STOCHASTIC DYNAMIC CUTTING PLANE FOR MULTISTAGE STOCHASTIC CONVEX PROGRAMS

Vincent Guigues
School of Applied Mathematics, FGV
Praia de Botafogo, Rio de Janeiro, Brazil
vincent.guigues@fgv.br

Renato D.C. Monteiro
Georgia Institute of Technology
Atlanta, Georgia 30332, USA,
renato.monteiro@isye.gatech

Abstract. We introduce StoDCuP (Stochastic Dynamic Cutting Plane), an extension of the Stochastic Dual Dynamic Programming (SDDP) algorithm to solve multistage stochastic convex optimization problems. At each iteration, the algorithm builds lower bounding affine functions not only for the cost-to-go functions, as SDDP does, but also for some or all nonlinear cost and constraint functions. We show the almost sure convergence of StoDCuP. We also introduce an inexact variant of StoDCuP where all subproblems are solved approximately (with bounded errors) and show the almost sure convergence of this variant for vanishing errors.

Keywords: Stochastic programming, Inexact cuts for value functions, SDDP, Inexact SDDP.

AMS subject classifications: 90C15, 90C90.

1. Introduction

Risk-neutral multistage stochastic programs (MSPs) aim at minimizing the expected value of the total cost over a given optimization period of T stages while satisfying almost surely for every stage some constraints depending on an underlying stochastic process. These optimization problems are useful for many real-life applications but are challenging to solve, see for instance [22] and references therein for a thorough discussion on MSPs. Popular solution methods for MSPs are based on decomposition techniques such as Approximate Dynamic Programming [16], Lagrangian relaxation, or Stochastic Dual Dynamic Programming (SDDP) [12]. Recently, several enhancements of SDDP have been proposed, see for instance [21], [8], [13] for risk-averse variants, [15], [2], [3] for convergence analysis, and [11], [5] to speed up the convergence of the method. In particular, in [5], Inexact SDDP was proposed, which incorporates inexact cuts in SDDP (for both linear and nonlinear programs). The idea of Inexact SDDP is to allow us to solve approximately some or all primal and dual subproblems in the forward and backward passes of SDDP. This extension and the study of Inexact SDDP was motivated by the following reasons:

(i) solving to a very high accuracy nonlinear programs can take a significant amount of time or may even be impossible whereas linear programs (of similar sizes) can be solved exactly or to high accuracy quicker. Therefore one has to study how to extend the SDDP algorithm to still derive valid cuts and a converging Inexact SDDP or an Inexact SDDP with controlled accuracy when only approximate primal and dual solutions are computed for nonlinear MSPs.
(ii) As explained in [5], numerical experiments (see for instance [4, 7, 10]) show that for both linear and nonlinear MSPs, for the first iterations and for the first stages, the cuts computed can be quite distant from the corresponding recourse function in the neighborhood of the trial point at which the cut was computed. Therefore, it makes sense, for both nonlinear and linear MSPs, to try and solve more quickly and less accurately (inexactly) all subproblems of the forward and backward passes corresponding to the first iterations and the first stages and to increase the precision of the computed solutions as the algorithm progresses. Using this strategy, it was shown in [5] that for several instances of a portfolio problem, Inexact SDDP can converge (i.e., satisfy the stopping criterion) quicker than SDDP.

In this paper, we extend [5] in two ways:

- a natural way of taking advantage of observation (i) above in the context of SDDP applied to nonlinear problems, consists in linearizing all nonlinear objective and constraint functions of the subproblems solved along the iterations of the method at the optimal solutions of these subproblems. However, to the best of our knowledge, this variant of SDDP, that we term as StoDCuP (Stochastic Dynamic Cutting Plane) has not been proposed and studied so far in the literature. In this context, the goal of this paper is to propose and study StoDCuP.
- As far as (ii) is concerned, it is interesting to notice that it is easy to incorporate inexact cuts in StoDCuP (i.e., to derive an inexact variant of StoDCuP), control the quality of these cuts (see Lemma 4.1), and show the convergence of this method (see Theorem 4.3 below). This comes from the fact that we can easily compute a cut for the value function of a linear program (and in StoDCuP all subproblems solved are linear programs) from any feasible primal-dual solution since the corresponding dual objective is linear, see Proposition 2.1 in [5]. On the contrary, deriving valid (inexact) cuts from approximate primal-dual solutions of the original problems solved in SDDP applied to nonlinear problems and showing the convergence of the corresponding variant of Inexact SDDP is technical and the computation of inexact cuts may require solving additional subproblems, see [5] for details.

The outline of the paper is the following. To ease the presentation and analysis of StoDCuP, we start in Section 2 with its deterministic counterpart, called DCuP (Dynamic Cutting Plane) which solves convex Dynamic Programming equations linearizing cost-to-go, constraint, and objective functions. Starting with the deterministic case allows us to focus on the differences between traditional Dual Dynamic Programming and its convergence analysis with DCuP and its convergence analysis. Two variants, a forward DCuP and a forward-backward DCuP, together with their convergence analysis, are presented. In Section 3, we introduce forward StoDCuP and prove the almost sure convergence of the method. Finally, in Section 4, we present two variants of StoDCuP: forward-backward StoDCuP and Inexact StoDCuP which builds inexact cuts on the basis of approximate primal-dual solutions of the subproblems solved along the iterations of the method. We also prove the almost sure convergence of Inexact StoDCuP for vanishing noises.

We will use the following notation:

- For a real-valued convex function f, we denote by $\ell_f(\cdot; x_0)$ an arbitrary lower bounding linearization of f at x_0, i.e., $\ell_f(\cdot; x_0) = f(x_0) + s_f(x_0)^\top(\cdot - x_0)$ where $s_f(x_0)$ is an arbitrary subgradient of f at x_0.
- The domain of a point to set operator $T : A \rightrightarrows B$ is given by Dom(T)= $\{a \in A : T(a) \neq \emptyset\}$.
- For vectors $x, y \in \mathbb{R}^n$, $\langle x, y\rangle = x^\top y$ is the usual scalar product between x and y.
- For $a \in \mathbb{R}^n$, $\bar{B}(a; \varepsilon) = \{x \in \mathbb{R}^n : \|x - a\|_2 \leq \varepsilon\}$.

2. The DCuP (Dynamic Cutting Plane) algorithm

2.1. Problem formulation and assumptions. Given $x_0 \in \mathbb{R}^n$, consider the optimization problem

$$
\left\{
\begin{array}{l}
\displaystyle\inf_{x_1,\ldots,x_T \in \mathbb{R}^n} \sum_{t=1}^{T} f_t(x_t, x_{t-1})\\
g_t(x_t, x_{t-1}) \leq 0, \;\; A_t x_t + B_t x_{t-1} = b_t, \; t = 1, \ldots, T,\\
x_t \in \mathcal{X}_t, \; t = 1, \ldots, T,
\end{array}
\right.
\tag{2.1}
$$

where A_t and B_t are matrices of appropriate dimensions, $f_t : \mathbb{R}^n \times \mathbb{R}^n \to (-\infty, \infty]$ and $g_t : \mathbb{R}^n \times \mathbb{R}^n \to (-\infty, \infty]^p$. In this problem, for each step t, we have nonlinear and linear coupling constraints, $g_t(x_t, x_{t-1}) \leq 0$ and $A_t x_t + B_t x_{t-1} = b_t$ respectively, and set constraints $x_t \in \mathcal{X}_t$. Without loss of generality, nonlinear noncoupling constraints $h_t(x_t) \leq 0$ can be dealt with by incorporating them into the constraint $g_t(x_t, x_{t-1}) \leq 0$. For convenience, we use the short notation

$$
X_t(x_{t-1}) := \{x_t \in \mathcal{X}_t : \; g_t(x_t, x_{t-1}) \leq 0, \;\; A_t x_t + B_t x_{t-1} = b_t\}
\tag{2.2}
$$

and

$$
X_t^0(x_{t-1}) = X_t(x_{t-1}) \cap \operatorname{ri} \mathcal{X}_t.
\tag{2.3}
$$

With this notation, the dynamic programming equations corresponding to problem (2.1) are

$$\mathcal{Q}_t(x_{t-1}) = \begin{cases} \inf_{x_t \in \mathbb{R}^n} \; F_t(x_t, x_{t-1}) := f_t(x_t, x_{t-1}) + \mathcal{Q}_{t+1}(x_t) \\ x_t \in X_t(x_{t-1}), \end{cases} \tag{2.4}$$

for $t = 1, \ldots, T$, and $\mathcal{Q}_{T+1} \equiv 0$. Cost-to-go function $\mathcal{Q}_{t+1}(x_t)$ represents the optimal total cost for time steps $t+1, \ldots, T$, starting from state x_t at the beginning of step $t+1$. Clearly, it follows from the above definition that

$$\mathrm{Dom}(X_t^0) \subset \mathrm{Dom}(X_t) \quad \forall t = 1, \ldots, T. \tag{2.5}$$

Setting $\mathcal{X}_0 = \{x_0\}$, the following assumptions are made throughout this section.

