CONVERGENCE ANALYSIS OF SAMPLING-BASED DECOMPOSITION METHODS FOR RISK-AVERSE MULTISTAGE STOCHASTIC CONVEX PROGRAMS

VINCENT GUIGUES

arXiv:1408.4439v3 [math.OC] 5 Feb 2016

Abstract. We consider a class of sampling-based decomposition methods to solve risk-averse multistage stochastic convex programs. We prove a formula for the computation of the cuts necessary to build the outer linearizations of the recourse functions. This formula can be used to obtain an efficient implementation of Stochastic Dual Dynamic Programming applied to convex nonlinear problems. We prove the almost sure convergence of these decomposition methods when the relatively complete recourse assumption holds. We also prove the almost sure convergence of these algorithms when applied to risk-averse multistage stochastic linear programs that do not satisfy the relatively complete recourse assumption. The analysis is first done assuming the underlying stochastic process is interstage independent and discrete, with a finite set of possible realizations at each stage. We then indicate two ways of extending the methods and convergence analysis to the case when the process is interstage dependent.

AMS subject classifications: 90C15, 90C90.

1. Introduction

Multistage stochastic convex optimization models have become a standard tool to deal with a wide range of engineering problems in which one has to make a sequence of decisions, subject to random costs and constraints, that arise from observations of a stochastic process. Decomposition methods are popular solution methods to solve such problems. These algorithms are based on dynamic programming equations and build outer linearizations of the recourse functions, assuming that the realizations of the stochastic process over the optimization period can be represented by a finite scenario tree. Exact decomposition methods such as the Nested Decomposition (ND) algorithm [2], [3], compute cuts at each iteration for the recourse functions at all the nodes of the scenario tree. However, in some applications, the number of scenarios may become so large that these exact methods entail prohibitive computational effort.

Monte Carlo sampling-based algorithms constitute an interesting alternative in such situations. For multistage stochastic linear programs (MSLP) whose number of immediate descendant nodes is small but with many stages, Pereira and Pinto [14] propose to sample in the forward pass of the ND. This sampling-based variant of the ND is the so-called Stochastic Dual Dynamic Programming (SDDP) algorithm, which has been the object of several recent improvements and extensions [22], [15], [9], [10], [8], [12].

In this paper, we are interested in the convergence of SDDP and related algorithms for risk-averse multistage stochastic convex programs (MSCP). A convergence proof of an enhanced variant of SDDP, the Cutting-Plane and Partial-Sampling (CUPPS) algorithm, was given in [5] for risk-neutral multistage stochastic linear programs with uncertainty in the right-hand side only. For this type of problems, the proof was later extended to a larger class of algorithms in [13], [17]. These proofs are directly applicable to show the convergence of SDDP applied to the risk-averse models introduced in [9]. Finally, more recently, Girardeau et al. proved the convergence of a class of sampling-based decomposition methods to solve some risk-neutral multistage stochastic convex programs [7]. We extend this latter analysis in several ways:

(A) The model is risk-averse, based on dynamic programming equations expressed in terms of conditional coherent risk functionals.

(B) Instead of using abstract sets, the dynamic constraints are expressed using equality and inequality constraints, a formulation needed when the algorithm is implemented for a real-life application. Regarding the problem formulation, the dynamic constraints also depend on the full history of decisions instead of just the previous decision. As a result, the recourse functions also depend on the

Key words and phrases. Stochastic programming and Risk-averse optimization and Decomposition algorithms and Monte Carlo sampling and Relatively complete recourse and SDDP.

the full history of decisions and the formulas of the optimality cuts for these functions and of the feasibility cuts built by the traditional implementation of SDDP (where recourse functions depend on the previous decision only) need to be updated, see Algorithms 1, 2, and 3.

(C) We use arguments and assumptions from [7], adapted to our risk-averse framework and problem formulation, but the proof uses Assumption (H2)-5), unstated in [7], which is necessary to obtain a description of the subdifferential of the recourse functions (see Lemma 2.1). Also, Assumption (H2)-4) is a necessary stronger assumption than Assumption (H1)-6) from [7].

(D) The argument $x_{1:t-1} = (x_1, \ldots, x_{t-1})$ of the recourse function $\mathcal{Q}_t$ for stage t takes values in $\mathbb{R}^{n(t-1)}$ (see Section 3 for details). To derive cuts for this function, we need the description of the subdifferential of a lower bounding convex function which is the value function of a convex problem. For that, proceeding as in [7], $n(t-1)$ additional variables $z^t \in \mathbb{R}^{n(t-1)}$ and the $n(t-1)$ constraints $z^t = (x_1^\top, \ldots, x_{t-1}^\top)^\top$ would be added. With the argument $x_{1:t-1}$ of the value function appearing only in the right-hand side of linear constraints, [7] then uses the (known) formula of the subdifferential of a value function whose argument is the right-hand side of linear constraints. On the contrary, we derive in Lemma 2.1 a formula for the subdifferential of the value function of a convex problem (with the argument of the value function in both the objective and nonlinear constraints) that does not need the introduction of additional variables and constraints. We believe that this lemma is a key tool for the implementation of SDDP applied to convex problems and is interesting per-se since subgradients of value functions of convex problems are computed at a large number of points at each iteration of the algorithm. The use of this formula should speed up each iteration. We are not aware of another paper proving this formula.

(E) A separate convergence proof is given for the case of interstage independent processes in which cuts can be shared between nodes of the same stage, assuming relatively complete recourse. The way to extend the algorithm and convergence proof to solve MSLPs that do not satisfy the relatively complete recourse assumption and to solve interstage dependent MSCPs is also discussed.

(F) It is shown that the optimal value of the approximate first stage problem converges almost surely to the optimal value of the problem and that almost surely any accumulation point of the sequence of approximate first stage solutions is an optimal solution of the first stage problem.

However, we use the traditional sampling process for SDDP ([14]), which is less general than the one from [7]. From the convergence analysis, we see that the main ingredients on which the convergence of SDDP relies (both in the risk-averse and risk-neutral settings) are the following:

(i) the decisions belong almost surely to compact sets.

(ii) The recourse functions and their lower bounding approximations are convex Lipschitz continuous on some sets. The subdifferentials of these functions are bounded on these sets.

(iii) The samples generated along the iterations are independent and at each stage, conditional to the history of the process, the number of possible realizations of the process is finite.

Since the recourse functions are expressed in terms of value functions of convex optimization problems, it is useful to study properties of such functions. This analysis is done in Section 2 where we provide a formula for the subdifferential of the value function of a convex optimization problem as well as conditions ensuring the continuity of this function and the boundedness of its subdifferential. Section 3 introduces the class of problems and decomposition algorithms we consider and prepares the ground showing (ii) above. Section 4 shows the convergence of these decomposition algorithms for interstage independent processes when relatively complete recourse holds. In Section 5, we explain how to extend the algorithm and convergence analysis for the special case of multistage stochastic linear programs that do not satisfy the relatively complete recourse assumption. Finally, while Sections 3-5 deal with interstage independent processes, Section 6 establishes the convergence when the process is interstage dependent.

We use the following notation and terminology:

- The tilde symbol will be used to represent realizations of random variables: for random variable ξ, $\tilde{\xi}$ is a realization of ξ.
- For vectors $x_1, \ldots, x_m \in \mathbb{R}^n$, we denote by $[x_1, \ldots, x_m]$ the $n \times m$ matrix whose i-th column is the vector x_i.

- For matrices A, B, we denote the matrix $\begin{pmatrix} A \\ B \end{pmatrix}$ by $[A; B]$ and the matrix $(A\ B)$ by $[A, B]$
- For sequences of n-vectors $(x_t)_{t \in \mathbb{N}}$ and $t_1 \leq t_2 \in \mathbb{N}$, $x_{t_1:t_2}$ will represent, depending on the context,
  (i) the Cartesian product $(x_{t_1}, x_{t_1+1}, \ldots, x_{t_2}) \in \underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{t_2 - t_1 + 1 \text{ times}}$ or
  (ii) the vector $[x_{t_1}; x_{t_1+1}; \ldots; x_{t_2}] \in \mathbb{R}^{n(t_2-t_1+1)}$.
- The usual scalar product in $\mathbb{R}^n$ is denoted by $\langle x, y \rangle = x^\top y$ for $x, y \in \mathbb{R}^n$. The corresponding norm is $\|x\| = \|x\|_2 = \sqrt{\langle x, x \rangle}$.
- $\mathbb{I}_A(\cdot)$ is the indicator function of the set A:

$$\mathbb{I}_A(x) := \left\{ \begin{array}{ll} 0, & \text{if } x \in A, \\ +\infty, & \text{if } x \notin A. \end{array} \right.$$

- $\mathrm{Gr}(f)$ is the graph of multifunction f.
- $A^* = \{x : \langle x, a \rangle \leq 0, \forall a \in A\}$ is the polar cone of A.
- $\mathcal{N}_A(x)$ is the normal cone to A at x.
- $\mathcal{T}_A(x)$ is the tangent cone to A at x.
- $\mathrm{ri}(A)$ is the relative interior of set A.
- $\mathbb{B}_n$ is the unit ball $\mathbb{B}_n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ in $\mathbb{R}^n$.
- $\mathrm{dom}(f)$ is the domain of function f.

2. Some properties of the value function of a convex optimization problem

We start providing a representation of the subdifferential of the value function of a convex optimization problem. This result plays a central role in the implementation and convergence analysis of SDDP applied to convex problems and will be used in the sequel.

Let $\mathcal{Q} : X \rightarrow \overline{\mathbb{R}}$, be the value function given by

$$\mathcal{Q}(x) = \left\{ \begin{array}{l} \inf_{y \in \mathbb{R}^n} \ f(x, y) \\ y \in S(x) := \{y \in Y \ : \ Ax + By = b, \ g(x, y) \leq 0\}. \end{array} \right. \tag{2.1}$$

Here, A and B are matrices of appropriate dimensions, and $X \subseteq \mathbb{R}^m$ and $Y \subseteq \mathbb{R}^n$ are nonempty, compact, and convex sets. Denoting by

$$X^\varepsilon := X + \varepsilon \mathbb{B}_m \tag{2.2}$$

the ε-fattening of the set X, we make the following assumption (H):

1) $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous, proper, and convex.
2) For $i = 1, \ldots, p$, the i-th component of function $g(x, y)$ is a convex lower semicontinuous function $g_i : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$.
3) There exists $\varepsilon > 0$ such that $X^\varepsilon \times Y \subset \mathrm{dom}(f)$.

Consider the Lagrangian dual problem

$$\sup_{(\lambda, \mu) \in \mathbb{R}^q \times \mathbb{R}^p_+} \theta_x(\lambda, \mu) \tag{2.3}$$

for the dual function

$$\theta_x(\lambda, \mu) = \inf_{y \in Y} \ f(x, y) + \lambda^\top (Ax + By - b) + \mu^\top g(x, y).$$

We denote by $\Lambda(x)$ the set of optimal solutions of the dual problem (2.3) and we use the notation

$$\mathrm{Sol}(x) := \{y \in S(x) : f(x, y) = \mathcal{Q}(x)\}$$

to indicate the solution set to (2.1).

It is well known that under Assumption (H), $\mathcal{Q}$ is convex and if f is uniformly convex then $\mathcal{Q}$ is uniformly convex too. The description of the subdifferential of $\mathcal{Q}$ is given in the following lemma:

Lemma 2.1. *Consider the value function $\mathcal{Q}$ given by (2.1) and take $x_0 \in X$ such that $S(x_0) \neq \emptyset$. Let*

$$C_1 = \Big\{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n \ : \ Ax + By = b\Big\} \text{ and } C_2 = \Big\{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n \ : \ g(x, y) \leq 0\Big\}.$$

Let Assumption (H) hold and assume the Slater-type constraint qualification condition:

$$\text{there exists } (\bar{x},\bar{y}) \in X\times \mathrm{ri}(Y) \text{ such that } (\bar{x},\bar{y}) \in C_1 \text{ and } (\bar{x},\bar{y}) \in \mathrm{ri}(C_2).$$

Then $s \in \partial \mathcal{Q}(x_0)$ if and only if

$$\begin{array}{l} (s,0) \in \partial f(x_0,y_0) + \Big\{[A^\top; B^\top]\lambda \ : \ \lambda \in \mathbb{R}^q\Big\} \\ \qquad +\Big\{ \displaystyle\sum_{i \in I(x_0,y_0)} \mu_i \partial g_i(x_0,y_0) \ : \ \mu_i \geq 0\Big\} + \{0\}\times \mathcal{N}_Y(y_0), \end{array} \tag{2.4}$$

where y_0 is any element in the solution set Sol(x_0), and with

$$I(x_0,y_0) = \Big\{i \in \{1,\ldots,p\} \ : \ g_i(x_0,y_0) = 0\Big\}.$$

In particular, if f and g are differentiable, then

$$\partial \mathcal{Q}(x_0) = \Big\{\nabla_x f(x_0,y_0) + A^\top \lambda + \sum_{i \in I(x_0,y_0)} \mu_i \nabla_x g_i(x_0,y_0) \ : \ (\lambda,\mu) \in \Lambda(x_0)\Big\}.$$

Proof. Observe that

$$\mathcal{Q}(x) = \left\{ \begin{array}{l} \inf \ f(x,y) + \mathbb{I}_{\mathrm{Gr}(S)}(x,y) \\ y \in \mathbb{R}^n \end{array} \right.$$

where $\mathbb{I}_{\mathrm{Gr}(S)}$ is the indicator function of the set

$$\mathrm{Gr}(S) := \Big\{(x,y) \in \mathbb{R}^m\times\mathbb{R}^n \ : \ Ax+By=b, \ g(x,y) \leq 0, \ y \in Y\Big\} = C_1 \bigcap C_2 \bigcap \mathbb{R}^m\times Y.$$

