



HAL
open science

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

Stéphane Crépey, Noufel Frikha, Azar Louzi

► **To cite this version:**

Stéphane Crépey, Noufel Frikha, Azar Louzi. A Multilevel Stochastic Approximation Algorithm for Value-at-Risk and Expected Shortfall Estimation. 2024. hal-04037328v3

HAL Id: hal-04037328

<https://hal.science/hal-04037328v3>

Preprint submitted on 29 Nov 2024

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

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

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

Stéphane Crépey*, Noufel Frikha†, Azar Louzi‡§

November 29, 2024

Abstract

We propose a multilevel stochastic approximation (MLSA) scheme for the computation of the value-at-risk (VaR) and 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 stochastic approximation problems with biased innovations. In this framework, for a prescribed accuracy ε , the optimal complexity of a nested 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 performance lost due to the nested nature of the problem.

Keywords. value-at-risk, expected shortfall, stochastic approximation, nested Monte Carlo, multilevel Monte Carlo, numerical finance.

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

Introduction

The post-great recession era has witnessed the implementation of multiple risk measures with the goal of better controlling financial losses. In a spirit of precaution and consistency, the Fundamental Review of the Trading Book [7] triggered a shift away from the value-at-risk (VaR) towards the expected shortfall (ES, i.e. the average loss given it exceeds the VaR) as a reference regulatory risk measure. The future valuation of a position in financial derivatives is defined by a conditional expectation. In the case of exotic products or non-linear portfolio computations (e.g. credit or funding valuation adjustments), this future valuation can only be computed by Monte Carlo simulation. VaR and ES computations on future portfolio losses are thus nested in nature. A brute force nested Monte Carlo computational approach à la Gordy and Juneja [23]

*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, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, 75013 Paris (azar.louzi@lpsm.paris).

§Alphabetical order.

may be too heavy for practical use. As a shortcut solution, a regression based estimator for the inner conditional expectation is implemented by Broadie et al. [9], but the resulting regression error is difficult to control. Barrera et al. [6] adopt the stochastic approximation (SA) point of view of Bardou et al. [3] (see also [4, 5, 15]), itself building on top of Rockafellar and Uryasev's representation of the VaR and ES as joint solutions to a convex optimization problem [25]. Barrera et al. [6] revisit both Gordy and Juneja's [23] and Broadie et al.'s [9] computational strategies under assumptions that are more realistic and easier to check, but we clarify in this paper that in order to reach an estimation accuracy of order ε , the complexity of the resulting nested SA algorithm is of order ε^{-3} . As for their regression strategy, its non-asymptotic error bounds involve some constants that are rather large, which renders them of theoretical interest only.

In the present paper, we propose a multilevel SA (MLSA) algorithm for the computation of the VaR and ES of a loss that writes as a conditional expectation. MLSA was introduced by Frikha [14] as an extension to the SA framework of multilevel Monte Carlo path simulation of Giles [18]. It was then revisited by Dereich and Müller-Gronbach [12] from the different perspective of approximating the objective function of an SA problem rather than its solutions. We also refer to Frikha and Huang [16] for a multi-step Richardson-Romberg stochastic approximation method. We stress that the uniform mean-reversion assumption on the objective function discussed by Frikha [14] is generally not satisfied in a VaR-ES setup, so the theoretical performance of the MLSA scheme applied to both risk measures is not directly guaranteed by the results therein. We also refer to Giles and Haji-Ali [19] who 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 do not provide a theoretical analysis of the latter. Our main contribution is to propose an optimized MLSA algorithm that achieves sharp theoretical complexities when focusing on estimating either the VaR or the ES. The VaR focused estimation achieves a complexity of order $\varepsilon^{-2-\delta}$, where $\delta \in (0, 1)$ 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 that show a considerable gain with respect to the nested SA approach in terms of computational time needed to estimate the VaR and ES.

The paper is organized as follows. Section 1 presents the problem of computing the VaR and ES by means of an SA scheme. Section 2 analyzes the nested SA scheme for the computation of the VaR and ES. Section 3 introduces and analyzes the MLSA algorithm. The theoretical analysis and the gain in performance achieved by our optimized MLSA scheme are illustrated numerically in Section 4. The proofs of the technical results are given in appendix form.

1 Baseline Setup

1.1 Value-at-Risk and Expected Shortfall

Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space on which is defined a real-valued financial loss

$$X_0 = \mathbb{E}[\varphi(Y, Z)|Y], \quad (1.1)$$

where $Y \in \mathbb{R}^d$, $Z \in \mathbb{R}^q$ are two independent random variables, and $\varphi: \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a measurable function such that $\varphi(Y, Z) \in L^1(\mathbb{P})$. Given a portfolio and a time horizon $\delta > 0$, Y and Z typically model the portfolio's risk factors respectively up to and beyond δ , $\varphi(Y, Z)$ the subsequent portfolio's future cash flows and X_0 the portfolio's future loss at time δ (see Sections 4.1 and 4.2.)

Hence, if $\varphi(y, Z) \in L^1(\mathbb{P})$ for all $y \in \mathbb{R}^d$, we can rewrite X_0 as

$$X_0 = \mathbb{E}[\varphi(Y, Z)|Y] = \Phi(Y), \quad \text{with} \quad \Phi(y) := \mathbb{E}[\varphi(y, Z)], \quad y \in \mathbb{R}^d. \quad (1.2)$$

We are interested in computing the VaR and ES of X_0 at a given confidence level $\alpha \in (0, 1)$, denoted respectively $\text{VaR}_\alpha(X_0)$ and $\text{ES}_\alpha(X_0)$. As per Föllmer and Schied [13] and Acerbi and Tasche [1], these risk measures are defined by

$$\text{VaR}_\alpha(X_0) := \inf \{ \xi \in \mathbb{R} : \mathbb{P}[X_0 \leq \xi] \geq \alpha \}, \quad \text{ES}_\alpha(X_0) := \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_a(X_0) da.$$

1.2 Stochastic Approximation

As established by Rockafellar and Uryasev [25], the VaR and ES are linked by a convex optimization problem. To accurately state this connection, we introduce the mapping

$$V_0(\xi) := \xi + \frac{1}{1-\alpha} \mathbb{E}[(X_0 - \xi)^+], \quad \xi \in \mathbb{R}. \quad (1.3)$$

If the distribution function F_{X_0} of X_0 is continuous, then V_0 is continuously differentiable on \mathbb{R} and

$$V_0'(\xi) = \frac{1}{1-\alpha} (F_{X_0}(\xi) - \alpha), \quad \xi \in \mathbb{R}.$$

If X_0 admits a continuous density function f_{X_0} , then V_0 is twice continuously differentiable on \mathbb{R} with

$$V_0''(\xi) = \frac{1}{1-\alpha} f_{X_0}(\xi), \quad \xi \in \mathbb{R}.$$

Lemma 1.1 ([5, Proposition 2.1]). *Suppose $X_0 \in L^1(\mathbb{P})$ and the distribution function of X_0 is continuous. Then V_0 is convex and continuously differentiable, $\lim_{|\xi| \rightarrow \infty} V_0(\xi) = \infty$ and*

$$\text{VaR}_\alpha(X_0) = \min \text{Argmin } V_0,$$

where

$$\text{Argmin } V_0 = \{V_0' = 0\} = \{\xi \in \mathbb{R} : \mathbb{P}[X_0 \leq \xi] = \alpha\}$$

is a closed bounded non-empty interval of \mathbb{R} . Moreover,

$$V_0'(\xi) = \mathbb{E}[H_1(\xi, X_0)], \quad \text{where } H_1(\xi, x) = 1 - \frac{1}{1-\alpha} \mathbf{1}_{\{x \geq \xi\}}, \quad \xi, x \in \mathbb{R}, \quad (1.4)$$

and

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

In general, the set $\text{Argmin } V_0$ does not reduce to a singleton. But if the distribution function of X_0 is increasing then it does and $\text{argmin } V_0 = \text{VaR}_\alpha(X_0)$.

Let $\xi_\star^0 \in \text{Argmin } V_0$ and $\chi_\star^0 = \min V_0 = V_0(\xi_\star^0)$. Under the approach initiated by Bardou et al. [5], in the case of an exactly computable function Φ in (1.2) and an increasing distribution function of X_0 , one may approximate the unique couple $(\xi_\star^0, \chi_\star^0)$ by a sequence $(\xi_n^0, \chi_n^0)_{n \geq 0}$ given by the two-time-scale SA dynamics

$$\begin{cases} \xi_{n+1}^0 = \xi_n^0 - \gamma_{n+1} H_1(\xi_n^0, X_0^{(n+1)}), \\ \chi_{n+1}^0 = \chi_n^0 - \frac{1}{n+1} H_2(\xi_n^0, \chi_n^0, X_0^{(n+1)}), \end{cases} \quad (1.5)$$

where

$$H_2(\xi, \chi, x) = \chi - \left(\xi + \frac{1}{1-\alpha} (x - \xi)_+ \right), \quad \xi, \chi, x \in \mathbb{R}.$$

The sequence $(X_0^{(n)})_{n \geq 1}$ stands for i.i.d. copies of X_0 and (ξ_0^0, χ_0^0) is a random vector independent of $(X_0^{(n)})_{n \geq 1}$ satisfying $\mathbb{E}[|\xi_0^0|^2] + \mathbb{E}[|\chi_0^0|^2] < \infty$. The learning rate sequence $(\gamma_n)_{n \geq 1}$ in (1.5) is deterministic, positive, non-increasing and such that

$$\sum_{n \geq 1} \gamma_n = \infty, \quad \sum_{n \geq 1} \gamma_n^2 < \infty.$$

While such behavior is typical for learning rates in the stochastic approximation literature, it is not required in our analyses as we allow $\sum_{n \geq 1} \gamma_n^2 = \infty$. See Algorithm 1.

Algorithm 1 SA for estimating (VaR, ES)

Require: $N \in \mathbb{N}$, a non-increasing positive sequence $(\gamma_n)_{n \geq 1}$.

1: Choose (ξ_0^0, χ_0^0) such that $\mathbb{E}[|\xi_0^0|^2] + \mathbb{E}[|\chi_0^0|^2] < \infty$

2: **for** $n = 0 \dots N - 1$ **do**

3: Simulate $X_0^{(n+1)} \sim X_0 = \Phi(Y)$

4: $\xi_{n+1}^0 \leftarrow \xi_n^0 - \gamma_{n+1} H_1(\xi_n^0, X_0^{(n+1)})$

5: $\chi_{n+1}^0 \leftarrow \chi_n^0 - \frac{1}{n+1} H_2(\chi_n^0, \xi_n^0, X_0^{(n+1)})$

6: **end for**

7: **return** (ξ_N^0, χ_N^0)

Usually however, one does not have access to an exact simulator of X_0 as given by (1.1), inasmuch as the law of $\varphi(Y, Z)$ conditionally on Y is not known and no analytical expression of Φ in (1.2) is available.

2 Nested Stochastic Approximation

The above discussion naturally suggests to replace the samples of X_0 in the dynamics (1.5) by approximate samples thereof.

We consider a bias parameter

$$h = \frac{1}{K} \in \mathcal{H} := \left\{ \frac{1}{K} : K \in \mathbb{N} \right\}.$$

We then approximate X_0 by the random variable X_h defined by

$$X_h = \frac{1}{K} \sum_{k=1}^K \varphi(Y, Z^{(k)}), \quad (2.1)$$

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

In the spirit of the previous section, assuming that the distribution function of X_h is continuous, we define the approximating optimization problem

$$\min_{\xi \in \mathbb{R}} V_h(\xi), \quad \text{where} \quad V_h(\xi) := \xi + \frac{1}{1-\alpha} \mathbb{E}[(X_h - \xi)^+], \quad \xi \in \mathbb{R}. \quad (2.2)$$

Since the distribution function F_{X_h} of X_h is continuous, then V_h is continuously differentiable on \mathbb{R} and $V_h'(\xi) = \frac{1}{1-\alpha}(F_{X_h}(\xi) - \alpha)$, $\xi \in \mathbb{R}$. If X_h admits a continuous density function f_{X_h} , then V_h is twice continuously differentiable on \mathbb{R} and $V_h''(\xi) = \frac{1}{1-\alpha} f_{X_h}(\xi)$, $\xi \in \mathbb{R}$.

By Lemma 1.1,

$$\text{Argmin } V_h = \{V_h' = 0\} \neq \emptyset \quad \text{and} \quad \chi_\star^h = \min V_h = V_h(\xi_\star^h).$$

Moreover, any minimizer ξ_\star^h of V_h satisfies $\mathbb{P}[X_h \leq \xi_\star^h] = \alpha$. If the distribution function of X_h is increasing, then $\xi_\star^h = \text{argmin } V_h$ is uniquely defined.

If the distribution function of X_h is increasing, in order to approximate the unique couple $(\xi_\star^h, \chi_\star^h)$, we devise the following two-time-scale nested SA (NSA) algorithm

$$\begin{cases} \xi_{n+1}^h = \xi_n^h - \gamma_{n+1} H_1(\xi_n^h, X_h^{(n+1)}), \\ \chi_{n+1}^h = \chi_n^h - \frac{1}{n+1} H_2(\xi_n^h, \chi_n^h, X_h^{(n+1)}), \end{cases} \quad (2.3)$$

where $(X_h^{(n)})_{n \geq 1}$ is a sequence of i.i.d. copies of X_h and (ξ_0^h, χ_0^h) is an \mathbb{R}^2 -valued random variable independent of $(X_h^{(n)})_{n \geq 1}$ satisfying $\mathbb{E}[|\xi_0^h|^2] + \mathbb{E}[|\chi_0^h|^2] < \infty$.

This scheme is nested in nature, in the sense that the update of the outer layer $(\xi_{n+1}^h, \chi_{n+1}^h)$ at step $n+1$ entails simulating the inner layer $X_h^{(n+1)}$ by Monte Carlo as in (2.1). Besides, this scheme is biased, as the target of $(\xi_n^h, \chi_n^h)_{n \geq 0}$ is $(\xi_\star^h, \chi_\star^h)$, which hopefully converges to $(\xi_\star^0, \chi_\star^0)$ as $\mathcal{H} \ni h \downarrow 0$.

Algorithm 2 summarizes the NSA procedure for approximating $(\xi_\star^0, \chi_\star^0)$.

Algorithm 2 Nested SA for estimating (VaR, ES)

Require: $K, N \in \mathbb{N}$, a positive non-increasing sequence $(\gamma_n)_{n \geq 1}$.

- 1: Choose (ξ_0^h, χ_0^h) such that $\mathbb{E}[|\xi_0^h|^2] + \mathbb{E}[|\chi_0^h|^2] < \infty$
 - 2: **for** $n = 0 \dots N-1$ **do**
 - 3: Simulate $Y^{(n+1)} \sim Y$ and $Z^{(n+1,1)}, \dots, Z^{(n+1,K)}$ i.i.d. Z independently of $Y^{(n+1)}$
 - 4: $X_h^{(n+1)} \leftarrow \frac{1}{K} \sum_{k=1}^K \varphi(Y^{(n+1)}, Z^{(n+1,k)})$
 - 5: $\xi_{n+1}^h \leftarrow \xi_n^h - \gamma_{n+1} H_1(\xi_n^h, X_h^{(n+1)})$
 - 6: $\chi_{n+1}^h \leftarrow \chi_n^h - \frac{1}{n+1} H_2(\chi_n^h, \xi_n^h, X_h^{(n+1)})$
 - 7: **end for**
 - 8: **return** (ξ_N^h, χ_N^h)
-

2.1 Convergence Analysis

We study here the NSA scheme (2.3). We first analyze its bias, then prove an $L^2(\mathbb{P})$ -error bound on the iterates, after that analyze its complexity and eventually establish a tuning method for the number of iterations with respect to the bias parameter $h \in \mathcal{H}$ in order to achieve some prescribed error.

For $h \in \overline{\mathcal{H}} := \mathcal{H} \cup \{0\}$, we denote $\Theta_h := \text{Argmin } V_h$, assuming that the distribution function of X_h is continuous.

Lemma 2.1. *Suppose that $\varphi(Y, Z) \in L^1(\mathbb{P})$, that for any $h \in \overline{\mathcal{H}}$, the distribution function 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*

$$\sup_{\xi \in \Theta_h} \text{dist}(\xi, \Theta_0) \rightarrow 0 \quad \text{as } \mathcal{H} \ni h \downarrow 0.$$

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

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

Proof. See Appendix B. □

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

Assumption 2.2. (i) For any $h \in \mathcal{H}$, F_{X_h} admits the first order Taylor expansion

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

for some functions $v, w(\cdot, h): \mathbb{R} \rightarrow \mathbb{R}$ satisfying, for any $\xi_\star^0 \in \Theta_0$,

$$\int_{\xi_\star^0}^{\infty} v(\xi) d\xi < \infty, \quad \lim_{\mathcal{H} \ni h \downarrow 0} w(\xi_\star^0, h) = \lim_{\mathcal{H} \ni h \downarrow 0} \int_{\xi_\star^0}^{\infty} w(\xi, h) d\xi = 0.$$

(ii) For any $h \in \overline{\mathcal{H}}$, the law of X_h admits a continuous 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} .

Remark 2.3. We refer to Giorgi et al.'s [22, Proposition 5.1(b)] for a result on the satisfaction of the first part of Assumption 2.2(i). The last part of Assumption 2.2(i) reads also $w(\xi_\star^0, h) = \int_{\xi_\star^0}^\infty w(\xi, h)d\xi = o(1)$ as $\mathcal{H} \ni h \downarrow 0$, i.e.

$$F_{X_h}(\xi_\star^0) - F_{X_0}(\xi_\star^0) = v(\xi_\star^0)h + o(h),$$

$$\int_{\xi_\star^0}^\infty (F_{X_h}(\xi) - F_{X_0}(\xi))d\xi = h \int_{\xi_\star^0}^\infty v(\xi)d\xi + o(h).$$

Assumption 2.2(ii) is a weakened postulate in comparison with the second part of Giorgi et al.'s [22, Proposition 5.1(a)].

The proposition below quantifies the weak error implied by approximating the unbiased problem with the biased one, in the form of a first order expansion in the bias parameter h of the error between $(\xi_\star^h, \chi_\star^h)$ and $(\xi_\star^0, \chi_\star^0)$.

Proposition 2.4. *Suppose that $\varphi(Y, Z) \in L^1(\mathbb{P})$, that Assumption 2.2 holds, and that the density function f_{X_0} is positive. Then, as $\mathcal{H} \ni h \downarrow 0$, for any $\xi_\star^h \in \Theta_h$,*

$$\xi_\star^h - \xi_\star^0 = -\frac{v(\xi_\star^0)}{f_{X_0}(\xi_\star^0)}h + o(h), \quad \chi_\star^h - \chi_\star^0 = -h \int_{\xi_\star^0}^\infty \frac{v(\xi)}{1-\alpha}d\xi + o(h).$$

Proof. See Appendix C. □

Assumption 2.5. (i) For any $R > 0$,

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

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

$$\sup_{h \in \overline{\mathcal{H}}} \{ \|f_{X_h}\|_\infty + [f_{X_h}]_{\text{Lip}} \} < \infty.$$

Remark 2.6. Assumption 2.5(i) is natural in view of Assumption 2.2(ii) if one assumes the distribution function F_{X_0} increasing. Assumption 2.5(ii) is in line with Assumption 2.2(ii).

Henceforth, we denote $k_\alpha := 1 \vee \frac{\alpha}{1-\alpha}$. Recalling the definition of $(\bar{\lambda}_{h,q})_{h \in \overline{\mathcal{H}}}$ in Lemma D.1(iii), for any positive integer q , we let

$$\bar{\lambda}_q := \inf_{h \in \overline{\mathcal{H}}} \bar{\lambda}_{h,q}. \tag{2.4}$$

Theorem 2.7. (i) *Suppose that $\varphi(Y, Z) \in L^2(\mathbb{P})$, that Assumptions 2.2 and 2.5 hold, and that*

$$\sup_{h \in \overline{\mathcal{H}}} \mathbb{E} \left[|\xi_0^h|^2 \exp \left(\frac{4}{1-\alpha} k_\alpha \sup_{h \in \overline{\mathcal{H}}} \|f_{X_h}\|_\infty |\xi_0^h| \right) \right] + \sup_{h \in \overline{\mathcal{H}}} \mathbb{E}[|\chi_0^h|] < \infty.$$

If $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0, 1]$, with $\bar{\lambda}_1 \gamma_1 > 1$ if $\beta = 1$, then, for any $h \in \mathcal{H}$ and any positive integer n ,

$$\mathbb{E}[(\xi_n^h - \xi_\star^h)^2] \leq \bar{K}_{h,2}^\beta \gamma_n, \tag{2.5}$$

for some constants $(\bar{K}_{h,2}^\beta)_{h \in \mathcal{H}}$ such that $\sup_{h \in \mathcal{H}} \bar{K}_{h,2}^\beta < \infty$. Moreover,

$$\mathbb{E}[|\chi_n^h - \chi_\star^h|] \leq \frac{C_h^\beta}{n^{\frac{1}{2} \wedge \beta}},$$

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

(ii) Assume further that

$$\sup_{h \in \mathcal{H}} \mathbb{E} \left[|\xi_0^h|^4 \exp \left(\frac{16}{1-\alpha} k_\alpha \sup_{h \in \mathcal{H}} \|f_{X_h}\|_\infty |\xi_0^h| \right) \right] + \sup_{h \in \mathcal{H}} \mathbb{E}[(\chi_0^h)^2] < \infty.$$

If $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0, 1]$, with $\bar{\lambda}_2 \gamma_1 > 2$ when $\beta = 1$, then, for any $h \in \mathcal{H}$ and any positive integer n ,

$$\mathbb{E}[(\xi_n^h - \xi_\star^h)^4] \leq \bar{K}_{h,4}^\beta \gamma_n^2, \quad (2.6)$$

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

$$\mathbb{E}[(\chi_n^h - \chi_\star^h)^2] \leq \frac{\bar{C}_h^\beta}{n^{1 \wedge 2\beta}},$$

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

Proof. See Appendix D. □

2.2 Complexity Analysis

Proposition 2.4 and Theorem 2.7(i) suggest a heuristic to balance the bias parameter $h \in \mathcal{H}$ with respect to the number of steps n in (2.3) in order to achieve a prescribed accuracy $\varepsilon \in (0, 1)$.

Theorem 2.8. *Let $\varepsilon \in (0, 1)$ a prescribed accuracy. Within the framework of Theorem 2.7(i), setting*

$$h = \frac{1}{\lceil \varepsilon^{-1} \rceil} \quad \text{and} \quad n = \lceil \varepsilon^{-\frac{2}{\beta}} \rceil$$

yields a global $L^1(\mathbb{P})$ error of order ε . The corresponding computational cost satisfies

$$\text{Cost}_{\text{NSA}} \leq C n h^{-1} \sim C \varepsilon^{-\frac{2}{\beta}-1} \quad \text{as } \varepsilon \downarrow 0,$$

for some constant C independent of ε . The minimal computational cost is of order ε^{-3} and is attained for $\gamma_n = \gamma_1 n^{-1}$, i.e. $\beta = 1$ under the restriction $\bar{\lambda}_1 \gamma_1 > 1$.

Proof. Following Proposition 2.4 and Theorem 2.7(i), if $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0, 1]$, with $\bar{\lambda}_1 \gamma_1 > 1$ if $\beta = 1$, then there exists $C > 0$ such that, for any $h \in \mathcal{H}$ and any positive integer n ,

$$\mathbb{E}[|\xi_n^h - \xi_\star^0|] \leq C(h + n^{-\frac{\beta}{2}}), \quad \mathbb{E}[|\chi_n^h - \chi_\star^0|] \leq C(h + n^{-\frac{1}{2} \wedge \beta}).$$

The sought results follow easily from these upper bounds. □

Remark 2.9. Barrera et al.'s [6, Algorithm 1] differs from our NSA scheme in two ways. First, they use a single-time-scale scheme for both the VaR and ES. We use in contrast a two-time-scale scheme, with a slower and more precise VaR component (independent from the ES) and a faster ES component. Second, the n -th iteration of their NSA algorithm uses a bias parameter h_n dependent upon n , where $(h_n)_{n \geq 1}$ tends to 0 as $n \uparrow \infty$, hence building iterates such that $(\xi_n^{h_n}, \chi_n^{h_n}) \rightarrow (\xi_\star^0, \chi_\star^0)$ as $n \uparrow \infty$. However, their approach results in a substantially increased complexity. Indeed, in view of Barrera et al.'s [6, Theorem 3.2], the error is of order $\sqrt{\gamma_n} = \gamma_1^{1/2} n^{-\beta/2}$, $\beta \in (\frac{1}{2}, 1]$, $n \geq 1$. For a prescribed accuracy $\varepsilon > 0$, one has to choose $n = \lceil \varepsilon^{-2/\beta} \rceil$. In [6, Section 3.2] Barrera et al. recommend selecting $h_n = n^{-\beta'}$, $\beta' > \beta$, $n \geq 1$. This results in a complexity of $\sum_{k=1}^n h_k^{-1} \leq C \varepsilon^{-2(1+\beta')/\beta}$, which is optimal when $\beta' \rightarrow \beta = 1$, yielding a complexity of order ε^{-4} . Our NSA algorithm (2.3), in contrast, has an optimal complexity of order ε^{-3} .