Assumption (H1):

1) For $t = 1, \ldots, T$:
   a) $\mathcal{X}_t \subset \mathbb{R}^n$ is nonempty, convex, and compact;
   b) f_t is a proper lower-semicontinuous convex function such that $\mathcal{X}_t \times \mathcal{X}_{t-1} \subset \mathrm{int}\,(\mathrm{dom}(f_t))$;
   c) each of the p components g_{ti}, $i = 1, \ldots, p$, of g_t is a proper lower-semicontinuous convex function such that $\mathcal{X}_t \times \mathcal{X}_{t-1} \subset \mathrm{int}\,(\mathrm{dom}(g_{ti}))$.
2) $X_1(x_0) \neq \emptyset$ and $\mathcal{X}_{t-1} \subset \mathrm{int}\,\left[\mathrm{Dom}(X_t^0)\right]$ for every $t = 2, \ldots, T$.

The following simple lemma states a few consequences of the above assumption.

Lemma 2.1. *The following statements hold:*

(a) *for every $t = 1, \ldots, T$, $\mathcal{Q}_{t+1}$ is a convex function such that*

$$\mathcal{X}_t \subset \mathrm{int}\; (\mathrm{dom}(\mathcal{Q}_{t+1}))\,;$$

(b) *for every $t = 1, \ldots, T$, $\mathcal{Q}_{t+1}$ is Lipschitz continuous on $\mathcal{X}_t$;*

(c) *for every $t = 1, \ldots, T$, $i = 1, \ldots, p$, and $(x_t, x_{t-1}) \in \mathcal{X}_t \times \mathcal{X}_{t-1}$,*

$$\partial f_t(x_t, x_{t-1}) \neq \emptyset, \quad \partial g_{ti}(x_t, x_{t-1}) \neq \emptyset;$$

(d) *for every $t = 1, \ldots, T$, $i = 1, \ldots, p$, the sets*

$$\cup \{\partial f_t(x_t, x_{t-1}) : (x_t, x_{t-1}) \in \mathcal{X}_t \times \mathcal{X}_{t-1}\}\,, \quad \cup \{\partial g_{ti}(x_t, x_{t-1}) : (x_t, x_{t-1}) \in \mathcal{X}_t \times \mathcal{X}_{t-1}\}$$

are bounded.

Proof: (a) The proof is by backward induction on t. The result clearly holds for $t = T$ since $\mathcal{Q}_{T+1} \equiv 0$. Assume now that $\mathcal{Q}_{t+1}$ is a convex function such that $\mathcal{X}_t \subset \mathrm{int}\; (\mathrm{dom}(\mathcal{Q}_{t+1}))$ for some $2 \le t \le T$. Then, condition 1) of Assumption (H1) implies that the function $(x_t, x_{t-1}) \mapsto F_t(x_t, x_{t-1}) + \delta_{X_t(x_{t-1})}(x_t)$ is convex. This conclusion together with the definition of $\mathcal{Q}_t$ and the discussion following Theorem 5.7 of [17] then imply that $\mathcal{Q}_t$ is a convex function. Moreover, conditions 1)b) and 2) of Assumption (H1) and relation (2.5) imply that there exists $\varepsilon > 0$ such that for every $x_{t-1} \in \mathcal{X}_{t-1} + \bar{B}(0, \varepsilon)$,

$$\mathrm{dom}(f_t(\cdot, x_{t-1})) \supset \mathcal{X}_t, \quad X_t(x_{t-1}) \neq \emptyset.$$

The induction hypothesis, the latter observation, and relations (2.2) and (2.4), then imply that

$$X_t(x_{t-1}) \cap \mathrm{dom}(F_t(\cdot, x_{t-1})) = X_t(x_{t-1}) \cap \mathrm{dom}(f_t(\cdot, x_{t-1})) \cap \mathrm{dom}(\mathcal{Q}_{t+1}) \supset X_t(x_{t-1}) \cap \mathcal{X}_t = X_t(x_{t-1}) \neq \emptyset$$

for every $x_{t-1} \in \mathcal{X}_{t-1} + \bar{B}(0, \varepsilon)$. Since by (2.4),

$$\mathrm{dom}(\mathcal{Q}_t) = \{x_{t-1} \in \mathbb{R}^n : X_t(x_{t-1}) \cap \mathrm{dom}(F_t(\cdot, x_{t-1})) \neq \emptyset\},$$

we then conclude that $\mathcal{X}_{t-1} + \bar{B}(0, \varepsilon) \subset \mathrm{dom}(\mathcal{Q}_t)$, and hence that $\mathcal{X}_{t-1} \subset \mathrm{int}\,(\mathrm{dom}(\mathcal{Q}_t))$. We have thus proved that (a) holds.

b) This statement follows from statement a) and Theorem 10.4 of [17].

c-d) These two statements follow from conditions 1)a), 1)b) and 1)c) of Assumption (H1) together with Theorem 23.4 and 24.7 of [17]. ∎

2.2. **Forward DCuP.** Before formally describing DCuP algorithm, we give some motivation for it. At iteration $k \geq 1$ and stage $t = 1, \ldots, T$, the algorithm uses the following approximation to function $\mathcal{Q}_t(\cdot)$ defined in (2.4):

$$\underline{\mathcal{Q}}_t^{k-1}(x_{t-1}) = \min\left\{ f_t^{k-1}(x_t, x_{t-1}) + \mathcal{Q}_{t+1}^{k-1}(x_t) : x_t \in X_t^{k-1}(x_{t-1}) \right\} \tag{2.6}$$

where

$$X_t^{k-1}(x_{t-1}) = \{ x_t \in \mathcal{X}_t : \; g_t^{k-1}(x_t, x_{t-1}) \leq 0, \; A_t x_t + B_t x_{t-1} = b_t \} \tag{2.7}$$

and f_t^{k-1}, g_t^{k-1}, and $\mathcal{Q}_{t+1}^{k-1}$ are polyhedral functions minorizing f_t, g_t and $\mathcal{Q}_{t+1}$, respectively, i.e.,

$$f_t^{k-1} \leq f_t, \quad g_t^{k-1} \leq g_t, \quad \mathcal{Q}_{t+1}^{k-1} \leq \mathcal{Q}_{t+1}. \tag{2.8}$$

For $t = T+1$, we actually assume that $\mathcal{Q}_{T+1}^{k-1} \equiv 0$, and hence that $\mathcal{Q}_{T+1}^{k} = \mathcal{Q}_{T+1}$. Moreover, we also assume that $\underline{\mathcal{Q}}_{T+1}^{k-1} \equiv 0$, and hence $\underline{\mathcal{Q}}_{T+1}^{k-1} = \mathcal{Q}_{T+1}$.

Observe that for every $k \geq 0$, $t = 1, \cdots, T$, and $x_{t-1} \in \mathcal{X}_{t-1}$, relations (2.7) and (2.8) imply that

$$X_t(x_{t-1}) \subset X_t^k(x_{t-1}) \subset \mathcal{X}_t \tag{2.9}$$

and

$$f_t^k(\cdot, x_{t-1}) + \mathcal{Q}_{t+1}^k(\cdot) \leq f_t(\cdot, x_{t-1}) + \mathcal{Q}_{t+1}(\cdot),$$

and hence that

$$\underline{\mathcal{Q}}_t^k \leq \mathcal{Q}_t, \quad \forall\, t = 1, 2, \ldots, T, \; \forall\, k \geq 0. \tag{2.10}$$

At iteration k, feasible points $x_1^k, \ldots, x_T^k$ are computed recursively as follows: for $t = 1, \ldots, T$, x_t^k is set to be an optimal solution of subproblem (2.6) with $x_{t-1} = x_{t-1}^k$ with the convention that $x_0^k = x_0$. These points in turn are used to compute new affine functions minorizing f_t, g_t and $\mathcal{Q}_{t+1}$ which are then added to the bundle of affine functions describing f_t^{k-1}, g_t^{k-1}, and $\mathcal{Q}_{t+1}^{k-1}$ to obtain new lower bounding approximations f_t^k, g_t^k, and $\mathcal{Q}_{t+1}^k$ for f_t, g_t and $\mathcal{Q}_{t+1}$, respectively.

The precise description of DCuP algorithm is as follows.

DCuP (Dynamic Cutting Plane) with linearizations computed in a forward pass.

Step 0. Initialization. For every $t = 1, \ldots, T$, let affine functions f_t^0 and g_t^0 such that $f_t^0 \leq f_t$ and $g_t^0 \leq g_t$, and a piecewise linear function $\mathcal{Q}_t^0 : \mathcal{X}_{t-1} \to \mathbb{R}$ such that $\mathcal{Q}_t^0 \leq \mathcal{Q}_t$ be given. Set $\mathcal{Q}_{T+1}^0 \equiv 0$ and $k = 0$.