Using Theorem 24(a) in Rockafellar [19], we have

$$\begin{array}{rcl} s \in \partial\mathcal{Q}(x_0) & \Leftrightarrow & (s,0) \in \partial(f+\mathbb{I}_{\mathrm{Gr}(S)})(x_0,y_0) \\ & \Leftrightarrow & (s,0) \in \partial f(x_0,y_0) + \mathcal{N}_{\mathrm{Gr}(S)}(x_0,y_0). \ (a) \end{array} \tag{2.5}$$

For equivalence (2.5)-(a), we have used the fact that f and $\mathbb{I}_{\mathrm{Gr}(S)}$ are proper, finite at (x_0,y_0), and

$$\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(\mathbb{I}_{\mathrm{Gr}(S)})) \neq \emptyset. \tag{2.6}$$

The set $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(\mathbb{I}_{\mathrm{Gr}(S)}))$ is nonempty because it contains the point $(\bar{x},\bar{y})$:

$$\begin{array}{rcl} (\bar{x},\bar{y}) \in C_1 \cap \mathrm{ri}(C_2) \cap \mathbb{R}^m\times \mathrm{ri}(Y) & = & \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \cap \mathbb{R}^m\times\mathrm{ri}(Y) \\ & = & \mathrm{ri}(C_1 \cap C_2 \cap \mathbb{R}^m\times Y) = \mathrm{ri}(\mathrm{dom}(\mathbb{I}_{\mathrm{Gr}(S)})), \\ (\bar{x},\bar{y}) \in X\times\mathrm{ri}(Y) \subseteq \mathrm{ri}(X^\varepsilon)\times\mathrm{ri}(Y) & = & \mathrm{ri}(X^\varepsilon\times Y) \overset{(H)}{\subseteq} \mathrm{ri}(\mathrm{dom}(f)). \end{array}$$

Using the fact C_1 is an affine space and C_2 and Y are closed and convex sets such that $(\bar{x},\bar{y}) \in \mathrm{ri}(C_2) \cap \mathrm{ri}(\mathbb{R}^m\times Y) \cap C_1 \neq \emptyset$, we have

$$\mathcal{N}_{\mathrm{Gr}(S)}(x_0,y_0) = \mathcal{N}_{C_1}(x_0,y_0) + \mathcal{N}_{C_2}(x_0,y_0) + \mathcal{N}_{\mathbb{R}^m\times Y}(x_0,y_0).$$

But $\mathcal{N}_{\mathbb{R}^m\times Y}(x_0,y_0) = \{0\}\times\mathcal{N}_Y(y_0)$ and standard calculus on normal and tangent cones show that

$$\begin{array}{rcl} \mathcal{T}_{C_1}(x_0,y_0) & = & \{(x,y) \ : \ Ax+By=0\} = \mathrm{Ker}([A,B]), \\ \mathcal{N}_{C_1}(x_0,y_0) & = & \mathcal{T}^*_{C_1}(x_0,y_0) = (\mathrm{Ker}([A,B])^\perp \\ & = & \mathrm{Im}[A^\top;B^\top] = \Big\{[A^\top;B^\top]\lambda \ : \ \lambda \in \mathbb{R}^q\Big\}, \\ \mathcal{N}_{C_2}(x_0,y_0) & = & \Big\{ \displaystyle\sum_{i \in I(x_0,y_0)} \mu_i \partial g_i(x_0,y_0) \ : \ \mu_i \geq 0\Big\}. \end{array}$$

This completes the announced characterization (2.4) of $\partial\mathcal{Q}(x_0)$. If f and g are differentiable then the condition (2.4) can be written

$$\begin{array}{l} s = \nabla_x f(x_0,y_0) + A^\top\lambda + \displaystyle\sum_{i \in I(x_0,y_0)} \mu_i \nabla_x g_i(x_0,y_0), \ (a) \\ -\left[\nabla_y f(x_0,y_0) + B^\top\lambda + \displaystyle\sum_{i \in I(x_0,y_0)} \mu_i \nabla_y g_i(x_0,y_0)\right] \in \mathcal{N}_Y(y_0), \ (b) \end{array} \tag{2.7}$$

for some $\lambda \in \mathbb{R}^q$ and $\mu \in \mathbb{R}_+^{|I(x_0,y_0)|}$.

Finally, note that a primal-dual solution (y_0, λ, μ) satisfies (2.7)-(b) and if (y_0, λ, μ) with $\mu \geq 0$ satisfies (2.7)-(b), knowing that y_0 is primal feasible, then under our assumptions (λ, μ) is a dual solution, i.e., $(\lambda, \mu) \in \Lambda(x_0)$. $\square$

The following proposition provides conditions ensuring the Lipschitz continuity of $\mathcal{Q}$ and the boundedness of its subdifferential at any point in X:

Proposition 2.2. *Consider the value function $\mathcal{Q}$ given by* (2.1)*. Let Assumption (H) hold and assume that for every $x \in X^\varepsilon$, the set $S(x)$ is nonempty, where ε is given in (H)-3). Then $\mathcal{Q}$ is finite on X^ε, Lipschitz continuous on X, and the set $\cup_{x\in X} \partial \mathcal{Q}(x)$ is bounded. More precisely, if $M_0 = \sup_{x \in X^\varepsilon} \mathcal{Q}(x)$ and $m_0 = \min_{x \in X} \mathcal{Q}(x)$, then for every $x \in X$ and every $s \in \partial \mathcal{Q}(x)$ we have*

$$\|s\| \leq \frac{1}{\varepsilon}(M_0 - m_0). \tag{2.8}$$

Proof. Finiteness of $\mathcal{Q}$ on X^ε follows from the fact that, under the assumptions of the lemma, for every $x \in X^\varepsilon$, the feasible set $S(x)$ of (2.1) is nonempty and compact and the objective function $f(x, \cdot)$ is finite valued on Y and lower semicontinuous. It follows that X is contained in the relative interior of the domain of $\mathcal{Q}$. Since $\mathcal{Q}$ is convex and since a convex function is Lipschitz continuous on the relative interior of its domain, $\mathcal{Q}$ is Lipschitz continuous on X.

Next, for every $x \in X$, for every $y \in X^\varepsilon$ and $s \in \partial \mathcal{Q}(x)$, we have

$$\mathcal{Q}(y) \geq \mathcal{Q}(x) + \langle s, y - x \rangle.$$

Observing that M_0 and m_0 are finite ($\mathcal{Q}$ is finite and lower semicontinuous on the compact set X^ε), for every $x \in X$ and $y \in X^\varepsilon$ we get

$$M_0 \geq m_0 + \langle s, y - x \rangle.$$

If $s = 0$ then (2.8) holds and if $s \neq 0$, taking $y = x + \varepsilon \frac{s}{\|s\|} \in X^\varepsilon$ in the above relation, we obtain (2.8), i.e., s is bounded. $\square$

3. Decomposition methods for risk-averse multistage stochastic convex programs

Consider a risk-averse multistage stochastic optimization problem of the form

$$\begin{array}{ll} \displaystyle\inf_{x_1 \in X_1(x_0, \xi_1)} & f_1(x_1, \Psi_1) + \rho_{2|\mathcal{F}_1}\left(\displaystyle\inf_{x_2 \in X_2(x_{0:1}, \xi_2)} f_2(x_{1:2}, \Psi_2) + \ldots \right. \\ & +\rho_{T-1|\mathcal{F}_{T-2}}\left(\displaystyle\inf_{x_{T-1} \in X_{T-1}(x_{0:T-2},\, \xi_{T-1})} f_{T-1}(x_{1:T-1}, \Psi_{T-1}) \right. \\ & \left. \left. +\rho_{T|\mathcal{F}_{T-1}} \left(\displaystyle\inf_{x_T \in X_T(x_{0:T-1},\, \xi_T)} f_T(x_{1:T}, \Psi_T) \right) \right) \ldots \right) \end{array} \tag{3.9}$$

for some functions f_t taking values in $\mathbb{R} \cup \{+\infty\}$, where

$$X_t(x_{0:t-1}, \xi_t) = \left\{ x_t \in \mathcal{X}_t \; : \; g_t(x_{0:t}, \Psi_t) \leq 0, \; \sum_{\tau=0}^{t} A_{t,\tau} x_\tau = b_t \right\}$$

for some vector-valued functions g_t, some random vectors Ψ_t and b_t, some random matrices $A_{t,\tau}$, and where ξ_t is a discrete random vector with finite support corresponding to the concatenation of the random variables $(\Psi_t, b_t, (A_{t,\tau})_{\tau=0,\ldots,t})$ in an arbitrary order. In this problem x_0 is given, ξ_1 is deterministic, (ξ_t) is a stochastic process, and setting $\mathcal{F}_t = \sigma(\xi_1, \ldots, \xi_t)$ and denoting by $\mathcal{Z}_t$ the set of $\mathcal{F}_t$-measurable functions, $\rho_{t+1|\mathcal{F}_t} : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ is a coherent and law invariant conditional risk measure.

In this section and the next two Sections 4 and 5, we assume that the stochastic process (ξ_t) satisfies the following assumption:

(H1) (ξ_t) is interstage independent and for $t = 2, \ldots, T$, ξ_t is a random vector taking values in $\mathbb{R}^K$ with discrete distribution and finite support $\{\xi_{t,1}, \ldots, \xi_{t,M}\}$ while ξ_1 is deterministic ($\xi_{t,j}$ is the vector corresponding to the concatenation of the elements in $(\Psi_{t,j}, b_{t,j}, (A_{t,\tau,j})_{\tau=0,\ldots,t})$).

Under Assumption (H1), $\rho_{t+1|\mathcal{F}_t}$ coincides with its unconditional counterpart $\rho_{t+1} : \mathcal{Z}_{t+1} \to \mathbb{R}$. To alleviate notation and without loss of generality, we assume that the number M of possible realizations of ξ_t, the size K of ξ_t, and n of x_t do not depend on t.

For problem (3.9), we can write the following dynamic programming equations: we set $\mathcal{Q}_{T+1} \equiv 0$ and for $t = 2, \ldots, T$, define

$$\mathcal{Q}_t(x_{1:t-1}) = \rho_t\Big(\mathfrak{Q}_t(x_{1:t-1}, \xi_t)\Big) \tag{3.10}$$

with

$$\begin{array}{lll}\mathfrak{Q}_t(x_{1:t-1}, \xi_t) & = & \left\{\begin{array}{l}\inf\limits_{x_t} \ F_t(x_{1:t}, \Psi_t) := f_t(x_{1:t}, \Psi_t) + \mathcal{Q}_{t+1}(x_{1:t}) \\ x_t \in \mathcal{X}_t, \ g_t(x_{0:t}, \Psi_t) \leq 0, \ \displaystyle\sum_{\tau=0}^{t} A_{t,\tau} x_\tau = b_t,\end{array}\right. \\ & = & \left\{\begin{array}{l}\inf\limits_{x_t} \ F_t(x_{1:t}, \Psi_t) \\ x_t \in X_t(x_{0:t-1}, \xi_t).\end{array}\right.\end{array} \tag{3.11}$$

With this notation, $F_t(x_{1:t}, \Psi_t)$ is the future optimal cost starting at time t from the history of decisions $x_{1:t-1}$ if Ψ_t and x_t are respectively the value of the process (Ψ_t) and the decision taken at stage t. Problem (3.9) can then be written

$$\left\{\begin{array}{l}\inf\limits_{x_1} \ F_1(x_1, \Psi_1) := f_1(x_1, \Psi_1) + \mathcal{Q}_2(x_1) \\ x_1 \in X_1(x_0, \xi_1) = \{x_1 \in \mathcal{X}_1 : g_1(x_0, x_1, \Psi_1) \leq 0, A_{1,1}x_1 = b_1 - A_{1,0}x_0\},\end{array}\right. \tag{3.12}$$

with optimal value denoted by $\mathcal{Q}_1(x_0) = \mathfrak{Q}_1(x_0, \xi_1)$.

Setting $\Phi_{t,j} = \mathbb{P}(\xi_t = \xi_{t,j}) > 0$ for $j = 1, \ldots, M$, we reformulate the problem as in [16] using the dual representation of a coherent risk measure [1]:

$$\mathcal{Q}_t(x_{1:t-1}) = \rho_t(\mathfrak{Q}_t(x_{1:t-1}, \xi_t)) = \sup_{p \in \mathcal{P}_t} \sum_{j=1}^{M} p_j \Phi_{t,j} \mathfrak{Q}_t(x_{1:t-1}, \xi_{t,j}) \tag{3.13}$$

for some convex subset $\mathcal{P}_t$ of

$$\mathcal{D}_t = \{p \in \mathbb{R}^M : p \geq 0, \ \sum_{j=1}^{M} p_j \Phi_{t,j} = 1\}.$$

Optimization problem (3.13) is convex and linear if $\mathcal{P}_t$ is a polyhedron. Such is the case when $\rho_t = CVaR_{1-\varepsilon_t}$ is the Conditional Value-at-Risk of level $1 - \varepsilon_t$ (see [18]) where

$$\mathcal{P}_t = \{p \in \mathcal{D}_t : p_j \leq \frac{1}{\varepsilon_t}, j = 1, \ldots, M\}.$$

In this case, the optimization problem (3.13) can be solved analytically, without resorting to an optimization step (once the values $\mathfrak{Q}_t(x_{1:t-1}, \xi_{t,j}), j = 1, \ldots, M$, are known, see [16] for details) and numerical simulations in Section 4.1.1 of [23] have shown that the corresponding subproblems are solved more quickly than if the minimization formula from [24], [18] for the Conditional Value-at-Risk was used. We refer to [20], [21], [6], [9] for the definition of the sets $\mathcal{P}_t$ corresponding to various popular risk measures.

