CVaR hedging using quantization based stochastic approximation algorithm

Olivier Bardou * Noufel Frikha † Gilles Pages ‡

January 27, 2011

Abstract

In this paper, we investigate a method based on risk minimization to hedge observable but non-tradable 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: stochastic approximation algorithm, optimal quantization and variance reduction techniques (importance sampling (IS) and linear control variable (LCV)) 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 time T. We devise a stochastic approximation algorithm and study its a.s. convergence and rate of convergence. Then, we extend to the dynamic case under the assumption that the process modelling the non-tradable source of risk and financial assets prices are Markov. Finally, we illustrate our approach by considering several portfolios in the incomplete energy market.

Keywords: *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 would like to have a more realistic view of financial or energy markets which are intrinsically incomplete for many reasons (stochastic volatility, jumps, temperature dependance of prices on energy markets, ...). 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 super-hedging criterion as studied in [8]. However, the price is often too high, actually, the trader can only hedge partially and often has to bear some risk of loss. Many authors studied pricing theory under a martingale measure which corresponds to an optimized criterion. For instance, one can refer to [13] for the minimal martingale measure, to [2], [16] and [21] for the minimal entropy martingale measure among others.

Another method widely used to adress this problem is based on expected utility maximization. Indeed, there is a huge litterature on hedging and pricing in incomplete markets using expected utility maximization method and utility indifference pricing. It consists in pricing an unhedgable

*GDFSuez Finance Division, Risk Department, 22 rue du Docteur Lancereaux, F-75008 Paris Cedex 8, E-mail: Olivier.Bardou@gdfsuez.com

†Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Pierre & Marie Curie, 4 pl. Jussieu, F-75252 Paris Cedex 5, France, E-mail: frikha.noufel@gmail.com

‡Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Pierre & Marie Curie, 4 pl. Jussieu, F-75252 Paris Cedex 5, France, E-mail: gilles.pages@upmc.fr,

claim so that the investor's utility remains unchanged between holding and not holding the contingent claim. We refer to [17], [18] and [9] among many others for some developments. Although, this approach has been studied for long, the main drawback for a pratician remains the lack of knowledge of his own utility function for hedging and pricing derivatives. Moreover, different agents may price and hedge a contingent claim differently according to their own risk preference so that it has little acceptance in practice.

In this article, we propose an alternative method based on risk minimization using stochastic approximation algorithm. To be more precise, we focus on minimizing dynamically the Conditional Value-at-Risk (CVaR). The CVaR is strongly linked to the famous risk measure called 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 Conditional Value-at-Risk 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 [1] and have been shown to share basic coherence properties (which is not the case of VaR_α. The extension to convex risk measures were introduced and extensively studied in [12].

Pricing and hedging using risk measures is a recent approach which has been investigated by many authors. Barrieu and El Karoui in [4] developped a risk minimization problem to hedge non-tradable risk on financial market using convex risk measures. Hedging strategy which maximizes the probability of successfull hedge is studied in [11] as an alternative to super-hedging strategy which requires a large amount of initial capital.

In [25], a portfolio optimization method which calculates the VaR_α and optimizes the CVaR_α using a linear programming approach is developed. Portfolio strategies with a low CVaR_α necessarily have a low VaR_α. The method first consists in generating loss scenarios and then in introducing them as constraints in the linear programming problem. The 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. In our approach, we are not limited by the number of simulated scenarios.

We consider an energy (or financial) market operating at discrete trading dates $t_0 = 0 < t_1 < \cdots < t_M = T$. We have d assets available for trade with price process $X = \left(X^1, \cdots, X^d\right)$ and $X^i = \left(X^i_{t_\ell}\right)_{0 \le \ell \le M}$ for $i = 1, \cdots, d$. We will denote X_ℓ for X_{t_ℓ}. 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 a probability space $(\Omega, \mathcal{G}, \mathbb{P})$. In our framework, the source of market incompleteness comes from the presence in L of a state process Z that is *observable* but *not available* for trade. Thus, it induces a source of risk that is not completely hedgable. Typically, in the electricity market, the loss L suffered by an energy company may be due to an anormal annual electricity (or gas) anormal consumption. This consumption depends on the temperature, which is an *observable* but non tradable source of risk. In this example the process $(Z_\ell)_{1 \le \ell \le M}$ can be considered as the temperature which may influence not only the loss but the assets available for trade, *i.e.* electricity prices of spot and forward contracts (which are in this example the only available assets for hedge). More generally, this kind of dependance with respect to an observable but non available source of risk is a particularly relevant source of incompleteness in financial and energy markets (stochastic volatility, default time, temperature for energy derivatives, weather contracts, ...). The probability space is equiped with a filtration $\mathbb{G} = (\mathcal{G}_\ell)_{0 \le \ell \le M}$. Intuitively, $\mathcal{G}_\ell$ represents the observable information at time t_ℓ by all investors, so that $\mathcal{G}_\ell = \sigma\left\{X_i, Z_i; 0 \le i \le \ell\right\}$.

In order to reduce its risk (or hedge the contingent claim), the holder of the portfolio uses a dynamic self-financed strategy represented by a d-dimensional predictable process $\theta = (\theta_\ell)_{0 \le \ell < M}$, where $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{G}_\ell, \mathbb{P})$ ($L^0_{\mathbb{R}^d}(\mathcal{G}_\ell, \mathbb{P})$ denotes the space of all $\mathcal{G}$-mesurable and $\mathbb{P} - a.s.$ finite random variables with values in $\mathbb{R}^d$). In such a strategy, we may regard θ_ℓ as the number of shares invested in the stock at time t_ℓ. The gains from a trading strategy θ with an initial investment of 0 are described by the discrete stochastic integral $\sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell$, where we denote by ΔX_ℓ the increments $X_\ell - X_{\ell-1}$. Throughout the paper, we will use the following main assumptions

Assumption 1. *The process $(X_\ell)_{0\le\ell\le M}$ is a $(\mathbb{G}, \mathbb{P})$-martingale.*

and,

Assumption 2. *The process $(X_\ell, Z_\ell)_{0\le\ell\le M}$ taking its values in $\mathbb{R}^d \times \mathbb{R}^q$ is Markovian with respect to the filtration $\mathbb{G}$.*

The basic problem for the holder of the portfolio is to find the optimal self-financed strategy θ^* which minimizes the residual risk of the portfolio's losses, *i.e.* the solution of the following minimization problem

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

where $\mathcal{A}_\mathcal{G} = \left\{\theta = (\theta_\ell)_{0\le\ell\le M-1} \mid \theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{G}_\ell, \mathbb{P}),\ \ell = 0, \cdots, \ M-1\right\}$ is the set of admissible strategies, that is minimizing the residual risk of the portfolio risk profile over all self-financed strategies.

A natural question which arises is how to measure dynamically the risk of the considered portfolio in this context. To measure the risk at a given time t_ℓ, we introduce in a quite natural way and for the first time to our knowledge, the definition of a dynamic version of the CVaR that will be denoted $\mathcal{G}_\ell$-CVaR based on the Rockafellar & Uryasev's static representation of the CVaR. In order to estimate at time 0, this random risk measure, which reads as a conditional expectation, we use integration cubature formula based on optimal quantization.

For many reasons (transaction costs, difficulties to store energy assets, ...), the holder of the portfolio may not want to trade every day but may be only interested by a rough hedge to reduce its risk. Consequently, we firstly investigate one step self-financed strategies. Decided at time t_{ℓ_0} such strategy is obtained by setting $\theta_k \equiv \theta_{\ell_0}$, for $k = \ell_0, \cdots, M-1$. Consequently, a one-step portfolio strategy decided at time t_{ℓ_0} is an $\mathbb{R}^d$-valued random variable $\theta_{\ell_0} \in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0}, \mathbb{P})$. The investor risk at time t_{ℓ_0} can be measured by the quantity $\mathcal{G}_{\ell_0}$-$\mathrm{CVaR}_\alpha(L - \theta_{\ell_0}.(X_M - X_{\ell_0}))$[2] which is only known at time t_{ℓ_0}. However, the investor can estimate this quantity at time 0 by numerically computing $\mathbb{E}[\mathcal{G}_{\ell_0}\text{-}\mathrm{CVaR}_\alpha(L - \theta_{\ell_0}.(X_M - X_{\ell_0}))]$. This quantity is a forward risk, *i.e.* it is the best estimation at time 0 of the risk at time t_{ℓ_0} while the quantity $\mathrm{CVaR}_\alpha(L - \theta_{\ell_0}.(X_M - X_{\ell_0}))$ represents the risk at time 0. Consequently, there are two optimization problems.

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

$$\inf_{\theta_{\ell_0}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0},\mathbb{P})} \mathbb{E}\left[\mathcal{G}_{\ell_0}\text{-}\mathrm{CVaR}_\alpha(L - \theta_{\ell_0}.(X_M - X_{\ell_0}))\right]. \tag{2}$$

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

$$\inf_{\theta_{\ell_0}\in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0},\mathbb{P})} \mathrm{CVaR}_\alpha(L - \theta_{\ell_0}.(X_M - X_{\ell_0})). \tag{3}$$

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

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

The VaR_α and the CVaR_α are disymetric risk measures unlike standard deviation. By CVaR hedging we aim at modifying the shape of the loss distribution L, *i.e.* we reduce the right-hand side of the distribution which corresponds to high loss greater than the left-hand side which 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 [14] and [26] among others.

Under a Markovian framework, *i.e.* under Assumption 1, we propose a stochastic approximation algorithm to compute the optimal self-financed portfolio strategy θ^* solution of (3), (2) and (1) (and both the VaR and the CVaR of the resulting portfolio).
However, in the case of dynamic self-financed strategies framework, when the number of trading dates M is too large (say $M \geq 10$, in practice) or when the dimension of the process (X, Z) is too large, the proposed algorithm to solve (1) turns out to be numerically untractable. We develop other approaches based on some majorations of the objective function of (1) in order to approximate the optimal solution.

All proposed algorithms are built on some Rockafellar & Uryasev's representation of the CVaR and spatial discretization of the process $(X_\ell, Z_\ell)_{0\leq\ell\leq M}$ using optimal vector quantization. This leads us to devise a global Robbins-Monro (RM) procedure to estimate all the quantities of interest. This kind of idea has already been used in [3] to propose an algorithm which simultaneously computes both the VaR and the CVaR. The estimator provided by the algorithm satisfies the standard Central Limit Theorem (CLT) for recursive stochastic algorithm. However, the proposed algorithm is just a first building block. When α is close to 1 (otherwise the original procedure behaves well), VaR and CVaR are fundamentaly related to rare events. As a matter of fact, in this kind of probem, we are interested in hedging extreme events, *i.e.* events that are observed with a very small probability (usually less than 5%, 1% or even 0.1%) thus we obtain few significant scenarios to update our estimates. As a crucial improvement, we need to introduce a recursive variance reduction method. To compute more accurate estimates, it is necessary to generate more samples in the area of interest, the tail of the distribution. A natural tool used in this situation is importance sampling (IS). Following the IS procedure developed in [20], which has already been used in [3] for the estimation of the VaR and the CVaR, our IS parameters are optimized adaptively by a companion (unconstrained) RM algorithm which is combined with our first procedure. We also propose another variance reduction method based on a linear control variable which can be used alone when IS is not necessary or can be combined with the IS algorithm. It dramatically accelerates the convergence of the original procedure. The weak convergence rate of the resulting procedure is ruled by a CLT with optimal rate and minimal variance.

The paper is organized as follows: in Section 2, we present the dynamic version of the CVaR and develop some fundamental theoretical results on the $\mathcal{G}$-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 the optimal strategy with its associated VaR and CVaR. We establish its *a.s.* convergence and rate of convergence. In order to approximate conditional expectation, we rely on optimal vector quantization. We present our several algorithms to approximate the optimal strategy solution of (1) and briefly mention the two variance reduction tools in the static framework. Finally, Section 5 is devoted to numerical examples. We focus on the energy market which is known to be incomplete. We propose several portfolios to challenge the algorithm and display dynamic CVaR estimations.

Acknowledgements: We would like to thank Sandrine Bouthemy for fruitful and stimulating discussions.

Notations: • $|.|$ will denote the canonical Euclidean norm on $\mathbb{R}^d$, $u.v$ will denote the canonical inner product of the two column vector $u,\ v \in \mathbb{R}^d$ and u^T denotes the transpose of the column vector $u \in \mathbb{R}^d$.

- $\xrightarrow{\mathcal{L}}$ 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.
- $L^p(\mathbb{P})$ will denote the sub-space of random variable U such that $(\mathbb{E}[|U|^p])^{1/p} < +\infty$.
- $L^p(du)$ will denote the sub-space of function f such that $\left(\int |f|^p du\right)^{1/p} < +\infty$.

2 Theoretical aspects of CVaR hedging

2.1 Definitions and preliminaries

We start this section by briefly recalling the definitions of the VaR and the CVaR (for more details, we refer to [3]). Then, we introduce the notion of dynamic CVaR that will be fundamental throughout the paper. To measure the risk associated to a loss (or a short position on the contingent claim with payoff) L, one usually considers the VaR at level $\alpha \in (0,1)$ *i.e.* the lowest α-quantile of the distribution L

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

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:

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

If the distribution function is (strictly) increasing, the above equation has a unique solution, otherwise there may (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 Conditional Value-at-Risk (at level α). Assuming that $L \in L^1_{\mathbb{R}}(\mathbb{P})$, it is defined by:

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

The next proposition shows that these two quantities are solutions to a convex optimization problem which value function can be represented as an expectation, as pointed out in [25]. It has already been used in [3] 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:*

$$V(\xi) = \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\xi)_+\right]. \tag{4}$$

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

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

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

$$\forall\, \xi \in \mathbb{R}, \quad V'(\xi) = \mathbb{E}\left[1 - \frac{1}{1-\alpha}\mathbf{1}_{\{L \ge \xi\}}\right]. \tag{5}$$

Furthermore,

$$\mathrm{CVaR}_\alpha(L) = \min_{\xi \in \mathbb{R}} V(\xi). \tag{6}$$

We refer to [25] or [3] for a proof. Now we are in position to define the dynamic CVaR. We consider a sub σ-field $\mathcal{F} \subseteq \mathcal{G}$, representative of the information observable by all investors. Given the above result concerning the CVaR, it is quite natural to define the $\mathcal{G}$-CVaR according to the following definition.

