HAL open science

Epiconvergence of relaxed stochastic optimization problem

Vincent Leclère

▶ To cite this version:

Vincent Leclère. Epiconvergence of relaxed stochastic optimization problem. 2013. hal-00848275v1

HAL Id: hal-00848275

https://hal.science/hal-00848275v1

Preprint submitted on 26 Jul 2013 (v1), last revised 2 Nov 2020 (v3)

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

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

Epiconvergence of relaxed stochastic optimization problem

Vincent Leclère, Université Paris-Est, CERMICS,
École des Ponts ParisTech,
6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2

16/07/2013

Abstract

In this paper we consider the relaxation of a dynamic stochastic optimization problem where we replace a stochastic constraint - for example an almost sure constraint - by a conditional expectation constraint. We show an epiconvergence result relying on the Kudo convergence of $\sigma-$algebra and continuity of the objective and constraint operators. We also present some classicals constraints in stochastic optimization and give some conditions insuring their continuity. We conclude with a decomposition algorithm that uses such a relaxation.

1 Introduction

Mathematical optimization is concerned with the problem of minimizing an objective function J mapping a set $\mathcal{U}$ into $\mathbb{R}\cup\{+\infty\}$ over a set of admissible controls $U^{ad} \subset \mathcal{U}$. In the field of stochastic optimization the space of control variable $\mathcal{U}$ is a space of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with value in a set $\mathbb{U}$, and most of the time the objective function J is given as $J(\boldsymbol{U}) = \mathbb{E}(j(\boldsymbol{U}))$, where j is a function mapping $\mathbb{U}$ into $\mathbb{R}\cup\{+\infty\}$. Consider the following problem

$$\min_{\boldsymbol{U}\in\mathcal{U}} \quad \mathbb{E}\big(j(\boldsymbol{U})\big) \ , \tag{1a}$$

$$s.t. \quad \Theta(\boldsymbol{U}) = 0 \ . \tag{1b}$$

If Ω is not finite solving this problem is extremely difficult as $\mathcal{U}$ is of infinite dimension. Thus there is a need to approximate Problem (1). In order to give theoretical results for the approximations of an optimization problem the right notion of convergence is the epi-convergence. Indeed, under some additional technical conditions, if a sequence of functions $\tilde{J}_n$ epi-converge towards $\tilde{J}$ then the sequence of minimizers of $\tilde{J}_n$ converges toward a minimizer of $\tilde{J}$.

A first idea consist in approximating the probability $\mathbb{P}$. Roughly speaking the Sample Average Approximation (see [8] and reference therein) procedure consist in simulating a set of scenarios under the real probability $\mathbb{P}$. Then we solve Problem (1) under the empirical probability uniform on the simulated scenarios. Intuitively the Law of Large Number gives theoretical foundation for this method and there is a lot of litterature on this approach. To be more precise there is a number of epi-convergence results for those approximations (see [4] [7] or [19]). More generally the idea of discretizing or quantizing Ω, for example by use of finite scenario trees have been largely studied in the field of Stochastic Programming (see [18] for a thorough presentation).

However in those approach the constraints, and especially the information constraints, are quite difficult to take into accounts. Therefore we propose a way to approximate constraints, especially almost sure constraints. The main idea is to replace a constraint by its conditional expectation. This is in some sense an aggregation of constraints. This approximation appears when considering duality schemes for dynamic stochastic optimization problem. More precisely if we replace an almost sure constraint by its conditional expectation with respect to (w.r.t.) a σ-algebra $\mathcal{B}$, then if there exist an optimal Lagrange multiplier, then there is an optimal Lagrange multiplier measurable w.r.t. $\mathcal{B}$. Consequently if $\mathcal{B}$ is well chosen then a decomposition-coordination approach can be used to solve the approximated problem. This is presented in the last part of our paper. Note that in this case the approximation can be seen as a decision rule approach in the dual, where we choose to restrict the multiplier in the class of $\mathcal{B}$-measurable random variables.

The paper is organized as follows. Section 2 presents the general form of the problem considered and its relaxation. Section 3 shows conditions on the sequence of approximate problems guaranteeing its convergence toward the initial problem. The main assumptions are the Kudo's convergence of $\sigma-$algebra, and the continuity - as operators - of the constraint function Θ and objective function J. Section 4 give some examples of continuous objective and constraint functions that represents some usual stochastic optimization problems. Finally section 5 presents a decomposition-coordination algorithm using this type of relaxation.

2 Problem statement

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote by $L^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{U})$ the set of random variables $\boldsymbol{U}$ with value in a Banach space $\mathbb{U}$ such that $\mathbb{E}\left(\left\|\boldsymbol{U}\right\|^p \right) < +\infty$.

We consider a stochastic optimization problem

$$\min_{\boldsymbol{U}\in\mathcal{U}} \quad J(\boldsymbol{U}) \ , \tag{2a}$$

$$s.t. \quad \Theta(\boldsymbol{U}) \in -C \ . \tag{2b}$$

Where the criterion J maps $\mathcal{U}$ into $\mathbb{R}\cup\{+\infty\}$; the operator of contraint Θ maps $\mathcal{U}$ into $\mathcal{V}$, with $\mathcal{V} := L^p\big(\Omega,\mathcal{F},\mathbb{P};\mathbb{V}\big)$, and $p\in[1,\infty)$. We specify $\mathcal{V}$ because the convergence result rely on a result on random variables in L^p; strictly speaking no assumption are required on $\mathcal{U}$, however it is often $L^p(\Omega,\mathcal{F},\mathbb{P};\mathbb{U})$ for some p in $[1,\infty]$. We assume that $C\subset\mathcal{V}$ is a closed convex cone of $\mathcal{V}$, and that $\mathbb{U}$ and $\mathbb{V}$ are separable Banach spaces with separable dual. The fact that C is a cone is not essential for our results.

The problem is static for now, but we will later consider a dynamic problem.

The usual choice for the criterion is the expected cost $J(\boldsymbol{U}) := \mathbb{E}\big(j(\boldsymbol{U})\big)$, where $j:\mathbb{U}\to\mathbb{R}$. Other choices could be risk measures (see [1] for example) like Conditional-Value-at-Risk (see [13] for a definition), worst-case or robust approaches and many others. Θ is also very general, and can represent, for example,

- almost sure constraints : $\Theta\big(\boldsymbol{U}\big)(\omega) := \theta\big(\boldsymbol{U}(\omega)\big)$, where θ maps $\mathbb{U}$ into $\mathbb{V}$ and $\theta\big(\boldsymbol{U}\big)\in C$ is realized almost surely;
- measurability constraints : $\Theta\big(\boldsymbol{U}\big) := \mathbb{E}\big(\boldsymbol{U}\ \big|\ \mathcal{B}\big)-\boldsymbol{U}$, with $C=\{0\}$, as $\boldsymbol{U}$ is measurable with respect to the σ-algebra $\mathcal{B}$ iff $\mathbb{E}\big(\boldsymbol{U}\ \big|\ \mathcal{B}\big)=\boldsymbol{U}$;
- risk constraint : $\Theta\big(\boldsymbol{U}\big) := \rho(\boldsymbol{U})-a$, where ρ is a risk measure, and $C=\mathbb{R}^+$,
- or probability constraint : $\Theta\big(\boldsymbol{U}\big) := \mathbb{P}\big(\boldsymbol{U}\in A\big)-p$, with $C=\mathbb{R}^+$, that is $\mathbb{P}\big(\boldsymbol{U}\in A\big)\geq p$.