The next section is devoted to the MLSA scheme which combines multiple paired NSA estimators obtained through a geometric progression of bias parameters in order to reduce the complexity.

3 Multilevel Stochastic Approximation

We first explain the construction of the multilevel SA algorithm to approximate the VaR and ES, then we provide an analysis thereof from both convergence and complexity viewpoints.

Let $h_0 = \frac{1}{K} \in \mathcal{H}$ be a bias parameter and M and L be two positive integers, with $M \geq 2$. L is called the number of levels. Consider the geometric progression of bias parameters

$$h_\ell = \frac{h_0}{M^\ell} = \frac{1}{KM^\ell}, \quad \ell \in \{0, \dots, L\}.$$

Assuming the distribution function of X_{h_ℓ} continuous and increasing for any $\ell \in \{0, \dots, L\}$, we let $(\xi_\star^{h_\ell}, \chi_\star^{h_\ell}) = (\operatorname{argmin} V_{h_\ell}, \min V_{h_\ell})$, $\ell \in \{0, \dots, L\}$. We thus have the telescopic decompositions

$$\xi_\star^{h_L} = \xi_\star^{h_0} + \sum_{\ell=1}^L \xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}, \quad \chi_\star^{h_L} = \chi_\star^{h_0} + \sum_{\ell=1}^L \chi_\star^{h_\ell} - \chi_\star^{h_{\ell-1}}.$$

While the idea of NSA is to put the entirety of the available computational power behind simulating the quantities on the left hand sides above, MLSA rather simulates the telescopic decompositions on the right hand sides. It starts by simulating the level 0 terms $\xi_\star^{h_0}$ and $\chi_\star^{h_0}$ which are, although of high bias, very fast to estimate. It then simulates independent incremental error corrections $(\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}, \chi_\star^{h_\ell} - \chi_\star^{h_{\ell-1}})_{1 \leq \ell \leq L}$ to the initial simulations.

In the previous section, we saw how NSA can approximate each pair $(\xi_\star^{h_\ell}, \chi_\star^{h_\ell})$, $\ell \in \{0, \dots, L\}$. Let $\mathbf{N} = (N_0, \dots, N_L) \in \mathbb{N}^{L+1}$. Following Frikha [14], we define the multilevel SA estimators $\xi_{\mathbf{N}}^{\text{ML}}$ and $\chi_{\mathbf{N}}^{\text{ML}}$ of $\xi_\star^{h_L}$ and $\chi_\star^{h_L}$ by

$$\begin{cases} \xi_{\mathbf{N}}^{\text{ML}} = \xi_{N_0}^{h_0} + \sum_{\ell=1}^L \xi_{N_\ell}^{h_\ell} - \xi_{N_\ell}^{h_{\ell-1}}, \\ \chi_{\mathbf{N}}^{\text{ML}} = \chi_{N_0}^{h_0} + \sum_{\ell=1}^L \chi_{N_\ell}^{h_\ell} - \chi_{N_\ell}^{h_{\ell-1}}, \end{cases} \quad (3.1)$$

where, at any level $\ell \in \{0, \dots, L\}$, the initializations $(\xi_0^{h_\ell}, \chi_0^{h_\ell})$ are generated such that $\mathbb{E}[|\xi_0^{h_\ell}|^2] + \mathbb{E}[|\chi_0^{h_\ell}|^2] < \infty$, and the iterates $(\xi_n^{h_\ell}, \chi_n^{h_\ell})_{n \geq 1}$ are computed using the NSA scheme (2.3). We stress that for any fixed level $\ell \in \{1, \dots, L\}$, the random variables $(X_{h_{\ell-1}}^{(n)}, X_{h_\ell}^{(n)})_{1 \leq n \leq N_\ell}$ used to compute $(\xi_{N_\ell}^{h_{\ell-1}}, \xi_{N_\ell}^{h_\ell})$ and $(\chi_{N_\ell}^{h_{\ell-1}}, \chi_{N_\ell}^{h_\ell})$ are i.i.d. with the same law as $(X_{h_{\ell-1}}, X_{h_\ell})$, where

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

and X_{h_ℓ} is obtained from $X_{h_{\ell-1}}$ via

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

Algorithm 3 summarizes this process.

Remark 3.1. Intuitively, the larger the ℓ , the closer the random variables $X_{h_{\ell-1}}$ and X_{h_ℓ} are to X_0 and the lesser iterations N_ℓ are required at the level ℓ of the MLSA scheme (3.1) to achieve a high accuracy. Hence $N_0 \geq \dots \geq N_L$.

Algorithm 3 Multilevel SA for estimating (VaR, ES)

Require: A number of levels $L \geq 1$, a bias parameter $h_0 = \frac{1}{K} \in \mathcal{H}$, a geometric step size $M \geq 2$, positive integers N_0, \dots, N_L , a positive non-increasing sequence $(\gamma_n)_{n \geq 1}$.

```

1: for  $\ell = 0 \dots L$  do
2:   Set  $h_\ell \leftarrow \frac{h_0}{M^\ell}$ 
3:   for  $j = (\ell - 1)^+ \dots \ell$  do
4:     Choose  $(\xi_0^{h_j}, \chi_0^{h_j})$  such that  $\mathbb{E}[\xi_0^{h_j}|^2] + \mathbb{E}[\chi_0^{h_j}|^2] < \infty$ 
5:   end for
6:   for  $n = 0 \dots N_\ell - 1$  do
7:     Simulate  $Y^{(n+1)} \sim Y$ 
8:     Simulate  $Z^{(n+1,1)}, \dots, Z^{(n+1, KM^\ell)}$  i.i.d.  $Z$  independently from  $Y^{(n+1)}$ 
9:     if  $\ell = 0$  then
10:       $X_{h_0}^{(n+1)} \leftarrow \frac{1}{K} \sum_{k=1}^K \varphi(Y^{(n+1)}, Z^{(n+1,k)})$ 
11:     else
12:       $X_{h_{\ell-1}}^{(n+1)} \leftarrow \frac{1}{KM^{\ell-1}} \sum_{k=1}^{KM^{\ell-1}} \varphi(Y^{(n+1)}, Z^{(n+1,k)})$ 
13:       $X_{h_\ell}^{(n+1)} \leftarrow \frac{1}{M} X_{h_{\ell-1}}^{(n+1)} + \frac{1}{KM^\ell} \sum_{k=KM^{\ell-1}+1}^{KM^\ell} \varphi(Y^{(n+1)}, Z^{(n+1,k)})$ 
14:     end if
15:     for  $j = (\ell - 1)^+ \dots \ell$  do
16:        $\xi_{n+1}^{h_j} \leftarrow \xi_n^{h_j} - \gamma_{n+1} H_1(\xi_n^{h_j}, X_{h_j}^{(n+1)})$ 
17:        $\chi_{n+1}^{h_j} \leftarrow \chi_n^{h_j} - \frac{1}{n+1} H_2(\chi_n^{h_j}, \xi_n^{h_j}, X_{h_j}^{(n+1)})$ 
18:     end for
19:   end for
20: end for
21:  $\xi_{\mathbf{N}}^{\text{ML}} \leftarrow \xi_{N_0}^{h_0} + \sum_{\ell=1}^L \xi_{N_\ell}^{h_\ell} - \xi_{N_\ell}^{h_{\ell-1}}$ 
22:  $\chi_{\mathbf{N}}^{\text{ML}} \leftarrow \chi_{N_0}^{h_0} + \sum_{\ell=1}^L \chi_{N_\ell}^{h_\ell} - \chi_{N_\ell}^{h_{\ell-1}}$ 
23: return  $(\xi_{\mathbf{N}}^{\text{ML}}, \chi_{\mathbf{N}}^{\text{ML}})$ 

```

3.1 Convergence Analysis

The global error between the multilevel estimator $(\xi_{\mathbf{N}}^{\text{ML}}, \chi_{\mathbf{N}}^{\text{ML}})$ and its target $(\xi_\star^0, \chi_\star^0)$ can be decomposed into a sum of a statistical error and a bias error:

$$\xi_{\mathbf{N}}^{\text{ML}} - \xi_\star^0 = (\xi_{\mathbf{N}}^{\text{ML}} - \xi_\star^{h_L}) + (\xi_\star^{h_L} - \xi_\star^0), \quad (3.2)$$

$$\chi_{\mathbf{N}}^{\text{ML}} - \chi_\star^0 = (\chi_{\mathbf{N}}^{\text{ML}} - \chi_\star^{h_L}) + (\chi_\star^{h_L} - \chi_\star^0). \quad (3.3)$$

In the ensuing analysis, we quantify each error appearing in the decompositions above in terms of the parameters of (3.1). We then propose an optimal choice for L and N_0, \dots, N_L to achieve a given prescribed accuracy.

Proposition 3.2. (i) *Assume that the real-valued random variables X_h admit density functions f_{X_h} that are bounded uniformly in $h \in \overline{\mathcal{H}}$.*

a. *If*

$$\mathbb{E}[|\varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]|^{p_\star}] < \infty \quad \text{holds for some } p_\star > 1, \quad (3.4)$$

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

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

where

$$C := B_{p_\star} \mathbb{E}[|\varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]|^{p_\star}]^{\frac{1}{p_\star+1}} \left(\sup_{h \in \overline{\mathcal{H}}} \|f_{X_h}\|_\infty \right)^{\frac{p_\star}{p_\star+1}},$$

with B_{p_\star} a positive constant that depends only upon p_\star .

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

$$\mathbb{E} \left[\exp \left(u(\varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]) \right) \middle| Y \right] \leq e^{C_g u^2} \quad \mathbb{P}\text{-as.} \quad (3.5)$$

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

$$\begin{aligned} & \mathbb{E} [|\mathbf{1}_{\{X_h > \xi\}} - \mathbf{1}_{\{X_{h'} > \xi\}}|] \\ & \leq 2\sqrt{C_g(h' - h)} \left(1 + \sup_{h \in \overline{\mathcal{H}}} \|f_{X_h}\|_\infty \sqrt{2|\ln(C_g(h' - h))|} \right). \end{aligned}$$

(ii) Let $G_\ell := h_\ell^{-\frac{1}{2}}(X_{h_\ell} - X_{h_{\ell-1}})$, and define $F_{X_{h_{\ell-1}}|G_\ell=g}: x \mapsto \mathbb{P}[X_{h_{\ell-1}} \leq x | G_\ell = g]$, $g \in \text{supp}(\mathbb{P}_{G_\ell})$, $\ell \geq 1$. Consider the sequence of random variables $(K_\ell)_{\ell \geq 1}$ given by $K_\ell := K_\ell(G_\ell)$, where

$$K_\ell(g) := \sup_{x \neq y} \frac{|F_{X_{h_{\ell-1}}|G_\ell=g}(x) - F_{X_{h_{\ell-1}}|G_\ell=g}(y)|}{|x - y|}, \quad \ell \geq 1, g \in \text{supp}(\mathbb{P}_{G_\ell}).$$

Assume that $(K_\ell)_{\ell \geq 1}$ satisfies

$$\sup_{\ell \geq 1} \mathbb{E}[|G_\ell|K_\ell] < \infty. \quad (3.6)$$

Then

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

Proof. See Appendix G. □

Remark 3.3. The scenarios formulated in Proposition 3.2 are ordered by strength. The scenario (3.4) is already described in Barrera et al.'s [6, Assumption H7], although rather for $p_\star > 2$: our formulation also includes heavy tailed setups where $p_\star < 2$. A variant of the Gaussian concentration scenario (3.5) is stated in Barrera et al.'s [6, Assumption H6]. According to Frikha and Menozzi [17, Section 1], it suffices that $\mathbb{E}[\exp(u_0\varphi(Y, Z)^2)|Y]$ be bounded for some $u_0 > 0$ for this condition to hold. Observe that (3.4) follows from (3.5) for any $p_\star > 1$ via a power series expansion of the exponential. The scenario (3.6) is a weakening of Haji-Ali et al.'s [24, Assumption 2.5] (and Gordy and Juneja's [23, Assumption 1]). As shown in Section 3.2, it describes a setup that is computationally optimal. Note that if $(g, \ell) \mapsto F_{X_{h_{\ell-1}}|G_\ell=g}$ is uniformly Lipschitz in g and ℓ , then, by (G.4) with $h = h_\ell$ and $h' = h_{\ell-1}$, (3.6) holds.

Assumption 3.4. There exist $C, \delta > 0$ such that, for any $h \in \mathcal{H}$ and any compact set $\mathcal{K} \subset \mathbb{R}$,

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

Remark 3.5. Similarly to the second part of Assumption 2.2(ii), Assumption 3.4 is a weakened postulate in comparison with Giorgi et al.'s [22, Proposition 5.1(a)].

We state below our main result regarding the non-asymptotic square statistical error of Algorithm 3.

Theorem 3.6. Suppose that $\varphi(Y, Z) \in L^2(\mathbb{P})$ and that Assumptions 2.2, 2.5 and 3.4 hold. Then, within the framework of Proposition 3.2, if $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0, 1]$, with $\bar{\lambda}_2 \gamma_1 > 2$ if $\beta = 1$, and

$$\sup_{h \in \overline{\mathcal{H}}} \mathbb{E} \left[|\xi_0^h|^4 \exp \left(\frac{16}{1 - \alpha} k_\alpha \sup_{h \in \overline{\mathcal{H}}} \|f_{X_h}\|_\infty |\xi_0^h| \right) \right] + \sup_{h \in \overline{\mathcal{H}}} \mathbb{E}[(\chi_0^h)^2] < \infty,$$

then there exists some constant $K < \infty$ such that for any positive integer L and any $\mathbf{N} = (N_0, \dots, N_L) \in \mathbb{N}^{L+1}$,

$$\mathbb{E}[(\xi_{\mathbf{N}}^{\text{ML}} - \xi_{\star}^{h_L})^2] \leq K \left(\gamma_{N_0} + \left(\sum_{\ell=1}^L \gamma_{N_\ell} \right)^2 + \sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{3}{2}} + \sum_{\ell=1}^L \gamma_{N_\ell} \epsilon(h_\ell) \right), \quad (3.7)$$

and

$$\begin{aligned} & \mathbb{E}[(\chi_{\mathbf{N}}^{\text{ML}} - \chi_{\star}^{h_L})^2] \\ & \leq K \left(\frac{1}{N_0^{1 \wedge 2\beta}} + \sum_{\ell=1}^L \frac{h_\ell}{N_\ell} + \sum_{\ell=1}^L \frac{\bar{\gamma}_{N_\ell}}{N_\ell} \epsilon(h_\ell) + \sum_{\ell=1}^L \frac{1}{N_\ell^2} \sum_{k=1}^{N_\ell} \gamma_k^{\frac{3}{2}} + \left(\sum_{\ell=1}^L \bar{\gamma}_{N_\ell} \right)^2 \right), \end{aligned} \quad (3.8)$$

where $\bar{\gamma}_n = \frac{1}{n} \sum_{k=1}^n \gamma_k$ and

$$\epsilon(h) := \begin{cases} h^{\frac{p^*}{2(1+p^*)}} & \text{if (3.4) is satisfied,} \\ h^{\frac{1}{2}} |\ln h|^{\frac{1}{2}} & \text{if (3.5) is satisfied,} \\ h^{\frac{1}{2}} & \text{if (3.6) is satisfied,} \end{cases} \quad h \in \mathcal{H}. \quad (3.9)$$

Proof. See Appendix H. □

3.2 Complexity Analysis

Considering the error decompositions (3.2) and (3.3), for $h_0 \in \mathcal{H}$ and $M \geq 2$, we rely on the result above to find the optimal number of levels $L \geq 1$ and number of steps $N_\ell \geq 1$ at each level $\ell \in \{0, \dots, L\}$ of the MLSA scheme (3.1) in order to achieve a prescribed accuracy $\varepsilon \in (0, 1)$.

Proposition 3.7. *Suppose Assumption 2.5(i) satisfied. Let $\varepsilon \in (0, 1)$ be some prescribed accuracy. If $h_0 > \varepsilon$, then setting*

$$L = \left\lceil \frac{\ln h_0 \varepsilon^{-1}}{\ln M} \right\rceil$$

achieves a bias error on the estimation of $(\xi_{\star}^0, \chi_{\star}^0)$ of order ε .

Proof. In view of Proposition 2.4, the bias error of the couple of estimators $(\xi_{\mathbf{N}}^{\text{ML}}, \chi_{\mathbf{N}}^{\text{ML}})$ is of order h_L . We thus select $L \geq 1$ such that $h_L = \frac{h_0}{M^L} \leq \varepsilon$. □

Lemma 3.8. *The computational cost of MLSA satisfies*

$$\text{Cost}_{\text{MLSA}} \leq C \sum_{\ell=0}^L \frac{N_\ell}{h_\ell}.$$

Proof. This follows straight from the scheme (3.1). □

Theorem 3.9. *Let $\varepsilon \in (0, 1)$ be some prescribed accuracy. Within the framework of Theorem 3.6,*

(i) (VaR focused parametrization) *setting*

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

i.e. $N_\ell =$

$$\begin{cases} \left\lceil (K\gamma_1)^{\frac{1}{\beta}} \varepsilon^{-\frac{2}{\beta}} h_\ell^{\frac{1}{1+\beta}(1+\frac{p^*}{2(1+p^*)})} \left(\sum_{\ell'=0}^L h_{\ell'}^{\frac{1}{1+\beta}(-\beta+\frac{p^*}{2(1+p^*)})} \right)^{\frac{1}{\beta}} \right\rceil, & \text{case (3.4),} \\ \left\lceil (K\gamma_1)^{\frac{1}{\beta}} \varepsilon^{-\frac{2}{\beta}} h_\ell^{\frac{3}{2(1+\beta)}} |\ln h_\ell|^{\frac{1}{2(1+\beta)}} \left(\sum_{\ell'=0}^L h_{\ell'}^{\frac{1-2\beta}{2(1+\beta)}} |\ln h_{\ell'}|^{\frac{1}{2(1+\beta)}} \right)^{\frac{1}{\beta}} \right\rceil, & \text{case (3.5),} \\ \left\lceil (K\gamma_1)^{\frac{1}{\beta}} \varepsilon^{-\frac{2}{\beta}} h_\ell^{\frac{3}{2(1+\beta)}} \left(\sum_{\ell'=0}^L h_{\ell'}^{\frac{1-2\beta}{2(1+\beta)}} \right)^{\frac{1}{\beta}} \right\rceil, & \text{case (3.6),} \end{cases}$$

where K is the constant appearing in (3.7), then MLSA achieves a statistical error on the estimation of ξ_\star^0 of order ε . The optimal computational cost of MLSA is achieved when $\beta = 1$, under the constraint $\bar{\lambda}_2\gamma_1 > 2$, in which case

$$\text{Cost}_{\text{MLSA}}^{\text{VaR}} \leq C \begin{cases} \varepsilon^{-3+\frac{p^\star}{2(1+p^\star)}} & \text{if (3.4) is satisfied,} \\ \varepsilon^{-\frac{5}{2}} |\ln \varepsilon|^{\frac{1}{2}} & \text{if (3.5) is satisfied,} \\ \varepsilon^{-\frac{5}{2}} & \text{if (3.6) is satisfied.} \end{cases}$$

(ii) (ES focused parametrization) under the constraint $\bar{\lambda}_2\gamma_1 > 2$, setting $\beta = 1$ and

$$N_\ell = \lceil K\varepsilon^{-2}Lh_\ell \rceil = \left\lceil K\varepsilon^{-2} \left\lceil \frac{\ln h_0\varepsilon^{-1}}{\ln M} \right\rceil h_\ell \right\rceil, \quad 0 \leq \ell \leq L,$$

where K is the constant appearing in (3.8), then MLSA achieves a statistical error on the estimation of χ_\star^0 of order ε . The corresponding computational cost satisfies

$$\text{Cost}_{\text{MLSA}}^{\text{ES}} \leq C\varepsilon^{-2} |\ln \varepsilon|^2.$$

Proof. (i) Similarly to the multilevel Monte Carlo algorithm for the computation of the probability $\mathbb{P}[X_0 > 0]$ by Giles and Haji-Ali [19], it is expected that the leading term in the global L^2 error (3.7) 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

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

where K is the constant appearing in (3.7).

We can easily check that, with the above choices of L and N_0, \dots, N_L , the first terms in (3.7), namely $\gamma_{N_0} + (\sum_{\ell=1}^L \gamma_{N_\ell})^2 + \sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{3}{2}}$, are in $\mathcal{O}(\varepsilon^2)$ as $\varepsilon \downarrow 0$. The related complexity computations are standard and are thereby omitted.

(ii) To parametrize the iterations amounts N_0, \dots, N_L optimally to compute the ES, we minimize the complexity of MLSA while constraining the global L^2 error of the multilevel ES estimator (3.8) to an order of ε^2 . We presume that the leading term of the upper estimate (3.8) is $\sum_{\ell=1}^L \frac{h_\ell}{N_\ell}$, which we check a posteriori. Hence, we solve

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

Under this parametrization, we check that $(\sum_{\ell=0}^L \bar{\gamma}_{N_\ell})^2 = \mathcal{O}(\varepsilon^{2\beta} |\ln \varepsilon|^{2(\mathbb{1}_{\{\beta=1\}} - \beta)})$ as $\varepsilon \downarrow 0$. 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 (3.8), 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 . \square

For the VaR estimation, in the best case scenario where (3.6) is satisfied, for a prescribed accuracy ε , it is possible to estimate the VaR with a complexity in $\mathcal{O}(\varepsilon^{-\frac{5}{2}})$ using the multilevel stochastic approximation approach. This is an order of magnitude lower than the optimal complexity in $\mathcal{O}(\varepsilon^{-3})$ of the nested stochastic approximation approach obtained in Theorem 2.8.

As for the ES estimation, note that the resulting complexity coincides exactly with the optimal complexity of the standard multilevel Monte Carlo algorithm derived by Giles [18].

4 Financial Case Studies

We now assess the convergence results of NSA and MLSA numerically. This is done through stylized financial setups, where the probability measure \mathbb{P} in the paper is used for all pricing and risk computations. For comparative and assessment purposes, the numerical settings we are interested in allow to retrieve the VaR and ES either analytically or through Algorithms 1, 2 or 3. Recall that real life scenarios nonetheless lack analytical formulas and prohibit the recourse to Algorithm 1. The code producing the results exhibited in the next sections is available at github.com/azarlouzi/mlsa.

4.1 European Option

We endorse the stylized financial setup of Giles and Haji-Ali [19, Section 3] and consider 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 given by

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

Let $\delta \in (0, 1)$ be a time horizon. We define the loss X_0 of the option by

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

Let $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$,

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

Then

$$X_0 \stackrel{\mathcal{L}}{=} \mathbb{E}[\varphi(Y, Z)] - \mathbb{E}[\varphi(Y, Z) | Y] = -1 - \mathbb{E}[\varphi(Y, Z) | Y] = \delta(Y^2 - 1), \quad (4.1)$$

where Y and Z are independent and of law $\mathcal{N}(0, 1)$.