Definition 2.1. Suppose that L satisfies $\mathbb{E}[L_+ \mid \mathcal{F}] < +\infty$ *a.s.* The $\mathcal{F}$-CVaR is a random risk measure defined by

$$\mathcal{F}\text{-CVaR}_\alpha(L) := \operatorname*{ess\,inf}_{\xi \in L^0(\mathcal{F})} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\xi)_+ \,|\mathcal{F}\right].$$

By construction, it is straightforward that it satisfies the following *coherence* properties

1. Sub-additivity: for every $L,\ L'$ such that $\mathbb{E}\left[L_+ + L'_+ \mid \mathcal{F}\right] < +\infty$, *a.s.*, $\mathcal{F}\text{-CVaR}_\alpha(L+L') \leq \mathcal{F}\text{-CVaR}_\alpha(L) + \mathcal{F}\text{-CVaR}_\alpha(L')$.

2. Positive homogeneity: If $\lambda \in L^0_{\mathbb{R}}(\mathcal{F})$ with $\lambda \geq 0$ *a.s.*, $\mathcal{F}\text{-CVaR}_\alpha(\lambda L) = \lambda \times \mathcal{F}\text{-CVaR}_\alpha(L)$.

3. Translation invariance: for all $Z \in L^0_{\mathbb{R}}(\mathcal{F})$, $\mathcal{F}\text{-CVaR}_\alpha(L+Z) = Z + \mathcal{F}\text{-CVaR}_\alpha(L)$.

4. Monotonicity: for every $L,\ L'$ such that $\mathbb{E}\left[L_+ + L'_+ \mid \mathcal{F}\right] < +\infty$ and $L \leq L'$ *a.s.*, $\mathcal{F}\text{-CVaR}_\alpha(L) \leq \mathcal{F}\text{-CVaR}_\alpha(L')$.

When $\mathcal{F} = \{\varnothing, \Omega\}$, the $\mathcal{F}\text{-CVaR}_\alpha(L)$ coincides with the usual $\text{CVaR}_\alpha(L)$. To estimate $\mathcal{F}\text{-CVaR}_\alpha(L)$ at time 0, one may compute the quantity $\mathbb{E}[\mathcal{F}\text{-CVaR}_\alpha(L)]$ which is still a coherent risk measure in the sense of [1].

2.2 General properties

In this section, we state some useful properties satisfied by the $\mathcal{G}_\ell\text{-CVaR}_\alpha$.

If one aims at measuring the risk at time t_ℓ of his financial strategy $\theta \in \mathcal{A}$ started at time $t_0 = 0$ using a CVaR criterion, one has to compute $\mathcal{G}_\ell\text{-CVaR}_\alpha\left(L - \sum_{p=1}^M \theta_{p-1}.\Delta X_p\right)$, which is only known at time t_ℓ. It is natural for the holder of the portfolio to ask how the risk evolves with time until maturity. Next result shows that the $\mathcal{G}_\ell$-CVaR risk measure is time consistent, *i.e.* the risk of any position decreases with time.

Proposition 2.2. *We set $M = +\infty$ for this result. Let $Y \in L^1_{\mathbb{R}}(\mathcal{G}_\infty, \mathbb{P})$ where $\mathcal{G}_\infty = \vee_\ell \mathcal{G}_\ell$. The sequence $(\mathcal{G}_\ell\text{-CVaR}_\alpha(Y))_{1\leq \ell \leq M}$ is a $\mathbb{G}$-supermartingale. Moreover, it satisfies,*

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

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

$$\mathcal{G}_\ell\text{-CVaR}_\alpha(Y) = \operatorname*{ess\,inf}_{\xi \in L^0(\mathcal{G}_\ell)} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y-\xi)_+ \,|\mathcal{G}_\ell\right] \leq \frac{1}{1-\alpha}\mathbb{E}\left[Y_+|\mathcal{G}_\ell\right] \in L^1(\mathbb{P}),$$

and by Jensen's inequality,

$$\mathbb{E}\left[Y|\mathcal{G}_\ell\right] = \operatorname*{ess\,inf}_{\xi \in L^0(\mathcal{G}_\ell)} \xi + \frac{1}{1-\alpha}\left(\mathbb{E}\left[Y|\mathcal{G}_\ell\right] - \xi\right)_+ \leq \mathcal{G}_\ell\text{-CVaR}_\alpha(Y), \tag{7}$$

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

$$\mathcal{G}_\ell\text{-CVaR}_\alpha(Y) \leq \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y-\xi)_+ \,|\mathcal{G}_\ell\right], \quad \text{for all } \xi \in L^0_{\mathbb{R}}(\mathcal{G}_{\ell-1}),$$

which implies that

$$\mathbb{E}\left[\mathcal{G}_\ell\text{-CVaR}_\alpha(Y)|\mathcal{G}_{\ell-1}\right] \leq \operatorname*{ess\,inf}_{\xi \in L^0(\mathcal{G}_{\ell-1})} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[(Y-\xi)_+ \,|\mathcal{G}_{\ell-1}\right] = \mathcal{G}_{\ell-1}\text{-CVaR}_\alpha(Y).$$

Consequently, the sequence $(\mathcal{G}_\ell\text{-CVaR}_\alpha(Y))_{1\leq \ell\leq M}$ is a $\mathbb{G}$-supermartingale.
Now, owing to (7), for $n \geq 1$

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

and

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

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 $(\mathbb{E}\left[Y|\mathcal{G}_n\right])_{n\geq 1}$ toward $\mathbb{E}\left[Y|\mathcal{G}_\infty\right] = Y$ (the convergence also holds in L^1), we get

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

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

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

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

$$\mathbb{E}\left[\mathbf{1}_A\mathcal{G}_n\text{-CVaR}_\alpha(Y)\right] \leq \mathbb{E}\left[\mathbf{1}_A\left(Y + \frac{1}{1-\alpha}\left(Y - \mathbb{E}\left[Y|\mathcal{G}_n\right]\right)_+\right)\right]. \tag{8}$$

It follows from Fatou's Lemma that

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

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

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

Combining these inequalities with (8) yields

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

which in turn implies that

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

This completes the proof. □

This result naturally implies that the sequence $(\mathbb{E}\left[\mathcal{G}_\ell\text{-CVaR}_\alpha(Y)\right])_{1\leq\ell\leq M}$ is non-increasing, 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 satisfies L is $\mathcal{G}_M$-measurable (as it is the case in our modelization) 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 2.3. *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 $\ell = 1, \cdots, M$. Let $\theta \in \mathcal{A}$ such that $\theta_\ell \in L^p_{\mathbb{R}^d}(\mathbb{P})$ with $p = \frac{p'}{p'-1}$. Then,*

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

and satisfies, for every $k \in \{0, \cdots, M-1\}$,

$$\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) = \mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) - \sum_{\ell=1}^{\ell} \theta_{\ell-1}\Delta X_\ell. \tag{9}$$

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

$$\begin{aligned}
\mathcal{G}_k\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right) &= \operatorname*{ess\,inf}_{\xi \in L^0(\mathcal{G}_k)} \xi + \frac{1}{1-\alpha}\mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+ \middle| \mathcal{G}_k\right] \\
&= \sum_{\ell=1}^{k} \theta_{\ell-1}\Delta X_\ell \\
&+ \operatorname*{ess\,inf}_{\tilde{\xi} \in L^0(\mathcal{G}_k)} \tilde{\xi} + \frac{1}{1-\alpha}\mathbb{E}\left[\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell - \tilde{\xi}\right)_+ \middle| \mathcal{G}_k\right] \\
&= \sum_{\ell=1}^{k} \theta_{\ell-1}\Delta X_\ell + \mathcal{G}_\ell\text{-CVaR}_\alpha\left(L - \sum_{\ell=k+1}^{M} \theta_{\ell-1}.\Delta X_\ell\right).
\end{aligned}$$

□

In particular, if X is a $(\mathcal{G}, \mathbb{P})$-martingale (9) implies that for every $k \in \{0, \cdots, M-1\}$

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

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

2.3 CVaR hedging using a one step self financed strategy

In this section, we adress the two problems (2) and (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 t_{ℓ_0}.

By one step strategy decided at time t_{ℓ_0}, we mean that the investor is restricted to rebalance its portfolio only once at time $t_\ell \in \{t_0, \cdots, t_{M-1}\}$. By a one step static strategy, we mean that the investor uses a one step strategy decided at time $t_0 = 0$.

This case of study is interesting since in energy markets, practiciens may be interested only 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 will be built on similar ideas used in this section.

Without loss of generality, we can suppose that the market operates with only two dates t_{ℓ_0} and $T = t_M$. We will denote X for $X_M - X_{\ell_0}$. Actually, we use a general σ-algebra $\mathcal{F} \subseteq \mathcal{A}$ with the possibility of setting $\mathcal{F}$ to $\mathcal{G}_{\ell_0}$ with $\ell_0 = 0, \cdots, M-1$. Consequently, we consider the two more general problems

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)\right], \tag{10}$$

and

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \text{CVaR}_\alpha\left(L - \theta.X\right). \tag{11}$$

Note that (11) can be written

$$\inf_{\xi \in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right] \tag{12}$$

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

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+\right]. \tag{13}$$

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 (13) is equivalent to minimizing the short fall risk

$$\inf_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\left(L - \theta.X\right)_+\right]. \tag{14}$$

First, we will show that there exists an optimal one step trading strategy $\tilde{\theta}$ solution to (14) thus for all $\xi \in \mathbb{R}$ there exists $\theta^*_\alpha(\xi)$ solution to (13). Finally, we will come back to (12) and deduce the existence of an optimal ξ^*_α solution of (12)

Now in order to derive the existence of solutions to (10) and (11), we assume the existence of a regular conditional distribution of the couple (L, X) given $\mathcal{F}$ 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 conditionnal distribution of the couple (L, X).

Assumption 3 (Static Case)**.**

i) The distribution of L *and* X *satisfies* $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $X \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$.

ii) $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathbb{E}\left[(u.X)_+ \mid \mathcal{F}\right] > 0$ *a.s.*

Assumption 4 (Forward Case)**.**

i) The distribution of L *and* X *satisfies* $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $X \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$.

ii) $\operatorname{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathcal{F}\text{-CVaR}_\alpha\left(u.X\right) > 0$ *a.s.*

The following proposition is the key result to solve our optimization problem. The proof is postponed to an appendix and relies on classical arguments from stochastic control theory.

Proposition 2.4. *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*

$$V_f(\omega, \xi, \theta) = \int v_f(\xi, \theta, y, x)\Pi(\omega, \mathrm{d}x, \mathrm{d}y), \tag{15}$$

$$V_s(\omega,\xi,\theta) = \int v_s(\theta,y,x)\Pi(\omega,\mathrm{d}x,\mathrm{d}y) \tag{16}$$

where

$$v_f(\xi,\theta,y,x) = \xi + \frac{1}{1-\alpha}\left(y-\theta.x-\xi\right)_+, \tag{17}$$

and

$$v_s(\theta,y,x) = \left(y-\theta.x\right)_+, \tag{18}$$

Then, we have

i) Static Risk: *Suppose that Assumption 3 is satisfied. Then, for all $\omega\in\Omega$, the function $V_s(\omega,.)$ is convex, lipschitz continuous and $\lim_{|\theta|\to+\infty} V_s(\omega,\theta) = +\infty$. Moreover, we have*

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[(L-\theta.X)_+\right] = \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[(L-\theta.X)_+ \,\middle|\, \mathcal{F}\right]\right], \tag{19}$$

and

$$\operatorname*{ess\,inf}_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[(L-\theta.X)_+ \,\middle|\, \mathcal{F}\right](\omega) = \min_{\theta\in\mathbb{R}^d} V_s(\omega,\theta). \tag{20}$$

ii) Forward Risk: *Suppose that Assumption 4 is satisfied. Then, for all $\omega\in\Omega$, the function $V_f(\omega,.,.)$ is convex, continuous and for all $\xi\in\mathbb{R}$, $\lim_{|(\xi,\theta)|\to+\infty} V_f(\omega,\xi,\theta) = +\infty$. Moreover, we have*

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\theta.X\right)\right] = \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ \xi\in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L-\theta.X-\xi\right)_+ \,\middle|\, \mathcal{F}\right]\right], \tag{21}$$

and

$$\operatorname*{ess\,inf}_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ \xi\in L^0_{\mathbb{R}}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L-\theta.X-\xi\right)_+ \,\middle|\, \mathcal{F}\right](\omega) = \min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\omega,\xi,\theta). \tag{22}$$

Remark 2.1. The non-degeneracy Assumptions 3 *ii)* and 4 *ii)* can be replaced by the stronger assumption:

- $\mathbb{E}\left[X \mid \mathcal{F}\right] = 0$ and $\mathbb{E}\left[XX^T \mid \mathcal{F}\right]$ is *a.s.* positive definite in $\mathcal{S}(d,\mathbb{R})$.
- The conditional distribution of X given $\mathcal{F}$ is continuous (no affine hyperplane has positive mass).

