Operations Research Letters 47 (2019) 553–559

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Epiconvergence of relaxed stochastic optimization problems

Vincent Leclère

Université Paris-Est, CERMICS (ENPC), F-77455 Marne-la-Vallée, France

ARTICLE INFO

Article history:
Received 30 April 2019
Received in revised form 27 July 2019
Accepted 28 September 2019
Available online 3 October 2019

Keywords:
Stochastic optimization
Epiconvergence
Linear decision rules
Dynamic programming

ABSTRACT

We consider relaxation of almost sure constraint in dynamic stochastic optimization problems and their convergence. We show an epiconvergence result relying on the Kudo convergence of σ-algebras and continuity of the objective and constraint operators. We present classical constraints and objective functions with conditions ensuring their continuity. We are motivated by a Lagrangian decomposition algorithm, known as Dual Approximate Dynamic Programming, that relies on relaxation, and can also be understood as a decision rule approach in the dual.

1. Introduction

Stochastic optimization problems often consist in minimizing a cost over a set of random variables belonging to an infinite dimensional space. Consequently, there is a need for approximation. We are interested in the approximation of almost sure constraints, say $\theta(\boldsymbol{u}) = 0$ almost surely (a.s.), by a conditional expectation constraint like $\mathbb{E}\big[\theta(\boldsymbol{u}) \bigm| \mathcal{F}_n\big] = 0$ a.s.

Consider the following problem,

$$\min_{\boldsymbol{u}\in\mathcal{U}} \quad J(\boldsymbol{u}) \,, \tag{1a}$$

$$s.t. \quad \theta(\boldsymbol{u}) = 0 \quad \text{a.s.} \,, \tag{1b}$$

where the set of controls $\mathcal{U}$ is a set of random variables over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $J(\boldsymbol{u}) := \int_\Omega j(u(\omega)) d\mathbb{P}(\omega)$. If Ω is not finite, $\mathcal{U}$ may be of infinite dimension. Moreover the constraint (1b) is a functional constraint that can roughly be seen as an infinite number of constraints. For tractability purposes, we consider approximations of this problem. In order to give theoretical results for the approximations of Problem (1) the right notion of convergence is epi-convergence. Indeed, under some additional technical conditions, the epi-convergence ensures the convergence of both the optimal value and the optimal solutions.

One way of approximating Problem (1) consists in approximating the probability $\mathbb{P}$. Roughly speaking the Sample Average Approximation procedure consists in drawing a set of scenarios under the true probability $\mathbb{P}$. We then solve Problem (1) under the empirical probability on the set of drawn scenarios. In this literature (see [7,10]) the authors are interested in problems where the controls are deterministic. However other epi-convergence results have been shown for more general spaces of controls, including spaces of random variables or random processes (see [27] and references therein, as well as Pennanen [15],Pennanen and Koivu [17], Pennanen [16]). More generally, the idea of discretizing or quantizing the set Ω, for example by use of finite scenario tree has been largely studied in the field of Stochastic Programming (see [26] for a thorough presentation).

Instead of approximating the probability space we propose a way to approximate constraints, especially almost sure constraints. The main idea is to replace a constraint by its conditional expectation with respect to (w.r.t.) a σ-algebra $\mathcal{B}$. This is in some sense an aggregation of constraints. This approximation appears when considering Lagrangian duality schemes with dual linear decision rules for dynamic stochastic optimization problem [5,14, 19].

More precisely, we relax the almost sure constraint (1b) by replacing it by its conditional expectation, i.e.

$$\mathbb{E}\big[\theta(\boldsymbol{u}) \bigm| \mathcal{B}\big] = 0 \,. \tag{2}$$

If $\boldsymbol{\lambda}$ is an integrable optimal multiplier for Constraint (1b), then $\boldsymbol{\lambda}_{\mathcal{B}} = \mathbb{E}\big[\boldsymbol{\lambda} \bigm| \mathcal{B}\big]$ is an optimal multiplier for Constraint (2). This leads to look for $\mathcal{B}$-measurable multiplier, which may authorize decomposition–coordination methods where the sub-problems are easily solvable. 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 exists an optimal Lagrange multiplier, then there is an optimal Lagrange multiplier measurable w.r.t. the σ-algebra $\mathcal{B}$. Consequently if $\mathcal{B}$ is well chosen then a decomposition–coordination approach can be used to solve the approximated problem. 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. Works using a decision rule approach on the dual problem are found in Kuhn et al. [12].

E-mail address: vincent.leclere@enpc.fr.

https://doi.org/10.1016/j.orl.2019.09.014
0167-6377/

The paper is organized as follows. Section 2 presents the general form of the problem considered and its approximation. Section 3 shows, after a few recalls on convergence notions of random variables, functions and σ-algebras, conditions on the sequence of approximate problems guaranteeing its convergence toward the initial problem. The main assumptions are the Kudo's convergence of σ-algebra, and the continuity – as operators – of the constraint function Θ and objective function J. Section 4 gives some examples of continuous objective and constraint functions that represent usual stochastic optimization problems. Finally Section 5 quickly presents a Lagrangian decomposition algorithm using this type of relaxation. The results presented here show consistency of this method: if we refine the approximation, the solution obtained converges toward solution of the original problem.

Notation

Bold letters are used for random variables. $\mathbb{I}_A(x) = 0$ if $x \in A$, and $\mathbb{I}_A(x) = +\infty$ otherwise. We denote by $[\![a, b]\!]$ the set of all integers between a and b. θ is used for the constraint function mapping Euclidean space $\mathbb{U}$ into $\mathbb{V}$, whereas Θ is used for the constraint operator generally mapping a set of functions on $\mathbb{U}$ into function a set of function on $\mathbb{V}$.

2. Problem statement

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a topological space of controls $\mathcal{U}$. Let $\mathcal{V}$ be the spaces of random variables with value in a Banach $\mathbb{V}$ with finite moment of order $p \in [1, \infty)$, denoted $\mathcal{V} = L^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{V})$.

We consider now a stochastic optimization problem

$$\min_{\boldsymbol{u}\in\mathcal{U}} \quad J(\boldsymbol{u}) , \tag{3a}$$

$$s.t. \quad \Theta(\boldsymbol{u}) \in \mathcal{C} , \tag{3b}$$

with J mapping $\mathcal{U}$ into $\mathbb{R} \cup \{+\infty\}$, and Θ mapping $\mathcal{U}$ into $\mathcal{V}$. We assume that $\mathcal{C} \subset \mathcal{V}$ is a subset of $\mathcal{V}$, and that $\mathbb{V}$ is a separable Banach space with separable dual.