Recalling definition (2.2) of the the ε-fattening of a set X, we also make the following Assumption (H2) for $t = 1, \ldots, T$:

1) $\mathcal{X}_t \subset \mathbb{R}^n$ is nonempty, convex, and compact.
2) For every $x_{1:t} \in \mathbb{R}^n \times \ldots \times \mathbb{R}^n$ the function $f_t(x_{1:t}, \cdot)$ is measurable and for every $j = 1, \ldots, M$, the function $f_t(\cdot, \Psi_{t,j})$ is proper, convex, and lower semicontinuous.
3) For every $j = 1, \ldots, M$, each component of the function $g_t(x_0, \cdot, \Psi_{t,j})$ is a convex lower semicontinuous function.
4) There exists $\varepsilon > 0$ such that:
   4.1) for every $j = 1, \ldots, M$,

$$\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\varepsilon} \times \mathcal{X}_t \subset \operatorname{dom} f_t\Big(\cdot, \Psi_{t,j}\Big);$$

4.2) for every $j=1,\ldots,M$, for every $x_{1:t-1} \in \Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$, the set $X_t(x_{0:t-1}, \xi_{t,j})$ is nonempty.

5) If $t \geq 2$, for every $j=1,\ldots,M$, there exists

$$\bar{x}_{t,j} = (\bar{x}_{t,j,1}, \ldots, \bar{x}_{t,j,t}) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1} \times \mathrm{ri}(\mathcal{X}_t) \cap \mathrm{ri}(\{g_t(x_0, \cdot, \Psi_{t,j}) \leq 0\})$$

such that $\bar{x}_{t,j,t} \in X_t(x_0, \bar{x}_{t,j,1}, \ldots, \bar{x}_{t,j,t-1}, \xi_{t,j})$.

As shown in Proposition 3.1, Assumption (H2) guarantees that for $t=2,\ldots,T$, recourse function $\mathcal{Q}_t$ is convex and Lipschitz continuous on the set $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$ for every $0<\hat{\varepsilon}<\varepsilon$.

Proposition 3.1. *Under Assumption (H2), for $t=2,\ldots,T+1$, for every $0<\hat{\varepsilon}<\varepsilon$, the recourse function $\mathcal{Q}_t$ is convex, finite on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$, and continuous on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$.*

Proof. The proof is by induction on t. The result holds for $t=T+1$ since $\mathcal{Q}_{T+1} \equiv 0$. Now assume that for some $t \in \{2,\ldots,T\}$, the function $\mathcal{Q}_{t+1}$ is convex, finite on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_t\Big]^{\hat{\varepsilon}}$, and continuous on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_t\Big]^{\hat{\varepsilon}}$ for every $0<\hat{\varepsilon}<\varepsilon$. Take an arbitrary $0<\hat{\varepsilon}<\varepsilon$, $x_{1:t-1} \in \Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$ and fix $j \in \{1,\ldots,M\}$. Consider the optimization problem (3.11) with $\xi_t=\xi_{t,j}$. Note that the feasible set $X_t(x_{0:t-1}, \xi_{t,j})$ of this problem is nonempty (invoking (H2)-4.2)) and compact, since it is the intersection of the compact set $\mathcal{X}_t$ (invoking (H2)-1)), an affine space, and a lower level set of $g_t(x_{0:t-1}, \cdot, \Psi_{t,j})$ which is closed since this function $g_t(x_{0:t-1}, \cdot, \Psi_{t,j})$ is lower semicontinuous (using Assumption (H2)-3)). Next observe that if $x_{1:t-1} \in \Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$ and $x_t \in \mathcal{X}_t^{\tilde{\varepsilon}}$ with $\tilde{\varepsilon} = \sqrt{\left(\frac{\varepsilon+\hat{\varepsilon}}{2}\right)^2 - \hat{\varepsilon}^2} > 0$, then $x_{1:t} \in \Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_t\Big]^{\frac{\varepsilon+\hat{\varepsilon}}{2}}$ with $(\varepsilon+\hat{\varepsilon})/2 < \varepsilon$. Using this observation and the induction hypothesis, we have that $\mathcal{Q}_{t+1}(x_{1:t-1}, \cdot)$ is finite (and convex) on $\mathcal{X}_t^{\tilde{\varepsilon}}$ which implies that $\mathcal{Q}_{t+1}(x_{1:t-1}, \cdot)$ is Lipschitz continuous on $\mathcal{X}_t$. It follows that the optimal value $\mathfrak{Q}_t(x_{1:t-1}, \xi_{t,j})$ of problem (3.11) with $\xi_t = \xi_{t,j}$ is finite because its objective function $x_t \to f_t(x_{1:t-1}, x_t, \Psi_{t,j}) + \mathcal{Q}_{t+1}(x_{1:t-1}, x_t)$ takes finite values on $\mathcal{X}_t$ (using (H2)-4.1), (H2)-2), and the induction hypothesis) and is lower semicontinuous (using (H2)-2)). Using Definition (3.13) of $\mathcal{Q}_t$, we deduce that $\mathcal{Q}_t(x_{1:t-1})$ is finite. Since $x_{1:t-1}$ was chosen arbitrarily in $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$, we have shown that $\mathcal{Q}_t$ is finite on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$.

Next, we deduce from Assumptions (H2)-1), (H2)-2), and (H2)-3) that for every $j \in \{1,\ldots,M\}$, $\mathfrak{Q}_t(\cdot, \xi_{t,j})$ is convex on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$. Since ρ_t is coherent, it is monotone and convex, and $\mathcal{Q}_t(\cdot) = \rho_t(\mathfrak{Q}_t(\cdot, \xi_t))$ is convex on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$. Since $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$ is a compact subset of the relative interior of the domain of convex function $\mathcal{Q}_t$, we have that $\mathcal{Q}_t$ is Lipschitz continuous on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\hat{\varepsilon}}$. $\square$

Recalling Assumption (H1), the distribution of $(\xi_2, \ldots, \xi_T)$ is discrete and the M^{T-1} possible realizations of $(\xi_2, \ldots, \xi_T)$ can be organized in a finite tree with the root node n_0 associated to a stage 0 (with decision x_0 taken at that node) having one child node n_1 associated to the first stage (with ξ_1 deterministic). Algorithm 1 below is a sampling algorithm which, for iteration $k \geq 1$, selects a set of nodes $(n_1^k, n_2^k, \ldots, n_T^k)$ of the scenario tree (with n_t^k a node of stage t) corresponding to a sample $(\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$. Since $\tilde{\xi}_1^k = \xi_1$, we have $n_1^k = n_1$ for all k.

In the sequel, we use the following notation: $\mathcal{N}$ is the set of nodes and $\mathcal{P}: \mathcal{N} \to \mathcal{N}$ is the function associating to a node its parent node (the empty set for the root node). We will denote by $\texttt{Nodes}(t)$ the set of nodes for stage t and for a node n of the tree, we denote by

- $C(n)$ the set of its children nodes (the empty set for the leaves);
- x_n a decision taken at that node;
- Φ_n the transition probability from the parent node of n to n;

- ξ_n the realization of process (ξ_t) at node n[1]: for a node n of stage t, this realization ξ_n is the concatenation of the realizations Ψ_n of Ψ_t, b_n of b_t, and $A_{\tau,n}$ of $A_{t,\tau}$ for $\tau = 0, 1, \ldots, t$;
- $\xi_{[n]}$ (resp. $x_{[n]}$) the history of the realizations of the process (ξ_t) (resp. the history of the decisions) from the first stage node n_1 to node n: for a node n of stage t, the i-th component of $\xi_{[n]}$ (resp. $x_{[n]}$) is $\xi_{\mathcal{P}^{t-i}(n)}$ (resp. $x_{\mathcal{P}^{t-i}(n)}$) for $i = 1, \ldots, t$;

We are now in a position to describe Algorithm 1 which is a decomposition algorithm solving (3.9). This algorithm exploits the convexity of recourse functions $\mathcal{Q}_t,\ t = 2, \ldots, T+1$, building polyhedral lower approximations $\mathcal{Q}_t^k,\ t = 2, \ldots, T+1$, of these functions of the form

$$\begin{array}{lll} \mathcal{Q}_t^k(x_{1:t-1}) & = & \max\limits_{0 \le \ell \le k} \Big(\theta_t^\ell + \langle \beta_t^\ell, x_{1:t-1} - x^\ell_{[n^\ell_{t-1}]} \rangle \Big) \\ & = & \max\limits_{0 \le \ell \le k} \Big(\theta_t^\ell + \langle \beta_{t,1}^\ell, x_{1:t-2} - x^\ell_{[n^\ell_{t-2}]} \rangle + \langle \beta_{t,2}^\ell, x_{t-1} - x^\ell_{n^\ell_{t-1}} \rangle \Big), \end{array}$$

where $\beta_{t,1}^\ell \in \mathbb{R}^{n(t-2)}$ (resp. $\beta_{t,2}^\ell \in \mathbb{R}^n$) gathers the first $n(t-2)$ (resp. last n) components of β_t^ℓ.

Since $\mathcal{Q}_{T+1} \equiv 0$ is known, we have $\mathcal{Q}_{T+1}^k \equiv 0$ for all k, i.e., θ_{T+1}^k and β_{T+1}^k are null for all $k \in \mathbb{N}$. At iteration k, decisions $(x^k_{n_1^k}, \ldots, x^k_{n_T^k})$ and coefficients $(\theta_t^k, \beta_t^k),\ t = 2, \ldots, T+1$, are computed for a sample of nodes $(n_1^k, n_2^k, \ldots, n_T^k)$: $x^k_{n_t^k}$ is the decision taken at node n_t^k replacing the (unknown) recourse function $\mathcal{Q}_{t+1}$ by $\mathcal{Q}_{t+1}^{k-1}$, available at the beginning of iteration k.

In Lemma 3.2 below, we show that the coefficients (θ_t^k, β_t^k) computed in Algorithm 1 define valid cuts for $\mathcal{Q}_t$, i.e., $\mathcal{Q}_t \ge \mathcal{Q}_t^k$ for all $k \in \mathbb{N}$. To describe Algorithm 1, it is convenient to introduce for $t = 2, \ldots, T$, the function $\mathfrak{Q}_t^{k-1}$ defined as follows: $\mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_t)$ is the optimal value of the optimization problem

$$(3.14) \qquad \left\{ \begin{array}{l} \inf\limits_{x_t} \ F_t^{k-1}(x_{1:t}, \Psi_t) := f_t(x_{1:t}, \Psi_t) + \mathcal{Q}_{t+1}^{k-1}(x_{1:t}) \\ x_t \in \mathcal{X}_t, g_t(x_{0:t}, \Psi_t) \le 0, \ \ \displaystyle\sum_{\tau=0}^{t} A_{t,\tau} x_\tau = b_t. \end{array} \right.$$

We also denote by $\mathfrak{Q}_1^{k-1}(x_0, \xi_1)$ the optimal value of the problem above for $t = 1$.

Algorithm 1: Multistage stochastic decomposition algorithm to solve (3.9).

Initialization. Set $\mathcal{Q}_t^0 \equiv -\infty$ for $t = 2, \ldots, T$, and $\mathcal{Q}_{T+1}^0 \equiv 0$, i.e., set $\theta_{T+1}^0 = 0$, $\beta_{t+1}^0 = 0$ for $t = 1, \ldots, T$, and $\theta_{t+1}^0 = -\infty$ for $t = 1, \ldots, T-1$.

Loop.
For $k = 1, 2, \ldots,$
  Sample a set of $T+1$ nodes $(n_0^k, n_1^k, \ldots, n_T^k)$ such that n_0^k is the root node, $n_1^k = n_1$ is the node corresponding to the first stage, and for every $t = 2, \ldots, T$, node n_t^k is a child node of node n_{t-1}^k.
  This set of nodes is associated to a sample $(\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$, realizations of random variables $(\xi_1^k, \xi_2^k, \ldots, \xi_T^k)$.
  For $t = 1, \ldots, T,$
    For every node n of stage $t-1$
      For every child node m of node n, compute an optimal solution x_m^k of (3.14) taking $(x_{0:t-1}, \xi_t) =$

[1]Note that to alleviate notation, the same notation $\xi_{\texttt{Index}}$ is used to denote the realization of the process at node `Index` of the scenario tree and the value of the process (ξ_t) for stage `Index`. The context will allow us to know which concept is being referred to. In particular, letters n and m will only be used to refer to nodes while t will be used to refer to stages.

$(x_0, x^k_{[n]}, \xi_m)$ solving

$$
(3.15)\quad \left\{\begin{array}{l}
\inf_{x_t} \; f_t(x^k_{[n]}, x_t, \Psi_m) + \mathcal{Q}^{k-1}_{t+1}(x^k_{[n]}, x_t)\\
g_t(x_0, x^k_{[n]}, x_t, \Psi_m) \leq 0,\\
A_{t,m} x_t = b_m - \displaystyle\sum_{\tau=0}^{t-1} A_{\tau,m} x^k_{\mathcal{P}^{t-\tau}(m)},\\
x_t \in \mathcal{X}_t,
\end{array}\right.
= \left\{\begin{array}{ll}
\inf_{x_t, z} \; f_t(x^k_{[n]}, x_t, \Psi_m) + z & \\
g_t(x_0, x^k_{[n]}, x_t, \Psi_m) \leq 0, & [\pi_{k,m,1}]\\
A_{t,m} x_t = b_m - \displaystyle\sum_{\tau=0}^{t-1} A_{\tau,m} x^k_{\mathcal{P}^{t-\tau}(m)}, & [\pi_{k,m,2}]\\
z \geq \theta^\ell_{t+1} + \langle \beta^\ell_{t+1,1}, x^k_{[n]} - x^\ell_{[n^\ell_{t-1}]} \rangle & \\
\quad + \langle \beta^\ell_{t+1,2}, x_t - x^\ell_{n^\ell_t} \rangle, 0 \leq \ell \leq k-1, & [\pi_{k,m,3}]\\
x_t \in \mathcal{X}_t, &
\end{array}\right.
$$

with optimal value $\mathfrak{Q}^{k-1}_t(x^k_{[n]}, \xi_m)$ for $t \geq 2$, $\mathfrak{Q}^{k-1}_1(x_0, \xi_1)$ for $t = 1$, and with the convention, for $t = 1$, that $(x_0, x^k_{[n^k_0]}, x_1) = (x_0, x_1)$ and $x^k_{[n^k_0]} = x^k_{n^k_0} = x^k_{n_0} = x_0$ for all k.
In the above problem, we have denoted by $\pi_{k,m,1}, \pi_{k,m,2}$ and $\pi_{k,m,3}$ the optimal Lagrange multipliers associated with respectively the first, second, and third group of constraints.