Indeed, for $\omega\in\Omega$, we have $\operatorname*{ess\,inf}_{u\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),|u|=1} \mathcal{F}\text{-CVaR}_\alpha\left(u.X\right)(\omega) = \inf_{\xi\in\mathbb{R},u\in\mathcal{S}_d(0,1)} V_f(\omega,\xi,u)$ where $\mathcal{S}_d\left(0,1\right) := \left\{u\in\mathbb{R}^d \mid |u|=1\right\}$ 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 $u,u'\in\mathcal{S}_d\left(0,1\right)$,

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

Consequently, for all $\omega \in \Omega$, the function $u \mapsto \inf_{\xi\in\mathbb{R}} V_f(\omega,\xi,u)$ is continuous on $\mathcal{S}_d(0,1)$. Thus, it remains to check that for all $u \in \mathcal{S}_d(0,1)$, $\inf_{\xi\in\mathbb{R}} V_f(\omega,\xi,u) > 0$, knowing that $\mathbb{E}[X \mid \mathcal{F}] = 0$ *a.s.* Proposition 2.1 implies that there exists ξ_α^* such that $\inf_{\xi\in\mathbb{R}} V_f(\omega,\xi,u) = V_f(\omega,\xi_\alpha^*,u)$. There are three cases to check:

- if $\xi_\alpha^* > 0$, then it is straighforward that $V_f(\omega,\xi_\alpha^*,u) \geq \xi_\alpha^* > 0$,
- if $\xi_\alpha^* < 0$, then Jensen's inequality leads to

$$V_f(\omega,\xi_\alpha^*,u) \geq \xi_\alpha^* + \frac{1}{1-\alpha}\left(u.\int x\Pi(\omega,\mathrm{d}x,\mathrm{d}y) - \xi_\alpha^*\right)_+ = -\frac{\alpha}{1-\alpha}\xi_\alpha^* > 0,$$

- if $\xi_\alpha^* = 0$, $\int v_f(\xi_\alpha^*,u,0,x)\,\Pi(\omega,\mathrm{d}x,\mathrm{d}y) = \frac{1}{1-\alpha}\mathbb{E}\left[(u.X)_+ \mid \mathcal{F}\right](\omega)$. Now, if $\mathbb{E}\left[(u.X)_+ \mid \mathcal{F}\right] = 0$, then $\mathbb{E}[|u.X| \mid \mathcal{F}] = 0$, since $\mathbb{E}[u.X \mid \mathcal{F}] = 0$. Then $u.X = 0$ *a.s.*, so that it implies that $u = 0$, which is impossible.

The right-hand sides of (20) and (22) show that the two optimization problems (10) and (11) can be written

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha(L-\theta.X)\right] = \mathbb{E}\left[\min_{(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d} V_f(\xi,\theta)\right], \tag{23}$$

and,

$$\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F})} \mathrm{CVaR}_\alpha(L-\theta.X) = \inf_{\xi\in\mathbb{R}} \mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right], \tag{24}$$

respectively. Consequently, for all $\omega \in \Omega$, we have to solve deterministic optimization problems. 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 2.5. *Suppose Assumption 3 is satisfied. Then, for all* $\xi \in \mathbb{R}$

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

If Assumption 4 is satisfied then

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

where the gradient of V_f *can be represented for every* $(\xi,\theta) \in \mathbb{R}\times\mathbb{R}^d$ *by*

$$\nabla_{(\xi,\theta)} V_f(\xi,\theta) = \int \nabla_{(\xi,\theta)} v_f(\xi,\theta,y,x)\Pi(\mathrm{d}x,\mathrm{d}y) \tag{25}$$

and,

$$\nabla_\theta V_f(\xi,\theta) = \int \nabla_\theta v_f(\xi,\theta,y,x)\Pi(\mathrm{d}x,\mathrm{d}y). \tag{26}$$

Moreover, $\xi \mapsto \mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right]$ *is Lipschitz continuous, convex, and* $\lim_{|\xi|\to+\infty} \mathbb{E}\left[\min_{\theta\in\mathbb{R}^d} V_f(\xi,\theta)\right] = +\infty$. *Consequently,* (10) *and* (11) *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 to get (25) and (26), we only need to check the domination property. First note that we have, for all $(y,x)\in\mathbb{R}\times\mathbb{R}^d$

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

so that there exists $C>0$ such that

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

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

$$\begin{aligned}\left|\mathbb{E}\left[\inf_{\theta\in\mathbb{R}^d}V_f(\xi,\theta)\right]-\mathbb{E}\left[\inf_{\theta\in\mathbb{R}^d}V_f(\xi',\theta)\right]\right| &\leq \mathbb{E}\left[\sup_{\theta\in\mathbb{R}^d}\left|V_f(\xi,\theta)-V_f(\xi',\theta)\right|\right]\\ &\leq K\ |\xi-\xi'|,\end{aligned}$$

and, owing to Jensen's inequality

$$\mathbb{E}\left[\min_{\theta\in\mathbb{R}^d}V_f(\xi,\theta)\right]\geq \xi+\frac{1}{1-\alpha}\left(\mathbb{E}\left[L\right]-\xi\right)_+,$$

so 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 CVaR hedging using a dynamic self financed strategy

In this section, we adress the main problem (1), that is hedging a contingent claim with a dynamic self-financed strategy starting with an initial investment of 0 using a static CVaR criterion at a fixed time $t=0$. Actually, in this theoretical section, we consider the more general multistage stochastic optimization problem:

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathrm{CVaR}_\alpha\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell\right)=\inf_{\xi\in\mathbb{R}}\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell-\xi\right)_+\right]\tag{27}$$

where $\mathcal{A}_{\mathcal{F}}$ is the set of all sequences $\theta=(\theta_0,\cdots,\theta_{M-1})$ such that $\theta_\ell\in L^0(\mathcal{F}_\ell)$, $\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\cdots\mathcal{F}_{M-1}\subseteq\mathcal{A}$ fixed σ-algebras. Later on, for numerical applications, we will set $\mathcal{F}_\ell$ to $\mathcal{G}_\ell$, $\ell=0,\cdots,M-1$.

Note that in order to solve (27), we may adress firstly the multistage stochastic optimization problem

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathbb{E}\left[\xi+\frac{1}{1-\alpha}\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell-\xi\right)_+\right],\quad\text{for each}\ \ \xi\in\mathbb{R}.\tag{28}$$

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 (28) is equivalent to minimizing the shortfall risk

$$\inf_{\theta\in\mathcal{A}_{\mathcal{F}}}\mathbb{E}\left[\left(L-\sum_{\ell=1}^M\theta_{\ell-1}.\Delta X_\ell\right)_+\right].\tag{29}$$

The optimization (29) is a classical stochastic control problem. One may think that it is possible to derive existence of solutions of problem (29) using results on dynamic programming. Unfortunately, in our case, standard assumptions of dynamic programming are not fulfilled (see e.g. [10]).

However, we can adapt this classical approach in order to derive the existence of an optimal shortfall-hedging sequence $\tilde{\theta} := (\tilde{\theta}_\ell)_{0\leq \ell \leq M-1}$ solution of (29), thus we obtain the existence of an optimal CVaR-hedging sequence $\theta^*_\alpha := (\theta^*_{\ell,\alpha})_{0\leq \ell\leq M-1}$ solution of (28). Finally, we will come back to (27) and using similar arguments to those of the static framework, we will deduce the existence of ξ^*_α solution of the problem

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

In order to derive similar results to those obtained in Section 2.3, we consider a family of regular conditional distributions $(\Pi_\ell)_{0\leq\ell\leq M-1}$ where $\Pi_\ell(\mathrm{d}y, \mathrm{d}x) = \Pi_\ell(\omega, \mathrm{d}y, \mathrm{d}x)$ denotes the regular conditional distribution of the couple $(L, \Delta X_1, \cdots, \Delta X_M)$ given $\mathcal{F}_\ell$ and we make the following assumption.

Assumption 5.

i) The distribution of $(L, \Delta X_1, \cdots, \Delta X_M)$ *satisfies* $L \in \mathbb{L}^1_{\mathbb{R}}(\mathbb{P})$, $\Delta X_\ell \in \mathbb{L}^1_{\mathbb{R}^d}(\mathbb{P})$, $\ell = 1, \cdots, M$

ii) $\operatorname{ess\,inf}_{u\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),\ |u|=1} \mathbb{E}\left[(u.\Delta X_\ell)_+ \mid \mathcal{F}_{\ell-1}\right] > 0$ *a.s.*

Remark 2.2. In the same way that Remark 2.1 in the one step framework, the non-degeneracy Assumption 5 can be replaced by the stronger assumption

- $\mathbb{E}\left[X_\ell \mid \mathcal{F}_{\ell-1}\right] = 0$ and $\mathbb{E}\left[X_\ell X_\ell^T \mid \mathcal{F}_{\ell-1}\right]$ is *a.s.* positive definite in $\mathcal{S}(d, \mathbb{R})$ for $\ell = 1, \cdots, M$.
- The conditional distribution of X_ℓ given $\mathcal{F}_{\ell-1}$ is continuous (no affine hyperplane has positive mass) for $\ell = 1, \cdots, M$.

In the spirit of the dynamic programming principle, we construct the solution of (29) using a step by step backward induction. To be more precise, using similar arguments to those used to prove (21), one first notices that (29) can be written

$$\inf_{\theta_\ell\in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell,\mathbb{P}),\ell=0,\cdots,M-2} \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta_{M-1}\in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1},\mathbb{P})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell\right)_+ \middle| \mathcal{F}_{M-1}\right]\right], \tag{30}$$

so that one may start by solving the following problem

$$\operatorname*{ess\,inf}_{\theta_{M-1}\in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1},\mathbb{P})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell\right)_+ \middle| \mathcal{F}_{M-1}\right](\omega) = \min_{\theta_{M-1}\in\mathbb{R}^d} V_{M-1}(\omega, \theta_{0:M-2}, \theta_{M-1}) \tag{31}$$

$$= V_{M-1}(\omega, \theta_{0:M-2}, \tilde{\theta}_{M-1}) \quad a.s. \tag{32}$$

where $\tilde{\theta}_{M-1} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1})$, V_{M-1} is defined for all $\omega \in \Omega$, $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\ell = 1, \cdots, M-1$, by

$$V_{M-1}(\omega, \theta_{0:M-2}, \theta_{M-1}) := \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M}\theta_{\ell-1}.\Delta X_\ell\right)_+ \middle| \mathcal{F}_{M-1}\right](\omega)$$

$$= \int \left(y - \sum_{\ell=1}^{M}\theta_{\ell-1}(\omega).\Delta x_\ell\right)_+ \Pi_{M-1}(\omega, \mathrm{d}x, \mathrm{d}y). \tag{33}$$

This follows from similar arguments to those of the proof of Proposition 2.4, *i.e.* from the fact that for all $\omega \in \Omega$, $\theta_\ell \in L^0(\mathcal{F}_\ell)$, $\ell = 1, \cdots, M-2$, the function (defined on $\mathbb{R}^d$) by $\theta_{M-1} \mapsto V_{M-1}(\xi, \theta_{0:M-2}, \theta_{M-1})$ is convex, Lipschitz continuous and $\lim_{|\theta_{M-1}| \to +\infty} V_{M-1}(\xi, \theta_{0:M-2}, \theta_{M-1}) = +\infty$. Thus, it implies that (31) has a solution that we denote by $\tilde{\theta}_{M-1} := \tilde{\theta}_{M-1}(\omega, \theta_0, \cdots, \theta_{M-2})$, which is $\mathcal{F}_{M-1}$-measurable (owing to measurable selection, see e.g. Lemma 3 and Lemma 4 in [10]) so that (32) holds.

Then we proceed by a backward induction: we denote by $\tilde{\theta}_{\ell:M-1} := (\tilde{\theta}_\ell, \cdots, \tilde{\theta}_{M-1})$ the solution built down to step ℓ. At step $\ell - 1$, we adress for every $\theta_{0:\ell-2} \in L^0_{\mathbb{R}^d}(\mathcal{F}_0, \mathbb{P}) \times \cdots \times L^0_{\mathbb{R}^d}(\mathcal{F}_{\ell-2}, \mathbb{P})$, the problem

$$\begin{aligned}\operatorname*{ess\,inf}_{\theta_{\ell-1} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{\ell-1}, \mathbb{P})} \mathbb{E}\left[V_\ell(\theta_{0:\ell-1}, \tilde{\theta}_\ell) \middle| \mathcal{F}_{\ell-1} \right](\omega) &= \min_{\theta_{\ell-1} \in \mathbb{R}^d} V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1}) \\ &= V_{\ell-1}(\omega, \theta_{0:\ell-2}, \tilde{\theta}_{\ell-1}) \quad a.s. \end{aligned} \tag{34}$$

where for all $\theta_k \in L^0_{\mathbb{R}^d}(\mathcal{F}_k, \mathbb{P})$, $k = 0, \cdots, \ell-1$, the functions V_ℓ and $V_{\ell-1}$ are defined by

$$\begin{aligned} V_\ell(\omega, \theta_{0:\ell-1}, \tilde{\theta}_\ell) &:= \mathbb{E}\left[\left(L - \sum_{k=1}^{\ell} \theta_{k-1}.\Delta X_k - \sum_{k=\ell+1}^{M} \tilde{\theta}_{k-1}.\Delta X_k \right)_+ \middle| \mathcal{F}_\ell \right](\omega) \\ &= \int \left(y - \sum_{k=1}^{\ell} \theta_{k-1}(\omega).\Delta x_k - \sum_{k=\ell+1}^{M} \tilde{\theta}_{k-1}.\Delta x_k \right)_+ \Pi_\ell(\omega, \mathrm{d}x, \mathrm{d}y) \end{aligned} \tag{35}$$

and,

$$V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1}) = \int \left(y - \sum_{k=1}^{\ell} \theta_{k-1}(\omega).\Delta x_k - \sum_{k=\ell+1}^{M} \tilde{\theta}_{k-1}.\Delta x_k \right)_+ \Pi_{\ell-1}(\omega, \mathrm{d}x, \mathrm{d}y) \tag{36}$$

The following proposition implies that (27) has an optimal solution $(\xi^*_\alpha, \theta^*_\alpha) \in \mathbb{R} \times \mathcal{A}_\mathcal{F}$.

