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

VINCENT GUIGUES

Abstract. We prove the almost sure convergence of a class of sampling-based decomposition methods to solve risk-averse multistage stochastic convex programs that satisfy the relatively complete recourse assumption. 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. Finally, we 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 problems have become a standard tool to model a wide range of engineering problems in which one has to make a sequence of decisions, subject to constraints, and 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 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 whose number of immediate descendant nodes is small but with many stages, Pereira and Pinto [13] propose to sample during the forward pass of the ND. This sampling-based variant of the ND is the so-called Stochastic Dual Dynamic Programming algorithm (SDDP), which has been the object of several recent improvements and extensions [17, 14, 8, 9, 7, 11].

In this paper, we are interested in the convergence of SDDP and related algorithms for risk-averse multistage stochastic convex programs. 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 problem, the proof was
later extended to a larger class of algorithms in [12], [15]. 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 [6]. 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 and used when a real-life application is modelled. Regarding the problem formulation, the dynamic constraints also depend on the full history of decisions instead of just the previous decision.

(C) We use key arguments and assumptions from [6], adapted to our risk-averse framework and problem formulation, but the proof is shorter and uses Assumption (H2)-5), unstated in [6], which is necessary to obtain a description of the subdifferential of the recourse functions. The necessity of this assumption is explained in Lemma 2.1. In particular, we do not use the fact that the recourse functions are Lipschitz but only their continuity. Also, Assumption (H2)-4) is a necessary stronger assumption than Assumption (H1)-6) from [6].

(D) In [6], the subdifferential of the recourse functions is obtained solving optimization problems having \( n \) more variables (assuming that decisions at all stages are \( n \)-dimensional) than the ones we solve (at each iteration of the algorithm, a large number of such problems is solved).

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

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

However, the sampling process is less general than the one used in [6]. 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 are convex continuous and the approximate recourse functions are convex Lipschitz continuous on some sets. The subdifferentials of these functions are bounded on these sets.

(iii) At each stage, conditional to the history of the process, the number of possible realizations of the process is finite, each realization having a positive probability.

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 algorithms we consider and prepares the ground
Section 6 establishes the convergence when the process is interstage dependent. Finally, while Sections 3-5 deal with interstage independent processes, Section 4 explains how to extend the algorithm and convergence analysis for the special case of multistage processes when relatively complete recourse holds. In Section 5, we extend the algorithm and convergence analysis for the special case of multistage processes when relatively complete recourse holds. In Section 4, we explain how to extend the algorithm and convergence analysis for the special case of multistage 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 processes when relatively complete recourse holds.

We use the following notation and terminology:

- The tilde symbol will be used to represent realizations of random variables:
  for random variable $\xi$, $\tilde{\xi}$ is a realization of $\xi$.
- 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 vectors $x, y$, we denote the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ by $[x; y]$.
- 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 $(x_{t_1}, x_{t_1+1}, \ldots, x_{t_2}) \in \mathbb{R}^n \times \ldots \times \mathbb{R}^n$ 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 $\langle x, y \rangle = x^T y$ for $x, y \in \mathbb{R}^n$.
- $I_A(\cdot)$ is the indicator function of the set $A$.
- $\text{Gr}(f)$ is the graph of function $f$.
- $A^* = \{ x : \langle x, a \rangle \leq 0, \forall a \in A \}$ is the polar cone of $A$.
- $N_A(x)$ is the normal cone to $A$ at $x$.
- $T_A(x)$ is the tangent cone to $A$ at $x$.
- $\text{ri}(A)$ is the relative interior of set $A$.
- $\mathbb{B}_n$ is the unit ball in $\mathbb{R}^n$.
- $\text{dom}(f)$ is the domain of function $f$.
- $\text{Aff}(X)$ is the affine hull of $X$.

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

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

$$Q(x) = \begin{cases} \inf_{y \in \mathbb{R}^n} f(x, y) \\ y \in S(x) := \{ y \in Y : Ax + By = b,\ g(x, y) \leq 0 \} \end{cases}.$$  

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

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

the $\varepsilon$-fattening of the set $X$, we make the following assumption (H):

1) $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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 $g_i : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.
3) There exists $\varepsilon > 0$ such that $X^\varepsilon \times Y \subseteq \text{dom}(f)$.

Consider the dual problem

$$\sup_{(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^p} \theta_x(\lambda, \mu)$$
For equivalence (2.5)-(a), we have used the fact that
\[
\theta_x(\lambda, \mu) = \inf_{y \in Y} f(x, y) + \lambda^\top (Ax + By - b) + \mu^\top g(x, y).
\]
We denote by \(\Lambda(x)\) the set of optimal solutions of the dual problem (2.3) and we use the notation
\[
\text{Sol}(x) := \{y \in S(x) : f(x, y) = Q(x)\}
\]
to indicate the solution set to (2.1).

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

**Lemma 2.1.** Consider the value function \(Q\) given by (2.1) and take \(x_0 \in X\) such that \(S(x_0) \neq \emptyset\). Let Assumption (H) hold and assume the Slater-type constraint qualification condition:

there exists \((\bar{x}, \bar{y}) \in X \times \text{ri}(Y)\) such that \(A\bar{x} + B\bar{y} = b\) and \((\bar{x}, \bar{y}) \in \text{ri}(\{g \leq 0\})\).

Then \(s \in \partial Q(x_0)\) if and only if
\[
(s, 0) \in \partial f(x_0, y_0) + \left\{[A^\top; B^\top] \lambda : \lambda \in \mathbb{R}^q\right\} + \left\{\sum_{i \in I(x_0, y_0)} \mu_i \partial g_i(x_0, y_0) : \mu_i \geq 0\right\} + \{0\} \times N_Y(y_0),
\]
where \(y_0\) is any element in the solution set \(\text{Sol}(x_0)\), and with
\[
I(x_0, y_0) = \left\{i \in \{1, \ldots, p\} : g_i(x_0, y_0) = 0\right\}.
\]

In particular, if \(f\) and \(g\) are differentiable, then
\[
\partial Q(x_0) = \left\{\nabla_x f(x_0, y_0) + A^\top \lambda + \sum_{i \in I(x_0, y_0)} \mu_i \nabla_x g_i(x_0, y_0) : (\lambda, \mu) \in \Lambda(x_0)\right\}.
\]

**Proof.** Observe that
\[
Q(x) = \left\{\inf_{y \in \mathbb{R}^n} f(x, y) + \mathbb{I}_{\text{Gr}(S)}(x, y)\right\}
\]
where \(\mathbb{I}_{\text{Gr}(S)}\) is the indicator function of the set
\[
\text{Gr}(S) := \left\{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n : Ax + By = b, g(x, y) \leq 0, y \in Y\right\}
\]
\[
= \left\{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n : Ax + By = b\right\} \cap \left\{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n : g(x, y) \leq 0\right\} \cap \mathbb{R}^m \times Y.
\]

Using Theorem 24(a) in Rockafellar [14], we have
\[
s \in \partial Q(x_0) \iff (s, 0) \in \partial (f + \mathbb{I}_{\text{Gr}(S)})(x_0, y_0)
\]
\[
\iff (s, 0) \in \partial f(x_0, y_0) + N_{\text{Gr}(S)}(x_0, y_0). \quad (a)
\]

For equivalence (2.5)-(a), we have used the fact that \(f\) and \(\mathbb{I}_{\text{Gr}(S)}\) are proper, finite at \((x_0, y_0)\), and
\[
\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(\mathbb{I}_{\text{Gr}(S)})) \neq \emptyset.
\]
The set \( \text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(I_{\text{Gr}(S)})) \) is nonempty because it contains the point \((\bar{x}, \bar{y})\):

\[
(\bar{x}, \bar{y}) \in C_1 \cap \text{ri}(C_2) \cap \mathbb{R}^m \times \text{ri}(Y) = \text{ri}(C_1) \cap \text{ri}(C_2) \cap \mathbb{R}^m \times \text{ri}(Y) = \text{ri}(C_1 \cap C_2) \cap \mathbb{R}^m \times \text{ri}(Y) = \text{ri}(\text{dom}(I_{\text{Gr}(S)})),
\]

(2.7)

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

\[
N_{\text{Gr}(S)}(x_0, y_0) = N_{C_1}(x_0, y_0) + N_{C_2}(x_0, y_0) + N_{\mathbb{R}^m \times Y}(x_0, y_0).
\]

But \( N_{\mathbb{R}^m \times Y}(x_0, y_0) = \{0\} \times N_Y(y_0) \) and standard calculus on normal and tangent cones show that

\[
N_{C_1}(x_0, y_0) = \{ (x, y) : Ax + By = 0, A_x y = 0 \} = \text{Ker}(\{A, B\}),
\]

\[
N_{C_2}(x_0, y_0) = \{ \sum_{i \in I(x_0, y_0)} \mu_i \nabla g_i(x_0, y_0) : \mu_i \geq 0 \}.
\]

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

\[
s = \nabla_x f(x_0, y_0) + A^\top \lambda + \sum_{i \in I(x_0, y_0)} \mu_i \nabla g_i(x_0, y_0), \quad (a)
\]

(2.7)

\[
- \left[ \nabla_y f(x_0, y_0) + B^\top \lambda + \sum_{i \in I(x_0, y_0)} \mu_i \nabla g_i(x_0, y_0) \right] \in N_Y(y_0), \quad (b)
\]

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

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

The following proposition provides conditions ensuring the continuity of \( Q \) and the boundedness of its subdifferential at any point in \( X \):

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

\[
\|s(x)\| \leq \frac{1}{\varepsilon}(M - m).
\]

\( \text{Proof.} \) Finiteness of \( Q \) on \( X^\varepsilon \) 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 \( Q \). Since \( Q \) is convex and since a convex function is Lipschitz continuous on the relative interior of its domain, \( Q \) is Lipschitz continuous and therefore continuous on \( X \).
Next, for every \( x \in X \), for every \( y \in X^\varepsilon \), and \( s(x) \in \partial Q(x) \), we have
\[
Q(y) \geq Q(x) + \langle s(x), y - x \rangle.
\]
Observing that \( M \) and \( m \) are finite (\( Q \) is finite on the compact set \( X^\varepsilon \)), for every \( x \in X \) and \( y \in X^\varepsilon \) we get
\[
M \geq m + \langle s(x), y - x \rangle.
\]
If \( s(x) = 0 \) then (2.8) holds and if \( s(x) \neq 0 \), taking \( y = x + \varepsilon \frac{s(x)}{\|s(x)\|} \in X^\varepsilon \) in the above relation, we obtain (2.8), i.e., \( s(x) \) is bounded.

**Remark 2.3.** If \( \text{Aff}(X) \neq \mathbb{R}^m \), the proof of (2.8) does not work if instead of \((H)-3\), we use the weaker assumption:

\[
\text{there exists } \varepsilon > 0 \text{ such that } (X^\varepsilon \cap \text{Aff}(X)) \times Y \subseteq \text{dom}(f).
\]

Indeed, assuming that \( 0 \in X \), then for any subgradient \( s(x) \in \partial Q(x) \), the orthogonal projection \( \Pi_{\text{Aff}(X)}[s(x)] \) of \( s(x) \) onto \( \text{Aff}(X) \) (which is a subspace since \( 0 \in X \)) is still a subgradient of \( Q \) at \( x \). However, an arbitrary subgradient \( s(x) \in \partial Q(x) \) does not necessarily belong to \( \text{Aff}(X) \) since \( \Pi_{\text{Aff}(X)}[s(x)] + y \in \partial Q(x) \) for any \( y \in \text{Aff}(X)^\perp \) and \( \Pi_{\text{Aff}(X)}[s(x)] + y \not\in \text{Aff}(X) \). For this reason, in the next section, we cannot replace Assumption \((H2)-4\) by Assumption \((H1)-6\) from [6].