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

$$1 - \alpha = \mathbb{P}[X_0 > \xi_\star^0] = \mathbb{P}\left[Y^2 > 1 + \frac{\xi_\star^0}{\delta}\right] = 2F\left(-\left(1 + \frac{\xi_\star^0}{\delta}\right)^{\frac{1}{2}}\right),$$

where F is the standard Gaussian distribution function. Hence

$$\xi_\star^0 = \delta \left\{ \left(F^{-1}\left(\frac{1-\alpha}{2}\right) \right)^2 - 1 \right\}. \quad (4.2)$$

We can also get an analytical formula for the ES χ_\star^0 at level α . Indeed, using the symmetry of the Gaussian distribution,

$$\chi_\star^0 = \mathbb{E}[X_0 | X_0 > \xi_\star^0] = \frac{\delta}{1-\alpha} \{2\mathbb{E}[Y^2 \mathbf{1}_{\{Y > \mu\}}] - (1-\alpha)\}, \quad \text{where } \mu := \left(1 + \frac{\xi_\star^0}{\delta}\right)^{\frac{1}{2}}.$$

Integrating by parts,

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

where f denotes the standard Gaussian density function. Hence

$$\chi_\star^0 = \frac{2\delta}{1-\alpha} \left(\mu f(\mu) + F(-\mu) - \frac{1-\alpha}{2} \right). \quad (4.3)$$

Stochastic Approximation. The loss $X_0 \stackrel{\mathcal{L}}{=} \delta(Y^2 - 1)$ (cf (4.1)) can be simulated directly, so that Algorithm 1 is applicable. But the goal here is to estimate the values of ξ_\star^0 and χ_\star^0 using Algorithms 2 and 3, and to study empirically the influence of the parametrization of each algorithm on their performances. The exact values obtained with (4.2) and (4.3) serve as convergence benchmarks.

The loss $X_0 \stackrel{\mathcal{L}}{=} -1 - \mathbb{E}[\varphi(Y, Z)|Y]$ can alternatively 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 $Y, Z^{(1)}, \dots, Z^{(K)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. We can then apply Algorithm 2 or 3 on this basis.

Numerical Results. In the following applications, the confidence level is taken as $\alpha = 97.5\%$ and the time horizon is set to $\delta = 0.5$. This setup yields $\xi_\star^0 \approx 2.012$ and $\chi_\star^0 \approx 2.901$ (computed using the explicit formulas (4.2) and (4.3)). The value of $\beta = 1$ is used for the step sequences $(\gamma_n)_{n \geq 1}$ since, according to Theorems 2.8 and 3.9, it leads to the optimal complexities for Algorithms 2 and 3.

Weak error expansion. To showcase the results of Proposition 2.4, we study the linearity of the errors $\xi_\star^h - \xi_\star^0$ and $\chi_\star^h - \chi_\star^0$ and the stability of the renormalized errors $h^{-1}(\xi_\star^h - \xi_\star^0)$ and $h^{-1}(\chi_\star^h - \chi_\star^0)$ as $\mathcal{H} \ni h \downarrow 0$. We run Algorithm 2 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 adopt the step size $\gamma_n = 0.1/(10^4 + n)$, with a smoothing applied to the denominator to avoid instability at early iterations. We average out 200 outcomes $(\xi_{10^6}^h, \chi_{10^6}^h)$ to form an estimate $(\bar{\xi}_{10^6}^h, \bar{\chi}_{10^6}^h)$ of $(\xi_\star^h, \chi_\star^h)$. The left panel of Figure 1 shows the evolution of the quantities $\bar{\xi}_{10^6}^h - \xi_\star^0$ and $\bar{\chi}_{10^6}^h - \chi_\star^0$ as h describes $\{\frac{1}{10}, \frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{200}\}$, while the right panel shows the evolution of the corresponding quantities $h^{-1}(\bar{\xi}_{10^6}^h - \xi_\star^0)$ and $h^{-1}(\bar{\chi}_{10^6}^h - \chi_\star^0)$.

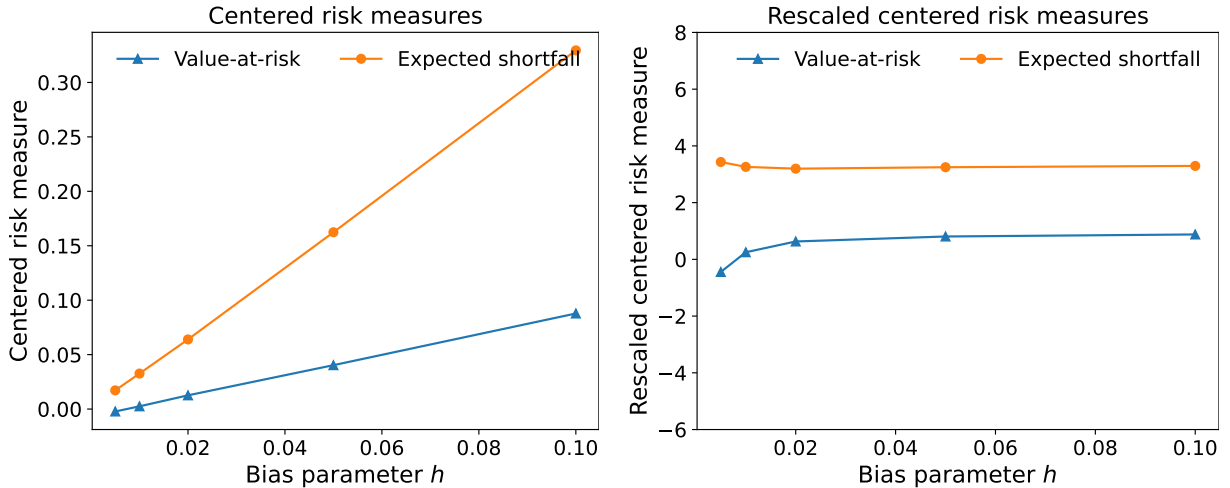


Figure 1. Centered and rescaled risk measures for a bias parameter h tending to 0

The left panel plot suggests that, asymptotically as $\mathcal{H} \ni h \downarrow 0$, the quantities $\xi_\star^h - \xi_\star^0$ and $\chi_\star^h - \chi_\star^0$ are linear in h . The right panel plot strengthens this observation, demonstrating that the quantities $h^{-1}(\xi_\star^h - \xi_\star^0)$ and $h^{-1}(\chi_\star^h - \chi_\star^0)$ are approximately constant in a neighborhood of 0. This checks empirically the validity of Proposition 2.4 on the weak error expansion.

Comparative study of the complexity. We aim here to compare the performances of Algorithms 1, 2 and 3.

For Algorithms 1 and 2, we set the step size $\gamma_n = 1/(100 + n)$ for the VaR simulation and $\gamma_n = 0.1/(2.5 \times 10^4 + n)$ for the ES simulation. We work under the scenario (3.4) for Algorithm 3, that clearly holds for any $p_\star > 1$, with the particular value $p_\star = 11$. We choose $M = 2$. The VaR and ES are computed using their respective optimal iterations amounts described in Theorem 3.9. The remaining hyperparameter setup is provided in Table 1. The $(\gamma_n)_{n \geq 1}$ given therein for the VaR estimation were obtained through a full grid search on their parameters.

For each algorithm and each risk measure, we plot on a logarithmic scale the average run times over 200 runs against the achieved RMSEs, for a prescribed accuracy ε that loops through $\{\frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$. We also plot the average run times against the prescribed accuracies ε themselves. See Figure 2. Table 2 reports the slopes fitted on these curves.

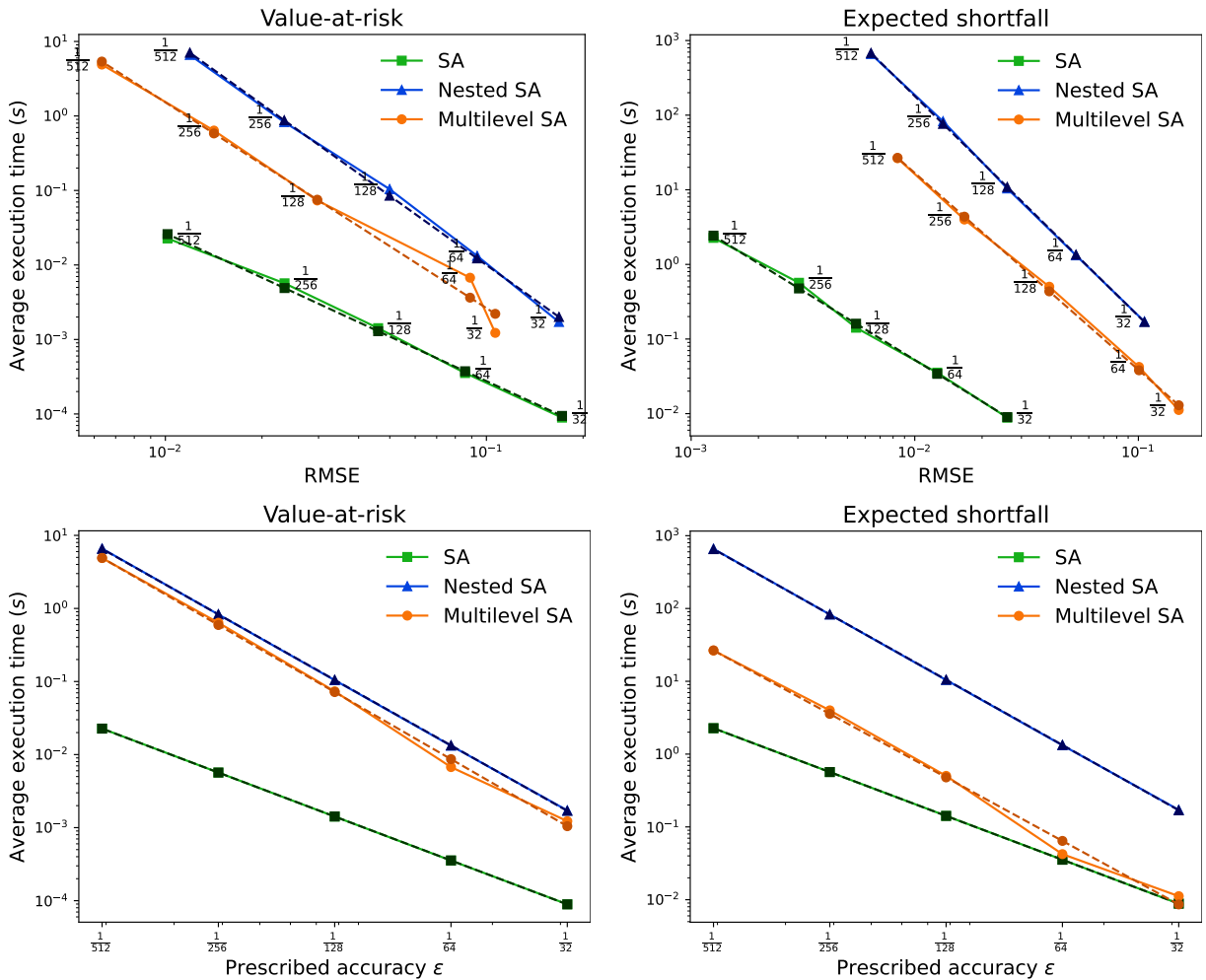


Figure 2. Performance comparison of Algorithms 1, 2 and 3

Prescribed accuracy ε	VaR estimation			ES estimation		
	h_0	L	γ_n	h_0	L	γ_n
$\frac{1}{32}$	$\frac{1}{16}$	1	$\frac{2}{2.5 \times 10^3 + n}$	$\frac{1}{16}$	1	$\frac{0.1}{10^4 + n}$
$\frac{1}{64}$	$\frac{1}{32}$	1	$\frac{2}{4 \times 10^3 + n}$	$\frac{1}{32}$	1	$\frac{0.1}{10^4 + n}$
$\frac{1}{128}$	$\frac{1}{32}$	2	$\frac{0.75}{9 \times 10^3 + n}$	$\frac{1}{32}$	2	$\frac{0.1}{10^4 + n}$
$\frac{1}{256}$	$\frac{1}{32}$	3	$\frac{0.25}{10^4 + n}$	$\frac{1}{32}$	3	$\frac{0.1}{2 \times 10^4 + n}$
$\frac{1}{512}$	$\frac{1}{32}$	4	$\frac{0.09}{10^4 + n}$	$\frac{1}{32}$	4	$\frac{0.1}{2.5 \times 10^4 + n}$

Table 1. Algorithm 3 parametrization by prescribed accuracy

SA scheme	VaR estimation		ES estimation	
	RMSE	Accuracy	RMSE	Accuracy
Nested SA (Alg. 2)	-3.08	-2.98	-2.95	-2.98
Multilevel SA (Alg. 3)	-2.76	-3.05	-2.64	-2.90
SA (Alg. 1)	-1.99	-2.00	-1.85	-2.00

Table 2. Reported slopes in Figure 2

For the VaR as for the ES, for any given scored RMSE, Algorithm 3 displays an execution time that is orders of magnitude lower than Algorithm 2. Take an RMSE of order 5×10^{-2} . MLSA computes the VaR in $10^{-2}s$, scoring a 10 fold speedup over NSA which computes it in $10^{-1}s$. Likewise, to compute the ES, MLSA requires $2 \times 10^{-1}s$, a 10 fold speedup over NSA which requires $2s$. This corroborates the theoretical optimal complexity of Algorithm 3 that is significantly lower than that of Algorithm 2 (c.f. Theorems 2.8 and 3.9). Finally, as one would expect, Algorithm 1 that simulates X_0 directly considerably outperforms both Algorithms 2 and 3. The spacing between the SA and NSA curves illustrates the performance gap occurring when using, in a Robbins-Monro instance, a nested simulation instead of a direct simulation for the loss. The placement of the MLSA curve in between these two demonstrates the partial yet significant regain on this performance gap, which is achieved by shifting from the nested to the multilevel scheme. Note however that the multilevel VaR estimation turned out rather unstable numerically. Small changes to the step size parametrization may fairly affect the results. To counter this, a thorough grid search on the parameters of the step size was performed, adding a non negligible tuning phase to the total execution time. The multilevel ES estimation, by contrast, is quite robust and requires only minor hyperparameter tuning. Up to a variation range, the slopes of the RMSE curves in Table 2 reflect the theoretical exponents in the complexity theorems. The curves of the average execution time as a function of the prescribed accuracy feature a log-linear behavior, which slopes in Table 2 roughly match the theoretical complexity exponents in Theorems 2.8 and 3.9.

4.2 Swap on a Rate

We adapt the setup of Albanese et al.'s [2, Section A.1]. Consider a stylized swap of nominal \bar{N} , strike \bar{S} and maturity \bar{T} on some underlying (FX or interest) rate. This rate follows a Black-Scholes model $(S_t)_{t \geq 0}$ of risk neutral drift $\bar{\kappa}$ and constant volatility $\bar{\sigma}$, or in other words, $\{\widehat{S}_t := e^{-\bar{\kappa}t} S_t, t \geq 0\}$ follows a Black martingale model of volatility $\bar{\sigma}$. Given coupon dates $0 < T_1 < \dots < T_d = \bar{T}$, the swap pays at T_i the cash flow $\Delta_i(S_{T_{i-1}} - \bar{S})$, 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 denote $i_t \in \{1, \dots, d+1\}$ 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

$$\begin{aligned} P_t &= \bar{N} \mathbb{E} \left[\rho_t^{-1} \rho_{T_{i_t}} \Delta_{i_t} (S_{T_{i_t-1}} - \bar{S}) + \sum_{i=i_t+1}^d \rho_t^{-1} \rho_{T_i} \Delta_i (S_{T_{i-1}} - \bar{S}) \middle| S_t \right] \\ &= \bar{N} \mathbb{E} \left[\rho_t^{-1} \rho_{T_{i_t}} \Delta_{i_t} (e^{\bar{\kappa} T_{i_t-1}} \widehat{S}_{T_{i_t-1}} - \bar{S}) + \sum_{i=i_t+1}^d \rho_t^{-1} \rho_{T_i} \Delta_i (e^{\bar{\kappa} T_{i-1}} \widehat{S}_{T_{i-1}} - \bar{S}) \middle| \widehat{S}_t \right] \end{aligned} \quad (4.4)$$

$$= \bar{N} \left(\rho_t^{-1} \rho_{T_{i_t}} \Delta_{i_t} (e^{\bar{\kappa} T_{i_t-1}} \widehat{S}_{T_{i_t-1}} - \bar{S}) + \sum_{i=i_t+1}^d \rho_t^{-1} \rho_{T_i} \Delta_i (e^{\bar{\kappa} T_{i-1}} \widehat{S}_t - \bar{S}) \right). \quad (4.5)$$

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

$$\bar{S} = \frac{\sum_{i=1}^d \rho_{T_i} \Delta_i e^{\bar{\kappa} T_{i-1}}}{\sum_{i=1}^d \rho_{T_i} \Delta_i} S_0. \quad (4.6)$$

Given some confidence level $\alpha \in (0, 1)$, we are interested in computing the risk measures $\xi_\star^0 := \text{VaR}_\alpha(\rho_\delta P_\delta)$ and $\chi_\star^0 := \mathbb{E} S_\alpha(\rho_\delta P_\delta)$ of a short position on the swap at some time horizon $\delta < \min_i \Delta_i$.

Analytical Formulas. In this lognormal setup, the values of ξ_\star^0 and χ_\star^0 can be obtained analytically.

On the one hand, observing that $i_\delta = 1$ and using (4.5) and (4.6),

$$\rho_\delta P_\delta = \bar{N} A (\widehat{S}_\delta - S_0), \quad \text{where } A := \sum_{i=2}^d \rho_{T_i} \Delta_i e^{\bar{\kappa} T_{i-1}}. \quad (4.7)$$

Hence

$$1 - \alpha = \mathbb{P}[\rho_\delta P_\delta > \xi_\star^0] = \mathbb{P} \left[\widehat{S}_\delta > S_0 + \frac{\xi_\star^0}{\bar{N} A} \right],$$

thus

$$\xi_\star^0 = \bar{N} A S_0 \left(\exp \left(F^{-1}(\alpha) \bar{\sigma} \sqrt{\delta} - \frac{\bar{\sigma}^2}{2} \delta \right) - 1 \right), \quad (4.8)$$

where F denotes the standard Gaussian distribution function.

On the other hand, setting

$$\omega := S_0 + \frac{\xi_\star^0}{\bar{N} A}, \quad \eta_\pm := \frac{1}{\bar{\sigma} \sqrt{\delta}} \left(\ln \frac{\omega}{S_0} \pm \frac{\bar{\sigma}^2 \delta}{2} \right),$$

we obtain

$$\chi_\star^0 = \mathbb{E}[\rho_\delta P_\delta | \rho_\delta P_\delta > \xi_\star^0] = \frac{\bar{N} A}{1 - \alpha} (\mathbb{E}[\widehat{S}_\delta \mathbf{1}_{\{\widehat{S}_\delta > \omega\}}] - (1 - \alpha) S_0),$$

hence

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

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

$$\chi_\star^0 = \bar{N} A S_0 \frac{\alpha - F(\eta_-)}{1 - \alpha}. \quad (4.9)$$

Stochastic Approximation. Define $X_0 := \rho_\delta P_\delta$.

Algorithm 1 can be applied to approximate the values of ξ_\star^0 and χ_\star^0 . Indeed, according to (4.7),

$$X_0 \stackrel{\mathcal{L}}{=} \bar{N} A S_0 \left(\exp \left(-\frac{\bar{\sigma}^2}{2} \delta + \bar{\sigma} \sqrt{\delta} U \right) - 1 \right),$$

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

Algorithms 2 and 3 are also applicable to approximate the values of ξ_\star^0 and χ_\star^0 . By (4.4) and (4.6),

$$X_0 = \bar{N} \mathbb{E} \left[\sum_{i=2}^d \rho_{T_i} \Delta_i e^{\bar{\kappa} T_{i-1}} (\widehat{S}_{T_{i-1}} - S_0) \middle| \widehat{S}_\delta \right] \stackrel{\mathcal{L}}{=} \mathbb{E}[\varphi(Y, Z) | Y], \quad (4.10)$$

where $Y \in \mathbb{R}$ and $Z = (Z_1, \dots, Z_{d-1}) \in \mathbb{R}^{d-1}$ are independent, with

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

and $(U_i)_{0 \leq i \leq d-1} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. For any bias parameter $h = \frac{1}{K} \in \mathcal{H}$, X_0 can be approximated using (4.10).

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

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

Given the above parameter set, the theoretical values of the VaR and ES obtained with (4.8) and (4.9) are $\xi_\star^0 \approx 219.64$ and $\chi_\star^0 \approx 333.91$.

Comparative study of the complexity. For the stochastic approximation of these quantities, we use the step sizes $\gamma_n = 100/n$, for the SA scheme and $\gamma_n = 50/n$ for the nested SA scheme. We run these schemes with their respective optimal iterations amounts. As for the multilevel SA scheme, we work under the scenario (3.4) with $p_\star = 4$, which is tuned using a grid search. The VaR and ES are simulated using their respective optimal amounts of iterations N_0, \dots, N_L . For each prescribed accuracy $\varepsilon \in \{\frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$, we perform a grid search to tune the initial bias parameter h_0 (giving the number of levels L) and the step sizes $(\gamma_n)_{n \geq 1}$. Table 3 lists the ensuing parametrizations by prescribed accuracy. Figure 3 plots, on a logarithmic scale, the joint evolution of the realized RMSE and the average execution time over 200 runs of each SA scheme, for an accuracy ε varying in $\{\frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$. The second row of Figure 3 displays the realized RMSE as a function of the prescribed accuracy. Table 3 reports the regressed slopes on these curves as depicted in dashed lines on Figure 3.