Proposition 2.6. *Suppose that Assumption 5 is satisfied. Then,*

i) (30) *is satisfied, the problem* (31) *has a solution and for $\ell = M-1, \cdots, 1$, one can find an $\mathcal{F}_{\ell-1}$-measurable random variable $\tilde{\theta}_{\ell-1}$ solution of* (34)*. Thus,* (28) *has an optimal solution denoted by $\theta^*_\alpha := (\theta^*_{\alpha,\ell})_{0 \le \ell \le M-1}$.*

ii) The function $\xi \mapsto \xi + \frac{1}{1-\alpha} \inf_{\theta \in \mathcal{A}_\mathcal{F}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^M \theta_{\ell-1}.\Delta X_\ell - \xi \right)_+ \right]$ is convex, Lipschitz continuous and satisfies $\lim_{|\xi| \to +\infty} \xi + \frac{1}{1-\alpha} \inf_{\theta \in \mathcal{A}_\mathcal{F}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^M \theta_{\ell-1}.\Delta X_\ell - \xi \right)_+ \right] = +\infty$ so that (27) *admits a solution.*

Proof. *i*) The proof of (30) uses similar arguments than those used in the proof of (19).
Let $\tilde{\theta}_{M-1}$ be a solution of (31). We go one step backward. For all $\omega \in \Omega$, $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell)$, $\ell = 0, \cdots, M-3$, (using the definition of V_{M-1}) we are interested by the function

$\theta_{M-2} \mapsto V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2}) :=$

$$\mathbb{E}\left[\underbrace{\operatorname*{ess\,inf}_{\theta_{M-1} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M-2} \theta_{\ell-1}.\Delta X_\ell - \theta_{M-2}.\Delta X_{M-1} - \tilde{\theta}_{M-1}.\Delta X_M \right)_+ \middle| \mathcal{F}_{M-1} \right]}_{=V_{M-1}(\theta_{0:M-2}, \tilde{\theta}_{M-1})} \middle| \mathcal{F}_{M-2} \right](\omega).$$

It is straightforward that this function is convex. Let $\theta_{M-2},\ \theta'_{M-2} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-2}, \mathbb{P})$, 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

$$\left|V_{M-2}(\theta_{0:M-3}, \theta_{M-2}) - V_{M-2}(\theta_{0:M-3}, \theta'_{M-2})\right| \le |\theta_{M-2} - \theta'_{M-2}| \mathbb{E}\left[|\Delta X_{M-1}| \mid \mathcal{F}_{M-2}\right] \quad a.s.$$

so that the function is Lipschitz continuous.

Lemma 2.7. *(Conditional Jensen's inequaility) Let F be a non-negative convex function and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. If X is random variable such that $\mathbb{E}\left[|X| \mid \mathcal{B}\right] < +\infty$, a.s. then*

$$\mathbb{E}\left[F(X) \mid \mathcal{B}\right] \ge F\left(\mathbb{E}\left[X \mid \mathcal{B}\right]\right) \quad a.s.$$

Proof. This is a straightforward adaptation of the proof of Jensen's inequality. □

Now, owing to Lemma 2.7 , we have

$$V_{M-2}(\theta_{0:M-3}, \theta_{M-2}) \ge \mathbb{E}\left[\left(\mathbb{E}\left[L \mid \mathcal{F}_{M-1}\right] - \sum_{\ell=1}^{M-2} \theta_{\ell-1}.\Delta X_\ell - \theta_{M-2}.\Delta X_{M-1}\right)_+ \middle| \mathcal{F}_{M-2}\right]. \tag{37}$$

We aim at showing that the right-hand side of (37) goes to infinity as $|\theta_{M-2}| \to +\infty$. First, the sub-additivity of the function $x \mapsto x_+$ implies that

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

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

$$|\theta_{M-2}| \operatorname*{ess\,inf}_{u \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-2}, \mathbb{P}),\ |u|=1} \mathbb{E}\left[(u.\Delta X_{M-1})_+ \middle| \mathcal{F}_{M-2}\right],$$

Consequently, Assumption 5 implies that for all $\omega \in \Omega$, $\theta_\ell \in L^0_{\mathbb{R}^d}(\mathcal{F}_\ell, \mathbb{P})$, $\ell = 0, \cdots, M-3$, $\lim_{|\theta_{M-2}| \to +\infty} V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2}) = +\infty$ and the function $\theta_{M-2} \mapsto V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2})$ has a minimum $\tilde{\theta}_{M-2}$ which is $\mathcal{F}_{M-2}$-measurable owing to measurable selection theorem.

Furthermore, using similar arguments to those used for the proof of Proposition 2.4, one shows that for all $\omega \in \Omega$, for every $\theta_{0:M-3} \in L^0_{\mathbb{R}^d}(\mathcal{F}_0) \times \cdots \times L^0_{\mathbb{R}^d}(\mathcal{F}_{M-3})$

$$\begin{aligned}&\operatorname*{ess\,inf}_{\theta_{M-2} \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-2})} V_{M-2}(\omega, \theta_{0:M-3}, \theta_{M-2}) = \\ &\qquad \operatorname*{ess\,inf}_{(\theta_{M-2}, \theta_{M-1}) \in L^0_{\mathbb{R}^d}(\mathcal{F}_{M-2}) \times L^0_{\mathbb{R}^d}(\mathcal{F}_{M-1})} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)_+ \mid \mathcal{F}_{M-2}\right] \quad a.s.\end{aligned}$$

Now if the solution is built down to step ℓ, for all $\omega \in \Omega$ and $\theta_k \in L^0(\mathcal{F}_k)$, $k = 0, \cdots, \ell-2$, one shows that the function $\theta_{\ell-1} \mapsto V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1})$ is convex, Lipschitz continuous and satisfies $\lim_{|\theta_{\ell-1}| \to +\infty} V_{\ell-1}(\omega, \theta_{0:\ell-2}, \theta_{\ell-1}) = +\infty$. Consequently, there exists $\tilde{\theta}_{\ell-1}$ solution of (34). Thus, (28) has an optimal $\theta^*_\alpha := (\theta^*_{\alpha,\ell})_{0 \le \ell \le M-1}$.

Now we come back to our original problem

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

For all $x \in \mathbb{R}$, the functions $\xi \mapsto \xi + \frac{1}{1-\alpha}(x-\xi)_+$ are convex and Lipschitz continuous so that $\xi \mapsto \xi + \frac{1}{1-\alpha} \inf_{\theta \in \mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+\right]$ is convex, Lipschitz continuous. Owing to Assumption 5 and using Lemma 2.7 with a backward induction, one shows

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

so that $\lim_{|\xi| \to +\infty} \xi + \frac{1}{1-\alpha} \inf_{\theta \in \mathcal{A}_{\mathcal{F}}} \mathbb{E}\left[\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell - \xi\right)_+\right] = +\infty$. 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 (3), (2) and (1). First, we will focus on (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).

3.1 Markovian framework and Optimal Vector Quantization

In order to simplify the numerical computation of conditional expectations that appear in the problem (2), we will work under Assumption 2.

To be more precise, from a modeling point of view, it is quite natural to consider that the random variable L can be written as a function of the process (X, Z), *i.e.* $L = \phi(X, Z)$. Typically, in the electricity market, Z can be considered as the temperature process and may influence electricity spot prices and electricity forward prices. Consequently, we assume that there exists two continuous functions $F : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}$ and $G : \left(\mathbb{R}^d\right)^{\ell_0+1} \times \left(\mathbb{R}^q\right)^{\ell_0+1} \times \mathbb{R}^{r_{\ell_0}} \to \mathbb{R}^d$ such that

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

where U_{ℓ_0+1} is a r_{ℓ_0}-dimensional random variable independent of $\mathcal{G}_{\ell_0} := \sigma\left(X_{\ell_0}, Z_{\ell_0}\right)$. We will denote U for U_{ℓ_0+1}. Under this markovian framework, the function (15) 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 (22) becomes

$$\operatorname*{ess\,inf}_{\theta \in L^0_{\mathbb{R}^d}(\mathcal{G}_{\ell_0}, \mathbb{P}), \xi \in L^0_{\mathbb{R}}(\mathcal{G}_{\ell_0}, \mathbb{P})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L - \theta.X - \xi\right)_+ \middle| \mathcal{G}_{\ell_0}\right] = \left(\min_{(\xi,\theta) \in \mathbb{R} \times \mathbb{R}^d} V(\xi, \theta, x, z)\right)_{(x,z)=(X_{\ell_0}, Z_{\ell_0})} \quad a.s. \tag{38}$$

Consequently, in order to solve the global problem (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

$$\mathbb{E}\left[\left(\inf_{(\theta,\xi) \in \mathbb{R}^d \times \mathbb{R}} V(\xi, \theta, x, z)\right)_{|x=X_{\ell_0}, z=Z_{\ell_0}}\right]. \tag{39}$$

When the dimension of the random variable (X_{ℓ_0}, Z_{ℓ_0}) is large (greater than $5, 10$), one can use Monte-Carlo simulations and estimates (39) 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 $5, 10$), 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 $\left(\hat{X}_{\ell_0}, \hat{Z}_{\ell_0}\right)$ of the random variable (X_{ℓ_0}, Z_{ℓ_0}), based on an optimal quantization grid $\Gamma_{\ell_0} := \Gamma^{N_{\ell_0}}_{(X_{\ell_0},Z_{\ell_0})} = \left(\left(x^1_{\ell_0}, z^1_{\ell_0}\right), \cdots, \left(x^{N_{\ell_0}}_{\ell_0}, z^{N_{\ell_0}}_{\ell_0}\right)\right)$. Then if we denote $CV^*_\alpha(x^j_{\ell_0}, z^j_{\ell_0})$ for $\inf_{(\theta,\xi)\in\mathbb{R}^d\times\mathbb{R}} V(\xi, \theta, (x^j_{\ell_0}, z^j_{\ell_0}))$, $j = 1, \cdots, N_{\ell_0}$, the quantization based quadrature formula to approximate (39) is given by

$$\sum_{j=1}^{N_{\ell_0}} CV^*_\alpha(x^j_{\ell_0}, z^j_{\ell_0})\mathbb{P}\left((X_{\ell_0}, Z_{\ell_0}) \in C_j(x_{\ell_0}, z_{\ell_0})\right), \tag{40}$$

where $(C_j(x_{\ell_0}, z_{\ell_0}))_{j=1,\cdots,N_{\ell_0}}$ is a Voronoi tesselation of the N_{ℓ_0}-quantizer Γ_{ℓ_0}. Fore more details about optimal vector quantization, including error bounds for cubature formulae, we refer to [22].

Consequently, we need to compute the solution as well as the value of the objective function for all nodes of a quantification 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 (38).

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 those distributions or functions.

3.2 Computing CVaR hedging by stochastic approximation: a first approach

The above local representation (38) naturally yields a stochastic gradient algorithm derived from the Lyapunov function V which will converge toward $(\xi^*_\alpha, \theta^*_\alpha) \in \operatorname{Arg\,min} V$. Then, following the procedure investigated in [3], a companion recursive procedure can be easily devised which has $CV^*_\alpha := CV^*_\alpha(x, z) = \mathrm{CVaR}_\alpha\left(F(U) - \theta^*_\alpha.G(U)\right)$ as a target, *i.e.*, the CVaR_α of the CVaR-hedged portfolio at the point (x, z). Finally, in order to compute the value function at time t_ℓ *i.e.* the global expectation (39), we will rely on the cubature formula based on optimal vector quantization given by (40).

First, we set

$$H_1(\xi, \theta, U) := \frac{\partial v}{\partial \xi}(\xi, \theta, U) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{F(U)-\theta.G(U)\ge\xi\}}, \tag{41}$$

$$H_{2:d+1}(\xi, \theta, U) := \frac{\partial v}{\partial \theta}(\xi, \theta, U) = -\frac{1}{1-\alpha}G(U)\mathbf{1}_{\{F(U)-\theta.G(U)\ge\xi\}}, \tag{42}$$

so that,

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

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

$$\xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, \theta_{n-1}, U_n), \tag{43}$$

$$\theta_n = \theta_{n-1} - \gamma_n H_{2:d+1}(\xi_{n-1}, \theta_{n-1}, U_n), \tag{44}$$

where $(U_n)_{n\ge1}$ 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 $(\gamma_n)_{n\ge1}$ is a positive deterministic step sequence satisfying

$$\sum_{n\geq 1}\gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1}\gamma_n^2 < +\infty. \tag{45}$$

Following [3], as a second step, in order to estimate $\mathrm{CVaR}_\alpha\left(F(U) - \theta_\alpha^*.G(U)\right)$, *i.e.* the CVaR_α of the local CVaR hedged loss, we devise a *companion* procedure using the same step sequence than (43) and (44), for $n \geq 1$

$$C_n = C_{n-1} - \gamma_n H_{d+2}\left(\xi_{n-1}, \theta_{n-1}, C_{n-1}, U_n\right), \tag{46}$$

with $H_{d+2}\left(\xi, \theta, c, u\right) := c - v(\xi, \theta, u)$. In order to derive the *a.s.* convergence of (43), (44) and (46), we introduce the following additional assumption on the distribution of $F(U)$ and $G(U)$.

Assumption 6. *Let $a > 0$. $F(U) \in \mathbb{L}^{2a}(\mathbb{P})$ 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 Robbins-Monro Theorem (see e.g. [7]). In fact we will use the following slight extension (which takes into acount the case of non-uniqueness of the target). For a proof, we refer e.g. to [20].

Theorem 3.1. *(Extended 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

$$\forall y \in \mathbb{R}^d \setminus \mathcal{T}^*, \forall y^* \in \mathcal{T}^*, \quad \langle y - y^*, h(y)\rangle > 0. \tag{47}$$

Let $(\gamma_n)_{n\geq 1}$ be a deterministic step sequence satisfying (45). Suppose that

$$\forall y \in \mathbb{R}^d, \quad \mathbb{E}[|H(y, X)|^2] \leq C(1 + |y|^2) \tag{48}$$

(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:

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

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

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 $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$.

Theorem 3.2. *Suppose that Assumptions 4 and 6 are satisfied (for $a = 1$), and that the step sequence $(\gamma_n)_{n\geq 1}$ satisfies the usual decreasing step assumption (45).*
Then the recursive procedure defined by (43), (44) and (46) satisfies:

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

such that

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

Moreover,

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

Proof. We first prove the *a.s.* convergence of $(\xi_n, \theta_n)_{n\geq 1}$ using the above Extended Robbins-Monro 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 [3] (Section 2.2). In order to apply the R.M. 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 (43) and (44) reads

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

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

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

since the function V is a convex differentiable function and $\operatorname{Arg\,min} V$ is non empty.