End For

If $n = n^k_{t-1}$ and $t \geq 2$, compute for every $m \in C(n)$

$$
\begin{array}{rcl}
\pi_{k,m} & = & f'_{t,x_{1:t-1}}\Big(x^k_{[n]}, x^k_m, \Psi_m\Big) + g'_{t,x_{1:t-1}}\Big(x_0, x^k_{[n]}, x^k_m, \Psi_m\Big)\pi_{k,m,1}\\
& & + \begin{pmatrix} A^\top_{1,m} \\ \vdots \\ A^\top_{t-1,m} \end{pmatrix} \pi_{k,m,2} + [\beta^0_{t+1,1}, \beta^1_{t+1,1}, \ldots, \beta^k_{t+1,1}]\pi_{k,m,3}
\end{array}
$$

where $f'_{t,x_{1:t-1}}(x^k_{[n]}, x^k_m, \Psi_m)$ is a subgradient of convex function $f_t(\cdot, x^k_m, \Psi_m)$ at $x^k_{[n]}$ and the i-th column of matrix $g'_{t,x_{1:t-1}}(x_0, x^k_{[n]}, x^k_m, \Psi_m)$ is a subgradient at $x^k_{[n]}$ of the i-th component of convex function $g_t(x_0, \cdot, x^k_m, \Psi_m)$.
Compute $p_{k,m}, m \in C(n)$, solving

$$
\begin{array}{rcl}
\rho_t\left(\mathfrak{Q}^{k-1}_t(x^k_{[n]}, \xi_t)\right) & = & \sup_{p \in \mathcal{P}_t} \displaystyle\sum_{m \in C(n)} p_m \Phi_m \mathfrak{Q}^{k-1}_t(x^k_{[n]}, \xi_m)\\
& = & \displaystyle\sum_{m \in C(n)} p_{k,m} \Phi_m \mathfrak{Q}^{k-1}_t(x^k_{[n]}, \xi_m).
\end{array}
$$

Compute coefficients

$$
(3.16)\quad \theta^k_t = \rho_t\left(\mathfrak{Q}^{k-1}_t(x^k_{[n]}, \xi_t)\right) = \sum_{m \in C(n)} p_{k,m} \Phi_m \mathfrak{Q}^{k-1}_t(x^k_{[n]}, \xi_m) \text{ and } \beta^k_t = \sum_{m \in C(n)} p_{k,m} \Phi_m \pi_{k,m},
$$

making up the new approximate recourse function

$$
\mathcal{Q}^k_t(x_{1:t-1}) = \max_{0 \leq \ell \leq k} \Big(\theta^\ell_t + \langle \beta^\ell_t, x_{1:t-1} - x^\ell_{[n^\ell_{t-1}]} \rangle\Big).
$$

End If

End For

End For

Compute $\theta^k_{T+1} = 0$ and $\beta^k_{T+1} = 0$.

End For

The convergence of Algorithm 1 is shown in the next section. Various modifications of Algorithm 1 have been proposed in the literature. For instance, it is possible to

(i) use a number of samples that varies along the iterations;
(ii) sample from the distribution of ξ_t (instead of using all realizations $\xi_{t,1}, \ldots, \xi_{t,M}$ of ξ_t) to build the cuts [5], [17];
(iii) generate the trial points $x^k_{n^k_t}$ using the Abridged Nested Decomposition Method [4].

The convergence proof of the next section can be extended to these variants of Algorithm 1.

We will assume that the sampling procedure in Algorithm 1 satisfies the following property:

(H3) for every $j = 1, \dots, M$, for every $t = 2, \dots, T$, and for every $k \in \mathbb{N}^*$, $\mathbb{P}(\xi_t^k = \xi_{t,j}) = \Phi_{t,j} > 0$ with $\sum_{j=1}^M \Phi_{t,j} = 1$. For every $t = 2, \dots, T$, and $k \geq 1$,

$$\xi_t^k \text{ is independent on } \sigma(\xi_2^1, \dots, \xi_T^1, \dots, \xi_2^{k-1}, \dots, \xi_T^{k-1}, \xi_2^k, \dots, \xi_{t-1}^k).$$

In the following three lemmas, we show item (ii) announced in the introduction: functions $\mathcal{Q}_t^k$ for $k \geq T - t + 1$ are Lipschitz continuous and have bounded subgradients.

Lemma 3.2. *Consider the sequences $\mathcal{Q}_t^k, \theta_t^k$, and β_t^k generated by Algorithm 1. Under Assumptions (H2), then almost surely, for $t = 2, \dots, T+1$, the following holds:*

(a) *$\mathcal{Q}_t^k$ is convex with $\mathcal{Q}_t^k \leq \mathcal{Q}_t$ on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$ for all $k \geq 1$;*

(b) *the sequences $(\theta_t^k)_{k \geq T-t+1}$, $(\beta_t^k)_{k \geq T-t+1}$, and $(\pi_{k,m})_{k \geq T-t+1}$ for all m, are bounded;*

(c) *for $k \geq T - t + 1$, $\mathcal{Q}_t^k$ is convex Lipschitz continuous on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$.*

Proof. We show the result by induction on k and t. For $t = T+1$, and $k \geq 0$, θ_t^k and β_t^k are bounded since they are null (recall that $\mathcal{Q}_{T+1}^k$ is null for all $k \geq 0$) and $\mathcal{Q}_{T+1}^k = \mathcal{Q}_{T+1} \equiv 0$ is convex and Lipschitz continuous on $\mathcal{X}_1 \times \dots \times \mathcal{X}_T$ for $k \geq 0$. Assume now that for some $t \in \{1, \dots, T\}$ and $k \geq T - t + 1$, the functions $\mathcal{Q}_{t+1}^j$ for $T - t \leq j \leq k - 1$ are convex Lipschitz continuous on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_t\Big]^{\varepsilon}$ with $\mathcal{Q}_{t+1}^j \leq \mathcal{Q}_{t+1}$. We show that (i) θ_t^k and β_t^k are well defined and bounded; (ii) $\mathcal{Q}_t^k$ is convex Lipschitz continuous on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$; (iii) $\mathcal{Q}_t \geq \mathcal{Q}_t^k$ on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$.

Take an arbitrary $x_{1:t-1} \in \Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$. Since $\mathcal{Q}_{t+1} \geq \mathcal{Q}_{t+1}^{k-1}$ on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_t\Big]^{\varepsilon}$, using definition (3.11) of $\mathfrak{Q}_t$ and the definition of $\mathfrak{Q}_t^k$, we have

$$\mathfrak{Q}_t(x_{1:t-1}, \cdot) \geq \mathfrak{Q}_t^{k-1}(x_{1:t-1}, \cdot) \tag{3.17}$$

and using the monotonicity of ρ_t (recall that ρ_t is coherent)

$$\begin{array}{rcl} \mathcal{Q}_t(x_{1:t-1}) = \rho_t\Big(\mathfrak{Q}_t(x_{1:t-1}, \xi_t)\Big) & \geq & \rho_t\Big(\mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_t)\Big) \\ & \geq & \displaystyle\sup_{p \in \mathcal{P}_t} \sum_{j=1}^M p_j \Phi_{t,j} \mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_{t,j}). \end{array} \tag{3.18}$$

Using Assumptions (H2)-1), 2), 3), 4.1), 4.2) and the fact that $\mathcal{Q}_{t+1}^{k-1}$ is Lipschitz continuous on the compact set $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_t\Big]^{\varepsilon}$ (induction hypothesis), we have that $\mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_{t,j})$ is finite for all $j \in \{1, \dots, M\}$. Recalling that

$$\theta_t^k = \sum_{m \in C(n)} p_{k,m} \Phi_m \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m)$$

with $x_{[n]}^k \in \mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1} \subset \Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$, it follows that θ_t^k is finite. Next, using Assumptions (H2)-2), 3), for every $j \in \{1, \dots, M\}$, the function $\mathfrak{Q}_t^k(\cdot, \xi_{t,j})$ is convex. Since it is finite on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$, it is Lipschitz continuous on $\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}$. This function is thus subdifferentiable on $\Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$ and using Lemma 2.1, whose assumptions are satisfied, $\pi_{k,m}$ is a subgradient of $\mathfrak{Q}_t^{k-1}(\cdot, \xi_m)$ at $x_{[n_{t-1}^k]}^k$, i.e., setting $n = n_{t-1}^k$, for every $x_{1:t-1} \in \Big[\mathcal{X}_1 \times \dots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$, we have

$$\mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_m) \geq \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m) + \langle \pi_{k,m}, x_{1:t-1} - x_{[n]}^k \rangle. \tag{3.19}$$

Plugging this inequality into (3.18), still denoting $n = n_{t-1}^k$, we obtain for $x_{1:t-1} \in \Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\varepsilon}$:

$$
\begin{array}{lcl}
\mathcal{Q}_t(x_{1:t-1}) & \geq & \displaystyle\sup_{p \in \mathcal{P}_t} \sum_{j=1}^{M} p_j \Phi_{t,j} \mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_{t,j}) \\
& = & \displaystyle\sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_m) \\
& \geq & \displaystyle\sum_{m \in C(n)} p_{k,m} \Phi_m \mathfrak{Q}_t^{k-1}(x_{1:t-1}, \xi_m) \text{ since } p_k = (p_{k,m})_{m \in C(n)} \in \mathcal{P}_t \\
& \stackrel{(3.19)}{\geq} & \displaystyle\sum_{m \in C(n)} p_{k,m} \Phi_m \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m) + \sum_{m \in C(n)} p_{k,m} \Phi_m \langle \pi_{k,m}, x_{1:t-1} - x_{[n]}^k \rangle \\
& \geq & \theta_t^k + \langle \beta_t^k, x_{1:t-1} - x_{[n]}^k \rangle
\end{array}
$$

using the definitions of θ_t^k and β_t^k. If $\beta_t^k = 0$ then β_t^k is bounded and if $\beta_t^k \neq 0$, plugging $x_{1:t-1} = x_{[n]}^k + \frac{\varepsilon}{2} \frac{\beta_t^k}{\|\beta_t^k\|} \in \Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\varepsilon/2}$ in the above inequality, where ε is defined in (H2)-4.2), we obtain

$$
\|\beta_t^k\| \leq \frac{2}{\varepsilon} \left(\mathcal{Q}_t\Big(x_{[n]}^k + \frac{\varepsilon}{2} \frac{\beta_t^k}{\|\beta_t^k\|}\Big) - \theta_t^k \right). \tag{3.20}
$$

From Proposition 3.1, $\mathcal{Q}_t$ is finite on $\Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\varepsilon/2}$. Since θ_t^k is finite, (3.20) shows that β_t^k is bounded:

$$
\|\beta_t^k\| \leq \frac{2}{\varepsilon} \left(\sup_{x_{1:t-1} \in [\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^{\varepsilon/2}} \mathcal{Q}_t\Big(x_{1:t-1}\Big) - \theta_t^k \right). \tag{3.21}
$$

This achieves the induction step. Gathering our observations, we have shown that $\mathcal{Q}_t \geq \mathcal{Q}_t^k$ for all $k \in \mathbb{N}$ and that $\mathcal{Q}_t^k$ is Lipschitz continuous for $k \geq T - t + 1$.

Finally, using Proposition 2.2, we have that $\pi_{k,m}$ is bounded. More precisely, If $\pi_{k,m} \neq 0$, then relation (3.19) written for $x_{1:t-1} = \tilde{x}_{1:t-1}^{k,m} = x_{[n]}^k + \frac{\varepsilon}{2} \frac{\pi_{k,m}}{\|\pi_{k,m}\|} \in \Big[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}\Big]^{\varepsilon/2}$ gives for $k \geq T - t + 2$,

$$
\|\pi_{k,m}\| \leq \frac{2}{\varepsilon} \Big(\mathfrak{Q}_t(\tilde{x}_{1:t-1}^{k,m}, \xi_m) - \mathfrak{Q}_t^{T-t+1}(x_{[n]}^k, \xi_m) \Big),
$$

where we have used the fact that $\mathfrak{Q}_t^k \leq \mathfrak{Q}_t$ and $\mathfrak{Q}_t^{k+1}(\cdot, \xi_m) \geq \mathfrak{Q}_t^k(\cdot, \xi_m)$, for all $k \in \mathbb{N}$. In the proof of Proposition 3.1, we have shown that for every $t = 2, \ldots, T$, and $j = 1, \ldots, M$, the function $\mathfrak{Q}_t(\cdot, \xi_{t,j})$ is finite on the compact set $[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^{\varepsilon/2}$. Also, we have just shown that for every $t = 2, \ldots, T$, and $j = 1, \ldots, M$, the function $\mathfrak{Q}_t^{T-t+1}(\cdot, \xi_{t,j})$ is continuous on the compact set $\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}$. It follows that for every $k \geq T - t + 1$ and node m, we have for $\pi_{k,m}$ the upper bound

$$
\begin{array}{l}
\|\pi_{k,m}\| \leq \displaystyle\max_{t=2,\ldots,T, j=1,\ldots,M} \frac{2M(t,j)}{\varepsilon} \text{ where} \\
M(t,j) = \displaystyle\max_{x_{1:t-1} \in [\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^{\varepsilon/2}} \mathfrak{Q}_t(x_{1:t-1}, \xi_{t,j}) - \min_{x_{1:t-1} \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}} \mathfrak{Q}_t^{T-t+1}(x_{1:t-1}, \xi_{t,j}).
\end{array} \tag{3.22}
$$

□

Remark 3.3. *In the case when the cuts are computed in a backward pass using approximate recourse functions $\mathcal{Q}_{t+1}^k$ instead of $\mathcal{Q}_{t+1}^{k-1}$ for iteration k, we can guarantee that θ_t^k and β_t^k are bounded for all $t = 2, \ldots, T+1$, and $k \geq 1$ (for $k = 0$, we have $\beta_t^k = 0$ but $\theta_t^k = -\infty$ is not bounded for $t \leq T$).*

