MATHEMATICAL FINANCE

An International Journal of Mathematics, Statistics and Financial Economics

Mathematical Finance, Vol. 26, No. 1 (January 2016), 184–229

CVaR HEDGING USING QUANTIZATION-BASED STOCHASTIC APPROXIMATION ALGORITHM

O. BARDOU

GDF Suez Finance Division, Risk Department

N. FRIKHA

LPMA, Université Paris-Diderot

G. PAGÈS

LPMA, Université Pierre & Marie Curie

In this paper, we investigate a method based on risk minimization to hedge observable but nontradable source of risk on financial or energy markets. The optimal portfolio strategy is obtained by minimizing dynamically the conditional value-at-risk (CVaR) using three main tools: a stochastic approximation algorithm, optimal quantization, and variance reduction techniques (importance sampling and linear control variable), as the quantities of interest are naturally related to rare events. As a first step, we investigate the problem of CVaR regression, which corresponds to a static portfolio strategy where the number of units of each tradable assets is fixed at time 0 and remains unchanged till maturity. We devise a stochastic approximation algorithm and study its a.s. convergence and weak convergence rate. Then, we extend our approach to the dynamic case under the assumption that the process modeling the nontradable source of risk and financial assets prices is Markovian. Finally, we illustrate our approach by considering several portfolios in connection with energy markets.

KEY WORDS: VaR, CVaR, stochastic approximation, Robbins–Monro algorithm, quantification.

1. INTRODUCTION

It is well known that in a complete financial market, an investor faced with a contingent claim can hedge perfectly on a finite horizon time T without any risk. However, from a practical standpoint, an agent needs to have a more realistic view of financial or energy markets which are intrinsically incomplete for many reasons: stochastic volatility, jumps, and more specifically the impact of the temperature in commodity prices in energy markets. Hence, there is no exact replication to provide a unique price. Thus, pricing and hedging contingent claims in such a framework require new approaches. One may still price and hedge using a superhedging criterion as studied in Rouge and El Karoui (2000). However, the price is often too high and, in practice, the trader can only hedge partially. Because of this, he often has to bear some risk of loss. An alternative is to develop a pricing theory under a martingale measure which corresponds to an optimized criterion.

We would like to thank Sandrine Bouthemy for fruitful and stimulating discussions.
Manuscript received January 2011; final revision received May 2013.
Address correspondence to N. Frikha, LPMA, Université Paris-Diderot, Bâtiment Sophie Germain, 5 rue Thomas Mann, 75025 Paris Cedex 13, France; e-mail: frikha@math.univ-paris-diderot.fr.

DOI: 10.1111/mafi.12049

For instance, one can refer to Föllmer and Schweizer (1991) for the minimal martingale measure and to Avellaneda (1998), Frittelli (2000), and Miyahara (2006) for the minimal entropy martingale measure among others.

Another widely used method to address this problem is based on expected utility maximization, such as utility indifference pricing. It consists of pricing a nonhedgeable claim so that the investor's utility remains unchanged between holding and not holding the contingent claim (see Hodges and Neuberger 1989; Karatzas et al. 1991; Rouge and El Karoui 2000; among many others). Although, this approach has been studied for a long time, its main drawback is that practitioners do not know their own utility function. Moreover, different agents may price and hedge a contingent claim differently according to their own risk preferences so that it has little acceptance in practice.

In this paper, we propose an alternative method based on risk minimization. To be more precise, we focus on minimizing dynamically the conditional value-at-risk (CVaR). The CVaR is strongly linked to the value-at-risk (VaR), which is certainly the most widely used risk measure in the practice of risk management. By definition, the VaR at level $\alpha \in (0, 1)$ (VaR_α) of a given portfolio loss distribution is the lowest amount not exceeded by the loss with probability α (usually $\alpha \in [0.95, 1)$). The CVaR at level α (CVaR_α) is the conditional expectation of the portfolio losses beyond the VaR_α level. Compared to VaR, the CVaR is known to have better properties. Risk measures of this type were introduced in Artzner et al. (1999) and have been shown to share basic coherence properties like subadditivity, which pleads for portfolio diversification by contrast with the VaR_α. The extension to general convex risk measures was extensively studied in Föllmer and Schied (2002).

Pricing and hedging using risk measures is a recent approach that has been investigated by many authors. Barrieu and El Karoui (2004) developed a risk minimization problem to hedge nontradable risks on financial market using convex risk measures. The hedging strategy which maximizes the probability of successful hedge is studied in Föllmer and Leukert (1999) as an alternative to superhedging strategies which require a large amount of initial capital.

The VaR and the CVaR are asymmetric risk measures unlike standard deviation. By CVaR hedging a loss distribution L, we mean that we aim at modifying its shape in order to reduce the right-hand side of the distribution, which corresponds to larger losses whereas the left-hand side corresponds to small losses or potential gains. That is the main difference between CVaR hedging and hedging by means of a quadratic criterion as developed in Föllmer and Sonderman (1986) and Schweizer (1991), among others.

With regards to the numerical aspects of VaR and CVaR computations, a linear programming approach has been developed in Rockafellar and Uryasev (2000) to optimize a portfolio based on a CVaR criterion. Portfolio strategies with a low CVaR necessarily have a low VaR. The method first consists of generating loss scenarios and introducing them as constraints in the linear programming problem. Its main drawback is that the dimension (number of constraints) of the linear programming problem to be solved is equal to the number of simulated scenarios so that this approach turns out to have strong limitations in practice. We will propose an alternative method based on a stochastic approximation algorithm, which is not limited by the number of simulated scenarios.

Let us first describe the discrete time financial market model with finite horizon on which we will work. We consider a filtered probability space $(\Omega, \mathcal{G}, (\mathcal{G}_\ell)_{0\leq\ell\leq M}, \mathbb{P})$, where $\mathcal{G}_N = \mathcal{G}$ and (for convenience) we set $\mathcal{G}_0 = \{\varnothing, \Omega\}$. On this space, d ($d \in \mathbb{N}$) risky assets are available for trade with (discounted) price process $X = (X^1, \ldots, X^d)$, where, at every time $\ell \geq 0$, $X^i_\ell \geq 0$, $i = 1, \ldots, d$ and X is adapted, that is, X_ℓ is $\mathcal{G}_\ell$-measurable. For simplicity, we assume that the risk-free rate is equal to zero. The portfolio loss (or the payoff of a financial instrument) with maturity T is described by an $\mathbb{R}$-valued random variable L defined on $(\Omega, \mathcal{G}, \mathbb{P})$.

One possible motivation of our study can be described as follows. The portfolio loss L may depend upon an adapted process $Z = (Z_\ell)_{0\leq \ell \leq M}$, which is *observable* but *not available* for trade. Thus, the market is incomplete since it induces a nonhedgeable exogenous source of risk. Typically, on the electricity market, the loss L suffered by an energy company may be due to an abnormal annual electricity (or gas) consumption. This consumption depends on the temperature, which is an *observable* but *nontradable* source of risk. In this market, the temperature can be modeled by a process Z that may influence not only the loss but also electricity, gas spot prices, as well as and possibly, but to a much lesser extent, forward contracts, which are the only assets available for trade on energy markets. More generally, this kind of dependence with respect to an observable but nontradable risk is a particularly relevant source of incompleteness in financial and energy markets. We refer to Benth et al. (2008) and Geman (2005) for a thorough introduction to energy markets. In this particular situation, the observable information at time ℓ is given by $\mathcal{G}_\ell = \sigma\{X_i, Z_i; 0 \leq i \leq \ell\}$.

In order to reduce his/her risk (or hedge the contingent claim), the holder of the portfolio uses a dynamic self-financing strategy represented by an $\mathbb{R}^d$-valued process $\theta = \left(\theta_\ell^1, \ldots, \theta_\ell^d\right)_{1\leq \ell\leq M}$, which is $\mathcal{G}$-predictable, that is, θ_ℓ is $\mathcal{G}_{\ell-1}$-measurable, $\ell = 1, \ldots, M$. In such a strategy, θ_ℓ is the number of units of risky assets held by the investor during $(l-1, l]$, $l = 1, \ldots, M$. In what follows, for $d \in \mathbb{N}$, Θ_ℓ^d will denote the space of all $\mathcal{G}_\ell$-measurable and $\mathbb{P}$ a.s. finite random variables with values in $\mathbb{R}^d$. The gain resulting from a trading strategy θ with an initial investment of 0 is given by the discrete stochastic integral $\sum_{\ell=1}^M \theta_\ell.\Delta X_\ell$, where $\Delta X_\ell = X_\ell - X_{\ell-1}$.

The basic problem for the holder of the portfolio is to find an optimal self-financed strategy θ^*, which minimizes the residual risk of portfolio losses among all self-financed strategies, i.e., the solution to the following minimization problem[1]

$$\inf_{\theta\in\mathcal{A}_\mathcal{G}} \mathrm{CVaR}_\alpha\left(L - \sum_{\ell=1}^M \theta_\ell.\Delta X_\ell\right), \tag{1.1}$$

where $\mathcal{A}_\mathcal{G} = \{\theta = (\theta_\ell)_{1\leq\ell\leq M} : \theta_\ell \in \Theta_{\ell-1}^d,\ \ell = 1, \ldots, M\}$ is the set of admissible strategies.

This approach is very natural from an economic perspective, especially to managers who are concerned by controlling the static risk since their objective is to control the residual loss beyond the critical area defined by the VaR_α. However, the fact that the CVaR_α is known not to be a time-consistent risk measure (in the sense of Geman and Ohana 2008, see also Boda and Filar 2006; Shapiro 2009; Street, Rudloff, and Valladao 2011), may induce numerical problems when the number M of trading dates increases.

A natural question that arises is then to measure dynamically the risk of the considered portfolio in this context. To measure the risk at a given time ℓ, it is natural to require that investors, who want to measure their risk by using a CVaR criterion, take into account the information available at time ℓ, i.e., $\mathcal{G}_\ell$. This naturally leads to the introduction of a conditional risk measure that will be denoted by $\mathcal{G}_\ell$-CVaR (see Section 2.2 for more details), which is a straightforward extension of the CVaR suggested by Rockafellar and Uryasev's static representation of the CVaR.

[1]We consider the general definition of expectation of a random variable Y, i.e., the quantity $\mathbb{E}[Y]$ exists in $\overline{\mathbb{R}}$ as soon as $\mathbb{E}[Y_+] < +\infty$ or $\mathbb{E}[Y_-] < +\infty$.

For various reasons, such as transaction costs and storing constraints of energy, the holder of the portfolio may not want to trade every day being only interested by a rough hedge to reduce his risk. Consequently, we first investigate one-step self-financed strategies. Initiated at time ℓ_0, such a strategy is obtained by setting $\theta_\ell \equiv \theta_{\ell_0+1}$, for $\ell = \ell_0 + 1, \dots, M$. Consequently, a one-step portfolio strategy decided at time ℓ_0 is an $\mathbb{R}^d$-valued random variable $\theta_{\ell_0+1} \in \Theta^d_{\ell_0}$. The investor risk at time ℓ_0 can be measured by the quantity $\mathcal{G}_{\ell_0}$-CVaR$_\alpha(L - \theta_{\ell_0+1}.(X_M - X_{\ell_0}))$,[2] which is only known at time ℓ_0. However, the investor can estimate this quantity at time 0 by numerically computing $\mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-CVaR}_\alpha\left(L - \theta_{\ell_0+1}.\left(X_M - X_{\ell_0}\right)\right)\right]$. This quantity measures a mean forward risk (posthedge), i.e., it is the best estimation at time 0 of the risk at time ℓ_0 while the quantity CVaR$_\alpha(L - \theta_{\ell_0+1}.(X_M - X_{\ell_0}))$ represents the risk at time 0. Consequently, this leads us to define two different optimization problems.

The first one is to minimize the mean forward risk, i.e., the expectation of the risk profile measured at time ℓ_0 of the portfolio losses using a self-financing one-step portfolio strategy starting from an initial wealth of 0, namely,

$$\inf_{\theta_{\ell_0+1} \in \Theta^d_{\ell_0}} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-CVaR}_\alpha\left(L - \theta_{\ell_0+1}.\left(X_M - X_{\ell_0}\right)\right)\right]. \tag{1.2}$$

The second problem consists in minimizing the risk measured at time 0 (i.e., using a static CVaR criterion) of the portfolio losses using a self-financing one-step portfolio strategy starting from an initial wealth of 0:

$$\inf_{\theta_{\ell_0+1} \in \Theta^d_{\ell_0}} \text{CVaR}_\alpha\left(L - \theta_{\ell_0+1}.\left(X_M - X_{\ell_0}\right)\right). \tag{1.3}$$

This second one-step stochastic control problem above arises for an investor who wants to minimize his *static risk at time* 0. The choice for an investor between (1.2) and (1.3) should be motivated by his requirement in terms of risk policy.

From a numerical point of view, we consider the incomplete discrete time market as described above and first propose a stochastic approximation algorithm to compute the optimal self-financing portfolio strategy θ^* solution of (1.3) and (1.2) (and both the VaR and the CVaR of the resulting portfolio) in a Markovian framework. In order to take into account the possible scenarios at time ℓ_0, we will rely on optimal quantization-based spatial discretization of the random vector (X_{ℓ_0}, Z_{ℓ_0}). It consists in replacing (X_{ℓ_0}, Z_{ℓ_0}) by a finitely supported discrete random variable which is close, with a known distribution (i.e., a grid Γ_{ℓ_0} of size N_{ℓ_0} and a weight vector). Each node of this grid describes a potential state of the random vector (X_{ℓ_0}, Z_{ℓ_0}). The solutions of (1.3) and (1.2) are then obtained by solving N_{ℓ_0} local convex optimization problems. As a second step, we propose a first stochastic approximation procedure to solve (1.1), based on a spatial discretization of the whole process $(X_\ell, Z_\ell)_{1 \le \ell \le M-1}$. However, when the number of trading dates M or the dimension of (X, Z) are too large (say $M \geq 10$, in practice, otherwise the algorithm behaves well), this first algorithm turns out to be numerically inefficient owing to the curse of dimensionality.

To circumvent this issue, we develop other procedures based on some upper bounds of the objective function of (1.1) in order to produce good suboptimal solutions and we compare them on numerical examples. To be more precise, our approach will be to

[2]We consider the general definition of conditional expectation of a random variable Y, i.e., the quantity $\mathbb{E}\left[Y|\mathcal{G}_{\ell_0}\right]$ as soon as $\mathbb{E}\left[Y_+|\mathcal{G}_{\ell_0}\right] < +\infty$ or $\mathbb{E}\left[Y_-|\mathcal{G}_{\ell_0}\right] < +\infty$.

devise closely connected (time-consistent) risk measures for which efficient and robust stochastic optimization procedures can be more easily designed. Typically, we propose (see Section 3.3) to compute an optimal strategy of the time-consistent multistage optimization problem

$$\sum_{\ell=1}^{M} \inf_{\theta_\ell \in \Theta^d_{\ell-1}} \mathbb{E}\left[\mathcal{F}_\ell\text{-CVaR}\left(\tilde{\Delta} L_\ell - \theta_\ell.\Delta X_\ell\right)\right],$$

with $\mathcal{F}_\ell = \mathcal{G}_\ell$ or $\mathcal{F}_\ell = \{\varnothing, \Omega\}$, $\tilde{\Delta} L_\ell = \mathbb{E}[.L|\mathcal{G}_\ell] - \mathbb{E}[L|\mathcal{G}_{\ell-1}]$ or $\tilde{\Delta} L_\ell = L_\ell - L_{\ell-1}$, $\ell = 1, \ldots, M$. We then evaluate the standard "residual" CVaR hedging risk induced by this strategy. We also propose a backward hedging strategy based on a dynamic programming principle using the $\mathcal{G}$-CVaR.

All proposed algorithms are inspired by Rockafellar and Uryasev's representation of the CVaR and optimal quantization-based spatial discretization schemes of the process $(X_\ell, Z_\ell)_{0\leq \ell \leq M}$. This leads us to devise a new stochastic algorithm "à la Robbins–Monro" (RM) to estimate the three quantities of interest (VaR_α, CVaR_α, and optimal strategy). This principle has already been successfully initiated in Bardou, Frikha, and Pagès (2009) to compute the VaR and CVaR in a static framework. The main innovation of our computational and numerical approach compared to Bardou et al. (2009) lies in the dynamic feature of the portfolio selection problem, which induces a spatial discretization and naturally leads to, first, devising several local optimization algorithms, and then to merging them in order to solve the global optimization problem. The estimator provided by the (local) algorithm satisfies a Gaussian Central Limit Theorem (CLT). However, the proposed algorithm is just a first step toward an efficient procedure.

When α is close to 1 (otherwise the former "naive" procedure behaves well), VaR and CVaR are fundamentally related to rare events. As a matter of fact, in this kind of problem, we are interested in hedging extreme events, that is, events that are observed with a very small probability so that we obtain few significant scenarios to update our estimates. In order to improve the efficiency of our simulations, we need to introduce a (recursive) variance reduction method. To be more precise, in this rare event framework, the challenge is to generate more samples in the area of interest, i.e., the tail of the distribution. A natural tool used in this situation is importance sampling (IS). In the same spirit as the recursive IS procedure developed in Lemaire and Pagès (2010) (and used in Bardou et al. 2009 for the estimation of the VaR and the CVaR), our IS parameters are optimized recursively by a companion (unconstrained) RM algorithm, which is combined in an adaptive way with our first procedure. We also propose another variance reduction method based on a linear control variable (LCV), which can be used either alone when IS is not necessary or combined with IS. It dramatically accelerates the convergence of the original procedure. The weak convergence rate of the resulting procedure is ruled by a CLT at an optimal rate and minimal asymptotic variance.

To sum up, the contribution of the paper is threefold. First, we introduce the $\mathcal{F}$-CVaR ($\mathcal{F} \subseteq \mathcal{G}$ being a sub σ-field representative of the information available for investors) and study its properties. In particular, we show that the dynamic risk measure $\mathcal{G}_\ell$-CVaR satisfies relevant properties. Second, we prove the existence of optimal strategies for the three stochastic control problems (1.1), (1.2), and (1.3) under mild assumptions, namely, the absence of arbitrage opportunity and the nondegeneracy of the price process. Third, we propose numerical procedures involving stochastic approximation algorithms, optimal vector quantization, and variance reduction techniques, such as IS in a Markovian framework. Compared to linear programming approaches, such as Rockafellar and Uryasev (2000), our procedures are robust and tractable since they are not limited by the number of simulated scenarios and can deal with the problem of rare events which is classical in this context.

The paper is organized as follows: in Section 2, we present the $\mathcal{F}$-CVaR and develop some fundamental theoretical results on the $\mathcal{F}$-CVaR and CVaR hedging. This will allow us to devise a RM algorithm. Section 3 is devoted to numerical aspects of CVaR hedging. We show how to devise a RM algorithm to compute an optimal strategy with its associated VaR and CVaR. We establish its $a.s.$ convergence and weak rate of convergence. In order to approximate conditional expectations, we rely on optimal vector quantization. We present several algorithms to approximate the optimal strategy solution of (1.1) and briefly mention the two variance reduction tools in the static framework. Finally, Section 4 is devoted to numerical examples. We focus on energy markets, which are known to be incomplete. We propose several portfolios to challenge the algorithm and display CVaR estimations.

Notations: • For $x, y \in \mathbb{R}^d$, the real number $x.y$ stands for their canonical scalar product, $|x| := \sqrt{x.x}$ denotes the Euclidean norm of x and x^T denotes the transpose of the column vector x.

• Given a sub σ-field $\mathcal{F} \subseteq \mathcal{G}$, $L_{\mathcal{F}}$ will denote the space of all $\mathcal{F}$-measurable and $\mathbb{P}$-$a.s.$ finite random variables with values in $\mathbb{R}$.

• $\xrightarrow{\mathcal{L}}$ will denote the convergence in distribution and $\xrightarrow{a.s.}$ will denote the almost sure convergence.

• $x_+ := \max(0, x)$ will denote the positive part function.

• For $d \in \mathbb{N}^*$ and $p > 0$, $L^p_{\mathbb{R}^d}(\mathbb{P})$ will denote the space of $\mathbb{R}^d$-valued random vector U such that $\mathbb{E}\left[|U|^p\right]^{1/p} < +\infty$.