- *Linear Growth of* $(\xi, \theta) \mapsto \mathbb{E}\left[\left|H_1\left(\xi, \theta, U\right)\right|^2 + \left|H_{2:d+1}\left(\xi, \theta, U\right)\right|^2\right]$. This conditions is clearly fulfilled since there exists a real constant $C > 0$ such that

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

so that,

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

Consequently, we have

$$(\xi_n, \theta_n) \stackrel{a.s.}{\longrightarrow} (\xi^*_\alpha, \theta^*_\alpha).$$

In order to prove the *a.s.* convergence of $(C_n)_{n\geq 1}$ toward CV^*_α, we set for convenience $\gamma_0 := 1 + \sup_{n\geq 1} \gamma_n$. Then, one defines recursively a sequence $(\Delta_n)_{n\geq 1}$ by

$$\Delta_{n+1} = \Delta_n \frac{\gamma_{n+1}}{\gamma_n} \frac{\gamma_0}{\gamma_0 - \gamma_{n+1}}, \; n \geq 0, \; \Delta_0 = 1.$$

Elementary computations show by induction that

$$\gamma_n = \gamma_0 \frac{\Delta_n}{S_n}, \; n \geq 0, \text{ with } S_n := \sum_{k=0}^{n} \Delta_k. \tag{49}$$

Furthermore, it follows from (49) that for every $n \geq 1$, $\log(S_n) \geq \frac{1}{\gamma_0}\sum_{k=1}^n \gamma_k$, which implies that $\lim_n S_n = +\infty$.

Now using (46) and (49), one gets for every $n \geq 1$

$$S_n C_n = S_{n-1} C_{n-1} + \Delta_n \left(\Delta N_n + V(\xi_{n-1}, \theta_{n-1})\right)$$

where $\Delta N_n := v(\xi_{n-1}, \theta_{n-1}, U_n) - V(\xi_{n-1}, \theta_{n-1})$, $n \geq 1$, defines a sequence of martingale increments with respect to the natural filtration of the algorithm $\mathcal{F}_n := \sigma(\xi_0, \theta_0, U_1, \cdots, U_n)$, $n \geq 0$. This implies that

$$C_n = \frac{1}{S_n}\left(\sum_{k=0}^{n-1} \Delta_{k+1} \Delta N_{k+1}\right) + \frac{1}{S_n}\left(\sum_{k=0}^{n-1} \Delta_{k+1} V(\xi_k, \theta_k)\right).$$

The second term in the right hand side of the above equality converges to $V(\xi_\alpha^*, \theta_\alpha^*) = C_\alpha^*$ owing to the continuity of V at $(\xi_\alpha^*, \theta_\alpha^*)$ and Cesàro's Lemma. The convergence to 0 of the first term will follow from the *a.s.* convergence of the series

$$N_n^\gamma := \sum_{k=1}^n \gamma_k \Delta N_k, \quad n \geq 1,$$

by the Kronecker Lemma since $\gamma_n = \gamma_0 \frac{\Delta_n}{S_n}$. The sequence $(N_n^\gamma)_{n\geq 1}$ is an $\mathcal{F}_n$-martingale since the ΔN_k's are martingale increments and

$$\mathbb{E}\left[(\Delta N_n)^2 | \mathcal{F}_{n-1}\right] \leq \frac{1}{(1-\alpha)^2}\mathbb{E}\left[(F(U) - \theta.G(U) - \xi)_+^2\right]_{|\xi=\xi_{n-1}, \theta=\theta_{n-1}}.$$

Assumption 6 and the *a.s.* convergence of (ξ_n, θ_n) toward $(\xi_\alpha^*, \theta_\alpha^*)$ imply that

$$\sup_{n\geq 1} \mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < +\infty \quad a.s.$$

Consequently, the step assumption (45) implies $\langle N^\gamma \rangle_\infty = \sum_{n\geq 1} \gamma_n^2 \mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < \infty$, which in term yields the *a.s.* convergence of $(N_n^\gamma)_{n\geq 1}$, so that $C_n \xrightarrow{a.s.} CV_\alpha^*$. □

As concerns the rate of convergence, the global procedure composed by (43), (44), (46) is a regular stochastic algorithm that behaves as described in usual Stochastic Approximation textbooks like [5], [6], [19]. As soon as $\mathcal{T}^*$ is reduced to a single point $(\xi_\alpha^*, \theta_\alpha^*)$ (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 depending on $(\xi_\alpha^*, \theta_\alpha^*)$ (which is unknown to the user). This always induces a more or less (sub-optimal) 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. [23]). 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

$$\phi_n = \phi_{n-1} - \gamma_n H(\phi_{n-1}, U_n), \tag{50}$$

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

$$\overline{\phi}_n = \frac{\phi_0 + \cdots + \phi_{n-1}}{n}, \quad n \geq 1, \tag{51}$$

where ϕ_n is defined by (50), *a.s.* converges to the same target. The Ruppert & Polyak's Averaging Principle says that an appropriate choice of the step yields for free the optimal asymptotic rate and the smallest possible asymptotic variance. We recall below this result following a version established in [24].

Theorem 3.3. *(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 $\mathcal{F}_n := \sigma(\xi_0, \theta_0, U_1, \cdots, U_n)$ 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*

$$(RP) \equiv \exists\ C > 0,\ \textit{such that a.s.}\ \begin{cases} (i)\ \mathbb{E}[\epsilon_{n+1}|\mathcal{F}_n]\mathbf{1}_{\{|\phi_n-\phi^*|\leq C\}} = 0, \\ (ii)\ \exists\ b > 2,\ \sup_n \mathbb{E}[|\epsilon_{n+1}|^b|\mathcal{F}_n]\ \mathbf{1}_{\{|\phi_n-\phi^*|\leq C\}} < +\infty, \\ (iii)\ \exists\ \Gamma \in \mathcal{S}^+(d,\mathbb{R})\ \textit{such that}\ \mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right] \xrightarrow{a.s.} \Gamma. \end{cases} \tag{52}$$

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 $\{\phi_n \to \phi^*\}$*:*

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

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 $p^{(X_{\ell_0},Z_{\ell_0})=(x_{\ell_0},z_{\ell_0})}_{(X_M - X_{\ell_0},L)}$ 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 the 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 7.

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

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

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

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

Example 1. We take the example (that will be studied in Section 5) of the energy provider which 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 geometrice Brownian motion correlated with $\rho \neq 0$ 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

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

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

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

Consequently, $p_{X,L}$ satisfies (i) and (iv). Moreover, for all $x > -S_0$, $x \neq 0$ and y, $\theta \in \mathbb{R}$,

$$p_{X,L}(x, \theta.x + y) \leq \frac{1}{2\pi\rho\sigma_g\sigma_C\sqrt{T}} \frac{1}{(x+S_0)|x|} \in L^1(\mathrm{d}x).$$

so that, (ii) is satisfied. Now, if $K \subset \mathbb{R}$ is a compact set, for all $\xi \in \mathbb{R}$, there exists a constant $A > 0$ such that for all $x > -S_0$, $x \neq 0$

$$(1+|x|^2)p_{X,L}(x, \theta.x + \xi) \leq \frac{1}{2\pi\rho\sigma_g\sigma_C\sqrt{T}} \frac{1+|x|^2}{(x+S_0)|x|} e^{-A\frac{1}{x^2}} \in L^1(\mathrm{d}x),$$

so that (iii) is satisfied.

In next theorem, we use notations of Theorem 3.3. We establish a CLT for the empirical mean sequence $\bar{\phi_n}$ defined by (51).

Theorem 3.4. *(Convergence rate of the procedure) Suppose that Assumptions 4, 6 (with $a > 1$) and 7 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 (51) satisfies*

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

where the asymptotic covariance matrix Σ is given by

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

with

$$P := \frac{1}{1-\alpha}\begin{pmatrix} \int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x & \left(\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^T & 0 \\ \int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x & \int_{\mathbb{R}^d} xx^T p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x & 0 \\ 0 & 0 & 1-\alpha \end{pmatrix}, \tag{53}$$

and

$$\Gamma := \begin{pmatrix} \frac{\alpha}{1-\alpha} & 0 & \frac{\alpha\mathbb{E}\left[(L-\theta_\alpha^*.X-\xi_\alpha^*)_+\right]}{(1-\alpha)^2} \\ 0 & \frac{\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta_\alpha^*.X\geq\xi_\alpha^*\}}\right]}{(1-\alpha)^2} & \frac{\mathbb{E}\left[X(L-\theta_\alpha^*.X-\xi_\alpha^*)_+\right]}{(1-\alpha)^2} \\ \frac{\alpha\mathbb{E}\left[(L-\theta_\alpha^*.X-\xi^*)_+\right]}{(1-\alpha)^2} & \frac{\mathbb{E}\left[X(L-\theta_\alpha^*.X-\xi^*)_+\right]^T}{(1-\alpha)^2} & \frac{\mathrm{Var}\left((L-\theta_\alpha^*.X-\xi^*)_+\right)}{(1-\alpha)^2} \end{pmatrix}. \tag{54}$$

Proof. First note that the procedure (50) 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 $h(\phi) := \mathbb{E}\left[H(\phi, U)\right] = \nabla V(\phi, x, z)$, and ϵ_n, $n \geq 1$, denotes the $\mathcal{F}_n$-adapted martingale increment sequence defined by:

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

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

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

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

$$\begin{aligned} h_1(\phi) &= 1 - \frac{1}{1-\alpha}\int_{\mathbb{R}^d\times\mathbb{R}} p_{X,L}(x,y)\mathbf{1}_{\{y\geq\xi+\theta.x\}}\mathrm{d}x\mathrm{d}y, \\ h_i(\phi) &= -\frac{1}{1-\alpha}\int_{\mathbb{R}^d\times\mathbb{R}} x_i p_{X,L}(x,y)\mathbf{1}_{\{y\geq\xi+\theta.x\}}\mathrm{d}x\mathrm{d}y, \quad i = 2,\cdots,d+1, \\ h_{d+2}(\phi) &= C - \left(\xi + \frac{1}{1-\alpha}\mathbb{E}\left[(L-\theta.X-\xi)_+\right]\right). \end{aligned}$$

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

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

Owing to Assumption 7, one can interchange integral and derivation. In order to differentiate $h_{2:d+1}$, first note that, by Fubini's Theorem,

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

so that owing to Assumption 7 and Lebesgue's differentiation Theorem, one can interchange integral and derivation. Consequently, the functions h_1 and $h_{2:d+1}$ are differentiable at $\phi^* := (\xi^*_\alpha, \theta^*_\alpha, C^*_\alpha)$ and for $i = 2,\cdots,d+1$,

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

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

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

so that M is given by (53). Let $u = (u_1, u_2, u3) \in \mathbb{R}\times\mathbb{R}^d\times\mathbb{R}$,

$$\begin{aligned} u^TMu = \frac{\int_{\mathbb{R}^d} p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}{1-\alpha}\Bigg(u_1^2 + 2u_1\frac{\int_{\mathbb{R}^d} x p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}\mathrm{d}x u_2 \\ + u_2^T\frac{\int_{\mathbb{R}^d} xx^T p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}u_2 + u_3^2\Bigg), \end{aligned}$$

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

$$\begin{aligned} u^TMu \geq \frac{\int_{\mathbb{R}^d} p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}{1-\alpha}\Bigg(u_2^T\int_{\mathbb{R}^d}\left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}\right) \\ \times\left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x}\right)^T p_{X,L}(x,\xi^*_\alpha+\theta^*_\alpha.x)\mathrm{d}x u_2 + u_3^2\Bigg), \\ > 0. \end{aligned}$$

Consequently, the matrix P is a uniformly repulsive matrix. To apply Theorem 3.3, we need to check assumptions (RP) of (52). Let $C > 0$. First note that

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

Thanks to Assumption 4 (with $a > 1$), there exists $A > 0$, such that for $i = 2, \cdots, d+1$,

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

and

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

Consequently, (ii) of (52) holds true with $b = 2a$ since

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

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

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

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

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. □

One may be interested by the asymptotic variance of each components 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

$$P^{-1} := \frac{1-\alpha}{\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x}\begin{pmatrix} 1 + V^T\Pi^{-1}V & -V^T\Pi^{-1} & 0 \\ -\Pi^{-1}V & \Pi^{-1} & 0 \\ 0 & 0 & \frac{1}{1-\alpha}\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x \end{pmatrix}, \tag{55}$$

where $\Pi := \int_{\mathbb{R}^d}\left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x,\xi_\alpha^*+\theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi_\alpha^*+\theta_\alpha^*.x)\mathrm{d}x}\right)\left(x - \frac{\int_{\mathbb{R}^d} x p_{X,L}(x,\xi_\alpha^*+\theta_\alpha^*.x)}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi_\alpha^*+\theta_\alpha^*.x)\mathrm{d}x}\right)^T p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x$, and $V := \frac{1}{\int_{\mathbb{R}^d} p_{X,L}(x,\xi_\alpha^*+\theta_\alpha^*.x)\mathrm{d}x}\int_{\mathbb{R}^d} x p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x$, so that, for $i = 2, \cdots, d+1$,

$$\begin{aligned}
\Sigma_{1,1} &= \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^2}\left(\left(1 + V^T\Pi^{-1}V\right)^2\alpha(1-\alpha)\right. \\
&\quad \left. + \left(\Pi^{-1}V\right)^T\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta_\alpha^*.X\ge\xi_\alpha^*\}}\right]\Pi^{-1}V\right),
\end{aligned} \tag{56}$$

$$\Sigma_{i,i} = \frac{1}{\left(\int_{\mathbb{R}^d} p_{X,L}(x, \xi_\alpha^* + \theta_\alpha^*.x)\mathrm{d}x\right)^2}\left(m_i^2\alpha(1-\alpha) + \tilde{m}_{.i}^T\mathbb{E}\left[XX^T\mathbf{1}_{\{L-\theta_\alpha^*.X\ge\xi_\alpha^*\}}\right]\tilde{m}_{.i}\right), \tag{57}$$

$$\Sigma_{d+2,d+2} = \left(\frac{1}{1-\alpha}\right)^2\mathrm{Var}\left((L - \theta_\alpha^*.X - \xi_\alpha^*)_+\right), \tag{58}$$

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

3.3 Dynamic CVaR hedging

In this section, we propose several methods to compute the optimal strategy of (1), the VaR and the CVaR of the CVaR-hedged portfolio. From a modeling point of view, under Assumption 2, we suppose for every $\ell = 1, \cdots, M$ that there exists two continuous functions $G_\ell : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_\ell} \to \mathbb{R}^d$, $F_\ell : \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^{r_\ell} \to \mathbb{R}$ such that

$$X_\ell - X_{\ell-1} = G_\ell(X_{\ell-1}, Z_{\ell-1}, U_\ell), \quad \text{and} \quad L = F_\ell\left(X_{\ell-1}, Z_{\ell-1}, U_\ell\right),$$

where U_ℓ is a r_ℓ-dimensional random variable independent of $\mathcal{G}_{l-1}$.