We consider a sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$ of sub-fields of $\mathcal{F}$, not necessarily a filtration. This sequence approximate in some sense $\mathcal{F}$, and might not be finitely generated. The closed convex cone of constraint C is said to be *eventually stable w.r.t. projection on* $\mathcal{F}_n$, if for all $n\in\mathbb{N}$ large enough we have

$$\forall\,\boldsymbol{V}\in C, \qquad \mathbb{E}(\boldsymbol{V}\mid\mathcal{F}_n)\in C\ . \tag{3}$$

Let us consider the relaxation of Problem (2)

$$\min_{\boldsymbol{U}\in\mathcal{U}} \quad J(\boldsymbol{U}) \ , \tag{4a}$$

$$s.t. \quad \mathbb{E}\big(\Theta(\boldsymbol{U})\ \big|\ \mathcal{F}_n\big)\in -C\ . \tag{4b}$$

We denote the set of admissible controls of Problem (2)

$$\mathcal{U}^{ad} := \left\{ \boldsymbol{U} \in \mathcal{U} \;\middle|\; \Theta(\boldsymbol{U}) \in -C \right\} , \tag{5}$$

and the corresponding set of admissible controls of problem (4)

$$\mathcal{U}_n^{ad} := \left\{ \boldsymbol{U} \in \mathcal{U} \;\middle|\; \mathbb{E}\big(\Theta(\boldsymbol{U}) \mid \mathcal{F}_n\big) \in -C \right\} . \tag{6}$$

Let us note that problems (2) and (4) can also be written[1]

$$\min_{\boldsymbol{U}\in\mathcal{U}} \quad \underbrace{J(\boldsymbol{U}) + \chi_{\mathcal{U}^{ad}}(\boldsymbol{U})}_{:=\tilde{J}(\boldsymbol{U})} , \tag{7}$$

and

$$\min_{\boldsymbol{U}\in\mathcal{U}} \quad \underbrace{J(\boldsymbol{U}) + \chi_{\mathcal{U}_n^{ad}}(\boldsymbol{U})}_{:=\tilde{J}_n(\boldsymbol{U})} . \tag{8}$$

Note that if we have $\mathcal{F}_n \subset \mathcal{F}_{n'}$ then $\mathcal{U}_n^{ad} \subset \mathcal{U}_{n'}^{ad} \subset \mathcal{U}^{ad}$, and in particular the problem (4) is indeed a relaxation of the original problem (2) as it has the same objective function but a wider set of admissible controls.

This is very similar to an aggregation constraint. Consider a finite set $\Omega = \{\omega_i\}_{i\in[\![1,n]\!]}$, with a probability $\mathbb{P}$ such that, for all $i \in [\![1,n]\!]$, we have $\mathbb{P}(\omega_i) = p_i > 0$. Consider a partition $\mathcal{B} = \{B_l\}_{l\in[\![1,L]\!]}$ of Ω, and the filtration $\mathcal{F}$ generated by $\mathcal{B}$. Assume that $C = \{0\}$, then the relaxation presented replace the constraint

$$\theta(\boldsymbol{U}) = 0 \qquad \text{a.s.}$$

explicitely written as n constraints

$$\theta(\boldsymbol{U}(\omega_i)) = 0 \qquad \forall i \in [\![1,n]\!] ,$$

by the collection of L constraints

$$\sum_{i\in B_l} p_i \theta(\boldsymbol{U}(\omega_i)) = 0 \qquad \forall l \in [\![1,L]\!] .$$

3 Epiconvergence result

Let fix some notations. Assume that $p \in [1,+\infty)$ and denote $q \in (1,+\infty]$ such that $1/q + 1/p = 1$. Let E be a separable Banach space with separable dual E'.

We say that a sequence $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ of $L^p(\Omega,\mathcal{F},\mathbb{P};E)$ converges strongly toward $\boldsymbol{X} \in L^p(\Omega,\mathcal{F},\mathbb{P};E)$, and write $\boldsymbol{X}_n \to_{L^p} \boldsymbol{X}$ if

$$\lim_{n\to\infty} \mathbb{E}\big(\left\| \boldsymbol{X}_n - \boldsymbol{X} \right\|^p \big) = 0 .$$

[1] We use the notation $\chi_A(x)$ for the characteristic function of A, that is the function with value 0 if $x \in A$, and $+\infty$ elsewhere.

We say that a sequence $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ of $L^p(\Omega,\mathcal{F},\mathbb{P};E)$ weakly converges toward $\boldsymbol{X}\in L^p(\Omega,\mathcal{F},\mathbb{P};E)$, and write $\boldsymbol{X}_n \rightharpoonup_{L^p} \boldsymbol{X}$ if

$$\forall \boldsymbol{X}' \in L^q(\Omega,\mathcal{F},\mathbb{P};E') \qquad \lim_{n\to\infty} \mathbb{E}\big(\langle \boldsymbol{X}_n - \boldsymbol{X}, \boldsymbol{X}'\rangle_{E,E'}\big) = 0 \ .$$

See [15] for more information or properties on this convergences.

A sequence of function $(J_n)_{n\in\mathbb{N}}$ epi-converge toward a function J if the sequence of epigraphs of J_n converges toward the epigraph of J, where the convergence of sets is the Painlevé-Kuratowski convergence. For precise definitions and properties of epi-convergence in finite dimension see Rockafellar-Wets [12], and Attouch [2] for infinite dimension.

If $\mathcal{F}$ is a σ-algebra and $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is a sequence of sub-fields of $\mathcal{F}$, we say that $(\mathcal{F}_n)_{n\in\mathbb{N}}$ Kudo-converges toward the sub-fields $\mathcal{F}_\infty$, and denote $\mathcal{F}_n \to \mathcal{F}_\infty$ if for all set $F\in\mathcal{F}$, we have that $\Big(\mathbb{E}\big(1_F \mid \mathcal{F}_n\big)\Big)_{n\in\mathbb{N}}$ converges in probability toward $\mathbb{E}\big(1_F \mid \mathcal{F}_\infty\big)$. In [9] Kudo shows that it is equivalent to saying that for all integrable random variable $\boldsymbol{X}$ we have $\mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big)$ converging in L^1 toward $\mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_\infty\big)$.

3.1 Preliminaries

We begin by a Lemma from Piccinini [10] that extends the convergence in L^1 to the convergence in L^p in the strong or weak sense.