The following lemma will be useful in the sequel:

Lemma 3.4. *Consider the sequences $\mathcal{Q}_t^k, x_{[n_t^k]}^k$, and θ_t^k generated by Algorithm 1. Under Assumptions (H2), for $t = 2, \ldots, T$, and for all $k \geq 1$, we have*

$$
\mathcal{Q}_t^k(x_{[n_{t-1}^k]}^k) = \theta_t^k. \tag{3.23}
$$

Proof. We use the short notation $x^k_{1:t-1} = x^k_{[n^k_{t-1}]}$ and $n = n^k_{t-1}$. Observe that by construction $\mathfrak{Q}^k_t \geq \mathfrak{Q}^{k-1}_t$ for every $t = 2, \ldots, T+1$, and every $k \in \mathbb{N}^*$. It follows that for fixed $0 \leq \ell \leq k$,

$$\begin{array}{lll}
\theta^k_t & = & \sup\limits_{p \in \mathcal{P}_t} \sum\limits_{m \in C(n)} p_m \Phi_m \mathfrak{Q}^{k-1}_t(x^k_{[n]}, \xi_m) \\
& \geq & \sup\limits_{p \in \mathcal{P}_t} \sum\limits_{m \in C(n)} p_m \Phi_m \mathfrak{Q}^{\ell-1}_t(x^k_{1:t-1}, \xi_m) \\
& = & \sup\limits_{p \in \mathcal{P}_t} \sum\limits_{m \in C(n^\ell_{t-1})} p_m \Phi_m \mathfrak{Q}^{\ell-1}_t(x^k_{1:t-1}, \xi_m) \quad \text{(since } (\xi_t) \text{ is interstage independent)} \\
& \geq & \sup\limits_{p \in \mathcal{P}_t} \sum\limits_{m \in C(n^\ell_{t-1})} p_m \Phi_m \left(\mathfrak{Q}^{\ell-1}_t(x^\ell_{1:t-1}, \xi_m) + \langle \pi_{\ell,m}, x^k_{1:t-1} - x^\ell_{1:t-1} \rangle \right)
\end{array}$$

using the convexity of function $\mathfrak{Q}^{\ell-1}_t(\cdot, \xi_m)$ and the fact that $\pi_{\ell,m}$ is a subgradient of this function at $x^\ell_{1:t-1}$. Recalling that

$$\theta^\ell_t = \rho_t \left(\mathfrak{Q}^{\ell-1}_t(x^\ell_{1:t-1}, \xi_t) \right) = \sum_{m \in C(n^\ell_{t-1})} p_{\ell,m} \Phi_m \mathfrak{Q}^{\ell-1}_t(x^\ell_{1:t-1}, \xi_m) \text{ and } \beta^\ell_t = \sum_{m \in C(n^\ell_{t-1})} p_{\ell,m} \Phi_m \pi_{\ell,m},$$

we get

$$\begin{array}{lll}
\theta^k_t & \geq & \sum\limits_{m \in C(n^\ell_{t-1})} p_{\ell,m} \Phi_m \left(\mathfrak{Q}^{\ell-1}_t(x^\ell_{1:t-1}, \xi_m) + \langle \pi_{\ell,m}, x^k_{1:t-1} - x^\ell_{1:t-1} \rangle \right) \\
& \geq & \theta^\ell_t + \langle \beta^\ell_t, x^k_{1:t-1} - x^\ell_{1:t-1} \rangle
\end{array}$$

and

$$\mathcal{Q}^k_t(x^k_{1:t-1}) = \max\Big(\theta^k_t, \theta^\ell_t + \langle \beta^\ell_t, x^k_{1:t-1} - x^\ell_{1:t-1} \rangle, \ \ell = 0, \ldots, k-1 \Big) = \theta^k_t.$$

□

Lemma 3.5. *For $t = 2, \ldots, T$, and $k \geq T - t + 1$, the functions $\mathcal{Q}^k_t$ are L-Lipschitz with L given by*

$$\frac{2}{\varepsilon} \max_{t=2,\ldots,T} \left(\sup_{x_{1:t-1} \in [\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^{\varepsilon/2}} \mathcal{Q}_t(x_{1:t-1}) - \min_{x_{1:t-1} \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}} \mathcal{Q}^{T-t+1}_t(x_{1:t-1}) \right).$$

Proof. This is an immediate consequence of (3.21) and (3.23). □

4. Convergence analysis for risk-averse multistage stochastic convex programs

Theorem 4.1 shows the convergence of the sequence $\mathfrak{Q}^k_1(x_0, \xi_1)$ to $\mathcal{Q}_1(x_0)$ and that any accumulation point of the sequence $(x^k_1)_{k \in \mathbb{N}^*}$ is an optimal solution of the first stage problem (3.12).

Theorem 4.1 (Convergence analysis of Algorithm 1)**.** *Consider the sequences of stochastic decisions x^k_n and of recourse functions $\mathcal{Q}^k_t$ generated by Algorithm 1 to solve dynamic programming equations* (3.10)-(3.11). *Let Assumptions (H1), (H2), and (H3) hold. Then*

(i) *almost surely, for $t = 2, \ldots, T+1$, the following holds:*

$$\mathcal{H}(t): \quad \forall n \in \texttt{Nodes}(t-1), \quad \lim_{k \to +\infty} \mathcal{Q}_t(x^k_{[n]}) - \mathcal{Q}^k_t(x^k_{[n]}) = 0.$$

(ii) *Almost surely, we have $\lim\limits_{k \to +\infty} \mathfrak{Q}^k_1(x_0, \xi_1) = \mathcal{Q}_1(x_0)$ and if $f_1(\cdot, \Psi_1)$ is continuous on $\mathcal{X}_1$, any accumulation point of the sequence $(x^k_1)_{k \in \mathbb{N}^*}$ is an optimal solution of the first stage problem* (3.12).

Proof. In this proof, all equalities and inequalities hold almost surely. We show $\mathcal{H}(2), \ldots, \mathcal{H}(T+1)$, by induction backwards in time. $\mathcal{H}(T+1)$ follows from the fact that $\mathcal{Q}_{T+1} = \mathcal{Q}^k_{T+1} = 0$. Now assume that $\mathcal{H}(t+1)$ holds for some $t \in \{2, \ldots, T\}$. We want to show that $\mathcal{H}(t)$ holds. Take a node $n \in \texttt{Nodes}(t-1)$. Let $\mathcal{S}_n = \{k \geq 1 : n^k_{t-1} = n\}$ be the set of iterations such that the sampled scenario passes through node n. Due to Assumption (H3), the set $\mathcal{S}_n$ is infinite. We first show that

$$\lim_{k \to +\infty, \, k \in \mathcal{S}_n} \mathcal{Q}_t(x^k_{[n]}) - \mathcal{Q}^k_t(x^k_{[n]}) = 0. \tag{4.24}$$

Take $k \in \mathcal{S}_n$. We have $n_{t-1}^k = n$ and using Lemma 3.4 and the definition of $\mathcal{Q}_t$, we get

$$0 \leq \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n]}^k) = \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n_{t-1}^k]}^k) = \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_t(x_{[n]}^k, \xi_m) - \theta_t^k.$$

It follows that

$$\begin{array}{ll} & \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n]}^k) \\ & = \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_t(x_{[n]}^k, \xi_m) - \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m) \text{ by definition of } \theta_t^k, \\ & \leq \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \Big[\mathfrak{Q}_t(x_{[n]}^k, \xi_m) - \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m)\Big] \\ (4.25) & = \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \Big[\mathfrak{Q}_t(x_{[n]}^k, \xi_m) - f_t(x_{[m]}^k, \Psi_m) - \mathcal{Q}_{t+1}^{k-1}(x_{[m]}^k)\Big] \text{ by definition of } x_m^k, \\ & = \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \Big[\mathfrak{Q}_t(x_{[n]}^k, \xi_m) - F_t(x_{[m]}^k, \Psi_m) + \mathcal{Q}_{t+1}(x_{[m]}^k) - \mathcal{Q}_{t+1}^{k-1}(x_{[m]}^k)\Big] \end{array}$$

using the definition of F_t. Observing that for every $m \in C(n)$ and $k \in \mathcal{S}_n$ the decision $x_m^k \in X_t(x_0, x_{[n]}^k, \xi_m)$, we obtain, using definition (3.11) of $\mathfrak{Q}_t$, that

$$F_t(x_{[n]}^k, x_m^k, \Psi_m) = F_t(x_{[m]}^k, \Psi_m) \geq \mathfrak{Q}_t(x_{[n]}^k, \xi_m).$$

Combining this relation with (4.25) gives for $k \in \mathcal{S}_n$

$$0 \leq \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n]}^k) \leq \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_m \Big[\mathcal{Q}_{t+1}(x_{[m]}^k) - \mathcal{Q}_{t+1}^{k-1}(x_{[m]}^k)\Big]. \tag{4.26}$$

Using the induction hypothesis $\mathcal{H}(t+1)$, we have for every child node m of node n that

$$\lim_{k \to +\infty} \mathcal{Q}_{t+1}(x_{[m]}^k) - \mathcal{Q}_{t+1}^k(x_{[m]}^k) = 0. \tag{4.27}$$

Now recall that $\mathcal{Q}_{t+1}$ is convex on the compact set $\mathcal{X}_1 \times \ldots \times \mathcal{X}_t$ (Proposition 3.1), $x_{[m]}^k \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_t$ for every child node m of node n, and the functions $\mathcal{Q}_{t+1}^k, k \geq T-t+1$, are L-Lipschitz (Lemma 3.5) with $\mathcal{Q}_{t+1} \geq \mathcal{Q}_{t+1}^k \geq \mathcal{Q}_{t+1}^{k-1}$ on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_t$ (Lemma 3.2). It follows that we can use Lemma A.1 in [7] to deduce from (4.27) that

$$\lim_{k \to +\infty} \mathcal{Q}_{t+1}(x_{[m]}^k) - \mathcal{Q}_{t+1}^{k-1}(x_{[m]}^k) = 0.$$

Combining this relation with (4.26), we obtain

$$\lim_{k \to +\infty, k \in \mathcal{S}_n} \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n]}^k) = 0. \tag{4.28}$$

To show $\mathcal{H}(t)$, it remains to show that

$$\lim_{k \to +\infty, k \notin \mathcal{S}_n} \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n]}^k) = 0. \tag{4.29}$$

To show (4.29), we proceed similarly to the end of the proof of Theorem 3.1 in [7], by contradiction and using the Strong Law of Large Numbers. For the sake of completeness, we apply here these arguments in our context, where the notation and the convergence statement $\mathcal{H}(t)$ is different from [7]. If (4.28) does not hold, there exists $\varepsilon > 0$ such that there is an infinite number of iterations $k \in \mathbb{N}$ satisfying $\mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n]}^k) \geq \varepsilon$. Since $\mathcal{Q}_t^k \geq \mathcal{Q}_t^{k-1}$, there is also an infinite number of iterations belonging to the set

$$\mathcal{K}_{n,\varepsilon} = \{k \in \mathbb{N} : \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^{k-1}(x_{[n]}^k) \geq \varepsilon\}.$$

Consider the stochastic processes $(w_n^k)_{k \in \mathbb{N}^*}$ and $(y_n^k)_{k \in \mathbb{N}^*}$ where $w_n^k = 1_{k \in \mathcal{K}_{n,\varepsilon}}$ and $y_n^k = 1_{k \in \mathcal{S}_n}$, i.e., y_n^k takes the value 1 if node n belongs to the sampled scenario for iteration k (when $n_{t-1}^k = n$) and 0 otherwise. Assumption (H3) implies that random variables $(y_n^k)_{k \in \mathbb{N}^*}$ are independent and setting $\tilde{\mathcal{F}}_k = \sigma(w_n^1, \ldots, w_n^k, y_n^1, \ldots, y_n^{k-1})$, by definition of $x_{[n]}^j$ and $\mathcal{Q}_t^j$ that y_n^k is independent on $((x_{[n]}^j, j = 1, \ldots, k), (\mathcal{Q}_t^j, j = 1, \ldots, k-1))$ and thus of $\tilde{\mathcal{F}}_k$. If z^j is the jth element in the set $\{y_n^k : k \in \mathcal{K}_{n,\varepsilon}\}$, using Lemma A.3 in [7], we

obtain that random variables z^j are i.i.d. and have the distribution of y_n^1. Using the Strong Law of Large Numbers, we get

$$\frac{1}{N}\sum_{j=1}^{N} z^j \xrightarrow{N\to+\infty} \mathbb{E}[z^1] = \mathbb{E}[y_n^1] = \mathbb{P}(y_n^1 > 0) \overset{(H3)}{>} 0.$$

Relation (4.28) and Lemma A.1 in [7] imply that $\lim_{k\to+\infty, k\in\mathcal{S}_n} \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^{k-1}(x_{[n]}^k) = 0$. It follows that the set $\mathcal{K}_{n,\varepsilon} \cap \mathcal{S}_n = \mathcal{K}_{n,\varepsilon} \cap \{k \in \mathbb{N}^* : y_n^k = 1\}$ is finite. This implies

$$\frac{1}{N}\sum_{j=1}^{N} z^j \xrightarrow{N\to+\infty} 0,$$

which yields the desired contradiction and achieves the proof of (i).