2. THEORETICAL ASPECTS OF CVaR HEDGING

2.1. A Short Background on VaR and CVaR

We start this section by briefly recalling the definitions of the VaR and the CVaR (for more details, we refer to Bardou et al. 2009). Then, we introduce the notion of dynamic CVaR that will be fundamental throughout the paper. To measure the risk of a loss L (or a short position on the contingent claim) at time 0, one usually considers the VaR_α, $\alpha \in (0, 1)$, i.e., the lowest α-quantile of the distribution L:

$$\mathrm{VaR}_\alpha(L) := \inf\left\{\xi \in \mathbb{R} \mid \mathbb{P}(L \leq \xi) \geq \alpha\right\}.$$

We assume that the distribution function of L is continuous (i.e., with no atom) so that the VaR is the lowest solution of the equation:

$$\mathbb{P}(L \leq \xi) = \alpha.$$

If the distribution function is (strictly) increasing, the above equation has a unique solution, otherwise there may be (infinitely) more. In fact, in what follows, we will consider that *any* solution of the previous equation is the $\mathrm{VaR}_\alpha(L)$. Another risk measure commonly used to provide information about the tail of the distribution of L is the CVaR (at level α). Assuming that $L \in L^1_{\mathbb{R}}(\mathbb{P})$, it is defined by:

$$\mathrm{CVaR}_\alpha(L) := \mathbb{E}\left[L | L \geq \mathrm{VaR}_\alpha(L)\right].$$

These two quantities are widely used in practice to measure the risk at time 0 of L. The next proposition shows that these two quantities are solutions to a convex optimization problem the value function of which can be represented as an expectation, as pointed out in Rockafellar and Uryasev (2000). It has already been used in Bardou et al. (2009) to devise a RM algorithm to compute both the VaR and the CVaR. We briefly recall this important result in order to justify the definition of the dynamic CVaR.

PROPOSITION 2.1. *Suppose that the distribution function of L is continuous and that $L \in L^1_{\mathbb{R}}(\mathbb{P})$. Let V be the function defined on $\mathbb{R}$ by:*

$$V(\xi) = \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\xi)_+\right]. \tag{2.1}$$

Then, the function V is convex, Lipschitz-continuous, differentiable, and $\mathrm{VaR}_\alpha(L)$ *is any point of the set*

$$\arg\min V = \left\{\xi \in \mathbb{R} \mid V'(\xi) = 0\right\} = \{\xi \in \mathbb{R} \mid \mathbb{P}(L \leq \xi) = \alpha\},$$

where V' denotes the derivative of V. This derivative V' can in turn be represented as an expectation by

$$\forall\, \xi \in \mathbb{R}, \quad V'(\xi) = \mathbb{E}\left[1 - \frac{1}{1-\alpha}\mathbf{1}_{\{L \geq \xi\}}\right]. \tag{2.2}$$

Furthermore,

$$\mathrm{CVaR}_\alpha(L) = \min_{\xi \in \mathbb{R}} V(\xi). \tag{2.3}$$

We refer to Rockafellar and Uryasev (2000) for a proof.

2.2. Definition and General Properties of $\mathcal{F}$-CVaR

Now, we are in a position to extend the notion of the static CVaR to the notion of $\mathcal{F}$-CVaR. We consider a σ-field $\mathcal{F} \subseteq \mathcal{G}$, which represents a nontrivial (consistent) set of information observable by all investors at a certain state.

DEFINITION 2.2. Suppose that L satisfies $\mathbb{E}[L_+|\mathcal{F}] < +\infty$ $a.s.$ The $\mathcal{F}$-CVaR is a random risk measure defined by

$$\mathcal{F}\text{-}\mathrm{CVaR}_\alpha(L) := ess\,inf_{\xi \in L_{\mathcal{F}}} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\xi)_+ \,|\mathcal{F}\right].$$

This is an $\mathcal{F}$-random variable, which describes the CVaR_α for all $\mathcal{F}$-measurable scenarios. By construction, it is straightforward that it satisfies the following randomized axioms of a *coherent* risk measure:

(1) Subadditivity: for every $L,\ L'$ satisfying $\mathbb{E}\left[L_+ + L'_+ \mid \mathcal{F}\right] < +\infty$, $a.s.$

$$\mathcal{F}\text{-CVaR}_\alpha(L + L') \le \mathcal{F}\text{-CVaR}_\alpha(L) + \mathcal{F}\text{-CVaR}_\alpha(L').$$

(2) Positive homogeneity: if $\lambda \in L_{\mathcal{F}}$ with $\lambda \ge 0$ $a.s.$, $\mathcal{F}\text{-CVaR}_\alpha(\lambda L) = \lambda \times \mathcal{F}\text{-CVaR}_\alpha(L)$.
(3) Translation invariance: for all $Z \in L_{\mathcal{F}}$, $\mathcal{F}\text{-CVaR}_\alpha(L + Z) = Z + \mathcal{F}\text{-CVaR}_\alpha(L)$.
(4) Monotonicity: for every $L,\ L'$ such that $\mathbb{E}\left[L_+ + L'_+ \mid \mathcal{F}\right] < +\infty$ and $L \le L'$ $a.s.$, $\mathcal{F}\text{-CVaR}_\alpha(L) \le \mathcal{F}\text{-CVaR}_\alpha(L')$.

When $\mathcal{F} = \{\varnothing, \Omega\}$, the $\mathcal{F}\text{-CVaR}_\alpha(L)$ coincides with the usual $\text{CVaR}_\alpha(L)$ in the sense of Artzner et al. (1999).

PROPOSITION 2.3. *The quantity* $\mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha(L)\right]$ *is a coherent risk measure.*

We will use this coherent risk measure to design several time-consistent risk measures, which will be used later on to produce strategies to be plugged in the original CVaR_α criterion.

If one aims at measuring the risk at time ℓ of his financial strategy $\theta \in \mathcal{A}_{\mathcal{G}}$ started at time 0 using a CVaR criterion, one has to compute $\mathcal{G}_\ell\text{-CVaR}_\alpha(L - \sum_{k=1}^{M} \theta_k.\Delta X_k)$, which is only known at time ℓ. It is natural for the holder of the portfolio to ask how the risk evolves with time until maturity. Next result points out the relevance of the $\mathcal{G}_\ell$-CVaR risk measure, more precisely, we prove that the risk of any position decreases with time.

PROPOSITION 2.4. *We set $M = +\infty$ for this result. Let $Y \in L_{\mathcal{G}_\infty}$ such that $\mathbb{E}[|Y|] < +\infty$, where $\mathcal{G}_\infty = \vee_\ell \mathcal{G}_\ell$.*
The sequence $(\mathcal{G}_\ell\text{-CVaR}_\alpha(Y))_{1 \le \ell \le M}$ is a $\mathbb{G}$-supermartingale. Moreover, it satisfies,

$$\mathcal{G}_n\text{-CVaR}_\alpha(Y) \xrightarrow{a.s.} Y, \quad as\ n \to +\infty.$$

Proof. First, note that for $\ell = 1, \ldots, M$,

$$\mathcal{G}_\ell\text{-CVaR}_\alpha(Y) = \operatorname*{ess\,inf}_{\xi \in \Theta^1_\ell} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y - \xi)_+ \,|\mathcal{G}_\ell\right] \le \frac{1}{1-\alpha}\mathbb{E}\left[Y_+ |\mathcal{G}_\ell\right] \in L^1(\mathbb{P}),$$

and by Jensen's inequality,

$$\mathbb{E}\left[Y|\mathcal{G}_\ell\right] = \operatorname*{ess\,inf}_{\xi \in \Theta^1_\ell} \xi + \frac{1}{1-\alpha}\left(\mathbb{E}\left[Y|\mathcal{G}_\ell\right] - \xi\right)_+ \le \mathcal{G}_\ell\text{-CVaR}_\alpha(Y), \tag{2.4}$$

so that, $\mathcal{G}_\ell\text{-CVaR}_\alpha(Y) \in L^1_{\mathbb{R}}(\mathbb{P})$. Then, by definition, we have

$$\mathcal{G}_\ell\text{−CVaR}_\alpha(\mathrm{Y}) \le \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(\mathrm{Y} - \xi)_+ \,|\mathcal{G}_\ell\right], \quad \text{for all } \xi \in \Theta^1_{\ell-1},$$

which implies that

$$\mathbb{E}\left[\mathcal{G}_\ell\text{-CVaR}_\alpha(Y)|\mathcal{G}_{\ell-1}\right] \le \operatorname*{ess\,inf}_{\xi \in \Theta^1_{\ell-1}} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y - \xi)_+ \,|\mathcal{G}_{\ell-1}\right] = \mathcal{G}_{\ell-1}\text{-CVaR}_\alpha(Y).$$

Consequently, the sequence $(\mathcal{G}_\ell\text{-CVaR}_\alpha(Y))_{1 \le \ell \le M}$ is a $\mathbb{G}$-supermartingale.

Now, owing to (2.4), for $n \geq 1$,

$$(\mathcal{G}_n\text{-CVaR}_\alpha(Y))_- \leq (\mathbb{E}[Y|\mathcal{G}_n])_- \leq (\mathbb{E}[Y_-|\mathcal{G}_n]) \leq \mathbb{E}[|Y||\mathcal{G}_n],$$

hence, we obtain: $\sup_{n\geq 0}\mathbb{E}\left[(\mathcal{G}_n\text{-CVaR}_\alpha(Y))_-\right] \leq \mathbb{E}[|Y|] < +\infty$.

Doob's martingale convergence theorem implies that the sequence $(\mathcal{G}_n\text{-CVaR}_\alpha(Y))_{n\geq 1}$ $a.s.$ converges toward $\tilde{Y}_\infty \in L^1(\mathbb{P})$. Now, from the first inequality and the $a.s.$ convergence of the sequence $(\mathbb{E}[Y|\mathcal{G}_n])_{n\geq 1}$ toward $\mathbb{E}[Y|\mathcal{G}_\infty] = Y$ (the convergence also holds in L^1), we get

$$\tilde{Y}_\infty \geq Y.$$

On the other hand, for every $n \geq 1$,

$$\mathcal{G}_n\text{-CVaR}_\alpha(Y) \leq \mathbb{E}[Y|\mathcal{G}_n] + \frac{1}{1-\alpha}\mathbb{E}\left[(Y - \mathbb{E}[Y|\mathcal{G}_n])_+ |\mathcal{G}_n\right],$$

so that, for every $n \geq m \geq 1$ and every $A \in \mathcal{G}_m$,

$$\mathbb{E}[\mathbf{1}_A\mathcal{G}_n\text{-CVaR}_\alpha(Y)] \leq \mathbb{E}\left[\mathbf{1}_A\left(Y + \frac{1}{1-\alpha}(Y - \mathbb{E}[Y|\mathcal{G}_n])_+\right)\right]. \tag{2.5}$$

It follows from Fatou's Lemma that

$$\mathbb{E}\left[\mathbf{1}_A\tilde{Y}_\infty\right] = \mathbb{E}\left[\mathbf{1}_A \liminf_n \mathcal{G}_n\text{-CVaR}_\alpha(Y)\right] \leq \liminf_n \mathbb{E}[\mathbf{1}_A\mathcal{G}_n\text{-CVaR}_\alpha(Y)],$$

since $\mathcal{G}_n\text{-CVaR}_\alpha(Y) \geq \mathbb{E}[Y|\mathcal{G}_n]$, $a.s.$, for every $n \geq 1$ and $\mathbb{E}[Y|\mathcal{G}_n]$ converges in $L^1(\mathbb{P})$. Now $(Y - \mathbb{E}[Y|\mathcal{G}_n])_+ \xrightarrow{L^1(\mathbb{P})} 0$, which shows that

$$\limsup_n \mathbb{E}\left[\mathbf{1}_A\left(Y + \frac{1}{1-\alpha}(Y - \mathbb{E}[Y|\mathcal{G}_n])_+\right)\right] \leq \mathbb{E}[\mathbf{1}_A Y].$$

Combining these inequalities with (2.5) yields

$$\forall\, m \geq 1,\ \forall\, A \in \mathcal{G}_m,\ \ \mathbb{E}\left[\mathbf{1}_A\tilde{Y}_\infty\right] \leq \mathbb{E}[\mathbf{1}_A Y],$$

which in turn implies that

$$\tilde{Y}_\infty \leq Y.$$

This completes the proof. □

This property ensures that $(\mathcal{G}_\ell\text{-CVaR}_\alpha(Y))_{\ell\geq 1}$ will asymptotically cover the risk of any position Y and will not cover a larger risk. This dynamic risk measure is said to be *asymptotic safe* and *asymptotic precise*. For more details, we refer to Föllmer and Penner (2006). This result naturally implies that the sequence $(\mathbb{E}[\mathcal{G}_\ell\text{-CVaR}_\alpha(Y)])_{1\leq \ell\leq M}$ is nonincreasing, thus the average risk (hopefully) decreases with time for any strategy $\theta \in \mathcal{A}$. The result concerning the convergence of the supermartingale is quite intuitive. If the loss of the considered portfolio is such that L is $\mathcal{G}_M$-measurable (as it is the case in our modeling), then the average risk associated to this position decreases toward the average loss itself.

Another useful result concerns the supermartingale property of the hedged portfolio.

COROLLARY 2.5. *Suppose that $L \in L^1_{\mathbb{R}}(\mathbb{P})$ and that there exists $p' > 1$ such that $\Delta X_\ell \in L^{p'}_{\mathbb{R}^d}(\mathbb{P})$ for $\ell = 1, \ldots, M$. Let $\theta \in \mathcal{A}$ such that $\theta_\ell \in L^p_{\mathbb{R}^d}(\mathbb{P})$ with $p = \frac{p'}{p'-1}$. Then,*

$$\left(\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell\right)\right)_{0\le k\le M}$$

is a supermartingale and satisfies, for every $k \in \{0, \ldots, M-1\}$,

$$\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_\ell.\Delta X_\ell\right) = \mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell\right) - \sum_{\ell=1}^{k} \theta_\ell \Delta X_\ell. \tag{2.6}$$

Proof. Hölder's inequality implies that $\sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell \in L^1_{\mathbb{R}}(\mathbb{P})$ so that in view of the definition of the $\mathcal{G}_k$-CVaR, $\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell\right) \in L^1(\mathbb{P})$. Now by the change of variable, $\xi = \tilde{\xi} + \sum_{\ell=1}^{k} \theta_\ell \Delta X_\ell$, we have

$$\begin{aligned}
\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell\right) &= \operatorname*{ess\,inf}_{\xi \in \Theta_k^1} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell - \xi\right)_+ \middle| \mathcal{G}_k\right] \\
&= \sum_{\ell=1}^{k} \theta_\ell \Delta X_\ell \\
&\quad + \operatorname*{ess\,inf}_{\tilde{\xi} \in \Theta_k^1} \tilde{\xi} + \frac{1}{1-\alpha}\mathbb{E}\left[\left(L - \sum_{\ell=k+1}^{M} \theta_\ell.\Delta X_\ell - \tilde{\xi}\right)_+ \middle| \mathcal{G}_k\right] \\
&= \sum_{\ell=1}^{k} \theta_\ell \Delta X_\ell + \mathcal{G}_\ell\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_\ell.\Delta X_\ell\right).
\end{aligned}$$

□

In particular, if X is a $(\mathcal{G}, \mathbb{P})$-martingale, (2.6) implies that for every $k \in \{0, \ldots, M-1\}$

$$\mathbb{E}\left[\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell\right)\right] = \mathbb{E}\left[\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_\ell.\Delta X_\ell\right)\right],$$

which means that the mean estimate at time 0 of the risk at time k does not depend on the decisions taken prior to time k. This property follows from the fact that the hedging strategy is self-financing.

2.3. CVaR Hedging Using a One-Step Self-Financed Strategy

In this section, we address the two problems (1.2) and (1.3), that is hedging a contingent claim using a one-step strategy starting with an initial investment of 0 and a CVaR or a $\mathcal{G}_{\ell_0}$-CVaR criterion at a fixed time ℓ_0.

By one-step strategy decided at time ℓ_0, we mean that the investor is restricted to rebalance its portfolio only once at time $\ell_0 \in \{0, \dots, M-1\}$. By a one-step static strategy, we mean that the investor uses a one-step strategy decided at time 0.

This case of study is interesting since on energy markets, practitioners may be interested by a rough hedge of their loss using only few forward contracts, especially when dealing with physical assets like gas storage or power plant. Moreover, theoretical results in the dynamic framework rely on ideas similar to those developed in this section.

Without loss of generality, we can suppose that the market operates with only two dates ℓ_0 and M. We set $X := X_M - X_{\ell_0}$. We consider the two one-period stochastic control problems

$$\inf_{\theta \in \Theta^d_{\ell_0}} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right], \tag{2.7}$$

and

$$\inf_{\theta \in \Theta^d_{\ell_0}} \text{CVaR}_\alpha\left(L - \theta.X\right). \tag{2.8}$$

Note that (2.8) can be written

$$\inf_{\xi \in \mathbb{R}} \inf_{\theta \in \Theta^d_{\ell_0}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right], \tag{2.9}$$

so that, in a first step, one may address the stochastic optimization problem

$$\inf_{\theta \in \Theta^d_{\ell_0}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right]. \tag{2.10}$$

Up to the change of variable $L := L - \xi$, we can suppose that $\xi = 0$ and $\alpha = 0$ so that, without loss of generality, the problem (2.10) is equivalent to minimizing the shortfall risk

$$\inf_{\theta \in \Theta^d_{\ell_0}} \mathbb{E}\left[(L - \theta.X)_+\right]. \tag{2.11}$$

First, we will show that there exists an optimal one-step trading strategy $\tilde{\theta}$ solution to (2.11), thus for all $\xi \in \mathbb{R}$ there exists $\theta^*_\alpha(\xi)$ solution to (2.10). Finally, we will come back to (2.9) and deduce the existence of an optimal ξ^*_α solution to (2.9).

Now in order to derive the existence of solutions to (2.7) and (2.8), we will impose the following absence of arbitrage property (NA):

$$(\text{NA}) : \forall \theta \in \mathcal{A}_{\mathcal{G}}, \ \left(V^\theta_M \geq 0 \ a.s. \Longrightarrow \ V^\theta_M = 0 \ a.s.\right), \tag{2.12}$$

where $V^\theta_M := \sum_{\ell=1}^M \theta_\ell.\Delta X_\ell$ denotes the final value of a (self-financed) portfolio θ with no initial capital. There is a well-known duality result between the no-arbitrage property (NA) and the existence of some equivalent martingale measure, it is the so-called fundamental theorem of asset pricing. Throughout the paper, we will denote by $\mathbb{P}^*$ a probability equivalent to $\mathbb{P}$ under which the (discounted) price process X is a martingale and such that $\frac{d\mathbb{P}^*}{d\mathbb{P}} < K \ a.s.$ for some deterministic constant $K < +\infty$. The existence of such a probability measure is established in Jacod and Shiryaev (1998, theorem 3, p. 264). Moreover, we assume the existence of a regular conditional distribution of the couple (L, X) given $\mathcal{G}_{\ell_0}$ denoted by $\Pi(\mathrm{d}y, \mathrm{d}x) = \Pi(\omega, \mathrm{d}y, \mathrm{d}x)$ and we make the following assumptions on the conditional distribution of (L, X).

ASSUMPTION 2.6. *The distribution of L and X satisfies $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $X \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$ and the two following properties hold:*

- $\mathbb{E}\left[XX^T|\mathcal{G}_{\ell_0}\right]$ *is a.s. positive definite in $\mathcal{S}(d, \mathbb{R})$.*
- *The conditional distribution of X given $\mathcal{G}_{\ell_0}$ is continuous, that is, no affine hyperplane has positive mass.*

The following proposition is the key result to solve our optimization problem. The proof is given in the Appendix and relies on classical arguments from stochastic control theory.

PROPOSITION 2.7. *Let V_f and V_s be the two functions defined, respectively, on $\Omega \times \mathbb{R} \times \mathbb{R}^d$ and $\Omega \times \mathbb{R}^d$ by*

$$V_f(\omega, \xi, \theta) = \int v_f(\xi, \theta, y, x)\Pi(\omega, \mathrm{d}x, \mathrm{d}y), \tag{2.13}$$

and

$$V_s(\omega, \theta) = \int v_s(\theta, y, x)\Pi(\omega, \mathrm{d}x, \mathrm{d}y), \tag{2.14}$$

where

$$v_f(\xi, \theta, y, x) := \xi + \frac{1}{1-\alpha}(y - \theta.x - \xi)_+ , \quad \textit{and} \quad v_s(\theta, y, x) := (y - \theta.x)_+ . \tag{2.15}$$

Assume that (NA) and Assumption 2.6 are satisfied, then we have

(i) Static Risk: For $\mathbb{P}$-almost all $\omega \in \Omega$, the function $V_s(\omega, .)$ is Lipschitz-continuous, convex, and satisfies

$$\lim_{|\theta|\to+\infty} V_s(\omega, \theta) = +\infty.$$

Moreover, we have

$$\inf_{\theta\in\Theta^d_{\ell_0}} \mathbb{E}\left[(L - \theta.X)_+\right] = \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta\in\Theta^d_{\ell_0}} \mathbb{E}\left[(L - \theta.X)_+ \middle| \mathcal{G}_{\ell_0}\right]\right], \tag{2.16}$$

and

$$\operatorname*{ess\,inf}_{\theta\in\Theta^d_{\ell_0}} \mathbb{E}\left[(L - \theta.X)_+ \middle| \mathcal{G}_{\ell_0}\right](\omega) = \min_{\theta\in\mathbb{R}^d} V_s(\omega, \theta). \tag{2.17}$$

(ii) Forward Risk: For $\mathbb{P}$-almost all $\omega \in \Omega$, the function $V_f(\omega, ., .)$ is Lipschitz-continuous, convex, and satisfies

$$\lim_{|(\xi,\theta)|\to+\infty} V_f(\omega, \xi, \theta) = +\infty.$$

Moreover, we have

$$\inf_{\theta\in\Theta^d_{\ell_0}} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-}CVaR_\alpha\,(L-\theta.X)\right]$$

$$= \mathbb{E}\left[\operatorname*{ess\,inf}_{(\xi,\theta)\in\Theta^{d+1}_{\ell_0}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\,(L-\theta.X-\xi)_+ \middle| \mathcal{G}_{\ell_0}\right]\right], \tag{2.18}$$

and

$$\operatorname*{ess\,inf}_{(\xi,\theta)\in\Theta^{d+1}_{\ell_0}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\,(L-\theta.X-\xi)_+ \middle| \mathcal{G}_{\ell_0}\right](\omega) = \min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\omega,\xi,\theta). \tag{2.19}$$

The right-hand sides of (2.17) and (2.19) show that the two optimization problems (2.7) and (2.8) can be written

$$\inf_{\theta\in\Theta^d_{\ell_0}} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-}\mathrm{CVaR}_\alpha\,(L-\theta.X)\right] = \mathbb{E}\left[\min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi,\theta)\right], \tag{2.20}$$

and

$$\inf_{\theta\in\Theta^d_{\ell_0}} \mathrm{CVaR}_\alpha\,(L-\theta.X) = \inf_{\xi\in\mathbb{R}} \mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right], \tag{2.21}$$

respectively. Consequently, for $\mathbb{P}$-almost all $\omega\in\Omega$, we have to solve deterministic optimization problems. The next result provides a characterization of those minima and will allow us to devise (later on) numerical procedures to estimate the quantities of interest.