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

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

\[
\inf_{x_1 \in X_1(x_0, \xi_1)} f_1(x_1, \Psi_1) + \rho_2 f_1 \left( \inf_{x_2 \in X_2(x_0, \xi_2)} f_2(x_2, \Psi_2) + \ldots + \rho_{T-1} f_T(x_{T-1}, \Psi_{T-1}) \right) + \rho_T f_T(x_T, \Psi_T)
\]

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

\[
X_t(x_0:t-1, \xi_t) = \left\{ x_t \in X_t : g_t(x_0:t, \Psi_t) \leq 0, \sum_{\tau=0}^{t} A_{t\tau} x_\tau = b_t \right\}
\]

for some vector-valued functions \( g_t \), some random vectors \( \Psi_t \) and \( b_t \), some random matrices \( A_{t\tau} \), and where \( \xi_t \) is a discrete random vector with finite support corresponding to the concatenation of the random variables in \( (\Psi_t, b_t, (A_{t\tau})_{\tau=0}^{t-1}) \) in an arbitrary order. In this problem \( x_0 \) is given, \( \xi_1 \) is deterministic, \( (\xi_t) \) is a stochastic process, and setting \( \mathcal{F}_t = \sigma(\xi_1, \ldots, \xi_t) \) and denoting by \( \mathbb{Z}_t \) the set of \( \mathcal{F}_t \)-measurable functions, \( \rho_{t+1, \mathcal{F}_t} : \mathbb{Z}_{t+1} \to \mathbb{R} \) 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 \( (\xi_t) \) is interstage independent. In this case, \( \rho_{t+1, \mathcal{F}_t} \) coincides with its unconditional counterpart \( \rho_{t+1} : \mathbb{Z}_{t+1} \to \mathbb{R} \). To alleviate notation and without loss of generality, we assume that the number \( M \) of possible realizations of \( \xi_t \), the size \( K \) of \( \xi_t \), and \( n \) of \( x_t \) do not depend on \( t \).
For problem (3.9), we can write the following dynamic programming equations:

\begin{equation}
Q_t(x_{1:t-1}) = Q_t(x_{1:t-1}, \xi_t)
\end{equation}

with

\begin{equation}
\begin{aligned}
\Omega_t(x_{1:t-1}, \xi_t) &= \inf_{x_t} \{ F_t(x_{1:t}, \Psi_t) : x_t \in X_t, g_t(x_{0:t}, \Psi_t) \leq 0, \sum_{\tau=0}^t A_{t\tau} x_\tau = b_t, \\
&= \inf_{x_t} F_t(x_{1:t}, \Psi_t) \\
&= \inf_{x_t} F_t(x_{1:t}, \Psi_t),
\end{aligned}
\end{equation}

We set \( \lambda_t \) for some convex subset \( P \).

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

\begin{equation}
\begin{cases}
\inf_{x_1} F_1(x_1, \Psi_1) := f_1(x_1, \Psi_1) + Q_1(x_1) \\
\end{cases}
\end{equation}

with optimal value denoted by \( Q_1(x_0) = \Omega_1(x_0, \xi_1) \).

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

\( \text{(H1) for } t = 2, \ldots, T, \text{ } \xi_t \text{ is a random vector taking values in } \mathbb{R}^K \text{ with discrete distribution and finite support } \{ \xi_t, \ldots, \xi_M \} \text{ while } \xi_1 \text{ is deterministic} \)

(\( \xi_t \)) is the vector corresponding to the concatenation of the elements in \( (\Psi_{tj}, b_{tj}, (A_{t\tau})_{\tau=0,\ldots,t}) \).

Setting \( \Phi_{tj} = P(\xi_t = \xi_{tj}) > 0 \) for \( j = 1, \ldots, M \), using the dual representation of a coherent risk measure \( [1] \), we have

\begin{equation}
Q_t(x_{1:t-1}) = \rho_t(\Omega_t(x_{1:t-1}, \xi_t)) = \sup_{\{ p \in P_t \}} \sum_{j=1}^M p_j \Phi_{tj} \Omega_t(x_{1:t-1}, \xi_{tj})
\end{equation}

for some convex subset \( P_t \) of

\( \mathcal{P}_t = \{ p \in \mathbb{R}^M : p \geq 0, \sum_{j=1}^M p_j \Phi_{tj} = 1 \} \).

Recalling definition (2.12) of the \( \varepsilon \)-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 \( f_t(\cdot, \Psi_{tj}) \) is proper, convex, and lower semicontinuous

3) For every \( j = 1, \ldots, M \), each component of the function \( g_t(x_{0:t}, \cdot, \Psi_{tj}) \) is a convex lower semicontinuous function.

4) There exists \( \varepsilon > 0 \) such that:

4.1) for every \( j = 1, \ldots, M \),

\( \left[ \mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1} \right] \varepsilon \mathcal{X}_t \subset \text{dom } f_t(\cdot, \Psi_{tj}) \).
4.2) for every $j = 1, \ldots, M$, for every $x_{1:t-1} \in \left[ X_1 \times \ldots \times X_{t-1} \right]$, the set $X_t(x_{0:t-1}, \xi_{tj})$ is nonempty.

5) If $t \geq 2$, for every $j = 1, \ldots, M$, there exists $\bar{x}_{tj} = (\bar{x}_{tj1}, \ldots, \bar{x}_{tjT}) \in X_t(x_{0:t-1}, \xi_{tj}) \cap \text{ri}(X_t) \cap \text{ri}(\{g_t(x_0, \cdots, \Psi_{tj}) \leq 0\})$ such that $\bar{x}_{tj} \in X_t(x_0, \bar{x}_{tj1}, \ldots, \bar{x}_{tjT}, \xi_{tj})$.

As shown in Proposition 3.1, Assumption (H2) guarantees that for $t = 1, \ldots, T$, recourse function $Q_t$ is convex and continuous on the set $X_1 \times \ldots \times X_{t-1}$.

**Proposition 3.1.** Under Assumption (H2), for $t = 1, \ldots, T+1$, the recourse function $Q_t$ is convex, finite on $\left[ X_1 \times \ldots \times X_{t-1} \right]$, and continuous on $X_1 \times \ldots \times X_{t-1}$.

**Proof.** The proof is by induction on $t$. The result holds for $t = T+1$ since $Q_{T+1} \equiv 0$. Now assume that for some $t \in \{1, \ldots, T\}$, the function $Q_{t+1}$ is convex, finite on $\left[ X_1 \times \ldots \times X_{t-1} \right]$, and continuous on $X_1 \times \ldots \times X_{t-1}$. Take an arbitrary $x_{1:t-1} \in \left[ X_1 \times \ldots \times X_{t-1} \right]$ and fix $j \in \{1, \ldots, M\}$. Consider the optimization problem \((3.11)\) with $\xi_t = \xi_{tj}$. The optimal value $\Omega_t(x_{1:t-1}, \xi_{tj})$ of this optimization problem is finite because its feasible set $X_t(x_{0:t-1}, \xi_{tj})$ is nonempty (invoking (H2)-4.2)) and compact (invoking (H2)-1) and (H2)-3), and its objective function $f_t(x_{1:t-1}, \Psi_{tj}) + Q_{t+1}(x_{1:t}, \cdot)$ takes finite values on $X_t$ (using (H2)-4.1), (H2)-2), and the induction hypothesis), is proper, and lower semicontinuous (using (H2)-2) and the induction hypothesis). Using Definition \((3.9)\) of $Q_t$, we deduce that $Q_t(x_{1:t-1})$ is finite. Since $x_{1:t-1}$ was chosen arbitrarily in $\left[ X_1 \times \ldots \times X_{t-1} \right]$, we have shown that $Q_t$ is finite on $\left[ X_1 \times \ldots \times X_{t-1} \right]$.

Next, we deduce from Assumptions (H2)-1), (H2)-2), and (H2)-3) that for every $j \in \{1, \ldots, M\}$, $Q_t(\cdot, \xi_{tj})$ is convex on $\left[ X_1 \times \ldots \times X_{t-1} \right]$. Since $\rho_t$ is coherent, it is monotone and convex, and $Q_t(\cdot) = \rho_t(\Omega_t(\cdot, \xi_t))$ is convex on $\left[ X_1 \times \ldots \times X_{t-1} \right]$. Since $X_1 \times \ldots \times X_{t-1}$ is a compact subset of the relative interior of the domain of convex function $Q_t$, we have that $Q_t$ is continuous on $X_1 \times \ldots \times X_{t-1}$. □

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 $Q_t$, $t = 2, \ldots, T+1$, building polyhedral lower approximations $Q_t^k$, $t = 2, \ldots, T+1$, of these functions of the form

$$Q_t^k(x_{1:t-1}) = \max_{0 \leq \ell \leq k} \left( \theta_t^\ell + \langle \beta_t^\ell, x_{1:t-1} - x_{1:t-1}^\ell \rangle \right) = \max_{0 \leq \ell \leq k} \left( \theta_t^\ell + \langle \beta_{1t}^\ell, x_{1:t-2} - x_{1:t-2}^\ell \rangle + \langle \beta_{2t}^\ell, x_{1:t-1} - x_{1:t-1}^\ell \rangle \right).$$

Since $Q_{T+1} \equiv 0$ is known, we have $Q_{T+1}^k \equiv 0$ for all $k$, i.e., $\theta_{T+1}^k$ and $\beta_{T+1}^k$ are null for all $k \in \mathbb{N}$. At iteration $k \geq 1$ of Algorithm 1, decisions $(x_1^k, \ldots, x_{T}^k)$ are computed on a sample $(\xi_1^k, \xi_2^k, \ldots, \xi_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$ replacing the (unknown) recourse functions $Q_t$, $t = 2, \ldots, T+1$, by $Q_t^{k-1}$, $t = 2, \ldots, T+1$ and the coefficients $(\theta_t^k, \beta_t^k)$, $t = 2, \ldots, T+1$, are computed. In Lemma 3.2 below, we show that the coefficients $(\theta_t^k, \beta_t^k)$ computed in Algorithm 1 define valid cuts for $Q_t$, i.e., $Q_t \geq Q_t^k$ for all $k \in \mathbb{N}$.
Algorithm 1: Multistage stochastic decomposition algorithm to solve (3.10).

Initialization. Set $Q^0_t = -\infty$ for $t = 2, \ldots, T$, and $Q^0_{T+1} = 0$, i.e., set $\theta^0_T = 0$, $\beta^0_{t+1} = 0$ for $t = 1, \ldots, T$, and $\theta^0_1 = -\infty$ for $t = 1, \ldots, T - 1$.

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

For $t = 2, \ldots, T$, let $Q^{k-1}_t(x_{1:t-1}, \xi_t)$ be the optimal value of the optimization problem

$$\inf_{x_t} P_t^{k-1}(x_{1:t}, \Psi_t) := f_t(x_{1:t}, \Psi_t) + Q^{k-1}_t(x_{1:t})$$

$$x_t \in \mathcal{X}_t, g_t(x_{0:t}, \Psi_t) \leq 0, \quad \sum_{\tau = 0}^{t} A_{\tau t} x_\tau = b_t$$

(3.14)

and let $Q^{k-1}_t(x_0, \xi_1)$ be the optimal value of the problem above for $t = 1$.