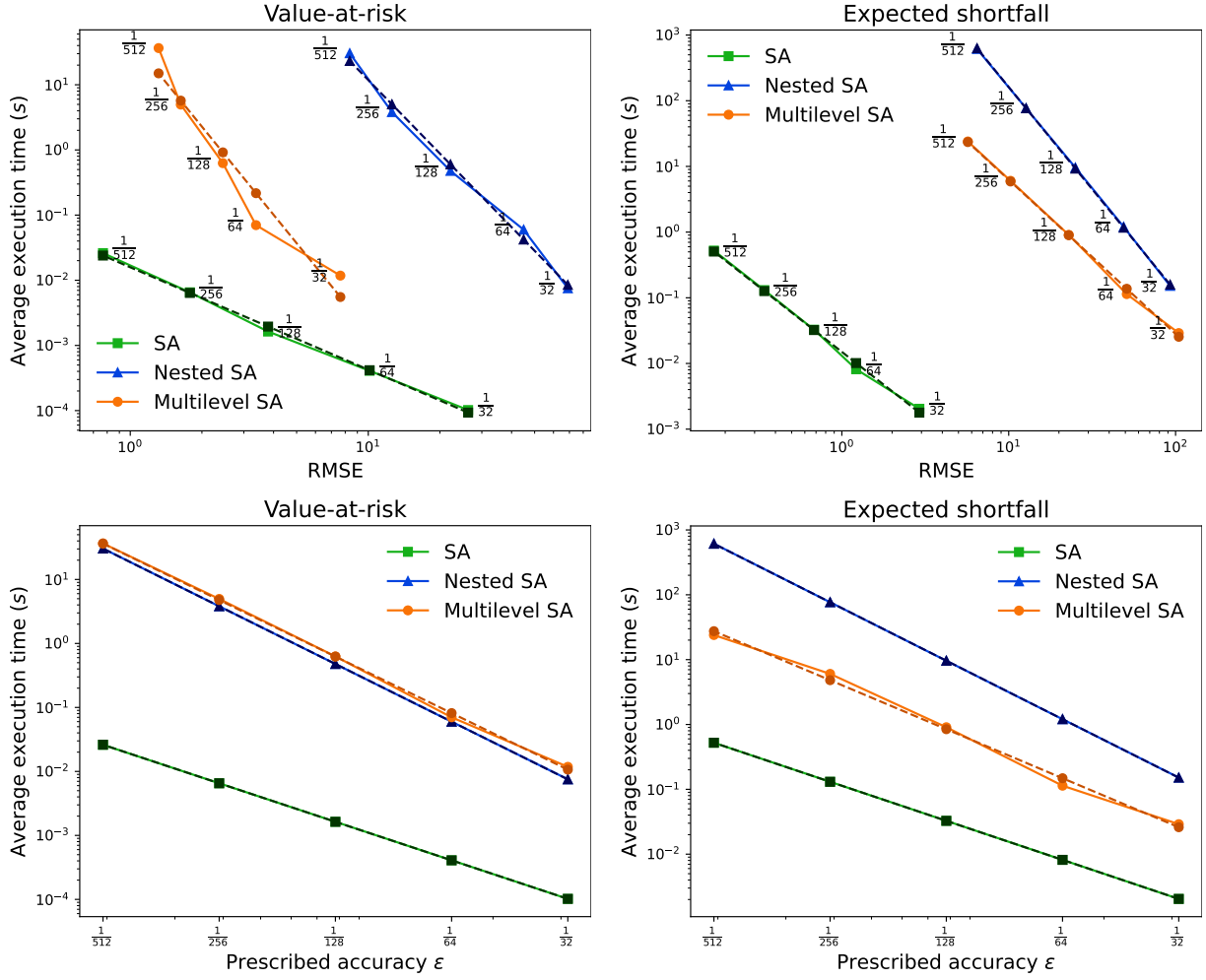


Figure 3. Performance comparison of Algorithms 1, 2 and 3

Prescribed accuracy ε	VaR estimation			ES estimation		
	h_0	L	γ_n	h_0	L	γ_n
$\frac{1}{32}$	$\frac{1}{8}$	2	$\frac{6}{10+n}$	$\frac{1}{8}$	2	$\frac{5}{10+n}$
$\frac{1}{64}$	$\frac{1}{16}$	2	$\frac{20}{500+n}$	$\frac{1}{16}$	2	$\frac{20}{500+n}$
$\frac{1}{128}$	$\frac{1}{16}$	3	$\frac{21}{10^3+n}$	$\frac{1}{16}$	3	$\frac{20}{500+n}$
$\frac{1}{256}$	$\frac{1}{16}$	4	$\frac{20}{2 \times 10^3+n}$	$\frac{1}{16}$	4	$\frac{20}{750+n}$
$\frac{1}{512}$	$\frac{1}{16}$	5	$\frac{21}{3 \times 10^3+n}$	$\frac{1}{32}$	4	$\frac{50}{2 \times 10^3+n}$

Table 3. Algorithm 3 parametrization by prescribed accuracy

SA scheme	VaR estimation		ES estimation	
	RMSE	Accuracy	RMSE	Accuracy
Nested SA (Alg. 2)	-3.75	-3.00	-3.11	-3.00
Multilevel SA (Alg. 3)	-4.50	-2.93	-2.35	-2.51
SA (Alg. 1)	-1.57	-2.00	-1.99	-2.00

Table 4. Reported slopes in Figure 3

We observe in the top panels of Figure 3 that the multilevel SA scheme significantly outperforms the nested SA scheme in terms of computational time and achieved error rate. Indeed, for a target RMSE of order 10 for the VaR, MLSA is 10^3 times faster than NSA as the former executes on average in approximately $10^{-2}s$ while the latter executes in $10s$. Similarly, for an RMSE of order 10 for the ES, MLSA is 10 times faster than NSA as the former executes on average in $10s$ while the latter executes in 10^2s . These results are very promising, as they apply to a more realistic scenario than the previous section. Eventually, the standard Robbins-Monro algorithm, that is only applicable if the loss X_0 is directly simulatable, outperforms both Algorithms 2 and 3 as expected. As indicated in the previous case study, the in-between positioning of the MLSA curve relatively to the two other ones depicts the partial regain by the multilevel scheme of the performance lost due to the nested simulation of X_0 . Once again, we point out that MLSA exhibits some numerical instability when simulating the VaR, while it remains very robust when simulating the ES. The VaR multilevel scheme is non-linearly sensitive to the parameterization of the step sizes $(\gamma_n)_{n \geq 1}$. In contrast, the ES multilevel scheme is significantly more stable with respect to potential mistuning of the step sizes $(\gamma_n)_{n \geq 1}$.

Within a variation range of order ± 0.2 , the RMSE slopes reported in Table 4 roughly match the exponents of the complexities in Theorems 2.8 and 3.9. The VaR focused MLSA slope may be improved with deeper hyperparameter tuning. This costly hyperparameter tuning phase is in fact an effect of the constraint $\bar{\lambda}_2 \gamma_1 > 2$ on the optimal choice $\beta = 1$ as stressed in the VaR focused theorem 3.9(i). Although $\bar{\lambda}_2$ is explicit in (2.4), it is tedious to compute, hence the necessary upstream tuning phase.

The bottom panels of Figure 3 display a quasi-log-linear dependency of the average execution time with respect to the prescribed accuracy. The corresponding slopes in Table 4 match quite accurately the theoretical exponents from the complexity Theorems 2.8 and 3.9.

Conclusion

In this work, we presented a stochastic approximation instance which solution retrieves the VaR and ES of the loss of a financial derivatives portfolio. This loss, however, can only be simulated in a nested Monte Carlo fashion. A naively nested algorithm performs one inner simulation layer within each outer update of a stochastic approximation scheme. A single (bias) parameter allows to control simultaneously the bias, the iterations amount and the complexity of the nested algorithm. But, considering a prescribed accuracy ε , this algorithm can only optimally achieve a complexity of order ε^{-3} . Our multilevel algorithm combines multiple paired estimates obtained by a geometric progression of biased nested schemes, in a manner that reduces the overall simulation complexity.

According to user intent, whether it is to simulate the VaR or the ES, our theoretical study identifies two corresponding optimal parameterizations. The VaR focused parameter setup results optimally in a complexity of order $\varepsilon^{-2-\delta}$, where $\delta \in (0, 1)$ is a specific parameter that depends on the integrability degree of the loss, while the ES focused parameter setup yields a complexity of order $\varepsilon^{-2} |\ln \varepsilon|^2$. These complexities are attained for a step size that is propor-

tional to $\frac{1}{n}$, albeit at the cost of a non trivial constraint on the proportionality factor. The empirical studies highlight the clear overperformance of the multilevel algorithm over the nested one. They demonstrate how, in situations where the inner simulations cannot be shortcutted by explicit formulas, utilizing the multilevel scheme results in a partial yet significant regain on the performance gap between the nested SA scheme and the classical (non-nested) Robbins-Monro scheme. This finding is of great interest when the latter scheme does not apply. The numerics have also underlined some degree of instability of MLSA when estimating the VaR, while they highlighted high robustness when estimating the ES. The former behavior is in line with the aforementioned proportionality factor constraint on the step sizes. As the ES is increasingly replacing the VaR as a fundamental financial risk measure, our ES focused multilevel algorithm, running in $O(\varepsilon^{-2}|\ln \varepsilon|^2)$ for a prescribed accuracy ε , should be of practical value for risk management.

As a followup to this paper, we have looked into complementing our non asymptotic error controls with asymptotic error distributions via central limit theorems [11]. We have also studied therein the Polyak-Ruppert variants of the NSA and MLSA methods as a means to rid the computationally optimal case of the non trivial constraint on the learning rate. Antithetic multilevel schemes are known to reduce the variances of the paired estimators within each independent level of MLSA [20, 21, 8, 26]. This incentivizes future research in this direction. A natural 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. A possible line of research in this direction would be to explore the added benefit of an adaptive selection of inner samples at each level of MLSA as proposed by Giles and Haji-Ali [19]. Eventually, in order to build stronger estimators, one could investigate other aggregation methods such as the multi-step Richardson-Romberg extrapolation of Frikha and Huang [16].

A Appendix

B Proof of Lemma 2.1

Step 1. Accumulation points of $(\xi_\star^h)_{h \in \mathcal{H}}$.

Let $h \in \mathcal{H}$. By Lemma 1.1, the set Θ_h is a closed non-empty bounded interval that coincides with the set of roots of V_h' .

Since $(X_h)_{h \in \mathcal{H}}$ converges in distribution to X_0 as $\mathcal{H} \ni h \downarrow 0$, the sequence of distribution 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 distribution function 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 $(V_h')_{h \in \overline{\mathcal{H}}}$, it follows that $(V_h')_{h \in \mathcal{H}}$ converges uniformly on \mathbb{R} towards V_0' . For any $\xi_\star^h \in \Theta_h$, $h \in \mathcal{H}$, one has

$$|V_0'(\xi_\star^h)| = |V_0'(\xi_\star^h) - V_h'(\xi_\star^h)| \leq \sup_{\xi \in \mathbb{R}} |V_0'(\xi) - V_h'(\xi)|.$$

Let $\bar{\xi}_\star^h = \operatorname{argmax}_{\xi \in \Theta_h} \operatorname{dist}(\xi, \Theta_0)$. Since $\lim_{\mathcal{H} \ni h \downarrow 0} \sup_{\xi \in \mathbb{R}} |V_0'(\xi) - V_h'(\xi)| = 0$, we eventually get

$$\lim_{\mathcal{H} \ni h \downarrow 0} V_0'(\bar{\xi}_\star^h) = 0. \tag{B.1}$$

$\{V_0' = 0\}$ being bounded according to Lemma 1.1, we deduce that the sequence $(\bar{\xi}_\star^h)_{h \in \mathcal{H}}$ is bounded. Using (B.1) and the continuity of V_0' , any accumulation point $\bar{\xi}_\star^0$ of the sequence $(\bar{\xi}_\star^h)_{h \in \mathcal{H}}$ satisfies $V_0'(\bar{\xi}_\star^0) = 0$ so that $\bar{\xi}_\star^0 \in \Theta_0$ and

$$\lim_{\mathcal{H} \ni h \downarrow 0} \operatorname{dist}(\bar{\xi}_\star^h, \Theta_0) = 0.$$

Step 2. Limit of $(\chi_\star^h)_{h \in \mathcal{H}}$.

We now assume that $(X_h)_{h \in \mathcal{H}}$ converges to X_0 in $L^1(\mathbb{P})$. By Lemma 1.1, for any $h \in \overline{\mathcal{H}}$,

$$|\chi_\star^h - \chi_\star^0| = \left| \min_{\xi} V_h(\xi) - \min_{\xi} V_0(\xi) \right| \leq \max_{\xi} |(V_h - V_0)(\xi)| \leq \frac{1}{1 - \alpha} \mathbb{E}[|X_h - X_0|],$$

so that $\lim_{\mathcal{H} \ni h \downarrow 0} \chi_\star^h = \chi_\star^0$.

C Proof of Proposition 2.4

Step 1. Asymptotic expansion of ξ_\star^h as $\mathcal{H} \ni h \downarrow 0$.

Since the density function f_{X_0} of X_0 is positive, the set $\Theta_0 = \text{Argmin } V_0 = \{\xi_\star^0\}$ is a singleton so that, by Lemma 2.1, $\xi_\star^h \rightarrow \xi_\star^0$ as $\mathcal{H} \ni h \downarrow 0$. Also note that

$$F_{X_h}(\xi_\star^h) - F_{X_h}(\xi_\star^0) = \alpha - F_{X_h}(\xi_\star^0) = F_{X_0}(\xi_\star^0) - F_{X_h}(\xi_\star^0). \quad (\text{C.1})$$

On the one hand, as $\mathcal{H} \ni h \downarrow 0$, by Assumption 2.2(i),

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

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

$$F_{X_h}(\xi_\star^h) - F_{X_h}(\xi_\star^0) = (\xi_\star^h - \xi_\star^0) \int_0^1 f_{X_h}(t\xi_\star^0 + (1-t)\xi_\star^h) dt. \quad (\text{C.3})$$

Since $\lim_{\mathcal{H} \ni h \downarrow 0} \xi_\star^h = \xi_\star^0$ and, by Assumption 2.2(ii), $(f_{X_h})_{h \in \mathcal{H}}$ converges locally uniformly towards the positive density function f_{X_0} , $\int_0^1 f_{X_h}(t\xi_\star^0 + (1-t)\xi_\star^h) dt > 0$ for $h \in \mathcal{H}$ small enough and $\lim_{h \downarrow 0} \int_0^1 f_{X_h}(t\xi_\star^0 + (1-t)\xi_\star^h) dt = f_{X_0}(\xi_\star^0)$. Combining (C.1), (C.2) and (C.3), we deduce

$$\begin{aligned} h^{-1}(\xi_\star^h - \xi_\star^0) &= -\frac{v(\xi_\star^0)}{\int_0^1 f_{X_h}(t\xi_\star^0 + (1-t)\xi_\star^h) dt} + o(1) \\ &= -\frac{v(\xi_\star^0)}{f_{X_0}(\xi_\star^0)} + o(1) \quad \text{as } \mathcal{H} \ni h \downarrow 0. \end{aligned}$$

Step 2. Asymptotic expansion of χ_\star^h as $\mathcal{H} \ni h \downarrow 0$.

Going back to the definitions of the functions $(V_h)_{h \in \overline{\mathcal{H}}}$ in (1.3) and (2.2), one has

$$\begin{aligned} \chi_\star^h - \chi_\star^0 &= V_h(\xi_\star^h) - V_0(\xi_\star^0) \\ &= \mathbb{E}[X_h | X_h > \xi_\star^h] - \mathbb{E}[X_0 | X_0 > \xi_\star^0] \\ &= \frac{1}{1 - \alpha} (\mathbb{E}[X_h \mathbf{1}_{\{X_h > \xi_\star^h\}}] - \mathbb{E}[X_0 \mathbf{1}_{\{X_0 > \xi_\star^0\}}]) \\ &= \frac{1}{1 - \alpha} \left(\int_{\xi_\star^h}^{\infty} \xi dF_{X_h}(\xi) - \int_{\xi_\star^0}^{\infty} \xi dF_{X_0}(\xi) \right) \\ &= \frac{1}{1 - \alpha} \left(\int_{\xi_\star^h}^{\xi_\star^0} \xi f_{X_h}(\xi) d\xi + \int_{\xi_\star^0}^{\infty} \xi d(F_{X_h} - F_{X_0})(\xi) \right). \end{aligned} \quad (\text{C.4})$$

On the one hand, via a change of variable and Proposition 2.4,

$$\begin{aligned} \int_{\xi_\star^h}^{\xi_\star^0} \xi f_{X_h}(\xi) d\xi &= (\xi_\star^0 - \xi_\star^h) \int_0^1 (t\xi_\star^h + (1-t)\xi_\star^0) f_{X_h}(t\xi_\star^h + (1-t)\xi_\star^0) dt \\ &= v(\xi_\star^0)\xi_\star^0 h + o(h). \end{aligned} \quad (\text{C.5})$$

On the other hand, for $h \in \overline{\mathcal{H}}$, integrating by parts yields

$$\begin{aligned} \int_{\xi_\star^0}^{\infty} \xi dF_{X_h}(\xi) &= -[\xi(1 - F_{X_h}(\xi))]_{\xi_\star^0}^{\infty} + \int_{\xi_\star^0}^{\infty} (1 - F_{X_h}(\xi)) d\xi \\ &= \xi_\star^0(1 - F_{X_h}(\xi_\star^0)) + \int_{\xi_\star^0}^{\infty} (1 - F_{X_h}(\xi)) d\xi, \end{aligned}$$

hence

$$\int_{\xi_\star^0}^{\infty} \xi d(F_{X_h} - F_{X_0})(\xi) = \xi_\star^0(F_{X_0}(\xi_\star^0) - F_{X_h}(\xi_\star^0)) + \int_{\xi_\star^0}^{\infty} (F_{X_0}(\xi) - F_{X_h}(\xi)) d\xi.$$

By plugging the first order Taylor expansion of $F_{X_h} - F_{X_0}$ in Assumption 2.2(i) into the right hand side of the above identity, we obtain

$$\int_{\xi_\star^0}^{\infty} \xi d(F_{X_h} - F_{X_0})(\xi) = -\left(\xi_\star^0 v(\xi_\star^0) + \int_{\xi_\star^0}^{\infty} v(\xi) d\xi\right) h + o(h). \quad (\text{C.6})$$

Eventually, by combining (C.4), (C.5) and (C.6),

$$\chi_\star^h - \chi_\star^0 = -h \int_{\xi_\star^0}^{\infty} \frac{v(\xi)}{1 - \alpha} d\xi + o(h), \quad \text{as } \mathcal{H} \ni h \downarrow 0.$$

D Proof of Theorem 2.7

We follow a strategy similar to the one developed in [10, Section 2.1] and define, for $h \in \overline{\mathcal{H}}$, $\mu \geq 0$ and an integer q , the Lyapunov function $\mathcal{L}_{h,q}^\mu: \mathbb{R} \rightarrow \mathbb{R}_+$ given by

$$\mathcal{L}_{h,q}^\mu(\xi) = (V_h(\xi) - V_h(\xi_\star^h))^q \exp\left(\mu(V_h(\xi) - V_h(\xi_\star^h))\right), \quad \xi \in \mathbb{R}. \quad (\text{D.1})$$

Hereafter, we denote $\mu_{h,q} = q^2 \|V_h''\|_\infty$ and $\bar{\mathcal{L}}_{h,q} = \mathcal{L}_{h,q}^{\mu_{h,q}}$, $h \in \overline{\mathcal{H}}$, q a positive integer.

Lemma D.1. *Under Assumption 2.5, for any $h \in \overline{\mathcal{H}}$, $\mu \geq 0$ and $q \geq 1$,*

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

(ii) for any $\xi \in \mathbb{R}$,

$$\bar{\mathcal{L}}_{h,q}(\xi) \leq k_\alpha^q |\xi - \xi_\star^h|^q \exp\left(\frac{q^2}{1 - \alpha} k_\alpha \sup_{h \in \overline{\mathcal{H}}} \|f_{X_h}\|_\infty |\xi - \xi_\star^h|\right).$$

(iii) for any $\xi \in \mathbb{R}$,

$$V_h'(\xi)(\mathcal{L}_{h,q}^\mu)'(\xi) \geq \lambda_{h,q}^\mu \mathcal{L}_{h,q}^\mu(\xi), \quad \text{where } \lambda_{h,q}^\mu := \frac{3}{8} q V_h''(\xi_\star^h) \wedge \mu \frac{V_h''(\xi_\star^h)^4}{4[V_h'']_{\text{Lip}}^2}.$$

Let $\bar{\lambda}_{h,q} := \lambda_{h,q}^{\mu_{h,q}}$. Then $\inf_{h \in \overline{\mathcal{H}}} \bar{\lambda}_{h,q} > 0$.

(iv) for any $\xi \in \mathbb{R}$,

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

where

$$\eta_{h,q}^\mu := (q \vee \mu) \|V_h''\|_\infty + k_\alpha^2 \mu (\mu \vee 2) + q(2\mu \vee (q-1)) \left(\frac{3k_\alpha^2 [V_h'']_{\text{Lip}}^2}{V_h''(\xi_\star^h)^3} \vee \frac{3 \|V_h''\|_\infty^2}{V_h''(\xi_\star^h)} \right).$$

Besides, introducing $\bar{\eta}_{h,q} := \eta_{h,q}^{\mu_{h,q}}$, one has $|\bar{\lambda}_{h,q}|^2 \leq \bar{\eta}_{h,q}$ and $\sup_{h \in \overline{\mathcal{H}}} \bar{\eta}_{h,q} < \infty$.

(v) for any $\xi \in \mathbb{R}$,

$$(\xi - \xi_\star^h)^{2q} \leq \kappa_{h,q} (\mathcal{L}_{h,q}^\mu(\xi) + \mathcal{L}_{h,2q}^\mu(\xi)), \quad \text{where } \kappa_{h,q} := \frac{3^q}{V_h''(\xi_\star^h)^q} \vee \frac{3^{2q} [V_h'']_{\text{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}}.$$

Moreover, $\sup_{h \in \bar{\mathcal{H}}} \kappa_{h,q} < \infty$.

Proof. See Appendix E. □

The next result concerns our special choice of learning sequence.

Lemma D.2. *Let*

$$\varphi_\eta(t) := \begin{cases} \eta^{-1}(t^\eta - 1), & \eta \neq 0, \\ \ln t + \gamma^\mathcal{E}, & \eta = 0, \end{cases} \quad (\text{D.2})$$

where $\gamma^\mathcal{E}$ is the Euler-Mascheroni constant. Let $\gamma_n = \gamma_1 n^{-\beta}$, $\gamma_1 > 0$, $\beta \in (0, 1]$, and, for $0 \leq k \leq n$,

$$\Pi_{k+1:n} := \prod_{j=k+1}^n (1 - \lambda \gamma_j + \zeta \gamma_j^2),$$

where λ, ζ are some positive parameters, with the convention $\prod_\emptyset = 1$. Assume that $1 - \lambda \gamma_n + \zeta \gamma_n^2 > 0$ for any positive integer n . Then, for $n \geq 2$,

(i) for any $0 \leq k \leq n$,

$$\Pi_{k+1:n} \leq \begin{cases} \exp(\zeta \frac{\pi^2}{6} \gamma_1^2 + \frac{\gamma_1 \lambda}{2}) \frac{(k+1)^{\lambda \gamma_1}}{(n+1)^{\lambda \gamma_1}}, & \beta = 1, \\ \exp(-\lambda \gamma_1 (\varphi_{1-\beta}(n+1) - \varphi_{1-\beta}(k+1))) \\ \quad \times \exp(2^{2\beta} \zeta \gamma_1^2 (\varphi_{1-2\beta}(n+1) - \varphi_{1-2\beta}(k+1))), & \beta \in (0, 1). \end{cases}$$

(ii) for any $p \geq 2$,

$$\sum_{k=1}^n \gamma_k^p \Pi_{k+1:n} \leq \begin{cases} \gamma_1^p \exp(\zeta \frac{\pi^2}{6} \gamma_1^2 + \frac{\gamma_1 \lambda}{2} + \ln 2 (\gamma_1 \lambda \vee p)) \frac{\varphi_{\lambda \gamma_1 - p + 1}(n+1)}{(n+1)^{\lambda \gamma_1}}, & \beta = 1, \\ 2^{2\beta} \gamma_1^p \exp(-2^{-(\beta+2)} \lambda \gamma_1 n^{1-\beta}) \varphi_{1-2\beta}(n+1) + \frac{2^{1+(p-1)\beta} \gamma_1^{p-1}}{\lambda n^{(p-1)\beta}}, & \beta \in (0, 1). \end{cases}$$

Proof. See Appendix F. □

As in the proof of Lemma D.1 in Appendix E, we drop the superscript μ from our notation and write $\mathcal{L}_{h,q}(\xi)$ for $\mathcal{L}_{h,q}^\mu(\xi)$. For any $h \in \mathcal{H}$, we introduce the filtration $\mathcal{F}^h = (\mathcal{F}_n^h)_{n \geq 0}$, where $\mathcal{F}_n^h = \sigma(\xi_0^h, \chi_0^h, X_h^{(1)}, \dots, X_h^{(n)})$, $n \geq 0$. Throughout, we fix $h \in \mathcal{H}$ a bias parameter, q a positive integer and $\mu > 0$.

