in the SA schemes for the computation of $(\xi_{N_\ell}^{h_{\ell-1}}, \xi_{N_\ell}^{h_\ell})$, $(C_{N_\ell}^{h_{\ell-1}}, C_{N_\ell}^{h_\ell})$ are iid with the same law as $(X_{h_{\ell-1}}, X_{h_\ell})$. In particular, one obtains X_{h_ℓ} from $X_{h_{\ell-1}}$ using the identity

$$X_{h_\ell} = \frac{1}{M} X_{h_{\ell-1}} + \frac{1}{KM^\ell} \sum_{k=KM^{\ell-1}+1}^{KM^\ell} \varphi(Y, Z^{(k)}). \tag{4.4}$$

Algorithm 2 summarizes this process.

Remark 1. *Intuitively, the larger the ℓ, the closer the random variables X_{h_ℓ} and $X_{h_{\ell-1}}$ are to X_0, so that less iterations N_ℓ of the SA scheme* (3.6) *are required at level ℓ.*

Algorithm 2 Multilevel SA algorithm for estimating $(\mathrm{VaR}, \mathrm{ES})$

Require: A number of levels $L \geq 1$, a bias parameter $h = \frac{1}{K}$, a geometric step size $M \geq 1$, a positive integer sequence $N_0, \dots, N_L$, a positive and non-increasing sequence $\{\gamma_n, 1 \leq n \leq \max_\ell N_\ell\}$.

for $\ell = 0 \,..\, L$ **do**
  Set $h_\ell \leftarrow \frac{h}{M^\ell}$
  for $j = (\ell-1)^+, \ell$ **do**
    Choose $(\xi_0^{h_j}, C_0^{h_j})$ such that $\mathbb{E}[|\xi_0^{h_j}|^2] + \mathbb{E}[|C_0^{h_j}|^2] < \infty$
  end for
  for $n = 0 \,..\, N_\ell - 1$ **do**
    Simulate $Y^{(n+1)} \sim Y$
    Simulate $Z^{(n+1,1)}, \dots, Z^{(n+1,KM^\ell)} \overset{\text{iid}}{\sim} Z$ independently from Y
    if $\ell = 0$ **then**
      $X_{h_0}^{(n+1)} \leftarrow \frac{1}{K}\sum_{k=1}^{K} \varphi(Y^{(n+1)}, Z^{(n+1,k)})$
    else
      $X_{h_{\ell-1}}^{(n+1)} \leftarrow \frac{1}{KM^{\ell-1}}\sum_{k=1}^{KM^{\ell-1}} \varphi(Y^{(n+1)}, Z^{(n+1,k)})$
      $X_{h_\ell}^{(n+1)} \leftarrow \frac{1}{M} X_{h_{\ell-1}}^{(n+1)} + \frac{1}{KM^\ell}\sum_{k=KM^{\ell-1}+1}^{KM^\ell} \varphi(Y^{(n+1)}, Z^{(n+1,k)})$
    end if
    for $j = (\ell-1)^+$ and $j = \ell$ **do**
      $\xi_{n+1}^{h_j} \leftarrow \xi_n^{h_j} - \gamma_{n+1} H_1(\xi_n^{h_j}, X_{h_j}^{(n+1)})$
      $C_{n+1}^{h_j} \leftarrow C_n^{h_j} - \frac{1}{n+1} H_2(C_n^{h_j}, \xi_n^{h_j}, X_{h_j}^{(n+1)})$
    end for
  end for
end for
$\xi_{\mathbf{N}}^{\mathrm{ML}} \leftarrow \xi_{N_0}^{h_0} + \sum_{\ell=1}^{L} \xi_{N_\ell}^{h_\ell} - \xi_{N_\ell}^{h_{\ell-1}}$
$\mathrm{C}_{\mathbf{N}}^{\mathrm{ML}} \leftarrow C_{N_0}^{h_0} + \sum_{\ell=1}^{L} C_{N_\ell}^{h_\ell} - C_{N_\ell}^{h_{\ell-1}}$
return $\xi_{\mathbf{N}}^{\mathrm{ML}}$ and $\mathrm{C}_{\mathbf{N}}^{\mathrm{ML}}$

4.1 Convergence Analysis

The global error between the multilevel estimator $(\xi_{\mathbf{N}}^{\mathrm{ML}}, \mathrm{C}_{\mathbf{N}}^{\mathrm{ML}})$ and its target $(\xi_\star, C_\star)$ can be decomposed into a sum of a statistical error and a bias error:

$$\begin{aligned} \xi_{\mathbf{N}}^{\mathrm{ML}} - \xi_\star &= \big(\xi_{\mathbf{N}}^{\mathrm{ML}} - \xi_\star^{h_L}\big) + (\xi_\star^{h_L} - \xi_\star), \\ \mathrm{C}_{\mathbf{N}}^{\mathrm{ML}} - C_\star &= \big(\mathrm{C}_{\mathbf{N}}^{\mathrm{ML}} - C_\star^{h_L}\big) + (C_\star^{h_L} - C_\star). \end{aligned}$$

In the following analysis, we quantify each error that appears in the decomposition above in terms of the parameters of Algorithm 2. We then propose a specific choice of L and $N_0, \dots, N_L$ to achieve to achieve some prescribed error. The next result is inspired from [17, Lemma 5.1, Proposition 5.2] and [6, Lemma A.4]. We provide the proof in Appendix D for the sake of completeness. The proof of Proposition 4.1 is postponed to Appendix E.

Lemma 4.1. *Let X and Y be two real-valued random variables with bounded densities denoted by f_X and f_Y respectively.*

(i) *Assume that $X - Y$ is in $L^{p_\star}(\mathbb{P})$ for some $p_\star > 1$. Then, for every $\xi \in \mathbb{R}$,*

$$\mathbb{E}\big[\big|\mathbf{1}_{X>\xi} - \mathbf{1}_{Y>\xi}\big|\big] \leq \Big(p_\star^{\frac{p_\star}{p_\star+1}} + p_\star^{\frac{1}{p_\star+1}}\Big)\big(\|f_X\|_\infty + \|f_Y\|_\infty\big)^{\frac{p_\star}{p_\star+1}} \mathbb{E}[|X-Y|^{p_\star}]^{\frac{1}{p_\star+1}}.$$

(ii) *Assume that there exists a positive constant $d_{X,Y}$ such that, for any $u \in \mathbb{R}$,*

$$\mathbb{E}\big[\exp\big(u(X-Y)\big)\big] \le \exp\big(d_{X,Y}u^2\big). \tag{4.5}$$

Then, for every $\xi \in \mathbb{R}$,

$$\mathbb{E}\big[\big|\mathbf{1}_{X>\xi} - \mathbf{1}_{Y>\xi}\big|\big] \le 2\sqrt{d_{X,Y}} + \big(\|f_X\|_\infty + \|f_Y\|_\infty\big)\sqrt{2d_{X,Y}\ln\big(d_{X,Y}^{-1} \vee 1\big)}.$$

Proposition 4.1. (i) *Assume that for any $h \in \mathcal{H}$, the real-valued random variable X_h admits a density f_{X_h}, bounded uniformly in $h \in \mathcal{H}$.*

a. *If*

$$\mathbb{E}\big[\big|\varphi(Y,Z) - \mathbb{E}[\varphi(Y,Z)|Y]\big|^{p_\star}\big] < \infty \quad \textit{holds for some } p_\star > 1, \tag{4.6}$$

then, for any $h, h' \in \mathcal{H}$ such that $0 < h \le h'$ and any $\xi \in \mathbb{R}$,

$$\mathbb{E}\big[\big|\mathbf{1}_{X_h>\xi} - \mathbf{1}_{X_{h'}>\xi}\big|\big] \le C(h'-h)^{\frac{p_\star}{2(p_\star+1)}},$$

with $C := B_{p_\star}\mathbb{E}\big[\big|\varphi(Y,Z) - \mathbb{E}[\varphi(Y,Z)|Y]\big|^{p_\star}\big]^{\frac{1}{p_\star+1}}(\sup_{h\in\mathcal{H}}\|f_{X_h}\|_\infty)^{\frac{p_\star}{p_\star+1}}$, where $B_{p_\star}$ is a positive constant which depends only upon $p_\star$.

b. *Assume that there exists a non negative constant $C_g < \infty$ such that, for all $u \in \mathbb{R}$,*

$$\mathbb{E}\big[\exp\big(u\big(\varphi(Y,Z) - \mathbb{E}[\varphi(Y,Z)|Y]\big)\big)\big|Y\big] \le \mathrm{e}^{C_g u^2} \quad \mathbb{P}\text{-as.} \tag{4.7}$$

Then, for any $h, h' \in \mathcal{H}$ such that $0 < h \le h'$ and any $\xi \in \mathbb{R}$,

$$\mathbb{E}\big[\big|\mathbf{1}_{X_h>\xi} - \mathbf{1}_{X_{h'}>\xi}\big|\big] \le 2\sqrt{C_g(h'-h)}\Big(1 + \sup_{h\in\mathcal{H}}\|f_{X_h}\|_\infty\sqrt{2\big|\ln\big(C_g(h'-h)\big)\big|}\Big).$$

(ii) *Define $G_\ell := h_\ell^{-\frac{1}{2}}(X_{h_\ell} - X_{h_{\ell-1}})$ and let $F_{X_{h_{\ell-1}}\,|\,G_\ell=g} : x \mapsto \mathbb{P}(X_{h_{\ell-1}} \le x \,|\, G_\ell = g)$, for $g \in \operatorname{supp}(\mathbb{P}_{G_\ell})$ and $\ell \ge 1$. Assume that the sequence of random variables $(K_\ell)_{\ell\ge1}$ defined by $K_\ell := K_\ell(G_\ell)$, where*

$$K_\ell(g) := \sup_{x\ne y}\frac{\big|F_{X_{h_{\ell-1}}\,|\,G_\ell=g}(x) - F_{X_{h_{\ell-1}}\,|\,G_\ell=g}(y)\big|}{|x-y|}, \quad \ell \ge 1,$$

satisfies

$$\sup_{\ell\ge1}\mathbb{E}[|G_\ell|K_\ell] < \infty. \tag{4.8}$$

Then, it holds

$$\sup_{\ell\ge1,\xi\in\mathbb{R}} h_\ell^{-\frac{1}{2}}\mathbb{E}\big[\big|\mathbf{1}_{X_{h_\ell}>\xi} - \mathbf{1}_{X_{h_{\ell-1}}>\xi}\big|\big] < \infty.$$

Remark 2. *We point out that if $F_{X_{h_{\ell-1}}\,|\,G_\ell=g}$ is Lipschitz continuous uniformly in g and ℓ then it follows from* (E.2)*, with $h' = h_{\ell-1}$ and $h = h_\ell$, that* (4.8) *is satisfied.*

Assumption 3. *There exist $C < \infty$ and $\delta > 0$ such that, for any $h \in \mathcal{H}$ and any compact set $\mathcal{K} \subset \mathbb{R}$,*

$$\sup_{\xi\in\mathcal{K}}|f_{X_h}(\xi) - f_{X_0}(\xi)| \le Ch^{\frac{1}{4}+\delta}.$$

We can now state the following result regarding the non-asymptotic $L^2(\mathbb{P})$ error of Algorithm 2.

Theorem 4.1. *Suppose that for any $h \in \mathcal{H} \cup \{0\}$ the random variable $X_h \in L^2(\mathbb{P})$ and that Assumptions 1, 2 and 3 are satisfied. Then, within the framework of Proposition 4.1, if $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0,1]$, with $\lambda_2\gamma_1 > 2$ if $\beta = 1$, and $\sup_{h\in\mathcal{H}} \mathbb{E}[|\xi_0^h|^4 \exp(\frac{16}{1-\alpha} k_\alpha \sup_{h\in\mathcal{H}} \|f_{X_h}\|_\infty |\xi_0^h|)] < \infty$, $\sup_{h\in\mathcal{H}} \mathbb{E}[(C_0^h)^2] < \infty$, then there exists some constant $K < \infty$ such that, for any positive integer L and any $\mathbf{N} = (N_0, \ldots, N_L) \in \mathbb{N}_0^{L+1}$,*

$$\mathbb{E}\big[\big(\xi^{\mathrm{ML}}_{\mathbf{N}} - \xi^{h_L}_\star\big)^2\big] \le K\bigg(\gamma_{N_0} + \bigg(\sum_{\ell=1}^{L} \gamma_{N_\ell}\bigg)^2 + \sum_{\ell=1}^{L} \gamma_{N_\ell}^{\frac{3}{2}} + \sum_{\ell=1}^{L} \gamma_{N_\ell}\epsilon(h_\ell)\bigg), \tag{4.9}$$

and

$$\mathbb{E}\big[\big(\mathrm{C}^{\mathrm{ML}}_{\mathbf{N}} - C^{h_L}_\star\big)^2\big] \le K\bigg(\frac{1}{N_0^{1\wedge 2\beta}} + \sum_{\ell=1}^{L} \frac{h_\ell}{N_\ell} + \sum_{\ell=1}^{L} \frac{\bar\gamma_{N_\ell}}{N_\ell}\epsilon(h_\ell) + \sum_{\ell=1}^{L} \frac{1}{N_\ell^2} \sum_{k=1}^{N_\ell} \gamma_k^{\frac{3}{2}} + \bigg(\sum_{\ell=1}^{L} \bar\gamma_{N_\ell}\bigg)^2\bigg), \tag{4.10}$$

where $\bar\gamma_n = n^{-1}\sum_{k=1}^n \gamma_k$ and

$$\epsilon(h_\ell) \coloneqq \begin{cases} h_\ell^{\frac{p_*}{2(1+p_*)}} & \text{if (4.6) is satisfied,} \\ h_\ell^{\frac{1}{2}} |\ln h_\ell|^{\frac{1}{2}} & \text{if (4.7) is satisfied,} \\ h_\ell^{\frac{1}{2}} & \text{if (4.8) is satisfied.} \end{cases} \tag{4.11}$$

Proof. ***Step 1: Proof of* (4.9).**
We follow a similar strategy to the one used in [12, Lemma 2.7] and decompose the dynamics of the sequence $(\xi_n^h)_{n\ge0}$ given by (3.6) as follows:

$$\begin{aligned} \xi_{n+1}^h - \xi_\star^h = \big(1 - \gamma_{n+1} V_0''(\xi_\star)\big)(\xi_n^h - \xi_\star^h) + \gamma_{n+1}\big(V_0''(\xi_\star) - V_h''(\xi_\star^h)\big)(\xi_n^h - \xi_\star^h) \\ + \gamma_{n+1} r_{n+1}^h + \gamma_{n+1}\rho_{n+1}^h + \gamma_{n+1} e_{n+1}^h, \end{aligned} \tag{4.12}$$

where

$$r_{n+1}^h \coloneqq V_h''(\xi_\star^h)(\xi_n^h - \xi_\star^h) - V_h'(\xi_n^h), \tag{4.13}$$

$$\rho_{n+1}^h \coloneqq V_h'(\xi_n^h) - V_h'(\xi_\star^h) - \big(H_1(\xi_n^h, X_h^{(n+1)}) - H_1(\xi_\star^h, X_h^{(n+1)})\big), \tag{4.14}$$

$$e_{n+1}^h \coloneqq V_h'(\xi_\star^h) - H_1(\xi_\star^h, X_h^{(n+1)}). \tag{4.15}$$

Iterating n times (4.12), we obtain

$$\begin{aligned} \xi_n^h - \xi_\star^h = (\xi_0^h - \xi_\star^h)\Pi_{1:n} + \sum_{k=1}^n \gamma_k \Pi_{k+1:n}\big(V_0''(\xi_\star) - V_h''(\xi_\star^h)\big)(\xi_k^h - \xi_\star^h) \\ + \sum_{k=1}^n \gamma_k \Pi_{k+1:n} r_k^h + \sum_{k=1}^n \gamma_k \Pi_{k+1:n} \rho_k^h + \sum_{k=1}^n \gamma_k \Pi_{k+1:n} e_k^h, \end{aligned} \tag{4.16}$$

where, for two positive integers i and n,

$$\Pi_{i:n} \coloneqq \prod_{j=i}^{n} \big(1 - \gamma_j V_0''(\xi_\star)\big),$$

with the convention that $\prod_{\varnothing} = 1$.