For some realization $(\xi^k_t, \xi^k_s, \ldots, \xi^k_T)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$, compute for $t = 1, \ldots, T$, an optimal solution $x^k_t$ of (3.14) taking $(x_{0:t-1}, \xi_t) = (x^k_{0:t-1}, \xi^k_t)$ solving

$$\inf_{x_t} f_t(x^k_{1:t-1}, x_t, \Psi_t) + z$$

$$g_t(x^k_{0:t-1}, x_t, \Psi_t) \leq 0, \quad [\pi_{tkj(k)1}]$$

$$\sum_{\tau = 0}^{t-1} \Delta^k_{\tau t} x^k_\tau + \Delta^k_{\tau t} x_t = \tilde{b}^k_t, \quad [\pi_{tkj(k)2}]$$

$$z \geq \theta^k_{t+1} + \langle \beta^k_{t+1}, x^k_{1:t-1} - x^k_{1:t-1} \rangle + \langle \beta^k_{t+12}, x_t - x^k_t \rangle, \quad 0 \leq \ell \leq k - 1, \quad [\pi_{tkj(k)3}]$$

(3.15)

with the convention that $x^k_0 = x_0$ and $\xi^k_t = \xi_1$ for all $k$. In the above problem, we have denoted by $\pi_{tkj(k)1}$, $\pi_{tkj(k)2}$, and $\pi_{tkj(k)3}$ the optimal Lagrange multipliers associated with respectively the first, second, and third group of constraints where $j(k)$ is an index such that $\xi^k_t = \xi_{j(k)}$.

For $t = 2, \ldots, T$

For each $j \in \{1, \ldots, M\}$ such that $\xi_{j} \neq \xi^k_t$, i.e., for $j \neq j(k)$,

solve the optimization problem

$$\inf_{x_t} f_t(x^k_{1:t-1}, x_t, \Psi_t) + Q^{k-1}_t(x^k_{1:t-1}, x_t)$$

$$g_t(x^k_{0:t-1}, x_t, \Psi_t) \leq 0, \quad [\pi_{tkj1}]$$

$$A_{tkj} x_t = b_{tj} - \sum_{\tau = 0}^{t-1} A_{\tau tkj} x^k_\tau, \quad [\pi_{tkj2}]$$

$$x^k_t \in \mathcal{X}_t,$$  

(3.16)

with optimal value $Q^{k-1}_t(x^k_{1:t-1}, \xi_{tj})$ and let $x^{k,j}_t$ be an optimal solution. In the above problem, we have denoted by $\pi_{tkj1}$, $\pi_{tkj2}$, and $\pi_{tkj3}$ the optimal Lagrange multipliers associated with respectively the first, second, and third group of constraints. We also set $x^{k,j(k)}_t = x^k_t$.

End For
For \( j = 1, \ldots, M \),
Compute \( p_{tkj} \), such that
\[
\rho_t \left( \Omega_t^{k-1}(x_{1:t-1}^k, \xi_t) \right) = \sup_{p \in P_t} \sum_{j=1}^{M} p_j \Phi_t^{k-1}(x_{1:t-1}^k, \xi_t) = \sum_{j=1}^{M} p_{tkj} \Phi_t^{k-1}(x_{1:t-1}^k, \xi_t).
\]
Compute
\[
\pi_{tkj} = f_t' \left( x_{1:t-1}^k, \xi_t; x_{1:t-1}^k, \Psi_{tj} \right) + g_t' \left( x_{0:t-1}^k, x_{1:t-1}^k, \Psi_{tj} \right) \pi_{tkj},
\]
where \( f_t' (x_{1:t-1}^k, \xi_t; x_{1:t-1}^k, \Psi_{tj}) \) is a subgradient of convex function \( f_t(x_{1:t-1}^k, \xi_t; x_{1:t-1}^k, \Psi_{tj}) \) at \( x_{1:t-1}^k \) and the \( i \)-th column of matrix \( g_t'(x_{0:t-1}^k, x_{1:t-1}^k, \Psi_{tj}) \) is a subgradient at \( x_{1:t-1}^k \) of the \( i \)-th component of convex function \( g_t(x_{0:t-1}^k, \xi_t; x_{1:t-1}^k, \Psi_{tj}) \).
End For
Compute coefficients
\[
(3.17)
\theta^k_t = \rho_t \left( \Omega_t^{k-1}(x_{1:t-1}^k, \xi_t) \right) = \sum_{j=1}^{M} p_{tkj} \Phi_t^{k-1}(x_{1:t-1}^k, \xi_t) \text{ and } \beta^k_t = \sum_{j=1}^{M} p_{tkj} \Phi_t \pi_{tkj},
\]
making up the new approximate recourse function
\[
Q_t^k(x_{1:t-1}) = \max_{0 \leq \ell \leq k} \left( \theta^k_t + \langle \beta^k_t, x_{1:t-1}^\ell - x_{1:t-1}^\ell \rangle \right).
\]
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
\begin{itemize}
\item[(i)] build the cuts in a backward pass using approximate cost-to-go functions \( Q_t^k \) instead of \( Q_t^{k-1} \);
\item[(ii)] use a number of samples that varies along the iterations;
\item[(iii)] sample from the distribution of \( \xi_t \) (instead of using all realizations \( \xi_{t1}, \ldots, \xi_{tM} \) of \( \xi_t \)) to build the cuts [5], [15];
\item[(iv)] generate the sequence \((x_t^k)_k\) using the Abridged Nested Decomposition Method [4].
\end{itemize}
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:
Lemma 3.2. Consider the sequences $Q_t^k$, $\theta_t^k$, and $\beta_t^k$ generated by Algorithm 1. Under Assumptions (H2), for $t = 2, \ldots, T + 1$, for all $k \in \mathbb{N}$, $Q_t^k$ is convex a.s.
with $Q_t^k \leq Q_t$ a.s. on $[X_1 \times \ldots \times X_{t-1}]^\varepsilon$, the sequences $(\theta_t^k)_{k \geq T-t+1}$, $(\beta_t^k)_{k \geq T-t+1}$, and $(\pi_t(k))_{k \geq T-t+1,j=1,\ldots,M}$, are bounded a.s. and for $k \geq T-t+1$, $Q_t^k$ is a.s.
convex Lipschitz continuous on $[X_1 \times \ldots \times X_{t-1}]^\varepsilon$.

Proof. We show the result by induction on $k$ and $t$. For $t = T+1$ and $k \geq 0$, $\theta_t^k$ and $\beta_t^k$ are bounded since they are null (recall that $Q_{T+1}^k$ is null for all $k \geq 0$) and $Q_{T+1}^k = Q_{T+1} = 0$ is convex and Lipschitz continuous on $X_1 \times \ldots \times X_T$ for $k \geq 0$. Assume now that for some $t \in \{1, \ldots, T\}$, and $k \geq T-t+1$, the functions $Q_t^{k-1}$ for $T-t \leq j \leq k-1$ are convex Lipschitz continuous on $[X_1 \times \ldots \times X_{t-1}]^\varepsilon$ with $Q_t^{k-1} \leq Q_t^{k+1}$. We show that (i) $\theta_t^k$ and $\beta_t^k$ are well defined and bounded;
(ii) $Q_t^k$ is convex Lipschitz continuous on $[X_1 \times \ldots \times X_{t-1}]^\varepsilon$; (iii) $Q_t \geq Q_t^k$ on $[X_1 \times \ldots \times X_{t-1}]^\varepsilon$.

Take an arbitrary $x_{1:t-1} \in [X_1 \times \ldots \times X_{t-1}]^\varepsilon$. Since $Q_t^{k+1} \geq Q_t^{k-1}$ on $[X_1 \times \ldots \times X_t]^\varepsilon$, using definition (3.11) of $Q_t$ and the definition of $Q_t^{k-1}$ given in Algorithm 1, we have

$$Q_t(x_{1:t-1}, \cdot) \geq Q_t^{k-1}(x_{1:t-1}, \cdot)$$

(3.18)

and using the monotonicity of $\rho_t$ (recall that $\rho_t$ is coherent)

$$Q_t(x_{1:t-1}) = \rho_t \left( Q_t(x_{1:t-1}, \xi_t) \right) \geq \rho_t \left( Q_t^{k-1}(x_{1:t-1}, \xi_t) \right) \geq \sup_{p \in \Pi_t} \sum_{j=1}^M p_j \Phi_{tj} Q_t^{k-1}(x_{1:t-1}, \xi_{tj}).$$

(3.19)

Using Assumptions (H2)-1, 2), 3), 4.1), 4.2) and the fact that $Q_t^{k-1}$ is Lipschitz continuous on the compact set $[X_1 \times \ldots \times X_t]^\varepsilon$ (induction hypothesis), we have that $Q_t^{k-1}(x_{1:t-1}, \xi_{tj})$ is finite for all $j \in \{1, \ldots, M\}$. It follows that

$$\theta_t^k = \sum_{j=1}^M p_{tj} \Phi_{tj} Q_t^{k-1}(x_{1:t-1}, \xi_{tj})$$

is finite. Next, using Assumptions (H2)-2), 3), for every $j \in \{1, \ldots, M\}$, the function $Q_t^{k-1}(\cdot, \xi_{tj})$ is convex. Since it is finite on $[X_1 \times \ldots \times X_{t-1}]^\varepsilon$, it is Lipschitz continuous on $X_1 \times \ldots \times X_{t-1}$. This function is thus subdifferentiable on
Also, we have just shown that for every $t \geq 1$:

\[ Q_t^{-1}(x_{1:t-1}, \xi_t) \geq Q_t^{-1}(x_{1:t-1}^*, \xi_t) + (\pi_{tkj}, x_{1:t-1} - x_{1:t-1}^*) \]

(3.20)

Plugging this inequality into (3.19), we obtain for $x_{1:t-1} \in [\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^\varepsilon$:

\[
\begin{align*}
Q_t(x_{1:t-1}) & \geq \sup_{p \in \mathcal{P}_t} \sum_{j=1}^M p_{tkj} \Phi_{tj} Q_t^{-1}(x_{1:t-1}, \xi_t) \\
& \geq \sum_{j=1}^M p_{tkj} \Phi_{tj} Q_t^{-1}(x_{1:t-1}, \xi_t) \\
& \geq \sum_{j=1}^M p_{tkj} \Phi_{tj} Q_t^{-1}(x_{1:t-1}, \xi_t) + \sum_{j=1}^M p_{tkj} \Phi_{tj} (\pi_{tkj}, x_{1:t-1} - x_{1:t-1}^*) \\
& \geq \theta_t^k + (\beta_t^k, x_{1:t-1} - x_{1:t-1}^*)
\end{align*}
\]

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

\[
\|\beta_t^k\| \leq \frac{1}{\varepsilon} \left( Q_t(x_{1:t-1}^* + \varepsilon \frac{\beta_t^k}{\|\beta_t^k\|} - \theta_t^k) \right).
\]

(3.21)

From Proposition 3.1, $Q_t$ is finite on $[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^\varepsilon$. Since $\theta_t^k$ is finite, (3.21) shows that $\beta_t^k$ is bounded:

\[
\|\beta_t^k\| \leq \frac{1}{\varepsilon} \left( \sup_{x_{1:t-1} \in [\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^\varepsilon} Q_t(x_{1:t-1}) - \theta_t^k \right).
\]

(3.22)