PROPOSITION 2.8. *Suppose that (NA) and Assumption 2.6 are satisfied. Then, for all* $\xi\in\mathbb{R}$

$$\operatorname{Arg\,min} V_f(\xi,.) = \left\{\theta\in\mathbb{R}^d \mid \nabla_\theta V_f(\xi,\theta) = 0\right\} \neq \varnothing$$

and

$$\operatorname{Arg\,min} V_f = \left\{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d \mid \nabla_{(\xi,\theta)} V_f(\xi,\theta) = 0\right\} \neq \varnothing,$$

where the gradient of V_f is given for every $(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d$ by

$$\nabla_{(\xi,\theta)} V_f(\xi,\theta) = \int \nabla_{(\xi,\theta)} v_f(\xi,\theta,y,x)\Pi(dx,dy)$$

and

$$\nabla_\theta V_f(\xi,\theta) = \int \nabla_\theta v_f(\xi,\theta,y,x)\Pi(dx,dy).$$

Moreover, $\xi\mapsto\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right]$ is Lipschitz-continuous, convex, and $\lim_{|\xi|\to+\infty}\mathbb{E}\times[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)] = +\infty$. *Consequently, (2.7) and (2.8) admit solutions.*

Proof. Since the functions $(\xi,\theta)\mapsto v_f(\xi,\theta,y,x)$, $(y,x)\in\mathbb{R}\times\mathbb{R}^d$, are convex, the function V_f is convex. To justify the formal differentiation of V_f, we only need to check the domination property. First, note that we have, for all $(y, x) \in \mathbb{R} \times \mathbb{R}^d$

$$\frac{\partial \xi}{\partial v_f}(\xi, \theta, y, x) = 1 - \frac{1}{1-\alpha} 1_{\{y-\theta.x \geq \xi\}},$$
$$\frac{\partial \theta}{\partial v_f}(\xi, \theta, y, x) = -\frac{1}{1-\alpha} x 1_{\{y-\theta.x \geq \xi\}},$$

so that there exists $C > 0$ such that

$$\left|\nabla_{(\xi,\theta)} v_f(\xi, \theta, L, X)\right| \leq C(1 + |X|) \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P}).$$

Now, let $\xi,\ \xi' \in \mathbb{R}$, there exists a real constant $C > 0$ such that

$$\left|\mathbb{E}\left[\inf_{\theta \in \mathbb{R}^d} V_f(\xi, \theta)\right] - \mathbb{E}\left[\inf_{\theta \in \mathbb{R}^d} V_f(\xi', \theta)\right]\right| \leq \mathbb{E}\left[\sup_{\theta \in \mathbb{R}^d} \left|V_f(\xi, \theta) - V_f(\xi', \theta)\right|\right]$$
$$\leq C\,|\xi - \xi'|.$$

Using Bayes' rule and Jensen's inequality, we have for all $\theta \in \mathcal{G}_{\ell_0}$

$$\mathbb{E}\left[\left(\xi + \frac{1}{1-\alpha}(L - \theta.X - \xi)_+\right)\middle|\, \mathcal{G}_{\ell_0}\right]$$
$$= \mathbb{E}\left[\left(\xi + \frac{1}{1-\alpha}(L - \theta.X - \xi)_+\right)\frac{d\mathbb{P}^*}{d\mathbb{P}}\frac{d\mathbb{P}}{d\mathbb{P}^*}\,\middle|\, \mathcal{G}_{\ell_0}\right]$$
$$\geq \frac{1}{K}\left(\xi + \frac{1}{1-\alpha}\left(\mathbb{E}_{\mathbb{P}^*}\left[L \mid \mathcal{G}_{\ell_0}\right] - \xi\right)_+\right)\mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}}\,\middle|\, \mathcal{G}_{\ell_0}\right].$$

It is clear that $\lim_{|\xi|\to+\infty} \xi + \frac{1}{1-\alpha}\left(\mathbb{E}_{\mathbb{P}^*}\left[L \mid \mathcal{G}_{\ell_0}\right] - \xi\right)_+ = +\infty$, hence, using Fatou's Lemma we conclude that $\lim_{|\xi|\to+\infty} \mathbb{E}\left[\min_{\theta \in \mathbb{R}^d} V_f(\xi, \theta)\right] = +\infty$. This completes the proof. □

2.4. Existence of an Optimal Multistep CVaR Hedging Strategy

In this section, we address the main problem (1.1), that is, hedging a contingent claim using a dynamic self-financing strategy (with an initial investment of 0) according to a static CVaR criterion at a fixed time $t = 0$. Actually, in this theoretical section, we consider the more general multistage stochastic optimization problem:

$$\inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathrm{CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell\right) = \inf_{\xi \in \mathbb{R}} \inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell - \xi\right)_+\right]. \tag{2.22}$$

Note that in order to solve (2.22), we may address first the multistage stochastic optimization problem

$$\inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_\ell.\Delta X_\ell - \xi\right)_+\right], \quad \text{for each } \xi \in \mathbb{R}. \tag{2.23}$$

Up to the change of variable $L := L - \xi$, we can suppose that $\xi = 0$ and $\alpha = 0$ so that, without loss of generality, the problem (2.23) is equivalent to minimizing the shortfall risk

$$\inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell\right)_+\right]. \tag{2.24}$$

The optimization problem (2.24) seems to be a classical stochastic control problem for which standard theorems ensuring the existence of a dynamic programming principle apply. Unfortunately, so is not the case. For instance, for $\mathbb{P}$-almost all w, the coercivity of the function $\theta \mapsto (L(w) - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell(w))_+$, which is crucial in our framework, is not satisfied, see assumption (iii) in Evstigneev (1976). Consequently, we will have to adapt the (classical) proof from Evstigneev (1976) to derive the existence of an optimal control $\tilde{\theta} := (\tilde{\theta}_\ell)_{1 \le \ell \le M}$ solution to (2.24). Hence, we will obtain the existence of an optimal CVaR-hedging sequence $\theta_\alpha^* := (\theta_{\ell,\alpha}^*)_{1 \le \ell \le M}$ solution to (2.23). Finally, we will come back to (2.22) and using similar arguments to those developed in the static framework, we will prove the existence of ξ_α^* solution to the optimization problem

$$\begin{aligned} &\inf_{\xi \in \mathbb{R}} \inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell - \xi\right)_+\right] \\ &= \inf_{\xi \in \mathbb{R}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \sum_{\ell=1}^{M} \theta_{\ell,\alpha}^* . \Delta X_\ell - \xi\right)_+\right]. \end{aligned}$$

ASSUMPTION 2.9. *The distribution of* $(L, \Delta X_1, \dots, \Delta X_M)$ *satisfies* $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$ *and the two following properties hold for* $\ell = 1, \dots, M$:

(i) $\Delta X_\ell \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$ *and* $\mathbb{E}\left[\Delta X_\ell \Delta X_\ell^T \middle| \mathcal{G}_{\ell-1}\right]$ *is a.s. positive definite in* $\mathcal{S}(d, \mathbb{R})$.
(ii) The conditional distribution of ΔX_ℓ *given* $\mathcal{G}_{\ell-1}$ *is continuous, that is, no affine hyperplane has positive mass.*

In the spirit of the dynamic programming principle, we construct the solution of (2.24) using a step-by-step backward induction. To be more precise, using similar arguments to those used to prove (2.18), one first notes that (2.24) can be written as

$$\inf_{\theta_\ell \in \Theta^d_{\ell-1}, \ell=1,\dots,M-1} \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta_M \in \Theta^d_{M-1}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell\right)_+ \middle| \mathcal{G}_{M-1}\right]\right], \tag{2.25}$$

so that one may start by solving the following problem

$$\operatorname*{ess\,inf}_{\theta_M \in \Theta^d_{M-1}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell\right)_+ \middle| \mathcal{G}_{M-1}\right](\omega) = \min_{\theta_M \in \mathbb{R}^d} V_{M-1}(\omega, \theta_{1:M-1}, \theta_M) \tag{2.26}$$

$$= V_{M-1}(\omega, \theta_{1:M-1}, \tilde{\theta}_M) a.s., \tag{2.27}$$

where $\tilde{\theta}_M \in \Theta^d_{M-1}$, V_{M-1} is defined for all $\omega \in \Omega$, $\theta_\ell \in \Theta^d_{\ell-1}$, $\ell = 1, \dots, M-1$, by

$$V_{M-1}(\omega, \theta_{1:M-1}, \theta_M) := \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell\right)_+ \,\middle|\, \mathcal{G}_{M-1}\right](\omega)$$

$$= \int \left(y - \sum_{\ell=1}^{M} \theta_\ell(\omega) . \Delta x_\ell\right)_+ \Pi_{M-1}(\omega, dx, dy). \tag{2.28}$$

This follows from similar arguments to those used for the proof of Proposition 2.7, i.e., from the fact that for all $\omega \in \Omega$, $\theta_\ell \in \Theta^d_{\ell-1}$, $\ell = 1, \dots, M-1$, the function (defined on $\mathbb{R}^d$) by $\theta_M \mapsto V_{M-1}(\xi, \theta_{1:M-1}, \theta_M)$ is convex, Lipschitz-continuous, and $\lim_{|\theta_M| \to +\infty} V_{M-1}(\xi, \theta_{1:M-1}, \theta_M) = +\infty$. Thus, it implies that (2.26) has a solution that we denote by $\tilde{\theta}_M := \tilde{\theta}_M(\omega, \theta_1, \dots, \theta_{M-1})$, which is $\mathcal{G}_{M-1}$-measurable (owing to measurable selection, see, e.g., lemmas 3 and 4 in Evstigneev 1976) so that (2.27) holds.

Then we proceed by a backward induction: we denote by $\tilde{\theta}_{\ell+1:M} := (\tilde{\theta}_{\ell+1}, \dots, \tilde{\theta}_M)$ the solution built down to step ℓ. At step $\ell - 1$, we address for every $\theta_{1:\ell-1} \in \Theta^d_1 \times \cdots \times \Theta^d_{\ell-2}$, the problem

$$\operatorname*{ess\,inf}_{\theta_\ell \in \Theta^d_{\ell-1}} \mathbb{E}\left[V_\ell(\theta_{1:\ell-1}, \theta_\ell) | \mathcal{G}_{\ell-1}\right](\omega) = \min_{\theta_\ell \in \mathbb{R}^d} V_{\ell-1}(\omega, \theta_{1:\ell-1}, \theta_\ell)$$

$$= V_{\ell-1}(\omega, \theta_{1:\ell-1}, \tilde{\theta}_\ell) \ \ a.s., \tag{2.29}$$

where for all $\theta_k \in \Theta^d_{k-1}$, $k = 1, \dots, \ell - 1$, the functions V_ℓ and $V_{\ell-1}$ are defined by

$$V_\ell(\omega, \theta_{1:\ell-1}, \tilde{\theta}_\ell) := \mathbb{E}\left[\left(L - \sum_{k=1}^{\ell-1} \theta_k . \Delta X_k - \sum_{k=\ell}^{M} \tilde{\theta}_k . \Delta X_k\right)_+ \,\middle|\, \mathcal{G}_\ell\right](\omega)$$

$$= \int \left(y - \sum_{k=1}^{\ell-1} \theta_k(\omega) . \Delta x_k - \sum_{k=\ell}^{M} \tilde{\theta}_k . \Delta x_k\right)_+ \Pi_\ell(\omega, dx, dy),$$

and $V_{\ell-1}(\omega, \theta_{1:\ell-1}, \theta_\ell) = \int \left(y - \sum_{k=1}^{\ell} \theta_k(\omega) . \Delta x_k - \sum_{k=\ell+1}^{M} \tilde{\theta}_k . \Delta x_k\right)_+ \Pi_{\ell-1}(\omega, dx, dy)$.

The following proposition implies that (2.22) has an optimal solution $(\xi^*_\alpha, \theta^*_\alpha) \in \mathbb{R} \times \mathcal{A}_{\mathcal{F}}$.

PROPOSITION 2.10. *Suppose that (NA) and Assumption 2.9 are satisfied. Then,*

(1) Equation (2.25) is satisfied, the problem (2.26) has a solution, and for $\ell = M - 1, \dots, 1$, *one can find a* $\mathcal{G}_{\ell-1}$*-measurable random variable* $\tilde{\theta}_\ell$ *solution of (2.29). Thus, (2.23) has an optimal solution denoted by* $\theta^*_\alpha := (\theta^*_{\alpha,\ell})_{1 \le \ell \le M}$.

(2) The function $\xi \mapsto \xi + \frac{1}{1-\alpha} \inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell - \xi\right)_+\right]$ *is Lipschitz-continuous, convex, and satisfies*

$$\lim_{|\xi| \to +\infty} \xi + \frac{1}{1-\alpha} \inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_\ell . \Delta X_\ell - \xi\right)_+\right] = +\infty,$$

so that (2.22) admits a solution.

Proof. (i) The proof of (2.25) relies on similar arguments to those already used for the proof of (2.16).

Let $\tilde{\theta}_M$ be a solution of (2.26). We go one step backward. For $\mathbb{P}$-almost all $\omega \in \Omega$, $\theta_\ell \in \Theta^d_{\ell-1}$, $\ell = 1, \dots, M-2$, (using the definition of V_{M-1}) we are interested by the function

$$\theta_{M-1} \mapsto V_{M-2}(\omega, \theta_{1:M-2}, \theta_{M-1}) := \mathbb{E}\left[\underbrace{\operatorname*{ess\,inf}_{\theta_M \in \Theta^d_{M-1}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M-2} \theta_\ell.\Delta X_\ell - \theta_{M-1}.\Delta X_{M-1} - \tilde{\theta}_M.\Delta X_M\right)_+ \middle| \mathcal{G}_{M-1}\right]}_{=V_{M-1}(\theta_{1:M-1},\tilde{\theta}_M)} \middle| \mathcal{G}_{M-2}\right](\omega).$$

It is straightforward that this function is convex. Let $\theta_{M-1},\ \theta^{'}_{M-2} \in \Theta^d_{M-2}$, using the standard inequality $|\operatorname{ess\,inf}_{i\in I} a_i - \operatorname{ess\,inf}_{i \in I} b_i| \le \operatorname{ess\,sup}_{i\in I}|a_i - b_i|$, we have

$$|V_{M-2}(\theta_{1:M-2}, \theta_{M-1}) - V_{M-2}(\theta_{1:M-2}, \theta^{'}_{M-1})| \le |\theta_{M-1} - \theta^{'}_{M-1}| \mathbb{E}[|\Delta X_{M-1}| \mid \mathcal{G}_{M-2}] \ \ a.s.,$$

so that the function is Lipschitz-continuous.

Now, owing to Jensen's inequality, Bayes' rule, and using similar arguments as those used in the proof of Propositions 2.7 and 2.8, we have

$$V_{M-1}(\theta_{1:M-1}, \tilde{\theta}_M) \ge \frac{1}{K}\left(\mathbb{E}_{\mathbb{P}^*}[L \mid \mathcal{G}_{M-1}] - \sum_{\ell=1}^{M-2} \theta_\ell.\Delta X_\ell - \theta_{M-1}.\Delta X_{M-1}\right)_+ \mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}} \middle| \mathcal{G}_{M-1}\right],$$

so that

$$\begin{aligned} &V_{M-2}(\theta_{1:M-2}, \theta_{M-1}) \\ &\quad \ge \frac{1}{K}\mathbb{E}\left[\left(\mathbb{E}_{\mathbb{P}^*}[L \mid \mathcal{G}_{M-1}] - \sum_{\ell=1}^{M-2} \theta_\ell.\Delta X_\ell - \theta_{M-1}.\Delta X_{M-1}\right)_+ \mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}} \middle| \mathcal{G}_{M-1}\right] \middle| \mathcal{G}_{M-2}\right] \\ &\quad = \frac{1}{K}\mathbb{E}_{\mathbb{P}^*}\left[\left(\mathbb{E}_{\mathbb{P}^*}[L \mid \mathcal{G}_{M-1}] - \sum_{\ell=1}^{M-2} \theta_\ell.\Delta X_\ell - \theta_{M-1}.\Delta X_{M-1}\right)_+ \middle| \mathcal{G}_{M-2}\right] \mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}} \middle| \mathcal{G}_{M-2}\right]. \end{aligned} \tag{2.30}$$

We aim at showing that the right-hand side of (2.30) goes to infinity as $|\theta_{M-1}| \to +\infty$. First, the subadditivity of the function $x \mapsto x_+$ implies that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^*}\left[(-\theta_{M-1}.\Delta X_{M-1})_+ \middle| \mathcal{G}_{M-2}\right] &\le \mathbb{E}_{\mathbb{P}^*}\left[\left(\mathbb{E}_{\mathbb{P}^*}[L|\mathcal{G}_{M-1}] - \sum_{\ell=1}^{M-1} \theta_\ell.\Delta X_\ell\right)_+ \middle| \mathcal{G}_{M-2}\right] \\ &\quad + \mathbb{E}_{\mathbb{P}^*}\left[\left(-\mathbb{E}_{\mathbb{P}^*}[L|\mathcal{G}_{M-1}] + \sum_{\ell=1}^{M-2} \theta_\ell.\Delta X_\ell\right)_+ \middle| \mathcal{G}_{M-2}\right]. \end{aligned}$$

We focus on the left-hand side of the above inequality, this quantity is lower bounded by

$$|\theta_{M-1}| \operatorname*{ess\,inf}_{u\in\Theta^d_{M-2},\ |u|=1} \mathbb{E}_{\mathbb{P}^*}\left[(u.\Delta X_{M-1})_+ \,\middle|\, \mathcal{G}_{M-2}\right].$$

Consequently, Assumption 2.9 implies that for $\mathbb{P}$-almost all $\omega\in\Omega$, the function $u\mapsto \mathbb{E}_{\mathbb{P}^*}\left[(u.\Delta X_{M-1})_+ \,\middle|\, \mathcal{G}_{M-2}\right](w)$ is continuous on $\mathcal{S}_d(0,1)$ so that there exists $u^*(w)$ such that $ess\,inf_{u\in\Theta^d_{M-2},\ |u|=1}\mathbb{E}_{\mathbb{P}^*}\left[(u.\Delta X_{M-1})_+ \,\middle|\, \mathcal{G}_{M-2}\right](w) = \mathbb{E}_{\mathbb{P}^*}\left[(u^*.\Delta X_{M-1})_+ \,\middle|\, \mathcal{G}_{M-2}\right](w)$. Now, if $u\in\Theta^d_{M-2}$ and $|u|=1$ is such that $\mathbb{E}_{\mathbb{P}^*}\left[(u.\Delta X_{M-1})_+ \,\middle|\, \mathcal{G}_{M-2}\right]=0$ $a.s.$, then $(u.\Delta X_{M-1})_+=0$ so that $(-u).\Delta X_{M-1}\geq 0$ $a.s.$ Using (NA), we clearly get $u.\Delta X_{M-1}=0$, $a.s.$ so that Assumption 2.9 implies $u=0$, which is impossible. Finally, $ess\,inf_{u\in\Theta^d_{M-2},\ |u|=1}\mathbb{E}_{\mathbb{P}^*}\left[(u.\Delta X_{M-1})_+ \,\middle|\, \mathcal{G}_{M-2}\right]>0$, $a.s.$ which in turn implies that for all $\theta_\ell\in\Theta^d_{\ell-1}$, $\ell=1,\ldots,M-2$, $\lim_{|\theta_{M-1}|\to+\infty} V_{M-2}(\omega,\theta_{1:M-2},\theta_{M-1})=+\infty$, and the function $\theta_{M-1}\mapsto V_{M-2}(\omega,\theta_{1:M-2},\theta_{M-1})$ has a minimum $\tilde{\theta}_{M-1}$, which is $\mathcal{G}_{M-2}$-measurable owing to measurable selection theorem.