Lemma 1. *Let $\big(\mathcal{F}_n\big)_{n\in\mathbb{N}}$ be a sequence of $\sigma-$algebra. The following statements are equivalent :*

1. $\mathcal{F}_n \to \mathcal{F}_\infty$.
2. $\forall \boldsymbol{X}\in L^p(\Omega,\mathcal{F},\mathbb{P};E), \quad \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big) \to_{L^p} \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_\infty\big)$.
3. $\forall \boldsymbol{X}\in L^p(\Omega,\mathcal{F},\mathbb{P};E), \quad \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big) \rightharpoonup_{L^p} \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_\infty\big)$.

And we have the following corollary where both the random variable and the σ-algebra are parametrized by n.

Corollary 1. *Let $\big(\mathcal{F}_n\big)_{n\in\mathbb{N}}$ be a sequence of σ-algebra. If $\mathcal{F}_n \to \mathcal{F}_\infty$, and $\boldsymbol{X}_n \to_{L^p} \boldsymbol{X}$ (resp. $\boldsymbol{X}_n \rightharpoonup_{L^p} \boldsymbol{X}$) then $\mathbb{E}\big(\boldsymbol{X}_n \mid \mathcal{F}_n\big) \to_{L^p} \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_\infty\big)$ (resp. $\mathbb{E}\big(\boldsymbol{X}_n \mid \mathcal{F}_n\big) \rightharpoonup_{L^p} \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_\infty\big)$).*

Proof. The weak-limit case is explicited in [10]. We show the strong convergence case. If $\boldsymbol{X}_n \to_{L^p} \boldsymbol{X}$, then

$$\begin{aligned}||\mathbb{E}\big(\boldsymbol{X}_n \mid \mathcal{F}_n\big) - \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}\big)||_{L^p} \leq ||\mathbb{E}\big(\boldsymbol{X}_n \mid \mathcal{F}_n\big) - \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big)||_{L^p} \\ + ||\mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big) - \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}\big)||_{L^p}\end{aligned}$$

And we have

$$||\mathbb{E}\big(\boldsymbol{X}_n \mid \mathcal{F}_n\big) - \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big)||_{L^p} \leq ||\boldsymbol{X}_n - \boldsymbol{X}||_{L^p} \to 0 \ ,$$

as the conditional expectation is a contraction. And we have

$$||\mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big) - \mathbb{E}\big(\boldsymbol{X} \mid \mathcal{F}_n\big)||_{L^p} \to 0$$

by Lemma 1. □

3.2 Epiconvergence in the general case

We can now present our main result.

Theorem 2. *Let $\mathcal{U}$ be endowed with a topology τ, and $\mathcal{V} = L^p\big(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{V}\big)$ be endowed with the strong or weak topology (p being in $[1, \infty)$). Assume that C is eventually stable by projection w.r.t $(\mathcal{F}_n)_{n\in\mathbb{N}}$ If Θ and J are continuous, and if $(\mathcal{F}_n)_{n\in\mathbb{N}}$ Kudo-converges to $\mathcal{F}$, then $\tilde{J}_{n\in\mathbb{N}}$ epi-converges to $\tilde{J}$.*

Proof. In a first place we recall the definition of the Painlevé-Kuratowski convergence of set. Let E be a topologic space and consider a sequence $(A_n)_{n\in\mathbb{N}}$ of subset of E. Then the inner limit of $(A_n)_{n\in\mathbb{N}}$ is the set of accumulation points of any sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_n \in A_n$, i.e,

$$\underline{\lim}_n A_n = \{x \in E \quad | \qquad \forall n \in \mathbb{N}, \quad x_n \in A_n, \quad \lim_{k\to\infty} x_n = x\} \ .$$

the outer limit of $(A_n)_{n\in\mathbb{N}}$ is the set of accumulation points of any sequence $(x_{n_k})_{k\in\mathbb{N}}$ such that $x_{n_k} \in A_k$, i.e,

$$\overline{\lim}_n A_n = \{x \in E \quad | \quad \exists (n_k)_{k\in\mathbb{N}}, \quad \forall k \in \mathbb{N}, \quad x_{n_k} \in A_k, \quad \lim_{k\to\infty} x_{n_k} = x\} \ .$$

We say that $(A_n)_{n\in\mathbb{N}}$ converges toward A in the Painlevé-Kuratowski sense if

$$A = \overline{\lim}_n A_n = \underline{\lim}_n A_n \ .$$

To prove the epi-convergence of $(\tilde{J}_n)_{n\in\mathbb{N}}$ toward $\tilde{J}$ it is sufficient to show that $\mathcal{U}_n^{ad}$ converges toward $\mathcal{U}^{ad}$ in the Painlevé-Kuratowski sense. Indeed it implies the epiconvergence of $(\chi_{\mathcal{U}_n^{ad}})_{n\in\mathbb{N}}$ toward $\chi_{\mathcal{U}^{ad}}$, and adding a continuous function conserve the epi-convergence (Attouch [2, Th 2.15]).

By stability of C for projection w.r.t. $(\mathcal{F}_n)_{n\in\mathbb{N}}$ we have that, for all $n \in \mathbb{N}$, $\mathcal{U}^{ad} \subset \mathcal{U}_n^{ad}$ and thus $\mathcal{U}^{ad} \subset \underline{\lim}_n \mathcal{U}_n^{ad}$.

We show that $\mathcal{U}^{ad} \supset \overline{\lim}_n \mathcal{U}_n^{ad}$. Let us take an element $\boldsymbol{U}$ of $\overline{\lim}_n \mathcal{U}_n^{ad}$. By definition there is a sequence $(\boldsymbol{U}_{n_k})_k$ that $\tau-$converges to $\boldsymbol{U}$, such that for all $k \in \mathbb{N}$, $\mathbb{E}\big(\Theta(\boldsymbol{U}_{n_k})|\mathcal{F}_{n_k}\big) \in -C$.

As Θ is continuous, we have $\Theta(\boldsymbol{U}_{n_k}) \to \Theta(\boldsymbol{U})$ strongly (resp weakly) in L^p. Moreover we have that $\mathcal{F}_{n_k} \to \mathcal{F}$, consequently by Corollary 1

$$\mathbb{E}\big(\Theta(\boldsymbol{U}_{n_k})|\mathcal{F}_{n_k}\big) \to \mathbb{E}\big(\Theta(\boldsymbol{U})|\mathcal{F}\big) \ ,$$

and thus $\Theta(\boldsymbol{U})$ is the limit of a sequence in $-C$, and by closedness of C (weak and strong as C is convex[2]) $\Theta(U) \in -C$. □

[2]if C is non-convex we need it to be sequentially closed.

The practical consequences on our relaxed problem is given in the following Corollary.