(ii) By definition of $\mathfrak{Q}_1^{k-1}$ (see Algorithm 1), we have

$$\mathfrak{Q}_1^{k-1}(x_0,\xi_1) \quad = \quad f_1(x_{[n_1]}^k,\Psi_1) + \mathcal{Q}_2^{k-1}(x_{[n_1]}^k) = F_1(x_{[n_1]}^k,\Psi_1) - \mathcal{Q}_2(x_{[n_1]}^k) + \mathcal{Q}_2^{k-1}(x_{[n_1]}^k). \tag{4.30}$$

Since $x_{[n_1]}^k \in X_1(x_0,\xi_1)$ we have $F_1(x_{[n_1]}^k,\Psi_1) \geq \mathfrak{Q}_1(x_0,\xi_1)$. Together with (4.30), this implies

$$0 \leq \mathfrak{Q}_1(x_0,\xi_1) - \mathfrak{Q}_1^{k-1}(x_0,\xi_1) \leq \mathcal{Q}_2(x_{[n_1]}^k) - \mathcal{Q}_2^{k-1}(x_{[n_1]}^k). \tag{4.31}$$

Using $\mathcal{H}(2)$ from item (i) and Lemma A.1 in [7], this implies

$$\lim_{k\to+\infty} \mathcal{Q}_2(x_{[n_1]}^k) - \mathcal{Q}_2^{k-1}(x_{[n_1]}^k) = 0. \tag{4.32}$$

Plugging this relation into (4.31), we get $\lim\limits_{k\to+\infty} \mathfrak{Q}_1^{k-1}(x_0,\xi_1) = \mathfrak{Q}_1(x_0,\xi_1)$.

Recalling that n_1 is the node associated to the first stage, consider now an accumulation point $x_{n_1}^*$ of the sequence $(x_{n_1}^k)_{k\in\mathbb{N}}$. There exists a set K such that the sequence $(x_{n_1}^k)_{k\in K}$ converges to $x_{n_1}^*$. By definition of $x_{n_1}^k$ and since $\Psi_{n_1} = \Psi_1$, we get

$$f_1(x_{n_1}^k,\Psi_1) + \mathcal{Q}_2^{k-1}(x_{n_1}^k) = \mathfrak{Q}_1^{k-1}(x_0,\xi_1). \tag{4.33}$$

Using (4.32) and the continuity of $\mathcal{Q}_2$ on $\mathcal{X}_1$, we have

$$\lim_{k\to+\infty,\, k\in K} \mathcal{Q}_2^{k-1}(x_{n_1}^k) = \mathcal{Q}_2(x_{n_1}^*).$$

Taking the limit in (4.33) when $k \to +\infty$ with $k \in K$, we obtain

$$\mathcal{Q}_1(x_0) = f_1(x_{n_1}^*,\Psi_1) + \mathcal{Q}_2(x_{n_1}^*) = F_1(x_{n_1}^*,\Psi_1).$$

Since for every $k \in K$, $x_{n_1}^k$ is feasible for the first stage problem, so is $x_{n_1}^*$ ($X_1(x_0,\xi_1)$ is closed) and $x_{n_1}^*$ is an optimal solution to the first stage problem. □

Remark 4.2. *In Algorithm 1, decisions are computed at every iteration for all the nodes of the scenario tree. However, in practice, decisions will only be computed for the nodes of the sampled scenarios and their children nodes (such is the case of SDDP). This variant of Algorithm 1 will build the same cuts and compute the same decisions for the nodes of the sampled scenarios as Algorithm 1. For this variant, for a node n, the decision variables* $(x_n^k)_k$ *are defined for an infinite subset* $\tilde{\mathcal{S}}_n$ *of iterations where the sampled scenario passes through the parent node of node n, i.e.,* $\tilde{\mathcal{S}}_n = \mathcal{S}_{\mathcal{P}(n)}$. *With this notation, for this variant of Algorithm 1, applying Theorem 4.1-(i), we get for* $t = 2, \ldots, T+1$,

$$\text{for all } n \in \texttt{Nodes}(t-1), \lim_{k\to+\infty, k\in\tilde{\mathcal{S}}_n} \mathcal{Q}_t(x_{[n]}^k) - \mathcal{Q}_t^k(x_{[n]}^k) = 0 \tag{4.34}$$

almost surely, while Theorem 4.1-(ii) still holds.

5. Convergence analysis for risk-averse multistage stochastic linear programs without relatively complete recourse

In this section, we consider the case when g_t is affine and f_t is linear. We replace assumption (H2)-1) by $\mathcal{X}_t = \mathbb{R}^n_+$ and we do not make Assumptions (H2)-4)-5). More precisely, instead of (3.11), we consider the following dynamic programming equations corresponding to multistage stochastic linear programs that do not satisfy the relatively complete recourse assumption: we set $\mathcal{Q}_{T+1} \equiv 0$ and for $t = 2, \ldots, T$, we define $\mathcal{Q}_t(x_{1:t-1}) = \rho_t\Big(\mathfrak{Q}_t(x_{1:t-1}, \xi_t)\Big)$ now with

$$
\mathfrak{Q}_t(x_{1:t-1}, \xi_t) = \left\{ \begin{array}{l} \inf_{x_t} \; F_t(x_{1:t}, \Psi_t) := \Psi_t^\top x_{1:t} + \mathcal{Q}_{t+1}(x_{1:t}) \\ \displaystyle\sum_{\tau=0}^{t} A_{t,\tau} x_\tau = b_t, \; x_t \geq 0, \end{array} \right. = \left\{ \begin{array}{l} \inf_{x_t} \; F_t(x_{1:t}, \Psi_t) \\ x_t \in X_t(x_{0:t-1}, \xi_t). \end{array} \right. \tag{5.35}
$$

At the first stage, we solve

$$
\left\{ \begin{array}{l} \inf_{x_1} \; F_1(x_1, \Psi_1) := \Psi_1^\top x_1 + \mathcal{Q}_2(x_1) \\ x_1 \in X_1(x_0, \xi_1) = \{x_1 \geq 0, A_{1,1} x_1 = b_1 - A_{1,0} x_0\}, \end{array} \right. \tag{5.36}
$$

with optimal value denoted by $\mathcal{Q}_1(x_0) = \mathfrak{Q}_1(x_0, \xi_1)$. If we apply Algorithm 1 to solve (5.36) (in the sense of Theorem 4.1), since Assumption (H2)-4) does not hold, it is possible that one of the problems (3.15) to be solved in the forward passes is infeasible. In this case, $x^k_{[n]}$ is not a feasible sequence of states from stage 1 to stage $t-1$ and we build a separating hyperplane separating x_n^k and the set of states that are feasible at stage $t-1$ (those for which there exist sequences of decisions on any future scenario, assuming that problem (5.36) is feasible). The construction of feasibility cuts for the nested decomposition algorithm is described in [2]. Feasibility cuts for sampling based decomposition algorithms were introduced in [8]. This latter reference also discusses how to share feasibility cuts among nodes of the same stage for some interstage independent processes and stochastic programs. In the case of problem (5.35), before solving problems (3.15) for all $m \in C(n)$ in the forward pass, setting $n = n^k_{t-1}$, we solve for every $m \in C(n)$ the optimization problem

$$
\begin{array}{ll} \min_{x_t, y_1, y_2} & e^\top (y_1 + y_2) \\ & A_{t,m} x_t + y_1 - y_2 = b_m - \displaystyle\sum_{\tau=0}^{t-1} A_{\tau,m} x^k_{\mathcal{P}^{t-\tau}(m)}, \qquad [\pi] \\ & (x^k_{[n]})^\top \tilde{\beta}^\ell_{t+1,1} + x_t^\top \tilde{\beta}^\ell_{t+1,2} \leq \tilde{\theta}^\ell_{t+1}, \; \ell = 1, \ldots, K_t, \qquad [\tilde{\pi}] \\ & x_t, y_1, y_2 \geq 0, \end{array} \tag{5.37}
$$

where e is a vector of ones. In the above problem, we have denoted by respectively π and $\tilde{\pi}$ optimal Lagrange multipliers for the first and second set of constraints[2]. If $\tilde{\mathfrak{Q}}_t(x^k_{[n]}, \xi_m)$ is the optimal value of (5.37), noting that $\tilde{\mathfrak{Q}}_t(\cdot, \xi_m)$ is convex with $s = [A^\top_{1,m}; \ldots; A^\top_{t-1,m}]\pi + [\tilde{\beta}^1_{t+1,1}, \ldots, \tilde{\beta}^{K_t}_{t+1,1}]\tilde{\pi}$ belonging to the subdifferential of $\tilde{\mathfrak{Q}}_t(\cdot, \xi_m)$ at $x^k_{[n]}$, if $x_{1:t-1}$ is feasible then

$$
\tilde{\mathfrak{Q}}_t(x_{1:t-1}, \xi_m) = 0 \geq \tilde{\mathfrak{Q}}_t(x^k_{[n]}, \xi_m) + s^\top (x_{1:t-1} - x^k_{[n]}). \tag{5.38}
$$

Inequality (5.38) defines a feasibility cut for $x_{1:t-1}$ of the form

$$
x_{1:t-1}^\top \tilde{\beta}^\ell_t \leq \tilde{\theta}^\ell_t \tag{5.39}
$$

with

$$
\tilde{\beta}^\ell_t = [(\tilde{\beta}^\ell_{t,1})^\top; (\tilde{\beta}^\ell_{t,2})^\top] = s \text{ and } \tilde{\theta}^\ell_t = -\tilde{\mathfrak{Q}}_t(x^k_{[n]}, \xi_m) + s^\top x^k_{[n]}, \tag{5.40}
$$

where $\tilde{\beta}^\ell_{t,1}$ (resp. $\tilde{\beta}^\ell_{t,2}$) is the vector containing the first $n(t-2)$ (resp. last n) components of $\tilde{\beta}^\ell_t$. Incorporating these cuts in the forward pass of Algorithm 1, we obtain Algorithm 2.

[2] We suppressed the dependency with respect to n, k to alleviate notation.

Algorithm 2: Multistage stochastic decomposition algorithm to solve (5.35) **without the relatively complete recourse assumption.**

Initialization. Set $k=1$ (iteration count), Out=0 (Out will be 1 if the problem is infeasible), $K_t=0$ (number of feasibility cuts at stage t), $\mathcal{Q}_t^0 \equiv -\infty$ for $t=2,\ldots,T$, and $\mathcal{Q}_{T+1}^0 \equiv 0$, i.e., set $\theta_{T+1}^0=0$, $\beta_{t+1}^0=0$ for $t=1,\ldots,T$, and θ_{t+1}^0 to $-\infty$ for $t=1,\ldots,T-1$.

While Out=0
  Sample a set of $T+1$ nodes $(n_0^k, n_1^k, \ldots, n_T^k)$ such that n_0^k is the root node, $n_1^k=n_1$ is the node corresponding to the first stage, and for every $t=1,\ldots,T$, node n_t^k is a child node of node n_{t-1}^k.
  This set of nodes is associated to a sample $(\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$, realization of random variables $(\xi_1^k, \ldots, \xi_T^k)$.
  Set $t=1$, $n=n_{t-1}^k=n_0$, and $x_n^k=x_0$.
  While $(t <= T)$ and (Out=0)
    Set OutAux= 0.
    While there remains a nonvisited child of n and (OutAux= 0),
      Take for m a nonvisited child of n.
      Solve problem (5.37).
      If the optimal value of (5.37) is positive then
        If $t=1$ then
          Out=1, OutAux= 1 //the problem is infeasible
        Else
          Compute (5.40) to build feasibility cut (5.39).
          Increase K_{t-1} by one.
          Set $n=n_{t-2}^k$ and OutAux= 1.
          Decrease t by one.
        End if
      End If
    End While
    If OutAux= 0,
      Setting $m=n_t^k$, compute an optimal solution x_m^k of

$$
(5.41) \qquad \left\{
\begin{array}{ll}
\inf\limits_{x_t} \, ([x_{[n]}^k; x_t])^\top \Psi_m + z & \\
\displaystyle\sum_{\tau=0}^{t-1} A_{\tau,m} x^k_{\mathcal{P}^{t-\tau}(m)} + A_{t,m} x_t = b_m, & [\pi_{k,m,1}] \\
z \geq \theta_{t+1}^\ell + \langle \beta_{t+1,1}^\ell, x_{[n]}^k - x^\ell_{[n_{t-1}^\ell]} \rangle & \\
\qquad + \langle \beta_{t+1,2}^\ell, x_t - x^\ell_{n_t^\ell} \rangle, \;\; 0 \leq \ell \leq k-1, & [\pi_{k,m,2}] \\
([x_{[n]}^k; x_t])^\top \tilde{\beta}_{t+1}^\ell \leq \tilde{\theta}_{t+1}^\ell, \;\; \ell = 1, \ldots, K_t, & [\pi_{k,m,3}] \\
x_t \geq 0. &
\end{array}
\right.
$$

      In the above problem, we have denoted by $\pi_{k,m,1}, \pi_{k,m,2}$, and $\pi_{k,m,3}$, the optimal Lagrange multipliers associated with respectively the first, second, and third group of constraints.
      Increase t by one and set $n=n_{t-1}^k$.
    End If
  End While
  If Out=0
    For $t=2,\ldots,T$,
     For each child node m of $n=n_{t-1}^k$ with $m \neq n_t^k$,
        compute an optimal solution x_m^k and the optimal value $\mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m)$
        of optimization problem (5.41).
     End For

For every child node m of n, compute

$$\begin{array}{rcl}\pi_{k,m} & = & \Psi_m(1:(n-1)t) + \begin{pmatrix} A_{1,m}^\top \\ \vdots \\ A_{t-1,m}^\top \end{pmatrix} \pi_{k,m,1} \\ & & +[\beta_{t+1,1}^0, \ldots \beta_{t+1,1}^{k-1}]\pi_{k,m,2} + [\tilde{\beta}_{t+1,1}^1, \ldots \tilde{\beta}_{t+1,1}^{K_t}]\pi_{k,m,3}.\end{array}$$

Compute $(p_{k,m})_{m\in C(n)}$ such that

$$\begin{array}{rcl}\rho_t\left(\mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_t)\right) & = & \sup\limits_{p\in\mathcal{P}_t} \sum\limits_{m\in C(n)} p_m \Phi_m \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m) \\ & = & \sum\limits_{m\in C(n)} p_{k,m}\Phi_m \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m)\end{array}$$

and coefficients

$$\theta_t^k = \rho_t\left(\mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_t)\right) = \sum_{m\in C(n)} p_{k,m}\Phi_m \mathfrak{Q}_t^{k-1}(x_{[n]}^k, \xi_m), \quad \beta_t^k = \sum_{m\in C(n)} p_{k,m}\Phi_m \pi_{k,m}, \tag{5.42}$$

making up the new approximate recourse function

$$\mathcal{Q}_t^k(x_{1:t-1}) = \max_{0\le \ell\le k} \left(\theta_t^\ell + \langle \beta_t^\ell, x_{1:t-1} - x_{[n_{t-1}^\ell]}^\ell\rangle\right).$$

End For

Compute $\theta_{T+1}^k = 0$ and $\beta_{T+1}^k = 0$.