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

Finally, using Proposition 2.2, we have that $\pi_{tkj}$ is bounded. More precisely, if $\pi_{tkj} \neq 0$, then relation (3.20) written for $x_{1:t-1}^* = x_{1:t-1}^* + \varepsilon \frac{\pi_{tkj}}{\|\pi_{tkj}\|}$ gives for $k \geq T - t + 2$,

\[
\|\pi_{tkj}\| \leq \frac{1}{\varepsilon} \left( Q_t(x_{1:t-1}^*, \xi_t) - Q_t^{T-t+1}(x_{1:t-1}^*, \xi_t) \right),
\]

where we have used the fact that $Q_t^{k-1} \leq Q_t$ and $Q_t^{k+1} \geq Q_t^k$, 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 $Q_t^k(\cdot, \xi_t)$ is finite on the compact set $[\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^\varepsilon$. Also, we have just shown that for every $t = 2, \ldots, T$, and $j = 1, \ldots, M$, the function $Q_t^{T-t+1}(\cdot, \xi_t)$ is continuous on the compact set $\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}$. It follows that we have for every $t = 2, \ldots, T, k \geq T - t + 2, j = 1, \ldots, M$, we have for $\pi_{tkj}$ the upper bound

\[
\|\pi_{tkj}\| \leq M_0 := \max_{t=2,\ldots,T} \max_{j=1,\ldots,M} \frac{M(t, j)}{\varepsilon} \quad \text{where}
\]

\[
M(t, j) = \max_{x_{1:t-1} \in [\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]^\varepsilon} Q_t(x_{1:t-1}, \xi_t) - \min_{x_{1:t-1} \in [\mathcal{X}_1 \times \ldots \times \mathcal{X}_{t-1}]} Q_t^{T-t+1}(x_{1:t-1}, \xi_t).
\]

\[\square\]
Remark 3.3. In the case when the cuts are computed in a backward pass, we can guarantee that $\theta^t_k$ and $\beta^t_k$ are bounded for all $t = 2, \ldots, T + 1$, and $k \geq 1$ (for $k = 0$, we have $\beta^t_0 = 0$ but $\theta^t_1 = -\infty$ is not bounded).

The following lemma will be useful in the sequel:

Lemma 3.4. Consider the sequences $Q^k_t, x^k_{0:t-1}$, and $\theta^k_t$ generated by Algorithm 1. Under Assumptions (H2), for $t = 2, \ldots, T$, and for all $k \in \mathbb{N}$, we have

$$Q^k_t(x^k_{0:t-1}) = \theta^k_t.$$  

Proof. Observe that by construction $\Omega^k_t \geq \Omega^{k-1}_t$ for every $t = 2, \ldots, T + 1$, and every $k \in \mathbb{N}$. It follows that for fixed $0 \leq \ell \leq k$,

$$\theta^k_t = \sup_{p \in \pi_t} \sum_{j=1}^M p_j \Phi_{tj} \Omega^{k-1}_t(x^k_{1:t-1}, \xi_{tj})$$

$$\geq \sup_{p \in \pi_t} \sum_{j=1}^M p_j \Phi_{tj} \Omega^{\ell-1}_t(x^\ell_{1:t-1}, \xi_{tj})$$

$$\geq \sup_{p \in \pi_t} \sum_{j=1}^M p_j \Phi_{tj} \left( \Omega^{\ell-1}_t(x^\ell_{1:t-1}, \xi_{tj}) + \langle \pi_{tj}, x^k_{1:t-1} - x^\ell_{1:t-1} \rangle \right)$$

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

$$\theta^\ell_t = p_t \left( \Omega^{\ell-1}_t(x^\ell_{1:t-1}, \xi_t) \right) = \sum_{j=1}^M p_{tj} \Phi_{tj} \Omega^{\ell-1}_t(x^\ell_{1:t-1}, \xi_{tj}), \quad \text{and} \quad \beta^\ell_t = \sum_{j=1}^M p_{tj} \Phi_{tj} \pi_{tj},$$

we get

$$\theta^k_t \geq \sum_{j=1}^M p_{tj} \Phi_{tj} \left( \Omega^{\ell-1}_t(x^\ell_{1:t-1}, \xi_{tj}) + \langle \pi_{tj}, x^k_{1:t-1} - x^\ell_{1:t-1} \rangle \right)$$

$$\geq \theta^\ell_t + \langle \beta^\ell_t, x^k_{1:t-1} - x^\ell_{1:t-1} \rangle$$

and

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

\(\square\)

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

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

Proof. This is an immediate consequence of (3.22) and (3.24).  

\(\square\)

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

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 associated to a stage 0 (with decision $x_0$ taken at that node) having one child node associated to the first stage (with $\xi_1$ deterministic).
In the sequel, we use the following notation: $\mathcal{N}$ is the set of nodes and $\mathcal{P} : \mathcal{N} \rightarrow \mathcal{N}$ is the function associating to a node its parent node (the empty set for the root node). For a node $n$ of the tree, we denote by

- $C(n)$ the set of its children nodes (the empty set for the leaves);
- $\xi_n$ the realization of process $(\xi_t)$ at node $n$;
- $\xi_{[n]}$ the history of the realizations of the process $(\xi_t)$ from the root node to node $n$: for a node $n$ of stage $t$, the $i$-th component of $\xi_{[n]}$ is $\xi_{n-i}$ for $i = 1, \ldots, t$.

We will also denote by $\text{Nodes}(t)$ the set of nodes for stage $t$.

**Proposition 4.1.** Consider the sequence of random variables $x^k_t$ generated by Algorithm 1. Let Assumptions (H1), (H2), and (H3) hold. Then with probability one, there exist $x^*_{n, n} \in \mathcal{N}$, and infinite subsets $K_n, n \in \mathcal{N}$, of integers, such that for $t = 1, \ldots, T$,

$$H_1(t) : \forall n \in \text{Nodes}(t), \quad \lim_{k \to +\infty, n \in K_n} (x^k_1, \ldots, x^k_t) = x^*_{[n]} \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_t$$

where the $i$-th component of $x^*_{[n]}$ is $x^*_{n-i}$ for $i = 1, \ldots, t$.

**Proof.** We show $H_1(t)$ for $t = 1, \ldots, T$, by induction on $t$. Since the sequence $(x^k_1)_{k \in \mathbb{N}}$ is a sequence of the compact set $\mathcal{X}_1$, there exists an infinite set $K_1$ such that the sequence $(x^k_1)_{k \in K_1}$ converges to some $x^*_{1}$. This shows $H_1(1)$. Assume now that $H_1(t)$ holds for some $t \in \{1, \ldots, T - 1\}$ and let us show that $H_1(t + 1)$ holds.

Let us take a node $m_0$ of stage $t + 1$ and consider its parent node $n$. Let us partition $K_n$ as $K_n = \bigcup_{m \in C(n)} K'_m$ where

$$(4.25) \quad K'_m = \{k \in K_n : (\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_{t+1}^k) = \xi_{[m]}\}.$$

Note that for every $m \in C(n)$, the set $K'_m$ has an infinite number of elements. Indeed, since $K_n$ is infinite, the set of samples $(\tilde{\xi}_1^k, \ldots, \tilde{\xi}_{t+1}^k)_{k \in K_n}$ constitute an infinite set of samples of $(\xi_1, \ldots, \xi_{t+1})$ that all pass through node $n$. Due to Assumption (H3) and the Borel-Cantelli lemma, for every child node $m$ of node $n$, the set of samples with indices in $K_n$ that pass through $m$ is infinite and this set is $K'_m$. In particular, since the sequence $(x^k_{t+1})_{k \in K'_m}$ is an infinite sequence from the compact set $\mathcal{X}_{t+1}$, there exists $x^*_{m_0} \in \mathcal{X}_{t+1}$ and an infinite subset $K_{m_0} \subset K'_m \subset K_n = \mathcal{P}(m_0)$ such that

$$(4.26) \quad \lim_{k \to +\infty, k \in K_{m_0}} x^k_{t+1} = x^*_{m_0}.$$

Since node $n$ belongs to stage $t$, using the induction hypothesis, we have

$$\lim_{k \to +\infty, k \in K_n} (x^k_1, \ldots, x^k_t) = x^*_{[n]} \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_t.$$

Since $K_{m_0}$ is an infinite set contained in $K_n$, we also have

$$\lim_{k \to +\infty, k \in K_{m_0}} (x^k_1, \ldots, x^k_t) = x^*_{[n]} \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_t,$$

which, combined with (4.26), implies

$$\lim_{k \to +\infty, k \in K_{m_0}} (x^k_1, \ldots, x^k_t, x^k_{t+1}) = (x^*_{[n]}, x^*_{m_0}) = x^*_{[m_0]} \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_{t+1}.$$
(since \(n\) is the parent node of \(m_0\)) and achieves the induction step.

□

For convenience, we shall denote by \((y(m, k))_{k \geq 0}\) the sequence of iterations, sorted in ascending order, belonging to \(K_m\). The following lemma will be useful in the sequel:

**Lemma 4.2.** Assume that for some \(t = 1, \ldots, T\), and some \(m \in \text{Nodes}(t)\), we have

\[
\lim_{k \to +\infty, k \in K_m} Q_{t+1}(x_{1:t}^k) - Q_{t+1}(x_{1:t}^k) = 0
\]

where sets \(K_m\) are defined in the proof of Proposition 4.7. Then

\[
\lim_{k \to +\infty, k \in K_m} Q_{t+1}(x_{1:t}^k) = Q_{t+1}(x_{1:t}^*)
\]

where \(x_{1:t}^*\) is defined in the proof of Proposition 4.4.

**Proof.** For some \(t \in \{1, \ldots, T\}\) and \(m \in \text{Nodes}(t)\), assume that (4.27) holds. We want to show that (4.28) holds. Fix \(\varepsilon > 0\) and let \(L\) be the Lipschitz constant for functions \(Q_{t+1}^k, k \geq T\), given by Lemma 5.6. From Proposition 4.1, we have \(\lim_{k \to +\infty} y_{1:t}^{y(m,k)} = x_{1:t}^*\). It follows that there exists \(k(m,0)\) such that for \(k \geq k(m,0)\), we have

\[
\|x_{1:t}^{y(m,k)} - x_{1:t}^*\| \leq \varepsilon/4L.
\]

Next, since \(Q_{t+1}\) is continuous on \(\mathcal{X}_1 \times \ldots \times \mathcal{X}_t\) with \(x_{1:t}^{y(m,k)}, x_{1:t}^* \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_t\), we have

\[
\lim_{k \to +\infty} y_{1:t}^{y(m,k)} = y_{1:t}^*
\]

Then for any \(k \geq k(m,1) + 1\), we deduce that

\[
Q_{t+1}^{y(m,k,m),1}(x_{1:t}^{y(m,k)}) - Q_{t+1}(x_{1:t}^*) \leq L\|x_{1:t}^{y(m,k)} - x_{1:t}^*\| \leq \varepsilon/2.
\]

In (4.33), the first inequality is due to \(Q_{t+1}^{y(m,k,m),1} \leq Q_{t+1}^{y(m,k,m)}\) because \(y(m, k) - 1 \geq y(m, k(m,1))\) and the second inequality to \(Q_{t+1}^{y(m,k,m)} \leq Q_{t+1}\). We deduce that for \(k \geq k(m,1) + 1\), we have

\[
|Q_{t+1}^{y(m,k,m),1}(x_{1:t}^{y(m,k)}) - Q_{t+1}(x_{1:t}^*)| \leq L\|x_{1:t}^{y(m,k)} - x_{1:t}^*\| \leq \varepsilon/2.
\]

using (4.32), the fact that \(Q_{t+1}^{y(m,k,m),1}\) is \(L\)-Lipschitz, and (4.29). From (4.33) and the above relation, we get \(|Q_{t+1}^{y(m,k,m),1}(x_{1:t}^{y(m,k)}) - Q_{t+1}(x_{1:t}^*)| \leq \varepsilon\) for \(k \geq k(m,1) + 1\) which achieves the proof. □
Theorem 4.3 shows the convergence of the sequence \( Q_k^i(x_0, \xi_1) \) to \( Q_1(x_0) \) and that any accumulation point of the sequence \( (x_1^k)_{k \in \mathbb{N}} \) is an optimal solution of the first stage problem (3.12).