Step 1. Forward pass. Set $\mathcal{C}_{T+1}^k = \mathcal{Q}_{T+1}^k \equiv 0$ and $x_0^k = x_0$. For $t = 1, 2, \ldots, T$, do:

a) find an optimal solution x_t^k of

$$\underline{\mathcal{Q}}_t^{k-1}(x_{t-1}^k) = \left\{ \begin{array}{l} \inf\limits_{x_t \in \mathbb{R}^n} f_t^{k-1}(x_t, x_{t-1}^k) + \mathcal{Q}_{t+1}^{k-1}(x_t) \\ x_t \in X_t^{k-1}(x_{t-1}^k), \end{array} \right. \tag{2.11}$$

where $X_t^k(\cdot)$ is as in (2.7);

b) compute function values and subgradients of f_t and g_{ti}, $i = 1, \ldots, p$, at (x_t^k, x_{t-1}^k), and let $\ell_{f_t}(\cdot; (x_t^k, x_{t-1}^k))$ and $\ell_{g_{ti}}(\cdot; (x_t^k, x_{t-1}^k))$ denote the corresponding linearizations;

c) set

$$f_t^k = \max\left(f_t^{k-1}\,,\, \ell_{f_t}\left((\cdot,\cdot); (x_t^k, x_{t-1}^k)\right) \right), \tag{2.12}$$

$$g_{ti}^k = \max\left(g_{ti}^{k-1}\,,\, \ell_{g_{ti}}\left((\cdot,\cdot); (x_t^k, x_{t-1}^k)\right) \right), \quad \forall i = 1, \ldots, p, \tag{2.13}$$

and define $g_t^k := (g_{t1}^k, \ldots, g_{tp}^k)$;

d) if $t \geq 2$, then compute $\beta_t^k \in \partial \underline{\mathcal{Q}}_t^{k-1}(x_{t-1}^k)$ and denote the corresponding linearization of $\underline{\mathcal{Q}}_t^{k-1}$ as

$$\mathcal{C}_t^k(\cdot) := \underline{\mathcal{Q}}_t^{k-1}(x_{t-1}^k) + \langle \beta_t^k, \cdot - x_{t-1}^k \rangle;$$

moreover, set

$$\mathcal{Q}_t^k = \max\{\mathcal{Q}_t^{k-1}, \mathcal{C}_t^k\}; \tag{2.14}$$

Step 2. Set $k \leftarrow k+1$ and go to Step 1.

We now make a few remarks about DCuP. First, Lemma 2.1(c) guarantees the existence of the subgradients, and hence the linearizations, of the functions f_t and g_{ti}, $i = 1, \ldots, p$, at any point $(x_t, x_{t-1}) \in \mathcal{X}_t \times \mathcal{X}_{t-1}$, and hence that the functions f_t^k and g_t^k in Step 1 are well-defined. Second, in view of the definition of x_t^k in Step a), we have that $x_t^k \in X_t^{k-1}(x_{t-1}^k) \subset \mathcal{X}_t$ for every $t = 1, \ldots, T$ and $k \geq 0$. Third, Lemma 2.2(b) below and the previous remark guarantee the existence of the subgradient β_t^k in Step d). Fourth, we dicuss in Subsection 2.4 ways of computing this subgradient.

In the remaining part of this subsection, we provide the convergence analysis of DCuP. The following result states some basic properties about the functions involved in DCuP.

Lemma 2.2. *The following statements hold:*

(a) *for every $k \geq 1$ and $t = 1, \ldots, T$, we have*

$$f_t^k \leq f_t^{k+1} \leq f_t, \quad g_t^k \leq g_t^{k+1} \leq g_t, \tag{2.15}$$

$$X_t(x_{t-1}) \subset X_t^{k+1}(x_{t-1}) \subset X_t^k(x_{t-1}) \subset \mathcal{X}_{t-1} \quad \forall x_{t-1} \in \mathbb{R}^n, \tag{2.16}$$

$$\mathcal{Q}_{t+1}^k \leq \mathcal{Q}_{t+1}^{k+1} \leq \mathcal{Q}_{t+1}, \tag{2.17}$$

$$\underline{\mathcal{Q}}_t^k \leq \underline{\mathcal{Q}}_t^{k+1} \leq \mathcal{Q}_t. \tag{2.18}$$

(b) *For every $k \geq 1$ and $t = 2, \ldots, T$, function $\underline{\mathcal{Q}}_t^k$ is convex and* $\operatorname{int}(\operatorname{dom}(\underline{\mathcal{Q}}_t^k)) \supset \mathcal{X}_{t-1}$*; as a consequence, $\partial \underline{\mathcal{Q}}_t^k(x_{t-1}) \neq \emptyset$ for every $x_{t-1} \in \mathcal{X}_{t-1}$.*

Proof: (a) Relations (2.15) and (2.16) follow immediately from the initialization of DCuP described in step 0, the recursive definitions of f_t^k and g_t^k in (2.12) and (2.13), respectively, the definition of $X_t^k(\cdot)$ in (2.7), and the fact that

$$\ell_{f_t}((\cdot,\cdot);(x_t^k, x_{t-1}^k)) \leq f_t(\cdot,\cdot), \qquad \ell_{g_{ti}}((\cdot\cdot);(x_t^k, x_{t-1}^k)) \leq g_{ti}(\cdot\cdot).$$

Next note that the inequalities in (2.18) follow immediately from the respective ones in (2.15), (2.16) and (2.17), together with relations (2.4) and (2.11). It then remains to show that the inequalities in (2.17) hold. Indeed, the inequality $\mathcal{Q}_{t+1}^k \leq \mathcal{Q}_{t+1}^{k+1}$ follows immediately from (2.14) with $t = t+1$. We will now show that inequalities $\mathcal{Q}_t^k \leq \mathcal{Q}_t$ for every $t = 2, \ldots, T+1$ implies that $\mathcal{Q}_t^{k+1} \leq \mathcal{Q}_t$ for every $t = 2, \ldots, T+1$, and hence that the second inequality in (2.17) follows from a simple inductive argument on k. Indeed, first observe that the inequality $\mathcal{Q}_{t+1}^k \leq \mathcal{Q}_{t+1}$ implies that $\underline{\mathcal{Q}}_t^k \leq \mathcal{Q}_t$. Next observe that the construction of $\mathcal{C}_t^{k+1}$ in Step d) of DCuP implies that $\mathcal{C}_t^{k+1} \leq \underline{\mathcal{Q}}_t^k$, and hence that $\mathcal{C}_t^{k+1} \leq \mathcal{Q}_t$. It then follows from (2.14) and the inequality $\mathcal{Q}_t^k \leq \mathcal{Q}_t$ that $\mathcal{Q}_t^{k+1} \leq \mathcal{Q}_t$. We have thus shown that $\mathcal{Q}_t^k \leq \mathcal{Q}_t$ for every $t = 2, \ldots, T+1$ implies that $\mathcal{Q}_t^{k+1} \leq \mathcal{Q}_t$ for every $t = 2, \ldots, T$. Since the latter inequality for $t = T+1$ is straightforward and $\mathcal{Q}_t^0 \leq \mathcal{Q}_t$ for $t = 2, \ldots, T$, (2.17) follows.

(b) The assertion that $\underline{\mathcal{Q}}_t^k$ is a convex function follows from the fact that $\mathcal{Q}_{t+1}^k$ is convex and the same arguments used in Lemma 2.1 to show that $\mathcal{Q}_t$ is convex. The assertion that $\operatorname{dom}(\underline{\mathcal{Q}}_t^k) \supset \mathcal{X}_{t-1}$ follows from the fact that by (2.18) we have $\underline{\mathcal{Q}}_t^k \leq \mathcal{Q}_t$, and hence that

$$\operatorname{int}\left(\operatorname{dom}(\underline{\mathcal{Q}}_t^k)\right) \supset \operatorname{int}\left(\operatorname{dom}(\mathcal{Q}_t)\right) \supset \mathcal{X}_{t-1},$$

where the last inclusion is due to Lemma 2.1(a). ∎

The following technical result is useful to establish uniform Lipschitz continuity of convex functions.

Lemma 2.3. *Assume that ϕ^- and ϕ^+ are proper convex functions such that $\phi^- \leq \phi^+$. Then, for any nonempty compact set $K \subset \operatorname{int}(\operatorname{dom}(\phi^+))$, there exists a scalar $L = L(K) \geq 0$ satisfying the following property: any convex function ϕ such that $\phi^- \leq \phi \leq \phi^+$ is L-Lipschitz continuous on K.*

Proof: Let ϕ be a convex function such that $\phi^- \leq \phi \leq \phi^+$ and let $K \subset \operatorname{int}(\operatorname{dom}(\phi^+))$ be a nonempty compact set. Since ϕ^- and ϕ^+ are proper, it then follows that ϕ is proper and $\operatorname{dom}(\phi) \supset \operatorname{dom}(\phi^+)$, and hence that $\text{int}\,(\text{dom}(\phi^-)) \supset \text{int}\,(\text{dom}(\phi)) \supset \text{int}\,(\text{dom}(\phi^+)) \supset K$. Hence, in view of Theorem 23.4 of [17], we conclude that $\partial\phi(x) \neq \emptyset$ for every $x \in K$. We now claim that there exists L such that $\|\beta\| \leq L$ for every $\beta \in \partial\phi(x)$ and $x \in K$. This claim in turn can be easily seen to imply that the conclusion of the lemma holds. To show the claim, let $x \in K$ and $0 \neq \beta \in \partial\phi(x)$ be given. The inclusion $K \subset \text{int}\,(\text{dom}(\phi^+))$ implies the existence of $\varepsilon > 0$ such that $K_\varepsilon := K + \bar{B}(0;\varepsilon) \subset \text{int}\,(\text{dom}(\phi^+))$. Let

$$y_\varepsilon := x + \varepsilon\frac{\beta}{\|\beta\|}, \quad \theta^+ := \max_{y\in K_\varepsilon} \phi^+(y), \quad \theta^- := \min_{y\in K} \phi^-(y).$$