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

We first prove a general inequality on $\mathbb{E}[\mathcal{L}_{h,q}(\xi_n^h)]$, $n \geq 0$. We decompose the dynamics of $(\xi_n^h)_{n \geq 0}$, given by the first component of (2.3), into

$$\xi_{n+1}^h = \xi_n^h - \gamma_{n+1} V_h'(\xi_n^h) - \gamma_{n+1} e_{n+1}^h, \quad (\text{D.3})$$

where

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

is an $(\mathcal{F}^h, \mathbb{P})$ -martingale increment. We test the Lyapunov function $\mathcal{L}_{h,q}$ along the dynamics (D.3). Via a second order Taylor expansion,

$$\begin{aligned}\mathcal{L}_{h,q}(\xi_{n+1}^h) &= \mathcal{L}_{h,q}(\xi_n^h - \gamma_{n+1}V'_h(\xi_n^h) - \gamma_{n+1}e_{n+1}^h) \\ &= \mathcal{L}_{h,q}(\xi_n^h) - \gamma_{n+1}\mathcal{L}'_{h,q}(\xi_n^h)(V'_h(\xi_n^h) + e_{n+1}^h) \\ &\quad + \gamma_{n+1}^2 H_1(\xi_n^h, X_h^{(n+1)})^2 \int_0^1 (1-t)\mathcal{L}''_{h,q}(t\xi_{n+1}^h + (1-t)\xi_n^h) dt.\end{aligned}$$

It follows from Lemmas D.1(iii,iv) that

$$\begin{aligned}\mathcal{L}_{h,q}(\xi_{n+1}^h) &\leq \mathcal{L}_{h,q}(\xi_n^h)(1 - \lambda_{h,q}^\mu \gamma_{n+1}) - \gamma_{n+1}\mathcal{L}'_{h,q}(\xi_n^h)e_{n+1}^h + \eta_{h,q}^\mu \gamma_{n+1}^2 H_1(\xi_n^h, X_h^{(n+1)})^2 \\ &\quad \times \int_0^1 (1-t) \left(\mathcal{L}_{h,q}(t\xi_{n+1}^h + (1-t)\xi_n^h) + \mathcal{L}_{h,q-1}(t\xi_{n+1}^h + (1-t)\xi_n^h) \right) dt.\end{aligned}\tag{D.4}$$

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

$$V_h(t\xi_{n+1}^h + (1-t)\xi_n^h) = V_h(\xi_n^h) + tV'_h(\tilde{\xi}_n^h(t))(\xi_{n+1}^h - \xi_n^h).\tag{D.5}$$

From (1.4) and (2.3),

$$|\xi_{n+1}^h - \xi_n^h| = \gamma_{n+1}|H_1(\xi_n^h, X_h^{(n+1)})| \leq k_\alpha \gamma_{n+1},\tag{D.6}$$

with $k_\alpha = \frac{\alpha}{1-\alpha} \vee 1$. Hence, applying the triangle inequality to (D.5), recalling that V_h is minimal at ξ_\star^h ,

$$0 \leq V_h(t\xi_{n+1}^h + (1-t)\xi_n^h) - V_h(\xi_\star^h) \leq V_h(\xi_n^h) - V_h(\xi_\star^h) + k_\alpha^2 \gamma_{n+1}.\tag{D.7}$$

Using the inequality $e^x \leq e\mathbb{1}_{\{x \leq 1\}} + x^q e^x \mathbb{1}_{\{x > 1\}} \leq e(1 + x^q e^x)$, $x \in \mathbb{R}$, and (D.1), for any $\xi \in \mathbb{R}$,

$$\mathcal{L}_{h,0}(\xi) \leq e(1 + \mu^q \mathcal{L}_{h,q}(\xi)).\tag{D.8}$$

Thus, using (D.7) then (D.8),

$$\begin{aligned}\mathcal{L}_{h,q}(t\xi_{n+1}^h + (1-t)\xi_n^h) &\leq (V_h(\xi_n^h) - V_h(\xi_\star^h) + k_\alpha^2 \gamma_{n+1})^q \exp\left(\mu(V_h(\xi_n^h) - V_h(\xi_\star^h) + k_\alpha^2 \gamma_{n+1})\right) \\ &\leq 2^{q-1} \exp(\mu k_\alpha^2 \gamma_{n+1}) (\mathcal{L}_{h,q}(\xi_n^h) + k_\alpha^{2q} \gamma_{n+1}^q \mathcal{L}_{h,0}(\xi_n^h)) \leq \sigma_q^\mu (\mathcal{L}_{h,q}(\xi_n^h) + \gamma_{n+1}^q),\end{aligned}\tag{D.9}$$

where

$$\sigma_q^\mu := 2^{q-1} \exp(\mu k_\alpha^2 \gamma_1) ((1 + e\mu^q k_\alpha^{2q} \gamma_1^q) \vee e k_\alpha^{2q}) \geq 2^{q-1}.$$

Besides, by (D.6),

$$\gamma_{n+1}^2 H_1(\xi_n^h, X_h^{(n+1)})^2 \leq k_\alpha^2 \gamma_{n+1}^2.\tag{D.10}$$

Plugging the upper bounds (D.9) and (D.10) into (D.4) yields

$$\begin{aligned}\mathcal{L}_{h,q}(\xi_{n+1}^h) &\leq \mathcal{L}_{h,q}(\xi_n^h)(1 - \lambda_{h,q}^\mu \gamma_{n+1} + \zeta_{h,q}^\mu \gamma_{n+1}^2) - \gamma_{n+1}\mathcal{L}'_{h,q}(\xi_n^h)e_{n+1}^h \\ &\quad + \zeta_{h,q}^\mu \gamma_{n+1}^2 \mathcal{L}_{h,q-1}(\xi_n^h) + \zeta_{h,q}^\mu \gamma_{n+1}^{q+1},\end{aligned}\tag{D.11}$$

with

$$\zeta_{h,q}^\mu := \frac{1}{2} \eta_{h,q}^\mu k_\alpha^2 ((\gamma_1 \sigma_q^\mu + \sigma_{q-1}^\mu) \vee \sigma_q^\mu),\tag{D.12}$$

which, according to the second part of Lemma D.1(iv), satisfies

$$\bar{\zeta}_q := \sup_{h \in \mathcal{H}} \{\bar{\zeta}_{h,q} := \zeta_{h,q}^{\mu_{h,q}}\} < \infty.$$

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

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

Hence, taking the expectation in both sides of inequality (D.11),

$$\begin{aligned} \mathbb{E}[\mathcal{L}_{h,q}(\xi_{n+1}^h)] &\leq \mathbb{E}[\mathcal{L}_{h,q}(\xi_n^h)](1 - \lambda_{h,q}^\mu \gamma_{n+1} + \zeta_{h,q}^\mu \gamma_{n+1}^2) \\ &\quad + \zeta_{h,q}^\mu \gamma_{n+1}^2 \mathbb{E}[\mathcal{L}_{h,q-1}(\xi_n^h)] + \zeta_{h,q}^\mu \gamma_{n+1}^{q+1}. \end{aligned} \quad (\text{D.13})$$

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

We prove here a sharper upper estimate on $\mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)]$. Taking $q = 1$ in (D.13) and (D.8), we get

$$\mathbb{E}[\mathcal{L}_{h,1}(\xi_{n+1}^h)] \leq \mathbb{E}[\mathcal{L}_{h,1}(\xi_n^h)](1 - \lambda_{h,1}^\mu \gamma_{n+1} + (e\mu + 1)\zeta_{h,1}^\mu \gamma_{n+1}^2) + (e + 1)\zeta_{h,1}^\mu \gamma_{n+1}^2. \quad (\text{D.14})$$

Using (D.12), that $|\bar{\lambda}_{h,q}|^2 \leq \bar{\eta}_{h,q}$ by Lemma D.1(iii) and that $x(1 - cx) \leq (4c)^{-1}$, $x \in \mathbb{R}$, $c > 0$,

$$\begin{aligned} \bar{\lambda}_{h,q} \gamma_{n+1} - (e\mu_{h,q}^q + 1)\bar{\zeta}_{h,q} \gamma_{n+1}^2 &\leq \sqrt{\bar{\eta}_{h,q}} \gamma_{n+1} - \bar{\eta}_{h,q} \bar{c}_{h,q} \gamma_{n+1}^2 \\ &= \sqrt{\bar{\eta}_{h,q}} \gamma_{n+1} (1 - \bar{c}_{h,q} \sqrt{\bar{\eta}_{h,q}} \gamma_{n+1}) \leq \frac{1}{4\bar{c}_{h,q}} \leq \frac{1}{2}, \end{aligned} \quad (\text{D.15})$$

where, recalling that $\mu > 0$, $k_\alpha \geq 1$ and $\sigma_{h,q}^\mu \geq 2^{q-1}$,

$$c_{h,q}^\mu := (e\mu^q + 1) \frac{\zeta_{h,q}^\mu}{\eta_{h,q}^\mu} = \frac{1}{2} (e\mu^q + 1) k_\alpha^2 ((\gamma_1 \sigma_q^\mu + \sigma_{q-1}^\mu) \vee \sigma_q^\mu) > \frac{1}{2}, \quad \bar{c}_{h,q} := c_{h,q}^{\mu_{h,q}}.$$

Hence, for $q = 1$,

$$1 - \bar{\lambda}_{h,1} \gamma_{n+1} + (e\mu_{h,1} + 1)\bar{\zeta}_{h,1} \gamma_{n+1}^2 \geq \frac{1}{2}.$$

Evaluating (D.14) at $\mu = \mu_{h,1}$ and iterating n times the inequality yields

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

where

$$\Pi_{k:n}^{h,q} := \prod_{j=k}^n (1 - \bar{\lambda}_{h,q} \gamma_j + (e\mu_{h,q}^q + 1)\bar{\zeta}_{h,q} \gamma_j^2),$$

with the convention $\prod_{\emptyset} = 1$.

If $\beta = 1$, invoking Lemma D.2 with $p = 2$ gives

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

where

$$\widehat{K}_{h,1} := \exp\left((e\mu_{h,q} + 1)\zeta_{h,1} \frac{\pi^2}{6} \gamma_1^2 + \frac{\gamma_1 \bar{\lambda}_{h,1}}{2}\right), \quad \bar{K}_{h,1} := \widehat{K}_{h,1} \gamma_1^2 (e+1) \bar{\zeta}_{h,1} 2^{\gamma_1 \bar{\lambda}_{h,1} \vee 2}.$$

Hence, if $\bar{\lambda}_{h,1} \gamma_1 > 1$, recalling (D.2),

$$\mathbb{E}[\bar{\mathcal{L}}_{h,1}(\xi_n^h)] \leq K_{h,1}^1 \gamma_n, \quad (\text{D.16})$$

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

Otherwise, if $\beta \in (0, 1)$, applying Lemma D.2 with $p = 2$ yields

$$\begin{aligned} \mathbb{E}[\bar{\mathcal{L}}_{h,1}(\xi_n^h)] &\leq \left(\mathbb{E}[\bar{\mathcal{L}}_{h,1}(\xi_0^h)] + \frac{e+1}{e\mu_{h,1}+1} \right) \\ &\times \exp \left(2^{2\beta+1}(e\mu_{h,1}+1)\bar{\zeta}_{h,1}\gamma_1^2\varphi_{1-2\beta}(n+1) - \frac{\bar{\lambda}_{h,1}}{2}\gamma_1\varphi_{1-\beta}(n+1) \right) \\ &+ (e+1)\bar{\zeta}_{h,1} \left(2^{2\beta}\gamma_1^2 \exp(-2^{-(\beta+2)}\bar{\lambda}_{h,1}\gamma_1 n^{1-\beta})\varphi_{1-2\beta}(n+1) + \frac{2^{\beta+1}\gamma_1}{\bar{\lambda}_{h,1}n^\beta} \right). \end{aligned}$$

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

$$\mathbb{E}[\bar{\mathcal{L}}_{h,1}(\xi_n^h)] \leq K_{h,1}^\beta \gamma_n, \quad (\text{D.17})$$

with

$$\begin{aligned} K_{h,1}^\beta &:= \gamma_1^{-1} \left(\mathbb{E}[\bar{\mathcal{L}}_{h,1}(\xi_0^h)] + \frac{e+1}{e\mu_{h,1}+1} \right) \\ &\times \sup_{n \geq 1} \left\{ n^\beta \exp \left(2^{2\beta+1}(e\mu_{h,1}+1)\bar{\zeta}_{h,1}\gamma_1^2\varphi_{1-2\beta}(n+1) - \frac{\bar{\lambda}_{h,1}}{2}\gamma_1\varphi_{1-\beta}(n+1) \right) \right\} \\ &+ (e+1)\bar{\zeta}_{h,1} \left(2^{2\beta}\gamma_1 \sup_{n \geq 1} \{ n^\beta \exp(-2^{-(\beta+2)}\bar{\lambda}_{h,1}\gamma_1 n^{1-\beta})\varphi_{1-2\beta}(n+1) \} + \frac{2^{\beta+1}}{\bar{\lambda}_{h,1}} \right). \end{aligned}$$

Step 3. Inequality on $\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)]$.

Take now $q = 2$ and $\mu = \mu_{h,2}$ in (D.13). Using either (D.16) or (D.17),

$$\begin{aligned} \mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_{n+1}^h)] &\leq \mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)](1 - \bar{\lambda}_{h,2}\gamma_{n+1} + \bar{\zeta}_{h,2}\gamma_{n+1}^2) + \bar{\zeta}_{h,2}\gamma_{n+1}^2 \mathbb{E}[\bar{\mathcal{L}}_{h,1}(\xi_n^h)] + \bar{\zeta}_{h,2}\gamma_{n+1}^3 \\ &\leq \mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)](1 - \bar{\lambda}_{h,2}\gamma_{n+1} + (e\mu_{h,2}^2 + 1)\bar{\zeta}_{h,2}\gamma_{n+1}^2) + (2^\beta e^4 K_{h,1}^\beta + 1)\bar{\zeta}_{h,2}\gamma_{n+1}^3. \end{aligned} \quad (\text{D.18})$$

From (D.15),

$$1 - \bar{\lambda}_{h,2}\gamma_{n+1} - (e\mu_{h,2}^2 + 1)\bar{\zeta}_{h,2}\gamma_{n+1}^2 \geq \frac{1}{2}.$$

Hence, iterating n times inequality (D.18),

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

If $\gamma_n = \gamma_1 n^{-1}$, by Lemma D.2 for $p = 2$,

$$\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)] \leq \widehat{K}_{h,2} \frac{\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_0^h)]}{(n+1)^{\bar{\lambda}_{h,2}\gamma_1}} + \bar{K}_{h,2} \frac{\varphi_{\bar{\lambda}_{h,2}\gamma_1-2}(n+1)}{(n+1)^{\bar{\lambda}_{h,2}\gamma_1}}, \quad (\text{D.19})$$

where

$$\widehat{K}_{h,2} := \exp \left((e\mu_{h,2}^2 + 1)\bar{\zeta}_{h,2} \frac{\pi^2}{6} \gamma_1^2 + \frac{\gamma_1 \bar{\lambda}_{h,2}}{2} \right), \quad \bar{K}_{h,2} := \widehat{K}_{h,2} \gamma_1^3 (2e^4 K_{h,1}^\beta + 1) \bar{\zeta}_{h,2} 2^{\gamma_1 \bar{\lambda}_{h,2} \vee 3}.$$

Since $\bar{\lambda}_{h,2} > \bar{\lambda}_{h,1} > 1$, one has $\gamma_1 \bar{\lambda}_{h,2} > 1$, so that for any positive integer n ,

$$\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)] \leq K_{h,2}^1 \gamma_n, \quad (\text{D.20})$$

with

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

Otherwise, if $\gamma_n = \gamma_1 n^{-\beta}$ with $\beta \in (0, 1)$, recalling Lemma D.2 with $p = 2$,

$$\begin{aligned} \mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)] &\leq \left(\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_0^h)] + \frac{\gamma_1(2^\beta K_{h,1}^\beta + 1)}{e\mu_{h,2}^2 + 1} \right) \\ &\times \exp \left(2^{2\beta+1} \bar{\zeta}_{h,2}(e\mu_{h,2}^2 + 1) \gamma_1^2 \varphi_{1-2\beta}(n+1) - \frac{\bar{\lambda}_{h,2}}{2} \gamma_1 \varphi_{1-\beta}(n+1) \right) \\ &+ (2^\beta K_{h,1}^\beta + 1) \bar{\zeta}_{h,2} \left(2^{2\beta} \gamma_1^3 \exp(-2^{-(\beta+2)} \bar{\lambda}_{h,2} \gamma_1 n^{1-\beta}) \varphi_{1-2\beta}(n+1) + \frac{2^{2\beta+1} \gamma_1^2}{\bar{\lambda}_{h,2} n^{2\beta}} \right). \end{aligned} \quad (\text{D.21})$$

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

$$\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)] \leq K_{h,2}^\beta \gamma_n, \quad (\text{D.22})$$

with

$$\begin{aligned} K_{h,2}^\beta &:= \left(\gamma_1^{-1} \mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_0^h)] + \frac{2^\beta K_{h,1}^\beta + 1}{e\mu_{h,2}^2 + 1} \right) \\ &\times \sup_{n \geq 1} \left\{ n^\beta \exp \left(2^{2\beta+1} \bar{\zeta}_{h,2}(e\mu_{h,2}^2 + 1) \gamma_1^2 \varphi_{1-2\beta}(n+1) - \frac{\bar{\lambda}_{h,2}}{2} \gamma_1 \varphi_{1-\beta}(n+1) \right) \right\} \\ &+ (2^\beta K_{h,1}^\beta + 1) \\ &\times \bar{\zeta}_{h,2} \left(\gamma_1^2 2^{2\beta} \sup_{n \geq 1} \left\{ n^\beta \exp(-2^{-(\beta+2)} \bar{\lambda}_{h,2} \gamma_1 n^{1-\beta}) \varphi_{1-2\beta}(n+1) \right\} + \gamma_1 \frac{2^{2\beta+1}}{\bar{\lambda}_{h,2}} \right). \end{aligned}$$

Step 4. Upper bound on $\mathbb{E}[(\xi_n^h - \xi_\star^h)^2]$.

Combining either the estimates (D.16) with (D.20) or (D.17) with (D.22), then invoking Lemma D.1(v) with $q = 1$, we conclude

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

with $\bar{K}_{h,2}^\beta := \kappa_{h,1}(K_{h,2}^\beta + K_{h,1}^\beta)$ satisfies $\sup_{h \in \mathcal{H}} \bar{K}_{h,2}^\beta < \infty$ for any $\beta \in (0, 1]$. This follows from that, for $q = 1, 2$,

$$\sup_{h \in \mathcal{H}} \mathbb{E}[\bar{\mathcal{L}}_{h,q}(\xi_0^h)] \leq \sup_{h \in \mathcal{H}} \mathbb{E} \left[(1 + |\xi_0^h|^2) \exp \left(\frac{4}{1-\alpha} k_\alpha \sup_{h \in \mathcal{H}} \|f_{X_h}\|_\infty |\xi_0^h| \right) \right] < \infty,$$

recalling Lemma D.1(ii).

Step 5. Upper bound on $\mathbb{E}[|\chi_n^h - \chi_\star^h|]$.

We now prove an $L^1(\mathbb{P})$ upper estimate for the difference $\chi_n^h - \chi_\star^h$, $n \geq 1$, $(\chi_n^h)_{n \geq 0}$ being given by (2.3). Without loss of generality, we assume $\chi_0^h = 0$, inasmuch as the general case $\chi_0^h \neq 0$ is handled similarly. Observe that

$$\begin{aligned} \chi_n^h - \chi_\star^h &= \frac{1}{n} \sum_{k=1}^n \left(\xi_{k-1}^h + \frac{1}{1-\alpha} (X_h^{(k)} - \xi_{k-1}^h)^+ \right) - V_h(\xi_\star^h) \\ &= \frac{1}{n} \sum_{k=1}^n \varepsilon_k^h + \frac{1}{n} \sum_{k=1}^n V_h(\xi_{k-1}^h) - V_h(\xi_\star^h), \end{aligned} \quad (\text{D.23})$$

where

$$\varepsilon_k^h := \xi_{k-1}^h + \frac{1}{1-\alpha} (X_h^{(k)} - \xi_{k-1}^h)^+ - V_h(\xi_{k-1}^h), \quad k \geq 1, \quad (\text{D.24})$$

is a sequence of $(\mathcal{F}^h, \mathbb{P})$ -martingale increments, that is, $\mathbb{E}[\varepsilon_k^h | \mathcal{F}_{k-1}^h] = 0$. Note that

$$\begin{aligned} \mathbb{E}[|\varepsilon_k^h|^2 | \mathcal{F}_{k-1}^h] &\leq \frac{1}{(1-\alpha)^2} \mathbb{E}[\left((X_h^{(k)} - \xi_{k-1}^h)^+ - \mathbb{E}[(X_h^{(k)} - \xi_{k-1}^h)^+ | \mathcal{F}_{k-1}^h]\right)^2 | \mathcal{F}_{k-1}^h] \\ &\leq \frac{1}{(1-\alpha)^2} \mathbb{E}[\left((X_h^{(k)} - \xi_{k-1}^h)^+\right)^2 | \mathcal{F}_{k-1}^h] \\ &\leq \frac{3}{(1-\alpha)^2} (\mathbb{E}[|X_h|^2] + (\xi_{k-1}^h - \xi_\star^h)^2 + |\xi_\star^h|^2). \end{aligned}$$

Using the above upper estimate, the Cauchy-Schwarz inequality and (2.5),

$$\begin{aligned} \frac{1}{n} \mathbb{E} \left[\left| \sum_{k=1}^n \varepsilon_k^h \right| \right] &\leq \frac{1}{n} \left(\sum_{k=1}^n \mathbb{E}[|\varepsilon_k^h|^2] \right)^{\frac{1}{2}} \\ &\leq \frac{\sqrt{3}}{(1-\alpha)n} \left(n \left(\sup_{h \in \mathcal{H}} \mathbb{E}[|X_h|^2] + \sup_{h \in \mathcal{H}} |\xi_\star^h|^2 \right) + \bar{K}_h^\beta \sum_{k=2}^n \gamma_{k-1} + \mathbb{E}[(\xi_0^h - \xi_\star^h)^2] \right)^{\frac{1}{2}} \\ &\leq \frac{\sqrt{3}}{(1-\alpha)n^{\frac{1}{2}}} \left(\left(\sup_{h \in \mathcal{H}} \mathbb{E}[|X_h|^2]^{\frac{1}{2}} + \sup_{h \in \mathcal{H}} |\xi_\star^h| \right) + (\gamma_1 \bar{K}_h^\beta)^{\frac{1}{2}} \frac{1}{(1-\beta)^{\frac{1}{2}} n^{\frac{\beta}{2}}} \mathbf{1}_{\{\beta \in (0,1)\}} \right. \\ &\quad \left. + (\gamma_1 \bar{K}_h^\beta)^{\frac{1}{2}} \frac{(\ln n)^{\frac{1}{2}}}{n^{\frac{1}{2}}} \mathbf{1}_{\{\beta=1\}} + \frac{\mathbb{E}[(\xi_0^h - \xi_\star^h)^2]^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right). \end{aligned}$$

For the second term appearing on the right hand side of (D.23), we use a second order Taylor expansion, that $\sup_{h \in \mathcal{H}} \|V_h''\|_\infty < \infty$ by Assumption 2.5(ii) and (2.5). Thus

$$\begin{aligned} \frac{1}{n} \mathbb{E} \left[\left| \sum_{k=1}^n V_h(\xi_{k-1}^h) - V_h(\xi_\star^h) \right| \right] &\leq \frac{\|V_h''\|_\infty}{2n} \sum_{k=1}^n \mathbb{E}[(\xi_{k-1}^h - \xi_\star^h)^2] \\ &\leq \frac{\|V_h''\|_\infty}{2} \left(\frac{\mathbb{E}[(\xi_0^h - \xi_\star^h)^2]}{n} + \gamma_1 K_h^\beta \left(\frac{1}{(1-\beta)n^\beta} \mathbf{1}_{\{\beta \in (0,1)\}} + \frac{\ln n}{n} \mathbf{1}_{\{\beta=1\}} \right) \right). \end{aligned}$$

Injecting the two previous estimates into (D.23) concludes the proof.

Step 6. Second inequality on $\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)]$.

The proof of (2.6) relies on similar arguments to those employed for (2.5). For the sake of brevity, we omit some technical details. Using either (D.19), recalling that $\gamma_1 \bar{\lambda}_{h,2} > 2$, or (D.21),

$$\mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)] \leq \tilde{K}_{h,2}^\beta \gamma_n^2, \quad (\text{D.25})$$

for some constants $(\tilde{K}_{h,2}^\beta)_{h \in \mathcal{H}}$ satisfying $\sup_{h \in \mathcal{H}} \tilde{K}_{h,2}^\beta < \infty$.

Step 7. Inequality on $\mathbb{E}[\bar{\mathcal{L}}_{h,3}(\xi_n^h)]$.

Then, setting $q = 3$ and $\mu = \mu_{h,3}$ in (D.13) and utilizing the previous inequality,

$$\begin{aligned} \mathbb{E}[\bar{\mathcal{L}}_{h,3}(\xi_{n+1}^h)] &\leq \mathbb{E}[\bar{\mathcal{L}}_{h,3}(\xi_n^h)] (1 - \bar{\lambda}_{h,3} \gamma_{n+1} + \bar{\zeta}_{h,3} \gamma_{n+1}^2) + \bar{\zeta}_{h,3} \gamma_{n+1}^2 \mathbb{E}[\bar{\mathcal{L}}_{h,2}(\xi_n^h)] + \bar{\zeta}_{h,3} \gamma_{n+1}^4 \\ &\leq \mathbb{E}[\bar{\mathcal{L}}_{h,3}(\xi_n^h)] (1 - \bar{\lambda}_{h,3} \gamma_{n+1} + (e\mu_{h,3}^3 + 1) \bar{\zeta}_{h,3} \gamma_{n+1}^2) + (2^{2\beta} e^{\frac{9}{4}} \tilde{K}_{h,2}^\beta + 1) \bar{\zeta}_{h,3} \gamma_{n+1}^4, \end{aligned}$$

so that

$$\mathbb{E}[\bar{\mathcal{L}}_{h,3}(\xi_n^h)] \leq \mathbb{E}[\bar{\mathcal{L}}_{h,3}(\xi_0^h)] \Pi_{1:n}^{h,3} + (2^{2\beta} e^{\frac{9}{4}} \tilde{K}_{h,2}^\beta + 1) \bar{\zeta}_{h,3} \sum_{k=1}^n \gamma_k^4 \Pi_{k+1:n}^{h,3}.$$

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 $\gamma_1 \bar{\lambda}_{h,3} > 2$ if $\beta = 1$, then

$$\mathbb{E}[\bar{\mathcal{L}}_{h,3}(\xi_n^h)] \leq K_{h,3}^\beta \gamma_n^2, \quad (\text{D.26})$$

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

Step 8. Inequality on $\mathbb{E}[\bar{\mathcal{L}}_{h,4}(\xi_n^h)]$.