To give an example of cost operator, assume that $\mathcal{U} \subset L^1\big(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{U}\big)$, where $\mathbb{U}$ is a Banach space. The usual choice for the objective function is the expected cost $J(\boldsymbol{u}) := \mathbb{E}\big[j(\boldsymbol{u})\big]$, for a suitable cost function $j : \mathbb{U} \to \mathbb{R}$. Other choices could be risk measures (see [2] for example) like Average-Value-at-Risk, worst-case or robust approaches. The constraint operator Θ covers various cases, for example

- almost sure constraint: $\Theta\big(\boldsymbol{u}\big)(\omega) := \theta\big(\boldsymbol{u}(\omega)\big)$, where θ maps $\mathbb{U}$ into $\mathbb{V}$ and $\theta\big(\boldsymbol{u}\big) \in C$ is realized almost surely, where C is a closed convex set;
- measurability constraint: $\Theta\big(\boldsymbol{u}\big) := \mathbb{E}\big[\boldsymbol{u} \bigm| \mathcal{B}\big] - \boldsymbol{u}$, with $\mathcal{C} = \{0\}$, expresses that $\boldsymbol{u}$ is measurable with respect to the σ-algebra $\mathcal{B}$, that is, $\mathbb{E}\big[\boldsymbol{u} \bigm| \mathcal{B}\big] = \boldsymbol{u}$;
- risk constraint: $\Theta\big(\boldsymbol{u}\big) := \rho(\boldsymbol{u}) - a$, where ρ is a conditional risk measure, and $\mathcal{C}$ is the cone of negative random variables.

We introduce a stability assumption of the set $\mathcal{C}$ that will be made throughout this paper.

Definition 1. We consider a sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$ of sub-fields of $\mathcal{F}$. The set C is said to be *stable w.r.t.* $\big(\mathcal{F}_n\big)_{n\in\mathbb{N}}$, if there exists a set-valued mapping S from Ω to $\mathbb{V}$ which is closed-convex valued and measurable with respect to $\mathcal{F}$ and all $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

In particular if $\mathcal{C}$ is stable, we have for all $n \in \mathbb{N}$ and all $\boldsymbol{v} \in C$, $\mathbb{E}[\boldsymbol{v} \mid \mathcal{F}_n] \in C$.

We now consider the following relaxation of Problem (3)

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

$$s.t. \quad \mathbb{E}\Big[\Theta(\boldsymbol{u}) \Bigm| \mathcal{F}_n\Big] \in \mathcal{C} , \tag{4b}$$

where C is assumed to be stable w.r.t. the sequence $\big(\mathcal{F}_n\big)_{n\in\mathbb{N}}$.

We denote the set of admissible controls of Problem (3)

$$\mathcal{U}^{ad} := \Big\{\boldsymbol{u} \in \mathcal{U} \Bigm| \Theta(\boldsymbol{u}) \in -\mathcal{C}\Big\} , \tag{5}$$

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

$$\mathcal{U}^{ad}_n := \Big\{\boldsymbol{u} \in \mathcal{U} \Bigm| \mathbb{E}\Big[\Theta(\boldsymbol{u}) \Bigm| \mathcal{F}_n\Big] \in -\mathcal{C}\Big\} . \tag{6}$$

Problems (3) and (4) can also be written as

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

and

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

Since $\mathcal{F}_n \subset \mathcal{F}$, and $\mathcal{C}$ is stable w.r.t. $(\mathcal{F}_n)_{n\in\mathbb{N}}$, we have $\mathcal{U}^{ad} \subset \mathcal{U}^{ad}_n$: Problem (4) is a relaxation of the original Problem (3).

Replacing an almost sure constraint by a conditional expectation constraint is similar to an aggregation of constraints. For example 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 $\mathcal{B} = \{B_l\}_{l\in[\![1,|\mathcal{B}|]\!]}$ of Ω, and the σ-algebra $\mathcal{F}_\mathcal{B}$ generated by the partition $\mathcal{B}$. Assume that $\mathcal{C} = \{0\}$, then the relaxation presented consists in replacing the constraint $\theta(\boldsymbol{u}) = 0$ almost surely, which is equivalent to N constraints $\theta(\boldsymbol{u}(\omega_i)) = 0$ for $i \in [\![1, N]\!]$, by the collection of $|\mathcal{B}| \leq N$ (where $|\mathcal{B}|$ is the number of sets in the partition $\mathcal{B}$) constraints

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

3. Epiconvergence result

In this section we show the epiconvergence of the sequence of approximated cost functions $(\tilde{J}_n)_{n\in\mathbb{N}}$ toward J. We start with some useful recalls.

3.1. Preliminaries

Assume that $p \in [1, +\infty)$ and denote $q \in (1, +\infty]$ such that $1/q + 1/p = 1$. Recall that $\mathbb{V}$ is a separable Banach space with separable dual $\mathbb{V}^*$. We denote $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{V})$ and $L^q = L^q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{V}^*)$.

Convergence of random variables

A sequence $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ of L^p is said to converge strongly toward $\boldsymbol{X} \in L^p$, and denoted $\boldsymbol{X}_n \to_{L^p} \boldsymbol{X}$ if $\lim_{n\to\infty} \mathbb{E}\big[\|\boldsymbol{X}_n - \boldsymbol{X}\|^p_\mathbb{V}\big] = 0$. A sequence $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ of L^p is said to weakly converge toward $\boldsymbol{X} \in L^p$, and denoted $\boldsymbol{X}_n \rightharpoonup_{L^p} \boldsymbol{X}$ if for all $\boldsymbol{X}' \in L^q$, we have $\lim_{n\to\infty} \mathbb{E}\big[\langle \boldsymbol{X}_n - \boldsymbol{X}, \boldsymbol{X}'\rangle_{\mathbb{V},\mathbb{V}^*}\big] = 0$. For more details we refer the reader to Rudin [23].

Epiconvergence of functions

Let E be a topological space and consider a sequence $(A_n)_{n\in\mathbb{N}}$ of subsets of E. Then the inner limit of $(A_n)_{n\in\mathbb{N}}$, denoted $\underline{\lim}_n A_n$, is the set of accumulation points of any sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_n \in A_n$, and the outer limit of $(A_n)_{n\in\mathbb{N}}$ denoted $\overline{\lim}_n A_n$, is the set of accumulation points of any sub-sequence $(x_{n_k})_{k\in\mathbb{N}}$ of a sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_n \in A_n$. We say that $(A_n)_{n\in\mathbb{N}}$ converges toward A in the Painlevé–Kuratowski sense if $A = \overline{\lim}_n A_n = \underline{\lim}_n A_n$.

A sequence $(J_n)_{n\in\mathbb{N}}$ of functions taking value into $\mathbb{R} \cup \{+\infty\}$ is said to epi-converge toward a function J if the sequence of epigraphs of J_n converges toward the epigraph of J, in the Painlevé–Kuratowski sense. For more details and properties of epi-convergence, see Rockafellar and Wets [22] in finite dimension, and Attouch [3] for infinite dimension.