Furthermore, using similar arguments to those used for the proof of Proposition 2.8, one shows that for $\mathbb{P}$-almost all $\omega\in\Omega$, for every $\theta_{1:M-2}\in\Theta^d_0\times\cdots\times\Theta^d_{M-3}$,

$$\begin{aligned}&\operatorname*{ess\,inf}_{\theta_{M-1}\in\Theta^d_{M-2}} V_{M-2}(\omega,\theta_{1:M-2},\theta_{M-1})\\ &\qquad = \operatorname*{ess\,inf}_{(\theta_{M-1},\theta_M)\in\Theta^d_{M-2}\times\Theta^d_{M-1}} \mathbb{E}\left[\left(L-\sum_{\ell=1}^{M}\theta_\ell.\Delta X_\ell\right)_+ \,\middle|\, \mathcal{G}_{M-2}\right] \quad a.s.\end{aligned}$$

Now if the solution is built down to step ℓ, for all $\theta_k\in\Theta^d_{k-1}$, $k=1,\ldots,\ell-1$, one shows that $\theta_\ell\mapsto V_{\ell-1}(\omega,\theta_{1:\ell-1},\theta_\ell)$ is convex, Lipschitz-continuous, and satisfies $\lim_{|\theta_\ell|\to+\infty} V_{\ell-1}(\omega,\theta_{1:\ell-1},\theta_\ell)=+\infty$. Consequently, there exists $\tilde{\theta}_\ell$ solution of (2.29). Thus, (2.23) has an optimal $\theta^*_\alpha := (\theta^*_{\alpha,\ell})_{0\leq\ell\leq M-1}$.

Now we come back to our original problem

$$\begin{aligned}&\inf_{\xi\in\mathbb{R}}\xi+\frac{1}{1-\alpha}\inf_{\theta\in\mathcal{A}_{\mathcal{G}}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^{M}\theta_\ell.\Delta X_\ell-\xi\right)_+\right]\\ &\qquad=\inf_{\xi\in\mathbb{R}}\xi+\frac{1}{1-\alpha}\mathbb{E}\left[\left(L-\sum_{\ell=1}^{M}\theta^*_{\alpha,\ell}.\Delta X_\ell-\xi\right)_+\right].\end{aligned}$$

For all $x\in\mathbb{R}$, the functions $\xi\mapsto\xi+\frac{1}{1-\alpha}(x-\xi)_+$ are convex and Lipschitz-continuous, so that $\xi\mapsto\xi+\frac{1}{1-\alpha}\inf_{\theta\in\mathcal{A}_{\mathcal{G}}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^{M}\theta_\ell.\Delta X_\ell-\xi\right)_+\right]$ is convex, Lipschitz-continuous. Owing to (NA) and Jensen's inequality, one easily obtains

$$\xi+\frac{1}{1-\alpha}\inf_{\theta\in\mathcal{A}_{\mathcal{G}}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^{M}\theta_\ell.\Delta X_\ell-\xi\right)_+\right]\geq\frac{1}{K}\left(\xi+\frac{1}{1-\alpha}\left(\mathbb{E}_{\mathbb{P}^*}[L]-\xi\right)_+\right),$$

so that $\lim_{|\xi|\to+\infty}\xi+\frac{1}{1-\alpha}\inf_{\theta\in\mathcal{A}_{\mathcal{G}}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^{M}\theta_\ell.\Delta X_\ell-\xi\right)_+\right]=+\infty$. This completes the proof. □

3. COMPUTATIONAL AND NUMERICAL ASPECTS OF CVaR HEDGING

In this section, we propose several methods to compute the optimal strategies of the three problems (1.3), (1.2), and (1.1). First, we will focus on (1.2) since it will be the main building block when we are going to propose several algorithms to approximate the optimal dynamic strategy solution of (1.1).

Note that, in order to reduce computational complexity of conditional expectations, we will assume that the underlying process is Markovian and compute them using optimal vector quantization grids. Doing so, the stochastic control problems (1.3), (1.2) will reduce to N_{ℓ_0} local convex optimization problems, where N_{ℓ_0} denotes the size of the quantization grid, that will be solved using stochastic gradient algorithm.

Concerning the multistep stochastic control problem (1.1), we will study several algorithms based on suboptimal time-consistent criteria and built on one-step procedures as a main tool.

3.1. Markovian Framework and Optimal Vector Quantization

For numerical considerations, as mentioned in the introduction, one relevant source of incompleteness of energy and financial markets is the case where there is a nonhedgeable source of risk represented by a process Z. Typically, in the electricity market, Z can be considered as the temperature process and may influence electricity spot prices and electricity forward prices. Hence, we will consider that $\mathcal{G}_\ell = \sigma\,(X_i, Z_i; 0 \leq i \leq \ell)$. As mentioned above, in order to simplify the numerical computation of conditional expectations that appear in (1.2), here as below, we will assume that (X, Z) is a discrete-time Markovian process. For instance, (X, Z) may represent an Euler-like scheme obtained by approximating the dynamic of a continuous time diffusion process.

To simplify the presentation of the numerical methodology, we will assume that (X_M, L) can be simulated from (X_{ℓ_0}, Z_{ℓ_0}) using a random vector U_{ℓ_0}, that is, there exists two continuous functions $F : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}$ and $G : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}^d$, such that

$$X_M - X_{\ell_0} = G(X_{\ell_0}, Z_{\ell_0}, U_{\ell_0+1}) \text{ and } L = F(X_{\ell_0}, Z_{\ell_0}, U_{\ell_0+1}),$$

where U_{ℓ_0+1} is a r_{ℓ_0}-dimensional random variable independent of $\mathcal{G}_{\ell_0}$. In this section, we will denote X for $X_M - X_{\ell_0}$ and U for U_{ℓ_0+1}. Under this simple Markovian framework, the function (2.13) can be written for all $(x, z) \in \mathbb{R}^d \times \mathbb{R}^q$

$$V(\xi, \theta, x, z) = \mathbb{E}\left[v(\xi, \theta, x, z, U)\right],$$

where $v(\xi, \theta, x, z, u) := \xi + \frac{1}{1-\alpha}\left(F(x, z, u) - \theta.G(x, z, u) - \xi\right)_+$ so that (2.19) becomes

$$\operatorname*{ess\,inf}_{(\xi,\theta)\in\Theta_{\ell_0}^{d+1}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ \Big| \mathcal{G}_{\ell_0}\right] = \left(\min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V(\xi, \theta, x, z)\right)_{(x,z)=(X_{\ell_0}, Z_{\ell_0})} \quad a.s. \tag{3.1}$$

Consequently, in order to solve the global problem (1.2) we need to solve the local optimization problem that appears in the right-hand side of the above equation for each

$(X_{\ell_0}(\omega), Z_{\ell_0}(\omega))$. Then, we have to estimate the quantity

$$\mathbb{E}\left[\left(\inf_{(\theta,\xi)\in\mathbb{R}^d\times\mathbb{R}} V(\xi,\theta,x,z)\right)_{|x=X_{\ell_0},z=Z_{\ell_0}}\right]. \tag{3.2}$$

When the dimension of the random variable (X_{ℓ_0}, Z_{ℓ_0}) is large (say greater than 10), one can use Monte Carlo simulations and estimates (3.2) using N samples by

$$\frac{1}{N}\sum_{k=1}^{N}\left(\inf_{(\theta,\xi)\in\mathbb{R}^d\times\mathbb{R}} V(\xi,\theta,x,z)\right)_{|x=X_{\ell_0,k},z=Z_{\ell_0,k}},$$

where $(X_{\ell_0,k}, Z_{\ell_0,k})_{1\le k\le N}$ are i.i.d. random vectors having the distribution of (X_{ℓ_0}, Z_{ℓ_0}).

When the dimension of the random variable (X_{ℓ_0}, Z_{ℓ_0}) is small (say less than 10), we can use an integration cubature formula based, for instance, on a spatial discretization of (X_{ℓ_0}, Z_{ℓ_0}). A commonly used method in such a framework is optimal vector quantization. Thus, we consider an optimal N_{ℓ_0}-quantization $(\hat{X}_{\ell_0}, \hat{Z}_{\ell_0})$ of the random variable (X_{ℓ_0}, Z_{ℓ_0}), based on an optimal quantization grid $\Gamma_{\ell_0} := \Gamma^{N_{\ell_0}}_{(X_{\ell_0},Z_{\ell_0})} = ((x^1_{\ell_0}, z^1_{\ell_0}), \ldots, (x^{N_{\ell_0}}_{\ell_0}, z^{N_{\ell_0}}_{\ell_0}))$. Then, if we denote $CV^*_\alpha(x^j_{\ell_0}, z^j_{\ell_0})$ for $\inf_{(\theta,\xi)\in\mathbb{R}^d\times\mathbb{R}} V(\xi,\theta,(x^j_{\ell_0}, z^j_{\ell_0}))$, $j = 1,\ldots,N_{\ell_0}$, the quantization-based quadrature formula to approximate (3.2) is given by

$$\sum_{j=1}^{N_{\ell_0}} CV^*_\alpha(x^j_{\ell_0}, z^j_{\ell_0})\mathbb{P}\left((X_{\ell_0}, Z_{\ell_0}) \in C_j(x_{\ell_0}, z_{\ell_0})\right), \tag{3.3}$$

where $(C_j(x_{\ell_0}, z_{\ell_0}))_{j=1,\ldots,N_{\ell_0}}$ is a Voronoi tessellation of the N_{ℓ_0}-quantizer Γ_{ℓ_0}. For more details about optimal vector quantization, including error bounds for cubature formulae, we refer to Pagès (1998).

Consequently, we need to compute the solution as well as the value of the objective function for all nodes of a quantization grid (or for all Monte Carlo samples). Thus, throughout this section, we will focus on the value function that appears within the brackets of the right-hand side of (3.1).

For the sake of simplicity, we will temporarily drop (x, z) in the notations so that we will denote $F(U)$ for $F(x,z,U)$, $G(U)$ for $G(x,z,U)$, $V(\xi,\theta)$ for $V(\xi,\theta,x,z)$, $V(\xi,\theta,U)$ for $v(\xi,\theta,x,z,U)$, and so on. Thus, we will omit "for all $(x,z)\in\mathbb{R}^d\times\mathbb{R}^q$" in any assumption or property given below about these distributions or functions.

3.2. Computing Optimal Strategies Using Stochastic Approximation: A First Naive Approach

The above local representation (3.1) naturally yields a stochastic gradient algorithm derived from the Lyapunov function V, which will converge toward $\left(\xi^*_\alpha, \theta^*_\alpha\right) \in Arg\,min\,V$. Then, following the procedure investigated in Bardou et al. (2009), a companion recursive procedure can be easily devised which has $CV^*_\alpha := CV^*_\alpha(x,z) = CVaR_\alpha\left(F(U) - \theta^*_\alpha.G(U)\right)$ as a target, i.e., the CVaR_α of the optimal portfolio at the point (x, z). Finally, in order to compute the value function at time ℓ_0, i.e., the global expectation (3.2), we will rely on the cubature formula based on optimal vector quantization given by (3.3).

First, we set

$$H_1(\xi, \theta, U) := \frac{\partial v}{\partial \xi}(\xi, \theta, U) = 1 - \frac{1}{1-\alpha} 1_{\{F(U)-\theta.G(U) \geq \xi\}}, \tag{3.4}$$

$$H_{2:d+1}(\xi, \theta, U) := \frac{\partial v}{\partial \theta}(\xi, \theta, U) = -\frac{1}{1-\alpha} G(U) 1_{\{F(U)-\theta.G(U) \geq \xi\}}, \tag{3.5}$$

so that

$$\nabla_{(\xi,\theta)} V(\xi, \theta) = \mathbb{E}\left[(H_1(\xi, \theta, U), H_{2:d+1}(\xi, \theta, U))\right].$$

Since we are looking for (ξ, θ) for which $\mathbb{E}[H_1(\xi, \theta, U)]$ and $\mathbb{E}[H_{2:d+1}(\xi, \theta, U)] = 0$, we implement a classical RM algorithm to approximate $(\xi_\alpha^*, \theta_\alpha^*)$, i.e., we define recursively for $n \geq 1$:

$$\xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, \theta_{n-1}, U_n), \tag{3.6}$$

$$\theta_n = \theta_{n-1} - \gamma_n H_{2:d+1}(\xi_{n-1}, \theta_{n-1}, U_n), \tag{3.7}$$

where $(U_n)_{n\geq 1}$ is an i.i.d. sequence of random vectors with the same distribution as U, independent of (ξ_0, θ_0), with $\xi_0 \in L^2_{\mathbb{R}}(\mathbb{P})$, $\theta_0 \in L^2_{\mathbb{R}^d}(\mathbb{P})$, and $(\gamma_n)_{n\geq 1}$ is a positive deterministic step sequence satisfying

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty. \tag{3.8}$$

Following Bardou et al. (2009), as a second step, in order to estimate $CVaR_\alpha(F(U) - \theta_\alpha^*.G(U))$, i.e., the CVaR_α of the local CVaR hedged loss, we devise a *companion* procedure using the same step sequence as (3.6) and (3.7), for $n \geq 1$,

$$C_n = C_{n-1} - \gamma_n H_{d+2}(\xi_{n-1}, \theta_{n-1}, C_{n-1}, U_n), \tag{3.9}$$

with $H_{d+2}(\xi, \theta, c, u) := c - v(\xi, \theta, u)$. In order to derive the $a.s.$ convergence of (3.6), (3.7), and (3.9), we introduce the following additional assumption on the distribution of $F(U)$ and $G(U)$.

ASSUMPTION 3.1. *There exists $a > 0$ such that $F(U) \in \mathbb{L}^{2a}(\mathbb{P})$ and $G(U) \in \mathbb{L}^{2a}(\mathbb{P})$.*

To establish the $a.s.$ convergence of $(\xi_n, \theta_n, C_n)_{n\geq 1}$, we will rely on RM Theorem (see, e.g., Duflo 1996). In fact, we will use a slight extension (which takes into account the case of nonuniqueness of the target). The statement of this classical theorem is given in the Appendix, we refer, e.g., to Lemaire and Pagès (2010) for a proof.

In the next proposition, we establish the $a.s.$ convergence of the sequence $(\xi_n, \theta_n, C_n)_{n\geq 1}$ toward its target $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$. The proof is given in the Appendix.

THEOREM 3.2. *Suppose that (NA), Assumptions 2.6 and 3.1 (with $a = 1$) hold. Assume that the step sequence $(\gamma_n)_{n\geq 1}$ satisfies the usual decreasing step assumption (3.8).*

Then the recursive procedure defined by (3.6), (3.7), and (3.9) satisfies:

$$\exists \ \left(\xi_\alpha^*, \theta_\alpha^*\right) : (\Omega, \mathcal{A}) \to \operatorname{Arg\,min} V \textit{(which is a compact set)},$$

such that

$$(\xi_n, \theta_n) \xrightarrow{a.s.} \left(\xi_\alpha^*, \theta_\alpha^*\right), \quad n \to +\infty.$$

Moreover,

$$C_n \xrightarrow{a.s.} CV_\alpha^* = \min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V(\xi, \theta) = V(\xi_\alpha^*, \theta_\alpha^*) \ \ n \to +\infty.$$

With regards to the rate of convergence, the global procedure composed by (3.6), (3.7), (3.9) is a regular stochastic algorithm that behaves as described in usual stochastic approximation textbooks like Benveniste, Métivier, and Priouret (1990), Duflo (1997), and Kushner and Yin (2003). As soon as $\mathcal{T}^*$ is reduced to a single point $\left(\xi_\alpha^*, \theta_\alpha^*\right)$ (the local CVaR CV_α^* is always unique), the procedure satisfies under quite standard assumptions a CLT at rate $\gamma_n^{-\frac{1}{2}}$. It is well known that the best asymptotic rate is obtained by specifying $\gamma_n = \frac{c}{b+n}$, $c, b > 0$. However, the choice of c is subject to a stringent condition involving $\left(\xi_\alpha^*, \theta_\alpha^*\right)$ (which is unknown to the user). This always induces a more or less (suboptimal) blind choice for the constant c.

To overcome this classical problem, we introduce the empirical mean of the global algorithm implemented with a slowly decreasing step "à la Ruppert & Polyak" (see, e.g., Pelletier 2000). First, we write the global algorithm in a more synthetic way by setting for $n \geq 1$,

$$\phi_n = (\xi_n, \theta_n, C_n), \quad \phi_0 = (\xi_0, \theta_0, C_0),$$

and

$$\phi_n = \phi_{n-1} - \gamma_n H(\phi_{n-1}, U_n), \tag{3.10}$$

where $H(\phi, u) = (H_1(\xi, \theta, u), H_{2:d+1}(\xi, \theta, u), H_{d+2}(\xi, \theta, c, u))$. Thus, the Cesàro mean of the procedure

$$\overline{\phi}_n = \frac{\phi_0 + \cdots + \phi_{n-1}}{n}, \quad n \geq 1, \tag{3.11}$$

where ϕ_n is defined by (3.10), *a.s.* converges to the same target. The Ruppert & Polyak (RP) Averaging Principle says that an appropriate choice of the step yields for free the optimal asymptotic rate and the smallest possible asymptotic variance. For sake of completeness, we recall in the Appendix this classical result following a version established in Pelletier (2000).

In order to derive the convergence rate of the averaged algorithm, we suppose that the conditional distribution of $(X_M - X_{\ell_0}, L)$ given $(X_{\ell_0}, Z_{\ell_0}) = (x_{\ell_0}, z_{\ell_0})$ has a probability density function $p_{(X_M - X_{\ell_0}, L)}^{(X_{\ell_0}, Z_{\ell_0}) = (x_{\ell_0}, z_{\ell_0})}$ for all $(x_{\ell_0}, z_{\ell_0}) \in \mathbb{R}^d \times \mathbb{R}^q$ that we will denote $p_{X,L}$ for the sake of simplicity. Moreover, in order to simplify notations, we will denote X the conditional distribution of $(X_M - X_{\ell_0})$ given $(X_{\ell_0}, Z_{\ell_0}) = (x_{\ell_0}, z_{\ell_0})$. Finally, we will denote by p_X and p_L the (conditional) marginal density functions of $X_M - X_{\ell_0}$ and L

(given $(X_{\ell_0}, Z_{\ell_0}) = (x_{\ell_0}, z_{\ell_0})$), respectively. Moreover, we make the following additional assumption on the joint conditional probability density function $p_{X,L}$.

ASSUMPTION 3.3.

(i) *For all* $x \in \mathbb{R}^d$, $y \mapsto p_{X,L}(x, y)$ *is continuous on* $\mathbb{R}$.

(ii) *For all* $\theta \in \mathbb{R}^d$, *for every compact set* $K \subset \mathbb{R}$, $\sup_{y \in K} p_{X,L}(x, \theta.x + y) \in L^1(dx)$.

(iii) *For all* $\xi \in \mathbb{R}$, *for every compact set* $K \subset \mathbb{R}^d$, $\sup_{\theta \in K}(1+|x|^2) p_{X,L}(x, \theta.x+\xi) \in L^1(dx)$.

(iv) $\int_{\mathbb{R}^d} p_{X,L}(x, \xi^*_\alpha + \theta^*_\alpha.x)dx > 0$, *for all* $(\xi^*_\alpha, \theta^*_\alpha) \in \operatorname{Arg\,min} V$.

EXAMPLE 3.4. We take the example (that will be studied in Section 5) of the energy provider who buys on an energy market a quantity $C_T = \mu_C + \sigma_C G_1$, where $G_1 \sim \mathcal{N}(0, 1)$ of gas at price S^g_T, where gas spot price is modeled as a geometric Brownian motion correlated with $\rho \in (0, 1)$ to the consumption, namely

$$S^g_T = S_0 e^{-\frac{\sigma_g^2}{2} T + \sigma_g \sqrt{T}\left(\rho G_1 + \sqrt{1-\rho^2} G_2\right)}.$$

This quantity is sold to consumers at a price $K = S_0$. The energy provider uses a one-step self-financed strategy based on S^g to reduce its risk so that $X = S^g_T - S_0$. The loss L can be written as

$$L = (S^g_T - K)C_T = (S^g_T - S_0)C_T.$$

Using the change of variable formula, one shows that the joint conditional distribution function writes for $x > -S_0$, $x \neq 0$,

$$p_{X,L}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_g\sigma_C\sqrt{T}} \frac{1}{(x+S_0)|x|} e^{-\frac{1}{2\sigma_g^2 T(1-\rho^2)}\left(\left(\log(\frac{x}{S_0}+1)+\frac{\sigma_g^2}{2}T\right)-\frac{\rho}{\sigma_C^2}\left(\frac{y}{x}-\mu_C\right)\right)^2} e^{-\frac{1}{2\sigma_C^2}\left(\frac{y}{x}-\mu_C\right)^2}.$$

In the next theorem, we use notations of Theorem 3.2. We establish a CLT for the empirical mean sequence $\overline{\phi}_n$ defined by (3.11).