End If

Increase k by one.

End While.

Theorem 5.1 which is a convergence analysis of Algorithm 2 is a corollary of the convergence analysis of Algorithm 1 from Theorem 4.1:

Theorem 5.1 (Convergence analysis of Algorithm 2)**.** *Let Assumptions (H1) and (H3) hold and assume that*

(H2') *for every $t = 1, \ldots, T$, for every realization $(\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_t)$ of $(\xi_1, \xi_2, \ldots, \xi_t)$, for every sequence of feasible decisions $x_{0:t-1}$ on that scenario, i.e., satisfying $x_\tau \in X_\tau(x_{0:\tau-1}, \tilde{\xi}_\tau)$ for $\tau = 1, \ldots, t-1$, the set $X_t(x_{0:t-1}, \tilde{\xi}_t)$ is bounded and nonempty.*

Then either Algorithm 2 terminates reporting that the problem is infeasible or for $t = 2, \ldots, T$, (4.34) holds almost surely and Theorem 4.1-(ii) holds.

Proof. Due to Assumption (H2'), recourse functions $\mathcal{Q}_t$ are convex polyhedral and Lipschitz continuous. Moreover, Assumption (H2') also guarantees that

(a) all linear programs (5.41) are feasible and have bounded primal and dual feasible sets. As a result, functions $(\mathcal{Q}_t^k)_{t,k}$ are also Lipschitz continuous convex and polyhedral.

(b) The feasible set of (5.37) is bounded and nonempty.

From (b) and Assumption (H1), we obtain that there is only a finite number of different feasibility cuts. From the definition of these feasibility cuts, the feasible set of (5.37) contains the first stage feasible set. As a result, if (5.37) is not feasible, there is no solution to (5.35). Otherwise, since only a finite number of different feasibility cuts can be generated, after some iteration k_0 no more feasibility cuts are generated. In this case, after iteration k_0, Algorithm 2 is the variant of Algorithm 1 described in Remark 4.2 and the proof can be achieved combining the proof of Algorithm 1 and Remark 4.2. □

6. Convergence analysis with interstage dependent processes

Consider a problem of form (3.9) with an interstage dependent process (ξ_t), and let Assumption (H2) hold. We assume that the stochastic process (ξ_t) is discrete with a finite number of realizations at each stage. The realizations of the process over the optimization period can still be represented by a finite scenario tree

with the root node n_0 associated to a fictitious stage 0 with decision x_0 taken at that node. The unique child node n_1 of this root node corresponds to the first stage (with ξ_1 deterministic). In addition to the notation introduced in Section 4, we also define τ_n to be the stage associated to node n.

For interstage dependent processes, Algorithm 1 can be extended in two ways. For some classes of processes, we can add in the state vectors past process values while preserving the convexity of the recourse functions. We refer to [11], [8] for more details. The convergence of Algorithm 1 applied to the corresponding dynamic programming equations can be proved following the developments of Sections 3 and 4.

It is also possible to deal with more general interstage dependent processes as in [7]. However, in this case, recourse functions are not linked to stages but to the nodes of the scenario tree. In this context, we associate to each node n of the tree a coherent risk measure $\rho_n : \mathbb{R}^{|C(n)|} \to \mathbb{R}$ and risk measure $\rho_{t+1|\mathcal{F}_t}$ in formulation (3.9) is given by the collection of the risk measures $(\rho_n)_{n\,:\,\tau_n=t}$. More precisely, we consider the following dynamic programming equations: for every node n which is neither the root node nor a leaf, using the dual representation of risk measure ρ_n, we define the recourse function

$$\mathcal{Q}_n(x_{[n]}) = \rho_n\Big(\mathfrak{Q}_n(x_{[n]}, (\xi_m)_{m\in C(n)})\Big) = \sup_{p\in\mathcal{P}_n} \sum_{m\in C(n)} p_m \Phi_m \mathfrak{Q}_n(x_{[n]}, \xi_m) \tag{6.43}$$

for some convex subset $\mathcal{P}_n$ of

$$\mathcal{D}_n = \{p \in \mathbb{R}^{|C(n)|} : p \geq 0, \quad \sum_{m\in C(n)} p_m\Phi_m = 1\},$$

where $\mathfrak{Q}_n(x_{[n]}, \xi_m)$ is given by

$$\left\{\begin{array}{l}\inf\limits_{x_m} \; F_{\tau_m}(x_{[n]}, x_m, \Psi_m)\\ x_m \in \mathcal{X}_{\tau_m},\; g_{\tau_m}(x_0, x_{[n]}, x_m, \Psi_m) \leq 0,\\ [\,A_{0,m}, \ldots, A_{\tau_m,m}][x_0; x_{[n]}; x_m] = b_m\end{array}\right. = \left\{\begin{array}{l}\inf\limits_{x_m} \; F_{\tau_m}(x_{[n]}, x_m, \Psi_m)\\ x_m \in X_{\tau_m}\Big(x_0, x_{[n]}, \xi_m\Big)\end{array}\right. \tag{6.44}$$

with

$$F_{\tau_m}(x_{[n]}, x_m, \Psi_m) = f_{\tau_m}(x_{[n]}, x_m, \Psi_m) + \mathcal{Q}_m(x_{[n]}, x_m).$$

If n is a leaf node then $\mathcal{Q}_n \equiv 0$. For the first stage, we solve problem (6.44) with $n = n_0$ and $m = n_1$, with optimal value denoted by $\mathcal{Q}_{n_0}(x_0) = \mathfrak{Q}_{n_0}(x_0, \xi_{n_1})$ where $\xi_{n_1} = \xi_1$.

Algorithm 3 solves these dynamic programming equations building at iteration k polyhedral lower approximation $\mathcal{Q}_n^k$ of $\mathcal{Q}_n$ where

$$\mathcal{Q}_n^k(x_{[n]}) = \max\Big(\theta_n^\ell + \langle \beta_n^\ell, x_{[n]} - x_{[n]}^\ell\rangle, 0 \leq \ell \leq k\Big)$$

for all node $n \in \mathcal{N}\backslash\{n_0\}$.

If node n is not a leaf, we introduce the function $\mathfrak{Q}_n^{k-1}$ such that $\mathfrak{Q}_n^{k-1}(x_{[n]}, \xi_m)$ is given by

$$\left\{\begin{array}{l}\inf\limits_{x_m} \; F_{\tau_m}^{k-1}(x_{[n]}, x_m, \Psi_m) := f_{\tau_m}(x_{[n]}, x_m, \Psi_m) + \mathcal{Q}_m^{k-1}(x_{[n]}, x_m)\\ g_{\tau_m}(x_0, x_{[n]}, x_m, \Psi_m) \leq 0,\\ [\,A_{0,m}, \ldots, A_{\tau_m,m}][x_0; x_{[n]}; x_m] = b_m\\ x_m \in \mathcal{X}_{\tau_m}.\end{array}\right. \tag{6.45}$$

Next, we write $\mathfrak{Q}_n^{k-1}(x_{[n]}, \xi_m)$ under the form

$$\left\{\begin{array}{ll}\inf\limits_{x_m} \; f_{\tau_m}(x_{[n]}, x_m, \Psi_m) + z & \\ g_{\tau_m}(x_0, x_{[n]}, x_m, \Psi_m) \leq 0, & [\pi_{k,m,1}]\\ [\,A_{0,m}, \ldots, A_{\tau_m,m}][x_0; x_{[n]}; x_m] = b_m & [\pi_{k,m,2}]\\ z \geq \theta_m^\ell + \langle \beta_{m,1}^\ell, x_{[n]} - x_{[n]}^\ell\rangle + \langle \beta_{m,2}^\ell, x_m - x_m^\ell\rangle, \;\; \ell \leq k-1, & [\pi_{k,m,3}]\\ x_m \in \mathcal{X}_{\tau_m}, & \end{array}\right. \tag{6.46}$$

where $\beta_{m,1}^\ell$ (resp. $\beta_{m,2}^\ell$) contains the first $n\tau_n$ (resp. last n) components of β_m^ℓ. In the above problem, we have denoted by $\pi_{k,m,1}, \pi_{k,m,2}$, and $\pi_{k,m,3}$ the optimal Lagrange multipliers associated with respectively the first, second, and third group of constraints. Finally, if n is a leaf, $\mathfrak{Q}_n^{k-1} = 0$.

For $n = n_0$, the optimal value of (6.45) is denoted by $\mathfrak{Q}_{n_0}^{k-1}(x_0, \xi_1)$.

Algorithm 3: Multistage stochastic decomposition algorithm to solve (6.43) **for interstage dependent processes.**

Initialization. Set $\mathcal{Q}_n^0 \equiv 0$ for all leaf node n and $\mathcal{Q}_n^0 \equiv -\infty$ for all other node n.

For $k = 1, 2, \ldots,$

Sample a scenario $(\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ for $(\xi_1, \xi_2, \ldots, \xi_T)$, realization of $(\xi_1^k, \ldots, \xi_T^k)$, i.e., sample a set of $T+1$ nodes $(n_0^k, n_1^k, n_2^k, \ldots, n_T^k)$ such that $n_0^k = n_0$ is the root node, $n_1^k = n_1$ is the node corresponding to the first stage, and for every $t = 2, \ldots, T$, node n_t^k is a child of node n_{t-1}^k.

For $t = 1, \ldots, T$,

For every node n of stage $t-1$,

For every child node m of node n,

solve (6.45) and denote by x_m^k be an optimal solution.

End For

If $t \geq 2$ and $n \neq n_{t-1}^k$, compute

$$\theta_n^k = \mathcal{Q}_n^{k-1}(x_{[n]}^k) \text{ and } \beta_n^k \in \partial \mathcal{Q}_n^{k-1}(x_{[n]}^k). \tag{6.47}$$

Else if $t \geq 2$ and $n = n_{t-1}^k$, then for every $m \in C(n)$ compute a subgradient $\pi_{k,m}$ of $\mathfrak{Q}_n^{k-1}(\cdot, \xi_m)$ at $x_{[n]}^k$:

$$\begin{array}{lll} \pi_{k,m} & = & f'_{\tau_m, x_{[n]}}\Big(x_{[n]}^k, x_m^k, \Psi_m\Big) + g'_{\tau_m, x_{[n]}}\Big(x_{[n]}^k, x_m^k, \Psi_m\Big)\pi_{k,m,1} \\ & & + \begin{pmatrix} A_{1,m}^\top \\ \vdots \\ A_{\tau_m - 1, m}^\top \end{pmatrix} \pi_{k,m,2} + [\beta_{m,1}^0, \ldots, \beta_{m,1}^{k-1}]\pi_{k,m,3}, \end{array} \tag{6.48}$$

where $f'_{\tau_m, x_{[n]}}\Big(x_{[n]}^k, x_m^k, \Psi_m\Big)$ is a subgradient of convex function $f_{\tau_m}(\cdot, x_m^k, \Psi_m)$ at $x_{[n]}^k$ and the i-th column of matrix $g'_{\tau_m, x_{[n]}}(x_{[n]}^k, x_m^k, \Psi_m)$ is a subgradient at $x_{[n]}^k$ of the i-th component of convex function $g'_{\tau_m, x_{[n]}}(\cdot, x_m^k, \Psi_m)$.

Update θ_n^k and β_n^k computing

$$\theta_n^k = \sum_{m \in C(n)} p_{k,m} \Phi_m \mathfrak{Q}_n^{k-1}(x_{[n]}^k, \xi_m) \text{ and } \beta_n^k = \sum_{m \in C(n)} p_{k,m} \Phi_m \pi_{k,m} \tag{6.49}$$

where $p_{k,m}$ satisfies:

$$\theta_n^k = \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_n^{k-1}(x_{[n]}^k, \xi_m) = \sum_{m \in C(n)} p_{k,m} \Phi_m \mathfrak{Q}_n^{k-1}(x_{[n]}^k, \xi_m).$$

End If

End For

End For

Set $\theta_n^k = 0$ and $\beta_n^k = 0$ for every leaf n.

End For

Theorem 6.1 (Convergence analysis of Algorithm 3)**.** *Consider the sequence of random variables $(x_n^k)_{k \in \mathbb{N}^*}, n \in \mathcal{N}$ and random functions $(\mathcal{Q}_n^k)_{k \in \mathbb{N}}, n \in \mathcal{N}$, generated by Algorithm 3. Let Assumptions (H2) hold and assume that (ξ_t) is a discrete random process with a finite set of possible realizations at each stage. Also assume that the samples generated along the iterations are independent: introducing the binary random variables y_n^k such that $y_n^k = 1$ if node n is selected at iteration k and 0 otherwise, the random variables $(y_{n_T^k}^k)_{k \in \mathbb{N}^*}$ are independent. Then,*

(i) *almost surely, for any node* $n \in \mathcal{N}\backslash\{n_0\}$, *we have*

$$\lim_{k\rightarrow+\infty} \ \mathcal{Q}_n^k(x_{[n]}^k) - \mathcal{Q}_n(x_{[n]}^k) = 0.$$

(ii) *Almost surely, we have*

$$\lim_{k\rightarrow+\infty} \ \mathfrak{Q}_{n_0}^k(x_0, \xi_1) = \mathcal{Q}_{n_0}(x_0),$$

i.e., the optimal value of the approximate first stage problems converges to the optimal value of the first stage problem. Moreover, if $f_1(\cdot, \Psi_1)$ *is continuous, almost surely any accumulation point of the sequence* $(x_{n_1}^k)_{k\in\mathbb{N}^*}$ *is an optimal solution of the first stage problem.*

Proof. We provide the main steps of the proof which follows closely the proofs of Sections 3 and 4.