Clearly, $y_\varepsilon \in K_\varepsilon$ due to the definition of K_ε and the facts that $x \in K$ and $\|y_\varepsilon - x\| \leq \varepsilon$. Moreover, using the fact that every proper convex function is continuous in the interior of its domain, we then conclude that the proper convex functions ϕ^+ and ϕ^- are continuous on K_ε and K, respectively, since these two sets lie in the interior of their domains, respectively. Hence, it follows from Weierstrass' theorem that θ^+ and θ^- are both finite due to the compactness of K and K_ε, respectively. Using the facts that $x \in K$, $y_\varepsilon \in K_\varepsilon$, $\beta \in \partial\phi(x)$ and $\phi^+ \geq \phi$, the definitions of θ^+ and θ^-, and the definition of subgradient, it then follows that

$$\theta^+ \geq \phi^+(y_\varepsilon) \geq \phi(y_\varepsilon) \geq \phi(x) + \langle\beta, y_\varepsilon - x\rangle = \phi(x) + \varepsilon\|\beta\| \geq \theta^- + \varepsilon\|\beta\|$$

and hence that the claim holds with $L = (\theta^+ - \theta^-)/\varepsilon$. ∎

Lemma 2.4. *The following statements hold:*

(a) *For each $t = 2,\ldots,T$, there exist $L_t \geq 0$ such that the functions $\mathcal{Q}_t^k$ and $\underline{\mathcal{Q}}_t^k$ are L_t-Lipschitz continuous on $\mathcal{X}_{t-1}$ for every $k \geq 1$;*

(b) *For each $t = 1,\ldots,T$, there exist $\hat{L}_t \geq 0$ such that the functions f_t^k and g_{ti}^k are $\hat{L}_t$-Lipschitz continuous functions on $\mathcal{X}_t \times \mathcal{X}_{t-1}$ for every $k \geq 1$ and $i = 1,\ldots,p$.*

Proof: Let $t \in \{2,\ldots,T\}$ be given. The existence of L_t satisfying (a) follows from Lemmas 2.1 and 2.2, and applying Lemma 2.3 twice, the first time with $K = \mathcal{X}_{t-1}$, $\phi^+ = \mathcal{Q}_t$ and $\phi^- = \mathcal{Q}_t^0$, and the second time with $K = \mathcal{X}_{t-1}$, $\phi^+ = \mathcal{Q}_t$ and $\phi^- = \underline{\mathcal{Q}}_t^0$. Moreover, the existence of $\hat{L}_t$ satisfying (b) follows from Lemma 2.2, and applying Lemma 2.3 twice, the first time with $K = \mathcal{X}_t \times \mathcal{X}_{t-1}$, $\phi^+ = f_t$ and $\phi^- = f_t^0$, and the second time with $K = \mathcal{X}_t \times \mathcal{X}_{t-1}$, $\phi^+ = g_{ti}$ and $\phi^- = g_{ti}^0$ for $i = 1,\ldots,p$. ∎

We now state a result whose proof is given in Lemma 5.2 of [2]. Even though the latter result assumes convexity of the functions involved in its statement, its proof does not make use of this assumption. For this reason, we state the result here in a slightly more general way than it is stated in Lemma 5.2 of [2].

Lemma 2.5. *Lemma 5.2 in* [2]. *Assume that $Y \subset \mathbb{R}^n$ is a compact set, $f : \mathbb{R}^n \to (-\infty,\infty]$ is a function and $\{f_k : \mathbb{R}^n \to (\infty,\infty]\}_{k=1}^\infty$ is a sequence of functions such that, for some integer $k_0 > 0$ and scalar $L > 0$, we have:*

(a) *$f^{k-k_0}(y) \leq f^k(y) \leq f(y) < \infty$ for every $k \geq k_0 + 1$ and $y \in Y$;*

(b) *f^k is L-Lipschitz continuous on Y for every $k \geq 1$.*

Then, for any infinite sequence $\{y^k\}_{k=1}^\infty \subset Y$, we have

$$\lim_{k\to+\infty} [f(y^k) - f^k(y^k)] = 0 \Longleftrightarrow \lim_{k\to+\infty} [f(y^k) - f^{k-k_0}(y^k)] = 0.$$

We are now ready to provide the main result of this subsection, i.e., the convergence analysis of DCuP.

Theorem 2.6. *Let Assumption (H1) hold. Define*

$$\mathcal{H}(t) \begin{cases} (i) & \overline{\lim}_{k\to+\infty} g_{ti}(x_t^k, x_{t-1}^k) \leq 0, \quad i = 1,\ldots,p, \\ (ii) & \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^k) - \underline{\mathcal{Q}}_t^{k-1}(x_{t-1}^k) = \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^k) - \underline{\mathcal{Q}}_t^k(x_{t-1}^k) = 0, \\ (iii) & \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^k) - \sum_{\tau=t}^T f_\tau(x_\tau^k, x_{\tau-1}^k) = 0, \\ (iv) & \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^k) - \mathcal{Q}_t^k(x_{t-1}^k) = 0. \end{cases}$$

Then $\mathcal{H}(t)$-(i) holds for $t = 1,\ldots,T$, $\mathcal{H}(t)$-(ii),(iii) hold for $t = 1,\ldots,T+1$, and $\mathcal{H}(t)$-(iv) holds for $t = 2,\ldots,T+1$. In particular, the limit of the sequence of upper bounds $(\sum_{t=1}^T f_t(x_t^k, x_{t-1}^k))_{k\geq1}$ and of lower bounds $\underline{\mathcal{Q}}_1^{k-1}(x_0)$ is the optimal value $\mathcal{Q}_1(x_0)$ of (2.1) and any accumulation point of the sequence $(x_1^k,\ldots,x_T^k)$ is an optimal solution to (2.1).

Proof: We first prove $\mathcal{H}(t)$-(i) for $t = 1, \ldots, T$. Let $t \in \{1, \ldots, T\}$ be given and define the sequence $\{y_t^k\}$ as $y_t^k = (x_t^k, x_{t-1}^k)$ for every $k \geq 1$. In view of Lemma 2.2, we have $g_{ti}(y_t^k) \geq g_{ti}^k(y_t^k) \geq \ell_{g_{ti}}(y_t^k; y_t^k) = g_{ti}(y_t^k)$, and hence

$$g_t^k(y_t^k) = g_t(y_t^k), \ \forall\, k \geq 1. \tag{2.19}$$

Due to Lemma 2.4-(b), functions g_{ti}^k are convex $\hat{L}_t$-Lipschitz continuous on $\mathcal{X}_t \times \mathcal{X}_{t-1}$. Therefore, recalling (2.19), we can apply Lemma 2.5 to $f = g_{ti}$, $f^k = g_{ti}^k$, $y^k = y_t^k$, $Y = \mathcal{X}_t \times \mathcal{X}_{t-1}$ for $i = 1, \ldots, p$, to obtain

$$\lim_{k \to +\infty} g_t(x_t^k, x_{t-1}^k) - g_t^{k-1}(x_t^k, x_{t-1}^k) = 0. \tag{2.20}$$

The latter conclusion together with the fact that $x_t^k \in X_t^{k-1}(x_{t-1}^k)$, and hence $g_t^{k-1}(x_t^k, x_{t-1}^k) \leq 0$, for every $k \geq 1$, then implies that $\mathcal{H}(t)$-(i) holds.

Let us now show $\mathcal{H}(1)$-(ii), (iii) and $\mathcal{H}(t)$-(ii)-(iii), (iv) for $t = 2, \ldots, T+1$ by backward induction on t. $\mathcal{H}(T+1)$-(ii), (iii), (iv) trivially holds. Now, fix $t \in \{1, \ldots, T\}$ and assume that $\mathcal{H}(t+1)$-(ii), (iii), (iv) holds. We will show that $\mathcal{H}(t)$-(ii), (iii) holds and that $\mathcal{H}(t)$-(iv) holds if $t \geq 2$. Indeed, since $f_t \geq f_t^k \geq \ell_{f_t}(\cdot; y_t^k)$ and $f_t(y_t^k) = \ell_{f_t}(y_t^k; y_t^k)$, we conclude that $f_t^k(y_t^k) = f_t(y_t^k)$ for every $k \geq 1$, and hence that $\lim_{k \to +\infty} f_t(y_t^k) - f_t^k(y_t^k) = 0$. Recalling by Lemma 2.4-(b) that f_t^k is $\hat{L}_t$-Lipschitz continuous on $\mathcal{X}_t \times \mathcal{X}_{t-1}$ and using Lemma 2.5 with $f = f_t$, $f^k = f_t^k$, $(y^k) = (y_t^k)$, and $Y = \mathcal{X}_t \times \mathcal{X}_{t-1}$, we conclude that

$$\lim_{k \to +\infty} f_t(x_t^k, x_{t-1}^k) - f_t^{k-1}(x_t^k, x_{t-1}^k) = 0. \tag{2.21}$$