Crude CVaR hedging Algorithm (C.H.)
The direct approach to solve (1) is to proceed as in the static framework and to devise a global stochastic gradient algorithm. To be more precise, at every time t_k, we consider an optimal N_k-quantization $\left(\hat{X}_k, \hat{Z}_k\right)$, $k = 1, \cdots, M-1$, based on an optimal quantization N_k grid $\Gamma_k = \left(\left(x_k^1, z_k^1\right), \cdots, \left(x_k^{N_k}, z_k^{N_k}\right)\right)$ of the state process at time t_k.

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 of the whole process (X, Z). For this method, we make the approximation which consists of making θ_k^* depending only of the state process at time k, (X_k, Z_k). Thus, we only need to estimate $\theta_k^{j,*}$ at each time t_k, for all nodes $\left(x_k^j, z_k^j\right)$, $j = 1, \cdots, N_k$ on the corresponding grid. This can be done by the following stochastic algorithm, namely,

$$\xi_n = \xi_{n-1} - \gamma_n H_1\left(\xi_{n-1}, \theta_{n-1}, U_n\right), \tag{59}$$

$$\theta_{0,n} = \theta_{0,n-1} - \gamma_n H_{2,0}\left(\xi_{n-1}, \theta_{n-1}, U_n\right), \tag{60}$$

$$\theta_{\ell,n}^j = \theta_{\ell,n-1}^j - \gamma_n H_{2,\ell}^j\left(\xi_{n-1}, \theta_{n-1}, U_n\right), \ j = 1, \cdots, N_\ell, \ k = 1, \cdots, M-1, \tag{61}$$

$$C_n = C_{n-1} - \gamma_n H_3\left(\xi_{n-1}, \theta_{n-1}, C_{n-1}, U_n\right), \tag{62}$$

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

$$H_1(\xi, \theta, u) = 1 - \frac{1}{1-\alpha} \mathbf{1}_{\left\{\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i \geq \xi\right\}},$$

$$H_{2,0}\left(\xi, \theta, u\right) = -\frac{G_1(X_0, Z_0, U_1)}{1-\alpha} \mathbf{1}_{\left\{\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i \geq \xi\right\}},$$

$$H_{2,\ell}^j\left(\xi, \theta, u\right) = -\frac{G_{\ell+1}(X_\ell, Z_\ell, U_{\ell+1})}{1-\alpha} \mathbf{1}_{\left\{\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i \geq \xi\right\}} \mathbf{1}_{\{(X_\ell, Z_\ell) \in C_j(x_\ell, z_\ell)\}},$$

$$H_3\left(\xi, \theta, C, u\right) = C - \xi - \frac{1}{1-\alpha}\left(\sum_{i=1}^M \Delta L_i - \theta_{i-1}.\Delta X_i - \xi\right)_+ .$$

The sequence $(\xi_n, \theta_n, C_n)_{n\geq 1}$ *a.s.* converges toward its target $(\xi_\alpha^*, \theta_\alpha^*, C_\alpha^*)$. 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 $D := 2 + d + \sum_{\ell=2}^M d \times N_\ell$.

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 (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.* $N = N_\ell = 5$, $\ell = 1, \cdots, M$, 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 $T_\ell = t_\ell$ that delivers 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 three months, the next two quarters and the next three 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 energy market, the dimension of the considered RM algorithm becomes a real issue.

From a numerical point of vue, we observe that in a high dimensional framework the algorithm "freezes" and "suffers", say as soon as the dimension is greater than 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 pratical implementation. To overcome this problem, we propose several approximate solutions to solve (1) which crucially relies on Assumptions 1, 2 and 5. These solutions have the major advantage to dramatically reduce the dimension of the above algorithm.

Backward dynamic hedging strategy (B.H.)
This strategy is based on (9) and consists in a backward resolution. To be more precise, if we consider M trading dates, then (9) and Assumption 1 imply that in order to hedge the risk at the last trading date t_{M-1}, we have to solve

$$\inf_{\theta\in\mathcal{A}} \mathbb{E}\left[\mathcal{G}_{M-1}\text{-CVaR}_\alpha\left(L - \sum_{\ell=1}^{M} \theta_{\ell-1}.\Delta X_\ell\right)\right] = \inf_{\theta_{M-1}\in L^0(\mathcal{G}_{t_{M-1}})} \mathbb{E}\left[\mathcal{G}_{M-1}\text{-CVaR}_\alpha\left(L - \theta_{M-1}.\Delta X_M\right)\right].$$

The optimization problem that appears in the right-hand side of the above equalitie can be easily solved using the static algorithm developed in Section 3. Now that we have the solution θ^b_{M-1} of this problem, we can go one step backward and solve the new problem

$$\inf_{\theta_{M-2}\in L^0(\mathcal{G}_{t_{M-2}})} \mathbb{E}\left[\mathcal{G}_{M-2}\text{-CVaR}_\alpha\left(L - \theta^b_{M-1}.\Delta X_M - \theta_{M-2}\Delta X_{M-1}\right)\right],$$

using again the algorithm developed in the static framework in order to obtain θ^b_{M-2}. 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. However we observe on numerical experiments that the resulting static CVaR related to this self-financed strategy θ^b, namely

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

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

$$\inf_{\theta_{M-k-1}\in L^0(\mathcal{G}_{M-k-1})} \mathbb{E}\left[\mathcal{G}_{M-k-1}\text{-CVaR}_\alpha\left(L - \sum_{\ell=M-k+1}^{M} \theta^b_{\ell-1}.\Delta X_\ell - \theta_{M-k-1}\Delta X_{M-k-1}\right)\right],$$

there is an error (compared to the original problem (1)) on the estimate $\theta^b_{M-\ell-1} \neq \theta^*_{M-\ell-1}$ 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.

Dynamic hedging strategy based on a martingale decomposition of L (M.D.H.)
This method is based on the sub-additivity of the CVaR and on the following decomposition of the loss L into a sum of $\mathbb{G}$-martingale increments, namely

$$L = \mathbb{E}\left[L\right] + \sum_{\ell=1}^{M} \tilde{\Delta} L_\ell, \tag{63}$$

where $\tilde{\Delta} L_\ell = \mathbb{E}\left[L \middle| \mathcal{G}_\ell\right] - \mathbb{E}\left[L \middle| \mathcal{G}_{\ell-1}\right]$, $1 \leq \ell \leq M$. Now, using the sub-additivity of the CVaR, we obtain

$$\inf_{\theta \in \mathcal{A}} \mathrm{CVaR}_\alpha \left(L - \sum_{\ell=1}^{M} \theta_{\ell-1} \Delta X_\ell \right) \leq \mathbb{E}\left[L\right] + \sum_{\ell=1}^{M} \inf_{\theta_{\ell-1} \in L^0(\mathcal{G}_{\ell-1})} \mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_{\ell-1} \Delta X_\ell \right). \tag{64}$$

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 $D \leq 150$), we observe that the CVaR obtained using this strategy is almost equal to the optimal one. When the dimension D becomes large, which is generally due to a large number of trading dates, 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 aproach. We use the inequality

$$\inf_{\theta_{\ell-1} \in L^0(\mathcal{G}_{\ell-1})} \mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-}\mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_{\ell-1} \Delta X_\ell \right) \right] \leq \inf_{\theta_{\ell-1} \in L^0(\mathcal{G}_{\ell-1})} \mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_{\ell-1} \Delta X_\ell \right), \tag{65}$$

and switch to the new optimization problem

$$\sum_{\ell=1}^{M} \inf_{\theta_{\ell-1} \in L^0(\mathcal{G}_{\ell-1})} \mathbb{E}\left[\mathcal{G}_{\ell-1}\text{-}\mathrm{CVaR}_\alpha \left(\tilde{\Delta} L_\ell - \theta_{\ell-1} \Delta X_\ell \right) \right]. \tag{66}$$

The main difference in solving the local problem in the left hand-side of (65) 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 (66), numerical experiments led us to the conclusion that this last algorithm behaves better than the one obtained by solving the right-hand side of (65) at each time step. To be more precise, the original CVaR estimated by the strategy obtained by (66) is lower than the one obtained by the strategy solution of the right-hand side in (64).

In order to solve (66), we use optimal vector quantization again to approximate the unknown random variable $\tilde{\Delta} L_\ell$, *i.e.*, we approximate $\mathbb{E}\left[L | \mathcal{G}_\ell\right]$ by using the cubature formula

$$\mathbb{E}\left[F_{\ell+1}(X_\ell, Z_\ell, U_{\ell+1}) \middle| (X_\ell, Z_\ell) \right] \approx \varphi(X_\ell, Z_\ell) = \sum_{j=1}^{N_\ell} F(X_\ell, Z_\ell, u^j_{\ell+1}) \mathbb{P}\left(U_{\ell+1} \in C_j(u_{\ell+1}) \right), \tag{67}$$

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

$$L = L_0 + \sum_{\ell=1}^{M} \Delta L_\ell \tag{68}$$

instead of (63). However, it is quite natural to approximate from the sequence of martingale increments ΔX_ℓ, $\ell = 1, \cdots, M$, another martingale sequence so that the decomposition (63) is more appropriate to our framework than the decomposition (68). 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 they are only updated on rare events since it tries to measure 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 [15], two variance reduction techniques have been developed in order to reduce the asymptotic variance in the CLT (53). The first one is based on the unconstrained importance sampling (IS) stochastic algorithm originally developed in [20] and then applied to both VaR and CVaR in [3]. 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 importance sampling (by translation) applied to stochastic approximation algorithm like (50) is to use the invariance of the Lebesgue measure by translation to show that for every $\mu \in (\mathbb{R}^r)^{d+2}$,

$$\mathbb{E}\left[H_i(\xi, \theta, U)\right] = \mathbb{E}\left[H_i(\xi, \theta, U + \mu_i)\frac{p(X + \mu_i)}{p(X)}\right], \quad i = 1, \cdots, d+1 \tag{69}$$

$$\mathbb{E}\left[H_{d+2}(\xi, \theta, c, U)\right] = \mathbb{E}\left[H_{d+2}(\xi, \theta, c, U + \mu_{d+2})\frac{p(X + \mu_{d+2})}{p(X)}\right]. \tag{70}$$

It is easy to obtain a new CLT using these new random variables with the same expectations following the lines of Theorem 3.4. 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 $\mu_i \in \mathbb{R}^r$, $i = 1, \cdots, d+1$ (idem with $i = d+2$). This yields a new minimization problem that can be solved by a Stochastic Approximation procedure. As $(\xi^*_\alpha, \theta^*_\alpha, C^*_\alpha)$ are unknown, we combine these new procedures in an adaptive way with the algorithm (50) in its averaged form following the ideas developped in [3] for the recursive computation of VaR and CVaR. To minimize the three variances $\mathrm{Var}\left(\mathbf{1}_{\{L - \theta^*_\alpha.X \geq \xi^*_\alpha\}}\right)$, $\mathbb{E}\left[X_i^2 \mathbf{1}_{\{L - \theta^*_\alpha.X \geq \xi^*_\alpha\}}\right]$ and $\mathrm{Var}\left((L - \theta^*_\alpha.X - \xi^*_\alpha)_+\right)$ respectively using (69) and (70), we are lead to minimize the functions $Q_i(., \xi^*_\alpha, \theta^*_\alpha)$ defined for $\mu = (\mu_1, \cdots, \mu_{d+2}) \in (\mathbb{R}^r)^{d+2}$ by,

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

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

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

where for 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 [20]), 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

$$\nabla Q_1(\mu,\xi_\alpha^*,\theta_\alpha^*)=\mathbb{E}\left[\mathbf{1}_{\left\{L^{(-\mu)}-\theta_\alpha^*.X^{(-\mu)}\geq\xi_\alpha^*\right\}}W(\mu,U)\right], \tag{71}$$

$$\nabla Q_i(\mu,\xi_\alpha^*,\theta_\alpha^*)=\mathbb{E}\left[\left(X_i^{(-\mu)}\right)^2\mathbf{1}_{\left\{L^{(-\mu)}-\theta_\alpha^*.X^{(-\mu)}\geq\xi_\alpha^*\right\}}W(\mu,U)\right], \tag{72}$$

$$\nabla Q_{d+2}(\mu,\xi_\alpha^*,\theta_\alpha^*)=\mathbb{E}\left[\left(L^{(-\mu)}-\theta_\alpha^*.X^{(-\mu)}-\xi_\alpha^*\right)_+^2W(\mu,U)\right], \tag{73}$$

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 $\arg\min Q_i(.,\xi_\alpha^*,\theta_\alpha^*)=\{\mu\in\mathbb{R}^r\mid\nabla_\mu Q_i(\mu,\xi_\alpha^*,\theta_\alpha^*)=0\}$ is non empty.