Corollary 3. *If $\left(\mathcal{F}_n\right) \to \mathcal{F}$, J and Θ are continuous, then we have the sequence of Problems (4) approximate Problem (2) in the following sense. If $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ is a sequence of control such that for all $n \in \mathbb{N}$,*

$$\tilde{J}_n(\boldsymbol{U}_n) < \inf_{\boldsymbol{U}\in\mathcal{U}} \tilde{J}_n(\boldsymbol{U}) + \varepsilon_n, \text{ where } \lim_n \varepsilon_n = 0,$$

then, for every converging sub-sequence $(\boldsymbol{U}_{n_k})_{k\in\mathbb{N}}$, we have

$$\tilde{J}(\lim_k \boldsymbol{U}_{n_k}) = \min_{\boldsymbol{U}\in\mathcal{U}} \tilde{J}(\boldsymbol{U}) = \lim_k \tilde{J}_{n_k}\left(\boldsymbol{U}_{n_k}\right)$$

Moreover if $\left(\mathcal{F}_n\right)$ is a filtration, then the convergences are monotonous in the sense that the optimal value is non-decreasing.

Proof. The convergence result is a direct application of Attouch [2, Th. 1.10, p. 27]. Monotonicity is given by the fact that, if $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is a filtration, then for $n > m$ then $\mathcal{U}_n^{ad} \subset \mathcal{U}_m^{ad}$. □

3.3 Dynamic Problem

We extend Problem (2) into the following dynamic problem

$$\begin{aligned} \min \quad & J(\boldsymbol{U}), \\ s.t. \quad & \Theta(\boldsymbol{U}_t) \in -C_t \quad \forall t \in [\![1,T]\!]. \end{aligned} \tag{9}$$

Where $\boldsymbol{U}$ is a stochastic process of control $(\boldsymbol{U}_t)_{t\in[\![1,T]\!]}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with value in $\mathbb{U}$. We have T constraints operators Θ_t with value in $L^p(\Omega, \mathcal{F}_t, \mathbb{P})$, where $(\mathcal{F}_t)_{t\in[\![1,T]\!]}$ is a sequence of σ-algebra[3]. Then, for each t we define a sequence of approximating σ-algebra $(\mathcal{F}_{n,t})_{n\in\mathbb{N}}$. For all $t \in [\![1,T]\!]$, C_t is a closed convex cone stable with respect to the projection on each $\mathcal{F}_{n,t}$ for $n \in \mathbb{N}$.

Finally we consider the sequence of approximated problem

$$\begin{aligned} \min \quad & J(\boldsymbol{U}) \ , \\ s.t. \quad & \mathbb{E}\big(\Theta(\boldsymbol{U}_t) \ \big| \ \mathcal{F}_{n,t}\big) \in -C_t \quad \forall t \in [\![1,T]\!] \ . \end{aligned} \tag{10}$$

Furthermore we denote

$$\mathcal{U}_t^{ad} := \left\{\boldsymbol{U}_t \in \mathcal{U}_t \ \middle| \ \Theta(\boldsymbol{U}_t) \in -C_t\right\} \ , \tag{11}$$

and

$$\mathcal{U}_{n,t}^{ad} := \left\{\boldsymbol{U}_t \in \mathcal{U}_t \ \middle| \ \mathbb{E}\big(\Theta(\boldsymbol{U}_t) \ \big| \ \mathcal{F}_n\big) \in -C_t\right\} \ . \tag{12}$$

[3] even if it is usually a filtration, it is not necessary.

Thus we have the set of admissible control for the original problem

$$\mathcal{U}^{ad} = \mathcal{U}_0^{ad} \times \cdots \times \mathcal{U}_T^{ad} \ ,$$

and accordingly for the relaxed problem

$$\mathcal{U}_n^{ad} = \mathcal{U}_{n,0}^{ad} \times \cdots \times \mathcal{U}_{n,T}^{ad} \ .$$

In order to show the convergence of the approximation proposed here we consider the functions

$$\tilde{J}(\boldsymbol{U}) = J(\boldsymbol{U}) + \chi_{\mathcal{U}^{ad}}(\boldsymbol{U}) \ ,$$

and

$$\tilde{J}_n(\boldsymbol{U}) = J(\boldsymbol{U}) + \chi_{\mathcal{U}_n^{ad}}(\boldsymbol{U}) \ ,$$

and show the epi-convergence of $\tilde{J}_n$ to $\tilde{J}$.

Theorem 4. *If Θ and J are continuous, and if for all $t \in [\![1, T]\!]$ $(\mathcal{F}_{t,n})_{n\in\mathbb{N}}$ Kudo-converges to $\mathcal{F}_t$, then $\tilde{J}_n$ epi-converges to $\tilde{J}$.*

Proof. The proof is deduced from the proof of Theorem 2. By following the same steps we obtain the Painlevé-Kuratowski convergence of $\mathcal{U}_{n,t}^{ad}$ to $\mathcal{U}_t^{ad}$, and thus the convergence of their cartesian products.

□

4 Examples

The assumption of continuity on J and Θ are not very usual as they are operators, that is they take random variables as arguments, consequently in this section we will give some examples of continuous Θ and J. We do not aim at exhaustivity but only gives some very classical examples and proofs that can be adapted to other cases.

4.1 A technical Lemma

We give a lemma that allow us to prove convergence for the topology of convergence in probability by considering sequence of random variables converging almost surely.

Lemma 2. *Let $\Theta : \mathcal{U} \to \mathcal{V}$, where $\mathcal{U}$ is a set of random variable endowed with the topology of convergence in probability, and $(\mathcal{V}, \tau)$ is a topological space. Assume that Θ is such that if $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ converges almost surely toward $\boldsymbol{U}$ then $\Theta(\boldsymbol{U}_n) \to_\tau \Theta(\boldsymbol{U})$. Then Θ is continuous for the topology of the convergence in probability.*

Proof. We recall first a well known property (see for example [5, Th 2.3.3]). Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in a topological space. If from any subsequence $\big(x_{n_k}\big)_{k\in\mathbb{N}}$ we can extract a sub-subsequence $\big(x_{\sigma(n_k)}\big)_{k\in\mathbb{N}}$ converging to x^*, then $(x_n)_{n\in\mathbb{N}}$ converges to x^*. Indeed suppose that $(x_n)_{n\in\mathbb{N}}$ does not converges toward x^*. Then there exist an open set $\mathcal{O}$ containing x^* and a subsequence $\big(x_{n_k}\big)_{k\in\mathbb{N}}$ such that for all $k\in\mathbb{N}$, $x_{n_k}\notin\mathcal{O}$, and no sub-subsequence can converges to x^*, which is a contradiction.

Let $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ be a sequence converging in probability to $\boldsymbol{U}$. We consider the sequence $\big(\Theta(\boldsymbol{U}_n)\big)_{n\in\mathbb{N}}$ in $\mathcal{V}$. We choose a subsequence $\big(\Theta\big(\boldsymbol{U}_{n_k}\big)\big)_{k\in\mathbb{N}}$, and we have $\boldsymbol{U}_{n_k}\to_{\mathbb{P}}\boldsymbol{U}$. Thus there exist a sub-subsequence $\boldsymbol{U}_{\sigma(n_k)}$ converging almost surely to $\boldsymbol{U}$, and consequently $\Theta\big(\boldsymbol{U}_{\sigma(n_k)}\big)\to\Theta\big(\boldsymbol{U}\big)$. Therefore Θ is sequentially continuous, and as the topology of convergence in probability is metrizable, Θ is continuous. $\square$

Let us note that the convergence in probability of $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ toward $\boldsymbol{U}$ is equivalent to saying that from any subsequence of $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ there is a further subsequence converging almost surely to $\boldsymbol{U}$ (see [5, Th 2.3.2]). Remark that this Lemma does not imply the equivalence between convergence almost sure and convergence in probability as you can not endow $\mathcal{U}$ with the "topology of almost sure convergence" as almost sure convergence is not induced by a topology.