Moreover, by the induction hypothesis $\mathcal{H}(t+1)$-(iv), we have $\lim_{k \to +\infty} \mathcal{Q}_{t+1}^k(x_t^k) - \mathcal{Q}_{t+1}(x_t^k) = 0$. Recalling by Lemma 2.4-(a) that functions $\mathcal{Q}_t^k$ are L_t-Lipschitz continuous on $\mathcal{X}_{t-1}$, we can use Lemma 2.5 with $k_0 = 1$, $f = \mathcal{Q}_{t+1}$, $f^k = \mathcal{Q}_{t+1}^k$, $y^k = x_t^k$ and $Y = \mathcal{X}_t$, to obtain

$$\lim_{k \to +\infty} \mathcal{Q}_{t+1}^{k-1}(x_t^k) - \mathcal{Q}_{t+1}(x_t^k) = 0. \tag{2.22}$$

Now, using Lemma 2.2, we easily see that the objective function $f_t^{k-1}(\cdot, x_{t-1}^k) + \mathcal{Q}_{t+1}^{k-1}(\cdot)$ and feasible region $X_t^{k-1}(x_{t-1}^k)$ of (2.11) satisfies $f_t^{k-1}(\cdot, x_{t-1}^k) + \mathcal{Q}_{t+1}^{k-1}(\cdot) \leq F_t(\cdot, x_{t-1}^k)$ and $X_t^{k-1}(x_{t-1}^k) \supseteq X_t(x_{t-1}^k)$. Since x_t^k is an optimal solution of (2.11) and $\mathcal{Q}_t(x_{t-1}^k)$ is the optimal value of $\min\{F_t(x_t, x_{t-1}^k) : x_t \in X_t(x_{t-1}^k)\}$ due to (2.4), we then conclude that $f_t^{k-1}(x_t^k, x_{t-1}^k) + \mathcal{Q}_{t+1}^{k-1}(x_t^k) \leq \mathcal{Q}_t(x_{t-1}^k)$. Hence, we conclude that

$$0 \geq \overline{\lim_{k \to +\infty}}\, f_t^{k-1}(x_t^k, x_{t-1}^k) + \mathcal{Q}_{t+1}^{k-1}(x_t^k) - \mathcal{Q}_t(x_{t-1}^k) = \overline{\lim_{k \to +\infty}}\, f_t(x_t^k, x_{t-1}^k) + \mathcal{Q}_{t+1}(x_t^k) - \mathcal{Q}_t(x_{t-1}^k)$$

where the equality is due to (2.21) and (2.22). We now claim that

$$\lim_{k \to +\infty} f_t(x_t^k, x_{t-1}^k) + \mathcal{Q}_{t+1}(x_t^k) - \mathcal{Q}_t(x_{t-1}^k) = 0. \tag{2.23}$$

Indeed, assume by contradiction that the above claim does not hold. Then, it follows from the last conclusion before the claim that

$$\underline{\lim}_{k \to +\infty}\ f_t(x_t^k, x_{t-1}^k) + \mathcal{Q}_{t+1}(x_t^k) - \mathcal{Q}_t(x_{t-1}^k) < 0. \tag{2.24}$$

Since $\{(x_t^k, x_{t-1}^k)\}$ is a sequence lying in the compact set $\mathcal{X}_t \times \mathcal{X}_{t-1}$, it has a subsequence $\{(x_t^k, x_{t-1}^k)\}_{k \in K}$ converging to some $(x_t^*, x_{t-1}^*) \in \mathcal{X}_t \times \mathcal{X}_{t-1}$. Hence, in view of $\mathcal{H}(t)$-(i), (2.24), and the fact that f_t and g_t are lower semi-continuous on $\mathcal{X}_t \times \mathcal{X}_{t-1}$ and Q_t (resp. Q_{t+1}) is lower semi-continuous on $\mathcal{X}_{t-1}$ (resp. $\mathcal{X}_t$), we conclude that

$$g_t(x_t^*, x_{t-1}^*) \leq 0, \quad f_t(x_t^*, x_{t-1}^*) + \mathcal{Q}_{t+1}(x_t^*) - \mathcal{Q}_t(x_{t-1}^*) < 0$$

and hence that $x_t^* \in X_t(x_{t-1}^*)$ (recall that $\mathcal{X}_t$ is closed) and $F_t(x_t^*, x_{t-1}^*) < \mathcal{Q}_t(x_{t-1}^*)$ due to the definition of X_t and F_t in (2.2) and (2.4), respectively. Since this contradicts the definition of Q_t in (2.4), the above claim follows. Combining

$$\begin{array}{rcl} 0 \leq \mathcal{Q}_t(x_{t-1}^k) - \underline{\mathcal{Q}}_t^k(x_{t-1}^k) & \leq & \mathcal{Q}_t(x_{t-1}^k) - \underline{\mathcal{Q}}_t^{k-1}(x_{t-1}^k), \\ & = & \mathcal{Q}_t(x_{t-1}^k) - f_t^{k-1}(x_t^k, x_{t-1}^k) - \mathcal{Q}_{t+1}^{k-1}(x_t^k) \text{ [by definition of } x_t^k] \end{array}$$

with relations (2.21), (2.22), (2.23) we obtain $\lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^k) - \underline{\mathcal{Q}}_t^k(x_{t-1}^k) = 0$. Also observe that

$$
\begin{aligned}
&\lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^k) - \sum_{\tau=t}^{T} f_\tau(x_\tau^k, x_{\tau-1}^k) \\
&= \underbrace{\lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^k) - f_t(x_t^k, x_{t-1}^k) - \mathcal{Q}_{t+1}(x_t^k)}_{=0 \text{ by } (2.23)} + \underbrace{\lim_{k\to+\infty} \mathcal{Q}_{t+1}(x_t^k) - \sum_{\tau=t+1}^{T} f_\tau(x_\tau^k, x_{\tau-1}^k)}_{=0 \text{ using } \mathcal{H}(t+1)-(iii)}, \\
&= 0,
\end{aligned}
$$

and we have shown $\mathcal{H}(t)$-(ii),(iii).

Finally, if $t \geq 2$, $\mathcal{H}(t)$-(iv) follows from

$$
\begin{array}{rcl}
0 \leq \mathcal{Q}_t(x_{t-1}^k) - \mathcal{Q}_t^k(x_{t-1}^k) & \leq & \mathcal{Q}_t(x_{t-1}^k) - \mathcal{C}_t^k(x_{t-1}^k) \text{ since } \mathcal{Q}_t^k \geq \mathcal{C}_t^k, \\
& = & \mathcal{Q}_t(x_{t-1}^k) - f_t^{k-1}(x_t^k, x_{t-1}^k) - \mathcal{Q}_{t+1}^{k-1}(x_t^k) \text{ [by definition of } x_t^k]
\end{array}
$$

combined with relations (2.21), (2.22), (2.23). ∎

2.3. **Forward-Backward DCuP.** It is also possible to have for each iteration both a forward and backward pass and compute cuts for $\mathcal{Q}_t$ in the backward passes and cuts for f_t, g_t in both the forward and backward passes. The corresponding extension of the algorithm is given below and the convergence of this variant of DCuP is given in Theorem 2.7.

Forward-Backward DCuP (Dynamic Cutting Plane) with linearizations computed in forward and backward passes.

Step 0. Initialization. Let $\mathcal{Q}_t^0 : \mathcal{X}_{t-1} \to \mathbb{R}$, $t = 2, \ldots, T+1$, be affine functions satisfying $\mathcal{Q}_t^0 \leq \mathcal{Q}_t, t = 2, \ldots, T$, and $\mathcal{Q}_{T+1}^0 \equiv 0$, and let $f_t^0, g_t^0 : \mathcal{X}_t \times \mathcal{X}_{t-1} \to \mathbb{R}$, $t = 1, \ldots, T$, be affine functions such that $f_t^0 \leq f_t, g_t^0 \leq g_t$. Set $k = 1$.

Step 1. Forward pass. Setting $x_0^{2k-1} = x_0$, for $t = 1, 2, \ldots, T$, compute an optimal solution x_t^{2k-1} of

$$
\left\{
\begin{array}{l}
\min\limits_{x_t} \; f_t^{2k-2}(x_t, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k-1}(x_t) \\
x_t \in X_t^{2k-2}(x_{t-1}^{2k-1}),
\end{array}
\right. \tag{2.25}
$$

where X_t^{2k-2} is given by (2.7) with $k-1$ replaced by $2k-2$. Compute $f_t(x_t^{2k-1}, x_{t-1}^{2k-1})$, $g_t(x_t^{2k-1}, x_{t-1}^{2k-1})$, and subgradients of f_t, g_{ti}, $i = 1, \ldots, p$, at $(x_t^{2k-1}, x_{t-1}^{2k-1})$ with corresponding linearizations $\ell_{f_t}(\cdot; (x_t^{2k-1}, x_{t-1}^{2k-1}))$ and $\ell_{g_{ti}}(\cdot; (x_t^{2k-1}, x_{t-1}^{2k-1}))$. Define

$$
f_t^{2k-1} = \max\Big(f_t^{2k-2}, \ell_{f_t}(\cdot; (x_t^{2k-1}, x_{t-1}^{2k-1}))\Big), \tag{2.26}
$$

$$
g_t^{2k-1} = (g_{t1}^{2k-1}, \ldots, g_{tp}^{2k-1}) \text{ where } g_{ti}^{2k-1} = \max\Big(g_{ti}^{2k-2}, \ell_{g_{ti}}(\cdot; (x_t^{2k-1}, x_{t-1}^{2k-1}))\Big), i = 1, \ldots, p. \tag{2.27}
$$