Finally, we take $q = 4$ and $\mu = \mu_{h,4}$ in (D.13) and use (D.26) to obtain

$$\begin{aligned} \mathbb{E}[\bar{\mathcal{L}}_{h,4}(\xi_{n+1}^h)] &\leq \mathbb{E}[\bar{\mathcal{L}}_{h,4}(\xi_n^h)](1 - \bar{\lambda}_{h,4}\gamma_{n+1} + \zeta_{h,4}\gamma_{n+1}^2) + \bar{\zeta}_{h,4}\gamma_{n+1}^2 \mathbb{E}[\mathcal{L}_{h,3}(\xi_n^h)] + \bar{\zeta}_{h,3}\gamma_{n+1}^5 \\ &\leq \mathbb{E}[\bar{\mathcal{L}}_{h,4}(\xi_n^h)](1 - \bar{\lambda}_{h,4}\gamma_{n+1} + (e\mu_{h,4}^4 + 1)\bar{\zeta}_{h,4}\gamma_{n+1}^2) + (2^{2\beta}e^{\frac{16}{9}}K_{h,3}^\beta + \gamma_1)\bar{\zeta}_{h,4}\gamma_{n+1}^4, \end{aligned}$$

so that

$$\mathbb{E}[\bar{\mathcal{L}}_{h,4}(\xi_n^h)] \leq \mathbb{E}[\bar{\mathcal{L}}_{h,4}(\xi_0^h)]\Pi_{1:n}^{h,4} + (2^{2\beta}e^{\frac{16}{9}}K_{h,3}^\beta + \gamma_1)\bar{\zeta}_{h,4} \sum_{k=1}^n \gamma_k^4 \Pi_{k+1:n}^{h,4}.$$

Analogously, we deduce from the previous inequality that if $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0, 1]$, with $\gamma_1 \bar{\lambda}_{h,4} > 2$ if $\beta = 1$, then for any $h \in \mathcal{H}$ and any positive integer n ,

$$\mathbb{E}[\bar{\mathcal{L}}_{h,4}(\xi_n^h)] \leq K_{h,4}^\beta \gamma_n^2, \quad (\text{D.27})$$

where $(K_{h,4}^\beta)_{h \in \mathcal{H}}$ are constants that satisfy $\sup_{h \in \mathcal{H}} K_{h,4}^\beta < \infty$.

Step 9. Upper bound on $\mathbb{E}[(\xi_n^h - \xi_\star^h)^4]$.

By (D.25), (D.27) and Lemma D.1(v) with $q = 2$, recalling that $\bar{\lambda}_{h,4} > \bar{\lambda}_{h,2}$, we conclude that if $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0, 1]$, with $\gamma_1 \bar{\lambda}_{h,2} > 2$ if $\beta = 1$, then for any $h \in \mathcal{H}$ and any positive integer n ,

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

where $\bar{K}_{h,4}^\beta := \kappa_{h,2}(K_{h,4}^\beta + \tilde{K}_{h,2}^\beta)$ satisfies $\sup_{h \in \mathcal{H}} \bar{K}_{h,4}^\beta < \infty$, $\beta \in (0, 1]$, as Lemma D.1(ii) implies that, for $q = 3, 4$,

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

Step 10. Upper bound on $\mathbb{E}[(\chi_n^h - \chi_\star^h)^2]$.

Assuming here again that $\chi_0^h = 0$, from the decomposition (D.23) and similar computations to the ones performed in Step 5, we get that for some constants $(\bar{C}_h^\beta)_{h \in \mathcal{H}}$ that may change from line to line,

$$\begin{aligned} \mathbb{E}[(\chi_n^h - \chi_\star^h)^2] &\leq \bar{C}_h^\beta \left(\frac{1}{n^2} \sum_{k=1}^n \mathbb{E}[|\varepsilon_k^h|^2] + \frac{1}{n^2} \left(\sum_{k=1}^n \mathbb{E}[(V_h(\xi_{k-1}^h) - V_h(\xi_\star^h))^2]^{\frac{1}{2}} \right)^2 \right) \\ &\leq \bar{C}_h^\beta \left(\frac{1}{n} + \frac{1}{n^2} \left(\sum_{k=1}^n \mathbb{E}[(\xi_{k-1}^h - \xi_\star^h)^4]^{\frac{1}{2}} \right)^2 \right) \leq \bar{C}_h^\beta \left(\frac{1}{n} + \frac{1}{n^2} \left(\sum_{k=1}^n \gamma_k \right)^2 \right) \leq \frac{\bar{C}_h^\beta}{n^{1+2\beta}}, \end{aligned}$$

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.

E Proof of Lemma D.1

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

(i) From the definition (D.1), $\mathcal{L}_{h,q}$ is continuously differentiable and satisfies

$$\begin{aligned}\mathcal{L}'_{h,q}(\xi) &= qV'_h(\xi)\mathcal{L}_{h,q-1}(\xi) + \mu V'_h(\xi)\mathcal{L}_{h,q}(\xi) \\ &= \mathcal{L}_{h,q-1}(\xi)V'_h(\xi)\left(q + \mu(V_h(\xi) - V_h(\xi_\star^h))\right).\end{aligned}\tag{E.1}$$

One may differentiate again the above identity. If $q \geq 2$,

$$\begin{aligned}\mathcal{L}''_{h,q}(\xi) &= \mathcal{L}_{h,q-2}(\xi)V'_h(\xi)^2\left(q-1 + \mu(V_h(\xi) - V_h(\xi_\star^h))\right)\left(q + \mu(V_h(\xi) - V_h(\xi_\star^h))\right) \\ &\quad + \mathcal{L}_{h,q-1}(\xi)V''_h(\xi)\left(q + \mu(V_h(\xi) - V_h(\xi_\star^h))\right) + \mathcal{L}_{h,q-1}(\xi)\mu V'_h(\xi)^2.\end{aligned}\tag{E.2}$$

If $q = 1$,

$$\mathcal{L}''_{h,1}(\xi) = (V''_h(\xi) + 2\mu V'_h(\xi)^2)\mathcal{L}_{h,0}(\xi) + (\mu^2 V'_h(\xi)^2 + \mu V''_h(\xi))\mathcal{L}_{h,1}(\xi).\tag{E.3}$$

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

(ii) The upper estimate follows directly from the fact that, for any $\xi \in \mathbb{R}$,

$$V_h(\xi) - V_h(\xi_\star^h) = (\xi - \xi_\star^h) \int_0^1 V'_h(\xi_\star^h + t(\xi - \xi_\star^h)) dt,$$

hence, since $\|V'_h\|_\infty \leq k_\alpha$, $V_h(\xi) - V_h(\xi_\star^h) \leq k_\alpha |\xi - \xi_\star^h|$.

The remaining properties of Lemma D.1 are trivially satisfied for $\xi = \xi_\star^h$. We thus assume, without loss of generality, that $\xi \neq \xi_\star^h$ in the rest of the proof. We let

$$\varepsilon_h := \frac{V''_h(\xi_\star^h)}{[V''_h]_{\text{Lip}}}.\tag{E.4}$$

Finally, denote $\mathcal{I}^h := [\xi_\star^h - \varepsilon_h, \xi_\star^h + \varepsilon_h]$.

(iii) **Step 1. Preliminaries.**

From (E.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

$$\mathcal{L}'_{h,q}(\xi) = \mathcal{L}_{h,q}(\xi)V'_h(\xi)\left(\frac{q}{V_h(\xi) - V_h(\xi_\star^h)} + \mu\right),$$

so that

$$\mathcal{L}'_{h,q}(\xi)V'_h(\xi) \geq \left(\frac{qV'_h(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \vee \mu V'_h(\xi)^2\right)\mathcal{L}_{h,q}(\xi).\tag{E.5}$$

Let us establish a lower bound for the parenthesised factor above. Since $V'_h(\xi_\star^h) = 0$, a first order Taylor expansion writes

$$\begin{aligned}V'_h(\xi) &= (\xi - \xi_\star^h) \int_0^1 V''_h(\xi_\star^h + t(\xi - \xi_\star^h)) dt \\ &= V''_h(\xi_\star^h)(\xi - \xi_\star^h) + (\xi - \xi_\star^h) \int_0^1 \left(V''_h(\xi_\star^h + t(\xi - \xi_\star^h)) - V''_h(\xi_\star^h)\right) dt,\end{aligned}\tag{E.6}$$

which, combined with the fact that $V_h'' = (1 - \alpha)^{-1} f_{X_h} \geq 0$ and the triangle inequality yields

$$|V_h'(\xi)| \geq V_h''(\xi_*^h) |\xi - \xi_*^h| - \frac{1}{2} [V_h'']_{\text{Lip}} (\xi - \xi_*^h)^2. \quad (\text{E.7})$$

Step 2. Case where $\xi \in \mathcal{I}^h$.

If $\xi \in \mathcal{I}^h$, the above inequality yields

$$V_h'(\xi)^2 \geq \frac{1}{4} V_h''(\xi_*^h)^2 (\xi - \xi_*^h)^2. \quad (\text{E.8})$$

Now, recalling that $V_h'(\xi_*^h) = 0$, a second order Taylor expansion gives

$$\begin{aligned} V_h(\xi) - V_h(\xi_*^h) &= \frac{1}{2} V_h''(\xi_*^h) (\xi - \xi_*^h)^2 \\ &\quad + (\xi - \xi_*^h)^2 \int_0^1 (1-t) \left(V_h''(\xi_*^h + t(\xi - \xi_*^h)) - V_h''(\xi_*^h) \right) dt, \end{aligned} \quad (\text{E.9})$$

so that

$$V_h(\xi) - V_h(\xi_*^h) \leq \frac{1}{2} V_h''(\xi_*^h) (\xi - \xi_*^h)^2 + \frac{1}{6} [V_h'']_{\text{Lip}} |\xi - \xi_*^h|^3.$$

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

$$V_h(\xi) - V_h(\xi_*^h) \leq \frac{2}{3} V_h''(\xi_*^h) (\xi - \xi_*^h)^2. \quad (\text{E.10})$$

Plugging (E.8) and (E.10) into (E.5), recalling that $V_h''(\xi_*^h) = (1 - \alpha)^{-1} f_{X_h}(\xi_*^h) > 0$ by Assumption 2.5(i),

$$\mathcal{L}'_{h,q}(\xi) V_h'(\xi) \geq \frac{q V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_*^h)} \mathcal{L}_{h,q}(\xi) \geq \frac{3}{8} q V_h''(\xi_*^h) \mathcal{L}_{h,q}(\xi). \quad (\text{E.11})$$

Step 3. Case where $\xi \notin \mathcal{I}^h$.

Assume now $\xi \notin \mathcal{I}^h$, say $\xi > \xi_*^h + \varepsilon_h$. The case $\xi < \xi_*^h - \varepsilon_h$ is similar and is omitted. Since V_h' is non-decreasing, $V_h'(\xi_*^h + \varepsilon_h) \geq V_h'(\xi_*^h) = 0$ and, evaluating (E.7) at $\xi_*^h + \varepsilon_h$ and recalling the definition (E.4),

$$V_h'(\xi) \geq V_h'(\xi_*^h + \varepsilon_h) \geq V_h''(\xi_*^h) \varepsilon_h - \frac{1}{2} [V_h'']_{\text{Lip}} \varepsilon_h^2 = \frac{V_h''(\xi_*^h)^2}{2 [V_h'']_{\text{Lip}}}.$$

Plugging this into (E.5), for any $\xi > \xi_*^h + \varepsilon_h$,

$$\mathcal{L}'_{h,q}(\xi) V_h'(\xi) \geq \mu \mathcal{L}_{h,q}(\xi) V_h'(\xi)^2 \geq \mu \frac{V_h''(\xi_*^h)^4}{4 [V_h'']_{\text{Lip}}^2} \mathcal{L}_{h,q}(\xi). \quad (\text{E.12})$$

The above inequality is valid also if $\xi < \xi_*^h - \varepsilon_h$.

Step 4. Conclusion.

Combining (E.11) and (E.12),

$$\mathcal{L}'_{h,q}(\xi) V_h'(\xi) \geq \left(\frac{3}{8} q V_h''(\xi_*^h) \wedge \mu \frac{V_h''(\xi_*^h)^4}{4 [V_h'']_{\text{Lip}}^2} \right) \mathcal{L}_{h,q}(\xi) = \lambda_{h,q}^\mu \mathcal{L}_{h,q}(\xi), \quad \xi \in \mathbb{R}.$$

By Lemma 2.1, there exists $R > 0$ such that $\xi_*^h \in B(\xi_*^0, R)$, $h \in \mathcal{H}$. By Assumption 2.5,

$$\inf_{h \in \mathcal{H}} V_h''(\xi_*^h) \geq \frac{1}{1 - \alpha} \inf_{\substack{h \in \mathcal{H} \\ \xi \in B(\xi_*^0, R)}} f_{X_h}(\xi) > 0, \quad \inf_{h \in \mathcal{H}} \frac{1}{[V_h'']_{\text{Lip}}} = \frac{1 - \alpha}{\sup_{h \in \mathcal{H}} [f_{X_h}]_{\text{Lip}}} > 0$$

and, by Assumption 2.5(i), $\inf_{h \in \overline{\mathcal{H}}} \|V_h''\|_\infty = (1-\alpha)^{-1} \inf_{h \in \overline{\mathcal{H}}} \|f_{X_h}\|_\infty > 0$. Thus $\inf_{h \in \overline{\mathcal{H}}} \bar{\lambda}_{h,q} > 0$.

(iv) **Step 1. Case where $q \geq 2$.**

Assume first that $q \geq 2$. Through (E.2) and the relation

$$\mathcal{L}_{h,q-1}(\xi) = (V_h(\xi) - V_h(\xi_\star^h))^{-1} \mathcal{L}_{h,q}(\xi),$$

one has

$$\begin{aligned} \mathcal{L}_{h,q}''(\xi) &= \mathcal{L}_{h,q-1}(\xi) \left(qV_h''(\xi) + \frac{q(q-1)V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \right) \\ &\quad + \mathcal{L}_{h,q}(\xi) \left(\mu V_h''(\xi) + \mu^2 V_h'(\xi)^2 + 2q\mu \frac{V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \right). \end{aligned} \quad (\text{E.13})$$

Step 1.1. Sub-case where $\xi \in \mathcal{I}^h$.

Applying the triangle inequality to (E.9) yields

$$V_h(\xi) - V_h(\xi_\star^h) \geq \frac{1}{2} V_h''(\xi_\star^h) (\xi - \xi_\star^h)^2 - \frac{1}{6} [V_h'']_{\text{Lip}} |\xi - \xi_\star^h|^3. \quad (\text{E.14})$$

Hence, if $\xi \in \mathcal{I}^h$, $|\xi - \xi_\star^h| \leq \varepsilon_h$ so that the above inequality and (E.4) yield

$$V_h(\xi) - V_h(\xi_\star^h) \geq \frac{1}{3} V_h''(\xi_\star^h) (\xi - \xi_\star^h)^2. \quad (\text{E.15})$$

Moreover, using (E.6) and that $\|V_h''\|_\infty = (1-\alpha)^{-1} \|f_{X_h}\|_\infty < \infty$ by Assumption 2.5(ii),

$$V_h'(\xi)^2 \leq \|V_h''\|_\infty^2 (\xi - \xi_\star^h)^2.$$

Combining the previous two inequalities, if $\xi \in \mathcal{I}^h$,

$$\frac{V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \leq \frac{3\|V_h''\|_\infty^2}{V_h''(\xi_\star^h)}. \quad (\text{E.16})$$

Step 1.2. Sub-case where $\xi \notin \mathcal{I}^h$.

Assume now $\xi \notin \mathcal{I}^h$, say $\xi > \xi_\star^h + \varepsilon_h$. The case $\xi < \xi_\star^h - \varepsilon_h$ is treated similarly and is hereby omitted. Using that V_h is increasing on $[\xi_\star^h, \infty)$, evaluating (E.14) at $\xi_\star^h + \varepsilon_h$ and recalling (E.4), we obtain

$$V_h(\xi) - V_h(\xi_\star^h) \geq V_h(\xi_\star^h + \varepsilon_h) - V_h(\xi_\star^h) \geq \frac{V_h''(\xi_\star^h)^3}{3[V_h'']_{\text{Lip}}^2},$$

This inequality together with the fact that $\|V_h'\|_\infty \leq k_\alpha$ give, for any $\xi > \xi_\star^h + \varepsilon_h$,

$$\frac{V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \leq \frac{3k_\alpha^2 [V_h'']_{\text{Lip}}^2}{V_h''(\xi_\star^h)^3}. \quad (\text{E.17})$$

The above upper estimate holds too if $\xi < \xi_\star^h - \varepsilon_h$.

Step 1.3. Sub-conclusion.

(E.16) and (E.17) show that, for any $\xi \neq \xi_\star^h$,

$$\frac{V_h'(\xi)^2}{V_h(\xi) - V_h(\xi_\star^h)} \leq \frac{3k_\alpha^2 [V_h'']_{\text{Lip}}^2}{V_h''(\xi_\star^h)^3} \vee \frac{3\|V_h''\|_\infty^2}{V_h''(\xi_\star^h)} =: \nu_h. \quad (\text{E.18})$$

Combining (E.13) and (E.18), recalling that $\|V_h'\|_\infty \leq k_\alpha$,

$$|\mathcal{L}_{h,q}''(\xi)| \leq \left((q \vee \mu) \|V_h''\|_\infty + \mu^2 k_\alpha^2 + q(2\mu \vee (q-1)) \nu_h \right) (\mathcal{L}_{h,q}(\xi) + \mathcal{L}_{h,q-1}(\xi)).$$

Step 2. Case where $q = 1$.

If $q = 1$, since $\|V_h'\|_\infty \leq k_\alpha$, (E.3) directly gives

$$|\mathcal{L}_{h,1}''(\xi)| \leq ((\mu \vee 1)\|V_h''\|_\infty + \mu(\mu \vee 2)k_\alpha^2)(\mathcal{L}_{h,0}(\xi) + \mathcal{L}_{h,1}(\xi)).$$

Step 3. Conclusion.

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

$$|\mathcal{L}_{h,q}''(\xi)| \leq \left((q \vee \mu)\|V_h''\|_\infty + \mu(\mu \vee 2)k_\alpha^2 + q(2\mu \vee (q-1))\nu_h \right) \times (\mathcal{L}_{h,q}(\xi) + \mathcal{L}_{h,q-1}(\xi)).$$

The property $|\bar{\lambda}_{h,q}|^2 \leq \bar{\eta}_{h,q}$ is a straightforward consequence of the definitions of $\bar{\lambda}_{h,q}$ and $\bar{\eta}_{h,q}$.

By Assumption 2.5(ii),

$$\sup_{h \in \bar{\mathcal{H}}} \|V_h''\|_\infty = \frac{\sup_{h \in \bar{\mathcal{H}}} \|f_{X_h}\|_\infty}{1 - \alpha} < \infty, \quad \sup_{h \in \bar{\mathcal{H}}} [V_h'']_{\text{Lip}} = \frac{\sup_{h \in \bar{\mathcal{H}}} [f_{X_h}]_{\text{Lip}}}{1 - \alpha} < \infty. \quad (\text{E.19})$$

Moreover, since $(\xi_\star^h)_{h \in \mathcal{H}}$ converges to ξ_\star^0 as $\mathcal{H} \ni h \downarrow 0$, there exists $R > 0$ such that $\xi_\star^h \in B(\xi_\star^0, R)$, $h \in \mathcal{H}$. Thus, by Assumption 2.5(i),

$$\sup_{h \in \bar{\mathcal{H}}} \frac{1}{V_h''(\xi_\star^h)} \leq \frac{1 - \alpha}{\inf_{\substack{h \in \bar{\mathcal{H}} \\ \xi \in B(\xi_\star^0, R)}} f_{X_h}(\xi)} < \infty. \quad (\text{E.20})$$

Coming back to (E.18), the two results above show that $\sup_{h \in \bar{\mathcal{H}}} \nu_h < \infty$. Eventually,

$$\sup_{h \in \bar{\mathcal{H}}} \left\{ \eta_{h,q}^{\mu_{h,q}} = (q \vee \mu_{h,q})\|V_h''\|_\infty + \mu_{h,q}(\mu_{h,q} \vee 2)k_\alpha^2 + q(2\mu_{h,q} \vee (q-1))\nu_h \right\} < \infty.$$

(v) On \mathcal{I}^h , (E.15) shows that every $\xi \neq \xi_\star^h$ satisfies

$$(\xi - \xi_\star^h)^{2q} \leq \frac{3^q}{V_h''(\xi_\star^h)^q} (V_h(\xi) - V_h(\xi_\star^h))^q \leq \frac{3^q}{V_h''(\xi_\star^h)^q} \mathcal{L}_{h,q}(\xi). \quad (\text{E.21})$$

Outside of \mathcal{I}^h , say if $\xi > \xi_\star^h + \varepsilon_h$ (the case $\xi < \xi_\star^h - \varepsilon_h$ is similar), given that V_h is increasing on $[\xi_\star^h, \infty)$, by (E.14) and (E.4),

$$\begin{aligned} V_h(\xi) - V_h(\xi_\star^h) &\geq V_h(\xi_\star^h + \varepsilon_h) - V_h(\xi_\star^h) \\ &\geq (\xi - \xi_\star^h) \left(\frac{1}{2} V_h''(\xi_\star^h) \varepsilon_h - \frac{1}{6} [V_h'']_{\text{Lip}} \varepsilon_h^2 \right) = (\xi - \xi_\star^h) \frac{V_h''(\xi_\star^h)^2}{3[V_h'']_{\text{Lip}}}, \end{aligned}$$

hence

$$(\xi - \xi_\star^h)^{2q} \leq \frac{3^{2q} [V_h'']_{\text{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}} (V_h(\xi) - V_h(\xi_\star^h))^{2q} \leq \frac{3^{2q} [V_h'']_{\text{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}} \mathcal{L}_{h,2q}(\xi). \quad (\text{E.22})$$

The above upper bound holds also if $\xi < \xi_\star^h - \varepsilon_h$. Putting together (E.21) and (E.22), for any $\xi \neq \xi_\star^h$,

$$\begin{aligned} (\xi - \xi_\star^h)^{2q} &\leq \left(\frac{3^q}{V_h''(\xi_\star^h)^q} \vee \frac{3^{2q} [V_h'']_{\text{Lip}}^{2q}}{V_h''(\xi_\star^h)^{4q}} \right) (\mathcal{L}_{h,q}(\xi) + \mathcal{L}_{h,2q}(\xi)) \\ &= \kappa_{h,q} (\mathcal{L}_{h,q}(\xi) + \mathcal{L}_{h,2q}(\xi)). \end{aligned}$$

Relations (E.19) and (E.20) yield $\sup_{h \in \bar{\mathcal{H}}} \kappa_{h,q} < \infty$.

F Proof of Lemma D.2

Let $n \geq 2$, $0 \leq k \leq n$ and $p \geq 2$. Using the bound $1 + x \leq e^x$, $x \in \mathbb{R}$,

$$\Pi_{k+1:n} \leq \exp\left(-\lambda \sum_{j=k+1}^n \gamma_j\right) \exp\left(\zeta \sum_{j=k+1}^n \gamma_j^2\right),$$

with the convention $\sum_{\emptyset} = 0$.

Step 1. Case where $\beta = 1$.