4.2 Objective function

Let $\mathcal{U}$ be a set of random variables on $(\Omega,\mathcal{F},\mathbb{P})$.

The most classical objective function is given as $J(\boldsymbol{U}):=\mathbb{E}\big(j(\boldsymbol{U})\big)$, where $j:\mathbb{U}\to\mathbb{R}$ is a pointwise bounded cost function. This objective function represent a risk-neutral attitude as a random cost with high variance or a deterministic cost with the same expectation are considered equivalent. Recently in order to modelize risk-averse attitude coherent risk measures as defined in [1], or more generally convex risk measures as defined in [6], have been proeminent in the litterature.

Following [17] and reference therein we call a *convex risk measure* an operator $\rho:L^\infty\big(\Omega,\mathcal{F},\mathbb{P}\big)\to\mathbb{R}\cup\{+\infty\}$ defined by

$$\rho(\boldsymbol{X}):=\sup_{\mathbb{Q}\in\mathcal{P}}\mathbb{E}_{\mathbb{Q}}\big(\boldsymbol{X}\big)-g(\mathbb{P})\ ,$$

where $\mathcal{P}$ is a set of probabilities on $(\Omega,\mathcal{F})$ absolutely continuous w.r.t. $\mathbb{P}$, and g maps $\mathcal{P}$ into $\mathbb{R}$.

Proposition 5. *Let $\mathcal{U}$ be a set of random variable endowed with the topology of convergence in probability, and $J(\boldsymbol{U}):=\rho\big(j(\boldsymbol{U})\big)$, where $j:\mathbb{U}\to\mathbb{R}$ is continuous and bounded, and ρ a convex measure of risk. We have that $J:\mathcal{U}\to\mathbb{R}$ is continuous.*

Proof. Consider a sequence $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ of $\mathcal{U}$ converging in probability toward $\boldsymbol{U} \in \mathcal{U}$. Note that as j is bounded we have $\rho\big(j(\boldsymbol{U})\big) < \infty$. By definition of ρ, for all $\varepsilon > 0$ there is a probability $\mathbb{P}_\varepsilon \in \mathcal{P}$ such that

$$\mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{U}) - g(\mathbb{P}_\varepsilon)\big) \geq \rho\big(j(\boldsymbol{U})\big) - \varepsilon \ .$$

As $\mathbb{P}_\varepsilon$ is absolutely continuous w.r.t $\mathbb{P}$, the convergence in probability under $\mathbb{P}$ of $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ imply the convergence of probability under $\mathbb{P}_\epsilon$ and in turn the convergence in law under $\mathbb{P}_\epsilon$. By definition of convergence in law we have that

$$\lim_n \mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{U}_n)\big) - g(\mathbb{P}_\varepsilon) = \mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{U})\big) - g(\mathbb{P}_\varepsilon) \ .$$

It is easy to see that $\lim_n \rho \circ j(\boldsymbol{U}_n) = \rho \circ j(\boldsymbol{U})$. Indeed let η be a positive real, and set $\epsilon = \eta/2$, and $N \in \mathbb{N}$ such that

$$|\mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{U}_n)\big) - \mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{U})\big)| \leq \frac{\eta}{2} \ .$$

Then, recalling that

$$\rho \circ j\big(\boldsymbol{U}\big) \geq \mathbb{E}_{\mathbb{P}_{\frac{\eta}{2}}}\big(j(\boldsymbol{U})\big) - g(\mathbb{P}_{\frac{\eta}{2}}) \geq \rho \circ j\big(\boldsymbol{U}\big) - \frac{\eta}{2} \ ,$$

we have that for all $n \geq N$,

$$\rho \circ j\big(\boldsymbol{U}\big) + \frac{\eta}{2} \geq \rho \circ j\big(\boldsymbol{U}_n\big) \geq \rho \circ j\big(\boldsymbol{U}\big) - \eta \ .$$

□

Note that if the convex risk measure ρ is simply the expectation then we can simply endow $\mathcal{U}$ with the topology of convergence in law.

More generally, if $\mathcal{U}$ is a set of random variables endowed with the topology of convergence in probability, and j continuous such that $j(\boldsymbol{U})$ is dominated by some integrable random variable, then $J : \mathcal{U} \to \mathbb{R}$ is continuous. Indeed from Lemma 2 we know that we can consider a sequence $\boldsymbol{U}_n$ almost surely converging to $\boldsymbol{U}$, and we show that $\lim_n \mathbb{E}\big(j(\boldsymbol{U}_n)\big) = \mathbb{E}\big(j(\boldsymbol{U})\big)$ using a dominated convergence theorem.

The continuity assumption on j can also be relaxed. Indeed if $\big(\boldsymbol{U}_n\big)_{n\in\mathbb{N}}$ converges in law toward $\boldsymbol{U}$, and if the set K of points where j is continuous is such that $\mathbb{P}(\boldsymbol{U} \in K) = 1$, then $\mathbb{E}\big(j(\boldsymbol{U}_n)\big)$ converges toward $\mathbb{E}\big(j(\boldsymbol{U})\big)$.

4.3 Constraint operator

We present some usual constraints and how they can be represented by an operator Θ that is continuous and take values into $\mathcal{V} = L^p(\Omega, \mathcal{F}, \mathbb{P})$.

4.3.1 Almost sure constraint

From Lemma 2 we obtain a first important example of continuous constraints,

Proposition 6. *If $\mathcal{U}$ is the set of random variable on $\big(\Omega, \mathcal{B}, \mathbb{P}\big)$, with the topology of convergence in probability, and if $\theta : \mathbb{U} \to \mathbb{V}$ is continuous and bounded, then $\Theta\big(\boldsymbol{U}\big)(\omega) := \theta\big(\boldsymbol{U}(\omega)\big)$ is continuous. Moreover if $\mathcal{B} \subset \mathcal{F}$, then $\Theta\big(\boldsymbol{U}\big) \in \mathcal{V}$.*

Proof. Suppose that $\boldsymbol{U}_n$ converges to $\boldsymbol{U}$ almost surely, then by boundeness of θ we have that $\Big(\Big\|\theta\big(\boldsymbol{U}_{n)}\big) - \theta\big(\boldsymbol{U}\big)\Big\|_{\mathbb{V}}^p\Big)_{n\in\mathbb{N}}$ is bounded, and thus by dominated convergence theorem we have that

$$\lim_{k\to\infty} \theta\big(\boldsymbol{U}_n\big) = \theta\big(\boldsymbol{U}\big) \quad \text{in } L^p \ .$$

Consequently by Lemma 2 we have the continuity of Θ. As θ is continuous it is Borel measurable thus for all $\boldsymbol{U} \in \mathcal{U}$, for all Borel set $V \subset \mathbb{V}$ we have

$$\big(\Theta(\boldsymbol{U})\big)^{-1}(V) = \{\omega \in \Omega \mid \boldsymbol{U}(\omega) \in \theta^{-1}(V)\} \in \mathcal{B} \ ,$$

thus $\Theta(\boldsymbol{U})$ is $\mathcal{B}$ measurable. Boundeness of θ insure the existence of moment of all order of $\Theta(\boldsymbol{U})$. Thus if $\mathcal{B} \subset \mathcal{F}$ then $\Theta(\boldsymbol{U}) \in \mathcal{V}$. □

We note that boundeness of θ is only necessary in order to use the dominated convergence theorem. Thus an alternative set of assumptions is given in the following proposition.