Step 2. Backward pass. For $t = T, T-1, \ldots, 1$, solve the problem

$$
\underline{\mathcal{Q}}_t^k(x_{t-1}^{2k-1}) := \left\{
\begin{array}{l}
\min\limits_{x_t} \; f_t^{2k-1}(x_t, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^k(x_t) \\
x_t \in X_t^{2k-1}(x_{t-1}^{2k-1}).
\end{array}
\right. \tag{2.28}
$$

Denoting by x_t^{2k} an optimal solution of (2.28), compute $f_t(x_t^{2k}, x_{t-1}^{2k-1})$, $g_t(x_t^{2k}, x_{t-1}^{2k-1})$, and subgradients of f_t and g_{ti}, $i = 1, \ldots, p$, at $(x_t^{2k}, x_{t-1}^{2k-1})$, with corresponding linearizations $\ell_{f_t}(\cdot; (x_t^{2k}, x_{t-1}^{2k-1}))$, $\ell_{g_{ti}}(\cdot; (x_t^{2k}, x_{t-1}^{2k-1}))$. Define

$$
f_t^{2k} = \max\Big(f_t^{2k-1}, \ell_{f_t}(\cdot; (x_t^{2k}, x_{t-1}^{2k-1}))\Big), \tag{2.29}
$$

$$
g_t^{2k} = (g_{t1}^{2k}, \ldots, g_{tp}^{2k}) \text{ where } g_{ti}^{2k} = \max\Big(g_{ti}^{2k-1}, \ell_{g_{ti}}(\cdot; (x_t^{2k}, x_{t-1}^{2k-1}))\Big), i = 1, \ldots, p. \tag{2.30}
$$

If $t \geq 2$, take a subgradient β_t^k of $\underline{\mathcal{Q}}_t^k(\cdot)$ at x_{t-1}^{2k-1}, and store the new cut

$$\mathcal{C}_t^k(x_{t-1}) := \underline{\mathcal{Q}}_t^k(x_{t-1}^{2k-1}) + \langle \beta_t^k, x_{t-1} - x_{t-1}^{2k-1} \rangle$$

for $\mathcal{Q}_t$, making up the new approximation $\mathcal{Q}_t^k = \max\{\mathcal{Q}_t^{k-1}, \mathcal{C}_t^k\}$.

Step 4. Do $k \leftarrow k+1$ and go to Step 1.

Theorem 2.7. *Let Assumption (H1) hold. Define*

$$\mathcal{H}(t) \left\{ \begin{array}{ll} (i) & \lim_{k\to+\infty} \max(g_t(x_t^{2k-1}, x_{t-1}^{2k-1}), 0) = 0,\ \lim_{k\to+\infty} \max(g_t(x_t^{2k}, x_{t-1}^{2k-1}), 0) = 0, \\ (ii) & \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^{2k-1}) - \underline{\mathcal{Q}}_t^k(x_{t-1}^{2k-1}) = 0, \\ (iii) & \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^{2k-1}) - \sum_{\tau=t}^{T} f_\tau(x_\tau^{2k-1}, x_{\tau-1}^{2k-1}) = 0, \\ (iv) & \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^{2k-1}) - \mathcal{Q}_t^k(x_{t-1}^{2k-1}) = 0. \end{array} \right.$$

Then $\mathcal{H}(t)$-(i) holds for $t = 1, \ldots, T$, $\mathcal{H}(t)$-(ii),(iii) hold for $t = 1, \ldots, T+1$, and $\mathcal{H}(t)$-(iv) holds for $t = 2, \ldots, T+1$. Moreover, the limit of the sequence $(\sum_{t=1}^{T} f_t(x_t^{2k-1}, x_{t-1}^{2k-1}))_{k\geq 1}$ is the optimal value $\mathcal{Q}_1(x_0)$ of (2.1) and any accumulation point of the sequence $(x_1^{2k-1}, \ldots, x_T^{2k-1})$ is an optimal solution to (2.1).

Proof: For $t = 1, \ldots, T$, let us define the sequence $(y_t^k)_k$ by $y_t^{2k} = (x_t^{2k}, x_{t-1}^{2k-1})$ and $y_t^{2k-1} = (x_t^{2k-1}, x_{t-1}^{2k-1})$ for all $k \geq 1$. Let us first show $\mathcal{H}(t)$-(i), $t = 1, \ldots, T$. Let $t \in \{1, \ldots, T\}$. Since $x_t^{2k-1} \in X_t^{2k-2}(x_{t-1}^{2k-1})$, we have $g_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) \leq 0$ and therefore if $g_t(x_t^{2k-1}, x_{t-1}^{2k-1}) \geq 0$ we have

$$\max(g_t(x_t^{2k-1}, x_{t-1}^{2k-1}), 0) = g_t(x_t^{2k-1}, x_{t-1}^{2k-1}) \leq g_t(x_t^{2k-1}, x_{t-1}^{2k-1}) - g_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}). \tag{2.31}$$

Since $g_t^{2k-2} \leq g_t$, the above relation also holds when $g_t(x_t^{2k-1}, x_{t-1}^{2k-1}) \leq 0$ and therefore holds for every $k \geq 1$. Similarly, since $x_t^{2k} \in X_t^{2k-1}(x_{t-1}^{2k-1})$ we have $g_t^{2k-1}(x_t^{2k}, x_{t-1}^{2k-1}) \leq 0$ which implies

$$\max(g_t(x_t^{2k}, x_{t-1}^{2k-1}), 0) \leq g_t(x_t^{2k}, x_{t-1}^{2k-1}) - g_t^{2k-1}(x_t^{2k}, x_{t-1}^{2k-1}) \tag{2.32}$$

for all $k \geq 1$. Next, $g_{ti} \geq g_{ti}^k \geq \ell_{g_{ti}}(\cdot; y_t^k)$, which implies

$$g_t^k(y_t^k) = g_t(y_t^k),\ \forall\, k \geq 1. \tag{2.33}$$

Using (2.33) and applying Lemma 2.5 to $f = g_t$, $f^k = g_t^k$, $y^k = y_t^k$ (observe that the assumptions of the lemma are satisfied), we obtain

$$\left\{ \begin{array}{l} \lim_{k\to+\infty} g_t(x_t^{2k-1}, x_{t-1}^{2k-1}) - g_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) = 0, \\ \lim_{k\to+\infty} g_t(x_t^{2k}, x_{t-1}^{2k-1}) - g_t^{2k-1}(x_t^{2k}, x_{t-1}^{2k-1}) = 0. \end{array} \right. \tag{2.34}$$

Combining (2.31), (2.32), and (2.34) we get

$$\lim_{k\to+\infty} \max(g_t(x_t^{2k-1}, x_{t-1}^{2k-1}), 0) = 0,\ \lim_{k\to+\infty} \max(g_t(x_t^{2k}, x_{t-1}^{2k-1}), 0) = 0, \tag{2.35}$$

which achieves the proof of $\mathcal{H}(t)$-(i).

Let us now show $\mathcal{H}(1)$-(ii), (iii) and $\mathcal{H}(t)$-(ii)-(iii), (iv) for $t = 2, \ldots, T+1$ by backward induction on t. Clearly, $\mathcal{H}(T+1)$-(ii),(iii), (iv) holds. Now, fix $t \in \{1, \ldots, T\}$ and assume that $\mathcal{H}(t+1)$-(ii), (iii), (iv) holds We will show that $\mathcal{H}(t)$-(ii), (iii) holds and that $\mathcal{H}(t)$-(iv) holds if $t \geq 2$. Since $f_t \geq f_t^k \geq \ell_{f_t}(\cdot; y_t^k)$, we have $f_t(y_t^k) \geq f_t^k(y_t^k) \geq \ell_{f_t}(y_t^k; y_t^k) = f_t(y_t^k)$ and therefore for all $k \geq 1$,

$$f_t^k(y_t^k) = f_t(y_t^k). \tag{2.36}$$

From (2.36), $\lim_{k\to+\infty} f_t(y_t^k) - f_t^k(y_t^k) = 0$ and applying Lemma 2.5 to $f = f_t$, $f^k = f_t^k$, and $(y^k) = (y_t^k)$ (observe that the assumptions of the lemma are satisfied), we obtain

$$\left\{ \begin{array}{l} \lim_{k\to+\infty} f_t(x_t^{2k-1}, x_{t-1}^{2k-1}) - f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) = 0, \\ \lim_{k\to+\infty} f_t(x_t^{2k}, x_{t-1}^{2k-1}) - f_t^{2k-1}(x_t^{2k}, x_{t-1}^{2k-1}) = 0. \end{array} \right. \tag{2.37}$$

Using the fact that $x_t^{2k-1} \in X_t^{2k-2}(x_{t-1}^{2k-1}) \supseteq X_t(x_{t-1}^{2k-1})$, the relation $f_t^{2k-2}(\cdot, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k-1}(\cdot) \leq f_t(\cdot, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}(\cdot)$, and recalling defintions of x_t^{2k-1} and $\mathcal{Q}_t$, we get

$$f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}) \leq \mathcal{Q}_t(x_{t-1}^{2k-1}).$$