Using the inequality $1 + x \leq \exp(x)$ for any $x \in \mathbb{R}$, a comparison between series and integrals and (3.37), we get

$$
\begin{aligned}
|\Pi_{i:n}| &\leq \exp\Big(- V_0''(\xi_\star) \sum_{j=i}^{n} \gamma_j \Big) \leq \\
&\begin{cases} \exp(\gamma_1 V_0''(\xi_\star))\, i^{\gamma_1 V_0''(\xi_\star)} (n+1)^{-\gamma_1 V_0''(\xi_\star)}, & \beta = 1, \\ \exp\big(- \gamma_1 V_0''(\xi_\star) \big(\varphi_{1-\beta}(n+1) - \varphi_{1-\beta}(i) \big) \big), & \beta \in (0,1). \end{cases}
\end{aligned}
\tag{4.17}
$$

Since $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0,1]$ with $\gamma_1 V_0''(\xi_\star) > \gamma_1 \lambda_2 > 1$ if $\beta = 1$, using (4.17) and a comparison between series and integrals (with computations similar to those performed in Step 2 of the proof of Theorem 3.1), we deduce that for any $b \geq 0$ and $a > 0$ such that $\gamma_1 a V_0''(\xi_\star) > b$ if $\beta = 1$, there exists $K < \infty$ such that for any integer n

$$
\sum_{k=1}^{n} \gamma_k^{1+b} |\Pi_{k+1:n}|^a \leq K \gamma_n^b, \tag{4.18}
$$

as well as

$$
|\Pi_{1:n}| \leq K \gamma_n. \tag{4.19}
$$

Using (4.16), the difference between the multilevel SA estimator $\xi_{\mathbf{N}}^{\mathrm{ML}}$ of the VaR in (4.3) and the solution $\xi_\star^{h_L} = \arg\min V_{h_L}$ can be decomposed as follows:

$$
\begin{aligned}
\xi_{\mathbf{N}}^{\mathrm{ML}} - \xi_\star^{h_L} &= \xi_{N_0}^{h_0} - \xi_\star^{h_0} + \sum_{\ell=1}^{L} \xi_{N_\ell}^{h_\ell} - \xi_\star^{h_\ell} - \big(\xi_{N_\ell}^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}} \big) \\
&= \xi_{N_0}^{h_0} - \xi_\star^{h_0} + \sum_{\ell=1}^{L} A_{N_\ell}^{\ell} + B_{N_\ell}^{\ell} + C_{N_\ell}^{\ell} + D_{N_\ell}^{\ell} + E_{N_\ell}^{\ell},
\end{aligned}
\tag{4.20}
$$

with

$$
A_n^\ell \coloneqq \big(\xi_0^{h_\ell} - \xi_\star^{h_\ell} - \big(\xi_0^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}} \big) \big) \Pi_{1:n}, \tag{4.21}
$$

$$
\begin{aligned}
B_n^\ell \coloneqq \sum_{k=1}^{n} \gamma_k \Pi_{k+1:n} \big\{ &\big(V_0''(\xi_\star) - V_{h_\ell}''(\xi_\star^{h_\ell}) \big) \big(\xi_k^{h_\ell} - \xi_\star^{h_\ell} \big) \\
&- \big(V_0''(\xi_\star) - V_{h_{\ell-1}}''(\xi_\star^{h_{\ell-1}}) \big) \big(\xi_k^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}} \big) \big\},
\end{aligned}
\tag{4.22}
$$

$$
C_n^\ell \coloneqq \sum_{k=1}^{n} \gamma_k \Pi_{k+1:n} \big(r_k^{h_\ell} - r_k^{h_{\ell-1}} \big), \tag{4.23}
$$

$$
D_n^\ell \coloneqq \sum_{k=1}^{n} \gamma_k \Pi_{k+1:n} \big(\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}} \big), \tag{4.24}
$$

$$
E_n^\ell \coloneqq \sum_{k=1}^{n} \gamma_k \Pi_{k+1:n} \big(e_k^{h_\ell} - e_k^{h_{\ell-1}} \big). \tag{4.25}
$$

We now quantify the contribution of each term appearing in the decomposition (4.20). We first note that Theorem 3.1 guarantees that the first term satisfies

$$
\mathbb{E}\big[\big(\xi_{N_0}^{h_0} - \xi_\star^{h_0} \big)^2 \big] \leq K_{h_0}^{\beta} \gamma_{N_0}. \tag{4.26}
$$

Step 1.1.
From (4.19), for some constant K that may change from line to line

$$\mathbb{E}\big[\big|A_n^\ell\big|^2\big] \leq 2K \sup_{\ell\geq 0} \mathbb{E}[|\xi_0^{h_\ell} - \xi_\star^{h_\ell}|^2]\gamma_n^2 \leq K\gamma_n^2. \tag{4.27}$$

Step 1.2.
Let $\mathcal{K}$ be a compact set of $\mathbb{R}$ containing the sequence $(\xi_\star^{h_\ell})_{\ell\geq 0}$. For any $\ell \geq 0$, one has

$$\begin{aligned}
\big|V_0''(\xi_\star) - V_{h_\ell}''(\xi_\star^{h_\ell})\big| &\leq \frac{1}{1-\alpha}\big(\big|f_{X_0}(\xi_\star) - f_{X_0}(\xi_\star^{h_\ell})\big| + \big|f_{X_0}(\xi_\star^{h_\ell}) - f_{X_{h_\ell}}(\xi_\star^{h_\ell})\big|\big) \\
&\leq \frac{1}{1-\alpha}\Big([f_{X_0}]_{\mathrm{Lip}}|\xi_\star^{h_\ell} - \xi_\star| + \sup_{\xi\in\mathcal{K}}\big|f_{X_0}(\xi) - f_{X_{h_\ell}}(\xi)\big|\Big) \\
&\leq K\big(h_\ell + h_\ell^{\frac{1}{4}+\delta}\big) \\
&\leq K h_\ell^{(\frac{1}{4}+\delta)\wedge 1},
\end{aligned} \tag{4.28}$$

where we used Assumptions 2(ii) and 3 together with (3.8). This estimate together with Theorem 3.1 and (4.18) give

$$\begin{aligned}
\mathbb{E}\big[\big|B_n^\ell\big|^2\big]^{\frac{1}{2}} &\leq 2\big(\big|V_0''(\xi_\star) - V_{h_\ell}''(\xi_\star^{h_\ell})\big| \vee \big|V_0''(\xi_\star) - V_{h_{\ell-1}}''(\xi_\star^{h_{\ell-1}})\big|\big) \\
&\quad\times\left(\sum_{k=1}^n \gamma_k|\Pi_{k+1:n}|\big(\mathbb{E}[|\xi_k^{h_\ell} - \xi_\star^{h_\ell}|^2]^{\frac{1}{2}} \vee \mathbb{E}[|\xi_k^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}}|^2]^{\frac{1}{2}}\big)\right) \\
&\leq K h_\ell^{(\frac{1}{4}+\delta)\wedge 1}\sum_{k=1}^n \gamma_k^{\frac{3}{2}}|\Pi_{k+1:n}| \\
&\leq K\gamma_n^{\frac{1}{2}} h_\ell^{(\frac{1}{4}+\delta)\wedge 1}.
\end{aligned} \tag{4.29}$$

Step 1.3.
Using a first order Taylor expansion together with the uniform Lipschitz regularity of $(f_{X_h})_{h\in\mathcal{H}}$ and again Theorem 3.1, we get

$$\begin{aligned}
\mathbb{E}\big[\big|r_k^{h_\ell}\big|^2\big]^{\frac{1}{2}} + \mathbb{E}\big[\big|r_k^{h_{\ell-1}}\big|^2\big]^{\frac{1}{2}} &\leq \frac{\sup_{h\in\mathcal{H}}[f_{X_h}]_{\mathrm{Lip}}}{(1-\alpha)}\big(\mathbb{E}\big[\big(\xi_{k-1}^{h_\ell} - \xi_\star^{h_\ell}\big)^4\big]^{\frac{1}{2}} + \mathbb{E}\big[\big(\xi_{k-1}^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}}\big)^4\big]^{\frac{1}{2}}\big) \\
&\leq K\gamma_k,
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E}\big[\big|C_n^\ell\big|^2\big]^{\frac{1}{2}} &\leq \sum_{k=1}^n \gamma_k|\Pi_{k+1:n}|\big(\mathbb{E}\big[\big|r_k^{h_\ell}\big|^2\big]^{\frac{1}{2}} + \mathbb{E}\big[\big|r_k^{h_{\ell-1}}\big|^2\big]^{\frac{1}{2}}\big) \\
&\leq K\sum_{k=1}^n \gamma_k^2|\Pi_{k+1:n}| \\
&\leq K\gamma_n,
\end{aligned} \tag{4.30}$$

where we again used (4.18) for the last inequality.

Step 1.4.

We first note that from the very definition (2.6) of H_1 and (3.16)

$$
\begin{aligned}
\mathbb{E}\big[\big(H_1(\xi_k^h, X_h^{(k+1)}) - H_1(\xi_\star^h, X_h^{(k+1)})\big)^2\big] \\
&= \frac{1}{(1-\alpha)^2}\mathbb{E}\big[\big|\mathbf{1}_{X_h^{(k+1)}>\xi_k^h} - \mathbf{1}_{X_h^{(k+1)}>\xi_\star^h}\big|\big] \\
&= \frac{1}{(1-\alpha)^2}\mathbb{E}\big[\mathbb{E}\big[\mathbf{1}_{\xi_k^h<X_h^{(k+1)}<\xi_\star^h} + \mathbf{1}_{\xi_\star^h<X_h^{(k+1)}<\xi_k^h}\big|\mathcal{F}_k^h\big]\big] \\
&= \frac{1}{(1-\alpha)^2}\mathbb{E}\big[\big|F_{X_h}(\xi_k^h) - F_{X_h}(\xi_\star^h)\big|\big] \\
&\le \frac{\sup_{h\in\mathcal{H}}\|f_{X_h}\|_\infty}{(1-\alpha)^2}\mathbb{E}[(\xi_k^h-\xi_\star^h)^2]^{\frac{1}{2}} \\
&\le |K_h^\beta|^{\frac{1}{2}}\frac{\sup_{h\in\mathcal{H}}\|f_{X_h}\|_\infty}{(1-\alpha)^2}\gamma_k^{\frac{1}{2}}.
\end{aligned}
\tag{4.31}
$$

Remark that $(\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}})_{k\ge1}$ is a sequence of martingale increments. Then, recalling (4.31) and using again (4.18), we get

$$
\begin{aligned}
\mathbb{E}\big[\big|D_n^\ell\big|^2\big] &= \mathbb{E}\bigg[\bigg(\sum_{k=1}^n \gamma_k\Pi_{k+1:n}\big(\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}}\big)\bigg)^2\bigg] \\
&= \sum_{k=1}^n \gamma_k^2|\Pi_{k+1:n}|^2\mathbb{E}\big[\big(\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}}\big)^2\big] \\
&\le 2\sum_{k=1}^n \gamma_k^2|\Pi_{k+1:n}|^2\big(\mathbb{E}\big[\big|\rho_k^{h_\ell}\big|^2\big] + \mathbb{E}\big[\big|\rho_k^{h_{\ell-1}}\big|^2\big]\big) \\
&\le 2\sum_{k=1}^n \gamma_k^2|\Pi_{k+1:n}|^2\Big(\mathbb{E}\big[\big(H_1(\xi_k^{h_\ell}, X_{h_\ell}^{(k+1)}) - H_1(\xi_\star^{h_\ell}, X_{h_\ell}^{(k+1)})\big)^2\big] \\
&\qquad + \mathbb{E}\big[\big(H_1(\xi_k^{h_{\ell-1}}, X_{h_{\ell-1}}^{(k+1)}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}}^{(k+1)})\big)^2\big]\Big) \\
&\le K_{h_\ell}^\beta \sum_{k=1}^n \gamma_k^{\frac{5}{2}}|\Pi_{k+1:n}|^2 \\
&\le K\gamma_n^{\frac{3}{2}},
\end{aligned}
\tag{4.32}
$$

since $2\gamma_1 V_0''(\xi_\star) > 3/2$. Then, since the random variables $(D_n^\ell)_{1\le\ell\le L}$ are independent and centered, the above estimate yields

$$
\mathbb{E}\bigg[\bigg(\sum_{\ell=1}^L D_{N_\ell}^\ell\bigg)^2\bigg] = \sum_{\ell=1}^L \mathbb{E}\big[\big|D_{N_\ell}^\ell\big|^2\big] \le K\sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{3}{2}}.
$$

Step 1.5.
We again note that $(e_k^{h_\ell} - e_k^{h_{\ell-1}})_{k\ge1}$ is a sequence of martingale increments so that

$$
\begin{aligned}
\mathbb{E}\big[\big|E_n^\ell\big|^2\big] &= \mathbb{E}\bigg[\bigg(\sum_{k=1}^n \gamma_k\Pi_{k+1:n}\big(e_k^{h_\ell} - e_k^{h_{\ell-1}}\big)\bigg)^2\bigg] \\
&= \sum_{k=1}^n \gamma_k^2|\Pi_{k+1:n}|^2\mathbb{E}\big[\big(e_k^{h_\ell} - e_k^{h_{\ell-1}}\big)^2\big] \\
&\le \sum_{k=1}^n \gamma_k^2|\Pi_{k+1:n}|^2\mathbb{E}\big[\big(H_1(\xi_\star^{h_\ell}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}})\big)^2\big].
\end{aligned}
\tag{4.33}
$$

The last term in the right hand side of the above inequality can be upper-bounded as follows

$$\begin{aligned}
&\mathbb{E}\big[\big(H_1(\xi_\star^{h_\ell}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}})\big)^2\big] \\
&\leq 2\big\{\mathbb{E}\big[\big(H_1(\xi_\star^{h_\ell}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_\ell})\big)^2\big] + \mathbb{E}\big[\big(H_1(\xi_\star^{h_{\ell-1}}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}})\big)^2\big]\big\} \\
&\leq \frac{2}{(1-\alpha)^2}\Big\{\mathbb{E}\Big[\Big(\mathbf{1}_{X_{h_\ell} > \xi_\star^{h_\ell}} - \mathbf{1}_{X_{h_\ell} > \xi_\star^{h_{\ell-1}}}\Big)^2\Big] + \mathbb{E}\Big[\Big(\mathbf{1}_{X_{h_\ell} > \xi_\star^{h_{\ell-1}}} - \mathbf{1}_{X_{h_{\ell-1}} > \xi_\star^{h_{\ell-1}}}\Big)^2\Big]\Big\}.
\end{aligned} \tag{4.34}$$