**Theorem 4.3** (Convergence analysis of Algorithm 1 - Interstage independent process). Consider the sequences of random variables \( x_k^i \) and random functions \( Q_k^i \) generated by Algorithm 1. Let Assumptions (H1), (H2), and (H3) hold. Then there exist \( x^*_n, n \in \mathcal{N} \), and infinite subsets \( K_n, n \in \mathcal{N} \), of integers, defined in the proof of Proposition 4.1, such that a.s.

(i) for \( t = 2, \ldots, T \),

\[
H_1(t) : \forall n \in \text{Nodes}(t), \quad \lim_{k \to +\infty, k \in K_n} (x_1^k, \ldots, x_t^k) = x^*_n \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_t
\]

where the \( i \)-th component of \( x^*_n \) is \( x^*_{t-1}(n) \) for \( i = 1, \ldots, t \).

(ii) For \( t = 2, \ldots, T \),

\[
H_2(t) : \forall n \in \text{Nodes}(t-1), \quad \lim_{k \to +\infty, k \in K_n} Q_k^i(x_1^k, \ldots, x_{t-1}^k) = Q_t(x^*_n) \text{ a.s.}
\]

(iii) \( \lim_{k \to +\infty} Q_k^i(x_0, \xi_1) = Q_1(x_0) \) and if \( f_1(\cdot, \Psi_1) \) is continuous on \( \mathcal{X}_1 \), any accumulation point of the sequence \( (x_1^k)_{k \in \mathbb{N}} \) is an optimal solution of the first stage problem (3.12).

**Proof.** Item (i) is Proposition 4.1.

We show (ii) by induction backwards in time. Recalling that \( Q_{T+1} = Q_{T+1}^k = 0 \), we have \( Q_T^k(x_k^1, \ldots, y) = Q_T^k(x_k^1, \ldots, y) \) for all \( k \in \mathbb{N} \). It follows that for all \( k \in \mathbb{N} \),

\[
Q_T^k(x_1^k, \ldots, y) = \lim_{k \to +\infty, k \in K_n} Q_T^k(x_1^k, \ldots, y) = Q_T^k(x_1^k, \ldots, y)
\]

From (i), for every node \( n \in \text{Nodes}(T-1) \), we have \( \lim_{k \to +\infty, k \in K_n} x_{1:T-1}^k = x^*_n \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_{T-1} \). From Proposition 3.1, \( \mathcal{Q}_T \) is continuous on \( \mathcal{X}_1 \times \cdots \times \mathcal{X}_{T-1} \). It follows that for every node \( n \in \text{Nodes}(T-1) \), taking the limit when \( k \to +\infty, k \in K_n \) in (3.12), we obtain

\[
\lim_{k \to +\infty, k \in K_n} Q_T^k(x_1^k, \ldots, y) = Q_T^k(x_1^k, \ldots, y)
\]

which is \( H_2(T) \).

Now assume that \( H_2(t+1) \) holds for some \( t \in \{2, \ldots, T-1\} \). We want to show that \( H_2(t) \) holds. Take a node \( n \in \text{Nodes}(t-1) \), consider an arbitrary child node \( m \) of \( n \), and \( k \in K_n \) with \( K_n \) defined in the proof of Proposition 4.1. For all \( k \in K_m \), since \( K_m \subset K_{m} \subset K_n \) with \( K_{m} \) given by (3.14), we have \( (\xi_1^k, \ldots, \xi_{t-1}^k) = \xi_m \).

In particular, \( \xi_t^k = \xi_m \) for the iterations \( k \in K_m \), the sampled scenarios pass through node \( m \). We obtain

\[
Q_t^k(x_{1:t-1}, \xi_m) = Q_t^k(x_{1:t-1}, \xi_t^k) = f_t(x_{1:t}^k, \Psi_m) + Q_t^k(x_{1:t}^k)\]

by definition of \( x_t^k \),

\[
= f_t(x_{1:t}^k, \Psi_m) - \Psi_t(x_{1:t}^k) + Q_t^k(x_{1:t}^k)
\]

by definition of \( F_t \).

Next, since \( x_t^k \in X_t(x_{0:t-1}^k, \xi_t^k) \), we have

\[
Q_t(x_{1:t-1}, \xi_m) = Q_t(x_{1:t-1}, \xi_t^k) = \inf_{x_t} F_t(x_{1:t-1}, x_t, \Psi_m) \leq F_t(x_{1:t}^k, \Psi_m).
\]


Plugging the above inequality into (4.35), we obtain for $k \in K_m$

$$0 \leq \Omega_t(x_{1:t-1}^k, \xi_m) - \Omega_t^{k-1}(x_{1:t-1}^k, \xi_m) \leq \Omega_{t+1}(x_{1:t}^k) - \Omega_{t+1}^{k-1}(x_{1:t}^k).$$

(4.37)

Since $m \in \text{Nodes}(t)$, the induction hypothesis, (i), and the continuity of $\Omega_{t+1}$ gives

$$\lim_{k \to +\infty, k \in K_m} \Omega_{t+1}(x_{1:t}^k) - \Omega_{t+1}^{k-1}(x_{1:t}^k) = 0.$$ 

As a result, using Lemma 4.2 (i), and the continuity of $\Omega_{t+1}$, we get

$$\lim_{k \to +\infty, k \in K_m} \Omega_{t+1}(x_{1:t}^k) - \Omega_{t+1}^{k-1}(x_{1:t}^k) = 0.$$ 

Combined with (4.37), we have shown that

(4.38) $$\lim_{k \to +\infty, k \in K_m} \Omega_t(x_{1:t-1}^k, \xi_m) - \Omega_t^{k-1}(x_{1:t-1}^k, \xi_m) = 0.$$ 

Let us now fix $\varepsilon > 0$. We have shown in the proof of Proposition 3.1 that for every $m \in C(n)$ (with $|C(n)| = M$ finite), the function $\Omega_t(\cdot, \xi_m)$ is continuous on $X_1 \times \cdots \times X_{t-1}$. Since the sequence $(x_{1:t-1}^k)_{k \in K_n}$ with $x_{1:t-1}^k$ converges to $x^*_n$, there exists $k_1$ such that for all $k \in K_n$ and $k \geq k_1$, we have for all $m \in C(n)$,

(4.39) $$|\Omega_t(x_{1:t-1}^k, \xi_m) - \Omega_t(x^*_n, \xi_m)| \leq \frac{\varepsilon}{10}.$$ 

Now recall that in Lemma 3.2 we obtained the upper bound $M_0$ given by (3.23) for $\|\tau_{ijk}\|$, valid for every $t = 2, \ldots, T$, $j = 1, \ldots, M$, and $k \geq T$. We can assume without loss of generality that $M_0 > 0$ (if $M_0 = 0$, it can be replaced by, say, $M_0 + 1 > 0$). Since the sequence $(x_{1:t-1}^k)_{k \in K_n}$ converges to $x^*_n$, there exists $k_2 \in K_n$ with $k_2 \geq T$ such that for $k \geq k_2$, we have for every $k \in K_n$

(4.40) $$\|x_{1:t-1}^k - x^*_n\| \leq \frac{\varepsilon}{10M_0}.$$ 

Using (4.38), for every $m \in C(n)$, there exists $k(3, m) \in K_m$ with $k(3, m) \geq \max(k_1, k_2, T)$ such that for every $k \in K_m$ satisfying $k \geq k(3, m)$, we have

(4.41) $$0 \leq \Omega_t(x_{1:t-1}^k, \xi_m) - \Omega_t^{k-1}(x_{1:t-1}^k, \xi_m) \leq \frac{\varepsilon}{10}.$$ 

From the continuity of $\Omega_t$ and the fact that the sequence $(x_{1:t-1}^k)_{k \in K_n}$ converges to $x^*_n$, there exists $k_4$ such that for $k \in K_n$ and $k \geq k_4$, we have

(4.42) $$|\Omega_t(x_{1:t-1}^k) - \Omega_t(x^*_n)| \leq \frac{\varepsilon}{2}.$$ 

Take now an arbitrary

$k \geq \max(k_1, k_2, \max(k(3, m), m \in C(n)), k_4) = \max(\max(k(3, m), m \in C(n)), k_4)$

with $k \in K_n$. Let $\text{Index}(m)$ be such that $\xi_{\text{Index}(m)} = \xi_m$. Since $k \geq k(3, m)$ for every $m \in C(n)$, using the convexity of $\Omega_t^{k(3,m)-1}(\cdot, \xi_m)$ and the fact that $\pi_{kk(3,m)}\text{Index}(m)$
is a subgradient of convex function $\Omega_t^{k(3,m)−1}(\cdot, \xi_m)$ at $x_{1:t−1}$, we have

(4.43)
\[
\theta_k^t = Q_t^k(x_{1:t−1}) = \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_{t,\text{Index}(n)} \Omega_t^{k(3,m)−1}(x_{1:t−1}, \xi_m)
\]
\[
\geq \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_{t,\text{Index}(n)} \Omega_t^{k(3,m)−1}(x_{1:t−1}, \xi_m)
\]
\[
\geq \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_{t,\text{Index}(n)} \left[ \Omega_t^{k(3,m)−1}(x_{1:t−1}, \xi_m) + \langle \pi_{tk(3,m)\text{Index}(n)}, x_{1:t−1}^k - x_{1:t−1}^k \rangle \right].
\]

This gives

(4.44)
\[
0 \leq Q_t(x_{1:t−1}^*) - Q_t^k(x_{1:t−1}) \leq \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_{t,\text{Index}(n)} \left[ \Omega_t(x_{1:t−1}, \xi_m) - \Omega_t^k(x_{1:t−1}, \xi_m) \right]
\]
\[
+ \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_{t,\text{Index}(n)} \left[ \pi_{tk(3,m)\text{Index}(n)}, x_{1:t−1}^k - x_{1:t−1}^k \right]
\]
\[
\leq \sup_{p \in \mathcal{P}_t} \sum_{m \in C(n)} p_m \Phi_{t,\text{Index}(n)} \left[ \left\| \pi_{tk(3,m)\text{Index}(n)} \right\| \left( \left\| x_{1:t−1}^k - x_{1:t−1}^k \right\| + \left\| x_{1:t−1}^k - x_{1:t−1}^k \right\| \right) \right]
\]
\[
\leq \frac{\varepsilon}{2}.
\]

The last inequality was obtained using relations (4.39), (4.40), (4.41), and the fact that $\sum_{m \in C(n)} p_m \Phi_{t,\text{Index}(n)} = 1$ for any $p \in \mathcal{P}_t$. In particular, (4.39) (resp. (4.40)) was used to bound from above the third (resp fourth) of the four terms of the right-hand side of the penultimate inequality. This is possible since (4.39) (resp. (4.40)) which holds for $k \in K_n$ with $k \geq k_1$ (resp. $k \in K_n$ with $k \geq k_2$) was used with $k = k(3, m) \geq \max(k_1, k_2) \geq k_1$ satisfying $k(3, m) \in K_m \subset K_n$ (resp. $k = k(3, m) \geq \max(k_1, k_2) \geq k_2$ satisfying $k(3, m) \in K_m \subset K_n$).