Since the sequence $f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}) - \mathcal{Q}_t(x_{t-1}^{2k-1})$ is bounded it has a finite limit sup which satisfies

$$\overline{\lim_{k\to+\infty}} f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}) - \mathcal{Q}_t(x_{t-1}^{2k-1}) \leq 0. \tag{2.38}$$

The induction hypothesis gives

$$\lim_{k\to+\infty} \mathcal{Q}_{t+1}^{k}(x_t^{2k-1}) - \mathcal{Q}_{t+1}(x_t^{2k-1}) = 0.$$

Applying Lemma 2.5 to $f = \mathcal{Q}_{t+1}, f^k = \mathcal{Q}_{t+1}^k, y^k = x_t^{2k-1}$ (observe that the assumptions of the lemma are satisfied), we obtain

$$\lim_{k\to+\infty} \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}) - \mathcal{Q}_{t+1}(x_t^{2k-1}) = 0.$$

Together with (2.37), (2.38) this relation, implies

$$\begin{array}{l}\overline{\lim}_{k\to+\infty} f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}) - \mathcal{Q}_t(x_{t-1}^{2k-1})\\ = \overline{\lim}_{k\to+\infty} f_t^{2k-1}(x_t^{2k}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k}(x_t^{2k-1}) - \mathcal{Q}_t(x_{t-1}^{2k-1})\\ = \overline{\lim}_{k\to+\infty} f_t(x_t^{2k-1}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}(x_t^{2k-1}) - \mathcal{Q}_t(x_{t-1}^{2k-1}) \leq 0.\end{array} \tag{2.39}$$

Let us now show by contradiction that

$$\underline{\lim}_{k\to+\infty}\ f_t(x_t^{2k-1}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}(x_t^{2k-1}) - \mathcal{Q}_t(x_{t-1}^{2k-1}) \geq 0. \tag{2.40}$$

If (2.40) does not hold, using the fact that $(x_t^{2k-1}, x_{t-1}^{2k-1})$ is a sequence from the compact set $\mathcal{X}_t \times \mathcal{X}_{t-1}$ and the lower semicontinuity of $f_t, g_t, \mathcal{Q}_{t+1}, \mathcal{Q}_t$, we can find a subsequence $(x_t^k, x_{t-1}^k)_{k\in K}$ converging to some $(x_t^*, x_{t-1}^*) \in \mathcal{X}_t \times \mathcal{X}_{t-1}$ such that

$$f_t(x_t^*, x_{t-1}^*) + \mathcal{Q}_{t+1}(x_t^*) < \mathcal{Q}_t(x_{t-1}^*),$$

$g_t(x_t^*, x_{t-1}^*) \leq 0$, and $x_t^* \in X_t(x_{t-1}^*)$, which is a contradiction. Therefore (2.40) must hold and we have shown

$$\lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^{2k-1}) - f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) - \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}) = \lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^{2k-1}) - f_t(x_t^{2k-1}, x_{t-1}^{2k-1}) - \mathcal{Q}_{t+1}(x_t^{2k-1}) = 0. \tag{2.41}$$

As before, note that the optimal value of (2.28) is larger than the optimal value of (2.25), i.e.,

$$\underline{\mathcal{Q}}_t^k(x_{t-1}^{2k-1}) \geq f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) + \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}), \tag{2.42}$$

implying

$$0 \leq \mathcal{Q}_t(x_{t-1}^{2k-1}) - \underline{\mathcal{Q}}_t^k(x_{t-1}^{2k-1}) \leq \mathcal{Q}_t(x_{t-1}^{2k-1}) - f_t^{2k-2}(x_t^{2k-1}, x_{t-1}^{2k-1}) - \mathcal{Q}_{t+1}^{k-1}(x_t^{2k-1}), \tag{2.43}$$

which, together with (2.41), gives $\mathcal{H}(t)$-(ii). Next,

$$\begin{array}{l}\displaystyle\lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^{2k-1}) - \sum_{\tau=t}^{T} f_\tau(x_\tau^{2k-1}, x_{\tau-1}^{2k-1})\\ \displaystyle= \underbrace{\lim_{k\to+\infty} \mathcal{Q}_t(x_{t-1}^{2k-1}) - f_t(x_t^{2k-1}, x_{t-1}^{2k-1}) - \mathcal{Q}_{t+1}(x_t^{2k-1})}_{=0 \text{ by } (2.41)} + \underbrace{\lim_{k\to+\infty} \mathcal{Q}_{t+1}(x_t^{2k-1}) - \sum_{\tau=t+1}^{T} f_\tau(x_\tau^{2k-1}, x_{\tau-1}^{2k-1})}_{=0 \text{ using } \mathcal{H}(t+1)-(iii)},\\ = 0,\end{array}$$

and we obtain $\mathcal{H}(t)$-(iii). Finally, for $t \geq 2$,

$$\begin{array}{rcl}0 \leq \mathcal{Q}_t(x_{t-1}^{2k-1}) - \mathcal{Q}_t^k(x_{t-1}^{2k-1}) & \leq & \mathcal{Q}_t(x_{t-1}^{2k-1}) - \mathcal{C}_t^k(x_{t-1}^{2k-1}) \text{ since } \mathcal{Q}_t^k \geq \mathcal{C}_t^k,\\ & = & \mathcal{Q}_t(x_{t-1}^{2k-1}) - \underline{\mathcal{Q}}_t^k(x_{t-1}^{2k-1}) \text{ by definition of } \mathcal{C}_t^k\end{array}$$

which combines with (2.43) to show $\mathcal{H}(t)$-(iv). ∎

2.4. **Computation of the subgradient in Step d) of DCuP.** This subsection explains how to compute a subgradient β_t^k of $\underline{\mathcal{Q}}_t^{k-1}(\cdot)$ at x_{t-1}^k in Step d) of DCuP.

Observe that we can express $\underline{\mathcal{Q}}_t^{k-1}$ as

$$\underline{\mathcal{Q}}_t^{k-1}(x_{t-1}) = \left\{ \begin{array}{l} \min\limits_{x_t \in \mathbb{R}^n, f, \theta \in \mathbb{R}} f + \theta \\ x_t \in \mathcal{X}_t, \\ f \geq \ell_{f_t}(x_t, x_{t-1}, (x_t^j, x_{t-1}^j)), j = 1, \ldots, k-1, \\ \theta \geq \underline{\mathcal{Q}}_{t+1}^{i-1}(x_t^i) + \langle \beta_{t+1}^i, x_t - x_t^i \rangle, \ i = 1, \ldots, k-1, \\ \ell_{g_{ti}}(x_t, x_{t-1}, (x_t^j, x_{t-1}^j)) \leq 0, \ j = 1, \ldots, k-1, i = 1, \ldots, p, \\ A_t x_t + B_t x_{t-1} = b_t. \end{array} \right. \tag{2.44}$$

Due to Assumption (H1)-2), for every $x_{t-1} \in \mathcal{X}_{t-1}$, there exists $x_t \in \mathrm{ri}(\mathcal{X}_t)$ such that $A_t x_t + B_t x_{t-1} = b_t$ and $g_t(x_t, x_{t-1}) \leq 0$, which implies that for every $i = 1, \ldots, p$, and $j = 1, \ldots, k-1$, we have

$$\ell_{g_{ti}}(x_t, x_{t-1}, (x_t^j, x_{t-1}^j)) \leq g_{ti}(x_t, x_{t-1}) \leq 0$$

and therefore Slater constraint qualification holds for problem (2.44) for every $x_{t-1} \in \mathcal{X}_{t-1}$. Next observe that the feasible set of (2.44) is compact and therefore the objective function is bounded on the feasible set. It follows that the optimal value of (2.44) is finite and by the Duality Theorem, we can write problem (2.44) as the optimal value of the corresponding dual problem. To write this dual, it is convenient to rewrite $\underline{\mathcal{Q}}_t^{k-1}$ on $\mathcal{X}_{t-1}$ as

$$\underline{\mathcal{Q}}_t^{k-1}(x_{t-1}) = \left\{ \begin{array}{l} \min\limits_{x_t \in \mathbb{R}^n, f, \theta \in \mathbb{R}} f + \theta \\ x_t \in \mathcal{X}_t, \\ f\mathbf{e} \geq A_t^{k-1} x_t + B_t^{k-1} x_{t-1} + C_t^{k-1}, \\ \theta\mathbf{e} \geq \theta_{t+1}^{1:k-1} + \beta_{t+1}^{1:k-1} x_t, \\ D_t^{k-1} x_t + E_t^{k-1} x_{t-1} + H_t^{k-1} \leq 0, \\ A_t x_t + B_t x_{t-1} = b_t, \end{array} \right. \tag{2.45}$$

where $\mathbf{e}$ is a vector of ones of dimension $k-1$ and $A_t^{k-1}, B_t^{k-1}, D_t^{k-1}, E_t^{k-1}, \beta_{t+1}^{1:k-1}$ (resp. $C_t^{k-1}, H_t^{k-1}, \theta_{t+1}^{1:k-1}$) are matrices (resp. vectors) of appropriate dimensions. In particular, $\beta_{t+1}^{1:k-1}$ is a matrix with $k-1$ rows with i-th row equal to $(\beta_{t+1}^i)^\top$ and $\theta_{t+1}^{1:k-1}$ is a vector of size $k-1$ with i-th component given by $\theta_{t+1}^i = \underline{\mathcal{Q}}_{t+1}^{i-1}(x_t^i) - \langle \beta_{t+1}^i, x_t^i \rangle$.