THEOREM 3.5. *(Convergence rate of the procedure) Suppose that (NA), Assumptions 2.6, 3.1 (with $a > 1$), and 3.3 are satisfied. If the step sequence is $\gamma_n = \frac{\gamma_1}{n^\beta}$, with $\frac{1}{2} < \beta < 1$ and $\gamma_1 > 0$, then the averaged procedure defined by (3.11) satisfies*

$$\sqrt{n}\left(\overline{\phi}_n - \phi^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma), \quad \text{as} \quad n \to +\infty,$$

where the asymptotic covariance matrix Σ is given by

$$\Sigma = P^{-1}\Gamma(P^{-1})^T,$$

with

$$P := \frac{1}{1-\alpha}\begin{pmatrix} \displaystyle\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx & \left(\displaystyle\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx\right)^T & 0 \\ \displaystyle\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx & \displaystyle\int_{\mathbb{R}^d} x x^T p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx & 0 \\ 0 & 0 & 1-\alpha \end{pmatrix}, \tag{3.12}$$

and

$$\Gamma := \begin{pmatrix} \dfrac{\alpha}{1-\alpha} & 0 & \dfrac{\alpha\mathbb{E}\left[\left(L-\theta_\alpha^*.X-\xi_\alpha^*\right)_+\right]}{(1-\alpha)^2} \\ 0 & \dfrac{\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]}{(1-\alpha)^2} & \dfrac{\mathbb{E}\left[X\left(L-\theta_\alpha^*.X-\xi_\alpha^*\right)_+\right]}{(1-\alpha)^2} \\ \dfrac{\alpha\mathbb{E}\left[\left(L-\theta_\alpha^*.X-\xi^*\right)_+\right]}{(1-\alpha)^2} & \dfrac{\mathbb{E}\left[X\left(L-\theta_\alpha^*.X-\xi^*\right)_+\right]^T}{(1-\alpha)^2} & \dfrac{Var\left(\left(L-\theta_\alpha^*.X-\xi^*\right)_+\right)}{(1-\alpha)^2} \end{pmatrix}. \tag{3.13}$$

One may be interested by the asymptotic variance of each component of the algorithm, namely, ξ_n, θ_n, and C_n rather than the whole asymptotic matrix. The inverse matrix P^{-1} can be written as

$$P^{-1} := \frac{1-\alpha}{\displaystyle\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx} \times \begin{pmatrix} 1 + V^T\Pi^{-1}V & -V^T\Pi^{-1} & 0 \\ -\Pi^{-1}V & \Pi^{-1} & 0 \\ 0 & 0 & \dfrac{1}{1-\alpha}\displaystyle\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx \end{pmatrix}, \tag{3.14}$$

where

$$\Pi := \int_{\mathbb{R}^d}\left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx}\right)\left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx}\right)^T \times p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx,$$

and

$$V := \frac{1}{\displaystyle\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx}\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx,$$

so that, for $i = 2, \ldots, d+1$,

$$\Sigma_{1,1} = \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx\right)^2} \left(\left(1 + V^T \Pi^{-1} V\right)^2 \alpha(1-\alpha) \right.$$

$$\left. + \left(\Pi^{-1} V\right)^T \mathbb{E}\left[XX^T 1_{\{L-\theta_\alpha^*.X \geq \xi_\alpha^*\}}\right] \Pi^{-1} V\right), \tag{3.15}$$

$$\Sigma_{i,i} = \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx\right)^2} \left(m_i^2 \alpha(1-\alpha) + \tilde{m}_{.i}^T \mathbb{E}\left[XX^T \mathbf{1}_{\{L-\theta_\alpha^*.X \geq \xi_\alpha^*\}}\right] \tilde{m}_{.i}\right), \tag{3.16}$$

$$\Sigma_{d+2,d+2} = \left(\frac{1}{1-\alpha}\right)^2 Var\left(\left(L - \theta_\alpha^*.X - \xi_\alpha^*\right)_+\right), \tag{3.17}$$

where $m = \Pi^{-1} V$ and $\Pi^{-1} = \left(\tilde{m}_{i,j}\right)_{1 \leq i \leq d, 1 \leq j \leq d}$.

3.3. Dynamic CVaR Hedging

In this section, we propose several methods to compute the optimal strategy of (1.1), the VaR, and the CVaR of the optimal portfolio. Except for the first procedure, the idea is to devise hybrid suboptimal strategies obtained as optimal strategies for a closely connected time-consistent multistage control stochastic optimization problem. Then, we will challenge these strategies from a numerical point of view by plugging them into the original CVaR criterion.

From a modeling point of view, we assume that (X, Z) is a discrete (inhomogeneous) Markov chain simulated by

$$X_\ell - X_{\ell-1} = G_\ell(X_{\ell-1}, Z_{\ell-1}, U_\ell),$$

where for all $\ell \in \{1, \ldots, M\}$, $G_\ell : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_\ell} \to \mathbb{R}^d$ is a continuous function and U_ℓ is a r_ℓ-dimensional random variable independent of $\mathcal{G}_{l-1}$. For the sake of simplicity, we also assume that the loss distribution is of the form $L = F(X_M, Z_M)$, where $F : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}$ is a continuous function.

3.3.1. Crude CVaR Hedging Algorithm (C.H.). A natural method to solve (1.1) is to proceed as in the one-step case by devising an optimal vector quantization-based stochastic gradient algorithm. Indeed, at every time t_k, we consider an optimal N_k-quantization $(\hat{X}_k, \hat{Z}_k)$, $k = 1, \ldots, M-1$, based on an optimal quantization grid $\Gamma_k = ((x_k^1, z_k^1), \ldots, (x_k^{N_k}, z_k^{N_k}))$ of the state process (X, Z) at time t_k. Once this sequence of grids is built, we associate an optimal local strategy $\theta_k^{j,*}$ to the node j of the grid Γ_k at time k. All local strategies will be estimated using a unique global stochastic gradient algorithm. This can be done by the following stochastic algorithm, namely, for all $k \in \{1, \ldots, M-1\}$

and all $j \in \{1, \ldots, N_k\}$,

$$\xi_n = \xi_{n-1} - \gamma_n H_1 \left(\xi_{n-1}, \theta_{n-1}, U_n\right), \tag{3.18}$$

$$\theta_{0,n} = \theta_{0,n-1} - \gamma_n H_{2,0} \left(\xi_{n-1}, \theta_{n-1}, U_n\right), \tag{3.19}$$

$$\theta_{k,n}^j = \theta_{k,n-1}^j - \gamma_n H_{2,k}^j \left(\xi_{n-1}, \theta_{n-1}, U_n\right), \tag{3.20}$$

$$C_n = C_{n-1} - \gamma_n H_3 \left(\xi_{n-1}, \theta_{n-1}, C_{n-1}, U_n\right), \tag{3.21}$$

where $\theta_n = (\theta_{0,n}, \ldots, \theta_{M-1,n})$, $(U_n)_{n\geq 1} = \left((U_{1,n}, \ldots, U_{M,n})_{n\geq 1}\right)$ are i.i.d. random variables with $U_{\ell,n} \sim U_\ell$ and for all $k \in \{1, \ldots, M\}$ and all $j \in \{1, \ldots, N_k\}$, the functions H_1, H_3, and $H_{2,k}^j$ are defined by

$$\begin{aligned}
H_1(\xi, \theta, u) &= 1 - \frac{1}{1-\alpha} 1_{\left\{L - \sum_{i=1}^M \theta_i . \Delta X_i \geq \xi\right\}}, \\
H_{2,0}(\xi, \theta, u) &= -\frac{G_1(X_0, Z_0, u_1)}{1-\alpha} 1_{\left\{L - \sum_{i=1}^M \theta_i . \Delta X_i \geq \xi\right\}}, \\
H_{2,k}^j(\xi, \theta, u) &= -\frac{G_{k+1}(X_k, Z_k, u_{k+1})}{1-\alpha} 1_{\left\{L - \sum_{i=1}^M \theta_i . \Delta X_i \geq \xi\right\}} 1_{\{(X_k, Z_k) \in C_j(x_k, z_k)\}}, \\
H_3(\xi, \theta, C, u) &= C - \xi - \frac{1}{1-\alpha}\left(L - \sum_{i=1}^M \theta_i . \Delta X_i - \xi\right)_+ .
\end{aligned}$$

The sequence $(\xi_n, \theta_n, C_n)_{n\geq 1}$ *a.s.* converges toward its target $\left(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*\right)$. Note that the dimension of the sequence (ξ_n, θ_n, C_n) to be updated at each step of the algorithm is equal to $D := 2 + d + \sum_{\ell=2}^M d \times N_\ell$.

Let us note that a careful reading of Section 2.4 shows that the optimal number of shares to be held over the time period $(k, k+1]$, θ_k^* depends on the whole process (X, Z), in fact, one can prove that it is a measurable function of $(X_k, Z_k, \sum_{\ell=1}^{k-1} \theta_\ell \Delta X_\ell)$. Hence, as a first approximation, we implicitly assumed that θ_k^* depends only on the state process at time k, (X_k, Z_k).

When the dimension is low ($D \leq 100$), which is often due to the fact that the number of trading dates is low (say $M \leq 5$) and the number of traded assets used for the hedging is small ($d \approx 1, 2$), the above algorithm is very efficient and we observe a great reduction of the CVaR compared to the static case (1.3).

However, if we consider a portfolio with a time horizon $T = 1$ year, 12 trading dates (one each month), if the investor hedges using 5 stocks and in the case where all layers in the quantization grids have the same size, i.e., $N = N_\ell = 5$, $\ell = 1, \ldots, M$, the dimension of the algorithm is $D = 282$. This example is a reasonable case in the energy sector when an energy company has to provide electricity or gas to consumers all year long and simultaneously needs to control and hedge its risk every month using electricity and/or gas forward contracts since the underlying spot is not storable. For instance, one may use 12 forward contracts with maturity ℓ that deliver electricity or gas over a period which corresponds to each month of the considered year. From a practical point of view, Electricity and Gas Futures market enable to trade: the next 3 months, the next 2 quarters, and the next 3 electricity or gas seasons (see the Powernext Gas Futures market, for instance), thus, one may proceed to a rough risk hedging using only some of these contracts. However, when dealing with a portfolio that depends on several energy commodities as it is often the case in the energy market, the dimension of the considered RM algorithm becomes a real issue.

From a numerical point of view, we observe that in a high-dimensional framework the algorithm "freezes" and "suffers" from a lack of exploration of the state space, as soon as the dimension is greater than, say, 100 or 150. Moreover, we observe that some components of θ_n are never updated by the algorithm. That is the bottleneck of this first algorithm in practical implementation. To overcome this problem, we propose several approximate solutions to solve (1.1). These solutions have the major advantage to dramatically reduce the dimension of the above algorithm.

3.3.2. Backward Dynamic Hedging Strategy (B.H.). This strategy is based on (2.6) and consists in a backward resolution. To be more precise, if we consider M trading dates, then (2.6) implies that in order to hedge the risk at the last trading date $M-1$, we have to solve

$$\inf_{\theta_M \in \Theta^d_{M-1}} \mathbb{E}\,[\mathcal{G}_{M-1}\text{--}CVaR_\alpha\,(L-\theta_M.\Delta X_M)]\,.$$

The optimization problem that appears in the right-hand side of the above equality can be easily solved using the static algorithm developed in Section 3.2. Now that we have the solution θ^b_M of this problem, we can go one step backward and solve the new problem

$$\inf_{\theta_{M-1} \in \Theta^d_{M-2}} \mathbb{E}\left[\mathcal{G}_{M-2}\text{--}CVaR_\alpha\left(L-\theta^b_M.\Delta X_M-\theta_{M-1}.\Delta X_{M-1}\right)\right],$$

using again the algorithm developed in the static framework in order to obtain θ^b_{M-1}. Following this idea till time 0, we obtain step by step the backward hedging strategy $\theta^b \equiv \left(\theta^b_\ell\right)_{1\le \ell\le M-1}$.

Although this method is not optimal from a theoretical point of view, it has the advantage to provide a strategy which controls the risk at each time step until maturity and which is time consistent. However, we observe on numerical experiments that the resulting static CVaR related to this self-financed strategy θ^b, namely,

$$\mathrm{CVaR}_\alpha\left(L-\sum_{\ell=1}^{M}\theta^b_\ell.\Delta X_\ell\right),$$

is significantly higher than the one obtained by the first global algorithm (C.H.). The reason is that by solving at step $k+1$, the optimization problem

$$\inf_{\theta_{M-k} \in \Theta^d_{M-k-1}} \mathbb{E}\left[\mathcal{G}_{M-k-1}\text{--}\mathrm{CVaR}_\alpha\left(L-\sum_{\ell=M-k+1}^{M}\theta^b_\ell.\Delta X_\ell-\theta_{M-k}\Delta X_{M-k}\right)\right],$$

there is an error (compared to the original problem (1.1)) on the estimate $\theta^b_{M-\ell} \neq \theta^*_{M-\ell}$, which propagates at each step and can become more and more important as the number of trading dates increases. That is the major drawback of this procedure.

3.3.3. Dynamic Hedging Strategy Based on a Martingale Decomposition of L (M.D.H.). This method is based on the subadditivity of the CVaR and on the following decomposition of the loss L into a sum of $\mathbb{G}$-martingale increments, namely,

$$L = \mathbb{E}[L] + \sum_{\ell=1}^{M} \tilde{\Delta} L_\ell, \tag{3.22}$$

where $\tilde{\Delta} L_\ell = \mathbb{E}[L|\mathcal{G}_\ell] - \mathbb{E}[L|\mathcal{G}_{\ell-1}]$, $1 \le \ell \le M$. Now, using the subadditivity of the CVaR, we obtain

$$\inf_{\theta \in \mathcal{A}_{\mathcal{G}}} \mathrm{CVaR}_\alpha \left(L - \sum_{\ell=1}^{M} \theta_\ell \Delta X_\ell \right) \le \mathbb{E}[L] + \sum_{\ell=1}^{M} \inf_{\theta_\ell \in \Theta_{\ell-1}^d} \mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_\ell . \Delta X_\ell \right). \tag{3.23}$$

The right-hand side of the last inequality shows that for each time step we have to solve a one-step static local CVaR-hedging problem. From a numerical point of view, indeed, when the dimension of the algorithm is not too large (say $D \le 150$), we observe that the CVaR obtained using this strategy is almost equal to the optimal one. When the dimension D becomes large, we observe a good behavior with a real improvement on the CVaR when the number of trading dates increases. An even better behavior is obtained by slightly modifying this second approach. We use the inequality

$$\inf_{\theta_\ell \in \Theta_{\ell-1}^d} \mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-}\mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_\ell . \Delta X_\ell \right) \right] \le \inf_{\theta_\ell \in \Theta_{\ell-1}^d} \mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_\ell . \Delta X_\ell \right), \tag{3.24}$$

and switch to the new time-consistent multistage optimization problem

$$\sum_{\ell=1}^{M} \inf_{\theta_\ell \in \Theta_{\ell-1}^d} \mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-}\mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_\ell . \Delta X_\ell \right) \right]. \tag{3.25}$$

The main difference in solving the local problem in the left-hand side of (3.24) compared to the right-hand side appears in the variable ξ_n of the two associated RM algorithms, which corresponds to the estimate at step n of the VaR_α. In this new version, like in the static case, the variable ξ_n is local and depends of the considered nodes whereas in the other version this variable is *global* and is the same for all nodes of the quantization tree.

Although, to our knowledge, there is neither an equality nor an inequality between the original problem and (3.25), numerical experiments led us to the conclusion that this last algorithm behaves better than the one obtained by solving the right-hand side of (3.24) at each time step. To be more precise, the original CVaR estimated by the strategy obtained by (3.25) is lower than the one obtained by the strategy solution of the right-hand side in (3.23).

In order to solve (3.25), we use optimal vector quantization again to approximate the unknown random variable $\tilde{\Delta} L_\ell$, i.e., we approximate $\mathbb{E}[L|\mathcal{G}_\ell]$ by using the cubature formula

$$\mathbb{E}[F_{\ell+1}(X_\ell, Z_\ell, U_{\ell+1}) | (X_\ell, Z_\ell)] \approx \varphi(X_\ell, Z_\ell) = \sum_{j=1}^{N_\ell} F(X_\ell, Z_\ell, u_{\ell+1}^j) \mathbb{P}\left(U_{\ell+1} \in C_j(u_{\ell+1}) \right), \tag{3.26}$$

and design at each time step a RM algorithm based on the procedure investigated in the static framework. One may have considered the classical decomposition

$$L = L_0 + \sum_{\ell=1}^{M} \Delta L_\ell, \tag{3.27}$$

instead of (3.22). The method based on this classical decomposition of L will be called *C.D.H.*

3.4. Design of Faster Procedures: Variance Reduction Techniques

In practice, the convergence of the different considered algorithms (static and dynamic frameworks) will be slow and chaotic when the confidence level α is close to 1. This is due to the fact that they are only updated on rare events since they try to measure and reduce the tail distribution: $\mathbb{P}\left(L - \theta_\alpha^*.X > \xi_\alpha^*\right) = 1 - \alpha \approx 0$. Another problem may be the simulation of L and X. Each evaluation may require a lot of computational efforts and takes a long time when L is representative of the loss of a huge and complex portfolio. So, for practical implementation, it is necessary to combine the above procedures with variance reduction techniques to achieve accurate estimates at a reasonable cost.

In Frikha (2010), two variance reduction techniques have been developed in order to reduce the asymptotic variance in the CLT (3.12). The first one is based on the unconstrained IS stochastic algorithm originally developed in Lemaire and Pagès (2010) and then applied to both VaR and CVaR in Bardou et al. (2009). Assume that U has an absolutely continuous distribution $\mathbb{P}_U(du) = p(u)\lambda_r(du)$, where λ_r denotes the Lebesgue measure on $(\mathbb{R}^r, \mathcal{B}or(\mathbb{R}^r))$. The starting idea of IS (by translation) applied to stochastic approximation algorithm like (3.10) is to use the invariance of the Lebesgue measure by translation to show that for every $\mu \in (\mathbb{R}^r)^{d+2}$,

$$\mathbb{E}\left[H_i(\xi, \theta, U)\right] = \mathbb{E}\left[H_i(\xi, \theta, U + \mu_i)\frac{p(X + \mu_i)}{p(X)}\right], \quad i = 1, \ldots, d+1 \tag{3.28}$$

$$\mathbb{E}\left[H_{d+2}(\xi, \theta, c, U)\right] = \mathbb{E}\left[H_{d+2}(\xi, \theta, c, U + \mu_{d+2})\frac{p(X + \mu_{d+2})}{p(X)}\right]. \tag{3.29}$$

It is easy to obtain a new CLT using these new random variables with the same expectations following the lines of Theorem 3.5. However, now we want to select μ so that the asymptotic variance covariance matrix is minimal in a specific sense, namely, that $\mathbb{E}\left[H_i^2(\xi_\alpha^*, \theta_\alpha^*, U + \mu_i)\frac{p^2(X+\mu_i)}{p^2(X)}\right]$ is minimal over all $\mu_i \in \mathbb{R}^r$, $i = 1, \ldots, d+1$ (idem with $i = d+2$). This yields a new minimization problem that can be solved by a new stochastic approximation procedure. As $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$ are unknown, we combine these new procedures in an adaptive way with the algorithm (3.10) in its averaged form following the ideas developed in Bardou et al. (2009) for the recursive computation of VaR and CVaR. To minimize the three variances $\mathrm{Var}(1_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}})$, $\mathbb{E}[X_i^2 1_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}]$, and $\mathrm{Var}((L - \theta_\alpha^*.X - \xi_\alpha^*)_+)$, respectively, using (3.28) and (3.29), we are led to minimize the functions

$Q_i(., \xi_\alpha^*, \theta_\alpha^*)$ defined for $\mu = (\mu_1, \dots, \mu_{d+2}) \in (\mathbb{R}^r)^{d+2}$ by,

$$Q_1(\mu_1, \xi_\alpha^*, \theta_\alpha^*) := \mathbb{E}\left[\mathbf{1}_{\{L^{(+\mu_1)} - \theta_\alpha^*.X^{(+\mu_1)} \geq \xi_\alpha^*\}} \frac{p^2(U+\mu_1)}{p^2(U)}\right],$$

$$Q_i(\mu_i, \xi_\alpha^*, \theta_\alpha^*) := \mathbb{E}\left[\left(X_i^{(+\mu_i)}\right)^2 \mathbf{1}_{\{L^{(+\mu_i)} - \theta_\alpha^*.X^{(+\mu_i)} \geq \xi_\alpha^*\}} \frac{p^2(U+\mu_i)}{p^2(U)}\right],$$

$$Q_{d+2}(\mu_{d+2}, \xi_\alpha^*, \theta_\alpha^*) := \mathbb{E}\left[\left(L^{(+\mu_{d+2})} - \theta_\alpha^*.X^{(+\mu_{d+2})} - \xi_\alpha^*\right)_+^2 \frac{p^2(U+\mu_{d+2})}{p^2(U)}\right],$$

where for the sake of simplicity we use the notations $L^{(\pm\mu)} = F(U \pm \mu) = F(x, z, U \pm \mu)$, $X^{(\pm\mu)} = G(U \pm \mu) = G(x, z, U \pm \mu) - x$, and $X_i^{(\pm\mu)} = G_i(U \pm \mu) - x_i$, for $\mu \in \mathbb{R}^r$.