On the one hand, it follows from Proposition 3.1 that

$$\begin{aligned}
\mathbb{E}\Big[\Big(\mathbf{1}_{X_{h_\ell} > \xi_\star^{h_\ell}} - \mathbf{1}_{X_{h_\ell} > \xi_\star^{h_{\ell-1}}}\Big)^2\Big] &= \mathbb{E}\Big[\mathbf{1}_{\xi_\star^{h_{\ell-1}} < X_{h_\ell} < \xi_\star^{h_\ell}} + \mathbf{1}_{\xi_\star^{h_\ell} < X_{h_\ell} < \xi_\star^{h_{\ell-1}}}\Big] \\
&= \big|F_{X_{h_\ell}}(\xi_\star^{h_\ell}) - F_{X_{h_\ell}}(\xi_\star^{h_{\ell-1}})\big| \\
&\leq \sup_{h\in\mathcal{H}} \|f_{X_h}\|_\infty \big|\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}\big| \\
&\leq K h_\ell.
\end{aligned} \tag{4.35}$$

On the other hand, Proposition 4.1 yields

$$\mathbb{E}\Big[\Big(\mathbf{1}_{X_{h_\ell} > \xi_\star^{h_{\ell-1}}} - \mathbf{1}_{X_{h_{\ell-1}} > \xi_\star^{h_{\ell-1}}}\Big)^2\Big] \leq K_{h_\ell}\epsilon(h_\ell), \tag{4.36}$$

where K_{h_ℓ} is a constant satisfying $\sup_{\ell\geq 1} K_{h_\ell} < \infty$, and $\epsilon(h_\ell)$ is defined in (4.11). Combining the two previous estimates with (4.34), (4.33) and using (4.18), we get

$$\begin{aligned}
\mathbb{E}\big[|E_n^\ell|^2\big] &= \mathbb{E}\Big[\Big(\sum_{k=1}^n \gamma_k \Pi_{k+1:n}\big(e_k^{h_\ell} - e_k^{h_{\ell-1}}\big)\Big)^2\Big] \\
&\leq K\big(h_\ell + \epsilon(h_\ell)\big)\sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}|^2 \\
&\leq K\epsilon(h_\ell)\gamma_n.
\end{aligned} \tag{4.37}$$

Observe again that the random variables $(E_n^\ell)_{1\leq\ell\leq L}$ are independent and centered. Hence, the above estimate directly yields

$$\mathbb{E}\Big[\Big(\sum_{\ell=1}^L E_{N_\ell}^\ell\Big)^2\Big] = \sum_{\ell=1}^L \mathbb{E}\big[|E_{N_\ell}^\ell|^2\big] \leq K\sum_{\ell=1}^L \gamma_{N_\ell}\epsilon(h_\ell).$$

Step 1.6.
Gathering the above estimates on each term, we eventually get

$$\mathbb{E}\big[\big(\xi_{\mathbf{N}}^{\mathrm{ML}} - \xi_\star^{h_L}\big)^2\big] \leq K\Big\{\gamma_{N_0} + \Big(\sum_{\ell=1}^L \gamma_{N_\ell} + \gamma_{N_\ell}^{\frac{1}{2}} h_\ell^{(\frac{1}{4}+\delta)\wedge 1}\Big)^2 + \sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{3}{2}} + \gamma_{N_\ell}\epsilon(h_\ell)\Big\}. \tag{4.38}$$

From the Cauchy-Schwarz inequality, one gets

$$\sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{1}{2}} h_\ell^{(\frac{1}{4}+\delta)\wedge 1} \leq \Big(\sum_{\ell=1}^L \gamma_{N_\ell} h_\ell^{\frac{1}{2}}\Big)^{\frac{1}{2}}\Big(\sum_{\ell=1}^L h_\ell^{2(\delta\wedge\frac{3}{4})}\Big)^{\frac{1}{2}} \leq K\Big(\sum_{\ell=1}^L \gamma_{N_\ell}\epsilon(h_\ell)\Big)^{\frac{1}{2}} \tag{4.39}$$

since $h_\ell^{\frac{1}{2}} \leq K\epsilon(h_\ell)$ and

$$\sup_{L\geq 1}\sum_{\ell=1}^{L} h_\ell^{2(\delta\wedge\frac{3}{4})} = \sup_{L\geq 1} h^{2(\delta\wedge\frac{3}{4})}\frac{1-M^{-2(\delta\wedge\frac{3}{4})L}}{M^{2(\delta\wedge\frac{3}{4})}-1} \leq \frac{h^{2(\delta\wedge\frac{3}{4})}}{M^{2(\delta\wedge\frac{3}{4})}-1} < \infty.$$

Plugging (4.39) into (4.38) concludes the proof of (4.9).

Step 2: Proof of (4.10)**.**

Step 2.1.
We first provide an $L^2(\mathbb{P})$-estimate for the quantity $(\xi_n^{h_\ell} - \xi_n^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}))$, $n \geq 1$. Recalling the decomposition (4.20) and combining (4.27), (4.29), (4.30), (4.32) and (4.37), we obtain

$$\mathbb{E}\big[\big(\xi_n^{h_\ell} - \xi_n^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}})\big)^2\big] \leq K\big(\gamma_n^{\frac{3}{2}} + \gamma_n\epsilon(h_\ell)\big), \quad n \geq 1. \tag{4.40}$$

Step 2.2.
Here again, for sake of simplicity, we assume that $C_0^{h_\ell} = 0$ for $\ell = 0, \cdots, L$. The general case follows from similar arguments. From the second part of Theorem 3.1, we also get

$$\mathbb{E}\big[\big(C_{N_0}^{h_0} - C_\star^{h_0}\big)^2\big] \leq \frac{K}{N_0^{1\wedge 2\beta}}. \tag{4.41}$$

From the decomposition (3.55) we also readily derive the identity

$$\begin{aligned} &C_n^{h_\ell} - C_\star^{h_\ell} - (C_n^{h_{\ell-1}} - C_\star^{h_{\ell-1}}) \\ &\quad = \frac{1}{n}\sum_{k=1}^{n}\varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}} + \frac{1}{n}\sum_{k=1}^{n} V_{h_\ell}(\xi_{k-1}^{h_\ell}) - V_{h_\ell}(\xi_\star^{h_\ell}) - \big(V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}}) - V_{h_{\ell-1}}(\xi_\star^{h_{\ell-1}})\big). \end{aligned} \tag{4.42}$$

We recall that for each $\ell = 0, \ldots, L$, $(\varepsilon_k^{h_\ell})_{1\leq k\leq n}$ is a sequence of $(\mathcal{F}^{h_\ell}, \mathbb{P})$-martingale increments satisfying $\mathbb{E}\big[\varepsilon_k^{h_\ell}\big|\mathcal{F}_{k-1}^{h_\ell}\big] = 0$, and

$$\begin{aligned} \mathbb{E}\big[\big(\varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}}\big)^2\big|\mathcal{F}_{k-1}\big] &\leq \frac{1}{(1-\alpha)^2}\mathbb{E}\big[\big((X_{h_\ell}^{(k)} - \xi_{k-1}^{h_\ell})^+ - (X_{h_{\ell-1}}^{(k)} - \xi_{k-1}^{h_{\ell-1}})^+\big)^2\big|\mathcal{F}_{k-1}\big] \\ &\leq \frac{3}{(1-\alpha)^2}\big(\mathbb{E}[(X_{h_\ell} - X_{h_{\ell-1}})^2] + \big(\xi_{k-1}^{h_\ell} - \xi_{k-1}^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}})\big)^2 + \big(\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}\big)^2\big) \\ &\leq K\big(h_\ell + \big(\xi_{k-1}^{h_\ell} - \xi_{k-1}^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}})\big)^2\big), \end{aligned}$$

where we used (3.8) and (E.2) with $h' = h_{\ell-1}$, $h = h_\ell$ and $p = 2$, for the last inequality. We now take expectation in both sides of the previous inequality and then use (4.40). Hence,

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k^{h_\ell}-\varepsilon_k^{h_{\ell-1}}\Big)^2\Big]\\
&\quad=\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\big[\big(\varepsilon_k^{h_\ell}-\varepsilon_k^{h_{\ell-1}}\big)^2\big]\\
&\quad\leq\frac{h_\ell+\sup_{\ell\geq 1}\mathbb{E}\big[\big(\xi_0^{h_\ell}-\xi_0^{h_{\ell-1}}-(\xi_\star^{h_\ell}-\xi_\star^{h_{\ell-1}})\big)^2\big]}{n^2}+\frac{1}{n^2}\sum_{k=2}^{n}\mathbb{E}\big[\big(\varepsilon_k^{h_\ell}-\varepsilon_k^{h_{\ell-1}}\big)^2\big]\\
&\quad\leq K\Big(\frac{1}{n^2}+\frac{h_\ell}{n}+\frac{1}{n^2}\sum_{k=1}^{n}\big(\gamma_k^{\frac{3}{2}}+\gamma_k\epsilon(h_\ell)\big)\Big)\\
&\quad\leq K\Big(\frac{h_\ell}{n}+\frac{1}{n^2}\sum_{k=1}^{n}\big(\gamma_k^{\frac{3}{2}}+\gamma_k\epsilon(h_\ell)\big)\Big).
\end{aligned}
\tag{4.43}
$$

Hence, since the random variables $\Big\{\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\varepsilon_k^{h_\ell}-\varepsilon_k^{h_{\ell-1}},\ell=1,\ldots,L\Big\}$ are independent and centered, we deduce from the previous inequality

$$
\mathbb{E}\Big[\Big(\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\varepsilon_k^{h_\ell}-\varepsilon_k^{h_{\ell-1}}\Big)^2\Big]\leq K\sum_{\ell=1}^{L}\Big(\frac{h_\ell}{N_\ell}+\frac{1}{N_\ell^2}\sum_{k=1}^{N_\ell}\big(\gamma_k^{\frac{3}{2}}+\gamma_k\epsilon(h_\ell)\big)\Big).\tag{4.44}
$$

In order to deal with the second term in the right hand side of (4.42), we use Minkowski's inequality and a second order Taylor expansion for both terms together with the fact that $\sup_{h\in\mathcal{H}}\|V_h''\|_\infty<\infty$ and then (3.17). We obtain

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{k=1}^{n}V_{h_\ell}(\xi_{k-1}^{h_\ell})-V_{h_\ell}(\xi_\star^{h_\ell})-\big(V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}})-V_{h_{\ell-1}}(\xi_\star^{h_{\ell-1}})\big)\Big)^2\Big]^{\frac{1}{2}}\\
&\qquad\leq K\Big(\frac{1}{n}+\frac{1}{n-1}\sum_{k=2}^{n}\Big(\mathbb{E}\big[\big(\xi_{k-1}^{h_\ell}-\xi_\star^{h_\ell}\big)^4\big]^{\frac{1}{2}}+\mathbb{E}\big[\big(\xi_{k-1}^{h_{\ell-1}}-\xi_\star^{h_{\ell-1}}\big)^4\big]^{\frac{1}{2}}\Big)\Big)\\
&\qquad\leq K\Big(\frac{1}{n}+\frac{1}{n-1}\sum_{k=1}^{n-1}\gamma_k\Big)\\
&\qquad\leq K\bar{\gamma}_n,
\end{aligned}
\tag{4.45}
$$

where we used the fact that $\sum_{k\geq 1}\gamma_k=\infty$ for the last inequality. We now come back to (4.42), use (4.43), (4.45). We thus obtain

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{\ell=1}^{L}C_{N_\ell}^{h_\ell}-C_\star^{h_\ell}-\big(C_{N_\ell}^{h_{\ell-1}}-C_\star^{h_{\ell-1}}\big)\Big)^2\Big]\\
&\quad\leq 2\,\mathbb{E}\Big[\Big(\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}\varepsilon_k^{h_\ell}-\varepsilon_k^{h_{\ell-1}}\Big)^2\\
&\qquad\quad+\Big(\sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell}V_{h_\ell}(\xi_{k-1}^{h_\ell})-V_{h_\ell}(\xi_\star^{h_\ell})-\big(V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}})-V_{h_{\ell-1}}(\xi_\star^{h_{\ell-1}})\big)\Big)^2\Big]\\
&\quad\leq K\Big(\sum_{\ell=1}^{L}\frac{h_\ell}{N_\ell}+\frac{1}{N_\ell^2}\sum_{k=1}^{N_\ell}\gamma_k^{\frac{3}{2}}+\gamma_k\epsilon(h_\ell)+\Big(\sum_{\ell=1}^{L}\bar{\gamma}_{N_\ell}\Big)^2\Big).
\end{aligned}
\tag{4.46}
$$

Putting together (4.41) and (4.46) concludes the proof of (4.10). $\square$

4.2 Complexity Analysis

For a fixed prescribed tolerance level $\varepsilon \in (0,1)$, a given $h \in \mathcal{H}$ and an integer $M \geq 2$, we rely on the above result to choose the number of levels L and the number of steps $N_0, \dots, N_L$ in each level in order to achieve a root mean square error for the computation of the VaR and ES of order ε using Algorithm 2.