Step 1.1. Inequality on $\Pi_{k+1:n}$.

Note that

$$\sum_{j=k+1}^n \gamma_j = \gamma_1(\psi(n+1) - \psi(k+1)),$$

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

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

yielding

$$\Pi_{k+1:n} \leq \exp\left(\zeta \frac{\pi^2}{6} \gamma_1^2 + \frac{\gamma_1 \lambda}{2}\right) \frac{(k+1)^{\lambda \gamma_1}}{(n+1)^{\lambda \gamma_1}}.$$

Step 1.2. Inequality on $\sum_{k=1}^n \gamma_k^p \Pi_{k+1:n}$.

The above estimate, together with a comparison between series and integrals, give

$$\sum_{k=1}^n \gamma_k^p \Pi_{k+1:n} \leq \gamma_1^p \exp\left(\zeta \frac{\pi^2}{6} \gamma_1^2 + \frac{\gamma_1 \lambda}{2} + \ln 2(\gamma_1 \lambda \vee p)\right) \frac{\varphi_{\lambda \gamma_1 - p + 1}(n+1)}{(n+1)^{\lambda \gamma_1}}.$$

Step 2. Case where $\beta \in (0, 1)$.

Step 2.1. Inequality on $\Pi_{k+1:n}$.

A comparison between series and integrals gives

$$\begin{aligned} \Pi_{k+1:n} &\leq \exp\left(-\lambda \gamma_1(\varphi_{1-\beta}(n+1) - \varphi_{1-\beta}(k+1))\right) \\ &\quad \times \exp\left(2^{2\beta} \zeta \gamma_1^2(\varphi_{1-2\beta}(n+1) - \varphi_{1-2\beta}(k+1))\right). \end{aligned}$$

Step 2.2. Inequality on $\sum_{k=1}^n \gamma_k^p \Pi_{k+1:n}$.

Introduce $n_0 := \inf\{n \in \mathbb{N} : \gamma_n \leq \lambda/(2\zeta)\} - 1$. Observe that

$$1 - \lambda \gamma_n + \zeta \gamma_n^2 \leq 1 - \frac{\lambda}{2} \gamma_n, \quad n \geq n_0 + 1,$$

so that

$$\begin{aligned} \sum_{k=1}^n \gamma_k^p \Pi_{k+1:n} &= \sum_{k=1}^{n_0 \wedge n} \gamma_k^p \Pi_{k+1:n_0 \wedge n} \Pi_{n_0 \wedge n + 1:n} + \sum_{k=n_0 \wedge n + 1}^n \gamma_k^p \Pi_{k+1:n} \\ &\leq \left(\sum_{k=1}^{n_0 \wedge n} \gamma_k^p \prod_{j=k+1}^{n_0 \wedge n} (1 + \zeta \gamma_j^2) \right) \prod_{j=n_0 \wedge n + 1}^n \left(1 - \frac{\lambda}{2} \gamma_j\right) + \sum_{k=1}^n \gamma_k^p \prod_{j=k+1}^n \left(1 - \frac{\lambda}{2} \gamma_j\right). \end{aligned} \tag{F.1}$$

We simplify the first term on the right hand side of the inequality above. We have

$$\begin{aligned}
\sum_{k=1}^{n_0 \wedge n} \gamma_k^p \prod_{j=k+1}^{n_0 \wedge n} (1 + \zeta \gamma_j^2) &\leq \gamma_1^{p-2} \sum_{k=1}^{n_0 \wedge n} \gamma_k^2 \prod_{j=k+1}^{n_0 \wedge n} (1 + \zeta \gamma_j^2) \\
&= \frac{\gamma_1^{p-2}}{\zeta} \sum_{k=1}^{n_0 \wedge n} \left(\prod_{j=k}^{n_0 \wedge n} (1 + \zeta \gamma_j^2) - \prod_{j=k+1}^{n_0 \wedge n} (1 + \zeta \gamma_j^2) \right) \\
&\leq \frac{\gamma_1^{p-2}}{\zeta} \prod_{j=1}^{n_0 \wedge n} (1 + \zeta \gamma_j^2) \leq \frac{\gamma_1^{p-2}}{\zeta} \exp \left(\zeta \sum_{j=1}^{n_0 \wedge n} \gamma_j^2 \right),
\end{aligned}$$

and

$$\begin{aligned}
\prod_{j=n_0 \wedge n+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) &\leq \exp \left(-\frac{\lambda}{2} \sum_{j=1}^n \gamma_j \right) \exp \left(\frac{\lambda}{2} \sum_{j=1}^{n_0 \wedge n} \gamma_j \right) \\
&\leq \exp \left(-\frac{\lambda}{2} \sum_{j=1}^n \gamma_j \right) \exp \left(\zeta \sum_{j=1}^{n_0 \wedge n} \gamma_j^2 \right),
\end{aligned}$$

where we used that $\gamma_j \geq \lambda/(2\zeta)$ for $j \leq n_0$. Combining the two preceding estimates and using a comparison between series and integrals,

$$\begin{aligned}
&\left(\sum_{k=1}^{n_0 \wedge n} \gamma_k^p \prod_{j=k+1}^{n_0 \wedge n} (1 + \zeta \gamma_j^2) \right) \prod_{j=n_0 \wedge n+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) \\
&\leq \frac{\gamma_1^{p-2}}{\zeta} \exp \left(2\zeta \sum_{j=1}^{n_0 \wedge n} \gamma_j^2 \right) \exp \left(-\frac{\lambda}{2} \sum_{j=1}^n \gamma_j \right) \\
&\leq \frac{\gamma_1^{p-2}}{\zeta} \exp \left(2^{2\beta+1} \zeta \gamma_1^2 \varphi_{1-2\beta}(n+1) \right) \exp \left(-\frac{\lambda}{2} \gamma_1 \varphi_{1-\beta}(n+1) \right).
\end{aligned}$$

Next, we deal with the second term appearing in the right hand side of (F.1). Since $(\gamma_n)_{n \geq 1}$ is a positive non-increasing sequence, for any integer $1 \leq m \leq n$, it holds

$$\begin{aligned}
&\sum_{k=1}^n \gamma_k^p \prod_{j=k+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) \\
&= \sum_{k=1}^m \gamma_k^p \prod_{j=k+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) + \sum_{k=m+1}^n \gamma_k^p \prod_{j=k+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) \\
&\leq \prod_{j=m+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) \sum_{k=1}^m \gamma_k^p + \gamma_m^{p-1} \sum_{k=m+1}^n \gamma_k \prod_{j=k+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) \\
&\leq \exp \left(-\frac{\lambda}{2} \sum_{j=m+1}^n \gamma_j \right) \gamma_1^{p-2} \sum_{k=1}^n \gamma_k^2 + \frac{2\gamma_m^{p-1}}{\lambda} \left(1 - \prod_{j=m+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) \right) \\
&\leq 2^{2\beta} \gamma_1^p \exp \left(-2^{-(\beta+1)} \lambda \gamma_1 (\varphi_{1-\beta}(n) - \varphi_{1-\beta}(m)) \right) \varphi_{1-2\beta}(n+1) + \frac{2\gamma_1^{p-1}}{\lambda m^{(p-1)\beta}}.
\end{aligned}$$

Select $m = \lceil n/2 \rceil$. By the concavity and the non-decreasing monotony of $\varphi_{1-\beta}$, for $n \geq 2$,

$$\varphi_{1-\beta}(n) - \varphi_{1-\beta}(n/2) \geq n^{1-\beta}/2.$$

Hence, for any integer n ,

$$\begin{aligned}
&\sum_{k=1}^n \gamma_k^p \prod_{j=k+1}^n \left(1 - \frac{\lambda}{2} \gamma_j \right) \\
&\leq 2^{2\beta} \gamma_1^p \exp(-2^{-(\beta+2)} \lambda \gamma_1 n^{1-\beta}) \varphi_{1-2\beta}(n+1) + \frac{2^{1+(p-1)\beta} \gamma_1^{p-1}}{\lambda n^{(p-1)\beta}}.
\end{aligned}$$

G Proof of Proposition 3.2

The next result draws inspiration from [22, Lemma 5.1 and Proposition 5.2] and [6, Lemma A.4].

Lemma G.1. *Let X and Y be two real-valued random variables with bounded density function denoted 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 $\xi \in \mathbb{R}$,*

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

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

$$\mathbb{E}[\exp(u(X - Y))] \leq \exp(d_{X,Y}u^2). \quad (\text{G.1})$$

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

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

Proof. Let $\xi \in \mathbb{R}$. For any $\delta > 0$,

$$\begin{aligned} \mathbb{E}[|\mathbf{1}_{\{X > \xi\}} - \mathbf{1}_{\{Y > \xi\}}|] &= \mathbb{E}[\mathbf{1}_{\{X > \xi\}}\mathbf{1}_{\{Y \leq \xi\}} + \mathbf{1}_{\{Y > \xi\}}\mathbf{1}_{\{X \leq \xi\}}] \\ &= \mathbb{P}[Y \leq \xi < X] + \mathbb{P}[X \leq \xi < Y] \\ &= \mathbb{P}[Y \leq \xi, \xi + \delta < X] + \mathbb{P}[Y \leq \xi < X \leq \xi + \delta] \\ &\quad + \mathbb{P}[X \leq \xi, \xi + \delta < Y] + \mathbb{P}[X \leq \xi < Y \leq \xi + \delta] \\ &\leq \mathbb{P}[X - Y > \delta] + \mathbb{P}[X - Y < -\delta] \\ &\quad + \mathbb{P}[\xi < X \leq \xi + \delta] + \mathbb{P}[\xi < Y \leq \xi + \delta] \\ &\leq \mathbb{P}[X - Y > \delta] + \mathbb{P}[X - Y < -\delta] + \delta(\|f_X\|_\infty + \|f_Y\|_\infty). \end{aligned} \quad (\text{G.2})$$

(i) By (G.2),

$$\begin{aligned} \mathbb{E}[|\mathbf{1}_{\{X > \xi\}} - \mathbf{1}_{\{Y > \xi\}}|] &\leq \mathbb{P}[|X - Y| > \delta] + \delta(\|f_X\|_\infty + \|f_Y\|_\infty) \\ &\leq \frac{\mathbb{E}[|X - Y|^{p_\star}]}{\delta^{p_\star}} + \delta(\|f_X\|_\infty + \|f_Y\|_\infty). \end{aligned}$$

A straightforward optimization yields the choice

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

giving the first result.

(ii) By (G.2), the exponential Markov inequality and (G.1), for any $u \geq 0$,

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

The right hand side of the above inequality above is minimized in u for $\delta/(2d_{X,Y})$, yielding

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

Evaluating the right hand side of the previous inequality at

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

gives

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

eventually completing the proof of the second result. \square

Let $h, h' \in \overline{\mathcal{H}}$. Without loss of generality and to avoid trivialities, we assume $0 \leq h < h'$. We study separately the cases $0 < h < h'$ and $0 = h < h'$.

The proof of the point (i)a is inspired from [22, Lemma 3.2].

(i)a. Assume that

$$\mathbb{E}[|\varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]|^p] < \infty \quad \text{for some } p \geq 1.$$

Then, for $h = \frac{1}{K} \in \mathcal{H}$,

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

Case 1. $0 < h < h'$.

Introduce $\tilde{\varphi}(Y, Z) := \varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]$ and take $h = \frac{1}{K} < h' = \frac{1}{K'}$. Then

$$X_h - X_{h'} = h \sum_{k=K'+1}^K \tilde{\varphi}(Y, Z^{(k)}) + (h - h') \sum_{k=1}^{K'} \tilde{\varphi}(Y, Z^{(k)}), \quad (\text{G.3})$$

so that by the triangle inequality,

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

By the tower law and the Burkholder-Davis-Gundy inequality,

$$\begin{aligned} \mathbb{E}[|X_h - X_{h'}|^p]^{\frac{1}{p}} &\leq B_p \mathbb{E}[|\varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]|^p]^{\frac{1}{p}} (h(K - K')^{\frac{1}{2}} + (h' - h)K'^{\frac{1}{2}}) \\ &\leq B_p \mathbb{E}[|\varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]|^p]^{\frac{1}{p}} (h' - h)^{\frac{1}{2}} \left(\left(\frac{h}{h'}\right)^{\frac{1}{2}} + \left(1 - \frac{h}{h'}\right)^{\frac{1}{2}} \right) \\ &\leq \sqrt{2} B_p \mathbb{E}[|\varphi(Y, Z) - \mathbb{E}[\varphi(Y, Z)|Y]|^p]^{\frac{1}{p}} (h' - h)^{\frac{1}{2}}. \end{aligned} \quad (\text{G.4})$$

Case 2. $0 < h$.

Note that, for $h = \frac{1}{K} \in \mathcal{H}$,

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

Since $(\tilde{\varphi}(Y, Z^{(k)}))_{1 \leq k \leq K}$ is a martingale increment sequence conditionally on Y , using successively the tower's law, the Burkholder-Davis-Gundy inequality and the triangle inequality, we obtain

$$\begin{aligned} \mathbb{E}[|X_h - X_0|^p]^{\frac{1}{p}} &= h \mathbb{E}\left[\left|\sum_{k=1}^K \tilde{\varphi}(Y, Z^{(k)})\right|^p\right]^{\frac{1}{p}} \\ &\leq Ch \mathbb{E}\left[\left|\sum_{k=1}^K \tilde{\varphi}(Y, Z^{(k)})^2\right|^{\frac{p}{2}}\right]^{\frac{1}{p}} \leq C \mathbb{E}[|\tilde{\varphi}(Y, Z)|^p]^{\frac{1}{p}} h^{\frac{1}{2}}. \end{aligned}$$

We conclude the proof of the first point by applying Lemma G.1(i) with $p = p_*$.

(i)b. **Case 1.** $0 < h < h'$.

Coming back to the decomposition (G.3), using that $(\tilde{\varphi}(Y, Z^{(k)}))_{1 \leq k \leq K}$ are independent conditionally to Y and (3.5),

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

Case 2. $0 < h$.

Let $h = \frac{1}{K} \in \mathcal{H}$. Via the independence of the sequence $(\tilde{\varphi}(Y, Z^{(k)}))_{1 \leq k \leq K}$ conditionally to Y and the Gaussian concentration assumption,

$$\begin{aligned} \mathbb{E}[\exp(u(X_h - X_0)) | Y] &= \mathbb{E}\left[\exp\left(uh \sum_{k=1}^K \tilde{\varphi}(Y, Z^{(k)})\right) \middle| Y\right] \\ &= \mathbb{E}[\exp(uh\tilde{\varphi}(Y, Z)) | Y]^K \leq \exp(C_g u^2 h). \end{aligned}$$

Taking the expectation on both sides of the above inequalities and setting $X = X_h$, $Y = X_{h'}$ and $d_{X,Y} = C_g(h' - h)$ in the first case, and $X = X_h$, $Y = X_0$ and $d_{X,Y} = C_g h$ in the second case, guarantee the condition (G.1). We eventually conclude the proof by applying Lemma G.1(ii).

(ii) We first write

$$\mathbb{E}[|\mathbb{1}_{\{X_{h_\ell} > \xi\}} - \mathbb{1}_{\{X_{h_{\ell-1}} > \xi\}}|] = \mathbb{P}[X_{h_{\ell-1}} \leq \xi < X_{h_\ell}] + \mathbb{P}[X_{h_\ell} \leq \xi < X_{h_{\ell-1}}]. \quad (\text{G.5})$$

Introducing the random variable G_ℓ ,

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

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

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

Similarly

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

Coming back to (G.5) and summing up the two preceding inequalities yields

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

completing the proof.

H Proof of Theorem 3.6

Throughout, K designates a constant that may change from line to line.

Step 1. Study of $\xi_{\mathbf{N}}^{\text{ML}} - \xi_{\star}^{h_L}$.

We follow a similar strategy to the one used in [14, Lemma 2.7] and decompose the dynamics of the sequence $(\xi_n^h)_{n \geq 0}$ given by (2.3) into

$$\begin{aligned} \xi_{n+1}^h - \xi_{\star}^h &= (1 - \gamma_{n+1}V_0''(\xi_{\star}^0))(\xi_n^h - \xi_{\star}^h) + \gamma_{n+1}(V_0''(\xi_{\star}^0) - V_h''(\xi_{\star}^h))(\xi_n^h - \xi_{\star}^h) \\ &\quad + \gamma_{n+1}r_{n+1}^h + \gamma_{n+1}\rho_{n+1}^h + \gamma_{n+1}e_{n+1}^h, \end{aligned} \quad (\text{H.1})$$

where

$$\begin{aligned} r_{n+1}^h &:= V_h''(\xi_{\star}^h)(\xi_n^h - \xi_{\star}^h) - V_h'(\xi_n^h), \\ \rho_{n+1}^h &:= V_h'(\xi_n^h) - V_h'(\xi_{\star}^h) - (H_1(\xi_n^h, X_h^{(n+1)}) - H_1(\xi_{\star}^h, X_h^{(n+1)})), \\ e_{n+1}^h &:= V_h'(\xi_{\star}^h) - H_1(\xi_{\star}^h, X_h^{(n+1)}). \end{aligned}$$

Iterating n times (H.1),

$$\begin{aligned} \xi_n^h - \xi_{\star}^h &= (\xi_0^h - \xi_{\star}^h)\Pi_{1:n} + \sum_{k=1}^n \gamma_k \Pi_{k+1:n} (V_0''(\xi_{\star}^0) - V_h''(\xi_{\star}^h))(\xi_k^h - \xi_{\star}^h) \\ &\quad + \sum_{k=1}^n \gamma_k \Pi_{k+1:n} r_k^h + \sum_{k=1}^n \gamma_k \Pi_{k+1:n} \rho_k^h + \sum_{k=1}^n \gamma_k \Pi_{k+1:n} e_k^h, \end{aligned} \quad (\text{H.2})$$

where, for two positive integers i and n ,

$$\Pi_{i:n} := \prod_{j=i}^n (1 - \gamma_j V_0''(\xi_{\star}^0)),$$

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

Using the inequality $1 + x \leq e^x$, $x \in \mathbb{R}$, a comparison between series and integrals and (D.2),

$$\begin{aligned} |\Pi_{i:n}| &\leq \exp\left(-V_0''(\xi_{\star}^0) \sum_{j=i}^n \gamma_j\right) \\ &\leq \begin{cases} \exp(\gamma_1 V_0''(\xi_{\star}^0)) i^{\gamma_1 V_0''(\xi_{\star}^0)} (n+1)^{-\gamma_1 V_0''(\xi_{\star}^0)}, & \beta = 1, \\ \exp(-\gamma_1 V_0''(\xi_{\star}^0)) (\varphi_{1-\beta}(n+1) - \varphi_{1-\beta}(i)), & \beta \in (0, 1). \end{cases} \end{aligned} \quad (\text{H.3})$$

Since $\gamma_n = \gamma_1 n^{-\beta}$, $\beta \in (0, 1]$ with $\gamma_1 V_0''(\xi_{\star}^0) > \gamma_1 \bar{\lambda}_2 > 1$ if $\beta = 1$, using (H.3) and a comparison between series and integrals (with computations similar to those performed in the proof of Lemma D.2), we deduce that for any $b \geq 0$ and $a > 0$ such that $\gamma_1 a V_0''(\xi_{\star}^0) > b$ if $\beta = 1$, there exists $K < \infty$ such that for any integer n

$$\sum_{k=1}^n \gamma_k^{1+b} |\Pi_{k+1:n}|^a \leq K \gamma_n^b, \quad (\text{H.4})$$

as well as

$$|\Pi_{1:n}| \leq K \gamma_n. \quad (\text{H.5})$$

Via (H.2), for $\ell \in \{1, \dots, L\}$,

$$\xi_n^{h_\ell} - \xi_{\star}^{h_\ell} - (\xi_n^{h_{\ell-1}} - \xi_{\star}^{h_{\ell-1}}) = A_n^\ell + B_n^\ell + C_n^\ell + D_n^\ell + E_n^\ell, \quad (\text{H.6})$$

where

$$\begin{aligned}
A_n^\ell &:= (\xi_0^{h_\ell} - \xi_\star^{h_\ell} - (\xi_0^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}}))\Pi_{1:n}, \\
B_n^\ell &:= \sum_{k=1}^n \gamma_k \Pi_{k+1:n} \{ (V_0''(\xi_\star^0) - V_{h_\ell}''(\xi_\star^{h_\ell}))(\xi_k^{h_\ell} - \xi_\star^{h_\ell}) \\
&\quad - (V_0''(\xi_\star^0) - V_{h_{\ell-1}}''(\xi_\star^{h_{\ell-1}}))(\xi_k^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}}) \}, \\
C_n^\ell &:= \sum_{k=1}^n \gamma_k \Pi_{k+1:n} (r_k^{h_\ell} - r_k^{h_{\ell-1}}), \\
D_n^\ell &:= \sum_{k=1}^n \gamma_k \Pi_{k+1:n} (\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}}), \\
E_n^\ell &:= \sum_{k=1}^n \gamma_k \Pi_{k+1:n} (e_k^{h_\ell} - e_k^{h_{\ell-1}}).
\end{aligned}$$

Hence, the difference between the multilevel SA estimator $\xi_{\mathbf{N}}^{\text{ML}}$ of the VaR in (3.1) and the solution $\xi_\star^{h_L} = \text{argmin } V_{h_L}$ can be decomposed into

$$\begin{aligned}
\xi_{\mathbf{N}}^{\text{ML}} - \xi_\star^{h_L} &= \xi_{N_0}^{h_0} - \xi_\star^{h_0} + \sum_{\ell=1}^L \xi_{N_\ell}^{h_\ell} - \xi_\star^{h_\ell} - (\xi_{N_\ell}^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}}) \\
&= \xi_{N_0}^{h_0} - \xi_\star^{h_0} + \sum_{\ell=1}^L A_{N_\ell}^\ell + B_{N_\ell}^\ell + C_{N_\ell}^\ell + D_{N_\ell}^\ell + E_{N_\ell}^\ell.
\end{aligned} \tag{H.7}$$

We now quantify the contribution of each term appearing in the decomposition (H.7).

Step 1.1. Study of $\xi_{N_0}^{h_0} - \xi_\star^{h_0}$.

Theorem 2.7(i) guarantees

$$\mathbb{E}[(\xi_{N_0}^{h_0} - \xi_\star^{h_0})^2] \leq K\gamma_{N_0}.$$

Step 1.2. Study of $\sum_{\ell=1}^L A_{N_\ell}^\ell$.

From (H.5),

$$\mathbb{E}[|A_n^\ell|^2] \leq 2K \sup_{\ell \geq 0} \mathbb{E}[|\xi_0^{h_\ell} - \xi_\star^{h_\ell}|^2] \gamma_n^2 \leq K\gamma_n^2. \tag{H.8}$$

Hence, by the triangle inequality,

$$\mathbb{E} \left[\left(\sum_{\ell=1}^L A_{N_\ell}^\ell \right)^2 \right]^{\frac{1}{2}} \leq K \sum_{\ell=1}^L \gamma_{N_\ell}.$$

Step 1.3. Study of $\sum_{\ell=1}^L B_{N_\ell}^\ell$.