Under some classical log-concavity hypothesis on p_U (for more details we refer to Lemaire and Pagès 2010), one shows that Q_i is finite, convex, differentiable on $\mathbb{R}^r$ so that if we define $W(\mu, u) = \frac{p^2(u-\mu)}{p(u)p(u-2\mu)} \frac{\nabla p(u-2\mu)}{p(u-2\mu)}$, we have

$$\nabla Q_1(\mu, \xi_\alpha^*, \theta_\alpha^*) = \mathbb{E}\left[\mathbf{1}_{\{L^{(-\mu)} - \theta_\alpha^*.X^{(-\mu)} \geq \xi_\alpha^*\}} W(\mu, U)\right], \tag{3.30}$$

$$\nabla Q_i(\mu, \xi_\alpha^*, \theta_\alpha^*) = \mathbb{E}\left[\left(X_i^{(-\mu)}\right)^2 \mathbf{1}_{\{L^{(-\mu)} - \theta_\alpha^*.X^{(-\mu)} \geq \xi_\alpha^*\}} W(\mu, U)\right], \tag{3.31}$$

$$\nabla Q_{d+2}(\mu, \xi_\alpha^*, \theta_\alpha^*) = \mathbb{E}\left[\left(L^{(-\mu)} - \theta_\alpha^*.X^{(-\mu)} - \xi_\alpha^*\right)_+^2 W(\mu, U)\right], \tag{3.32}$$

for all $\mu \in \mathbb{R}^r$. Moreover, one shows that $\lim_{|\mu| \to +\infty} Q_i(\mu, \xi_\alpha^*, \theta_\alpha^*) = +\infty$ so that $\arg\min Q_i(., \xi_\alpha^*, \theta_\alpha^*) = \left\{\mu \in \mathbb{R}^r \mid \nabla_\mu Q_i(\mu, \xi_\alpha^*, \theta_\alpha^*) = 0\right\}$ is nonempty.

Equations (3.30), (3.31), and (3.32) may look complicated at first glance but in fact the weight term $W(\mu, U)$ can be easily controlled by a deterministic function of μ since

$$|W(\mu, u)| \leq e^{2\rho|\mu|^b}(A|u|^{b-1} + A|\mu|^{b-1} + B), \tag{3.33}$$

for some real constants ρ, A, and B (for more details we refer to Lemaire and Pagès 2010 and Frikha 2010). In the case of a normal distribution $U \stackrel{d}{=} \mathcal{N}(0; 1)$,

$$W(\mu, U) = e^{\mu^2}(2\mu - U).$$

Now, if we have a control on the growth of the function F and G, typically for some positive constants C and c

$$\begin{cases} \forall u \in \mathbb{R}^r, |F(u)| \leq \tilde{F}(u) \quad \text{and} \quad \tilde{F}(u+v) \leq C(1+\tilde{F}(u))^c(1+\tilde{F}(v))^c, \\ \forall u \in \mathbb{R}^r, |G(u)| \leq \tilde{G}(u) \quad \text{and} \quad \tilde{G}(u+v) \leq C(1+\tilde{G}(u))^c(1+\tilde{G}(v))^c, \\ \qquad \mathbb{E}\left[|U|^{2(b-1)}\left(\tilde{F}(U)^{4c} + \tilde{G}(U)^{4c}\right) + \tilde{G}(U)^{4c}\right] < +\infty, \end{cases} \tag{3.34}$$

then we can define, for $\mu \in (\mathbb{R}^r)^{d+2}$

$$K_1(\mu_1, \xi_\alpha^*, \theta_\alpha^*, U) = e^{-2\rho|\mu_1|^b} \mathbf{1}_{\{L^{(-\mu_1)} - \theta_\alpha^*.X^{(-\mu_1)} \geq \xi_\alpha^*\}} W(\mu_1, U), \tag{3.35}$$

$$K_i(\mu_i, \xi^*_\alpha, \theta^*_\alpha, U) = \frac{e^{-2\rho|\mu_i|^b}}{1+\tilde{G}(-\mu_i)^{2c}} \left(X_i^{(-\mu_i)}\right)^2 \mathbf{1}_{\{L^{(-\mu_i)}-\theta^*_\alpha . X^{(-\mu_i)} \geq \xi^*_\alpha\}} W(\mu_i, U),\ i=2,\dots,d+1 \tag{3.36}$$

$$K_{d+2}(\mu_{d+2}, \xi^*_\alpha, \theta^*_\alpha, U) = \frac{e^{-2\rho|\mu_{d+2}|^b}}{1+\tilde{F}(-\mu_{d+2})^{2c}+|\theta^*_\alpha|^{2c}\tilde{G}(-\mu_{d+2})^{2c}} \left(L^{(-\mu_{d+2})}-\theta^*_\alpha . X^{(-\mu_{d+2})}-\xi^*_\alpha\right)_+^2 W(\mu_{d+2}, U), \tag{3.37}$$

so that it satisfies the linear growth assumption (A.2) of the RM Theorem and for $i = 1, \dots, d+2$

$$\left\{\mu_i \in \mathbb{R}^r \mid \mathbb{E}\left[K_i(\mu_i, \xi^*_\alpha, \theta^*_\alpha, U)\right] = 0\right\} = \left\{\mu_i \in \mathbb{R}^r \mid \nabla_{\mu_i} Q_i(\mu_i, \xi^*_\alpha, \theta^*_\alpha) = 0\right\}.$$

Moreover, since Q_i is convex, $\nabla_{\mu_i} Q_i$ satisfies (A.1). Now, the RM algorithms defined for $n \geq 1$ by

$$\mu_{i,n} = \mu_{i,n-1} - \gamma_n K_i(\mu_{i,n-1}, \xi^*_\alpha, \theta^*_\alpha, U_n),\ \mu_{i,0} \in \mathbb{R}^r,$$

a.s. converges to an $\operatorname{Arg\,min} Q_i(., \xi^*_\alpha, \theta^*_\alpha)$ (square integrable) random variable $\mu^*_{i,\alpha}$ (for more details about unconstrained recursive IS, we refer to Lemaire and Pagès 2010 and Bardou et al. 2009). Since we do not know either ξ^*_α and θ^*_α, respectively, we make the whole procedure adaptive by replacing at step n, these unknown parameters by their running approximation at step $n-1$. This finally justifies to introduce the following global procedure. One defines the state variable, for $n \geq 0$,

$$\phi_n = (\xi_n, \theta_n, C_n, \mu_{1,n}, \dots, \mu_{d+2,n}), \tag{3.38}$$

where $\xi_n,\ \theta_n,\ C_n$ denotes the VaR_α, the regression vector, and the CVaR_α estimates at step n, μ_1 denotes the variance reducer for the VaR_α, μ_i denotes the variance reducer for the ith component of θ^*_α, i.e., $\theta^*_{i,\alpha}$ and μ_{d+2} denote the variance reducer for the CVaR_α. We update this state variable recursively by

$$\phi_n = \phi_{n-1} - \gamma_n L(\phi_{n-1}, U_n),\ n \geq 1, \tag{3.39}$$

where $(U_n)_{n\geq 1}$ is an i.i.d. sequence with distribution U (and probability density p) and for $i = 2, \dots, d+1$,

$$L_1(\xi, \theta, \mu_1, u)) = e^{-\rho|\mu_1|^b}\left(1 - \frac{1}{1-\alpha}\mathbf{1}_{\{L^{(+\mu_1)}-\theta . X^{(+\mu_1)} \geq \xi\}} \frac{p(u+\mu_1)}{p(u)}\right), \tag{3.40}$$

$$L_i(\xi, \theta, \mu_i, u)) = \frac{e^{-\rho|\mu_i|^b}}{\left(1+\tilde{G}^{2c}(-\mu_i)\right)^{1/2}} X^{(+\mu_i)} \mathbf{1}_{\{L^{(+\mu_i)}-\theta . X^{(+\mu_i)} \geq \xi\}} \frac{p(u+\mu_i)}{p(u)}, \tag{3.41}$$

$$L_{d+2}(\xi, \theta, C, \mu_{d+2}, u) = C - \xi - \frac{1}{1-\alpha}\left(L^{(+\mu_{d+2})} - \theta . X^{(+\mu_{d+2})} - \xi\right)_+ \frac{p(u+\mu_{d+2})}{p(u)}, \tag{3.42}$$

$$L_{d+2+j}\left(\xi,\theta,\mu_j,u\right) = K_j(\mu_j,\xi,\theta,u), \quad j = 1,\ldots,d+2. \tag{3.43}$$

In Frikha (2010), it is shown that the algorithm (3.39) behaves as expected, i.e., it *a.s.* converges toward its target and that its empirical mean satisfies a Gaussian CLT with optimal rate and minimal asymptotic variances.

The second variance reduction tool is based on linear control variate. We use a control variable based on X, if $\mathbb{E}[X]$ is known and is easy to compute, e.g., if the price process is a $(\mathcal{G},\mathbb{P})$-martingale. For more details, we refer to Frikha (2010) where we only develop and study these two methods in the static self-financed strategy framework though it can be easily generalized to the other considered algorithm.

4. NUMERICAL EXAMPLES

4.1. Static Setting

First, we consider two simple examples in the static framework in order to show the efficiency of the CVaR hedging algorithm and of the two variance reduction techniques. For all examples, we use RM algorithm with two phases (see Remark 4.1) combined with the RP averaging principle. In all examples, we define the step sequence by $\gamma_n = \frac{1}{n^p}$, with $p = \frac{3}{4}$.

4.1.1. Spark Spread. We consider a short position on an exchange option between gas and electricity (called spark spread). Since electricity has very limited storage possibilities, the seller of this option hedges by trading only gas spot contracts. The process Z can be considered as the electricity spot price since it is observable on the energy market but cannot be used to set up hedging strategies. We choose to model the price of the two spot contracts by the Black–Scholes model with a correlation $\rho = 0.8$ between the two Brownian motions. The loss L can be written

$$L = \left(S_T^e - h_R S_T^g - C\right)_+,$$

where the time horizon $T = 1$ (year), the heat rate $h_R = 4$ BTU/kWh (BTU: British Thermal Unit), the generation costs $C = 3\$/\text{MWh}$, the two volatilities $\sigma_g = 0.4, \sigma_e = 0.8$, and the electricity and gas initial spot prices are $S_0^e = 40\$/\text{MWh}$, $S_0^g = 3\$/\text{MMBTU}$. The seller of the option uses a self-financed static strategy based on the gas spot price in order to reduce its risk at time $t_0 = 0$. Thus, its optimal strategy is given by the solution of (1.3) with $\ell = 0$. A crude Monte Carlo yields $\mathbb{E}[L] = 11.86$ with a variance of 3,692 after 3,000,000 trials.

The variance ratios correspond to an estimate of the asymptotic variance obtained without any variance reduction techniques, i.e., (3.15), (3.16), and (3.17) divided by an estimate of the asymptotic variance using IS (column (IS)) or LCV (column (LCV)) (see the asymptotic matrix obtained using IS and LCV in Frikha 2010).

In this example, the LCV method based on X does not provide any variance reduction. However, for the CVaR component, we use the control

$$\Lambda = 1_{\{S_T^e \geq q_\delta^e\}} - (1-\delta),$$

TABLE 4.1
One-Step Self-Financed Static CVaR-Hedging of Spark Spread Option

	No hedging		Static hedging						
α	VaR	CVaR	VaR	θ^*_α	CVaR	VR_{VaR} (IS)	VR_{Reg} (IS)	VR_{CVaR} (IS)	VR_{CVaR} (LCV)
95%	65.1	114.4	63.1	7.8	98.3	3.0	1.9	16.7	2.0
99%	142.2	208.3	120.2	13.6	163.2	3.7	2.3	19.0	1.7
99.5%	183.1	257.8	146.8	16.4	190.2	4.5	3.0	20.2	1.5

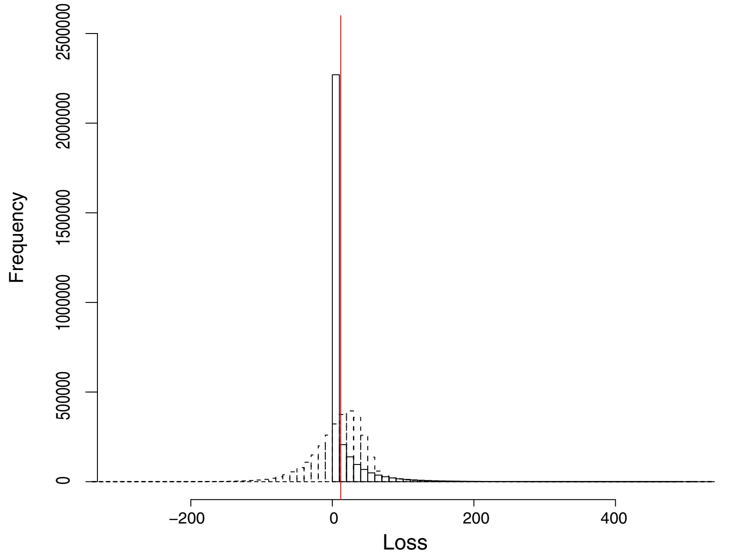

FIGURE 4.1. Histogram of consumption loss with (dashed lines) and without (normal lines) one-step CVaR-hedging at level $\alpha = 95\%$. The vertical line is the mean of the portfolio loss distribution.

where q^e_δ is the quantile of S^e_T at level δ. We choose: $\delta = 0.995$ ($q^e_\delta \approx 228.04$) for $\alpha = 0.95$, $\delta = 0.999$ ($q^e_\delta \approx 344.15$) for $\alpha = 0.99$, and $\delta = 0.9995$ ($q^e_\delta \approx 403.95$) for $\alpha = 0.9995$. The results obtained for three different values of the confidence level $\alpha = 95\%,\ 99\%,\ 99.5\%$ after 3,000,000 iterations of the RM procedure are specified in Table 4.1. We provide the VaR and CVaR of the loss without any hedging strategy, which are computed using the RM procedure developed in Bardou et al. (2009).

To complete this numerical example, we provide the histograms of the loss obtained with and without hedging. We clearly see in Figure 4.1 that the asymmetry of the histogram has been changed from right (loss) to left (gain) so that gains are more likely to occur with the hedged portfolio. In order to change the right tail distribution of the loss, the mode of the original portfolio has been greatly reduced and slightly translated to the right. Figure 4.2 confirms this idea: in order to hedge rare events that happen in the right tail distribution, the strategy consists in enlarging the left tail distribution. This induces a slight reduction of the mode and its translation to the right.

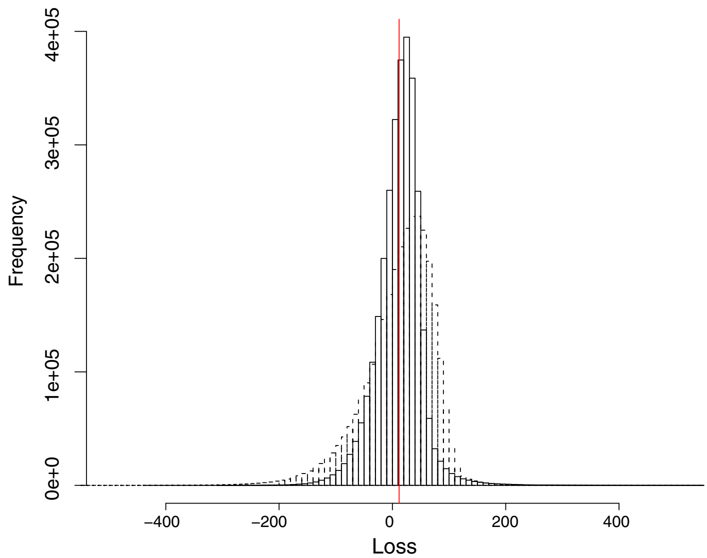

FIGURE 4.2. Histogram of one-step CVaR-hedged loss at level $\alpha = 95\%$ (normal lines) and $\alpha = 99\%$ (dashed lines). The vertical line is the mean of the portfolio loss distribution.

4.1.2. Consumption Hedging. At time $T = 1$ (year), an energy provider buys on an energy market a quantity C_T of gas at price S_T^g and sells it to consumers at a fixed price $K = 11$€/MWh. The quantity C_T denotes the consumption at time T and is equal to $C_T = a - bT_T$, with $a = 100$ Mwh and $b = 3$ MWh/°C. The temperature is modeled as an Ornstein–Uhlenbeck process so that the temperature at time T is given by

$$T_T = e^{-\lambda T} T_0 + m(1 - e^{-\lambda T}) + \sigma_T \sqrt{\frac{1 - e^{-2\lambda T}}{2\lambda}} G_1,$$

with $T_0 = 11$ °C, $\lambda = 0.02$, $m = 11$ °C, $\sigma_T = 6$ °C, and $G_1 \sim \mathcal{N}(0, 1)$. Gas spot price is modeled as a geometric Brownian motion with $S_0 = 11$€/MWh and the Brownian motion of gas spot price is correlated with the one of the temperature, $\rho = -0.8$, namely

$$S_T = S_0 e^{-\frac{\sigma_g^2}{2} T + \sigma_g \sqrt{T}\left(\rho G_1 + \sqrt{1-\rho^2} G_2\right)},$$

where $\sigma_g = 0.4$, $G_2 \sim \mathcal{N}(0, 1)$, and is independent of G_1. Consequently, the loss suffered by the energy provider at time T is given by

$$L = (S_T - K)C_T.$$

The energy provider uses a self-financed static strategy based on the gas spot price in order to reduce its risk at time $t_0 = 0$. A crude Monte Carlo gives $\mathbb{E}[L] = 62.6$ with a variance of 10,747.4 after 3,000,000 trials.

In this example, IS algorithms and the LCV method based on X do not achieve any significative variance reduction. Consequently, for the hedged portfolio, we do not provide any variance reduction ratio. However, we note that in order to estimate both

TABLE 4.2
Self-Financed Static Hedging of Consumption

	No hedging		Static hedging		
α	VaR	CVaR	VaR	θ^*_α	CVaR
95%	784.6	1,226.3	259.6	81.6	366.5
99%	1,452.4	2,012.3	437.1	89.9	537.3
99.5%	1,769.9	2,382.8	505.7	92.3	608.6

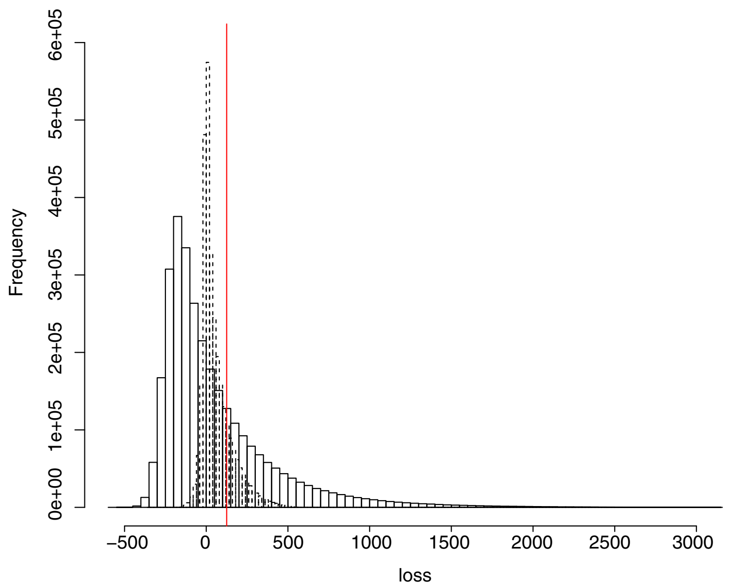

FIGURE 4.3. Histogram of loss with (dashed lines) and without (normal lines) one step CVaR-hedging at level $\alpha = 95\%$. The vertical line is the mean of the portfolio loss distribution.

VaR and CVaR of the loss L without hedging portfolio, the IS algorithm and the LCV method provide significant variance reduction. These results are due to the fact that, in this example, CVaR hedging already appears as a way to (optimally) reduce the variance of the loss. Consequently, reducing again the different variances by IS does not provide any further variance reduction whereas in the first example, CVaR hedging did not reduce the variance of the original loss but tries to capture some gains in order to reduce the global CVaR so that IS and LCV succeed in reducing the considered variances. Results are summarized in Table 4.2.