Proposition 7. *If $\mathcal{U} = L^{p'}\big(\Omega, \mathcal{B}, \mathbb{P}\big)$, with the topology of convergence in probability, and if θ is $\gamma-$Holdër, with $\gamma \leq p'/p$ then $\Theta\big(\boldsymbol{U}\big)(\omega) := \theta\big(\boldsymbol{U}(\omega)\big)$ is continuous. Moreover if $\mathcal{B} \subset \mathcal{F}$, then $\Theta\big(\boldsymbol{U}\big) \in \mathcal{V}$.*

Proof. By definition a function θ mapping $\mathbb{U}$ into $\mathbb{V}$ is $\gamma-$Holdër iff there exist a constant $C > 0$ such that for all u, u' in $\mathbb{U}$ we have

$$\big\|\theta(u) - \theta(u')\big\|_{\mathbb{V}} \leq C \left\|u - u'\right\|_{\mathbb{U}} \ ,$$

in particular the 1-Holdër continuity is the Lipschitz continuity.

Following the preceding proof we just have to verify that the sequence $\Big(\big\|\theta\big(\boldsymbol{U}_n\big) - \theta\big(\boldsymbol{U}\big)\big\|_{\mathbb{V}}^p\Big)_{n\in\mathbb{N}}$ is dominated by some integrable variable. With our assumption we have

$$\Big\|\theta\big(\boldsymbol{U}_{n_k}\big) - \theta\big(\boldsymbol{U}\big)\Big\|_{\mathbb{V}}^p \leq C^p \left\|\boldsymbol{U}_{n_k} - \boldsymbol{U}\right\|_{\mathbb{U}}^{p\gamma} \ .$$

And as $p\gamma \leq p'$, and $\boldsymbol{U}_n$ and $\boldsymbol{U}$ are elements of $L^{p'}(\Omega, \mathcal{F}, \mathbb{P})$, $\left\|\boldsymbol{U}_{n_k} - \boldsymbol{U}\right\|_{\mathbb{U}}^{p\gamma}$ is integrable. □

4.3.2 Measurability constraint

Measurability constraints are extremely important in a dynamic setting. They can be represented by stating that a random variable and its conditional expectation are equal.

Proposition 8. *We set $\mathcal{U} = L^{p'}\big(\Omega, \mathcal{F}', \mathbb{P}\big)$, with $p' \geq p$ and $\mathcal{F}' \subset \mathcal{F}$. If $\mathcal{U}$ is equipped with the strong topology then $\mathcal{V}$ is either equipped with the strong or weak topology, and if $\mathcal{U}$ is equipped with the weak topology then $\mathcal{V}$ is equipped the weak topology. If $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{F}$, then $\Theta\big(\boldsymbol{U}\big) := \mathbb{E}\big(\boldsymbol{U} \mid \mathcal{B}\big) - \boldsymbol{U}$, is continuous. Moreover for all $\boldsymbol{U} \in \mathcal{U}$, we have $\Theta\big(\boldsymbol{U}\big) \in \mathcal{V}$.*

Proof. In a first place note that as $p' \geq p$, and $\mathcal{F}' \subset \mathcal{F}, \mathcal{U} \subset \mathcal{V}$; and if $\boldsymbol{V} \in \mathcal{V}$ then $\mathbb{E}\big(\boldsymbol{V} \mid \mathcal{B}\big) \in \mathcal{V}$ as the conditional expectation is a contraction.Thus for all $\boldsymbol{U} \in \mathcal{U}$, we have$\Theta(\boldsymbol{U}) \in \mathcal{V}$.

Consider a sequence $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ of $\mathcal{U}$ converging strongly in L^p toward $\boldsymbol{U} \in \mathcal{U}$. We have

$$\begin{aligned} ||\Theta\big(\boldsymbol{U}_n\big) - \Theta\big(\boldsymbol{U}\big)||_p &\leq ||\boldsymbol{U}_n - \boldsymbol{U}||_p + ||\mathbb{E}\big(\boldsymbol{U}_n - \boldsymbol{U} \mid \mathcal{B}\big)||_p \ , \\ &\leq 2||\boldsymbol{U}_n - \boldsymbol{U}||_p \ , \\ &\leq 2||\boldsymbol{U}_n - \boldsymbol{U}||_{p'} \to 0 \ . \end{aligned}$$

Thus the strong continuity of Θ is proven.

Now consider $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ converging weakly in L^p toward $\boldsymbol{U} \in \mathcal{U}$, we have, for all $\boldsymbol{Y} \in L^q$,

$$\begin{aligned} \mathbb{E}\Big(\mathbb{E}\big(\boldsymbol{U}_n \mid \mathcal{B}\big)Y\Big) = &\quad \mathbb{E}\Big(\boldsymbol{U}_n\mathbb{E}\big(Y \mid \mathcal{B}\big)\Big) \ , \\ \underset{n}{\longrightarrow} &\ \mathbb{E}\Big(\boldsymbol{U}\mathbb{E}\big(Y \mid \mathcal{B}\big)\Big) \ , \\ = &\quad \mathbb{E}\Big(\mathbb{E}\big(\boldsymbol{U} \mid \mathcal{B}\big)Y\Big) \ . \end{aligned}$$

Thus we have the weak convergence of the conditional expectation and therefore of Θ. Finally as the strong convergence imply the weak convergence we have the continuity from $\mathcal{U}$-strong into $\mathcal{V}$-weak. □

Until now the topology of convergence in probability has been largely used. In order to be endow $\mathcal{U}$ with the topology of convergence in probability in the previous proposition we need some stronger assumption on $\mathcal{U}$. Indeed if $\mathcal{U}$ is a set of probabilities such that there exist a random variable in $L^{p'}\big(\Omega, \mathcal{F}', \mathbb{P}\big)$ dominating every random variable in $\mathcal{U}$, then a sequence converging almost surely will converge in for the $L^{p'}$ norm and we can follow the previous proof to show the continuity of Θ.

4.3.3 Risk constraints

There is a lot of different ways to incorporate risk constraints in an optimization problem, we have seen that a risk measure can be choosen as objective function, and we show that it can also be used as a constraint.