Combining the above relation (4.44) with (4.42), we have shown that for every $\varepsilon > 0$, for every $k \geq \max(\max(k(3, m), m \in C(n)), k_4)$ with $k \in K_n$ we have $|Q_t^k(x_{1:t−1}^k) - Q_t^k(x_{1:t−1}^k)| \leq \varepsilon$. Since the node $n$ was arbitrarily chosen in $\text{Nodes}(t−1)$, we have shown $H(t)$, which achieves the induction step and the proof of (ii).

(iii) By definition of $\Omega_1^{k−1}$ (see Algorithm 1), we have

\[
\Omega_1^{k−1}(x_0, \xi_1) = F_1(x_1^k, \Psi_1) − Q_2(x_1^k) + Q_2^{k−1}(x_1^k)
\]

which implies

(4.45)
\[
0 \leq \Omega_1(x_0, \xi_1) - \Omega_1^{k−1}(x_0, \xi_1) \leq Q_2(x_1^k) - Q_2^{k−1}(x_1^k).
\]

From (ii), $H_2(2)$ holds, which, combined with the continuity of $Q_2$, gives

$$
\lim_{k \to +\infty, k \in K_{n_1}} Q_2^k(x_1^k) - Q_2(x_1^k) = 0
$$
where \( n_1 \) is the unique child of the root node (remember that \( \xi_1 \) is deterministic), i.e., the node associated to the first stage. Using Lemma 4.2 with \( t = 1 \) and the continuity of \( Q_2 \), we obtain that \( \lim_{k \to +\infty, k \in K_{n_1}} Q_2(x_t^k) - Q_2^{k-1}(x_t^k) = 0 \). Combined with (4.35), we have shown that \( \lim_{k \to +\infty, k \in K_{n_1}} Q_1^{k-1}(x_0, \xi_1) = Q_1(x_0, \xi_1) = Q_1(x_0) \).

It follows that for every \( \varepsilon > 0 \), there exists \( k_0 \in K_{n_1} \) such that for every \( k \in K_{n_1} \) with \( k \geq k_0 \), we have \( 0 \leq Q_1(x_0) - Q_1^{k-1}(x_0, \xi_1) \leq \varepsilon \). As a result, for every \( k \in \mathbb{N} \) with \( k \geq k_0 - 1 \), we have

\[
0 \leq Q_1(x_0) - Q_1^k(x_0, \xi_1) \leq Q_1(x_0) - Q_1^{k_0-1}(x_0, \xi_1) \leq \varepsilon,
\]

which shows the convergence of the whole sequence \( (Q_1^k(x_0, \xi_1))_{k \in \mathbb{N}} \) to \( Q_1(x_0) \), i.e., the optimal value of the approximate optimization problem solved at the first stage converges with probability one to the optimal value \( Q_1(x_0) \) of the optimization problem.

Next, consider an accumulation point \( x^* \) of the sequence \( (x_t^k)_{k \in \mathbb{N}} \). There exists a set \( K \) such that the sequence \( (x_t^k)_{k \in K} \) converges to \( x^* \). Recalling that \( n_1 \) is the node associated to the first stage, we can take \( K_{n_1} = K \) in the proof of Proposition 4.1 and by definition of \( x^*_{[n_1]} \) we have \( x^*_{[n_1]} = x^* \). Also, by definition of \( x_t^k \),

\[
(4.46) \quad f_1(x_t^k, \Psi_1^k) + Q_2^{k-1}(x_t^k) = Q_1^{k-1}(x_0, \xi_1^k) = \Omega_1^{k-1}(x_0, \xi_1).
\]

Using \( H_2(2) \), Lemma 4.2 and the continuity of \( Q_2 \) on \( X_1 \), we have

\[
\lim_{k \to +\infty, k \in K_{n_1}} Q_2^{k-1}(x_t^k) = Q_2(x^*_{[n_1]}).
\]

Taking the limit in (4.40) when \( k \to +\infty \) with \( k \in K_{n_1} \), we obtain

\[ Q_1(x_0) = f_1(x^*, \Psi_1) + Q_2(x^*) = F_1(x^*, \Psi_1). \]

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

5. Convergence analysis for 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 \( X_t = \mathbb{R}^n \) and we do not make Assumptions (H2)-4)-5). More precisely, instead of (4.11), we consider the following dynamic programming equations corresponding to multistage stochastic linear programs that do not satisfy the relatively complete recourse assumption: we set \( Q_{T+1} \equiv 0 \) and for \( t = 2, \ldots, T \), we define \( Q_t(x_{1:t-1}, \xi_t) = \rho_t \left( \Omega_t(x_{1:t-1}, \xi_t) \right) \) now with

\[
\Omega_t(x_{1:t-1}, \xi_t) = \begin{cases} 
\inf_{x_t} F_t(x_{1:t}, \Psi_t) := \Psi_t^\top x_{1:t} + Q_{t+1}(x_{1:t}) \\
\sum_{\tau=0}^{\infty} A_{t\tau} x_\tau = b_t, \quad x_1 \geq 0, \\
\inf_{x_t} F_t(x_{1:t}, \Psi_t) \\
x_t \in X_t(x_{0:t-1}, \xi_t).
\end{cases}
\]

At the first stage, we solve

\[
(5.48) \quad \begin{cases} 
\inf_{x_1} F_1(x_1, \Psi_1) := \Psi_1^\top x_1 + Q_2(x_1) \\
x_1 \in X_1(x_0, \xi_1),
\end{cases}
\]
with optimal value denoted by $Q_1(x_0) = \Omega(x_0, \xi_1)$. If we apply Algorithm 1 to solve (5.48) (in the sense of Theorem 4.3), since Assumption (H2)-4) does not hold, it is possible that one of the problems (5.47) to be solved in the forward passes is unfeasible. In this case, $x^k_{t:t-1}$ is not a feasible sequence of states from stage 1 to stage $t-1$ and we build a separating hyperplane separating $x^k_{t-1}$ 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 is (5.48) 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 [7]. 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 (5.47), before solving (3.15) in the forward pass, we solve for $j = 1, \ldots, M$, the optimization problem

$$\min_{x, y_1, y_2} e^\top (y_1 + y_2)$$

$$= \sum_{\tau = 0}^T A_{\tau j} x^k_\tau + A_{\tau j} x_\tau + y_1 - y_2 = b_{\tau j}, \quad [\pi]$$

$$(x^k_{1:t-1})^\top \tilde{\beta}^\top_{t+1} + x^k_t \tilde{\beta}^\top_{t+2} \leq \tilde{\theta}^\top_{t+1}, \quad \ell = 1, \ldots, K_t, \quad [\tilde{\pi}]$$

$$x_\tau, y_1, y_2 \geq 0,$$

where $e$ is a vector of ones. In the above problem, we have denoted by respectively $\pi$ and $\tilde{\pi}$ optimal Lagrange multipliers for the first and second set of constraints.

If $\hat{\Omega}_t(x^k_{1:t-1}, \xi_{tj})$ is the optimal value of (5.48), noting that $\hat{\Omega}_t(\cdot, \xi_{tj})$ is convex with $s = [A_{11j}; \ldots; A_{4t-1j}] \pi + [\tilde{\beta}^1_t; \ldots; \tilde{\beta}^{K_t}_t] \tilde{\pi}$ belonging to the subdifferential of $\hat{\Omega}_t(\cdot, \xi_{tj})$ at $x^k_{1:t-1}$, if $x^k_{1:t-1}$ is feasible then

$$(5.50) \quad \hat{\Omega}_t(x^k_{1:t-1}, \xi_{tj}) = 0 \geq \hat{\Omega}_t(x^k_{1:t-1}, \xi_{tj}) + s^\top (x^k_{1:t-1} - x^k_{1:t-1}).$$

Inequality (5.50) defines a feasibility cut for $x^k_{1:t-1}$ of the form

$$(5.51) \quad x^k_{1:t-1} \tilde{\beta}^\top_t \leq \tilde{\theta}^\top_t$$

with

$$(5.52) \quad \tilde{\beta}^\top_t = [\tilde{\beta}^\top_{t1}; \tilde{\beta}^\top_{t2}] = s \text{ and } \tilde{\theta}^\top_t = -\hat{\Omega}_t(x^k_{1:t-1}, \xi_{tj}) + s^\top x^k_{1:t-1}.$$

Incorporating these cuts in the forward pass of Algorithm 1, we obtain Algorithm 2.

**Algorithm 2: Multistage stochastic decomposition algorithm to solve (5.48) without the relatively complete recourse assumption.**

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

**While Out = 0**

Set $t = 1$, $x^k_0 = x_0$.

**While** ($t < T$) and (Out = 0)

Set $j = 1$ and OutAux = 0.

**While** ($j \leq M$) and (OutAux = 0),
Solve problem \((5.49)\).

If the optimal value of \((5.49)\) is positive

If \(t = 1\) then

\[\text{Out}=1, \text{OutAux}=1 // \text{the problem is infeasible}\]

Else

compute \((5.52)\) to build feasibility cut \((5.51)\).

Increase \(K_{t-1}\) by one, decrease \(t\) by one, \(\text{OutAux}=1\).

End if

Else \(j \leftarrow j + 1\).
End If

End While

If \(\text{OutAux}=0\),

Sample a realization \(\xi^k_t\) from the distribution of \(\xi_t\) and solve

\[
\begin{aligned}
\inf_{x_t^k} & \left( (x_{1:t}^k - z \right) \\
\text{s.t.} & \sum_{\tau=0} \mathcal{A}_{t\tau} x_{\tau}^k + A_{tj} x_t = b_j, [\pi_{tkj}]_1 \\
& z \geq \theta_{t+1}^\ell + \langle \beta_{t+11}^\ell, x_{t+1}^k - x_{t:t-1}^k \rangle + \langle \beta_{t+12}^\ell, x_t^k - x_{t}^k \rangle, 0 \leq \ell \leq k - 1, [\pi_{tkj}]_2 \\
& (x_{1:t}^k) \geq \theta_{t+1}^\ell, \ell = 1, \ldots, K_t, [\pi_{tkj}]_3 \\
x_t^k \geq 0.
\end{aligned}
\]

\[(5.53)\]

In the above problem, we have denoted by \(\pi_{tkj(1)}, \pi_{tkj(2)}, \text{ and } \pi_{tkj(3)}\)
the optimal Lagrange multipliers associated with respectively the first, second, and third group of constraints where \(j(k)\) is an index such that

\(\xi^k_t = \xi_{t(j(k))}\).

Increase \(t\) by one.