To complete this numerical example, we provide the histograms of the loss obtained with and without CVaR hedging using 3,000,000 samples. We can see on Figure 4.3 that the right tail distribution (which corresponds to high loss) is greatly reduced. The deformation provided by a CVaR hedging at 95% level is very impressive. The mode of the hedged loss distribution has been translated to the right near 0 whereas without hedging it was negative, which means that the most frequently occurring loss has changed

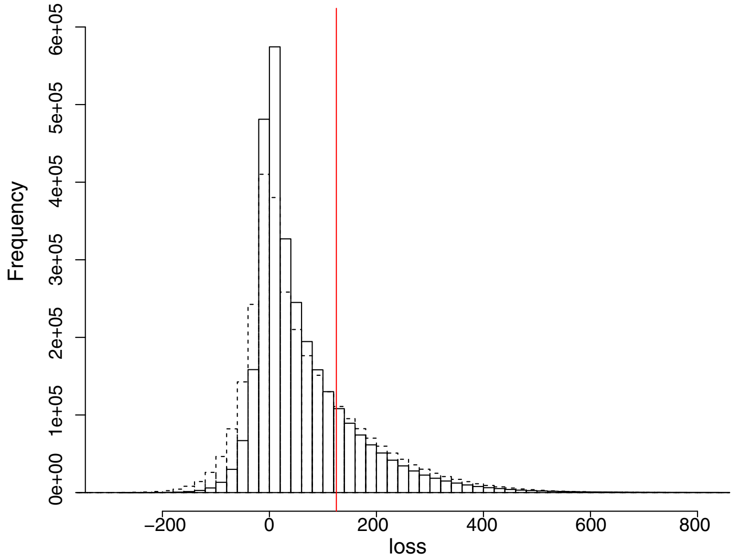

FIGURE 4.4. Histogram of one-step CVaR-hedged loss at level $\alpha = 95\%$ (normal lines) and $\alpha = 99\%$ (dashed lines).

from negative (gain) to positive value (loss). In order to reduce the right heavy tail, which corresponds to high loss, the CVaR hedging strategy translates the mode near the mean and thus gives more probability to small losses. Figure 4.4 illustrates the histograms obtained with a CVaR hedging at 95% and 99% levels. We remark that the distribution which corresponds to a CVaR hedging at 99% level has heavier tails than the one corresponding to a CVaR hedging at 95% level. The more α is close to 1, the heavier CVaR-hedged loss distribution tails are. Note that the mode of the distribution slightly translated to the left.

4.2. Dynamic Setting

We keep on studying the consumption hedging example and now, we experiment with our four different algorithms to compute the optimal self-financed dynamic strategy: C.H., B.H., M.D.H., and C.D.H. (see Section 3 for more details about each strategy and the RM algorithm associated). The parameters of the last example remain unchanged.

We consider three different values for the number of trading dates: $M = 4$ (one trade each trimester), $M = 12$ (one trade each month), $M = 52$ (one trade each week), and the CVaR-hedging level is 95%. All layers in the quantization tree of the process $(X_\ell, Z_\ell)_{1\leq\ell\leq M}$ have the same size, i.e., $N = N_\ell = 10$, $\ell = 1, \ldots, M$. Note that we do not quantify the process $(S_\ell, T_\ell)_{1\leq\ell\leq M-1}$ but only the two Gaussian random variables $(G1, G2)$, so that our quantization trees are obtained as a transform of the two-dimensional normal distribution optimal grid. It is crucial to have a good approximate of the random variable $\tilde{\Delta} L_\ell$, $\ell = 1, \ldots, M$, for the method M.D.H. so that we use an optimized quantization grid of size 100 in (3.26). Results are summarized in Table 4.3.

We clearly see that the optimal strategy is given by the M.D.H. method when the number of trading dates becomes large. The C.H. method for $M \leq 12$ is optimal but

TABLE 4.3
Self-Financed Dynamic CVaR Hedging of Consumption at Level 95% with Different Strategies

method	C.H.		B.H.		M.D.H.		C.D.H.	
M	VaR	CVaR	VaR	CVaR	VaR	CVaR	VaR	CVaR
4	178.3	240.9	175.9	252.5	177.8	252.9	178.9	259.2
12	163.2	214.1	160.7	233.8	158.7	221.7	161.9	232.9
52	272.6	395.1	158	233.2	148.7	210.1	153.1	223.7

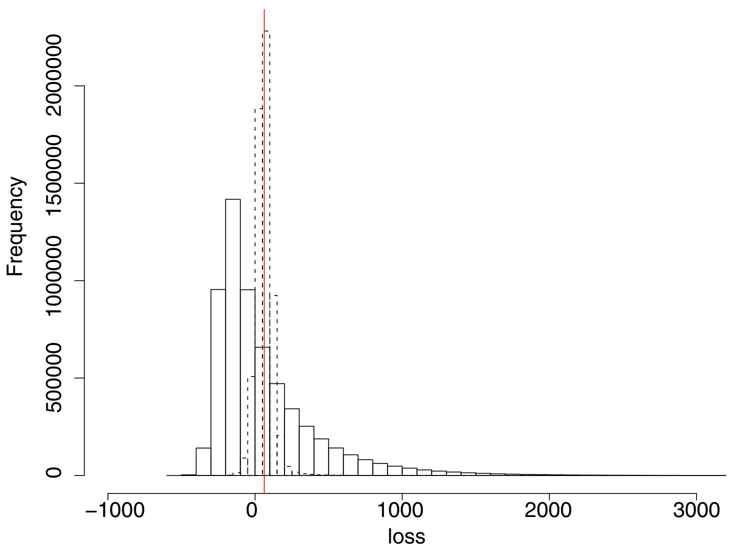

FIGURE 4.5. Histogram of consumption's loss with (dashed lines) and without (normal lines) dynamic CVaR-hedging at level $\alpha = 95\%$ using the M.D.H. strategy (52 trading dates).

suffers from convergence when $M \geq 12$. When M is large enough, the dimension of the algorithm in the C.H. method becomes too high and the estimate of the optimal strategy does not converge anymore. The larger is the number of trading dates, the greater is the difference between the M.D.H. and C.D.H. methods. Figure 4.5 presents the histograms of the loss without any hedging strategy and with a CVaR-hedging at 95% level using the M.D.H. method with 52 trading dates. The deformation of the loss distribution is very impressive. Like in the static framework, the mode of the CVaR-hedged loss distribution has been translated near the mean and in order to reduce the right tail distribution, the CVaR hedging strategy makes middle loss more likely. Figure 4.6 compares the CVaR-hedged loss distribution at 95% level using the static strategy and the dynamic strategy M.D.H. with 52 dates. The dynamic strategy translates the mode on the mean and removes losses under the mean to reduce the right tail distribution. Note that the very left tail of the two distributions (which corresponds to gains) are quite similar: the dynamic strategy reduces greatly high losses but only slightly high gains.

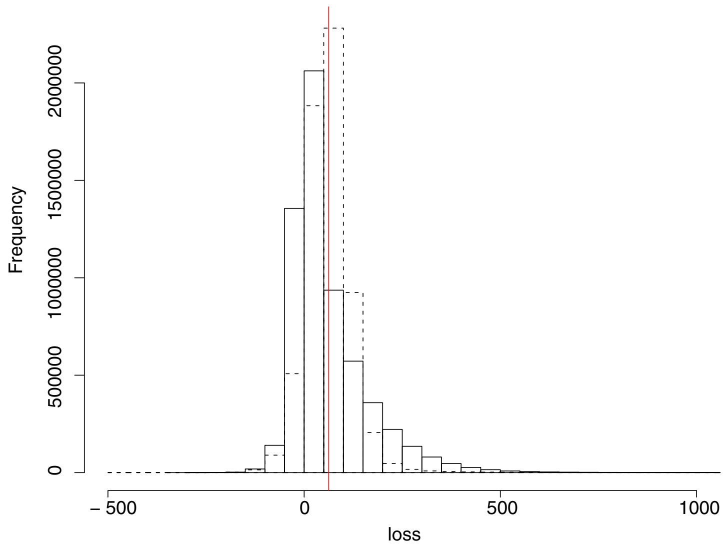

FIGURE 4.6. Histogram of CVaR-hedged loss at level $\alpha = 95\%$ with one-step static (normal lines) and dynamic (dashed lines, M.D.H. with 52 trading dates) self-financed strategies.

APPENDIX

A.1 Proof of Proposition 2.7

We propose below the proof of (2.18) and (2.19), which are the key results in order to derive our RM algorithm. The proof of (2.16) and (2.17) will follow using similar arguments.

Proof. It is clear that we always have

$$\inf_{\theta\Theta^d_{\ell_0}} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{–CVaR}_\alpha\left(L-\theta.X\right)\right] \geq \mathbb{E}\left[\operatorname*{ess\,inf}_{(\xi,\theta)\in\Theta^{d+1}_{\ell_0}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L-\theta.X-\xi\right)_+ |\mathcal{G}_{\ell_0}\right]\right].$$

Let $(\theta_n)_{n\geq 1}$ be a sequence in $\Theta^d_{\ell_0}$, such that

$$\operatorname*{ess\,inf}_{(\xi,\theta)\in\Theta^{d+1}_{\ell_0}} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L-\theta.X-\xi\right)_+ |\mathcal{G}_{\ell_0}\right] = \inf_{n\geq 1} \mathcal{G}_{\ell_0}\text{-CVaR}_\alpha\left(L-\theta_n.X\right),$$

and consider the sequence $(\Xi_n)_{n\geq 1}$ with $\Xi_1 = \theta_0 := 0$, and defined recursively for $n \geq 1$ by

$$\Xi_{n+1} := \begin{cases} \Xi_n & \text{, if } \mathcal{G}_{\ell_0}\text{–CVaR}_\alpha\left(L-\Xi^n.X\right) \leq \mathcal{G}_{\ell_0}\text{–CVaR}_\alpha\left(L-\theta_n.X\right), \\ \theta_n & \text{, if } \mathcal{G}_{\ell_0}\text{–CVaR}_\alpha\left(L-\Xi^n.X\right) \geq \mathcal{G}_{\ell_0}\text{–CVaR}_\alpha\left(L-\theta_n.X\right). \end{cases}$$

Note that $\Xi_n \in \Theta^d_{\ell_0}$ for $n \geq 1$ and

$$\mathcal{G}_{\ell_0}\text{–CVaR}_\alpha\left(L-\Xi_{n+1}.X\right) = \min_{0\leq p\leq n} \mathcal{G}_{\ell_0}\text{–CVaR}_\alpha\left(L-\theta_p.X\right) \quad a.s.,$$

so that the sequence $\left(\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\Xi_n.X)\right)_{n\geq 0}$ is nonincreasing. Using the Bayes' rule, Jensen's inequality, and the risk-neutral equivalent probability measure $\mathbb{P}^*$ (keeping in mind that $\frac{d\mathbb{P}^*}{d\mathbb{P}} < K$ $a.s.$ for some deterministic constant $K < +\infty$), we have for all $\xi \in \mathcal{G}_{\ell_0}$

$$
\begin{aligned}
&\xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\Xi_n.X-\xi)_+\big|\,\mathcal{G}_{\ell_0}\right]\\
&\quad = \mathbb{E}\left[\left(\xi + \frac{1}{1-\alpha}(L-\Xi_n.X-\xi)_+\right)\frac{d\mathbb{P}^*}{d\mathbb{P}}\frac{d\mathbb{P}}{d\mathbb{P}^*}\bigg|\,\mathcal{G}_{\ell_0}\right]\\
&\quad \geq \left(\xi + \frac{1}{1-\alpha}\mathbb{E}_{\mathbb{P}^*}\left[(L-\Xi_n.X-\xi)|\,\mathcal{G}_{\ell_0}\right]\right)\times\frac{1}{K}\times\mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}}\bigg|\,\mathcal{G}_{\ell_0}\right]\\
&\quad \geq \left(\xi + \frac{1}{1-\alpha}\left(\mathbb{E}_{\mathbb{P}^*}\left[L|\,\mathcal{G}_{\ell_0}\right]-\xi\right)_+\right)\times\frac{1}{K}\times\mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}}\bigg|\,\mathcal{G}_{\ell_0}\right]\\
&\quad \geq \frac{1}{1-\alpha}\mathbb{E}_{\mathbb{P}^*}\left[L|\,\mathcal{G}_{\ell_0}\right]\times\frac{1}{K}\times\mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}}\bigg|\,\mathcal{G}_{\ell_0}\right] = \frac{1}{1-\alpha}\frac{1}{K}\mathbb{E}\left[L\frac{d\mathbb{P}^*}{d\mathbb{P}}\bigg|\,\mathcal{G}_{\ell_0}\right],
\end{aligned}
$$

so that, $\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\Xi_n.X) \geq \frac{1}{1-\alpha}\frac{1}{K}\mathbb{E}\left[L\frac{d\mathbb{P}^*}{d\mathbb{P}}\,|\,\mathcal{G}_{\ell_0}\right] \in L^1(\mathbb{P})$.

Moreover, by definition for $n \geq 0$

$$
\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\Xi_n.X) \leq \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L) \leq \frac{1}{1-\alpha}\mathbb{E}\left[L_+|\,\mathcal{G}_{\ell_0}\right] \in L^1(\mathbb{P}).
$$

The sequence $(\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha(L-\Xi_n.X))_{n\geq 0}$ converges in $L^1(\mathbb{P})$ owing to Beppo–Levi's Theorem and

$$
\begin{aligned}
\inf_{\theta\in\Theta^d_{\ell_0}}\mathbb{E}\left[\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\theta.X)\right] &\geq \inf_{n\geq 0}\mathbb{E}\left[\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\Xi_n.X)\right]\\
&\geq \mathbb{E}\left[\inf_n \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\Xi_n.X)\right]\\
&= \mathbb{E}\left[ess\,inf_{\theta\in\Theta^d_{\ell_0}}\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\theta.X)\right].
\end{aligned}
$$

The proof of (2.16) follows from similar arguments.

For $\mathbb{P}$-almost all $\omega \in \Omega$, $V_f(\omega,\dots)$ is convex. Using the Bayes' rule and Jensen's inequality, we have for all $\theta \in \mathcal{G}_{\ell_0}$

$$
\begin{aligned}
\mathbb{E}\left[\left(\xi+\frac{1}{1-\alpha}(L-\theta.X-\xi)_+\right)\bigg|\,\mathcal{G}_{\ell_0}\right] &= \mathbb{E}\left[\left(\xi+\frac{1}{1-\alpha}(L-\theta.X-\xi)_+\right)\frac{d\mathbb{P}^*}{d\mathbb{P}}\frac{d\mathbb{P}}{d\mathbb{P}^*}\bigg|\,\mathcal{G}_{\ell_0}\right]\\
&\geq \frac{1}{K}\left(\xi+\frac{1}{1-\alpha}\left(\mathbb{E}_{\mathbb{P}^*}\left[L|\,\mathcal{G}_{\ell_0}\right]-\xi\right)_+\right)\mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}}\bigg|\,\mathcal{G}_{\ell_0}\right],
\end{aligned}
$$

and it is clear that $\lim_{\xi\to+\infty}\left(\xi + \frac{1}{1-\alpha}\left(\mathbb{E}_{\mathbb{P}^*}\left[L|\,\mathcal{G}_{\ell_0}\right](w)-\xi\right)_+\right) = +\infty$, which finally yields, for all $\theta \in \mathbb{R}^d$,

$$
\lim_{|\xi|\to+\infty} V_f(\omega,\xi,\theta) = +\infty.
$$

Now, in order to establish that the function $V_f(\omega,\xi,.)$ goes to infinity as $|\theta|$ goes to infinity for all $\xi \in \mathbb{R}$, we show that $\inf_{\xi\in\mathbb{R}} V_f(\omega,\xi,\theta) = \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\theta.X)(\omega)$

satisfies

$$\lim_{|\theta|\to+\infty} \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\theta.X)\,(\omega) = +\infty.$$

First, note that the subadditivity of the function $x \mapsto x_+$ implies that

$$\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(-\theta.X) \le \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\theta.X) + \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(-L),$$

so that,

$$|\theta|\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\left(-\frac{\theta}{|\theta|}.X\right) - \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(-L) \le \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\theta.X),$$

which finally yields

$$|\theta| \operatorname*{ess\,inf}_{u\in\Theta^d_{\ell_0},|u|=1} \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X) - \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(-L) \le \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(L-\theta.X).$$

For $\mathbb{P}$-almost all $\omega\in\Omega$, we have $\operatorname{ess\,inf}_{u\in\Theta^d_{\ell_0},|u|=1}\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X)\,(\omega) = \inf_{\xi\in\mathbb{R},u\in\mathcal{S}_d(0,1)} V_f(\omega,\xi,u)$, where $\mathcal{S}_d\,(0,1) := \left\{u\in\mathbb{R}^d \mid |u|=1\right\}$ denotes the (compact) unit sphere. Furthermore, since the function $v_f(\xi,.,y,x)$ is Lipschitz-continuous for all $\xi, y\in\mathbb{R},\ x\in\mathbb{R}^d$, it follows that for any $u,u'\in\mathcal{S}_d\,(0,1)$,

$$\begin{aligned}
&\left|\inf_{\xi\in\mathbb{R}}\int v_f\,(\xi,u,0,x)\,\Pi(dx,dy) - \inf_{\xi\in\mathbb{R}}\int v_f\left(\xi,u',0,x\right)\Pi(dx,dy)\right| \\
&\qquad\le \sup_{\xi\in\mathbb{R}}\left|\int \left(v_f(\xi,u,y,x) - v_f(\xi,u',y,x)\right)\Pi(dx,dy)\right| \\
&\qquad\le \frac{\left|u-u'\right|}{1-\alpha}\int |x|\Pi(dx,dy),\ a.s.
\end{aligned}$$

Consequently, for $\mathbb{P}$-almost all $\omega\in\Omega$, the function $u\mapsto \mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X)\,(\omega)$ is continuous on $\mathcal{S}_d\,(0,1)$. Thus, it remains to check that for all $u\in\mathcal{S}_d\,(0,1)$, $\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X)\,(\omega) > 0$. Proposition 2.1 implies that there exists $\xi^*_\alpha(w)$ such that $\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X)\,(\omega) = \xi^*_\alpha(w) + \frac{1}{1-\alpha}\mathbb{E}\left[(u.X-\xi^*_\alpha(w))_+\middle|\,\mathcal{G}_{\ell_0}\right](w)$. There are three cases to check:

- if $\xi^*_\alpha(w) > 0$, then it is straightforward that $\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X)\,(\omega) \ge \xi^*_\alpha(w) > 0$,
- if $\xi^*_\alpha(w) < 0$, then (NA), Bayes' rule, and Jensen's inequality (see the end of the proof of Proposition 2.7) lead to

$$\begin{aligned}
\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X)\,(\omega) &\ge \frac{1}{K}\left(\xi^*_\alpha(w) + \frac{1}{1-\alpha}\left(-\xi^*_\alpha(w)\right)_+\right)\mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}}\middle|\,\mathcal{G}_{\ell_0}\right](w) \\
&= -\frac{\alpha}{1-\alpha}\xi^*_\alpha(w)\frac{1}{K}\mathbb{E}\left[\frac{d\mathbb{P}^*}{d\mathbb{P}}\middle|\,\mathcal{G}_{\ell_0}\right](w) > 0,
\end{aligned}$$

- if $\xi^*_\alpha(w) = 0$, $\mathcal{G}_{\ell_0}\text{−CVaR}_\alpha\,(u.X)\,(\omega) = \frac{1}{1-\alpha}\mathbb{E}\left[(u.X)_+ \mid \mathcal{G}_{\ell_0}\right](\omega)$. Now, if $\mathbb{E}\left[(u.X)_+ \mid \mathcal{G}_{\ell_0}\right] = 0$, then $(u.X)_+ = 0$ $a.s.$, i.e., $(-u).X \ge 0$. Using (NA), we clearly get $u.X = 0$ $a.s.$ so that it implies that $u=0$, which is impossible.

The proof of (2.17) follows using similar arguments.

Consequently, there exists $(\xi_\alpha^*, \theta_{1,\alpha}^*) := (\xi_\alpha^*(\omega), \theta_\alpha^*(\omega))$ and for all $\xi \in \mathbb{R}$, $\theta_{2,\alpha}^* := \theta_\alpha^*(\omega, \xi)$, which are $\mathcal{G}_{\ell_0}$-measurable owing to measurable selection theorem (see, e.g., Lemmas 3 and 4 in Evstigneev 1976), such that

$$\inf_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi,\theta) = \min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi,\theta) = V_f(\xi_\alpha^*, \theta_{1,\alpha}^*) \quad \text{and}$$
$$\inf_{\theta\in\mathbb{R}^d} V_f(\xi,\theta) = \min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta) = V_f(\xi, \theta_{2,\alpha}^*) \ \ a.s. \qquad \square$$

A.2 Classical Results on Stochastic Approximation Algorithms

In this section, we provide two standard results related to $a.s.$ convergence and weak rate of convergence of stochastic approximation algorithms. For detailed proofs and further developments, we refer to classical textbooks on stochastic approximation like Duflo (1996), Duflo (1997), and Kushner and Yin (2003).