In view of Proposition 3.1, the bias of our couple of estimators $(\xi^{\mathrm{ML}}_{\mathbf{N}}, \mathrm{C}^{\mathrm{ML}}_{\mathbf{N}})$ is of order h_L. We thus select $L \geq 1$ such that

$$L = \left\lceil \frac{\ln h\varepsilon^{-1}}{\ln M} \right\rceil \quad \Rightarrow \quad h_L = \frac{h}{M^L} \leq \varepsilon. \tag{4.47}$$

The computational cost of Algorithm 2 is given by

$$\mathrm{Cost}_{\mathrm{ML}} = C \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell}.$$

VaR Focused Parametrization. We first focus on estimating the VaR. Similarly to the multilevel Monte Carlo algorithm for the computation of the probability $\mathbb{P}(X_0 > 0)$ (see e.g. [16]), it is expected that the leading term in the global L^2 error (4.9) is the last one, namely, $\sum_{\ell=1}^{L} \gamma_{N_\ell} \epsilon(h_\ell)$. Following this heuristic, in order to obtain the optimal values for $N_0, \dots, N_L$, we minimize the complexity under a mean square error constraint, namely, we solve the following constrained optimization problem

$$\begin{aligned} &\mathrm{minimize}_{N_0,\dots,N_L>0} && \textstyle\sum_{\ell=0}^{L} N_\ell h_\ell^{-1}, \\ &\text{subject to} && \textstyle\sum_{\ell=0}^{L} \gamma_{N_\ell} \epsilon(h_\ell) = K^{-1}\varepsilon^2, \end{aligned}$$

where K is the constant appearing in (4.9). Under the assumption that $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0,1]$, the optimal values for N_ℓ, $\ell = 0, \dots, L$, are given by

$$N_\ell = \left\lceil (K\gamma_1)^{\frac{1}{\beta}} \varepsilon^{-\frac{2}{\beta}} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{-\frac{\beta}{1+\beta}} \epsilon(h_{\ell'})^{\frac{1}{1+\beta}} \right)^{\frac{1}{\beta}} h_\ell^{\frac{1}{1+\beta}} \epsilon(h_\ell)^{\frac{1}{1+\beta}} \right\rceil,$$

or $N_\ell =$

$$\begin{cases} \left\lceil (K\gamma_1)^{\frac{1}{\beta}} \varepsilon^{-\frac{2}{\beta}} h_\ell^{\frac{1}{1+\beta}(1+\frac{p_*}{2(1+p_*)})} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{\frac{1}{1+\beta}(-\beta+\frac{p_*}{2(1+p_*)})} \right)^{\frac{1}{\beta}} \right\rceil & \text{if (4.6) is satisfied,} \\ \left\lceil (K\gamma_1)^{\frac{1}{\beta}} \varepsilon^{-\frac{2}{\beta}} h_\ell^{\frac{3}{2(1+\beta)}} |\ln h_\ell|^{\frac{1}{2(1+\beta)}} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{\frac{1-2\beta}{2(1+\beta)}} |\ln h_{\ell'}|^{\frac{1}{2(1+\beta)}} \right)^{\frac{1}{\beta}} \right\rceil & \text{if (4.7) is satisfied,} \\ \left\lceil (K\gamma_1)^{\frac{1}{\beta}} \varepsilon^{-\frac{2}{\beta}} h_\ell^{\frac{3}{2(1+\beta)}} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{\frac{1-2\beta}{2(1+\beta)}} \right)^{\frac{1}{\beta}} \right\rceil & \text{if (4.8) is satisfied.} \end{cases}$$

We can easily check that, with the above choice of L and $N_0, \dots, N_\ell$, the first terms in (4.9), namely $\gamma_{N_0} + \big(\sum_{\ell=1}^{L} \gamma_{N_\ell}\big)^2 + \sum_{\ell=1}^{L} \gamma_{N_\ell}^{\frac{3}{2}}$, are $O(\varepsilon^2)$ as $\varepsilon \downarrow 0$.

Standard computations that we omit show that the optimal computational cost of Algorithm 2 is achieved when $\beta = 1$, in which case the complexity is given by

$$\mathrm{Cost}_{\mathrm{ML}} \leq C \begin{cases} \varepsilon^{-3+\frac{p_*}{2(1+p_*)}} & \text{if (4.6) is satisfied,} \\ \varepsilon^{-\frac{5}{2}} |\ln \varepsilon|^{\frac{1}{2}} & \text{if (4.7) is satisfied,} \\ \varepsilon^{-\frac{5}{2}} & \text{if (4.8) is satisfied.} \end{cases}$$

Therefore, in the best case where (4.8) is satisfied, for a given target quadratic error ε^2, it is possible to estimate the VaR with a complexity of $O(\varepsilon^{-\frac{5}{2}})$ using the multilevel stochastic approximation approach, which is lower than the optimal complexity of $O(\varepsilon^{-3})$ for the nested stochastic approximation approach.

ES Focused Parametrization. A possible approach for parametrizing the amounts of iterations $N_0, \dots, N_L$ optimally to compute the ES would be to choose these quantities to minimize the complexity of Algorithm 1 while constraining the global L^2 error of the multilevel ES estimator (4.10) to an order of ε^2. We presume that the leading term of the upper estimate in (4.10) is $\sum_{\ell=1}^{L} \frac{h_\ell}{N_\ell}$. We will verify this assumption a posteriori. Under this assumption, we solve the optimization problem

$$\begin{array}{ll} \text{minimize}_{N_0,\dots,N_L>0} & \sum_{\ell=0}^{L} N_\ell h_\ell^{-1}, \\ \text{subject to} & \sum_{\ell=0}^{L} h_\ell N_\ell^{-1} = K^{-1}\varepsilon^2. \end{array}$$

The optimal values are given by

$$N_\ell = \lceil K\varepsilon^{-2} L h_\ell \rceil = \left\lceil K\varepsilon^{-2} \left\lceil \frac{\ln h\varepsilon^{-1}}{\ln M} \right\rceil h_\ell \right\rceil.$$

Under this parametrization, we check that as $\varepsilon \downarrow 0$, $\big(\sum_{\ell=0}^{L} \bar{\gamma}_{N_\ell}\big)^2 = O(\varepsilon^{2\beta} \left|\ln \varepsilon\right|^{2(\mathbf{1}_{\beta=1}-\beta)})$. Hence, in order to achieve a mean squared error of order ε^2, one has to choose $\beta = 1$. One can also verify that the remaining terms of the upper estimate of (4.10), namely $\frac{1}{N_0} + \sum_{\ell=1}^{L} \frac{\bar{\gamma}_{N_\ell}}{N_\ell} \epsilon(h_\ell) + \sum_{\ell=1}^{L} \frac{1}{N_\ell^2} \sum_{k=1}^{N_\ell} \gamma_k^{\frac{3}{2}}$, are of order ε^2. With this specific choice of $N_0, \dots, N_L$, one achieves a mean squared error of order ε^2 for the computation of the ES, with a complexity of

$$\mathrm{Cost}_{\mathrm{ML}} \leq C\varepsilon^{-2} \left|\ln \varepsilon\right|^2.$$

Note that it exactly coincides with the optimal complexity of the standard multilevel Monte Carlo algorithm derived in [15].

We now assess the above results numerically. This is done in stylized financial setups, where the probability measure $\mathbb{P}$ in the paper is used for all pricing and risk computations. The code used to produce the results exhibited in the following sections is available here: github.com/azarlouzi/mlsa.

5 Financial Case Study: Option

We endorse the stylized financial setup of [16, Section 3], considering an option with payoff $-W_T^2$ at maturity $T = 1$, where W is a standard Brownian motion. Assuming zero interest rates, the value of the option $v(t, y)$ at time $t \in [0, 1]$ is defined by

$$v(t, y) = \mathbb{E}[-W_1^2 | W_t = y].$$

Let $\tau \in (0, 1)$ be a time horizon. We define the loss X_0 (in our previous notation) on the option as

$$X_0 = v(0, 0) - v(\tau, W_\tau).$$

Let $\varphi \colon \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

$$\varphi(y, z) \coloneqq -\big(\sqrt{\tau} y + \sqrt{1-\tau} z\big)^2 = -\tau y^2 - 2\sqrt{\tau(1-\tau)} yz - (1-\tau) z^2, \quad y, z \in \mathbb{R}.$$

Then

$$X_0 \overset{\mathcal{L}}{=} \mathbb{E}[\varphi(Y,Z)] - \mathbb{E}[\varphi(Y,Z)|Y] = -1 - \mathbb{E}[\varphi(Y,Z)|Y] = \tau(Y^2-1), \tag{5.1}$$

where Y and Z are independent and follow a normal distribution $\mathcal{N}(0,1)$.

5.1 Some Analytical Formulas

The VaR $\xi_\star$ at level $\alpha \in (0,1)$ of the loss X_0 can be obtained analytically. It satisfies

$$1-\alpha = \mathbb{P}(X_0 > \xi_\star) = \mathbb{P}\Big(Y^2 > 1 + \frac{\xi_\star}{\tau}\Big) = 2F\Big(-\Big(1+\frac{\xi_\star}{\tau}\Big)^{\frac{1}{2}}\Big),$$

where F is the standard Gaussian cdf. Hence

$$\xi_\star = \tau\Big\{\Big(F^{-1}\Big(\frac{1-\alpha}{2}\Big)\Big)^2 - 1\Big\}. \tag{5.2}$$

We can also get an analytical formula for the ES $C_\star$ at level α. Indeed, using the symmetry of the Gaussian distribution,

$$C_\star = \mathbb{E}[X_0|X_0 > \xi_\star] = \frac{\tau}{1-\alpha}\big\{2\,\mathbb{E}\big[Y^2\mathbf{1}_{Y>\mu}\big] - (1-\alpha)\big\}, \quad \text{where} \quad \mu := \Big(1+\frac{\xi_\star}{\tau}\Big)^{\frac{1}{2}}.$$

Integrating by parts, we obtain

$$\mathbb{E}\big[Y^2\mathbf{1}_{Y>\mu}\big] = \mu f(\mu) + F(-\mu),$$

where f denotes the standard Gaussian pdf. Hence,

$$C_\star = \frac{2\tau}{1-\alpha}\Big(\mu f(\mu) + F(-\mu) - \frac{1-\alpha}{2}\Big). \tag{5.3}$$

5.2 Stochastic Approximation

The loss $X_0 \overset{\mathcal{L}}{=} \tau(Y^2-1)$ (cf (5.1)) can be simulated directly. But the goal here is to estimate the values of $\xi_\star$ and $C_\star$ using Algorithms 1 and 2, and to study empirically the influence of the parametrization of each algorithm in their performances. The exact values (5.2) and (5.3) are used to evaluate the convergence of the algorithm.

The loss $X_0 \overset{\mathcal{L}}{=} -1 - \mathbb{E}[\varphi(Y,Z)|Y]$ can be approximated, for a given bias parameter $h = \frac{1}{K} \in \mathcal{H}$, by

$$X_h = -1 - \frac{1}{K}\sum_{k=1}^{K}\varphi(Y, Z^{(k)}),$$

where $Y, Z^{(1)}, \dots, Z^{(K)} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. We can then apply Algorithm 1 or 2 on this basis.

5.3 Numerical Results

In the following applications, the confidence level is taken equal to $\alpha = 97.5\%$ and the time horizon is set to $\tau = 0.5$. This setup yields $\xi_\star \approx 2.012$ and $C_\star \approx 2.901$. The value of $\beta = 1$ is used for the step sequence $(\gamma_n)_{n\geq 1}$ since, according to Sections 3.2 and 4.2, it leads to the optimal complexities of Algorithms 1 and 2.

Weak Error Expansion. To highlight the results of Proposition 3.1, we propose to study the linearity of the quantities $\xi^h_\star - \xi_\star$ and $C^h_\star - C_\star$ as $\mathcal{H} \ni h \downarrow 0$. We additionally study the degree of stability of the quantities $h^{-1}(\xi^h_\star - \xi_\star)$ and $h^{-1}(C^h_\star - C_\star)$ as $\mathcal{H} \ni h \downarrow 0$. To do so, we run Algorithm 1 for multiple values of $h \in \{\frac{1}{10}, \frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{200}\}$ and for a very large number of iterations $N = 10^6$. We use the outcome of each run $(\xi^h_{10^6}, C^h_{10^6})$ as an estimate of $(\xi^h_\star, C^h_\star)$. The value of $(\xi_\star, C_\star)$ is computed using the explicit formulas (5.2) and (5.3). Finally, we adopt the step function $\{\gamma_n = 0.1/(10^4 + n), n \geq 1\}$, where a smoothing is applied in order to avoid instability during the first few iterations. The left panel of Figure 1 shows the values of $\xi^h_{10^6} - \xi_\star$ and $C^h_{10^6} - C_\star$ as h varies in $\{\frac{1}{10}, \frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{200}\}$, while the right panel shows the corresponding values of $h^{-1}(\xi^h_{10^6} - \xi_\star)$ and $h^{-1}(C^h_{10^6} - C_\star)$.

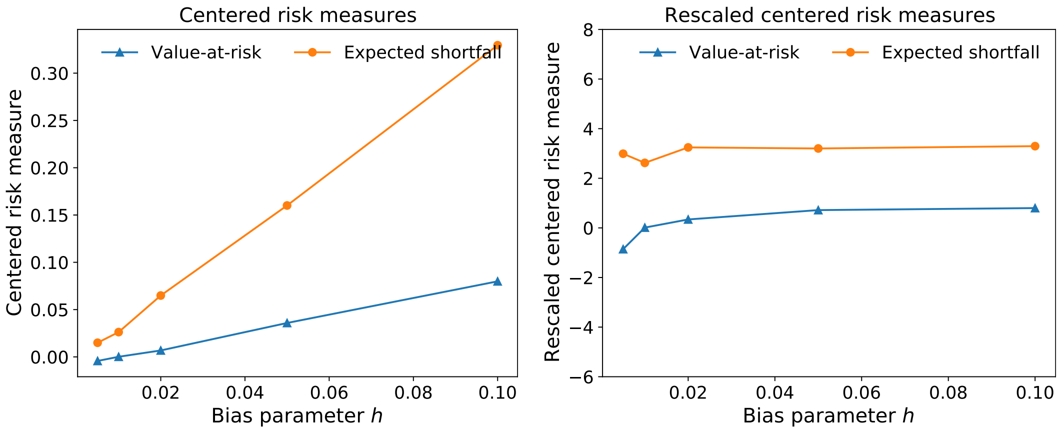

Figure 1: Centered and rescaled risk measures as the bias parameter goes to 0.

The left panel plot suggests that, asymptotically as $\mathcal{H} \ni h \downarrow 0$, the quantities $\xi^h_\star - \xi_\star$ and $C^h_\star - C_\star$ are linear in h. The right panel plot strengthens this observation, showing that the quantities $h^{-1}(\xi^h_\star - \xi_\star)$ and $h^{-1}(C^h_\star - C_\star)$ are approximately constant in a neighborhood of 0. This verifies empirically the validity of Proposition 3.1 on the weak error expansion.

Comparative Complexity Study. We aim here to compare the running times and achieved performances of the scheme (2.7) and Algorithms 1 and 2. The scheme (2.7) and Algorithm 1 are run with the step function $\{\gamma_n = 0.1/(2.5 \times 10^4 + n), n \geq 1\}$, and with their respective theoretical optimal numbers of iterations. Algorithm 2 is run under Assumption (4.6), that is clearly satisfied for any $p_\star > 1$, with the particular value of $p_\star = 11$. We also choose $M = 2$. As for the number of iterations in each level, we use the VaR focused parametrization for the VaR computation, and the ES focused parametrization for the ES computation. Further parametrization of Algorithm 2 is described in Table 1. For each algorithm, we plot the average run time against the achieved RMSE for the VaR and ES over 200 runs. The results are reported in Figure 2.

For the VaR as for the ES, we notice that for any given scored RMSE, Algorithm 2 displays an execution time that is orders of magnitude lower than Algorithm 1. This is corroborated by the theoretical optimal complexity of Algorithm 2 that is significantly lower than that of Algorithm 1. Finally, as one would expect, the scheme (2.7) that simulates X_0 directly considerably outperforms both Algorithms 1 and 2. The VaR and ES curves offer an illustration of the performance gain achieved by shifting from the nested SA scheme to the multilevel SA scheme. Indeed, part of the the performance loss occurring when using a nested simulation instead of a direct simulation of the loss, as illustrated by the performance gap between the SA scheme and the nested SA scheme, is regained when utilizing the multilevel SA scheme.

accuracy ε	VaR estimation			ES estimation		
	h	L	γ_n	h	L	γ_n
$\frac{1}{64}$	$\frac{1}{32}$	1	$\frac{0.1}{10^4+n}$	$\frac{1}{32}$	1	$\frac{0.1}{10^4+n}$
$\frac{1}{128}$	$\frac{1}{32}$	2	$\frac{0.1}{10^4+n}$	$\frac{1}{32}$	2	$\frac{0.1}{10^4+n}$
$\frac{1}{256}$	$\frac{1}{32}$	3	$\frac{0.1}{2\times10^4+n}$	$\frac{1}{32}$	3	$\frac{0.1}{2\times10^4+n}$
$\frac{1}{512}$	$\frac{1}{32}$	4	$\frac{0.1}{2.5\times10^4+n}$	$\frac{1}{32}$	4	$\frac{0.1}{2.5\times10^4+n}$
$\frac{1}{1024}$	$\frac{1}{32}$	5	$\frac{0.1}{5\times10^4+n}$	$\frac{1}{32}$	5	$\frac{0.1}{5\times10^4+n}$
$\frac{1}{2048}$				$\frac{1}{16}$	7	$\frac{0.1}{2\times10^4+n}$

Table 1: Parametrization of Algorithm 2 for different prescribed accuracies.