Let ρ be a conditional risk mapping as defined in [16], that is ρ maps $\mathcal{U}$ into $\mathcal{V}$ where $\mathcal{V} \subset \mathcal{U}$ are linear spaces of real-valued functions $\phi(\omega)$, $\omega \in \Omega$, measurable with respect to $\mathcal{F}_U$ and $\mathcal{F}_V$ respectively ($\mathcal{F}_V \subset \mathcal{F}_U$). For each $\omega \in \Omega$ we define

$$\rho_\omega(\boldsymbol{U}) := [\rho(\boldsymbol{U})](\omega)$$

and we have in particular that ρ_ω is a convex function.

Proposition 9. *If ρ is a conditional risk mapping, j is a cost function mapping $\mathbb{U}$ into $\mathbb{R}$, and $a \in \mathcal{V}$, then $\Theta\big(\boldsymbol{U}\big) := \rho \circ j\big(\boldsymbol{U}\big) - a$ is continuous on the interior of its domain if $\mathcal{U}$ and $\mathcal{V}$ are equipped with the topology of convergence in probability.*

Proof. Let $\boldsymbol{U}_n$ be a sequence of random variables in the interior of the domain of $\rho \circ j$ converging almost surely to $\boldsymbol{U}$. Thus for almost all $\omega \in \Omega$, $\boldsymbol{U}_n(\omega) \in int(dom(\rho_\omega \circ j))$, and as $\rho_\omega \circ j$ is convex, we obtain $\rho_\omega \circ j(\boldsymbol{U}_n) \to \rho_\omega \circ j(\boldsymbol{U})$. Thus $\rho \circ j(\boldsymbol{U}_n)$ converges almost surely toward $\rho \circ j(\boldsymbol{U})$ and Lemma 2 achieve the proof. □

Note that if j is bounded then $\mathrm{dom}(\rho \circ j) = \mathcal{U}$.

Another widely used risk measure, even if it has some serious drawback, is the Value-at-Risk. If X is a real random variable it's value at risk of level α can be defined as $VaR_\alpha(\boldsymbol{X}) := \inf\{F_{\boldsymbol{X}}^{-1}(\alpha)\}$ where $F_{\boldsymbol{X}}(x) := \mathbb{P}(\boldsymbol{X} \leq x)$.

Proposition 10. *If $j : \mathbb{U} \to \mathbb{R}$ is continuous, and $\mathcal{U}$ such that every $\boldsymbol{U} \in \mathcal{U}$ have a continuous distribution function, then $\Theta(\boldsymbol{U}) := VaR_\alpha\Big(j\big(\boldsymbol{U}\big)\Big)$ is continuous if we have endowed $\mathcal{U}$ with the topology of convergence in law, and a fortiori for the topology of convergence in probability.*

Proof. By definition of convergence in law, if $\boldsymbol{U}_n \to \boldsymbol{U}$ in law, we have, for all $x \in \mathbb{R}$, $F_{\boldsymbol{U}_n}(x) \to F_{\boldsymbol{U}}(x)$, which means that $\Theta(\boldsymbol{U}_n) \to \Theta(\boldsymbol{U})$ almost surely, and as $\Theta(j(\boldsymbol{U}))$ is deterministic, Θ is continuous. □

5 Dual Approximate Dynamic Programming

In this section we will say a few words about how this approximation can be used.

5.1 Presentation of the problem

We are interested in an electricity production problem with N power stations. Each power station i have an internal state $\boldsymbol{X}_t^i$ at time step t, and is affected by a random exogenous noise $\boldsymbol{W}^i$. For each power station, and each time step, we have a control $\boldsymbol{U}_t^i \in \mathcal{U}_{t,i}^{ad}$ that must be measurable with respect to $\mathcal{F}_t$ (denoted $\boldsymbol{U}_t^i \preceq \mathcal{F}_t$) where $\mathcal{F}_t$ is the sigma algebra generated by all past noises : $\mathcal{F}_t = \sigma\big(\boldsymbol{W}_s^i\big)_{1\leq i\leq n, 0\leq s\leq t}$. Moreover there is a coupling constraint saying that the total production must be equal to the demand. It is represented as $\sum_{i=1}^N \theta_t^i(\boldsymbol{U}_t^i) = 0$, where θ_i is a continuous bounded function from $\mathbb{U}$ into $\mathbb{V}$, for all $i \in [\![1, n]\!]$. The cost to be minimized is a sum over time and power stations of all current local cost $L_t^i\big(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\big)$.

Finally the problem reads

$$
\begin{aligned}
\min_{\boldsymbol{X},\boldsymbol{U}} \quad & \mathbb{E}\Big(\sum_{i=1}^N \sum_{t=0}^T L_t^i\big(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\big)\Big) && && \text{(13a)}\\
& \boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i) && \forall t, \quad \forall i, && \text{(13b)}\\
& \boldsymbol{X}_0^i = x_0^i && \forall i, && \text{(13c)}\\
& \boldsymbol{U}_t^i \in \mathcal{U}_{t,i}^{ad} && \forall t, \quad \forall i, && \text{(13d)}\\
& \boldsymbol{U}_t^i \preceq \mathcal{F}_t && \forall t, \quad \forall i, && \text{(13e)}\\
& \sum_{i=1}^N \theta_t^i(\boldsymbol{U}_t^i) = 0 && \forall t, \quad \forall i, && \text{(13f)}
\end{aligned}
$$

Let assume that all random variable are in L^2 spaces and dualize the coupling constraint. We do not study here the relation between the primal and the following dual problem (see [11] and [14] for an alternative formulation involving duality between L^1 and its dual).

$$
\begin{aligned}
\max_{\lambda_t \in L^2} \quad \min_{\boldsymbol{X},\boldsymbol{U}} \quad & \mathbb{E}\Big(\sum_{i=1}^N \sum_{t=0}^T L_t^i\big(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\big) + \lambda_t \theta_t^i(\boldsymbol{U}_t^i)\Big) \\
& \boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i) && \forall t, \quad \forall i, \\
& \boldsymbol{X}_0^i = x_0^i && \forall i, \\
& \boldsymbol{U}_t^i \in \mathcal{U}_{t,i}^{ad} && \forall t, \quad \forall i, \\
& \boldsymbol{U}_t^i \preceq \mathcal{F}_t && \forall t, \quad \forall i,
\end{aligned}
\tag{14}
$$

We solve this problem with a gradient-type algorithm on λ. Thus for a fixed $\lambda^{(k)}$ we have to solve N problems of smaller size

$$
\begin{aligned}
\min_{\boldsymbol{X},\boldsymbol{U}} \quad & \mathbb{E}\Big(\sum_{t=0}^{T} L_t^i\big(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\big) + \lambda_t^{(k)}\theta_t^i(\boldsymbol{U}_t^i)\Big) \\
& \boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i) && \forall t, \\
& \boldsymbol{X}_0^i = x_0^i \\
& \boldsymbol{U}_t^i \in \mathcal{U}_{t,i}^{ad} && \forall t, \\
& \boldsymbol{U}_t^i \preceq \mathcal{F}_t && \forall t,
\end{aligned} \tag{15}
$$

However the process $\lambda^{(k)}$ can be choosen to be adapted to the filtration $(\mathcal{F}_t)_{t=1,..,T}$. Consequently solving Problem (15) by Dynamic Programming is possible but numerically difficult as we need to keep all the past realisations of the noises in the state. Consequently the so-called curse of dimensionality prevent us to solve numerically this problem.