End if

End While

If \(\text{Out}=0\)

For \(t = 2, \ldots, T,\)

For each \(j \in \{1, \ldots, M\}\) such that \(\xi_{tj} \neq \xi^k_t,\)

solve the optimization problem

\[
\begin{aligned}
\inf_{x_t^k} & \left( (x_{1:t-1}^k - x_t) \right) \\
\text{s.t.} & \sum_{\tau=0} \mathcal{A}_{t\tau} x_{\tau}^k + A_{tj} x_t = b_j, [\pi_{tkj}]_1 \\
& z \geq \theta_{t+1}^\ell + \langle \beta_{t+11}^\ell, x_{t+1}^k - x_{t:t-1}^k \rangle + \langle \beta_{t+12}^\ell, x_t^k - x_{t}^k \rangle, 0 \leq \ell \leq k - 1, [\pi_{tkj}]_2 \\
& (x_{1:t}^k) \geq \theta_{t+1}^\ell, \ell = 1, \ldots, K_t, [\pi_{tkj}]_3 \\
x_t^k \geq 0.
\end{aligned}
\]

\[(5.54)\]
with optimal value $\Omega_{t-1}^{k-1}(x_{1:t-1}, \xi_t)$ and let $x_{k}^{k,j}$ be an optimal solution. Compute $p_{tkj}$ such that

$$\rho_t(\Omega_{t-1}^{k-1}(x_{1:t-1}, \xi_t)) = \sup_{p \in P_t} \sum_{j=1}^M p_j \Phi_{tj} \Omega_{t-1}^{k-1}(x_{1:t-1}, \xi_t)$$

$$= \sum_{j=1}^M p_{tkj} \Phi_{tj} \Omega_{t-1}^{k-1}(x_{1:t-1}, \xi_t).$$

Compute

$$\pi_{tkj} = \Psi_{tj}(1 : (n-1)t) + \begin{bmatrix} A_{1tj}^T \\ \vdots \\ A_{(t-1)j}^T \end{bmatrix} \pi_{tkj}$$

$$+ [\beta_{t+1}^0, \ldots, \beta_{t+1}^{k-1}] \pi_{tkj}^2 + [\tilde{\beta}_{t+1}^0, \ldots, \tilde{\beta}_{t+1}^K] \pi_{tkj}^3$$

and coefficients

$$\theta_t^k = \rho_t(\Omega_{t-1}^{k-1}(x_{1:t-1}, \xi_t)) = \sum_{j=1}^M p_{tkj} \Phi_{tj} \Omega_{t-1}^{k-1}(x_{1:t-1}, \xi_t), \quad \beta_t^k = \sum_{j=1}^M p_{tkj} \Phi_{tj} \pi_{tkj},$$

making up the new approximate recourse function

$$Q_t^k(x_{1:t-1}) = \max_{0 \leq \ell \leq k} \left( \theta_t^\ell + \langle \beta_t^\ell, x_{1:t-1} - x_{1:t-1}^\ell \rangle \right).$$

End For
End For
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 4.3:

**Theorem 5.1 (Convergence analysis of Algorithm 2).** Let Assumption (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 (i), (ii), and (iii) stated in Theorem 4.3 hold.

**Proof.** Due to Assumption (H2'), recourse functions $Q_t$ are convex polyhedral and Lipschitzian. Moreover, Assumption (H2') also guarantees that

(a) all linear programs (5.53), (5.54) are feasible and have bounded primal and dual feasible sets. As a result, $Q_t^k$ are also Lipschitzian convex and polyhedral.

(b) The feasible set of (5.49) 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 (5.49) contains the first stage feasible set. As a result, if (5.49) is not feasible, there is no solution to (5.47). 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 just Algorithm 1 and (i),(ii), (iii) can be proved following the proof of Theorem 4.3. □

6. Convergence analysis with interstage dependent processes

Consider a problem of form (3.9) and let Assumption (H2) hold. We assume that the stochastic process $(\xi_t)$ is discrete, interstage dependent, 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 $\xi_1$ deterministic).

In addition to the notation introduced in Section 4, we also define:

- $[n]$: the set of ancestor nodes of $n$ including $n$, i.e., all the nodes that are on the unique scenario in the tree going from node $n_1$ to node $n$;
- $\Phi_m$: the transition probability from node $P(m)$ to node $m$;
- $\tau_n$: the stage associated to node $n$;
- $\xi_n := (\Psi_n, b_n, (A_{nk})_{k=0,\ldots,\tau_n})$ the realization of process $(\xi_t)$ at node $n$;
- $x_n$: the decision taken at node $n$;
- $x_m$: the vector or Cartesian product whose components are all decisions $x_m$ with $m \in [n]$.

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 [10], [7] 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 associating recourse functions to each node 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 $\rho_{t+1|t}$ in formulation (6.5) is given by the collection of the risk measures $(\rho_n)_{n : \tau_n = t+1}$. More precisely, we consider the following dynamic programming equations: for every node $n$ which is neither the root node nor a leaf, we define the recourse function

$$(6.56) \quad Q_n(x_{[n]}) = \sup_{p \in C_n} \sum_{m \in C(n)} p_m \Phi_m Q_n(x_{[n]}, \xi_m)$$

for some convex subset $C_n$ of $D_n = \{ p \in \mathbb{R}^{C(n)} : p \geq 0, \sum_{m \in C(n)} p_m \Phi_m = 1 \}$. 
where \( \Omega_n(x_n, \xi_m) \) is given by

\[
(6.57) \quad \begin{cases}
\min_{x_m} F_{\tau_m}(x_n, x_m, \Psi_m) \\
x_m \in X_{\tau_m}, g_{\tau_m}(x_0, x_n, x_m, \Psi_m) \leq 0,
\end{cases}
\]

with

\[
F_{\tau_m}(x_n, x_m, \Psi_m) = f_{\tau_m}(x_n, x_m, \Psi_m) + Q_m(x_n, x_m).
\]

If \( n \) is a leaf node then \( Q_n \equiv 0 \). For the first stage, we solve problem (6.57) with \( n = n_0 \) and \( m = m_1 \), with optimal value denoted by \( \Omega_{n_0}(x_0) = \Omega_{n_0}(x_0, \xi_{n_1}) \) where \( \xi_{n_1} = \xi_1 \).

Algorithm 3 solves these dynamic programming equations building at iteration \( k \) polyhedral lower approximation \( Q_n^k \) of \( Q_n \) where

\[
Q_n^k(x_n) = \max \left( \theta_n^\ell + \langle \beta_n^\ell, x_n - x_n^0 \rangle, \ell \in I_n, \ell \leq k \right)
\]

for all node \( n \in N \backslash \{ n_0 \} \) and for some set \( I_n \) containing \( \{0\} \) and all the iterations, up to the current iteration, where the sampled scenario passes through node \( n \).

To describe this algorithm, if node \( n \) is not a leaf, it is convenient to introduce the function \( \Omega_n^{k-1}(x_n, \xi_m) \) is given by

\[
(6.58) \quad \begin{cases}
\min_{x_m} F_{\tau_m}^{k-1}(x_n, x_m, \Psi_m) \\
g_{\tau_m}(x_0, x_n, x_m, \Psi_m) \leq 0,
\end{cases}
\]

with

\[
F_{\tau_m}^{k-1}(x_n, x_m, \Psi_m) = f_{\tau_m}(x_n, x_m, \Psi_m) + Q_n^{k-1}(x_n, x_m).
\]

Next, we write \( \Omega_n^{k-1}(x_n, \xi_m) \), i.e., (6.58), under the form

\[
(6.59) \quad \begin{cases}
\min_{x_m} f_{\tau_m}(x_n, x_m, \Psi_m) + z \\
g_{\tau_m}(x_0, x_n, x_m, \Psi_m) \leq 0,
\end{cases}
\]

with

\[
F_{\tau_m}^{k-1}(x_n, x_m, \Psi_m) = f_{\tau_m}(x_n, x_m, \Psi_m) + Q_n^{k-1}(x_n, x_m).
\]

In the above problem, we have denoted by \( \pi_{km1}, \pi_{km2}, \) and \( \pi_{km3} \) the optimal Lagrange multipliers associated with respectively the first, second, and third group of constraints. Finally, if \( n \) is a leaf, \( \Omega_n^{k-1} = 0 \).

For \( n = n_0 \), the optimal value of (6.58) is denoted by \( \Omega_{n_0}^{k-1}(x_0, \xi_1) \).

---

Algorithm 3: Multistage stochastic decomposition algorithm to solve for interstage dependent processes.

**Initialization.** Set \( Q_n^0 \equiv 0 \) for all leaf node \( n \) and \( Q_n^0 \equiv -\infty \) for all other node \( n \), i.e., \( \theta_n^0 = -\infty \) and \( \beta_n^0 = 0 \) for all \( n \in N \) which is not a leaf node. For all \( n \in N \), set \( I_n = \{0\} \).

For \( k = 1, 2, \ldots \),

Sample a scenario \( (\xi_1^k, \xi_2^k, \ldots, \xi_T^k) \) for \( (\xi_1, \xi_2, \ldots, \xi_T) \), 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 $t = 1, \ldots, T$,

Solve (6.58) with $n$ and $m$ respectively replaced by $n_{t-1}^k$ and $n_t^k$ (note that $\xi_m = \xi_{n_t^k} = \xi_t^k$) and add $k$ to $I_{n_t^k}$.

Let $x_{n_t^k}$ be an optimal solution.

If $t \geq 2$,

For every $m \in C(n_{t-1}^k)$ with $m \neq n_t^k$

Solve (6.58) for that node $m$ with $n$ replaced by $n_{t-1}^k$.

Let $x_m$ be an optimal solution.

End If

End For

Compute a subgradient $\pi_{km}$ of $\Omega_n^{k-1}(:, \xi_m)$ at $x_n^k$:

$$
\pi_{km} = f'_{r_m x_n^k} \left( x_n^k, x_m^k, \Psi_m \right) + g'_{r_m x_n^k} \left( x_n^k, x_m^k, \Psi_m \right) \pi_{km1} + \left[ \begin{array}{c} \vdots \\ A_{m1} \\ \vdots \\ A_{m,\tau_m-1} \end{array} \right] \pi_{km2} + [(\beta_{m1}^k)_{i \in I_m, i \leq k-1}] \pi_{km3},
$$

(6.60)

where $f'_{r_m x_n^k} \left( x_n^k, x_m^k, \Psi_m \right)$ is a subgradient of convex function $f_{r_m}(:, x_m^k, \Psi_m)$ at $x_n^k$ and the $i$-th column of matrix $g'_{r_m x_n^k} (x_n^k, x_m^k, \Psi_m)$ is a subgradient at $x_n^k$ of the $i$-th component of convex function $g'_{r_m x_n^k} (:, x_m^k, \Psi_m)$.

Setting $n = n_{t-1}^k$, update $\theta_n^k$ and $\beta_n^k$ computing

$$
\theta_n^k = \sum_{m \in C(n)} p_{km} \Phi_m \Omega_n^{k-1}(x_n^k, \xi_m)
$$

and

$$
\beta_n^k = \sum_{m \in C(n)} p_{km} \Phi_m \pi_{km}.
$$

where $p_{km}$ satisfies:

$$
\theta_n^k = \sup_{\nu \in \mathcal{P}_n} \sum_{m \in C(n)} p_{m \nu} \Phi_m \Omega_n^{k-1}(x_n^k, \xi_m)
$$

$$
= \sum_{m \in C(n)} p_{km} \Phi_m \Omega_n^{k-1}(x_n^k, \xi_m).
$$

End If

End For

$\theta_{n^k}^k = 0$ and $\beta_{n^k}^k = 0$.