Convergences of σ-algebras

Let $\mathcal{F}$ be a σ-algebra and $(\mathcal{F}_n)_{n\in\mathbb{N}}$ a sequence of sub-fields of $\mathcal{F}$ (not necessarily finite nor a filtration). It is said that the sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$ Kudo-converges toward the σ-algebra $\mathcal{F}_\infty$, and denoted $\mathcal{F}_n \to \mathcal{F}_\infty$, if for each set $F \in \mathcal{F}$, $\Big(\mathbb{E}\big[1_F \mid \mathcal{F}_n\big]\Big)_{n\in\mathbb{N}}$ converges in probability toward $\mathbb{E}\big[1_F \mid \mathcal{F}_\infty\big]$.

It is shown by Kudo [11] that $\mathcal{F}_n \to \mathcal{F}_\infty$ if and only if for each integrable random variable $\boldsymbol{x}$, $\mathbb{E}\big[\boldsymbol{x} \mid \mathcal{F}_n\big]$ converges in L^1 toward $\mathbb{E}\big[\boldsymbol{x} \mid \mathcal{F}_\infty\big]$. Piccinini [18] extends this result to the convergence in L^p (where $p < +\infty$) in the strong or weak sense with the following lemma.

Lemma 1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-σ-algebras of $\mathcal{F}$. The following statements are equivalent:*

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

We have the following useful proposition where both the random variable and the σ-algebra are parametrized by n.

Proposition 2. *Assume that $\mathcal{F}_n \to \mathcal{F}_\infty$, 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 detailed in Piccinini [18]. We show the strong convergence case. If $\boldsymbol{X}_n \to_{L^p} \boldsymbol{X}$, then

$$\|\mathbb{E}\big[\boldsymbol{X}_n \mid \mathcal{F}_n\big] - \mathbb{E}\big[\boldsymbol{X} \mid \mathcal{F}\big]\|_{L^p} \le \|\mathbb{E}\big[\boldsymbol{X}_n \mid \mathcal{F}_n\big] - \mathbb{E}\big[\boldsymbol{X} \mid \mathcal{F}_n\big]\|_{L^p} + \|\mathbb{E}\big[\boldsymbol{X} \mid \mathcal{F}_n\big] - \mathbb{E}\big[\boldsymbol{X} \mid \mathcal{F}\big]\|_{L^p}$$

As the conditional expectation is a contraction and by Lemma 1 we have the result. □

We end with a few properties on the Kudo-convergence of σ-algebras (for more details we refer to Kudo [11] and Cotter [6]):

1. the topology associated with the Kudo-convergence is metrizable;
2. the set of σ-fields generated by the partitions of Ω is dense in the set of all σ-algebras;
3. if a sequence of random variables $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges in probability toward $\boldsymbol{x}$ and for all $n \in \mathbb{N}$ we have $\sigma(\boldsymbol{x}_n) \subset \sigma(\boldsymbol{x})$, then we have the Kudo-convergence of $\big(\sigma(\boldsymbol{x}_n)\big)_{n\in\mathbb{N}}$ toward $\sigma(\boldsymbol{X})$.

3.2. Main result

Denote τ the topology of $\mathcal{U}$, and recall that $\mathcal{V} = L^p$, with $p \in [1, \infty)$.

Theorem 3. *Let $\mathcal{V}$ be endowed with the strong or weak topology. Assume that $\mathcal{C}$ is closed and stable w.r.t. $(\mathcal{F}_n)_{n\in\mathbb{N}}$. If the two mappings Θ and J are continuous, and if $(\mathcal{F}_n)_{n\in\mathbb{N}}$ Kudo-converges toward $\mathcal{F}$, then $(\tilde{J}_n)_{n\in\mathbb{N}}$ (defined in (7)) epi-converges toward $\tilde{J}$ (defined in (8)).*

Note that $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is not assumed to be a filtration and that $\mathcal{F}_n$ is not assumed to be finite.

Proof. To prove the epi-convergence of $(\tilde{J}_n)_{n\in\mathbb{N}}$ toward $\tilde{J}$ it is sufficient to show that $\mathcal{U}^{ad}_n$ (defined in (6)) converges toward $\mathcal{U}^{ad}$ (defined in (5)) in the Painlevé–Kuratowski sense. Indeed it implies the epi-convergence of $(\mathbb{I}_{\mathcal{U}^{ad}_n})_{n\in\mathbb{N}}$ toward $\mathbb{I}_{\mathcal{U}^{ad}}$, and adding a continuous function preserves the epi-convergence (Attouch [3, Th 2.15]).

By stability of $\mathcal{C}$ w.r.t. $(\mathcal{F}_n)_{n\in\mathbb{N}}$ we have that, for all $n \in \mathbb{N}$, $\mathcal{U}^{ad} \subset \mathcal{U}^{ad}_n$ and thus $\mathcal{U}^{ad} \subset \liminf_n \mathcal{U}^{ad}_n$ (for any $x \in \mathcal{U}^{ad}$ take the constant sequence equal to x).

We now show that $\mathcal{U}^{ad} \supset \limsup_n \mathcal{U}^{ad}_n$. Let $\boldsymbol{u}$ be an element of $\limsup_n \mathcal{U}^{ad}_n$. By definition of outer-limit of sets, there exists a sequence $(\boldsymbol{u}_{n_k})_{k\in\mathbb{N}}$ that τ-converges to $\boldsymbol{u}$, such that for all $k \in \mathbb{N}$, $\mathbb{E}\big(\Theta(\boldsymbol{u}_{n_k})|\mathcal{F}_{n_k}\big) \in \mathcal{C}$. As Θ is continuous, we have $\Theta(\boldsymbol{u}_{n_k}) \to \Theta(\boldsymbol{u})$ strongly (resp. weakly) in L^p. Since $\mathcal{F}_{n_k} \to \mathcal{F}$, by Proposition 2,

$$\mathbb{E}\big(\Theta(\boldsymbol{u}_{n_k})|\mathcal{F}_{n_k}\big) \to_{L^p} \mathbb{E}\big(\Theta(\boldsymbol{u})|\mathcal{F}\big) = \Theta(U) \ .$$

Thus $\Theta(\boldsymbol{u})$ is the limit of a sequence in $\mathcal{C}$. By closedness of $\mathcal{C}$, we have that $\Theta(\boldsymbol{u}) \in -C$ and thus $\boldsymbol{u} \in \mathcal{U}^{ad}$. □

The practical consequences for the convergence of the approximation (4) toward the original (3) are given in the following Corollary.