Nevertheless it has recently been proposed in [3] to use $\mathbb{E}\big(\lambda_t \mid \boldsymbol{Y}_t\big)$ where $\boldsymbol{Y}_t$ is a random variable measurable with respect to $(\boldsymbol{Y}_{t-1}, \boldsymbol{W}_t)$ instead of λ_t. This is comparable to a decision rule approach for the dual as we are restraining the control to a certain class, the $\boldsymbol{Y}_t$-measurable λ in our case. Thus $(\boldsymbol{X}_t^i, \boldsymbol{Y}_t)$ is a state for Problem (15), which can be solved by Dynamic Programming (or another method). It has also been shown that, under some non-trivial conditions, using $\mathbb{E}\big(\lambda_t \mid \boldsymbol{Y}_t\big)$ instead of λ_t is equivalent to solving

$$
\begin{aligned}
\min_{\boldsymbol{X},\boldsymbol{U}} \quad & \mathbb{E}\Big(\sum_{i=1}^{N}\sum_{t=0}^{T} L_t^i\big(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\big)\Big) && && \text{(16a)} \\
& \boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i) && \forall t, \quad \forall i, && \text{(16b)} \\
& \boldsymbol{X}_0^i = x_0^i && \forall i, && \text{(16c)} \\
& \boldsymbol{U}_t^i \in \mathcal{U}_{t,i}^{ad} && \forall t, \quad \forall i, && \text{(16d)} \\
& \boldsymbol{U}_t^i \preceq \mathcal{F}_t && \forall t, \quad \forall i, && \text{(16e)} \\
& \mathbb{E}\Big(\sum_{i=1}^{N} \theta_t^i(\boldsymbol{U}_t^i) \;\Big|\; Y_t\Big) = 0 && \forall t, \quad \forall i. && \text{(16f)}
\end{aligned}
$$

Thus Problem (16) is a relaxation of Problem (13) where the almost sure constraint (13f) is replaced by the constraint (16f). Now consider a sequence of information processus $(\boldsymbol{Y}^{(n)})_{n\in\mathbb{N}}$ each generating a $\sigma-$algebra $\mathcal{F}_n$, and their associated relaxation $(\mathcal{P}_n)$ (as specified in Problem 16) of Problem (13) (denoted $(\mathcal{P})$). Those problems corresponds to Problems (9) and (10) with

$$
J(\boldsymbol{U}) = \mathbb{E}\Big(\sum_{i=1}^{N}\sum_{t=0}^{T} L_t^i\big(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\big)\Big),
$$

where $\boldsymbol{U} = (\boldsymbol{U}^{(i)})$ and $\boldsymbol{X}_t^i$ follow the dynamic equation (13b). We also have

$$\Theta_t(\boldsymbol{U}_t) = \sum_{i=1}^{N} \theta_t^i(\boldsymbol{U}_t^i)$$

and $C_t = \{0\}$.

From now on we assume that for all $t \in [\![1, T]\!]$, and all $i \in [\![1, N]\!]$ the cost functions L_t^i and constraint function Θ_t^i are continuous, and that $\mathcal{U}_{t,i}^{ad}$ is a compact subset of an euclidian space. Moreover we assume that the noise variables $\boldsymbol{W}_t^i$ are essentially bounded. Finally we endow the space of control processes with the topology of convergence in probability. Then by induction we have that the state processes and the control processes are essentially bounded, thus so is the cost $L_t^i\big(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\big)$. Thus the cost function can be effectively replaced by bounded functions. Consequently Proposition 5 insure that J is continuous if $\mathcal{U}$ is equipped with the topology of convergence in probability. Similarly Proposition 6 insure that Θ is continuous.

Thus Theorem 4 imply that our sequence of approximated problems $(\mathcal{P}_n)$ converges toward the initial problem $(\mathcal{P})$. More precisely assume that $(\boldsymbol{U}_n)_{n\in\mathbb{N}}$ is a sequence of ε_n-optimal solution of $\mathcal{P}_n$, i.e. $\boldsymbol{U}_n$ verifying constraint (16f) and $J(\boldsymbol{U}_n) < \inf_{\boldsymbol{U}\in\mathcal{U}_n^{ad}} + \varepsilon_n$, with $(\varepsilon_n)_{n\in\mathbb{N}}$ a sequence of positive real number converging to 0. Then we can extract a subsequence $(\boldsymbol{U}_{n_k})_{k\in\mathbb{N}}$ converging almost surely to an optimal solution of $(\mathcal{P})$, and the limit of the approximated value of $(\mathcal{P}_n)$ converges to the value of $(\mathcal{P})$.

References

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

[2] H. Attouch. *Variational convergence for functions and operators*, volume 1. Pitman Advanced Pub. Program, 1984.

[3] K. Barty, P. Carpentier, and P. Girardeau. Decomposition of large-scale stochastic optimal control problems. *RAIRO. Recherche opérationnelle*, 44(3):167–183, 2010.

[4] J. Dupacová and R.J.B. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The annals of statistics*, 16(4):1517–1549, 1988.

[5] R. Durrett and R. Durrett. *Probability: theory and examples*. Cambridge Univ Pr, 2010.

[6] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002. 10.1007/s007800200072.

[7] A.J. King and R.J.B. Wets. Epi-consistency of convex stochastic programs. *Stochastics and Stochastic Reports*, 34(1-2):83–92, 1991.

[8] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[9] H. Kudo. A note on the strong convergence of σ-algebras. *The Annals of Probability*, 2(1):76–83, 1974.

[10] L. Piccinini. A new version of the multivalued fatou lemma. *Journal of Applied Analysis*, 4(2):231–244, 1998.

[11] R. T. Rockafellar and R. J.B. Wets. The optimal recourse problem in discrete time: l^1-multipliers for inequality constraints. *SIAM Journal on Control and Optimization*, 16(1):16–36, 1978.

[12] R Tyrrell Rockafellar and Roger J-B Wets. *VARIATIONAL ANALYSIS*, volume 317. Springer Verlag, 1998.

[13] R.T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

[14] R.T. Rockafellar and R.J.B. Wets. Measures as lagrange multipliers in multistage stochastic programming. *Journal of mathematical analysis and applications*, 60(2):301–313, 1977.

[15] W. Rudin. Functional analysis. international series in pure and applied mathematics, 1991.

[16] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, pages 544–561, 2006.

[17] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, pages 433–452, 2006.

[18] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming: modeling and theory*, volume 9. Society for Industrial and Applied Mathematics, 2009.

[19] M. Zervos. On the epiconvergence of stochastic optimization problems. *Mathematics of Operations Research*, 24(2):495–508, 1999.