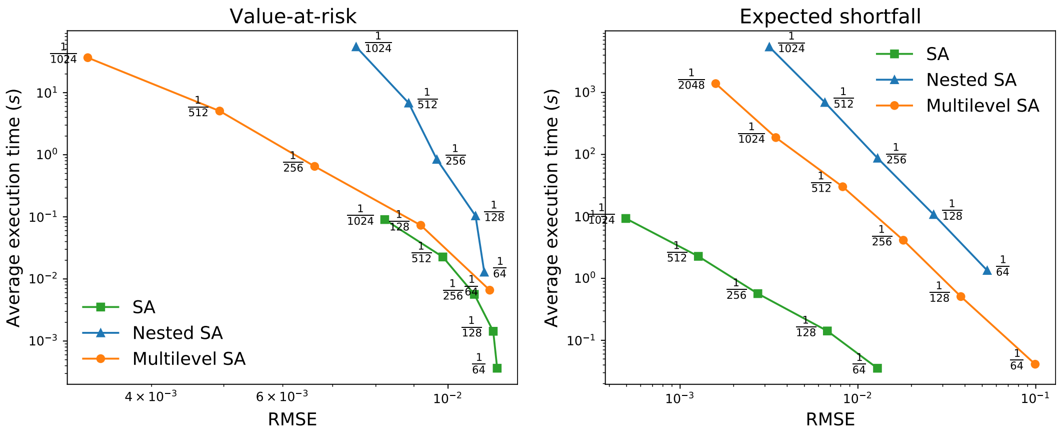

Figure 2: Performance comparison of Scheme (2.7) and Algorithms 1 and 2.

6 Financial Case Study: Swap

We adapt the setup of [2, Section A.1]. Consider a long position on a stylized swap with nominal $\bar{N}$, strike $\bar{S}$ and maturity $\bar{T}$ on some underlying (FX or interest rate) rate. This rate follows a Black-Scholes model $(S_t)_{t\geq0}$ of risk neutral drift $\bar{\kappa}$ and volatility $\bar{\sigma}$, so that $\big\{\widehat{S}_t \coloneqq \mathrm{e}^{-\bar{\kappa}t}S_t, t\geq0\big\}$ follows a Black martingale model of volatility $\bar{\sigma}$. Given times $0=T_0<T_1<\cdots<T_d=\bar{T}$, the swap pays at the positive coupon dates $T_i>0$ the cash flows $\Delta_i(S_{T_{i-1}}-\bar{S})$, where $\Delta_i=T_i-T_{i-1}$. For $t\in[0,\bar{T}]$, let $\rho_t=e^{-rt}$, for some constant risk-free rate r, and let $i_t\in\{1,\dots,d+1\}$ denote the integer that satisfies $T_{i_t-1}\leq t<T_{i_t}$. The value P_t of the swap at time

$t \in [0, \bar{T}]$ is thus expressed as

$$P_t = \bar{N}\,\mathbb{E}\bigg[\rho_t^{-1}\rho_{T_{i_t}}\Delta_{i_t}(S_{T_{i_t-1}} - \bar{S}) + \sum_{i=i_t+1}^{d} \rho_t^{-1}\rho_{T_i}\Delta_i(S_{T_{i-1}} - \bar{S})\bigg| S_t\bigg]$$

$$= \bar{N}\,\mathbb{E}\bigg[\rho_t^{-1}\rho_{T_{i_t}}\Delta_{i_t}\big(\mathrm{e}^{\bar{\kappa}T_{i_t-1}}\widehat{S}_{T_{i_t-1}} - \bar{S}\big) + \sum_{i=i_t+1}^{d} \rho_t^{-1}\rho_{T_i}\Delta_i\big(\mathrm{e}^{\bar{\kappa}T_{i-1}}\widehat{S}_{T_{i-1}} - \bar{S}\big)\bigg| \widehat{S}_t\bigg] \tag{6.1}$$

$$= \bar{N}\bigg(\rho_t^{-1}\rho_{T_{i_t}}\Delta_{i_t}\big(\mathrm{e}^{\bar{\kappa}T_{i_t-1}}\widehat{S}_{T_{i_t-1}} - \bar{S}\big) + \sum_{i=i_t+1}^{d} \rho_t^{-1}\rho_{T_i}\Delta_i\big(\mathrm{e}^{\bar{\kappa}T_{i-1}}\widehat{S}_t - \bar{S}\big)\bigg). \tag{6.2}$$

We assume the swap issued at par at time 0, i.e. $P_0 = 0$, hence

$$\bar{S} = \frac{\sum_{i=1}^{d} \rho_{T_i}\Delta_i \mathrm{e}^{\bar{\kappa}T_{i-1}}}{\sum_{i=1}^{d} \rho_{T_i}\Delta_i} S_0. \tag{6.3}$$

Given some confidence level $\alpha \in (0, 1)$, we are interested in the risk measures $\xi_\star := \mathbb{V}\mathrm{a}\mathbb{R}_\alpha(\rho_\delta P_\delta)$ and $C_\star := \mathbb{E}\mathbb{S}_\alpha(\rho_\delta P_\delta)$ of the position at some time horizon $\delta < \Delta$.

6.1 Some Analytical Formulas

In this lognormal setup, the values of $\xi_\star$ and $C_\star$ can be obtained analytically. On the one hand, observe that $i_\delta = 1$ and, using (6.2) and (6.3),

$$\rho_\delta P_\delta = \bar{N}A\big(\widehat{S}_\delta - S_0\big), \qquad \text{where} \qquad A := \sum_{i=2}^{d} \rho_{T_i}\Delta_i \mathrm{e}^{\bar{\kappa}T_{i-1}}. \tag{6.4}$$

Hence

$$1 - \alpha = \mathbb{P}(\rho_\delta P_\delta > \xi_\star) = \mathbb{P}\Big(\widehat{S}_\delta > S_0 + \frac{\xi_\star}{\bar{N}A}\Big),$$

thus

$$\xi_\star = \bar{N}AS_0\Big(\exp\Big(F^{-1}(\alpha)\bar{\sigma}\sqrt{\delta} - \frac{\bar{\sigma}^2}{2}\delta\Big) - 1\Big), \tag{6.5}$$

where F denotes the standard Gaussian cdf. On the other hand, set

$$\omega := S_0 + \frac{\xi_\star}{\bar{N}A}, \qquad \eta_\pm := \frac{1}{\bar{\sigma}\sqrt{\delta}}\Big(\ln\frac{\omega}{S_0} \pm \frac{\bar{\sigma}^2\delta}{2}\Big).$$

We can then write

$$C_\star = \mathbb{E}[\rho_\delta P_\delta | \rho_\delta P_\delta > \xi_\star] = \frac{\bar{N}A}{1-\alpha}\big(\mathbb{E}\big[\widehat{S}_\delta \mathbf{1}_{\widehat{S}_\delta > \omega}\big] - (1-\alpha)S_0\big).$$

Note that

$$\mathbb{E}\big[\widehat{S}_\delta \mathbf{1}_{\widehat{S}_\delta > \omega}\big] = S_0\mathbb{E}\Big[\exp\Big(\bar{\sigma}\sqrt{\delta}U - \frac{\bar{\sigma}^2}{2}\delta\Big)\mathbf{1}_{U > \eta_+}\Big] = S_0(1 - F(\eta_-)),$$

where $U \sim \mathcal{N}(0, 1)$. Therefore,

$$C_\star = \bar{N}AS_0\frac{\alpha - F(\eta_-)}{1-\alpha}. \tag{6.6}$$

6.2 Stochastic Approximation

Define now $X_0 \coloneqq \rho_\delta P_\delta$. The scheme (2.7) can be applied to approximate the values of $\xi_\star$ and $C_\star$. Indeed, according to (6.4),

$$X_0 \overset{\mathcal{L}}{=} \bar{N} A S_0 \Big(\exp\Big(-\frac{\bar{\sigma}^2}{2}\delta + \bar{\sigma}\sqrt{\delta} U \Big) - 1 \Big), \tag{6.7}$$

where U is a standard Gaussian variable. Going back to (6.1) and using (6.3) yields

$$X_0 = \bar{N}\, \mathbb{E}\Bigg[\sum_{i=2}^{d} \rho_{T_i} \Delta_i \mathrm{e}^{\bar{\kappa} T_{i-1}} \big(\widehat{S}_{T_{i-1}} - S_0\big) \Bigg| \widehat{S}_\delta \Bigg] \overset{\mathcal{L}}{=} \mathbb{E}[\varphi(Y, Z)|Y], \tag{6.8}$$

where Y is a real valued random variable, $Z = (Z_1, \dots, Z_{d-1})$ is an $\mathbb{R}^{d-1}$-valued random variable given by

$$\begin{aligned}
Y &\coloneqq \exp\Big(-\frac{\bar{\sigma}^2}{2}\delta + \bar{\sigma}\sqrt{\delta} U_0 \Big) \sim \frac{\widehat{S}_\delta}{S_0}, \\
Z_1 &\coloneqq \exp\Big(-\frac{\bar{\sigma}^2}{2}(T_1 - \delta) + \bar{\sigma}\sqrt{T_1 - \delta}\, U_1 \Big) \sim \frac{\widehat{S}_{T_1}}{\widehat{S}_\delta}, \\
Z_i &\coloneqq \exp\Big(-\frac{\bar{\sigma}^2}{2}\Delta_i + \bar{\sigma}\sqrt{\Delta_i}\, U_i \Big) \sim \frac{\widehat{S}_{T_i}}{\widehat{S}_{T_{i-1}}}, \quad 2 \le i \le d-1, \\
\varphi(y, z) &\coloneqq \bar{N} S_0 \sum_{i=2}^{d} \rho_{T_i} \Delta_i \mathrm{e}^{\bar{\kappa} T_{i-1}} \Big(y \prod_{j=1}^{i-1} z_j - 1 \Big), \quad y \in \mathbb{R}, z = (z_1, \dots, z_{d-1}) \in \mathbb{R}^{d-1},
\end{aligned}$$

where $(U_i)_{0 \le i \le d-1}$ are iid random variables with law $\mathcal{N}(0, 1)$. Note that Y and Z are independent. Algorithms 1 and 2 are hence applicable to approximate the values of $\xi_\star$ and $C_\star$.

6.3 Numerical Results

We set $r = 2\%$, $S_0 = 1\%$, $\bar{\kappa} = 12\%$, $\bar{\sigma} = 20\%$, $\Delta_i = 3\,\text{months}$, $\bar{T} = 1\,\text{year}$, $\delta = 1\,\text{week} = 7\,\text{days}$ and $\alpha = 85\%$. We use the 30/360 day count convention, so that $1\,\text{month} = 30\,\text{days}$ and $1\,\text{year} = 360\,\text{days}$. Finally, we set the nominal $\bar{N}$ such that the value of each leg of the swap is equal to 1 at time 0, that is,

$$\bar{N} = \frac{1}{S_0 \sum_{i=1}^{d} \rho_{T_i} \Delta_i \mathrm{e}^{\bar{\kappa} T_{i-1}}}.$$

Given the above set of parameters, the theoretical values of the VaR and ES obtained with (6.5) and (6.6) are $\xi_\star \approx 219.64$ and $C_\star \approx 333.91$.

Comparative Complexity Study. For the stochastic approximation of these quantities, we use the step function $\{\gamma_n = 100/n, n \ge 1\}$ for the SA scheme, and $\{\gamma_n = 50/n, n \ge 1\}$ for the nested SA scheme. We also run these schemes with their optimal numbers of iterations. To apply the multilevel SA scheme, we work under Assumption (4.6), with $p_\star = 8$. The VaR and ES are simulated using their respective optimal amounts of iterations $N_0, \dots, N_L$. For each prescribed accuracy $\varepsilon \in \{\frac{1}{8}, \frac{1}{32}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$, we adapt the hyperparameter setup by performing a grid search on the bias parameter h, the number of levels L and the step sequence

$(\gamma_n)_{n\geq 1}$. Table 2 lists these parametrizations by accuracy. Figure 3 plots the joint evolution of the realized RMSEs and the average execution times over 200 runs of each SA scheme, as the accuracy ε varies in $\{\frac{1}{8}, \frac{1}{32}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$.

We observe in Figure 3 that the multilevel SA scheme for the VaR and ES significantly outperforms the nested SA scheme in terms of computational time and achieved error rate. Indeed, for a given target RMSE of order 10, the MLSA scheme for the VaR has a computational time of order 10^{-2} seconds, while the nested SA scheme has one of order 10 seconds. Similarly, for the ES, an RMSE of 10 is achieved by the nested SA scheme in 10^2 seconds, while it is attained by the multilevel SA scheme in only 10 seconds. These results are very promising, as they apply to a more realistic scenario than the previous case study. Eventually, as expected, the standard Robbins-Monro algorithm, that is only applicable if the loss X_0 is directly simulatable, outperforms both Algorithms 1 and 2. Again, as we have pointed out in the previous case study, the curves show how one can regain some of the performance loss due to nested simulation of the loss X_0 by implementing a multilevel SA scheme.

accuracy ε	VaR estimation			ES estimation		
	h	L	γ_n	h	L	γ_n
$\frac{1}{8}$	$\frac{1}{4}$	1	$\frac{10}{n}$	$\frac{1}{4}$	1	$\frac{10}{n}$
$\frac{1}{32}$	$\frac{1}{8}$	2	$\frac{5}{10+n}$	$\frac{1}{8}$	2	$\frac{5}{10+n}$
$\frac{1}{128}$	$\frac{1}{16}$	3	$\frac{20}{500+n}$	$\frac{1}{16}$	3	$\frac{20}{500+n}$
$\frac{1}{256}$	$\frac{1}{16}$	4	$\frac{20}{10^3+n}$	$\frac{1}{16}$	4	$\frac{25}{500+n}$
$\frac{1}{512}$	$\frac{1}{32}$	4	$\frac{25}{750+n}$	$\frac{1}{32}$	4	$\frac{50}{2\times 10^3+n}$

Table 2: Parametrization of Algorithm 2 for different prescribed accuracies.