We prove (i) by backward induction on the number of stages. Following the proof of Proposition 3.1, we show that $\mathcal{Q}_n$ is continuous on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_{\tau_n}$ for all $n \in \mathcal{N}\backslash\{n_0\}$. Following the proof of Lemma 3.2, we show that for all $n \in \mathcal{N}\backslash\{n_0\}$ and k sufficiently large, say $k \geq T_0$, $\mathcal{Q}_n^k$ is Lipschitz continuous and for every m the sequence $(\pi_{k,m})_k$ given by (6.48) is bounded.

We also observe that

$$\mathcal{Q}_n^k(x_{[n]}^k) = \theta_n^k, \forall k \geq 1, \ \forall n \in \mathcal{N}\backslash\{n_0\}. \tag{6.50}$$

Indeed, if n is a leaf the above relation holds by definition of θ_n^k and $\mathcal{Q}_n^k$. Let us show (6.50) when n is not a leaf. In this case, if $k \notin \mathcal{S}_n = \{k \geq 1 : n_{\tau_n}^k = n\}$ then using (6.47), we have $\theta_n^k = \mathcal{Q}_n^{k-1}(x_{[n]}^k)$ and $\mathcal{Q}_n^k(x_{[n]}^k) = \max(\theta_n^k, \mathcal{Q}_n^{k-1}(x_{[n]}^k)) = \theta_n^k$. To show (6.50) for $k \in \mathcal{S}_n$, observe that for a given node n, a new cut is added at iteration k for $\mathcal{Q}_n$ only when $k \in \mathcal{S}_n$. It follows that

$$\mathcal{Q}_n^k(x) = \max_{\ell\in\mathcal{S}_n^k} \theta_n^\ell + \langle \beta_n^\ell, x - x_{[n]}^\ell \rangle$$

where $\mathcal{S}_n^k = \{1 \leq j \leq k : j \in \mathcal{S}_n\} \bigcup \{0\}$. For $k \in \mathcal{S}_n$ and $\ell \in \mathcal{S}_n^k$ with $1 \leq \ell < k$ we have

$$\begin{array}{lll}
\theta_n^k & = & \displaystyle\sup_{p\in\mathcal{P}_n} \sum_{m\in C(n)} p_m \Phi_m \mathfrak{Q}_n^{k-1}(x_{[n]}^k, \xi_m) \text{ using (6.49) and the fact that } k \in \mathcal{S}_n,\\
& \geq & \displaystyle\sup_{p\in\mathcal{P}_n} \sum_{m\in C(n)} p_m \Phi_m \mathfrak{Q}_n^{\ell-1}(x_{[n]}^k, \xi_m) \text{ by monotonicity,}\\
& \geq & \displaystyle\sup_{p\in\mathcal{P}_n} \sum_{m\in C(n)} p_m \Phi_m \Big(\mathfrak{Q}_n^{\ell-1}(x_{[n]}^\ell, \xi_m) + \langle \pi_{\ell,m}, x_{[n]}^k - x_{[n]}^\ell \rangle \Big) \text{ by definition of } \pi_{\ell,m},\\
& \geq & \displaystyle\sum_{m\in C(n)} p_{\ell,m} \Phi_m \Big(\mathfrak{Q}_n^{\ell-1}(x_{[n]}^\ell, \xi_m) + \langle \pi_{\ell,m}, x_{[n]}^k - x_{[n]}^\ell \rangle \Big) \text{ since } (p_{\ell,m})_{m\in C(n)} \in \mathcal{P}_n,\\
& \geq & \theta_n^\ell + \langle \beta_n^\ell, x_{[n]}^k - x_{[n]}^\ell \rangle \text{ using (6.49) and the fact that } \ell \in \mathcal{S}_n.
\end{array}$$

Observing that $k \in \mathcal{S}_n^k$, we get $\mathcal{Q}_n^k(x_{[n]}^k) = \max\Big(\theta_n^k, \max\Big(\theta_n^\ell + \langle \beta_n^\ell, x_{[n]}^k - x_{[n]}^\ell \rangle, \ell < k, \ell \in \mathcal{S}_n^k\Big)\Big) = \theta_n^k$ which shows (6.50). We now prove (i) by induction (the proof is similar to the proof of (i) in Theorem 4.1).

The induction hypothesis is that for each node m of stage $t+1$,

$$\lim_{k\rightarrow+\infty} \ \mathcal{Q}_m(x_{[m]}^k) - \mathcal{Q}_m^k(x_{[m]}^k) = 0. \tag{6.51}$$

The above relation is satisfied for every leaf m of the tree. Now assume that the induction hypothesis is true for each node m of stage $t+1$ for some $t \in \{1, \ldots, T\}$. We want to show that for each node n of stage t,

$$\lim_{k\rightarrow+\infty} \ \mathcal{Q}_n(x_{[n]}^k) - \mathcal{Q}_n^k(x_{[n]}^k) = 0. \tag{6.52}$$

We first show that for each node n of stage t,

$$\lim_{k\rightarrow+\infty,\, k\in\mathcal{S}_n} \ \mathcal{Q}_n(x_{[n]}^k) - \mathcal{Q}_n^k(x_{[n]}^k) = 0. \tag{6.53}$$

We deduce from the induction hypothesis (6.51) and Lemma A.1 in [7] that for each node m of stage $t+1$

$$\lim_{k\rightarrow+\infty} \ \mathcal{Q}_m(x_{[m]}^k) - \mathcal{Q}_m^{k-1}(x_{[m]}^k) = 0. \tag{6.54}$$

We then have for $k \in \mathcal{S}_n$

$$\begin{array}{rcl} \mathfrak{Q}_n^{k-1}\Big(x_{[n]}^k, \xi_m\Big) & = & F_{\tau_m}\Big(x_{[m]}^k, \Psi_m\Big) - \mathcal{Q}_m\Big(x_{[m]}^k\Big) + \mathcal{Q}_m^{k-1}\Big(x_{[m]}^k\Big) \\ & \geq & \mathfrak{Q}_n\Big(x_{[n]}^k, \xi_m\Big) - \mathcal{Q}_m\Big(x_{[m]}^k\Big) + \mathcal{Q}_m^{k-1}\Big(x_{[m]}^k\Big) \end{array} \tag{6.55}$$

where the last inequality comes from the relation

$$\mathfrak{Q}_n\Big(x_{[n]}^k, \xi_m\Big) = \left\{ \begin{array}{l} \inf_{x_m} \ F_{\tau_m}(x_{[n]}^k, x_m, \Psi_m) \\ x_m \in X_{\tau_m}\Big(x_{[n]}^k, \xi_m\Big) \end{array} \right\} \leq F_{\tau_m}\Big(x_{[m]}^k, \Psi_m\Big)$$

which holds since $x_m^k \in X_{\tau_m}\Big(x_{[n]}^k, \xi_m\Big)$ for $k \in \mathcal{S}_n, m \in C(n)$. It follows that

$$0 \ \leq \ \mathfrak{Q}_n\Big(x_{[n]}^k, \xi_m\Big) - \mathfrak{Q}_n^{k-1}\Big(x_{[n]}^k, \xi_m\Big) \leq \mathcal{Q}_m\Big(x_{[m]}^k\Big) - \mathcal{Q}_m^{k-1}\Big(x_{[m]}^k\Big). \tag{6.56}$$

Combining (6.54) and (6.56) we obtain that for every node $m \in C(n)$

$$\lim_{k \to +\infty, k \in \mathcal{S}_n} \ \mathfrak{Q}_n\Big(x_{[n]}^k, \xi_m\Big) - \mathfrak{Q}_n^{k-1}\Big(x_{[n]}^k, \xi_m\Big) = 0. \tag{6.57}$$

Next, for $k \in \mathcal{S}_n$,

$$\begin{array}{l} 0 \leq \mathcal{Q}_n(x_{[n]}^k) - \mathcal{Q}_n^k(x_{[n]}^k) = \sup\limits_{p \in \mathcal{P}_n} \sum\limits_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_n(x_{[n]}, \xi_m) - \theta_n^k \\ = \sup\limits_{p \in \mathcal{P}_n} \sum\limits_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_n(x_{[n]}, \xi_m) - \sup\limits_{p \in \mathcal{P}_n} \sum\limits_{m \in C(n)} p_m \Phi_m \mathfrak{Q}_n^{k-1}(x_{[n]}^k, \xi_m) \\ \leq \sup\limits_{p \in \mathcal{P}_n} \sum\limits_{m \in C(n)} p_m \Phi_m (\mathfrak{Q}_n(x_{[n]}, \xi_m) - \mathfrak{Q}_n^{k-1}(x_{[n]}^k, \xi_m)) \end{array}$$

Combining the above relation with (6.57) we have shown (6.53).

Next, following the end of the proof of Theorem 4.1, we show by contradiction and using the Strong Law of Large Numbers that

$$\lim_{k \to +\infty, \, k \notin \mathcal{S}_n} \mathcal{Q}_n(x_{[n]}^k) - \mathcal{Q}_n^{k-1}(x_{[n]}^k) = 0,$$

implying by monotonicity

$$\lim_{k \to +\infty, \, k \notin \tilde{\mathcal{S}}_n} \mathcal{Q}_n(x_{[n]}^k) - \mathcal{Q}_n^k(x_{[n]}^k) = 0, \tag{6.58}$$

which achieves the proof of (ii).

Finally, the proof of (ii) is analogous to the proof of (ii) in Theorem 4.1. □

We have an analogue of Remark 4.2 for Algorithm 3:

Remark 6.2. *Similarly to Algorithm 1, in Algorithm 3, decisions are computed at every iteration for all the nodes of the scenario tree. However, in practice, decisions will only be computed for the nodes of the sampled scenarios and their children nodes. This variant of Algorithm 3 will build the same cuts and compute the same decisions for the nodes of the sampled scenarios as Algorithm 3. For this variant, for a node n, the decision variables* $(x_n^k)_k$ *are defined for an infinite subset* $\tilde{\mathcal{S}}_n$ *of iterations where the sampled scenario passes through the parent node of node n, i.e.,* $\tilde{\mathcal{S}}_n = \mathcal{S}_{\mathcal{P}(n)}$*. With this notation, applying Theorem 6.1-(i), we get for* $t = 2, \ldots, T$,

$$\lim_{k \to +\infty, k \in \tilde{\mathcal{S}}_n} \mathcal{Q}_n^k(x_{[n]}^k) - \mathcal{Q}_n(x_{[n]}^k) = 0,$$

almost surely, while Theorem 6.1-(ii) still holds.

Acknowledgments

The author's research was partially supported by an FGV grant, CNPq grant 307287/2013-0, FAPERJ grants E-26/110.313/2014, and E-26/201.599/2014. We would like to thank the two anonymous reviewers for their suggestions and comments.

References

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

[2] J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, New York, 1997.

[3] J.R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, 33:989–1007, 1985.

[4] J.R. Birge and C. J. Donohue. The Abridged Nested Decomposition Method for Multistage Stochastic Linear Programs with Relatively Complete Recourse. *Algorithmic of Operations Research*, 1:20–30, 2001.

[5] Z.L. Chen and W.B. Powell. Convergent Cutting-Plane and Partial-Sampling Algorithm for Multistage Stochastic Linear Programs with Recourse. *Journal of Optimization Theory and Applications*, 102:497–524, 1999.

[6] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16:69–95, 2005.

[7] P. Girardeau, V. Leclere, and A.B. Philpott. On the convergence of decomposition methods for multistage stochastic convex programs. *Mathematics of Operations Research*, 40:130–145, 2015.

[8] V. Guigues. SDDP for some interstage dependent risk-averse problems and application to hydro-thermal planning. *Computational Optimization and Applications*, 57:167–203, 2014.

[9] V. Guigues and W. Römisch. Sampling-based decomposition methods for multistage stochastic programs based on extended polyhedral risk measures. *SIAM Journal on Optimization*, 22:286–312, 2012.

[10] V. Guigues and W. Römisch. SDDP for multistage stochastic linear programs based on spectral risk measures. *Operations Research Letters*, 40:313–318, 2012.

[11] G. Infanger and D. Morton. Cut sharing for multistage stochastic linear programs with interstage dependency. *Mathematical Programming*, 75:241–256, 1996.

[12] V. Kozmik and D. Morton. Risk-Averse Stochastic Dual Dynamic Programming. *Optimization OnLine*, 2013.

[13] K. Linowsky and A. B. Philpott. On the convergence of sampling-based decomposition algorithms for multistage stochastic programs. *Journal of optimization theory and applications*, 125(2):349–366, 2005.

[14] M.V.F. Pereira and L.M.V.G Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

[15] A. Philpott and V. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research*, 218:470–483, 2012.

[16] A. Philpott, V. de Matos, and E. Finardi. On solving multistage stochastic programs with coherent risk measures. *Operations Research*, 61:957–970, 2013.

[17] A. B. Philpott and Z. Guan. On the convergence of stochastic dual dynamic programming and related methods. *Operations Research Letters*, 36:450–455, 2008.

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

[19] T. Rockafellar. Conjugate Duality and Optimization. *No 16 in Conference Board of Math. Sciences Series, SIAM Publications*, pages 1–79, 1974.

[20] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31:544–561, 2006.

[21] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.

[22] A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209:63–72, 2011.

[23] A. Shapiro, W. Tekaya, J.P. da Costa, and M.P. Soares. Worst-case-expectation approach to optimization under uncertainty. *Operations Research*, 61:1435–1449, 2013.

[24] S. Uryasev. Conditional value-at-risk: Optimization algorithms and applications. *Financial Engineering News*, 14, 2000.

Vincent Guigues: FGV/EMAp, Praia de Botafogo, Rio de Janeiro, Brazil, vguigues@fgv.br