Equations (71), (72) and (73) 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

$$|W(\mu,u)|\leq e^{2\rho|\mu|^b}(A|u|^{b-1}+A|\mu|^{b-1}+B), \tag{74}$$

for some real constants ρ, A and B (for more details we refer to [20] and [15]). 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

$$\begin{cases}\forall u\in\mathbb{R}^r,|F(u)|\leq\tilde{F}(u)\quad\text{and}\quad\tilde{F}(u+v)\leq C(1+\tilde{F}(u))^c(1+\tilde{F}(v))^c,\\ \forall u\in\mathbb{R}^r,|G(u)|\leq\tilde{G}(u)\quad\text{and}\quad\tilde{G}(u+v)\leq C(1+\tilde{G}(u))^c(1+\tilde{G}(v))^c,\\ \qquad\mathbb{E}\left[|U|^{2(b-1)}\left(\tilde{F}(U)^{4c}+\tilde{G}(U)^{4c}\right)+\tilde{G}(U)^{4c}\right]<+\infty,\end{cases} \tag{75}$$

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

$$K_1(\mu_1,\xi_\alpha^*,\theta_\alpha^*,U)=e^{-2\rho|\mu_1|^b}\mathbf{1}_{\left\{L^{(-\mu_1)}-\theta_\alpha^*.X^{(-\mu_1)}\geq\xi_\alpha^*\right\}}W(\mu_1,U), \tag{76}$$

$$K_i(\mu_i,\xi_\alpha^*,\theta_\alpha^*,U)=\frac{e^{-2\rho|\mu_i|^b}}{1+\tilde{G}(-\mu_i)^{2c}}\left(X_i^{(-\mu_i)}\right)^2\mathbf{1}_{\left\{L^{(-\mu_i)}-\theta_\alpha^*.X^{(-\mu_i)}\geq\xi_\alpha^*\right\}}W(\mu_i,U),\ i=2,\cdots,d+1 \tag{77}$$

$$K_{d+2}(\mu_{d+2},\xi_\alpha^*,\theta_\alpha^*,U)=\frac{e^{-2\rho|\mu_{d+2}|^b}}{1+\tilde{F}(-\mu_{d+2})^{2c}+|\theta_\alpha^*|^{2c}\tilde{G}(-\mu_{d+2})^{2c}}\left(L^{(-\mu_{d+2})}-\theta_\alpha^*.X^{(-\mu_{d+2})}-\xi_\alpha^*\right)_+^2W(\mu_{d+2},U), \tag{78}$$

so that it satisfies the linear growth assumption (48) of the RM Theorem and for $i=1,\cdots,d+2$

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

Moreover, since Q_i is convex $\nabla_{\mu_i}Q_i$ satisfies (47). Now the RM algorithms defined for $n\geq1$ by

$$\mu_{i,n}=\mu_{i,n-1}-\gamma_nK_i(\mu_{i,n-1},\xi_\alpha^*,\theta_\alpha^*,U_n),\ \mu_{i,0}\in\mathbb{R}^r,$$

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 [20] and [3]). Now, 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 $n \geq 0$,

$$\phi_n = (\xi_n, \theta_n, C_n, \mu_{1,n}, \cdots, \mu_{d+2,n}), \tag{79}$$

where ξ_n, θ_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} denotes the variance reducer for the CVaR_α. We update this state variable recursively by

$$\phi_n = \phi_{n-1} - \gamma_n L(\phi_{n-1}, U_n), \ n \geq 1, \tag{80}$$

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

$$L_1(\xi, \theta, \mu_1, u)) = e^{-\rho|\mu_1|^b}\left(1 - \frac{1}{1-\alpha}\mathbf{1}_{\{L^{(+\mu_1)} - \theta.X^{(+\mu_1)} \geq \xi\}}\frac{p(u+\mu_1)}{p(u)}\right), \tag{81}$$

$$L_i(\xi, \theta, \mu_i, u)) = \frac{e^{-\rho|\mu_i|^b}}{\left(1 + \tilde{G}^{2c}(-\mu_i)\right)^{1/2}} X^{(+\mu_i)}\mathbf{1}_{\{L^{(+\mu_i)} - \theta.X^{(+\mu_i)} \geq \xi\}}\frac{p(u+\mu_i)}{p(u)}, \tag{82}$$

$$L_{d+2}(\xi, \theta, C, \mu_{d+2}, u) = C - \xi - \frac{1}{1-\alpha}\left(L^{(+\mu_{d+2})} - \theta.X^{(+\mu_{d+2})} - \xi\right)_+ \frac{p(u+\mu_{d+2})}{p(u)}, \tag{83}$$

$$L_{d+2+j}(\xi, \theta, \mu_j, u) = K_j(\mu_j, \xi, \theta, u), \quad j = 1, \cdots, d+2, \tag{84}$$

In [15] it is shown that the algorithm (80) 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 variances.

The second variance reduction tool is based on Linear Control Variate. We use a control variable based on X, since under Assumption 1 we have $\mathbb{E}[X] = 0$. For more details, we refer to [15] where we only develop and study those 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 example, we use RM algorithm with two phases (see Remark 4.1) combined with the Ruppert & Polyak's averaging principle. In all examples, we define the step sequence by $\gamma_n = \frac{1}{n^p}$, with $p = \frac{3}{4}$.

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

$$L = \left(S^e_T - h_R S^g_T - C\right)_+,$$

	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

Table 1: One step Self-financed static CVaR-hedging of Spark Spread option

where the time horizon $T = 1$ (year), the heat rate $h_R = 4$BTU/kWh (BTU: British Thermal Unit), the generation costs $C = 3$\$/MWh, the two volatilities $\sigma_g = 0.4$, $\sigma_e = 0.8$ and the electricity and gas initial spot prices are $S_0^e = 40$\$/MWh, $S_0^g = 3$\$/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 (3) with $\ell = 0$. A crude Monte Carlo gives $\mathbb{E}[L] = 11.86$ with a variance of 3692 after 3 000 000 trials. The variance ratios correspond to an estimate of the asymptotic variance obtained without any variance reduction techniques, *i.e.* (56), (57) and (58) 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 [15]).

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

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

where q_δ^e is the quantile of S_T^e at level δ. We choose: $\delta = 0.995$ ($q_\delta^e \approx 228.04$) for $\alpha = 0.95$, $\delta = 0.999$ ($q_\delta^e \approx 344.15$) for $\alpha = 0.99$ and $\delta = 0.9995$ ($q_\delta^e \approx 403.95$) 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 Robbins-Monro procedure are specified in Table 1. We provide the VaR and CVaR of the loss without any hedging strategy which are computed using the Robbins-Monro procedure developed in [3].

To complete this numerical example, we provide the histograms of the loss obtained with and without hedging. We clearly see on Figure 1 that the asymetry of the histogram has been changed from right(loss) to left (gain) so that gains are more likely to occured 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 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.

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

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

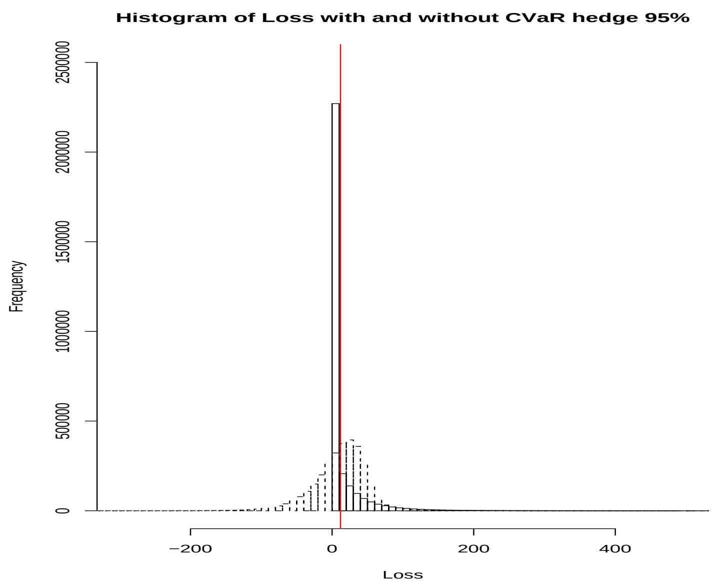

Figure 1: Histogram of loss with (dashed lines) and without (normal lines) one step CVaR-hedging at level $\alpha = 95\%$. The vertical line is the mean of the portfolio loss distribution.

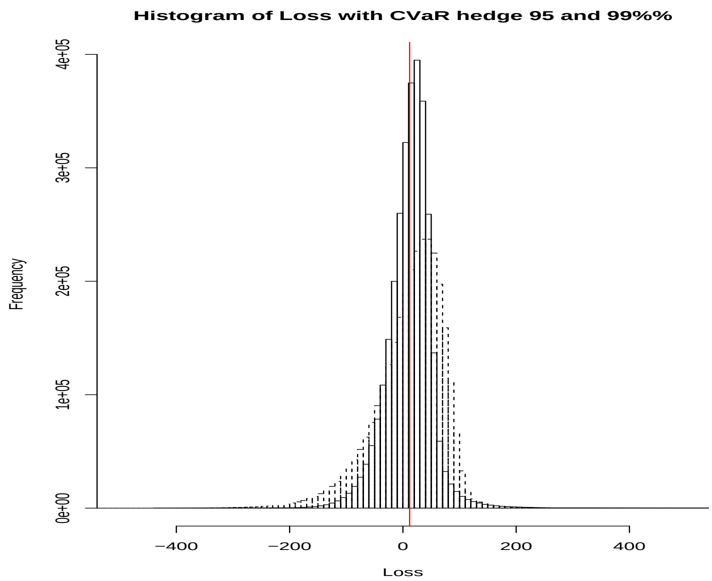

Figure 2: Histogram of one step CVaR-hedged loss at level $\alpha = 95\%$ (normal lines) and $\alpha = 99\%$ (dashed lines). The vertical line is the mean of the portfolio loss distribution.

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

	No hedging		Static hedging		
α	VaR	CVaR	VaR	θ^*_α	CVaR
95%	784.6	1226.3	259.6	81.6	366.5
99%	1452.4	2012.3	437.1	89.9	537.3
99.5%	1769.9	2382.8	505.7	92.3	608.6

Table 2: Self-financed static hedging of Consumption

reduce its risk at time $t_0 = 0$. A crude Monte Carlo gives $\mathbb{E}[L] = 62.6$ with a variance of 10747.4 after 3 000 000 trials.

In this example, IS algorithms and the LCV method based on X don't achieve any significative variance reduction. Consequently, for the hedged portfolio, we don't provide any variance reduction ratio. However, we notice that in order to estimate both VaR and CVaR of the loss L without hedging portfolio, the IS algorithm and the LCV method provides significant variance reduction. Those 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 doesn't 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 succeeds in reducing the considered variances. Results are summarized in Table 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 Fig 3. that the right tail distribution (which corresponds to high loss) is greatly reduced. The deformation provided by a CVaR hedging at level 95% is very impresive. 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 loss occuring the most frequently has changed 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. Fig 4. illustrates the histrograms obtained with a CVaR hedging at level 95% and 99%. We remark that the distribution which corresponds to a CVaR hedging at level 99% has heavier tails than the one corresponding to a CVaR hedging at level 95%. 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 our 4 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 3 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.* $N = N_\ell = 10$, $\ell = 1, \cdots, M$. Note that we do not quantify the process $(S_{t_\ell}, T_{t_\ell})_{1\leq M-1}$ but only the two gaussian random variables $(G1, G2)$ so that our quantization trees are obtained as a transform of the 2-dimensional normal distribution optimal grid. It is crucial to have a good approximate of the random variable $\tilde{\Delta}L_\ell$, $\ell = 1, \cdots, M$, for the method M.D.H. so that we use an optimized quantization grid of size 100 in (67). Results are summarized in Table 3.

We clearly see that the optimal strategy is given by the M.D.H. method when the number of

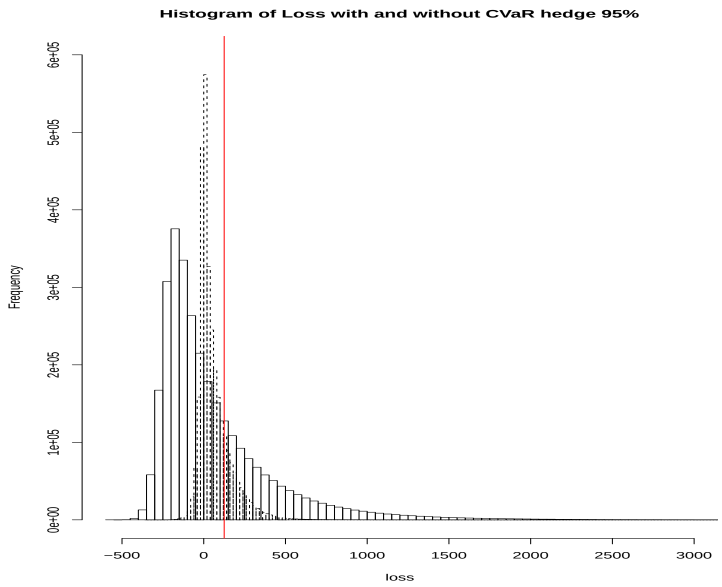

Figure 3: Histogram of loss with (dashed lines) and without (normal lines) one step CVaR-hedging at level $\alpha = 95\%$. The vertical line is the mean of the portfolio loss distribution.

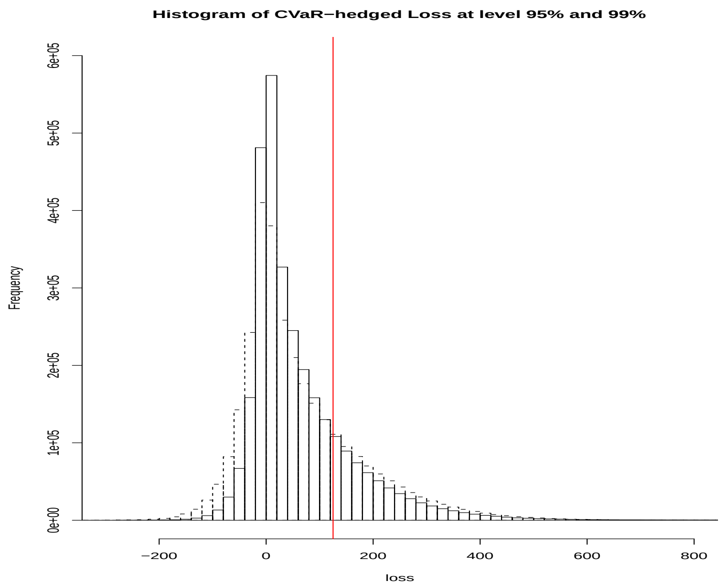

Figure 4: Histogram of one step CVaR-hedged loss at level $\alpha = 95\%$ (normal lines) and $\alpha = 99\%$ (dashed lines).

trading dates becomes large. The C.H. method for $M \leq 12$ is optimal but 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 doesn't 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 5 presents the histograms of the loss without any hedging strategy and with a

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

Table 3: Self-financed dynamic CVaR hedging of Consumption at level 95% with different strategies.