THEOREM A.1. *(Robbins–Monro Theorem) Let $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function and let X be an $\mathbb{R}^d$-valued random vector such that $\mathbb{E}[|H(y, X)|] < +\infty$ for every $y \in \mathbb{R}^d$. Then set*

$$\forall y \in \mathbb{R}^d, \quad h(y) = \mathbb{E}[H(y, X)].$$

Suppose that the function h is continuous and that $\mathcal{T}^* := \{h = 0\}$ satisfies

$$\forall y \in \mathbb{R}^d \setminus \mathcal{T}^*, \forall y^* \in \mathcal{T}^*, \quad \langle y - y^*, h(y)\rangle > 0. \tag{A.1}$$

Let $(\gamma_n)_{n\geq 1}$ be a deterministic step sequence satisfying (3.8). Suppose that

$$\forall y \in \mathbb{R}^d, \ \ \mathbb{E}[|H(y, X)|^2] \leq C(1 + |y|^2) \tag{A.2}$$

(which implies that $|h(y)|^2 \leq C(1 + |y|)$).

Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence of random vectors having the distribution of X, let y_0 be a random vector independent of $(X_n)_{n\geq 1}$ satisfying $\mathbb{E}|y_0|^2 < +\infty$, all defined on the same probability space $(\Omega, A, \mathbb{P})$.

Then, the recursive procedure defined for $n \geq 1$ by

$$y_n = y_{n-1} - \gamma_n H(y_{n-1}, X_n),$$

satisfies:

$$\exists\, y_\infty : (\Omega, \mathcal{A}) \to \mathcal{T}^*, y_\infty \in \mathbb{L}^2(\mathbb{P}) \quad \text{such that } \ y_n \xrightarrow{a.s.} y_\infty.$$

The convergence also holds in $L^p(\mathbb{P})$, $p \in (0, 2)$.

THEOREM A.2. *(Ruppert and Polyak's Averaging Principle) Suppose that the $\mathbb{R}^d$-sequence $(\phi_n)_{n\geq 0}$ is defined recursively by*

$$\phi_n = \phi_{n-1} - \gamma_n \left(h(\phi_{n-1}) + \epsilon_n\right),$$

where h is a Borel function. Let $\mathcal{F}_n := \sigma(\xi_0, \theta_0, U_1, \ldots, U_n)$ be the natural filtration of the algorithm. Suppose that h is $\mathcal{C}^1$ in the neighborhood of a zero ϕ^* of h and that $P = Dh(\phi^*)$

is a uniformly repulsive matrix (all its eigenvalues have positive real parts), and that $(\epsilon_n)_{n\geq 1}$ *is a random* $\mathcal{F}_n$*-adapted sequence satisfying*

$$\exists C > 0, \text{ such that a.s.}$$

$$\text{(A.3)} \qquad \begin{cases} (i)\ \mathbb{E}[\epsilon_{n+1}|\mathcal{F}_n]\mathbf{1}_{\{|\phi_n-\phi^*|\leq C\}} = 0, \\ (ii)\ \exists\, b > 2,\ \ \sup_n \mathbb{E}[|\epsilon_{n+1}|^b|\mathcal{F}_n]\mathbf{1}_{\{|\phi_n-\phi^*|\leq C\}} < +\infty, \\ (iii)\ \exists\, \Gamma \in \mathcal{S}^+(d,\mathbb{R}) \textit{ such that } \mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right] \xrightarrow{a.s.} \Gamma. \end{cases}$$

Set $\gamma_n = \frac{\gamma_1}{n^\beta}$ *with* $\frac{1}{2} < \beta < 1$*, and*

$$\overline{\phi}_n := \frac{\phi_0 + \cdots + \phi_{n-1}}{n} = \overline{\phi}_{n-1} - \frac{1}{n}\left(\overline{\phi}_{n-1} - \phi_{n-1}\right),\ n \geq 1.$$

Then, on the set of convergence $\{\phi_n \to \phi^*\}$*:*

$$\sqrt{n}\left(\overline{\phi}_n - \phi^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\ P^{-1}\Gamma(P^{-1})^T\right) \qquad \textit{as } n \to +\infty.$$

A.3 Proof of Theorem 3.2

We first prove the $a.s.$ convergence of $(\xi_n, \theta_n)_{n\geq 1}$ using the so-called RM Theorem; that of $(C_n)_{n\geq 1}$ will follow the lines of the proof of the $a.s.$ convergence of the CVaR algorithm in Bardou et al. (2009) (see Section 2.2). In order to apply the RM theorem, we have to check the following facts:

- *Mean reversion*: For the sake of simplicity, we denote by y the couple (ξ, θ). The mean function of the algorithm defined by (3.6) and (3.7) reads

$$l(y) := \mathbb{E}\left[(H_1(y, U), H_{2:d+1}(y, U))\right] = \nabla V(y),$$

  so that $\mathcal{T}^* := \{l = 0\} = \{\nabla V = 0\}$. Moreover, if $y^* \in \mathcal{T}^*$ and $y \in \mathbb{R} \times \mathbb{R}^d \mathcal{T}^*$,

$$\langle y - y^*, l(y)\rangle = \langle y - y^*, \nabla V(y)\rangle > 0,$$

  since the function V is a convex differentiable function and $\operatorname{Arg\,min} V$ is nonempty.
- *Linear Growth of* $(\xi, \theta) \mapsto \mathbb{E}\left[|H_1(\xi, \theta, U)|^2 + |H_{2:d+1}(\xi, \theta, U)|^2\right]^{1/2}$: This condition is clearly fulfilled since there exists a real constant $C > 0$ such that

$$\mathbb{E}\left[|H_1(\xi, \theta, U)|^2\right] < C \text{ and } \mathbb{E}\left[|H_{2:d+1}(\xi, \theta, U)|^2\right] < \frac{1}{(1-\alpha)^2}\mathbb{E}\left[|G(U))|^2\right] < C,$$

  so that

$$\mathbb{E}\left[|H_1(\xi, \theta, U)|^2 + |H_{2:d+1}(\xi, \theta, U)|^2\right] \leq C\left(1 + |y|^2\right).$$

Consequently, we have

$$(\xi_n, \theta_n) \xrightarrow{a.s.} \left(\xi_\alpha^*, \theta_\alpha^*\right),\ n \to +\infty.$$

The $a.s.$ convergence of $(C_n)_{n\geq 1}$ toward CV_α^* follows by similar arguments as those used for the $a.s.$ convergence of the CVaR procedure in Bardou et al. (2009), section 2.2.

A.4 Proof of Theorem 3.5

First, note that the procedure (3.10) can be written

$$\forall\, n \geq 1,\quad \phi_n = \phi_{n-1} - \gamma_n\left(h(\phi_{n-1}) + \epsilon_n\right),\quad \phi_0 = (\xi_0, \theta_0, C_0),$$

where $h(\phi) := \mathbb{E}[H(\phi, U)] = \nabla V(\phi, x, z)$, and ϵ_n, $n \geq 1$, denotes the martingale increment sequence defined by:

$$\epsilon_{1,n} := \frac{1}{1-\alpha}\left(\mathbb{P}\left(L - \theta.X \geq \xi\right)_{|\xi=\xi_{n-1},\theta=\theta_{n-1}} - 1_{\{L_n - \theta_{n-1}.X_n \geq \xi_{n-1}\}}\right),$$

$$\epsilon_{i,n} := \frac{1}{1-\alpha}\left(\mathbb{E}\left[X_{i-1}1_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_{n-1},\theta=\theta_{n-1}} - X_{i-1,n}1_{\{L_n-\theta_{n-1}.X_n\geq\xi_{n-1}\}}\right),\ i=2,\ldots,d+1,$$

$$\epsilon_{d+2,n} := \Delta N_n = \frac{1}{1-\alpha}\left(\mathbb{E}\left[(L-\theta.X-\xi)_+\right]_{|\xi=\xi_{n-1},\theta=\theta_{n-1}} - (L_n - \theta_{n-1}.X_n - \xi_{n-1})_+\right),$$

where $L_n = F(U_n)$ and $X_n := G(U_n)$. Since the function V is convex, its Hessian matrix P is positive as soon as h is differentiable. Now, in order to differentiate h, we write

$$h_1(\phi) = 1 - \frac{1}{1-\alpha}\int_{\mathbb{R}^d\times\mathbb{R}} p_{X,L}(x,y)1_{\{y\geq\xi+\theta.x\}}dxdy,$$

$$h_i(\phi) = -\frac{1}{1-\alpha}\int_{\mathbb{R}^d\times\mathbb{R}} x_i\, p_{X,L}(x,y)1_{\{y\geq\xi+\theta.x\}}dxdy,\quad i = 2,\ldots,d+1,$$

$$h_{d+2}(\phi) = C - \left(\xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\theta.X-\xi)_+\right]\right).$$

In order to differentiate h_1, note that by Fubini's Theorem,

$$h_1(\phi) = 1 - \frac{1}{1-\alpha}\int_{\mathbb{R}^d} dx \int_{\xi+\theta.x}^{+\infty} p_{X,L}(x,y)dy.$$

Owing to Assumption 3.3, one can interchange integral and derivation. In order to differentiate $h_{2:d+1}$, first, note that by Fubini's Theorem,

$$h_{2:d+1}(\phi) = -\frac{1}{1-\alpha}\int_{\mathbb{R}^d} dx \int_{\xi+\theta.x}^{+\infty} x p_{X,L}(x,y)dy,$$

so that owing to Assumption 3.3 and Lebesgue's differentiation theorem, one can interchange integral and derivation. Consequently, the functions h_1 and $h_{2:d+1}$ are differentiable at $\phi^* := \left(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*\right)$ and for $i = 2,\ldots,d+1$,

$$\frac{\partial h_1}{\partial\xi}(\phi^*) = \frac{1}{1-\alpha}\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx,\quad \frac{\partial h_1}{\partial\theta_{i-1}}(\phi^*) = \frac{\partial h_i}{\partial\xi}(\phi^*)$$
$$= \frac{1}{1-\alpha}\int_{\mathbb{R}^d} x_i\, p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx,$$

$$\frac{\partial h_1}{\partial C}(\phi^*) = \frac{\partial h_{d+2}}{\partial \xi}(\phi^*) = 0, \quad \frac{\partial h_i}{\partial \theta_j}(\phi^*) = \frac{\partial h_j}{\partial \theta_i}(\phi^*) = \frac{1}{1-\alpha}\int_{\mathbb{R}^d} x_i x_j p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx,$$

$$\frac{\partial h_i}{\partial C}(\phi^*) = \frac{\partial h_{d+2}}{\partial \theta_i}(\phi^*) = 0, \quad \frac{\partial h_{d+2}}{\partial C}(\phi^*) = 1,$$

so that M is given by (3.12). Let $u = (u_1, u_2, u3) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$,

$$u^T M u = \frac{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx}{1-\alpha}\left(u_1^2 + 2u_1 \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx} dx u_2 \right.$$
$$\left. + u_2^T \frac{\int_{\mathbb{R}^d} x x^T p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx} u_2 + u_3^2 \right),$$

using the inequality $2u_1 \frac{\int_{\mathbb{R}^d} x p_{X,L}(x,\xi_\alpha^*+\theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi_\alpha^*+\theta_\alpha^*.x)dx} dx u_2 \geq -u_1^2 - u_2^T \mathbb{E}_{\mathbb{Q}}[X] \mathbb{E}_{\mathbb{Q}}[X]^T u_2$, we obtain

$$u^T M u \geq \frac{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx}{1-\alpha}\left(u_2^T \int_{\mathbb{R}^d} \left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx} \right) \right.$$
$$\left. \times \left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx} \right)^T p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)dx u_2 + u_3^2 \right) > 0.$$

Consequently, the matrix P is uniformly repulsive. To apply Theorem A.2, we need to check assumption (A.3). Let $C > 0$. First, note that

$$\mathbb{E}\left[|\epsilon_{1,n+1}|^{2a} | \mathcal{F}_n\right] 1_{\{|\phi_n - \phi^*| \leq C\}} \leq \left(\frac{1}{1-\alpha}\right)^{2a} 2^{2a} < +\infty.$$

Thanks to Assumption 3.1 (with $a > 1$), there exists $A > 0$, such that for $i = 2, \ldots, d+1$,

$$\mathbb{E}\left[|\epsilon_{i,n+1}|^{2a} | \mathcal{F}_n\right] 1_{\{|\phi_n - \phi^*| \leq C\}} \leq A\mathbb{E}\left[X_{i-1}^{2a}\right] < +\infty,$$

and

$$\mathbb{E}\left[|\epsilon_{d+2,n+1}|^{2a} | \mathcal{F}_n\right] 1_{\{|\phi_n - \phi^*| \leq C\}} \leq A\left(\mathbb{E}\left[|L|^{2a}\right] + \mathbb{E}\left[|X|^{2a}\right]\right) < +\infty.$$

Consequently, (ii) of (A.3) holds true with $b = 2a$ since

$$\sup_{n \geq 0} \mathbb{E}\left[|\epsilon_{n+1}|^{2a} | \mathcal{F}_n\right] 1_{\{|\phi_n - \phi^*| \leq C\}} < +\infty.$$

It remains to check (iii) for some positive definite symmetric matrix Γ.

The continuity of the functions $(\xi,\theta)\mapsto \mathbb{E}\left[X_{i-1}X_{j-1}1_{\{L-\theta.X\geq\xi\}}\right]$ and $(\xi,\theta)\mapsto \mathbb{E}\left[X_{i-1}1_{\{L-\theta.X\geq\xi\}}\right]$ at $(\xi_\alpha^*,\theta_\alpha^*)$, which follows from the continuity of the joint distribution (L, X), combined with the equality $\mathbb{E}\left[X_{i-1}1_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]=0$, $i=2,\dots,d+1$, imply that

$$
\begin{aligned}
\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{i,j}|\mathcal{F}_n\right] = \mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{j,i}|\mathcal{F}_n\right]
&= \frac{1}{(1-\alpha)^2}\Big(\mathbb{E}\left[X_{i-1}X_{j-1}1_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_n,\theta=\theta_n} \\
&\quad - \mathbb{E}\left[X_{i-1}1_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_n,\theta=\theta_n}\mathbb{E}\left[X_{j-1}1_{\{L-\theta.X\geq\xi\}}\right]_{|\xi=\xi_n,\theta=\theta_n}\Big), \\
&\xrightarrow{a.s.} \frac{1}{(1-\alpha)^2}\mathbb{E}\left[X_{i-1}X_{j-1}1_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right].
\end{aligned}
$$

Using similar arguments, one shows that $\mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right]\xrightarrow{a.s.}\Gamma$. This completes the proof.

REFERENCES

ARTZNER, P., F. DELBAEN, J.-M. EBER, and D. HEATH (1999): Coherent Measures of Risk, *Math. Finance* 9(3), 203–228.

AVELLANEDA, M. (1998): The Minimum-Entropy Algorithm and Related Methods for Calibrating Asset-Pricing Models, in *Proceedings of the International Congress of Mathematicians, Documenta Mathematica*, Vol. III. Berlin: pp. 545–563 (electronic).

BARDOU, O., N. FRIKHA, and G. PAGÈS (2009): Computing VaR and CVaR Using Stochastic Approximation and Adaptive Unconstrained Importance Sampling, *Monte Carlo Methods Appl.* 15(3), 173–210.

BARRIEU, P., and N. EL KAROUI (2004): Optimal Derivatives Design under Dynamic Risk Measures, in *Mathematics of Finance*, Vol. 351 of Contemporary Mathematics, Providence, RI: American Mathematical Society, pp. 13–25.

BENTH, F. E., J. ŠALTYTĖ BENTH, and S. KOEKEBAKKER (2008): Stochastic Modelling of Electricity and Related Markets. *Advanced Series on Statistical Science & Applied Probability*, 11, Hackensack, NJ: World Scientific Publishing Co. Pte. Ltd.

BENVENISTE, A., M. MÉTIVIER, and P. PRIOURET (1990): *Adaptive Algorithms and Stochastic Approximations*, Vol. 22 of Applications of Mathematics (New York), Berlin: Springer-Verlag. Translated from the French by Stephen S. Wilson.

BODA, K., and J. A. FILAR (2006): Time Consistent Dynamic Risk Measures, *Math. Methods Oper. Res.* 63(1), 169–186.

DUFLO, M. (1996): *Algorithmes Stochastiques*, Vol. 23 of Mathématiques & Applications (Berlin) [Mathematics & Applications], Berlin: Springer-Verlag.

DUFLO, M. (1997): *Random Iterative Models*, Vol. 34 of Applications of Mathematics (New York), Berlin: Springer-Verlag. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

EVSTIGNEEV, I. V. (1976): Measurable Selection and Dynamic Programming, *Math. Oper. Res.* 1(3), 267–272.

FÖLLMER, H., and P. LEUKERT (1999): Quantile Hedging, *Finance Stoch.* 3(3), 251–273.

FÖLLMER, H., and I. PENNER (2006): Convex Risk Measures and the Dynamics of Their Penalty Functions, *Statist. Decisions* 24(1), 61–96.

FÖLLMER, H., and A. SCHIED (2002): Convex Measures of Risk and Trading Constraints, *Finance Stoch.* 6(4), 429–447.

FÖLLMER, H., and M. SCHWEIZER (1991): Hedging of Contingent Claims under Incomplete Information, in *Applied Stochastic Analysis (London, 1989)*, M. H. A Davis and R. J. Elliott, eds. Vol. 5 of Stochastics Monographs, New York: Gordon and Breach, pp. 389–414.

FÖLLMER, H., and D. SONDERMAN (1986): Hedging of Non-Redundant Contingent Claims, in *Contributions to Mathematical Economics*, W. Hildenbrand and A. Mas-Colell, eds. Amsterdam: North Holland, pp. 205–223.

FRIKHA, N. (2010): *Contribution à la modélisation et à la gestion dynamique du risque sur les marchés de l'énergie*. PhD thesis, Université Pierre & Marie Curie.

FRITTELLI, M. (2000): The Minimal Entropy Martingale Measure and the Valuation Problem in Incomplete Markets, *Math. Finance* 10(1), 39–52.

GEMAN, H. (2005): *Commodities and Commodity Derivatives: Pricing and Modeling Agricultural, Metals and Energy*, Chichester, UK: Wiley Finance.

GEMAN, H., and S. OHANA (2008): Time-Consistency in Managing a Commodity Portfolio: A Dynamic Risk Measure Approach, *J. Banking Finance* 32, 1991–2005.

HODGES, S. D., and A. NEUBERGER (1989): Optimal Replication of Contingent Claims under Transaction Costs, *Rev. Futures Markets* 8, 222–239.

JACOD, J., and A. N. SHIRYAEV (1998): Local Martingales and the Fundamental Asset Pricing Theorems in the Discrete-Time Case, *Finance Stoch.* 2(3), 259–273.

KARATZAS, I., J. P. LEHOCZKY, S. E. SHREVE, and G.-L. XU (1991): Martingale and Duality Methods for Utility Maximization in an Incomplete Market, *SIAM J. Control Optim.* 29(3), 702–730.

KUSHNER, H. J., and G. G. YIN (2003): *Stochastic Approximation and Recursive Algorithms and Applications*, Vol. 35 of Applications of Mathematics (New York), 2nd ed., New York: Springer-Verlag.

LEMAIRE, V., and G. PAGÈS (2010): Unconstrained Recursive Importance Sampling, *Ann. Appl. Probab.* 20(3), 1029–1067.

MIYAHARA, Y. (2006): [GLP & MEMM] Pricing Models and Related Problems, in *Stochastic Processes and Applications to Mathematical Finance*, J. Akahori, S. Ogawa, and S. Watanabe, eds. Hackensack, NJ: World Scientific Publications, pp. 125–156.

PAGÈS, G. (1998): A Space Quantization Method for Numerical Integration, *J. Comput. Appl. Math.* 89(1), 1–38.

PELLETIER, M. (2000): Asymptotic Almost Sure Efficiency of Averaged Stochastic Algorithms, *SIAM J. Control Optim.* 39(1), 49–72 (electronic).

ROCKAFELLAR, R. T., and S. URYASEV (2000): Optimization of Conditional Value-at-Risk, *J. Risk* 2, 493–517.

ROUGE, R., and N. EL KAROUI (2000): Pricing via Utility Maximization and Entropy, *Math. Finance* 10(2), 259–276. *INFORMS Applied Probability Conference* (Ulm, 1999).

SCHWEIZER, M. (1991): Option Hedging for Semimartingales. *Stochastic Process. Appl.* 37(2), 339–363.

SHAPIRO, A. (2009): On a Time Consistency Concept in Risk Averse Multistage Stochastic Programming. *Oper. Res. Lett.* 37(3), 143–147.

STREET, A., B. RUDLOFF, and D. VALLADAO (2011): On the Economic Interpretation of Time Consistent Risk Averse Dynamic Stochastic Programming Problems.