Corollary 4. *Assume that $\mathcal{F}_n \to \mathcal{F}$, and that J and Θ are continuous. Then the sequence of Problems (4) approximates Problem (3) in the following sense. If $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ is a sequence of controls 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

$$\tilde{J}\big(\lim_k \boldsymbol{u}_{n_k}\big) = \min_{\boldsymbol{u}\in\mathcal{U}} \tilde{J}(\boldsymbol{u}) = \lim_k \tilde{J}_{n_k}\big(\boldsymbol{u}_{n_k}\big) \ .$$

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

Proof. The convergence result is a direct application of Attouch [3, 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 $\mathcal{U}^{ad}_n \subset \mathcal{U}^{ad}_m$. □

3.3. Dynamic problem

We cast Problem (3) into the following dynamic problem

$$\min_{\boldsymbol{u}\in\mathcal{U}} \quad J(\boldsymbol{u}) \ , \tag{9a}$$

$$s.t. \quad \Theta_t(\boldsymbol{u}_t) \in \mathcal{C}_t \qquad \forall t \in [\![1, T]\!] \ , \tag{9b}$$

$$\boldsymbol{u}_t \preceq \mathcal{F}_t \ , \tag{9c}$$

where $\boldsymbol{u}_t \preceq \mathcal{F}_t$ stands for "$\boldsymbol{u}_t$ is $\mathcal{F}_t$-measurable". Here $\boldsymbol{u}$ is a stochastic process of control $(\boldsymbol{u}_t)_{t\in[\![1,T]\!]}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with value in a space $\mathbb{U}$. We have T constraints operators Θ_t taking values in $L^p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{V}_t)$, where $(\mathcal{F}_t)_{t\in[\![1,T]\!]}$ is a sequence of σ-algebra. Note that $(\mathcal{F}_t)_{t\in[\![1,T]\!]}$ is not necessarily a filtration. Then, for each $t \in [\![1, T]\!]$ we define a sequence of approximating σ-algebra $(\mathcal{F}_{n,t})_{n\in\mathbb{N}}$. For all $t \in [\![1, T]\!]$, $\mathcal{C}_t$ is a closed convex cone stable w.r.t. $\big(\mathcal{F}_{n,t}\big)_{n\in\mathbb{N}}$. The interaction between the different time-step is integrated into the objective function J (usually a sum over time).

Finally, we consider the sequence of approximated problem

$$\min \quad J(\boldsymbol{u}) \ , \tag{10a}$$

$$s.t. \quad \mathbb{E}\big[\Theta(\boldsymbol{u}_t) \mid \mathcal{F}_{n,t}\big] \in \mathcal{C}_t \qquad \forall t \in [\![1, T]\!] \ . \tag{10b}$$

Furthermore we denote

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

and

$$\mathcal{U}_{n,t}^{ad} := \left\{ \boldsymbol{u}_t \in \mathcal{U}_t \;\middle|\; \mathbb{E}\left[\Theta(\boldsymbol{u}_t) \;\middle|\; \mathcal{F}_{n,t}\right] \in -C_t \right\} .$$

We define the set of admissible controls 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

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

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

Theorem 5. *Let $\mathcal{U}$ be endowed with a product 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 $[1, \infty)$). 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 $\big(\tilde{J}_n\big)_{n\in\mathbb{N}}$ epi-converges to $\tilde{J}$.*

Proof. The proof is deduced from the one of Theorem 3. 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 of continuous operators

The continuity of J and Θ as operators required in Theorem 3 is an abstract assumption. This section presents conditions for some classical constraint and objective functions to be representable by continuous operators. Before presenting those results we prove a technical lemma that allows us to prove convergence for the topology of convergence in probability by considering sequences of random variables converging almost surely.

4.1. A technical lemma

Lemma 6. *Let $\Theta : E \to F$, where $(E, \tau_{\mathbb{P}})$ is a space of random variables endowed with the topology of convergence in probability, and (F, τ) 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 a continuous operator from $(E, \tau_{\mathbb{P}})$ into (F, τ).*

Proof. Recall that if $(x_n)_{n\in\mathbb{N}}$ is a sequence in a topological space, such that from any sub-sequence $\big(x_{n_k}\big)_{k\in\mathbb{N}}$ we can extract a sub-sub-sequence $\big(x_{\sigma(n_k)}\big)_{k\in\mathbb{N}}$ 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 sub-sequence $\big(x_{n_k}\big)_{k\in\mathbb{N}}$ such that for all $k \in \mathbb{N}$, $x_{n_k} \notin \mathcal{O}$, and no sub-sub-sequence can converges to x^*, hence 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 F. We choose a sub-sequence $\big(\Theta\big(\boldsymbol{u}_{n_k}\big)\big)_{k\in\mathbb{N}}$. By assumption $\big(\boldsymbol{u}_n\big)_{n\in\mathbb{N}}$ converges in probability toward $\boldsymbol{u}$, thus we have $\boldsymbol{u}_{n_k} \to_{\mathbb{P}} \boldsymbol{u}$. Consequently there exists a sub-sub-sequence $\boldsymbol{u}_{\sigma(n_k)}$ 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. □

Remark 1. This lemma does not imply the equivalence between convergence almost sure and convergence in probability as one cannot endow $\mathcal{U}$ with the "topology of almost sure convergence" as almost sure convergence is not generally induced by a topology.

However note that $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ converges in probability toward $\boldsymbol{u}$ iff from any sub-sequence of $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ we can extract a further sub-sequence converging almost surely to $\boldsymbol{u}$ (see [8, Th 2.3.2]).

4.2. Objective function

Let $\mathcal{U}$ be a space of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, with value in a Banach space $\mathbb{U}$.

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 measurable, bounded cost function. This objective function expresses a risk-neutral attitude; indeed a random cost with high variance or a deterministic cost with the same expectation are considered equivalent. Recently in order to capture risk-averse attitudes, coherent risk measures (as defined in Artzner et al. [2]), or more generally convex risk measures (as defined in Föllmer and Schied [9]), have been prominent in the literature.

Following Ruszczyński and Shapiro [25], we call *convex risk measure* an operator $\rho : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ verifying

- Convexity: for all $\lambda \in [0, 1]$ and all $X, Y \in \mathcal{X}$, we have
$$\rho\big(\lambda X + (1-\lambda)Y\big) \le \lambda\rho\big(X\big) + (1-\lambda)\rho\big(Y\big) ;$$
- Monotonicity: for all $X, Y \in \mathcal{X}$ such that $X \le Y$ we have $\rho(X) \le \rho(Y)$;
- Translation equivariance: for all constant $c \in \mathbb{R}$ and all $X \in \mathcal{X}$, we have $\rho(X + c) = \rho(X) + c$,

where $\mathcal{X}$ is a linear space of measurable functions. We focus on the case where $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R})$.

Proposition 7. *Let $\mathcal{U}$ be a set of random variables 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 proper lower semi-continuous convex risk measure. Then, $J : \mathcal{U} \to \mathbb{R}$ is continuous.*

Proof. Note that as j is bounded, $j(\boldsymbol{u}) \in \mathcal{X}$ for any $u \in \mathcal{U}$. Then we know that [25] there is a convex set of probabilities $\mathcal{P}$ such that

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

where g is convex and weak*-lowersemicontinuous on the space of finite signed measures on $(\Omega, \mathcal{F})$. Moreover any probability in $\mathcal{P}$ is absolutely continuous w.r.t. $\mathbb{P}$.