CVaR-hedging at level 95% using the M.D.H. method with 52 trading dates. The deformation of the loss distribution is very impresive. 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 6 compares the CVaR hedged loss distribution at level 95% 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: dynamic strategy reduces greatly high losses and slightly high gains. Figure 7 shows the 10 components of the optimal trading strategy using the M.D.H with 52 trading dates.

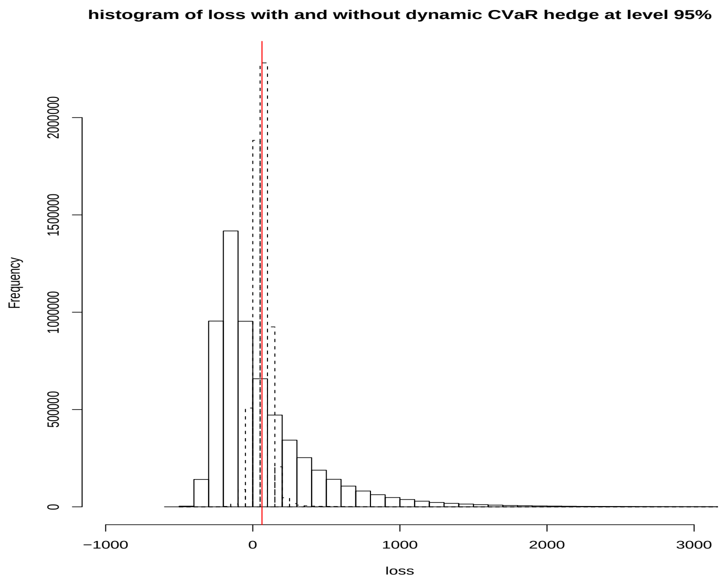

Figure 5: Histogram of Consumption's Loss with (dashed lines) and without (normal lines) dynamic CVaR-hedging at level $\alpha = 95\%$ using the M.D.H. strategy (52 trading dates).

References

[1] P. Artzner, F. Delbaen, E. J.M., and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[2] M. Avellaneda. Minimum-relative-entropy calibration of asset pricing models. *International Journal of Theoretical and Applied Finance*, 1:447–472, 1998.

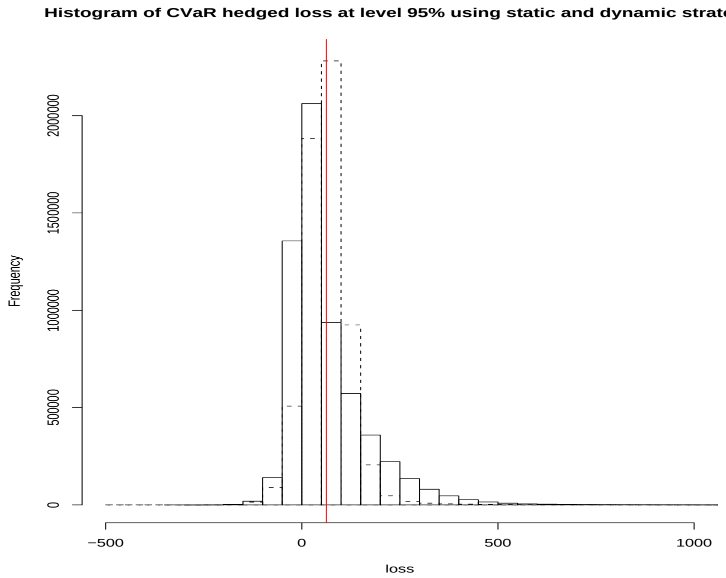

Figure 6: Histogram of CVaR hedged Loss at level $\alpha = 95\%$ with one step static (normal lines) and dynamic (dashed lines, M.D.H. with 52 trading dates) self-financed strategies.

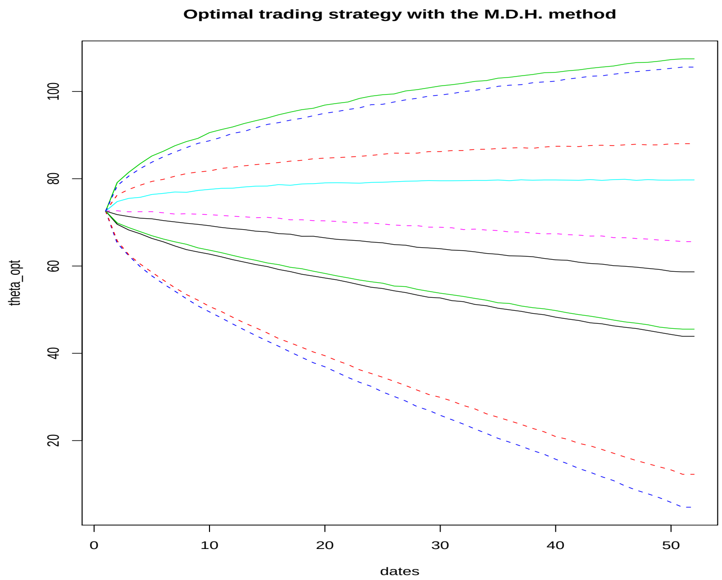

Figure 7: Optimal trading strategy using the M.D.H. at level $\alpha = 95\%$ with 52 trading dates and 10 quantization nodes per dates.

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

[4] P. Barrieu and N. El Karoui. Optimal derivatives design under dynamic risk measures. *Mathematics of Finance*, A.M.S. Proceedings 351, 2004.

[5] M. Benveniste, M. Metivier, and P. Priouret. *Adaptive algorithms and stochastic approximation.*

Springer-Verlag, 1990.

[6] M. Duflo. *Iterative random models.* Springer-Verlag, 1988.

[7] M. Duflo. *Algorithmes Stochastiques.* Springer, Berlin, 1996.

[8] N. El Karoui and M.-C. Quenez. Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM J. Control and Optim.*, 33:29–66, 1995.

[9] N. El Karoui and R. Rouge. Pricing via utility maximsation and entropy. *Mathematical Finance*, 10:250–276, 2000.

[10] V. Evstigneev, I. Measurable selection and dynamic programming. *Mathematics of operations Research*, 1:267–275, 1976.

[11] H. Foellmer and P. Leukert. Quantile hedging. *Finance and Stochastics*, 3:251–273, 1999.

[12] H. Foellmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[13] H. Foellmer and M. Schweizer. Hedging of contingent claims under incomplete information. *Applied Stochastic Analysis*, pages 389–414, 1991.

[14] H. Foellmer and D. Sonderman. Hedging of non-redundant contingent claims. *Contributions to Mathematical Economics*, pages 205–223, 1986.

[15] N. Frikha. *Contribution à la modélisation et à la gestion dynamique du risque sur les marchés de l'énergie.* PhD thesis, Université Pierre & Marie Curie, 2010.

[16] M. Fritteli. The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10:39–52, 2002.

[17] S. D. Hodges and A. Neuberger. Optimal replication of contingent claims under transaction costs. *The Review of Futures Markets*, 8:222–239, 1989.

[18] I. Karatzas, J. Lehoczky, S. Shreve, and G. Xu. Martingale and duality methods for utility maximization in an incomplete market. *SIAM J. Control and Optimization*, 29:702–730, 1991.

[19] H. J. Kushner and G. Yin. *Stochastic Approximation and Recursive algorithms and Applications.* Springer, 2003.

[20] V. Lemaire and G. Pagès. Unconstrained recursive importance sampling. *Annals of Applied Probability*, 20:1029–1067, 2010.

[21] Y. Miyahara. Pricing model and related estimation problems. *Asia-Pacific Financial Markets*, 8:45–60, 2001.

[22] G. Pagès. A space vector quantization method for numerical integration. *J. Computational and Applied Mathematics*, 89:1–38, 1998.

[23] M. Pelletier. Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39(1):49–72 (electronic), 2000.

[24] M. Pelletier. Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39:49–72, 2000.

[25] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:493–517, 2000.

[26] M. Schweizer. Option hedging for semimartingales. *Stochastic Processes & Their Applications*, pages 339–363, 1991.

5 Appendix: Proof of Proposition 2.4

We propose below the proof of (21) and (22) which are the key results in order to derive our R.M. algorithm. The proof of (19) and (20) will follow using similar arguments.

Proof. First note that since $L \in L^1(\mathbb{P})$, $\mathbb{E}\left[L|\,\mathcal{F}\right] \in L^0\left(\mathcal{F}\right)$ so that

$$\mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\theta.X\right)\right] = \mathbb{E}[L] + \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\mathbb{E}\left[L|\,\mathcal{F}\right]-\theta.X\right)\right].$$

Consequently, we can suppose that $\mathbb{E}\left[L|\,\mathcal{F}\right]=0$ for the rest of the proof. It is straightforward that

$$\inf_{\theta L^0_{\mathbb{R}^d}(\mathcal{F})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\theta.X\right)\right] \geq \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta,\xi\in L^0(\mathcal{F})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L-\theta.X-\xi\right)_+|\mathcal{F}\right]\right].$$

Let $(\theta_n)_{n\geq 1}$ be a sequence in $L^0_{\mathbb{R}^d}\left(\mathcal{F}\right)$ such that

$$\operatorname*{ess\,inf}_{\theta,\xi\in L^0_{\mathbb{R}^d}(\mathcal{F})} \mathbb{E}\left[\xi + \frac{1}{1-\alpha}\left(L-\theta.X-\xi\right)_+|\mathcal{F}\right] = \inf_{n\geq 1} \mathcal{F}\text{-CVaR}_\alpha\left(L-\theta_n.X\right),$$

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

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

Note that $\Xi_n \in L^0_{\mathbb{R}^d}(\mathcal{F})$ for $n\geq 1$ and

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

so that the sequence $\left(\mathcal{F}\text{-CVaR}_\alpha\left(L-\Xi_n.X\right)\right)_{n\geq 0}$ is non increasing and by Jensen's inequality

$$\mathcal{F}\text{-CVaR}_\alpha\left(L-\Xi_n.X\right) \geq \frac{1}{1-\alpha}\mathbb{E}\left[L_+|\,\mathcal{F}\right] \geq \frac{1}{1-\alpha}\mathbb{E}\left[L|\,\mathcal{F}\right]_+ = 0.$$

Moreover, by definition for $n\geq 0$

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

The sequence $\left(\mathcal{F}\text{-CVaR}_\alpha\left(L-\Xi_n.X\right)\right)_{n\geq 0}$ converges in $L^1(\mathbb{P})$ ought to Beppo-Levi's Theorem and

$$\begin{aligned}\inf_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\theta.X\right)\right] &\geq \inf_{n\geq 0}\mathbb{E}\left[\mathcal{F}\text{-CVaR}_\alpha\left(L-\Xi_n.X\right)\right] \\ &\geq \mathbb{E}\left[\inf_n \mathcal{F}\text{-CVaR}_\alpha\left(L-\Xi_n.X\right)\right] \\ &= \mathbb{E}\left[\operatorname*{ess\,inf}_{\theta\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P})} \mathcal{F}\text{-CVaR}_\alpha\left(L-\theta.X\right)\right].\end{aligned}$$

The proof of (19) follows using similar arguments.

Let $\omega\in\Omega$. The convexity of $V_f(\omega,.,.)$ is a consequence of the convexity of $(\xi,\theta)\mapsto v_f(\xi,\theta,y,x)$ for all $(y,x)\in\mathbb{R}\times\mathbb{R}^d$. Owing to Jensen's inequality, for all $(\xi,\theta)\in\mathbb{R}\times\mathbb{R}^d$,

$$V_f(\omega,\xi,\theta) \geq \xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega,\mathrm{d}x,\mathrm{d}y)-\xi\right)_+.$$

Now, if $\xi < \int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y)$, we have $\xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi\right)_+ = -\frac{\alpha}{1-\alpha}\xi + \frac{1}{1-\alpha}\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) \to +\infty$, $\xi \to -\infty$. Moreover for all $\xi \in \mathbb{R}$, $\xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi\right)_+ \geq \xi$, thus it imples that $\lim_{\xi\to+\infty} \xi + \frac{1}{1-\alpha}\left(\int y\Pi(\omega, \mathrm{d}x, \mathrm{d}y) - \xi\right)_+ = +\infty$, 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 $\inf_{\xi\in\mathbb{R}} V_f(\omega, \xi, \theta) = \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)(\omega)$ satisfies

$$\lim_{|\theta|\to+\infty} \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)(\omega) = +\infty.$$

First note that the sub-additivity of the function $x \mapsto x_+$ implies that

$$\mathcal{F}\text{-CVaR}_\alpha\left(-\theta.X\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right) + \mathcal{F}\text{-CVaR}_\alpha\left(-L\right),$$

so that,

$$|\theta|\mathcal{F}\text{-CVaR}_\alpha\left(-\frac{\theta}{|\theta|}.X\right) - \mathcal{F}\text{-CVaR}_\alpha\left(-L\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right),$$

which finally yields,

$$|\theta| \operatorname*{ess\,inf}_{u\in L^0_{\mathbb{R}^d}(\mathcal{F},\mathbb{P}),|u|=1} \mathcal{F}\text{-CVaR}_\alpha\left(u.X\right) - \mathcal{F}\text{-CVaR}_\alpha\left(-L\right) \leq \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right),$$

so that owing to Assumption 4 *ii*), $\lim_{|\theta|\to+\infty} \mathcal{F}\text{-CVaR}_\alpha\left(L - \theta.X\right)(\omega) = +\infty$. The proof of (20) 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 $\mathcal{F}$-measurable owing to measurable selection theorem (see e.g. Lemma 3 and Lemma 4 in [10]), such that

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

□