By Lemma 2.1, $(\xi_\star^{h_\ell})_{\ell \geq 0}$ is bounded. Let \mathcal{K} be a compact set of \mathbb{R} containing the sequence $(\xi_\star^{h_\ell})_{\ell \geq 0}$. For any $\ell \geq 0$,

$$\begin{aligned}
|V_0''(\xi_\star^0) - V_{h_\ell}''(\xi_\star^{h_\ell})| &\leq \frac{1}{1-\alpha} (|f_{X_0}(\xi_\star^0) - f_{X_0}(\xi_\star^{h_\ell})| + |f_{X_0}(\xi_\star^{h_\ell}) - f_{X_{h_\ell}}(\xi_\star^{h_\ell})|) \\
&\leq \frac{1}{1-\alpha} \left([f_{X_0}]_{\text{Lip}} |\xi_\star^{h_\ell} - \xi_\star^0| + \sup_{\xi \in \mathcal{K}} |f_{X_0}(\xi) - f_{X_{h_\ell}}(\xi)| \right) \\
&\leq K(h_\ell + h_\ell^{\frac{1}{4}+\delta}) \leq Kh_\ell^{(\frac{1}{4}+\delta) \wedge 1},
\end{aligned}$$

where we used Assumptions 2.5(ii) and 3.4 together with Proposition 2.4. Using this estimate, Theorem 2.7(i) and (H.4),

$$\begin{aligned} \mathbb{E}[|B_n^\ell|^2]^{\frac{1}{2}} &\leq 2(|V_0''(\xi_\star^0) - V_{h_\ell}''(\xi_\star^{h_\ell})| \vee |V_0''(\xi_\star^0) - V_{h_{\ell-1}}''(\xi_\star^{h_{\ell-1}})|) \\ &\quad \times \left(\sum_{k=1}^n \gamma_k |\Pi_{k+1:n}| (\mathbb{E}[|\xi_k^{h_\ell} - \xi_\star^{h_\ell}|^2]^{\frac{1}{2}} \vee \mathbb{E}[|\xi_k^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}}|^2]^{\frac{1}{2}}) \right) \\ &\leq K h_\ell^{(\frac{1}{4}+\delta)\wedge 1} \sum_{k=1}^n \gamma_k^{\frac{3}{2}} |\Pi_{k+1:n}| \leq K \gamma_n^{\frac{1}{2}} h_\ell^{(\frac{1}{4}+\delta)\wedge 1}. \end{aligned} \quad (\text{H.9})$$

Therefore, by the triangle inequality,

$$\mathbb{E} \left[\left(\sum_{\ell=1}^L B_{N_\ell}^\ell \right)^2 \right]^{\frac{1}{2}} \leq K \sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{1}{2}} h_\ell^{(\frac{1}{4}+\delta)\wedge 1}.$$

Step 1.4. Study of $\sum_{\ell=1}^L C_{N_\ell}^\ell$.

Using a first order Taylor expansion, the uniform Lipschitz regularity of $(f_{X_h})_{h \in \mathcal{H}}$ under Assumption 2.5(ii) and Theorem 2.7(ii),

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

Hence, reusing (H.4),

$$\mathbb{E}[|C_n^\ell|^2]^{\frac{1}{2}} \leq \sum_{k=1}^n \gamma_k |\Pi_{k+1:n}| (\mathbb{E}[|r_k^{h_\ell}|^2]^{\frac{1}{2}} + \mathbb{E}[|r_k^{h_{\ell-1}}|^2]^{\frac{1}{2}}) \leq K \sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}| \leq K \gamma_n. \quad (\text{H.10})$$

Thus, by the triangle inequality,

$$\mathbb{E} \left[\left(\sum_{\ell=1}^L C_{N_\ell}^\ell \right)^2 \right]^{\frac{1}{2}} \leq K \sum_{\ell=1}^L \gamma_{N_\ell}.$$

Step 1.5. Study of $\sum_{\ell=1}^L D_{N_\ell}^\ell$.

By the definition (1.4) of H_1 and (2.5)

$$\begin{aligned} \mathbb{E}[(H_1(\xi_k^h, X_h^{(k+1)}) - H_1(\xi_\star^h, X_h^{(k+1)}))^2] &= \frac{1}{(1-\alpha)^2} \mathbb{E}[|\mathbb{1}_{\{X_h^{(k+1)} > \xi_k^h\}} - \mathbb{1}_{\{X_h^{(k+1)} > \xi_\star^h\}}|] \\ &= \frac{1}{(1-\alpha)^2} \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{\xi_k^h < X_h^{(k+1)} < \xi_\star^h\}} + \mathbb{1}_{\{\xi_\star^h < X_h^{(k+1)} < \xi_k^h\}} | \mathcal{F}_k^h]] \\ &= \frac{1}{(1-\alpha)^2} \mathbb{E}[|F_{X_h}(\xi_k^h) - F_{X_h}(\xi_\star^h)|] \leq \frac{\sup_{h \in \mathcal{H}} \|f_{X_h}\|_\infty}{(1-\alpha)^2} \mathbb{E}[(\xi_k^h - \xi_\star^h)^2]^{\frac{1}{2}} \\ &\leq K \frac{\sup_{h \in \mathcal{H}} \|f_{X_h}\|_\infty}{(1-\alpha)^2} \gamma_k^{\frac{1}{2}}. \end{aligned} \quad (\text{H.11})$$

Observe that $(\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}})_{k \geq 1}$ is a sequence of $(\mathcal{F}^{h_\ell}, \mathbb{P})$ -martingale increments. Thus, recalling

that $2\gamma_1 V_0''(\xi_\star^0) > \frac{3}{2}$, by (H.11) and (H.4),

$$\begin{aligned}
\mathbb{E}[|D_n^\ell|^2] &= \mathbb{E}\left[\left(\sum_{k=1}^n \gamma_k \Pi_{k+1:n}(\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}})\right)^2\right] \\
&= \sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}|^2 \mathbb{E}[(\rho_k^{h_\ell} - \rho_k^{h_{\ell-1}})^2] \\
&\leq 2 \sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}|^2 (\mathbb{E}[|\rho_k^{h_\ell}|^2] + \mathbb{E}[|\rho_k^{h_{\ell-1}}|^2]) \\
&\leq 2 \sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}|^2 \left(\mathbb{E}[(H_1(\xi_k^{h_\ell}, X_{h_\ell}^{(k+1)}) - H_1(\xi_\star^{h_\ell}, X_{h_\ell}^{(k+1)}))^2] \right. \\
&\quad \left. + \mathbb{E}[(H_1(\xi_k^{h_{\ell-1}}, X_{h_{\ell-1}}^{(k+1)}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}}^{(k+1)})^2] \right) \\
&\leq K \sum_{k=1}^n \gamma_k^{\frac{5}{2}} |\Pi_{k+1:n}|^2 \leq K \gamma_n^{\frac{3}{2}}.
\end{aligned} \tag{H.12}$$

Hence, since the random variables $(D_n^\ell)_{1 \leq \ell \leq L}$ are independent and centered, the above estimate yields

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

Step 1.6. Study of $\sum_{\ell=1}^L E_{N_\ell}^\ell$.

Note again that $(e_k^{h_\ell} - e_k^{h_{\ell-1}})_{k \geq 1}$ is a sequence of $(\mathcal{F}^{h_\ell}, \mathbb{P})$ -martingale increments, so that

$$\begin{aligned}
\mathbb{E}[|E_n^\ell|^2] &= \mathbb{E}\left[\left(\sum_{k=1}^n \gamma_k \Pi_{k+1:n} (e_k^{h_\ell} - e_k^{h_{\ell-1}})\right)^2\right] = \sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}|^2 \mathbb{E}[(e_k^{h_\ell} - e_k^{h_{\ell-1}})^2] \\
&\leq \sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}|^2 \mathbb{E}[(H_1(\xi_\star^{h_\ell}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}}))^2].
\end{aligned} \tag{H.13}$$

The last term on the right hand side above can be upper bounded by

$$\begin{aligned}
&\mathbb{E}[(H_1(\xi_\star^{h_\ell}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}}))^2] \\
&\leq 2(\mathbb{E}[(H_1(\xi_\star^{h_\ell}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_\ell}))^2] \\
&\quad + \mathbb{E}[(H_1(\xi_\star^{h_{\ell-1}}, X_{h_\ell}) - H_1(\xi_\star^{h_{\ell-1}}, X_{h_{\ell-1}}))^2]) \\
&\leq \frac{2}{(1-\alpha)^2} (\mathbb{E}[(\mathbb{1}_{\{X_{h_\ell} > \xi_\star^{h_\ell}\}} - \mathbb{1}_{\{X_{h_\ell} > \xi_\star^{h_{\ell-1}}\}})^2] \\
&\quad + \mathbb{E}[(\mathbb{1}_{\{X_{h_\ell} > \xi_\star^{h_{\ell-1}}\}} - \mathbb{1}_{\{X_{h_{\ell-1}} > \xi_\star^{h_{\ell-1}}\}})^2]).
\end{aligned} \tag{H.14}$$

On the one hand, by Proposition 2.4,

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

On the other hand, Proposition 3.2 yields

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

where K_{h_ℓ} is a constant satisfying $\sup_{\ell \geq 1} K_{h_\ell} < \infty$, and $\epsilon(h_\ell)$ is defined in (3.9). Combining the two previous estimates, recalling (H.13), (H.14) and (H.4),

$$\begin{aligned} \mathbb{E}[|E_n^\ell|^2] &= \mathbb{E}\left[\left(\sum_{k=1}^n \gamma_k \Pi_{k+1:n}(e_k^{h_\ell} - e_k^{h_{\ell-1}})\right)^2\right] \\ &\leq K(h_\ell + \epsilon(h_\ell)) \sum_{k=1}^n \gamma_k^2 |\Pi_{k+1:n}|^2 \leq K\epsilon(h_\ell)\gamma_n. \end{aligned} \quad (\text{H.15})$$

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

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

Step 1.7. Conclusion.

Gathering the previous estimates on each term of (H.7),

$$\mathbb{E}[(\xi_{\mathbf{N}}^{\text{ML}} - \xi_\star^{h_L})^2] \leq K \left(\gamma_{N_0} + \left(\sum_{\ell=1}^L \gamma_{N_\ell} + \gamma_{N_\ell}^{\frac{1}{2}} h_\ell^{(\frac{1}{4} + \delta) \wedge 1} \right)^2 + \sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{3}{2}} + \gamma_{N_\ell} \epsilon(h_\ell) \right). \quad (\text{H.16})$$

From the Cauchy-Schwarz inequality and the fact that $h_\ell^{\frac{1}{2}} \leq K\epsilon(h_\ell)$,

$$\sum_{\ell=1}^L \gamma_{N_\ell}^{\frac{1}{2}} h_\ell^{(\frac{1}{4} + \delta) \wedge 1} \leq \left(\sum_{\ell=1}^L \gamma_{N_\ell} h_\ell^{\frac{1}{2}} \right)^{\frac{1}{2}} \left(\sum_{\ell=1}^L h_\ell^{2(\delta \wedge \frac{3}{4})} \right)^{\frac{1}{2}} \leq K \left(\sum_{\ell=1}^L \gamma_{N_\ell} \epsilon(h_\ell) \right)^{\frac{1}{2}}, \quad (\text{H.17})$$

where we used that $\sum_{\ell=1}^\infty h_\ell^{2(\delta \wedge \frac{3}{4})} = \frac{h_0^{2(\delta \wedge \frac{3}{4})}}{M^{2(\delta \wedge \frac{3}{4}) - 1}} < \infty$. Plugging (H.17) into (H.16) concludes the proof of (3.7).

Step 2. Study of $\chi_{\mathbf{N}}^{\text{ML}} - \chi_\star^{h_L}$.

For the sake of simplicity, we assume $\chi_0^{h_\ell} = 0$, $\ell \in \{0, \dots, L\}$. The general case follows from similar arguments. Recalling the definition (D.24), we derive from the decomposition (D.23)

$$\begin{aligned} \chi_n^{h_\ell} - \chi_\star^{h_\ell} - (\chi_n^{h_{\ell-1}} - \chi_\star^{h_{\ell-1}}) \\ = \frac{1}{n} \sum_{k=1}^n \varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}} + \frac{1}{n} \sum_{k=1}^n V_{h_\ell}(\xi_{k-1}^{h_\ell}) - V_{h_\ell}(\xi_\star^{h_\ell}) - (V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}}) - V_{h_{\ell-1}}(\xi_\star^{h_{\ell-1}})). \end{aligned}$$

Thus

$$\begin{aligned} \chi_{\mathbf{N}}^{\text{ML}} - \chi_\star^{h_L} &= \chi_{N_0}^{h_0} - \chi_\star^{h_0} + \sum_{\ell=1}^L \chi_{N_\ell}^{h_\ell} - \chi_\star^{h_\ell} - (\chi_{N_\ell}^{h_{\ell-1}} - \chi_\star^{h_{\ell-1}}) \\ &= \chi_{N_0}^{h_0} - \chi_\star^{h_0} + \sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}} \\ &\quad + \sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} V_{h_\ell}(\xi_{k-1}^{h_\ell}) - V_{h_\ell}(\xi_\star^{h_\ell}) - (V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}}) - V_{h_{\ell-1}}(\xi_\star^{h_{\ell-1}})). \end{aligned} \quad (\text{H.18})$$

Step 2.1. Study of $\chi_{N_0}^{h_0} - \chi_\star^{h_0}$.

By Theorem 2.7(ii),

$$\mathbb{E}[(\chi_{N_0}^{h_0} - \chi_\star^{h_0})^2] \leq \frac{K}{N_0^{1 \wedge 2\beta}}. \quad (\text{H.19})$$

Step 2.2. Study of $\sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}}$.

Via the definition (D.24), the 1-Lipschitz property of $x \mapsto x^+$, $x \in \mathbb{R}$, the fact that $\mathbb{E}[(X_{h_\ell} - X_{h_{\ell-1}})^2] \leq Kh_\ell$ by (G.4) with $p = 2$, and the fact that $|\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}| \leq Kh_\ell$ by Proposition 2.4,

$$\begin{aligned} \mathbb{E}[(\varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}})^2 | \mathcal{F}_{k-1}] &= \frac{1}{(1-\alpha)^2} \text{Var}((X_{h_\ell}^{(k)} - \xi_{k-1}^{h_\ell})^+ - (X_{h_{\ell-1}}^{(k)} - \xi_{k-1}^{h_{\ell-1}})^+ | \mathcal{F}_{k-1}) \\ &\leq \frac{1}{(1-\alpha)^2} \mathbb{E}[(X_{h_\ell}^{(k)} - \xi_{k-1}^{h_\ell} - (X_{h_{\ell-1}}^{(k)} - \xi_{k-1}^{h_{\ell-1}}))^2 | \mathcal{F}_{k-1}] \\ &\leq \frac{3}{(1-\alpha)^2} \left(\mathbb{E}[(X_{h_\ell} - X_{h_{\ell-1}})^2] + (\xi_{k-1}^{h_\ell} - \xi_{k-1}^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}))^2 + (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}})^2 \right) \\ &\leq K \left(h_\ell + (\xi_{k-1}^{h_\ell} - \xi_{k-1}^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}))^2 \right). \end{aligned} \tag{H.20}$$

We seek an $L^2(\mathbb{P})$ -estimate for the quantity $(\xi_k^{h_\ell} - \xi_k^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}))$. By the decomposition (H.6), using (H.8), (H.9), (H.10), (H.12) and (H.15), we obtain

$$\mathbb{E}[(\xi_k^{h_\ell} - \xi_k^{h_{\ell-1}} - (\xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}}))^2] \leq K(\gamma_k^{\frac{3}{2}} + \gamma_k \epsilon(h_\ell)). \tag{H.21}$$

Recall that, for each $\ell \in \{1, \dots, L\}$, $(e_k^{h_\ell} - e_k^{h_{\ell-1}})_{1 \leq k \leq n}$ are $(\mathcal{F}^{h_\ell}, \mathbb{P})$ -martingale increments. Hence, taking the expectation on both sides of the inequality (H.20), using (H.21) and the fact that, by assumption and Lemma 2.1, $\sup_{\ell \geq 0} \mathbb{E}[|\xi_0^{h_\ell}|] + |\xi_\star^{h_\ell}| < \infty$,

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

Thus, the random variables $(\frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}})_{1 \leq \ell \leq L}$ being independent and centered,

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}} \right)^2 \right] &= \sum_{\ell=1}^L \mathbb{E} \left[\left(\frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}} \right)^2 \right] \\ &\leq K \sum_{\ell=1}^L \left(\frac{h_\ell}{N_\ell} + \frac{1}{N_\ell^2} \sum_{k=1}^{N_\ell} \gamma_k^{\frac{3}{2}} + \gamma_k \epsilon(h_\ell) \right). \end{aligned} \tag{H.22}$$

Step 2.3. Study of

$$\sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} V_{h_\ell}(\xi_{k-1}^{h_\ell}) - V_{h_\ell}(\xi_\star^{h_\ell}) - (V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}}) - V_{h_{\ell-1}}(\xi_\star^{h_{\ell-1}})).$$

We deal now with the last term on the right hand side of (H.18). Using Minkowski's inequality, a second order Taylor expansions of V_{h_ℓ} and $V_{h_{\ell-1}}$, that $V_{h_\ell}'(\xi_\star^{h_\ell}) = 0$, that $\sup_{h \in \mathcal{H}} \|V_h''\|_\infty < \infty$ by Assumption 2.5(ii), (2.6) and that $\sum_{k \geq 1} \gamma_k = \infty$,

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{n} \sum_{k=1}^n V_{h_\ell}(\xi_{k-1}^{h_\ell}) - V_{h_\ell}(\xi_\star^{h_\ell}) - (V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}}) - V_{h_{\ell-1}}(\xi_\star^{h_{\ell-1}})) \right)^2 \right]^{\frac{1}{2}} \\ \leq K \left(\frac{1}{n} + \frac{1}{n-1} \sum_{k=2}^n \mathbb{E}[(\xi_{k-1}^{h_\ell} - \xi_\star^{h_\ell})^4]^{\frac{1}{2}} + \mathbb{E}[(\xi_{k-1}^{h_{\ell-1}} - \xi_\star^{h_{\ell-1}})^4]^{\frac{1}{2}} \right) \\ \leq K \left(\frac{1}{n} + \frac{1}{n-1} \sum_{k=1}^{n-1} \gamma_k \right) \leq K \bar{\gamma}_n. \end{aligned}$$

Thus

$$\mathbb{E} \left[\left(\sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} V_{h_\ell}(\xi_{k-1}^{h_\ell}) - V_{h_\ell}(\xi_{\star}^{h_\ell}) - (V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}}) - V_{h_{\ell-1}}(\xi_{\star}^{h_{\ell-1}})) \right)^2 \right]^{\frac{1}{2}} \leq K \sum_{\ell=1}^L \bar{\gamma}_{N_\ell}. \quad (\text{H.23})$$

Step 2.4. Conclusion.

Coming back to (H.18), by (H.19), (H.22) and (H.23),

$$\begin{aligned} \mathbb{E}[(\chi_{\mathbf{N}}^{\text{ML}} - \chi_{\star}^{h_L})^2] &\leq 3 \left(\mathbb{E}[(\chi_{N_0}^{h_0} - \chi_{\star}^{h_0})^2] + \mathbb{E} \left[\left(\sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} \varepsilon_k^{h_\ell} - \varepsilon_k^{h_{\ell-1}} \right)^2 \right] \right. \\ &\quad \left. + \mathbb{E} \left[\left(\sum_{\ell=1}^L \frac{1}{N_\ell} \sum_{k=1}^{N_\ell} V_{h_\ell}(\xi_{k-1}^{h_\ell}) - V_{h_\ell}(\xi_{\star}^{h_\ell}) - (V_{h_{\ell-1}}(\xi_{k-1}^{h_{\ell-1}}) - V_{h_{\ell-1}}(\xi_{\star}^{h_{\ell-1}})) \right)^2 \right] \right) \\ &\leq K \left(\sum_{\ell=1}^L \frac{h_\ell}{N_\ell} + \frac{1}{N_\ell^2} \sum_{k=1}^{N_\ell} \gamma_k^{\frac{3}{2}} + \gamma_k \epsilon(h_\ell) + \left(\sum_{\ell=1}^L \bar{\gamma}_{N_\ell} \right)^2 \right). \end{aligned}$$

This concludes the proof of (3.8).

References

- [1] C. Acerbi and D. Tasche. “On the coherence of expected shortfall”. In: *Journal of Banking & Finance* 26.7 (2002), pp. 1487–1503.
- [2] C. Albanese, Y. Armenti, and S. Crépey. “XVA metrics for CCP optimization”. In: *Statistics & Risk Modeling* 37.1-2 (2020), pp. 25–53.
- [3] O. Bardou, N. Frikha, and G. Pagès. “Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling”. In: *Monte Carlo Methods and Applications* 15.3 (2009), pp. 173–210.
- [4] O. Bardou, N. Frikha, and G. Pagès. “CVaR hedging using quantization-based stochastic approximation algorithm”. In: *Mathematical Finance* 26.1 (2016), pp. 184–229.
- [5] O. Bardou, N. Frikha, and G. Pagès. “Recursive computation of value-at-risk and conditional value-at-risk using MC and QMC”. In: *Monte Carlo and Quasi-Monte Carlo Methods*. Springer Berlin Heidelberg, 2009, pp. 193–208.
- [6] D. Barrera, S. Crépey, B. Diallo, G. Fort, E. Gobet, and U. Staszynski. “Stochastic approximation schemes for economic capital and risk margin computations”. In: *ESAIM: Proceedings and Surveys* 65 (2019), pp. 182–218.
- [7] Basel Committee on Banking Supervision. *Consultative document: Fundamental Review of the Trading Book: A revised market risk framework*. 2013. URL: [bis.org/publ/bcbs265.pdf](https://www.bis.org/publ/bcbs265.pdf).
- [8] F. Bourgey, S. De Marco, E. Gobet, and A. Zhou. “Multilevel Monte Carlo methods and lower–upper bounds in initial margin computations”. In: *Monte Carlo Methods and Applications* 26.2 (2020), pp. 131–161.
- [9] M. Broadie, Y. Du, and C. Moallemi. “Risk estimation via regression”. In: *Operations Research* 63.5 (2015), pp. 1077–1097.
- [10] M. Costa and S. Gadat. “Non asymptotic controls on a recursive superquantile approximation”. In: *Electronic Journal of Statistics* 15.2 (2021), pp. 4718–4769.
- [11] S. Crépey, N. Frikha, A. Louzi, and G. Pagès. *Asymptotic error analysis of multilevel stochastic approximations for the value-at-risk and expected shortfall*. 2023. arXiv: 2311.15333.

- [12] S. Dereich and T. Müller-Gronbach. “General multilevel adaptations for stochastic approximation algorithms of Robbins–Monro and Polyak–Ruppert type”. In: *Numerische Mathematik* 142.2 (2019), pp. 279–328.
- [13] H. Föllmer and A. Schied. *Convex Risk Measures*. John Wiley & Sons, Ltd, 2010.
- [14] N. Frikha. “Multi-level stochastic approximation algorithms”. In: *The Annals of Applied Probability* 26.2 (2016), pp. 933–985.
- [15] N. Frikha. “Shortfall risk minimization in discrete time financial market models”. In: *SIAM Journal on Financial Mathematics* 5.1 (2014), pp. 384–414.
- [16] N. Frikha and L. Huang. “A multi-step Richardson–Romberg extrapolation method for stochastic approximation”. In: *Stochastic Processes and their Applications* 125.11 (2015), pp. 4066–4101.
- [17] N. Frikha and S. Menozzi. “Concentration bounds for stochastic approximations”. In: *Electronic Communications in Probability* 17 (2012), pp. 1–15.
- [18] M. B. Giles. “Multilevel Monte Carlo path simulation”. In: *Operations Research* 56.3 (2008), pp. 607–617.
- [19] M. B. Giles and A.-L. Haji-Ali. “Multilevel nested simulation for efficient risk estimation”. In: *SIAM/ASA Journal on Uncertainty Quantification* 7.2 (2019), pp. 497–525.
- [20] M. B. Giles, A.-L. Haji-Ali, and J. Spence. *Efficient risk estimation for the credit valuation adjustment*. 2024. arXiv: 2301.05886.
- [21] M. B. Giles and L. Szpruch. “Antithetic multilevel Monte Carlo estimation for multidimensional SDEs”. In: *Monte Carlo and Quasi-Monte Carlo Methods 2012*. Springer Berlin Heidelberg, 2013, pp. 367–384.
- [22] D. Giorgi, V. Lemaire, and G. Pagès. “Weak error for nested multilevel Monte Carlo”. In: *Methodology and Computing in Applied Probability* 22 (3 2020), pp. 1325–1348.
- [23] M. B. Gordy and S. Juneja. “Nested simulation in portfolio risk measurement”. In: *Management Science* 56.10 (2010), pp. 1833–1848.
- [24] A.-L. Haji-Ali, J. Spence, and A. L. Teckentrup. “Adaptive multilevel Monte Carlo for probabilities”. In: *SIAM Journal on Numerical Analysis* 60.4 (2022), pp. 2125–2149.
- [25] R. T. Rockafellar and S. Uryasev. “Optimization of conditional value-at-risk”. In: *Journal of Risk* 2.3 (2000), pp. 21–41.
- [26] Z. Xu, Z. He, and X. Wang. “Efficient risk estimation via nested multilevel quasi-Monte Carlo simulation”. In: *Journal of Computational and Applied Mathematics* 443 (2024), p. 115745.