Consider a sequence $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ of elements 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 monotonicity of ρ. By definition of ρ, for all $\varepsilon > 0$ there is a probability $\mathbb{P}_\varepsilon \in \mathcal{P}$ such that

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

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

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

Let η be a positive real, and set $\varepsilon = \eta/2$, and $N \in \mathbb{N}$ such that for all $n \ge N$,

$$|\mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{u}_n)\big) - \mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{u})\big)| \le \frac{\eta}{2} . \tag{11}$$

Then, recalling that

$$\rho\Big(j(\boldsymbol{u})\Big) \geq \mathbb{E}_{\mathbb{P}_{\frac{\eta}{2}}}\big(j(\boldsymbol{u})\big) - g(\mathbb{P}_{\frac{\eta}{2}}) \geq \rho\big(j(\boldsymbol{u})\big) - \frac{\eta}{2}\,, \tag{12}$$

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

$$\begin{aligned}\rho\big(j(\boldsymbol{u}_n)\big) &= \sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}_{\mathbb{Q}}\Big(j(\boldsymbol{u}_n)\Big) - g(\mathbb{Q}) \\ &\geq \mathbb{E}_{\mathbb{P}_{\frac{\eta}{2}}}\Big(j(\boldsymbol{u}_n)\Big) - g\big(\mathbb{P}_{\frac{\eta}{2}}\big) \\ &\geq \mathbb{E}_{\mathbb{P}_{\frac{\eta}{2}}}\Big(j(\boldsymbol{u})\Big) - g\big(\mathbb{P}_{\frac{\eta}{2}}\big) && \text{by (11)}, \\ &\geq \rho\Big(j(\boldsymbol{u})\Big) - \eta && \text{by (12)},\end{aligned}$$

and thus

$$\rho\Big(j(\boldsymbol{u})\Big) + \frac{\eta}{2} \geq \rho\Big(j(\boldsymbol{u}_n)\Big) \geq \rho\Big(j(\boldsymbol{u})\Big) - \eta\,.$$

Thus $\lim_n \rho\big(j(\boldsymbol{u}_n)\big) = \rho\big(j(\boldsymbol{u})\big)$. Hence the continuity of J. □

The assumptions of this proposition can be relaxed in different ways.

In the first place, if the convex risk measure ρ is simply the expectation then we can simply endow $\mathcal{U}$ with the topology of convergence in law. In this case 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 $\mathbb{E}\big[j(\boldsymbol{u})\big]$.

Otherwise assume that $\mathcal{U}$ is a set of random variables endowed with the topology of convergence in probability and that j continuous. Moreover, if we can ensure that $j(\boldsymbol{u})$ is dominated by some integrable (for all probability of $\mathcal{P}$) random variable, then $J : \mathcal{U} \to \mathbb{R}$ is continuous. Indeed we consider a sequence $\big(\boldsymbol{u}_n\big)_{n\in\mathbb{N}}$ almost surely converging to $\boldsymbol{u}$. We modify the proof of Proposition 7 by using a dominated convergence theorem to show that $\lim_n \mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{u}_n)\big) = \mathbb{E}_{\mathbb{P}_\varepsilon}\big(j(\boldsymbol{u})\big)$, and end with Lemma 6.

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 $\mathcal{V}$.

4.3.1. Almost sure constraint

From Lemma 6, we obtain a first important example of continuous constraints, which can also be obtained and extended from results on Nemytskii operators (see, e.g. [1]).

Proposition 8. *Suppose that $\mathcal{U}$ is the set of random variables on $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$, with value in $\mathbb{U}$, endowed with the topology of convergence in probability. Assume that $\theta : \mathbb{U} \to \mathbb{V}$ is continuous and bounded. Then the operator $\Theta\big(\boldsymbol{u}\big)(\omega) := \theta\big(\boldsymbol{u}(\omega)\big)$ maps $\mathcal{U}$ into $\mathcal{V}$ and is continuous.*

Proof. The function θ being continuous, is also 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 $\mathcal{F}$-measurable. Boundedness of θ ensure the existence of moment of all order of $\Theta(\boldsymbol{u})$. Thus Θ is well defined.

Suppose that $\big(\boldsymbol{u}_n\big)_{n\in\mathbb{N}}$ converges to $\boldsymbol{u}$ almost surely. Then by boundedness 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

$$\lim_{n\to\infty} \theta\big(\boldsymbol{u}_n\big) = \theta\big(\boldsymbol{u}\big) \quad \text{in } L^p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{V})\,,$$

which is exactly

$$\lim_{n\to\infty} \Theta\big(\boldsymbol{u}_n\big) = \Theta\big(\boldsymbol{u}\big)\,.$$

Consequently by Lemma 6 we have the continuity of Θ. □

We note that boundedness 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 9. *Let $\mathcal{B}$ be a sub-field of $\mathcal{F}$. If $\mathcal{U} = L^{p'}\big(\Omega, \mathcal{B}, \mathbb{P}\big)$, with the topology of convergence in probability, and if θ is γ-Hölder, with $\gamma \leq p'/p$ then $\Theta\big(\boldsymbol{u}\big)(\omega) := \theta\big(\boldsymbol{u}(\omega)\big)$ is well defined and continuous as an operator mapping $\mathcal{U}$ into $\mathcal{V}$.*

Proof. By definition a function θ mapping $\mathbb{U}$ into $\mathbb{V}$ is γ-Hölder if there exists a constant $M > 0$ such that for all u, u' in $\mathbb{U}$ we have

$$\big\|\theta(u) - \theta(u')\big\|_{\mathbb{V}} \leq M\, \big\|u - u'\big\|_{\mathbb{U}}^{\gamma}\,,$$

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

Following the previous proof we just have to check 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. The Hölder assumption implies

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

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

4.3.2. Measurability constraint

When considering a dynamic stochastic optimization problem, measurability constraints are used to represent the nonanticipativity constraints. They can be expressed by stating that a random variable and its conditional expectation are equal.

Proposition 10. *We set $\mathcal{U} = L^{p'}\big(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{U}\big)$, with $p' \geq p$. Assume that*

- *either $\mathcal{U}$ is equipped with the strong topology, and $\mathcal{V}$ is equipped with the strong or weak topology,*
- *or $\mathcal{U}$ and $\mathcal{V}$ are equipped with the weak topology.*

If $\mathcal{B}$ is a sub-field of $\mathcal{F}$, then $\Theta\big(\boldsymbol{u}\big) := \mathbb{E}\big[\boldsymbol{u} \bigm| \mathcal{B}\big] - \boldsymbol{u}$, is well defined and continuous.