End For

**Theorem 6.1** (Convergence analysis of Algorithm 3). Consider the sequence of random variables $(x_n^k)_{k \in \mathbb{N}}$, $n \in \mathcal{N}$, generated by Algorithm 3. Let Assumptions (H2) hold and assume that $(\xi_t)$ is a discrete random process with a finite set of possible realizations at each stage. Also assume that at each node $m$, the transition probability $\Phi_m$ to go from node $\mathcal{P}(m)$ to node $m$ is positive. Then, with probability one:
(i) there exist \( x_n^* \), \( n \in \mathcal{N} \), and infinite subsets \( K_n, n \in \mathcal{N} \) of integers, such that for all \( n \in \mathcal{N} \):
\[
\lim_{k \to +\infty, k \in K_n} x_n^k = x_n^* \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n.
\]

(ii) For any node \( n \in \mathcal{N} \setminus \{ n_0 \} \), we have
\[
\lim_{k \to +\infty, k \in K_n} \Omega_n^k(x_n^k) = \Omega_n(x_n^*).
\]

(iii) We have
\[
\lim_{k \to +\infty} \Omega_{n_0}^{k-1}(x_0, \xi_1) = \Omega_{n_0}(x_0),
\]
that is, 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, any accumulation point of the sequence \( (x_{n}^k)_{k \in \mathbb{N}} \) 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. Item (i) can be shown following the proof of Proposition 4.1 introducing the sets \( K_n, K_n', n \in \mathcal{N} \), of this proof.

We prove (ii) by backward induction on the number of stages. Following the proof of Proposition 3.1, we show that \( \Omega_n \) is continuous on \( \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \) for all \( n \in \mathcal{N} \setminus \{ n_0 \} \). Following the proof of Lemma 3.2, we show that for all \( n \in \mathcal{N} \setminus \{ n_0 \} \) and \( k \) sufficiently large, say \( k \geq T_0 \), \( \Omega_n^k \) is Lipschitz continuous and \( \pi_{k,m} \) given by (6.60) is bounded.

We also observe that for every stage \( t = 1, \ldots, T - 1 \), for every node \( n \in \text{Nodes}(t) \) and for every \( k, \ell \in I_n \) with \( k \geq \ell \), we have
\[
\theta_n^k = \sup_{p \in P_n} \sum_{m \in C(n)} p_m \Phi_n \Omega_n^{k-1}(x_n^k, \xi_m) \\
\geq \sup_{p \in P_n} \sum_{m \in C(n)} p_m \Phi_n \Omega_n^\ell \Omega_n^{\ell-1}(x_n^k, \xi_m) \\
\geq \sup_{p \in P_n} \sum_{m \in C(n)} p_m \Phi_n \left( \Omega_n^{\ell-1}(x_n^\ell, \xi_m) + \langle \pi_{\ell,m}, x_n^k - x_n^\ell \rangle \right) \\
\geq \sum_{m \in C(n)} \pi_{\ell,m} \left( \Omega_n^{\ell-1}(x_n^\ell, \xi_m) + \langle \pi_{\ell,m}, x_n^k - x_n^\ell \rangle \right) \\
\geq \theta_n^\ell + \langle \beta_n^\ell, x_n^k - x_n^\ell \rangle.
\]

It follows that for \( k \in I_n \),
\[
\Omega_n^k(x_n^k) = \max \left( \theta_n^k, \max \left( \theta_n^\ell + \langle \beta_n^\ell, x_n^k - x_n^\ell \rangle, \ell \in I_n, \ell < k \right) \right) = \theta_n^k.
\]

Finally, note that identity (6.61) also holds for all leaf \( n \). Let us now prove (ii) by induction. The induction hypothesis is that for each node \( m \) of stage \( t + 1 \),
\[
\Omega_m^k(x_m^k) = \Omega_m(x_m^k).
\]

For every leaf \( m \) of the tree, we have for every iteration \( k \),
\[
\Omega_m^k(x_m^k) = \Omega_m(x_m^k) = 0.
\]

Using item (i), the above relation, and the continuity of \( \Omega_m \), we have shown the induction hypothesis for every leaf, i.e., (6.62) for every node \( m \) of stage \( T \).
Now assume that the induction hypothesis is true for each node \( m \) of stage \( t + 1 \) for some \( t \in \{1, \ldots, T - 1\} \). We want to show that for each node \( n \) of stage \( t \),

\[
\lim_{k \to +\infty, k \in K_m} \Omega_n(x^k_{[n]}) - Q_n(x^k_{[n]}) = 0.
\]

Following the proof of Lemma 4.2, we deduce from the induction hypothesis \([6.62]\) that for each node \( m \) of stage \( t + 1 \)

\[
\lim_{k \to +\infty, k \in K_m} \Omega^{m-1}_n(x^k_{[m]}) = Q_m(x^*_{[m]}).
\]

We then have for \( k \in K_m \) and \( n = P(m) \),

\[
\Omega_n^{k-1}(x^k_{[n]}, \xi_{t+1}) = \Omega_n^{k-1}(x^k_{[n]}, \xi_m)
\]

\[
= F_{r_m}(x^k_{[m]}, \Psi_m) - Q_m(x^k_{[m]}) + \Omega^{k-1}_m(x^k_{[m]})
\]

\[
\geq \Omega_n(x^k_{[n]}, \xi_m) - Q_m(x^k_{[m]}) + \Omega^{k-1}_m(x^k_{[m]})
\]

where the last inequality comes from the relation

\[
\Omega_n(x^k_{[n]}, \xi_m) = \left\{ \inf_{x_m} F_{r_m}(x^k_{[n]}, x_m, \Psi_m) \right\} \leq F_{r_m}(x^k_{[m]}, \Psi_m)
\]

which holds since \( x^k_m \in X_{r_m}(x^k_{[n]}, \xi_m) \) for \( k \in K_m \). It follows that

\[
\lim_{k \to +\infty, k \in K_m} \Omega_n(x^k_{[n]}, \xi_m) - \Omega^{k-1}_n(x^k_{[n]}, \xi_m) = 0
\]

where \( n = P(m) \).

Let us now fix \( \varepsilon > 0 \) and \( n \in \text{Nodes}(t) \). From the continuity of \( \Omega_n(\cdot, \xi_m) \) and the fact that the sequence \( (x^k_{[n]})_{k \in K_n} \) converges to \( x^*_{[n]} \), there exists \( k_1 \) such that for \( k \in K_n \) and \( k \geq k_1 \), we have for all \( m \in C(n) \),

\[
|\Omega_n(x^k_{[n]}, \xi_m) - \Omega_n(x^*_{[n]}, \xi_m)| \leq \frac{\varepsilon}{10}.
\]

Denoting by \( 0 < M_1 < +\infty \) an upper bound on \( \|\pi_{k,m}\| \) (valid for all \( m \) and \( k \geq T_0 \)),

since the sequence \( (x^k_{[n]})_{k \in K_n} \) converges to \( x^*_{[n]} \), there exists \( k_2 \in K_n \) with \( k_2 \geq T_0 \) such that for \( k \geq k_2 \) and \( k \in K_n \)

\[
\|x^k_{[n]} - x^*_{[n]}\| \leq \frac{\varepsilon}{10M_1}.
\]

Using \([6.64]\), for every \( m \in C(n) \), there exists \( k(3, m) \in K_m \) with \( k(3, m) \geq \max(k_1, k_2) \geq T_0 \) such that for every \( k \in K_m \) with \( k \geq k(3, m) \), we have

\[
0 \leq \Omega_n(x^k_{[n]}, \xi_m) - \Omega^{k-1}_n(x^k_{[n]}, \xi_m) \leq \frac{\varepsilon}{10}.
\]

From the continuity of \( Q_n \) and the fact that the sequence \( (x^k_{[n]})_{k \in K_n} \) converges to \( x^*_{[n]} \), there exists \( k_4 \) such that for \( k \in K_n \) and \( k \geq k_4 \), we have

\[
|Q_n(x^k_{[n]}) - Q_n(x^*_{[n]})| \leq \frac{\varepsilon}{2}.
\]
Take now an arbitrary $k \in K_n$ with
\[ k \geq \max(T_0, k_1, k_2, \max(k(3, m), m \in C(n)), k_4) = \max(\max(k(3, m), m \in C(n)), k_4). \]
Since $k \geq k(3, m)$ for every $m \in C(n)$, we have
\begin{align*}
\Phi^k_n &= \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \Omega^{k-1}_n(x^k_{[n]}, \xi_m) \\
&\geq \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \Omega^{k(3,m)-1}_n(x^k_{[n]}, \xi_m) \\
&\geq \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \left[ \Omega^{k(3,m)-1}_n(x^k_{[n]}, \xi_m) + \langle \pi_{k(3,m)} m, x^k_{[n]} - x^{k(3,m)} \rangle \right]
\end{align*}
using the convexity of $\Omega^{k(3,m)-1}_n(\cdot, \xi_m)$ and the fact that $\pi_{k(3,m)} m$ is a subgradient of $\Omega^{k(3,m)-1}_n(\cdot, \xi_m)$ at $x^{k(3,m)}_{[n]}$ (recall that $k(3, m) \in K_m$). We obtain
\begin{align*}
0 \leq Q_n(x^k_{[n]}) - \Phi^k_n(x^k_{[n]}) &\leq \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \left[ \Omega_n\left(x^k_{[n]}, \xi_m\right) - \Omega^{k(3,m)-1}_n(x^k_{[n]}, \xi_m) - \langle \pi_{k(3,m)} m, x^k_{[n]} - x^{k(3,m)} \rangle \right] \\
&\leq \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \left[ \Omega_n(x^k_{[n]}, \xi_m) - \Omega_n\left(x^*_{[n]}, \xi_m\right) \right] \\
&\quad + \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \left[ \Omega_n(x^k_{[n]}, \xi_m) - \Omega^{k(3,m)-1}_n(x^k_{[n]}, \xi_m) \right] \\
&\quad + \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \left[ \Omega_n(x^*_{[n]}, \xi_m) - \Omega_n(x^*_{[n]}, \xi_m) \right] \\
&\quad + \sup_{p \in \mathcal{P}_n} \sum_{m \in C(n)} p_m \Phi_m \left[ \pi_{k(3,m)} m \left( \|x^k_{[n]} - x^*_{[n]}\| + \|x^{k(3,m)} - x^*_{[n]}\| \right) \right] \leq \frac{\varepsilon}{2}.
\end{align*}
The last inequality was obtained using relations (6.68), (6.69), (6.70), and the fact that $\sum_{m \in C(n)} p_m \Phi_m = 1$ for any $p \in \mathcal{P}_n$. Combining the above relation with (6.71), we have shown that for every $\varepsilon > 0$, for every $k \geq \max(k_1, k_2, \max(k(3, m), m \in C(n)), k_4)$ with $k \in K_n$, we have $|Q_n(x^k_{[n]}) - Q_n(x^*_{[n]})| \leq \varepsilon$. We have thus shown (6.66), which achieves the induction step and the proof of (ii).

Finally, the proof of (iii) is analogous to the proof of (ii) in Theorem 1.3

Acknowledgments

The author’s research was partially supported by an FGV grant, CNPq grant 307287/2013-0, and PRONEX-Optimization.

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] P. Girardeau, V. Leclere, and A.B. Philpott. On the convergence of decomposition methods for multistage stochastic convex programs. *Optimization OnLine*, 2013.

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

[8] 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.

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

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

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

[12] 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.

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

[14] 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.

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

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

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

Vincent Guigues: FGV/EMAP, Praia de Botafogo, Rio de Janeiro, Brazil, vguigues@fgv.br, UFRJ, Escola Politécnica, Departamento de Engenharia Industrial, Ilha do Fundão, CT, Bloco F, Rio de Janeiro, Brazil