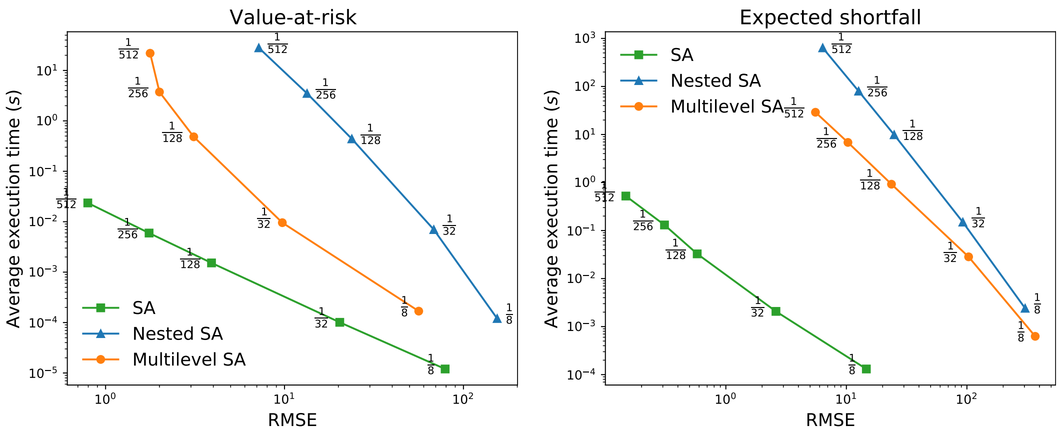

Figure 3: Performance comparison of Scheme (2.7) and Algorithms 1 and 2.

7 Conclusion

In this work, we presented a stochastic approximation problem which solution retrieves the Var and ES of the loss on a financial derivatives portfolio. The loss of the portfolio, however, can only be simulated in a nested Monte Carlo fashion. A naively nested algorithm performs one inner simulation layer within each update of a stochastic gradient descent scheme. A single (bias) parameter allows controlling both the bias and the complexity of the nested algorithm. But, considering a prescribed error ε, this algorithm can only optimally achieve a complexity of order ε^{-3}. Our MLSA algorithm combines multiple estimates obtained by the nested approach for different bias parameters, while reducing the overall complexity of the total entailed simulations.

According to the intent of the user of this algorithm, whether it is to simulate the VaR or the ES, our complexity study shows that there are two corresponding sets of optimal parameters. The VaR focused parameter setup results in a complexity of order $\varepsilon^{-2-\delta}$, where $\delta < 1$ is some specific parameter depending on the integrability degree of the loss while the ES focused parameterization yields a complexity of order $\varepsilon^{-2} |\ln \varepsilon|^2$. The numerical studies highlight the overperformance of the multilevel algorithm, with a performance gap that widens exponentially as the prescribed accuracy ε goes to 0. The study of the joint evolution of the RMSE and the average execution time demonstrates how, in situations where the inner simulations can be shortcutted by explicit formulas, utilizing the multilevel scheme results in a partial, nonetheless significant, performance regain with respect to some classical (non-nested) Robbins-Monro scheme. This finding is of great interest especially when the latter scheme does not apply.

A first question that arises from our study is whether, in the case of the VaR estimation, one could simply close the performance gap between the multilevel scheme and the classical Robbins-Monro scheme. Besides, in order to build a stronger estimator, one could investigate other aggregation methods than the multilevel scheme. A couple of possible ways that may deserve further exploration are the multistep Richardson-Romberg extrapolation method [13] and the adaptive selection the number of inner samples on each multilevel sample as proposed in [16]. Finally, one could look at how to complement our nonasymptotic error controls with asymptotic error distribution results such as central limit theorems.

A Proof of Lemma 3.1

Step 1: Cluster points of $(\xi^h_\star)_{h\in\mathcal{H}}$**.**
Let $h \in \mathcal{H} \cup \{0\}$. By Lemma 2.1, the set Θ_h is a non empty bounded interval that coincides with the set of roots of V'_h.

Since $(X_h)_{h\in\mathcal{H}}$ converges in distribution towards X_0 as $\mathcal{H} \ni h \downarrow 0$, the sequence of cdf functions $(F_{X_h})_{h\in\mathcal{H}}$ converges pointwise on $\mathbb{R}$ towards F_{X_0} as $\mathcal{H} \ni h \downarrow 0$. The cdf F_{X_0} being continuous, the second Dini theorem implies that $(F_{X_h})_{h\in\mathcal{H}}$ converges uniformly on $\mathbb{R}$ towards F_{X_0} as $\mathcal{H} \ni h \downarrow 0$. According to the definition of the functions V'_h, $h \in \mathcal{H} \cup \{0\}$, it follows that $(V'_h)_{h\in\mathcal{H}}$ converges uniformly on $\mathbb{R}$ towards V'_0. For any $\xi^h_\star \in \Theta_h$, $h \in \mathcal{H}$, one has

$$|V'_0(\xi^h_\star)| = |V'_0(\xi^h_\star) - V'_h(\xi^h_\star)| \leq \sup_{\xi\in\mathbb{R}} |V'_0(\xi) - V'_h(\xi)|. \tag{A.1}$$

Since $\lim_{h\downarrow 0} \sup_{\xi\in\mathbb{R}} |V'_0(\xi) - V'_h(\xi)| = 0$, we eventually get

$$\lim_{h\downarrow 0} V'_0(\xi^h_*) = 0. \tag{A.2}$$

We thus deduce from (2.6) that the sequence $(\xi^h_\star)_{h\in\mathcal{H}}$ is bounded. Using (A.2) and the continuity of V'_0, any *valeur d'adhérence* $\xi^0_\star$ of the sequence $(\xi^h_\star)_{h\in\mathcal{H}}$ satisfies

$V_0'(\xi_\star^0) = 0$ so that $\xi_\star^0 \in \Theta_0$ and

$$\lim_{h\downarrow 0} \operatorname{dist}(\xi_\star^h, \Theta_0) = 0. \tag{A.3}$$

Step 2: Limit of $(C_\star^h)_{h\in\mathcal{H}}$.
We now assume that $(X_h)_{h\in\mathcal{H}}$ converges to X_0 in $L^1(\mathbb{P})$. By Lemma 2.1, for any $h \in \mathcal{H}\cup\{0\}$,

$$|C_\star^h - C_\star| = |\min_\xi V_h(\xi) - \min_\xi V_0(\xi)| \le \max_\xi |(V_h - V_0)(\xi)| \le \frac{1}{1-\alpha}\mathbb{E}[|X_h - X_0|],$$

so that

$$\lim_{h\downarrow 0} C_\star^h = C_\star.$$

B Proof of Proposition 3.1

We first observe that since the density function f_{X_0} of X_0 is positive the set $\Theta_0 = \arg\min V_0 = \{\xi_\star\}$ is a singleton so that it follows from Lemma 3.1 that $\xi_\star^h \to \xi_\star$ as $\mathcal{H} \ni h \downarrow 0$.

Step 1: Asymptotic expansion of $\xi_\star^h$ as $\mathcal{H} \ni h \downarrow 0$.
Now observe that

$$F_{X_h}(\xi_\star^h) - F_{X_h}(\xi_\star) = \alpha - F_{X_h}(\xi_\star) = F_{X_0}(\xi_\star) - F_{X_h}(\xi_\star). \tag{B.1}$$

On the one hand, as $\mathcal{H} \ni h \downarrow 0$, Assumption 1(i) gives

$$F_{X_0}(\xi_\star) - F_{X_h}(\xi_\star) = -v(\xi_\star)h + o(h), \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0. \tag{B.2}$$

On the other hand, a first order Taylor expansion with integral remainder gives

$$F_{X_h}(\xi_\star^h) - F_{X_h}(\xi_\star) = (\xi_\star^h - \xi_\star)\int_0^1 f_{X_h}(t\xi_\star + (1-t)\xi_\star^h)\,\mathrm{d}t. \tag{B.3}$$

Since $\lim_{h\downarrow 0}\xi_\star^h = \xi_\star$ and, by Assumption 1(ii), $(f_{X_h})_{h\in\mathcal{H}}$ converges locally uniformly towards the positive density function f_{X_0}, we deduce that $\int_0^1 f_{X_h}(t\xi_\star + (1-t)\xi_\star^h)\,\mathrm{d}t > 0$ for $h \in \mathcal{H}$ small enough and $\lim_{h\downarrow 0}\int_0^1 f_{X_h}(t\xi_\star + (1-t)\xi_\star^h)\,\mathrm{d}t = f_{X_0}(\xi_\star)$. Combining (B.1), (B.2) and (B.3), we thus deduce

$$h^{-1}(\xi_\star^h - \xi_\star) = -\frac{v(\xi_\star)}{\int_0^1 f_{X_h}(t\xi_\star + (1-t)\xi_\star^h)\,\mathrm{d}t} + o(1) = -\frac{v(\xi_\star)}{f_{X_0}(\xi_\star)} + o(1) \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0.$$

Step 2: Asymptotic expansion of $C_\star^h$ as $\mathcal{H} \ni h \downarrow 0$.
Going back to the definitions of the functions V_0 and V_h in (2.5) and (3.4), one has

$$\begin{aligned}
C_\star^h - C_\star &= V_h(\xi_\star^h) - V_0(\xi_\star)\\
&= \mathbb{E}[X_h | X_h > \xi_\star^h] - \mathbb{E}[X_0 | X_0 > \xi_\star]\\
&= \frac{1}{1-\alpha}\Big(\mathbb{E}[X_h \mathbf{1}_{X_h > \xi_\star^h}] - \mathbb{E}[X_0 \mathbf{1}_{X_0 > \xi_\star}]\Big)\\
&= \frac{1}{1-\alpha}\Big(\int_{\xi_\star^h}^\infty \xi\,\mathrm{d}F_{X_h}(\xi) - \int_{\xi_\star}^\infty \xi\,\mathrm{d}F_{X_0}(\xi)\Big)\\
&= \frac{1}{1-\alpha}\Big(\int_{\xi_\star^h}^{\xi_\star} \xi f_{X_h}(\xi)\,\mathrm{d}\xi + \int_{\xi_\star}^\infty \xi\,\mathrm{d}(F_{X_h} - F_{X_0})(\xi)\Big).
\end{aligned} \tag{B.4}$$

On the one hand, combining a change of variable with (3.8) yields

$$\begin{aligned}\int_{\xi_\star^h}^{\xi_\star} \xi f_{X_h}(\xi)\,\mathrm{d}\xi &= (\xi_\star - \xi_\star^h)\int_0^1 (t\xi_\star^h + (1-t)\xi_\star) f_{X_h}(t\xi_\star^h + (1-t)\xi_\star)\,\mathrm{d}t \\ &= \upsilon(\xi_\star)\xi_\star h + o(h).\end{aligned} \tag{B.5}$$

On the other hand, for $h \in \mathcal{H} \cup \{0\}$, by integrating by parts, we get

$$\begin{aligned}\int_{\xi_\star}^{\infty} \xi\,\mathrm{d}F_{X_h}(\xi) &= -\big[\xi\big(1 - F_{X_h}(\xi)\big)\big]_{\xi_\star}^{\infty} + \int_{\xi_\star}^{\infty}\big(1 - F_{X_h}(\xi)\big)\,\mathrm{d}\xi \\ &= \xi_\star\big(1 - F_{X_h}(\xi_\star)\big) + \int_{\xi_\star}^{\infty}\big(1 - F_{X_h}(\xi)\big)\,\mathrm{d}\xi,\end{aligned} \tag{B.6}$$

so that

$$\int_{\xi_\star}^{\infty} \xi\,\mathrm{d}(F_{X_h} - F_{X_0})(\xi) = \xi_\star\big(F_{X_0}(\xi_\star) - F_{X_h}(\xi_\star)\big) + \int_{\xi_\star}^{\infty}\big(F_{X_0}(\xi) - F_{X_h}(\xi)\big)\,\mathrm{d}\xi. \tag{B.7}$$

We now plug the first order Taylor expansion of $F_{X_h} - F_{X_0}$ in Assumption 1(i) into the right hand side of the above identity. We obtain

$$\int_{\xi_\star}^{\infty} \xi\,\mathrm{d}(F_{X_h} - F_{X_0})(\xi) = -\left(\xi_\star \upsilon(\xi_\star) + \int_{\xi_\star}^{\infty} \upsilon(\xi)\mathrm{d}\xi\right)h + o(h). \tag{B.8}$$

Eventually, combining the results from (B.4), (B.5) and (B.8), we conclude

$$C_\star^h - C_\star = -h\int_{\xi_\star}^{\infty} \frac{\upsilon(\xi)}{1-\alpha}\mathrm{d}\xi + o(h), \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0. \tag{B.9}$$

C Proof of Lemma 3.2

Throughout this proof, we consider $q \geq 1$, $h \in \mathcal{H}$ and $\upsilon \geq 0$. We drop the superscript υ from our notation and write $\mathcal{L}_{h,q}(\xi)$ for $\mathcal{L}_{h,q}^{\upsilon}(\xi)$.

(i) From definition (3.10), the function $\mathcal{L}_{h,q}$ is continuously differentiable and satisfies

$$\mathcal{L}_{h,q}'(\xi) = qV_h'(\xi)\mathcal{L}_{h,q-1}(\xi) + \upsilon V_h'(\xi)\mathcal{L}_{h,q}(\xi) = \mathcal{L}_{h,q-1}(\xi)V_h'(\xi)\big(q + \upsilon\big(V_h(\xi) - V_h(\xi_\star^h)\big)\big). \tag{C.1}$$

If $q \geq 2$, one may again differentiate both sides of the above identity to deduce that $\mathcal{L}_{h,q}$ is twice continuously differentiable with

$$\begin{aligned}\mathcal{L}_{h,q}''(\xi) &= \mathcal{L}_{h,q-2}(\xi)V_h'(\xi)^2\big(q - 1 + \upsilon\big(V_h(\xi) - V_h(\xi_\star^h)\big)\big)\big(q + \upsilon\big(V_h(\xi) - V_h(\xi_\star^h)\big)\big) \\ &\quad + \mathcal{L}_{h,q-1}(\xi)V_h''(\xi)\big(q + \upsilon\big(V_h(\xi) - V_h(\xi_\star^h)\big)\big) + \mathcal{L}_{h,q-1}(\xi)\upsilon V_h'(\xi)^2.\end{aligned} \tag{C.2}$$

If $q = 1$, we obtain

$$\mathcal{L}_{h,1}''(\xi) = \big(V_h''(\xi) + 2\upsilon V_h'(\xi)^2\big)\mathcal{L}_{h,0}(\xi) + \big(\upsilon^2 V_h'(\xi)^2 + \upsilon V_h''(\xi)\big)\mathcal{L}_{h,1}(\xi). \tag{C.3}$$

Therefore, for all $q \geq 1$, $\mathcal{L}_{h,q}$ is twice continuously differentiable.