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 $\mathbb{E}\big[\boldsymbol{v} \bigm| \mathcal{B}\big] \in \mathcal{V}$ 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 $\mathcal{U}$ strongly converging in $L^{p'}$ toward $\boldsymbol{u} \in \mathcal{U}$. We have

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

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 $\boldsymbol{y} \in L^q$,

$$\begin{aligned}\mathbb{E}\Big[\mathbb{E}\big[\boldsymbol{u}_n \bigm| \mathcal{B}\big] \cdot Y\Big] &= \mathbb{E}\Big[\boldsymbol{u}_n \mathbb{E}\big[Y \bigm| \mathcal{B}\big]\Big]\,, \\ &\underset{n}{\longrightarrow} \mathbb{E}\Big[\boldsymbol{u}\mathbb{E}\big[Y \bigm| \mathcal{B}\big]\Big]\,, \\ &= \mathbb{E}\Big[\mathbb{E}\big[\boldsymbol{u} \bigm| \mathcal{B}\big]Y\Big]\,.\end{aligned}$$

Thus we have the weak convergence of the conditional expectation and therefore of Θ. Finally, as the strong convergence implies 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. If we endow $\mathcal{U}$ with the topology of convergence in probability in the previous proposition we will obtain continuity of Θ on a subset of $\mathcal{U}$. Indeed if a set of random variables $\mathcal{U}^{ad}$ such that there exists a random variable in $L^{p'}(\Omega,\mathcal{F},\mathbb{P})$ dominating every random variable in $\mathcal{U}^{ad}$, then a sequence converging almost surely will converge for the $L^{p'}$ norm and we can follow the previous proof to show the continuity of Θ on $\mathcal{U}^{ad}$.

4.3.3. Risk constraints

Risk attitude can be expressed through the objective function or through constraints. We have seen that a risk measure can be chosen as the objective function, we now show that conditional risk measure can used as constraints.

Let ρ be a conditional risk mapping as defined in Ruszczyński and Shapiro [24], and more precisely ρ maps $\mathcal{U}$ into $\mathcal{V}$ where $\mathcal{U}=L^p(\Omega,\mathcal{F},\mathbb{P};\mathbb{U})$ and $\mathcal{V}=L^p(\Omega,\mathcal{B},\mathbb{P};\mathbb{V})$, with $\mathcal{B}\subset\mathcal{F}$, and verifies the following properties

- Convexity: for all $\lambda\in\mathcal{U}$, $\lambda\in[0,1]$ and all $X,Y\in\mathcal{V}$, we have
$$\rho\big(\lambda X+(1-\lambda)Y\big)\leq\lambda\rho\big(X\big)+(1-\lambda)\rho\big(Y\big)\,;$$
- Monotonicity: for all $X,Y\in\mathcal{V}$ such that $X\leq Y$ we have $\rho(X)\leq\rho(Y)$;
- Translation equivariance: for all $c\in\mathcal{V}$ and all $X\in\mathcal{U}$, we have $\rho(X+c)=\rho(X)+c$.

Proposition 11. *Let $\mathcal{U}$ be endowed with the topology of convergence in probability, and $\mathcal{V}$ endowed with the strong topology. If ρ is a conditional risk mapping, θ is a continuous bounded cost function mapping $\mathbb{U}$ into $\mathbb{R}$, and $a\in\mathcal{V}$, then $\Theta\big(\boldsymbol{u}\big):=\rho\Big(\theta\big(\boldsymbol{u}\big)\Big)-a$ is continuous.*

Proof. Consider a sequence of random variables $\big(\boldsymbol{u}_n\big)_{n\in\mathbb{N}}$ converging in probability toward $\boldsymbol{u}_\infty$. Let $\pi\ :\ L^p(\Omega,\mathcal{B},\mathbb{P};\mathbb{U})\to L^p(\Omega,\mathcal{B},\mathbb{P};\mathbb{U})$ be a selector of $\mathcal{V}=L^p(\Omega,\mathcal{B},\mathbb{P};\mathbb{U})$, i.e. for any $\boldsymbol{x}\in L^p(\Omega,\mathcal{F},\mathbb{P};\mathbb{U})$, $\pi(\boldsymbol{X})\in\boldsymbol{x}$. For any $\omega\in\Omega$, any $\boldsymbol{x}\in L^p(\Omega,\mathcal{F},\mathbb{P};\mathbb{U})$ we define

$$\rho_\omega(\boldsymbol{u}):=\pi\big(\rho\big(\boldsymbol{u}\big)\big)(\omega)\ .$$

Note that for $\mathbb{P}$-almost all $\omega\in\Omega$, the function $\Theta_\omega(\boldsymbol{u}):=\rho_\omega\big(\theta(\boldsymbol{u})\big)$, satisfies the conditions of Proposition 7. Thus for $\mathbb{P}$-almost all $\omega\in\Omega$, $\big(\Theta_\omega(\boldsymbol{u}_n)\big)_{n\in\mathbb{N}}$ converges toward $\Theta_\omega(\boldsymbol{u}_\infty)$. Thus we have shown that $\big(\Theta(\boldsymbol{u}_n)\big)_{n\in\mathbb{N}}$ converges almost surely toward $\Theta\big(\boldsymbol{u}_\infty\big)$. By boundedness of θ and monotonicity of ρ we obtain the boundedness of $\big(\Theta(\boldsymbol{u}_n)\big)_{n\in\mathbb{N}}$. Thus almost sure convergence and dominated convergence theorem ensure that $\big(\Theta(\boldsymbol{u}_n)\big)_{n\in\mathbb{N}}$ converges in L^p toward $\Theta\big(\boldsymbol{u}_\infty\big)$, hence the continuity of Θ. □

Another widely used risk measure, even if it has some serious drawbacks, is the Value-at-Risk. If $\boldsymbol{X}$ is a real random variable its 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 12. *If $\theta:\mathbb{U}\to\mathbb{R}$ is continuous, and if $\mathcal{U}$ is such that every $\boldsymbol{u}\in\mathcal{U}$ has a continuous distribution function, then $\Theta(\boldsymbol{u}):=VaR_\alpha\Big(\theta\big(\boldsymbol{u}\big)\Big)$ 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, then $\big(\theta\big(\boldsymbol{u}_n\big)\big)_{n\in\mathbb{N}}$ converges in law toward $\theta\big(\boldsymbol{u}\big)$ and we have, for all $x\in\mathbb{R}$, $F_{\theta(\boldsymbol{u}_n)}(x)\to F_{\theta(\boldsymbol{u})}(x)$. Thus $\big(\Theta(\boldsymbol{u}_n)\big)_{n\in\mathbb{N}}$ converges almost surely toward $\Theta(\boldsymbol{u})$, and as $\Theta(\boldsymbol{u})$ is deterministic, Θ is continuous. □

Note that in Proposition 12 the constraint function takes deterministic values. Thus considering the conditional expectation of this constraint yields exactly the same constraint. However consider a constraint $\Theta_1\ :\ \mathcal{U}\to\mathbb{R}$ of this form, and another constraint $\Theta_2:\mathcal{U}\to\mathcal{V}$. Then if Θ_1 and Θ_2 are continuous, then so is the constraint $\Theta=(\Theta_1,\Theta_2)\to\mathbb{R}\times\mathcal{V}$. Thus we can apply Theorem 3 on the coupled constraint.