Observe that the remaining properties of Lemma 3.2 are trivially satisfied for $\xi = \xi_\star^h$. We thus assume that $\xi \neq \xi_\star^h$ for the rest of the proof. We let

$$\varepsilon_h := \frac{V_h''(\xi_\star^h)}{[V_h'']_{\mathrm{Lip}}} = \frac{f_{X_h}(\xi_\star^h)}{[f_{X_h}]_{\mathrm{Lip}}}, \tag{C.4}$$

where we used the fact that

$$V_h''(\xi) = \frac{f_{X_h}(\xi)}{1-\alpha}. \tag{C.5}$$

Finally, we denote by I^h the interval given by

$$I^h := \left[\xi_\star^h - \varepsilon_h, \xi_\star^h + \varepsilon_h\right]. \tag{C.6}$$

(ii) From (C.1) and the relation $\mathcal{L}_{h,q-1}(\xi) = (V_h(\xi) - V_h(\xi_\star^h))^{-1}\mathcal{L}_{h,q}(\xi)$, we have that

$$\mathcal{L}_{h,q}'(\xi) = \mathcal{L}_{h,q}(\xi) V_h'(\xi) \left\{ \frac{q}{V_h(\xi) - V_h(\xi_\star^h)} + \upsilon \right\}, \tag{C.7}$$

so that

$$\mathcal{L}_{h,q}'(\xi) V_h'(\xi) \geq \left\{ \frac{q V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \vee \upsilon V_h'(\xi)^2 \right\} \mathcal{L}_{h,q}(\xi). \tag{C.8}$$

We now establish a lower bound for the term inside the parentheses on the right hand side of the above inequality. Since $V_h'(\xi_\star^h) = 0$, a first order Taylor expansion gives

$$V_h'(\xi) = V_h''(\xi_\star^h)(\xi - \xi_\star^h) + (\xi - \xi_\star^h) \int_0^1 \left(V_h''(\xi_\star^h + t(\xi - \xi_\star^h)) - V_h''(\xi_\star^h) \right) \mathrm{d}t, \tag{C.9}$$

which, combined with the fact that $V_h'' \geq 0$ and the triangle inequality, yields

$$|V_h'(\xi)| \geq V_h''(\xi_\star^h)|\xi - \xi_\star^h| - \frac{1}{2}[V_h'']_{\mathrm{Lip}}(\xi - \xi_\star^h)^2. \tag{C.10}$$

If $\xi \in I^h$, the above inequality directly gives

$$V_h'(\xi)^2 \geq \frac{1}{4} V_h''(\xi_\star^h)^2 (\xi - \xi_\star^h)^2. \tag{C.11}$$

Now, a second order Taylor expansion gives

$$V_h(\xi) - V_h(\xi_\star^h) = \frac{1}{2} V_h''(\xi_\star^h)(\xi - \xi_\star^h)^2 + (\xi - \xi_\star^h)^2 \int_0^1 (1-t)\left(V_h''(\xi_\star^h + t(\xi - \xi_\star^h)) - V_h''(\xi_\star^h) \right) \mathrm{d}t, \tag{C.12}$$

so that

$$V_h(\xi) - V_h(\xi_\star^h) \leq \frac{1}{2} V_h''(\xi_\star^h)(\xi - \xi_\star^h)^2 + \frac{1}{6}[V_h'']_{\mathrm{Lip}}|\xi - \xi_\star^h|^3. \tag{C.13}$$

Hence, if $\xi \in I^h$, the above inequality implies

$$V_h(\xi) - V_h(\xi_\star^h) \leq \frac{2}{3} V_h''(\xi_\star^h)(\xi - \xi_\star^h)^2. \tag{C.14}$$

Plugging (C.11) and (C.14) into (C.8), we conclude

$$\mathcal{L}_{h,q}'(\xi) V_h'(\xi) \geq \frac{q \mathcal{L}_{h,q}(\xi) V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \geq \frac{3}{8} q V_h''(\xi_\star^h) \mathcal{L}_{h,q}(\xi), \quad \xi \in I^h. \tag{C.15}$$

Assume now that $\xi \notin I^h$, say $\xi > \xi_\star^h + \varepsilon_h$. The case $\xi < \xi_\star^h - \varepsilon_h$ is treated similarly and is omitted. Since V_h' is non decreasing, we get that $V_h'(\xi_\star^h + \varepsilon_h) \geq V_h'(\xi_\star^h) = 0$, and also that

$$V_h'(\xi) \geq V_h'(\xi_\star^h + \varepsilon_h) \geq V_h''(\xi_\star^h)\varepsilon_h - \frac{1}{2}[V_h'']_{\mathrm{Lip}}\varepsilon_h^2 = \frac{V_h''(\xi_\star^h)^2}{2[V_h'']_{\mathrm{Lip}}}, \tag{C.16}$$

where we evaluated (C.10) at $\xi_\star^h + \varepsilon_h \in I^h$ for the last inequality. Plugging this inequality into (C.8), we obtain

$$\mathcal{L}_{h,q}'(\xi)V_h'(\xi) \geq \upsilon\mathcal{L}_{h,q}(\xi)V_h'(\xi)^2 \geq \upsilon\frac{V_h''(\xi_\star^h)^4}{4[V_h'']_{\mathrm{Lip}}^2}\mathcal{L}_{h,q}(\xi), \tag{C.17}$$

for any $\xi > \xi_\star^h + \varepsilon_h$. The above inequality is still valid if $\xi < \xi_\star^h - \varepsilon_h$.

Hence, combining (C.15) and (C.17) yield

$$\mathcal{L}_{h,q}'(\xi)V_h'(\xi) \geq \left\{\frac{3}{8}qV_h''(\xi_\star^h) \wedge \upsilon\frac{V_h''(\xi_\star^h)^4}{4[V_h'']_{\mathrm{Lip}}^2}\right\}\mathcal{L}_{h,q}(\xi), \quad \xi \in \mathbb{R}. \tag{C.18}$$

By Lemma 3.1, there exists $\widetilde{R} > 0$ such that, $\forall h \in \mathcal{H}$, $\xi_\star^h \in B(\xi_\star, \widetilde{R})$. Assumptions 2(i) and 2(ii) imply that

$$\inf_{h\in\mathcal{H}} V_h''(\xi_\star^h) \geq \frac{1}{1-\alpha} \inf_{\substack{h\in\mathcal{H} \\ \xi\in B(\xi_\star,\widetilde{R})}} f_{X_h}(\xi) > 0, \tag{C.19}$$

and

$$\inf_{h\in\mathcal{H}} \frac{1}{[V_h'']_{\mathrm{Lip}}} \geq \frac{1-\alpha}{\sup_{h\in\mathcal{H}}[f_{X_h}]_{\mathrm{Lip}}} > 0, \tag{C.20}$$

which eventually gives (3.11). The upper-estimate (3.12) follows directly from the fact that for any $\xi \in \mathbb{R}$, $V_h(\xi) - V_h(\xi_\star^h) \leq k_\alpha|\xi - \xi_\star^h|$ recalling that $k_\alpha = 1 \vee \frac{\alpha}{1-\alpha}$ and the very definition of υ_h.

(iii) We first assume that $q \geq 2$. It follows from (C.2) and the relation $\mathcal{L}_{h,q-1}(\xi) = (V_h(\xi) - V_h(\xi_\star^h))^{-1}\mathcal{L}_{h,q}(\xi)$ that

$$\begin{aligned}\mathcal{L}_{h,q}''(\xi) = \mathcal{L}_{h,q-1}(\xi)&\left\{qV_h''(\xi) + \frac{q(q-1)V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)}\right\} \\ &+ \mathcal{L}_{h,q}(\xi)\left\{\upsilon V_h''(\xi) + \upsilon^2V_h'(\xi)^2 + 2q\upsilon\frac{V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)}\right\}.\end{aligned} \tag{C.21}$$

The triangle inequality applied to (C.12) yields

$$V_h(\xi) - V_h(\xi_\star^h) \geq \frac{1}{2}V_h''(\xi_\star^h)(\xi - \xi_\star^h)^2 - \frac{1}{6}[V_h'']_{\mathrm{Lip}}|\xi - \xi_\star^h|^3. \tag{C.22}$$

Hence, if $\xi \in I^h$ the above inequality directly yields

$$V_h(\xi) - V_h(\xi_\star^h) \geq \frac{1}{3}V_h''(\xi_\star^h)(\xi - \xi_\star^h)^2. \tag{C.23}$$

Moreover, since $V_h'(\xi_\star^h) = 0$ and V_h'' is bounded,

$$V_h'(\xi)^2 \leq \|V_h''\|_\infty^2(\xi - \xi_\star^h)^2. \tag{C.24}$$

Combining (C.23) and (C.24) yields

$$\frac{V_h'(\xi)^2}{V_h(\xi)-V_h(\xi_\star^h)} \le \frac{3\|V_h''\|_\infty^2}{V_h''(\xi_\star^h)}, \quad \xi \in I^h. \tag{C.25}$$

Assume now that $\xi \notin I^h$, say $\xi > \xi_\star^h + \varepsilon_h$. The case $\xi < \xi_\star^h - \varepsilon_h$ can be treated similarly and is thus omitted. Observe that V_h is increasing on $[\xi_\star^h, \infty)$. Given that for all $t \in [0,1]$, $\xi_\star^h + t\varepsilon_h \in I^h$, evaluating (C.10) for $\xi = \xi_\star^h + t\varepsilon_h$, for $t \in [0,1]$, we obtain

$$\begin{aligned}
V_h(\xi)-V_h(\xi_\star^h) &\ge V_h(\xi_\star^h+\varepsilon_h)-V_h(\xi_\star^h) \\
&= \varepsilon_h \int_0^1 V_h'(\xi_\star^h + t\varepsilon_h)\,\mathrm{d}t \\
&\ge \frac{1}{2}V_h''(\xi_\star^h)\varepsilon_h^2 - \frac{1}{6}[V_h'']_{\mathrm{Lip}}\varepsilon_h^3 \\
&= \frac{V_h''(\xi_\star^h)^3}{3[V_h'']_{\mathrm{Lip}}^2},
\end{aligned}$$

by (C.4). Using this inequality together with the fact that

$$\|V_h'\|_\infty \le 1 \vee \frac{\alpha}{1-\alpha} =: k_\alpha, \tag{C.26}$$

we deduce

$$\frac{V_h'(\xi)^2}{V_h(\xi)-V_h(\xi_\star^h)} \le \frac{3k_\alpha^2[V_h'']_{\mathrm{Lip}}^2}{V_h''(\xi_\star^h)^3}, \tag{C.27}$$

for any $\xi > \xi_\star^h + \varepsilon^h$. The above estimate still holds for $\xi < \xi_\star^h + \varepsilon^h$.

From (C.25) and (C.27), setting

$$\nu_h := \frac{3k_\alpha^2[V_h'']_{\mathrm{Lip}}^2}{V_h''(\xi_\star^h)^3} \vee \frac{3\|V_h''\|_\infty^2}{V_h''(\xi_\star^h)}, \tag{C.28}$$

we get:

$$\frac{V_h'(\xi)^2}{V_h(\xi)-V_h(\xi_\star^h)} \le \nu_h, \tag{C.29}$$

for any $\xi \neq \xi_\star^h$. Combining (C.21) with (C.29) yields

$$|\mathcal{L}_{h,q}''(\xi)| \le \big\{(q\vee\upsilon)\|V_h''\|_\infty + \upsilon^2 k_\alpha^2 + q\big(2\upsilon \vee (q-1)\big)\nu_h\big\}\big(\mathcal{L}_{h,q}(\xi)+\mathcal{L}_{h,q-1}(\xi)\big). \tag{C.30}$$

Now if $q=1$. From (C.3), we directly get

$$|\mathcal{L}_{h,1}''(\xi)| \le \big\{(\upsilon\vee 1)\|V_h''\|_\infty + \upsilon(\upsilon\vee 2)k_\alpha^2\big\}\big(\mathcal{L}_{h,0}(\xi)+\mathcal{L}_{h,1}(\xi)\big). \tag{C.31}$$

Therefore, combining the last two inequalities, we conclude that, for all $q \ge 1$ and $\xi \neq \xi_\star^h$,

$$|\mathcal{L}_{h,q}''(\xi)| \le \big\{(q\vee\upsilon)\|V_h''\|_\infty + \upsilon(\upsilon\vee 2)k_\alpha^2 + q\big(2\upsilon \vee (q-1)\big)\nu_h\big\}\big(\mathcal{L}_{h,q}(\xi)+\mathcal{L}_{h,q-1}(\xi)\big). \tag{C.32}$$

Inequality (3.13) for $\upsilon = \bar{\upsilon}_h := q^2\|V_h''\|_\infty$ is easily verified using the very definitions of $\lambda_{h,q}^\upsilon$ and $\eta_{h,q}^\upsilon$. Also, note that

$$\sup_{h\in\mathcal{H}} \|V_h''\|_\infty = \frac{1}{1-\alpha}\sup_{h\in\mathcal{H}}\|f_{X_h}\|_\infty < \infty, \tag{C.33}$$

while Assumption 2(ii) gives

$$\sup_{h\in\mathcal{H}}[V_h'']_{\mathrm{Lip}} = \frac{1}{1-\alpha}\sup_{h\in\mathcal{H}}[f_{X_h}]_{\mathrm{Lip}} < \infty. \tag{C.34}$$

Moreover, since $(\xi_\star^h)_{h\in\mathcal{H}}$ converges to $\xi_\star$ as $\mathcal{H}\ni h\downarrow 0$, there exists $\widetilde{R}>0$ such that, $\forall h\in\mathcal{H}$, $\xi_\star^h\in B(\xi_\star,\widetilde{R})$. Thus, using Assumption 2(i), we get

$$\sup_{h\in\mathcal{H}}\frac{1}{V_h''(\xi_\star^h)} \leq \frac{1-\alpha}{\inf_{\substack{h\in\mathcal{H}\\ \xi\in B(\xi_\star,\widetilde{R})}} f_{X_h}(\xi)} < \infty. \tag{C.35}$$

Coming back to (C.28) and combining (C.33), (C.34) and (C.35) we obtain

$$\sup_{h\in\mathcal{H}}\nu_h < \infty. \tag{C.36}$$

Eventually, properties (C.33) and (C.36) show that

$$\sup_{h\in\mathcal{H}}\Big\{\eta_{h,q}^{\bar{\upsilon}_h} \coloneqq (q\vee\bar{\upsilon}_h)\|V_h''\|_\infty + \bar{\upsilon}_h(\bar{\upsilon}_h\vee 2)k_\alpha^2 + q\big(2\bar{\upsilon}_h\vee(q-1)\big)\nu_h\Big\} < \infty. \tag{C.37}$$

(iv) On I^h, relation (C.23) shows that

$$(\xi-\xi_\star^h)^{2q} \leq \frac{3^q}{V_h''(\xi_\star^h)^q}\big(V_h(\xi)-V_h(\xi_\star^h)\big)^q \leq \frac{3^q}{V_h''(\xi_\star^h)^q}\mathcal{L}_{h,q}(\xi). \tag{C.38}$$

Outside of I^h, say for $\xi>\xi_\star^h+\varepsilon_h$ (the case of $\xi<\xi_\star^h-\varepsilon_h$ is similarly treated), since V_h' is increasing and nonnegative on $[\xi_\star^h,\infty)$, one obtains

$$\forall t\in[0,1],\quad V_h'(\xi_\star^h+t(\xi-\xi_\star^h)) \geq V_h'(\xi_\star^h+t\varepsilon_h), \tag{C.39}$$

and since for all $t\in[0,1]$, $\xi_\star^h+t\varepsilon_h\in I^h$, (C.10) gives

$$\begin{aligned}
V_h(\xi)-V_h(\xi_\star^h) &= (\xi-\xi_\star^h)\int_0^1 V_h'(\xi_\star^h+t(\xi-\xi_\star^h))\mathrm{d}t \\
&\geq (\xi-\xi_\star^h)\int_0^1 V_h'(\xi_\star^h+t\varepsilon_h)\mathrm{d}t \\
&\geq (\xi-\xi_\star^h)\Big\{\frac{1}{2}V_h''(\xi_\star^h)\varepsilon_h - \frac{1}{6}[V_h'']_{\mathrm{Lip}}\varepsilon_h^2\Big\} \\
&= (\xi-\xi_\star^h)\frac{V_h''(\xi_\star^h)^2}{3[V_h'']_{\mathrm{Lip}}}.
\end{aligned} \tag{C.40}$$

Hence

$$(\xi-\xi_\star^h)^{2q} \leq \frac{3^{2q}[V_h'']_{\mathrm{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}}\big(V_h(\xi)-V_h(\xi_\star^h)\big)^{2q} \leq \frac{3^{2q}[V_h'']_{\mathrm{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}}\mathcal{L}_{h,2q}(\xi), \tag{C.41}$$

for any $\xi>\xi_\star^h+\varepsilon_h$. The above estimate still holds if $\xi<\xi_\star^h-\varepsilon_h$. Putting together (C.38) and (C.41), one gets

$$(\xi-\xi_\star^h)^{2q} \leq \Big\{\frac{3^q}{V_h''(\xi_\star^h)^q}\vee\frac{3^{2q}[V_h'']_{\mathrm{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}}\Big\}\big(\mathcal{L}_{h,q}(\xi)+\mathcal{L}_{h,2q}(\xi)\big). \tag{C.42}$$

for any $\xi\neq\xi_\star^h$.