5. Dual approximate dynamic programming

In this section, we say a few words about how the approximation of an almost sure constraint by a conditional expectation – as presented in Section 3 – can be used. More details and numerical experiment of this algorithm can be found in Barty et al. [4], Leclère [13], Carpentier et al. [5], Ramakrishnan and Luedtke [19].

5.1. Presentation of the problem

We are interested in an electricity production problem with N power stations coupled by an equality constraint. At time step t, each power station i has an internal state $\boldsymbol{X}_t^i$, and is affected by a random exogenous noise $\boldsymbol{\xi}_t^i$. For each power station, and each time step t, we have a control $\boldsymbol{q}_t^i\in\mathcal{Q}_{t,i}^{ad}$ that must be measurable with respect to $\mathcal{F}_t$ where $\mathcal{F}_t$ is the σ-algebra generated by all past noises: $\mathcal{F}_t=\sigma\big(\boldsymbol{\xi}_s^i\big)_{1\leq i\leq n,0\leq s\leq t}$. Moreover, there is a coupling constraint expressing that the total production must be equal to the demand. This constraint is represented as $\sum_{i=1}^N\theta_t^i(\boldsymbol{q}_t^i)=0$, where θ_t^i is a continuous bounded function from $\mathbb{Q}_{t,i}^{ad}$ 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 $L_t^i\big(\boldsymbol{x}_t^i,\boldsymbol{q}_t^i,\boldsymbol{\xi}_t^i\big)$.

Finally the problem reads

$$\min_{\boldsymbol{x},\boldsymbol{q}}\quad\mathbb{E}\Bigg[\sum_{i=1}^N\sum_{t=0}^T L_t^i\big(\boldsymbol{x}_t^i,\boldsymbol{q}_t^i,\boldsymbol{\xi}_t^i\big)\Bigg]\tag{13a}$$
$$s.t.\quad \boldsymbol{x}_{t+1}^i=f_t^i(\boldsymbol{x}_t^i,\boldsymbol{q}_t^i,\boldsymbol{\xi}_t^i)\qquad\forall t,\quad\forall i,\tag{13b}$$
$$\boldsymbol{x}_0^i=x_0^i\qquad\forall i,\tag{13c}$$
$$\boldsymbol{q}_t^i\in\mathcal{Q}_{t,i}^{ad}\qquad\forall t,\quad\forall i,\tag{13d}$$
$$\boldsymbol{q}_t^i\preceq\mathcal{F}_t\qquad\forall t,\quad\forall i,\tag{13e}$$
$$\sum_{i=1}^N\theta_t^i(\boldsymbol{q}_t^i)=0\qquad\forall t,\quad\forall i.\tag{13f}$$

For the sake of brevity, we denote by $\mathcal{A}$ the set of random processes $(\boldsymbol{X},\boldsymbol{q})$ verifying constraints (13b)–(13d).

Let us assume that all random variables are in L^2 spaces and dualize the coupling constraint (13f). We do not study here the relation between the primal and the following dual problem (see [20,21] for an alternative formulation involving duality between L^1 and its dual).

$$\max_{\boldsymbol{\lambda}\in L^2}\quad\min_{(\boldsymbol{x},\boldsymbol{q})\in\mathcal{A}}\quad\mathbb{E}\Bigg[\sum_{i=1}^N\sum_{t=0}^T L_t^i\big(\boldsymbol{x}_t^i,\boldsymbol{q}_t^i,\boldsymbol{\xi}_t^i\big)+\boldsymbol{\lambda}_t\theta_t^i(\boldsymbol{q}_t^i)\Bigg]\tag{14a}$$
$$s.t.\qquad\boldsymbol{q}_t^i\preceq\mathcal{F}_t\qquad\forall t,\quad\forall i.\tag{14b}$$

Note that, for fixed $\boldsymbol{\lambda}$, the inner minimization problem is decomposable. Thus for a fixed $\boldsymbol{\lambda}^{(k)}$ we have to solve N problems of smaller size than Problem (14), $\boldsymbol{\lambda}^{(k)}$ being updated in a gradient-like scheme.

$$(\mathcal{P})\qquad\min_{(\boldsymbol{x},\boldsymbol{u})\in\mathcal{A}}\quad\mathbb{E}\Bigg[\sum_{t=0}^T L_t^i\big(\boldsymbol{x}_t^i,\boldsymbol{q}_t^i,\boldsymbol{\xi}_t^i\big)+\boldsymbol{\lambda}_t^{(k)}\theta_t^i(\boldsymbol{q}_t^i)\Bigg]\tag{15a}$$
$$s.t.\qquad\boldsymbol{q}_t^i\preceq\mathcal{F}_t\qquad\forall t,\quad\forall i.\tag{15b}$$

Note that the process $\boldsymbol{\lambda}^{(k)}$ has no given dynamics but can be chosen to be adapted to the filtration $(\mathcal{F}_t)_{t=1,\dots,T}$. Consequently solving Problem (15) by Dynamic Programming is possible but numerically difficult as we need to keep all the past realizations of the noises in the state. In fact, the so-called curse of dimensionality prevents us to solve numerically this problem.

Nevertheless it has been proposed in Barty et al. [4] to replace $\boldsymbol{\lambda}_t$ by $\mathbb{E}\big[\boldsymbol{\lambda}_t \mid \boldsymbol{Y}_t\big]$, where $\boldsymbol{Y}_t$ is a random variable measurable with respect to $(\boldsymbol{y}_{t-1}, \boldsymbol{\xi}_t)$ instead of $\boldsymbol{\lambda}_t$. This is similar to a decision rule approach for the dual as we are restraining the control to a certain class, the $\boldsymbol{Y}_t$-measurable $\boldsymbol{\lambda}$ in our case. Thus Problem (15) can be solved by Dynamic Programming with the augmented state $(\boldsymbol{x}_t^i, \boldsymbol{y}_t)$. It has also been shown that, under some non-trivial conditions, replacing $\boldsymbol{\lambda}_t$ by its conditional expectation $\mathbb{E}\big[\boldsymbol{\lambda}_t \mid \boldsymbol{Y}_t\big]$ is equivalent to solving

$$\min_{(\boldsymbol{x},\boldsymbol{q})\in\mathcal{A}} \quad \mathbb{E}\Big[\sum_{i=1}^N \sum_{t=0}^T L_t^i\big(\boldsymbol{x}_t^i, \boldsymbol{q}_t^i, \boldsymbol{\xi}_t^i\big)\Big] \tag{16a}$$

$$s.t. \qquad \boldsymbol{q}_t^i \preceq \mathcal{F}_t \qquad \forall t, \quad \forall i, \tag{16b}$$

$$\mathbb{E}\Big[\sum_{i=1}^N \theta_t^i(\boldsymbol{q}_t^i) \;\Big|\; Y_t\Big] = 0 \qquad \forall t, \quad \forall i. \tag{16c}$$

Problem (16) is a relaxation of Problem (13) where the almost sure constraint (13f) is replaced by the constraint (16c). Now consider a sequence of information processes $(\boldsymbol{Y}^{(n)})_{n\in\mathbb{N}}$ each generating a σ-algebra $\mathcal{F}_n$, and their associated relaxation $(\mathcal{P}_n)$ (as specified in Problem (16)) of Problem (13) (denoted $(\mathcal{P})$). Those problems correspond to Problems (9) and (10) with $J(\boldsymbol{u}) = \mathbb{E}\Big[\sum_{i=1}^N \sum_{t=0}^T L_t^i\big(\boldsymbol{x}_t^i, \boldsymbol{q}_t^i, \boldsymbol{\xi}_t^i\big)\Big]$, where $\boldsymbol{u} = (\boldsymbol{q}^{(i)})_{i\in[\![1,N]\!]}$ and $\boldsymbol{x}_t^i$ follow the dynamic equation (13b). We also have $\Theta_t(\boldsymbol{u}_t) = \sum_{i=1}^N \theta_t^i(\boldsymbol{q}_t^i)$ and $C_t = \{0\}$.