Relations (C.34) and (C.35) yield (3.15).

D Proof of Lemma 4.1

The proof of the point (i) is inspired from [17, Lemma 5.1].

(i) Let $\xi \in \mathbb{R}$. For any $\delta > 0$, it holds

$$\begin{aligned}
\mathbb{E}[|\mathbf{1}_{X>\xi} - \mathbf{1}_{Y>\xi}|] &= \mathbb{E}[\mathbf{1}_{X>\xi}\mathbf{1}_{Y\le\xi} + \mathbf{1}_{Y>\xi}\mathbf{1}_{X\le\xi}] \\
&= \mathbb{P}(Y \le \xi < X) + \mathbb{P}(X \le \xi < Y) \\
&= \mathbb{P}(Y \le \xi, \xi + \delta < X) + \mathbb{P}(Y \le \xi < X \le \xi + \delta) \\
&\quad + \mathbb{P}(X \le \xi, \xi + \delta < Y) + \mathbb{P}(X \le \xi < Y \le \xi + \delta) \\
&\le \mathbb{P}(X - Y > \delta) + \mathbb{P}(X - Y < -\delta) \\
&\quad + \mathbb{P}(\xi < X \le \xi + \delta) + \mathbb{P}(\xi < Y \le \xi + \delta) \\
&\le \mathbb{P}(|X - Y| > \delta) + \delta\big(\|f_X\|_\infty + \|f_Y\|_\infty\big) \\
&\le \frac{\mathbb{E}[|X - Y|^{p_\star}]}{\delta^{p_\star}} + \delta\big(\|f_X\|_\infty + \|f_Y\|_\infty\big).
\end{aligned} \tag{D.1}$$

We come back to (D.1). Under the current assumptions, after a straightforward optimization, we take

$$\delta = \frac{p_\star^{\frac{1}{p_\star+1}} \mathbb{E}[|X - Y|^{p_\star}]^{\frac{1}{p_\star+1}}}{\big(\|f_X\|_\infty + \|f_Y\|_\infty\big)^{\frac{1}{p_\star+1}}}. \tag{D.2}$$

Injecting the above value of δ in the right hand side of (D.1) gives the first result.

(ii) Under the current Gaussian concentration assumption, we first employ the exponential Markov inequality and then (4.5), so that, for any $\delta > 0$ and any $u \ge 0$,

$$\begin{aligned}
\mathbb{E}[|\mathbf{1}_{X>\xi} - \mathbf{1}_{Y>\xi}|] &\le \mathbb{P}(X - Y > \delta) + \mathbb{P}(-(X - Y) > \delta) + \delta\big(\|f_X\|_\infty + \|f_Y\|_\infty\big) \\
&\le \exp(-u\delta)\big(\mathbb{E}[\exp(u(X - Y))] + \mathbb{E}[\exp(-u(X - Y))]\big) + \delta\big(\|f_X\|_\infty + \|f_Y\|_\infty\big) \\
&\le 2\exp(d_{X,Y}u^2 - u\delta) + \delta\big(\|f_X\|_\infty + \|f_Y\|_\infty\big).
\end{aligned} \tag{D.3}$$

The first term appearing in the right hand side of the above inequality above is minimized in u for $\delta/(2d_{X,Y})$, in which case the inequality becomes

$$\mathbb{E}[|\mathbf{1}_{X>\xi} - \mathbf{1}_{Y>\xi}|] \le 2\exp\Big(-\frac{\delta^2}{4d_{X,Y}}\Big) + \delta\big(\|f_X\|_\infty + \|f_Y\|_\infty\big). \tag{D.4}$$

Evaluating the right hand side of the previous inequality at

$$\delta = \sqrt{2d_{X,Y}\ln\big(d_{X,Y}^{-1} \vee 1\big)}$$

eventually gives

$$\mathbb{E}[|\mathbf{1}_{X>\xi} - \mathbf{1}_{Y>\xi}|] \le 2\sqrt{d_{X,Y}} + \sqrt{2}\big(\|f_X\|_\infty + \|f_Y\|_\infty\big)\sqrt{d_{X,Y}\ln\big(d_{X,Y}^{-1} \vee 1\big)}. \tag{D.5}$$

This completes the proof of the second result.

E Proof of Proposition 4.1

The proof of the point (i)a is inspired from [17, Lemma 3.2].

(i)a. Assume that $\mathbb{E}\big[\big|\varphi(Y,Z)-\mathbb{E}[\varphi(Y,Z)|Y]\big|^p\big]<\infty$ for some $p\geq 1$. Then,

$$\mathbb{E}[|X_h-X_0|^p]\leq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\big[\big|\varphi(Y,Z^{(k)})-\mathbb{E}[\varphi(Y,Z)|Y]\big|^p\big]=\mathbb{E}\big[\big|\varphi(Y,Z)-\mathbb{E}[\varphi(Y,Z)|Y]\big|^p\big].$$

Hence, introducing $\widetilde{\varphi}(Y,Z)\coloneqq\varphi(Y,Z)-\mathbb{E}[\varphi(Y,Z)|Y]$ and assuming that $h=\frac{1}{K}\leq h'=\frac{1}{K'}$, one has

$$X_h-X_{h'}=h\sum_{k=K'+1}^{K}\widetilde{\varphi}(Y,Z^{(k)})+(h-h')\sum_{k=1}^{K'}\widetilde{\varphi}(Y,Z^{(k)}),\tag{E.1}$$

so that, by the triangle inequality,

$$\mathbb{E}[|X_h-X_{h'}|^p]^{\frac{1}{p}}\leq h\mathbb{E}\Big[\Big|\sum_{k=K'+1}^{K}\widetilde{\varphi}(Y,Z^{(k)})\Big|^p\Big]^{\frac{1}{p}}+(h'-h)\,\mathbb{E}\Big[\Big|\sum_{k=1}^{K'}\widetilde{\varphi}(Y,Z^{(k)})\Big|^p\Big]^{\frac{1}{p}}.$$

Using the tower law property of conditional expectation and Burkholder-Davis-Gundy's inequality we obtain

$$\begin{aligned}\mathbb{E}[|X_h-X_{h'}|^p]^{\frac{1}{p}}&\leq B_p\,\mathbb{E}\big[\big|\varphi(Y,Z)-\mathbb{E}[\varphi(Y,Z)|Y]\big|^p\big]^{\frac{1}{p}}\left(h(K-K')^{\frac{1}{2}}+(h'-h){K'}^{\frac{1}{2}}\right)\\&\leq B_p\,\mathbb{E}\big[\big|\varphi(Y,Z)-\mathbb{E}[\varphi(Y,Z)|Y]\big|^p\big]^{\frac{1}{p}}(h'-h)^{\frac{1}{2}}\left(\Big(\frac{h}{h'}\Big)^{\frac{1}{2}}+\Big(1-\frac{h}{h'}\Big)^{\frac{1}{2}}\right)\\&\leq\sqrt{2}B_p\,\mathbb{E}\big[\big|\varphi(Y,Z)-\mathbb{E}[\varphi(Y,Z)|Y]\big|^p\big]^{\frac{1}{p}}(h'-h)^{\frac{1}{2}}.\end{aligned}\tag{E.2}$$

We then conclude the proof of the first point by applying Lemma 4.1(i) for $p=p_\star$.

(i)b. We go back to the decomposition (E.1), use the fact that the random variables $\widetilde{\varphi}(Y,Z^{(k)})$, $k=1,\dots,K$, are independent conditionally to Y, and then use (4.7):

$$\begin{aligned}\mathbb{E}\big[\exp\big(u(X_h-X_{h'})\big)\big|Y\big]&=\mathbb{E}\big[\exp\big(uh\widetilde{\varphi}(Y,Z)\big)\big|Y\big]^{K-K'}\mathbb{E}\big[\exp\big(u(h-h')\widetilde{\varphi}(Y,Z)\big)\big|Y\big]^{K'}\\&\leq\exp\Big(C_gu^2h^2\Big(\frac{1}{h}-\frac{1}{h'}\Big)\Big)\exp\Big(C_gu^2\frac{(h-h')^2}{h'}\Big)\\&=\exp\big(C_g(h'-h)u^2\big).\end{aligned}$$

Taking expectation in both sides of the above inequality guarantees that (4.5) is satisfied with $X=X_h$, $Y=X_{h'}$ and $d_{X,Y}=C_g(h'-h)$. We eventually conclude the proof by applying Lemma 4.1(ii).

(ii) We first write

$$\mathbb{E}\big[\big|\mathbf{1}_{X_{h_\ell}>\xi}-\mathbf{1}_{X_{h_{\ell-1}}>\xi}\big|\big]=\mathbb{P}(X_{h_{\ell-1}}\leq\xi<X_{h_\ell})+\mathbb{P}(X_{h_\ell}\leq\xi<X_{h_{\ell-1}}).\tag{E.3}$$

Introducing the random variable G_ℓ, it holds

$$\begin{aligned}\mathbb{P}(X_{h_{\ell-1}}\leq\xi<X_{h_\ell})&=\mathbb{P}\big(X_{h_{\ell-1}}\leq\xi<X_{h_{\ell-1}}+h_\ell^{\frac{1}{2}}G_\ell\big)\\&=\mathbb{P}\big(X_{h_{\ell-1}}\leq\xi<X_{h_{\ell-1}}+h_\ell^{\frac{1}{2}}G_\ell,G_\ell>0\big)\\&=\mathbb{E}\big[\mathbb{P}\big(\xi-h_\ell^{\frac{1}{2}}G_\ell<X_{h_{\ell-1}}\leq\xi,G_\ell>0\big|G_\ell\big)\big]\\&=\mathbb{E}\big[\mathbf{1}_{G_\ell>0}\big(F_{X_{h_{\ell-1}}\,|\,G_\ell}(\xi)-F_{X_{h_{\ell-1}}\,|\,G_\ell}(\xi-h_\ell^{\frac{1}{2}}G_\ell)\big)\big].\end{aligned}\tag{E.4}$$

The $\mathbb{P}$-as-Lipschitz regularity of $F_{X_{h_{\ell-1}}|G_\ell}$ now yields

$$\mathbb{P}(X_{h_{\ell-1}} \leq \xi < X_{h_\ell}) \leq h_\ell^{\frac{1}{2}} \mathbb{E}[G_\ell^+ K_\ell].$$

A similar argument gives

$$\mathbb{P}(X_{h_\ell} \leq \xi \leq X_{h_{\ell-1}}) \leq h_\ell^{\frac{1}{2}} \mathbb{E}[G_\ell^- K_\ell].$$

Coming back to (E.3) and summing up the two preceding inequalities yields

$$\mathbb{E}\big[\big|\mathbf{1}_{X_{h_\ell}>\xi} - \mathbf{1}_{X_{h_{\ell-1}}>\xi}\big|\big] \leq h_\ell^{\frac{1}{2}} \mathbb{E}[|G_\ell| K_\ell],$$

which completes the proof.

References

[1] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.

[2] C. Albanese, Y. Armenti, and S. Crépey. Xva metrics for ccp optimization. *Statistics & Risk Modeling*, 37(1-2):25–53, 2020.

[3] O. Bardou, N. Frikha, and G. Pagès. Recursive computation of value-at-risk and conditional value-at-risk using mc and qmc. In Pierre L'Ecuyer and Art B. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 193–208, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[4] O. Bardou, N. Frikha, and G. Pagès. CVaR hedging using quantization-based stochastic approximation algorithm. *Mathematical Finance*, 26(1):184–229, 2016.

[5] O. Bardou, N. Frikha, and G. Pagès. Computing var and cvar using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009.

[6] D. Barrera, S. Crépey, B. Diallo, G. Fort, E. Gobet, and U. Stazhynski. Stochastic Approximation Schemes for Economic Capital and Risk Margin Computations. *ESAIM: Proceedings and Surveys*, 65:182–218, 2019.

[7] Basel Committee on Banking Supervision. Consultative document: Fundamental Review of the Trading Book: A revised market risk framework. 2013.

[8] M. Broadie, Y. Du, and C. Moallemi. Risk estimation via regression. *Operations Research*, 63(5):1077–1097, 2015.

[9] M. Costa and S. Gadat. Non asymptotic controls on a recursive superquantile approximation. *Electronic Journal of Statistics*, 15(2):4718 – 4769, 2021.

[10] S. Dereich and T. Müller-Gronbach. General multilevel adaptations for stochastic approximation algorithms of robbins–monro and polyak–ruppert type. *Numerische Mathematik*, 142(2):279–328, 2019.

[11] N. Frikha. Shortfall risk minimization in discrete time financial market models. *SIAM Journal on Financial Mathematics*, 5(1):384–414, 2014.

[12] N. Frikha. Multi-level stochastic approximation algorithms. *The Annals of Applied Probability*, 26(2):933 – 985, 2016.

[13] N. Frikha and L. Huang. A multi-step Richardson-Romberg extrapolation method for stochastic approximation. 31 pages, 1 figure, March 2015.

[14] H. Föllmer and A. Schied. *Convex Risk Measures.* John Wiley & Sons, Ltd, 2010.

[15] M. B. Giles. Multilevel monte carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[16] M. B. Giles and A.-L. Haji-Ali. Multilevel nested simulation for efficient risk estimation. *SIAM/ASA Journal on Uncertainty Quantification*, 7(2):497–525, 2019.

[17] D. Giorgi, V. Lemaire, and G. Pagès. Weak error for nested multilevel monte carlo. *Methodology and Computing in Applied Probability*, 22:1325–1348, 2020.

[18] M. B. Gordy and S. Juneja. Nested simulation in portfolio risk measurement. *Management Science*, 56(10):1833–1848, 2010.

[19] R. T. Rockafellar and S. Uryasev. Optimization of Conditional Value-at-Risk. *Journal of risk*, 2(3):21–41, 2000.