Assume that for all $t \in [\![1,T]\!]$, and all $i \in [\![1,N]\!]$ the cost functions L_t^i, dynamic functions f_t and constraint functions θ_t^i are continuous, and that $\mathcal{Q}_{t,i}^{ad}$ is a compact subset of a Euclidean space. Moreover we assume that the noise variables $\boldsymbol{\xi}_t^i$ 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 $L_t^i\big(\boldsymbol{x}_t^i, \boldsymbol{u}_t^i, \boldsymbol{\xi}_t^i\big)$. Thus the cost function can be effectively replaced by bounded functions. Consequently Proposition 7 ensures that J is continuous if $\mathcal{U}$ is equipped with the topology of convergence in probability. Similarly Proposition 8 ensures that Θ is continuous. Theorem 5 implies that our sequence of approximated problems $(\mathcal{P}_n)$ converges toward the initial problem $(\mathcal{P})$. Thus, let $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ be a sequence of ε_n-optimal solution of $\mathcal{P}_n$, i.e. $\boldsymbol{u}_n$ verifying constraint (16c) and $J(\boldsymbol{u}_n) < \inf_{\boldsymbol{u}\in\mathcal{U}_n^{ad}} J(\boldsymbol{u}) + \varepsilon_n$, 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 $(\mathcal{P})$.

Remark 2. To get an idea of the numerical interest of such an approach fix all discretization (in space, control, time and number of units) to 10, frontal dynamic programming requires 10^{31} operations, whereas, in the decomposed approach, each subgradient iteration requires only 10^6 iterations. The subgradient method being applied in $\mathbb{R}^{10}$ requires a few thousand iterations to give a reasonable solution, hence the approximated problem can be solved in around 10^{10} operations.

Acknowledgments

The author would like to thank anonymous referees and R. Wets for their insightful comments on previous version of this work. This research benefited from the support of the FMJH 'Program Gaspard Monge for optimization and operations research and their interactions with data science', and from the support from EDF.

References

[1] J. Appell, P.P. Zabrejko, Nonlinear Superposition Operators, in: Cambridge Tracts in Mathematics, vol. 95, Cambridge University Press, Cambridge, 1990, viii+311.
[2] P. Artzner, F. Delbaen, J. Eber, D. Heath, Coherent measures of risk, Math. Finance 9 (1999) 203–228.
[3] H. Attouch, Variational Convergence for Functions and Operators, Vol. 1, Pitman Advanced Pub. Program, 1984.
[4] K. Barty, P. Carpentier, P. Girardeau, Decomposition of large-scale stochastic optimal control problems, RAIRO Rech. Opér. 44 (3) (2010) 167–183.
[5] P. Carpentier, J.P. Chancelier, V. Leclère, F. Pacaud, Stochastic decomposition applied to large-scale hydro valleys management, European J. Oper. Res. 270 (3) (2018) 1086–1098.
[6] K.D. Cotter, Similarity of information and behavior with a pointwise convergence topology, J. Math. Econ. 15 (1) (1986) 25–38.
[7] J. Dupacová, R. Wets, Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems, Ann. Statist. 16 (4) (1988).
[8] R. Durrett, R. Durrett, Probability: Theory and Examples, Cambridge Univ Pr, 2010.
[9] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance Stoch. 6 (2002) 429–447.
[10] A. King, R. Wets, Epi-consistency of convex stochastic programs, Stoch. Stoch. Rep. 34 (1–2) (1991).
[11] H. Kudo, A note on the strong convergence of σ-algebras, Ann. Probab. 2 (1) (1974) 76–83.
[12] D. Kuhn, W. Wiesemann, A. Georghiou, Primal and dual linear decision rules in stochastic and robust optimization, Math. Program. 130 (1) (2011).
[13] V. Leclère, Contributions to Decomposition Methods in Stochastic Optimization (Ph.D. thesis), Université Paris-Est, 2014.
[14] F. Pacaud, P. Carpentier, J.-P. Chancelier, M. De Lara, Stochastic optimal control of a domestic microgrid equipped with solar panel and battery, 2018, arXiv preprint arXiv:1801.06479.
[15] T. Pennanen, Epi-convergent discretizations of multistage stochastic programs, Math. Oper. Res. 30 (1) (2005) 245–256.
[16] T. Pennanen, Epi-convergent discretizations of multistage stochastic programs via integration quadratures, Math. Program. 116 (1–2) (2009) 461–479.
[17] T. Pennanen, M. Koivu, Epi-convergent discretizations of stochastic programs via integration quadratures, Numer. Math. 100 (1) (2005) 141–163.
[18] L. Piccinini, A new version of the multivalued Fatou lemma, J. Appl. Anal. 4 (2) (1998).
[19] J. Ramakrishnan, J. Luedtke, A dual approximate dynamic programming approach to multi-stage stochastic unit commitment, 2018, arXiv preprint arXiv:1801.02259.
[20] R. Rockafellar, R. Wets, Measures as Lagrange multipliers in multistage stochastic programming, J. Math. Anal. Appl. 60 (2) (1977).
[21] R.T. Rockafellar, R.J. Wets, The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints, SIAM J. Control Optim. 16 (1) (1978).
[22] R.T. Rockafellar, R. Wets, Variational Analysis, Vol. 317, Springer Verlag, 1998.
[23] W. Rudin, Functional Analysis, in: International Series in Pure and Applied Mathematics, McGraw-Hill, New York, 1991.
[24] A. Ruszczyński, A. Shapiro, Conditional risk mappings, Math. Oper. Res. (2006) 544–561.
[25] A. Ruszczyński, A. Shapiro, Optimization of convex risk functions, Math. Oper. Res. (2006) 433–452.
[26] A. Shapiro, D. Dentcheva, A. Ruszczyński, Lectures on Stochastic Programming: Modeling and Theory, Vol. 9, Society for Industrial and Applied Mathematics, 2009.
[27] M. Zervos, On the epiconvergence of stochastic optimization problems, Math. Oper. Res. 24 (2) (1999).