We now write the dual of (2.45) as

$$\underline{\mathcal{Q}}_t^{k-1}(x_{t-1}) = \left\{ \begin{array}{l} \max\limits_{\alpha, \mu, \delta, \lambda} h_{t, x_{t-1}}(\alpha, \lambda, \mu, \delta) \\ \alpha \geq 0, \mu \geq 0, \delta \geq 0, \lambda, \end{array} \right. \tag{2.46}$$

where dual function $h_{t, x_{t-1}}$ is given by

$$h_{t, x_{t-1}}(\alpha, \lambda, \mu, \delta) = \left\{ \begin{array}{l} \min\limits_{x_t \in \mathbb{R}^n, f, \theta \in \mathbb{R}} L_{t, x_{t-1}}(x_t, f, \theta; \alpha, \lambda, \mu, \delta) \\ x_t \in \mathcal{X}_t, \end{array} \right. \tag{2.47}$$

with Lagrangian $L_{t, x_{t-1}}(x_t, f, \theta; \alpha, \lambda, \mu, \delta)$ given by

$$\begin{array}{rcl} L_{t, x_{t-1}}(x_t, f, \theta; \alpha, \lambda, \mu, \delta) & = & f + \theta + \langle \alpha, A_t^{k-1} x_t + B_t^{k-1} x_{t-1} + C_t^{k-1} - f\mathbf{e} \rangle + \langle \lambda, A_t x_t + B_t x_{t-1} - b_t \rangle \\ & & + \langle \mu, D_t^{k-1} x_t + E_t^{k-1} x_{t-1} + H_t^{k-1} \rangle + \langle \delta, \theta_{t+1}^{1:k-1} + \beta_{t+1}^{1:k-1} x_t - \theta\mathbf{e} \rangle. \end{array}$$

Next, let $(\alpha_t^k, \lambda_t^k, \mu_t^k, \delta_t^k)$ be an optimal solution of (2.46) written for $x_{t-1} = x_{t-1}^k$. With this notation, we have

$$\beta_t^k = (B_t^{k-1})^\top \alpha_t^k + B_t^\top \lambda_t^k + (E_t^{k-1})^\top \mu_t^k \in \partial \underline{\mathcal{Q}}_t^{k-1}(x_{t-1}^k). \tag{2.48}$$

When $\mathcal{X}_t$ is polyhedral, formula (2.48) follows from Duality for linear programming. For a more general convex set $\mathcal{X}_t$, formula (2.48) directly follows from applying to value function $\underline{\mathcal{Q}}_t^{k-1}$ Lemma 2.1 in [3] or Proposition 3.2 in [6] which respectively provide a characterization of the subdifferential and subgradients for value functions of general convex optimization problems (whose argument is in the objective function and in linear and nonlinear coupling constraints of the corresponding optimization problem). For the interested reader and for the sake of completeness, we provide in the Appendix a proof of relation (2.48) specializing to the particular case of value function $\underline{\mathcal{Q}}_t^{k-1}$ the proof of Lemma 2.1 in [3].

3. The StoDCuP (Stochastic Dynamic Cutting Plane) algorithm

3.1. **Problem formulation and assumptions.** We consider multistage stochastic nonlinear optimization problems of the form

(3.49)

$$\min_{x_1 \in X_1(x_0,\xi_1)} f_1(x_1, x_0, \xi_1) + \mathbb{E}\left[\min_{x_2 \in X_2(x_1,\xi_2)} f_2(x_2, x_1, \xi_2) + \mathbb{E}\left[\ldots + \mathbb{E}\left[\min_{x_T \in X_T(x_{T-1},\xi_T)} f_T(x_T, x_{T-1}, \xi_T)\right]\right]\right],$$

where x_0 is given, $(\xi_t)_{t=2}^T$ is a stochastic process, ξ_1 is deterministic, and

$$X_t(x_{t-1}, \xi_t) = \{x_t \in \mathbb{R}^n : A_t x_t + B_t x_{t-1} = b_t, g_t(x_t, x_{t-1}, \xi_t) \le 0, x_t \in \mathcal{X}_t\}.$$

In the constraint set above, $\mathcal{X}_t$ is polyhedral and ξ_t contains in particular the random elements in matrices A_t, B_t, and vector b_t.

We make the following assumption on (ξ_t):

(H0) (ξ_t) is interstage independent and for $t = 2, \ldots, T$, ξ_t is a random vector taking values in $\mathbb{R}^K$ with a discrete distribution and a finite support $\Theta_t = \{\xi_{t1}, \ldots, \xi_{tM_t}\}$ with $p_{ti} = \mathbb{P}(\xi_t = \xi_{ti}), i = 1, \ldots, M_t$, while ξ_1 is deterministic.

For this problem, we can write Dynamic Programming equations: the first stage problem is

$$\mathcal{Q}_1(x_0) = \left\{ \begin{array}{l} \min_{x_1 \in \mathbb{R}^n} f_1(x_1, x_0, \xi_1) + \mathcal{Q}_2(x_1) \\ x_1 \in X_1(x_0, \xi_1) \end{array} \right. \tag{3.50}$$

for x_0 given and for $t = 2, \ldots, T$, $\mathcal{Q}_t(x_{t-1}) = \mathbb{E}_{\xi_t}[\mathfrak{Q}_t(x_{t-1}, \xi_t)]$ with

$$\mathfrak{Q}_t(x_{t-1}, \xi_t) = \left\{ \begin{array}{l} \min_{x_t \in \mathbb{R}^n} f_t(x_t, x_{t-1}, \xi_t) + \mathcal{Q}_{t+1}(x_t) \\ x_t \in X_t(x_{t-1}, \xi_t), \end{array} \right. \tag{3.51}$$

with the convention that $\mathcal{Q}_{T+1}$ is null.

We set $\mathcal{X}_0 = \{x_0\}$ and make the following assumptions (H1)-Sto on the problem data:

(H1)-Sto: for $t = 1, \ldots, T$,

1) $\mathcal{X}_t$ is a nonempty, compact, and polyhedral set.
2) For every $j = 1, \ldots, M_t$, the function $f_t(\cdot, \cdot, \xi_{tj})$ is convex, proper, lower semicontinuous on $\mathcal{X}_t \times \mathcal{X}_{t-1}$ and $\mathcal{X}_t \times \mathcal{X}_{t-1} \subset \text{int}(\text{dom}(f_t(\cdot, \cdot, \xi_{tj})))$.
3) For every $j = 1, \ldots, M_t$, each component $g_{ti}(\cdot, \cdot, \xi_{tj}), i = 1, \ldots, p$, of function $g_t(\cdot, \cdot, \xi_{tj})$ is convex, proper, lower semicontinuous such that $\mathcal{X}_t \times \mathcal{X}_{t-1} \subset \text{int}(\text{dom}(g_{ti}(\cdot, \cdot, \xi_{tj})))$.
4) $X_1(x_0, \xi_1) \neq \emptyset$ and for every $t = 2, \ldots, T$, for every $j = 1, \ldots, M_t$, $\mathcal{X}_{t-1} \subset \text{int}(\text{dom}(X_t(\cdot, \xi_{tj})))$.

Remark 3.1. *Nonlinear constraints of form $h_{ti}(x_t, \xi_t) \le 0$ or $h_{ti}(x_t) \le 0$ at stage t can be handled, adding the corresponding component functions h_{ti} in g_t, as long as (H1)-Sto is satisfied. In particular, convexity of $h_{ti}(\cdot, \xi_{tj})$ is required for $j = 1, \ldots, M_t$.*

It is easy to show that under Assumption (H1)-Sto, functions $\mathcal{Q}_t$ are convex and Lipschitz continuous on $\mathcal{X}_{t-1}$:

Lemma 3.2. *Let Assumption (H1)-Sto hold. Then $\mathcal{Q}_t$ is convex Lipschitz continuous on $\mathcal{X}_{t-1}$ for $t = 2, \ldots, T+1$.*

Proof: The proof is analogue to the proof of Lemma 2.1. ∎

3.2. **Algorithm.** The algorithm to be presented in this section for solving (3.49) is an extension of the DCuP algorithm to the stochastic case. All inequalities and equalities between random variables in the rest of the paper hold almost surely.

Due to Assumption (H0), the $\prod_{t=2}^{T} M_t$ realizations of $(\xi_t)_{t=1}^T$ form a scenario tree of depth $T+1$ where the root node n_0 associated to a stage 0 (with decision x_0 taken at that node) has one child node n_1 associated to the first stage (with ξ_1 deterministic).

We denote by $\mathcal{N}$ the set of nodes, by $\texttt{Nodes}(t)$ the set of nodes for stage t and for a node n of the tree, we define:

- $C(n)$: the set of children nodes (the empty set for the leaves);
- x_n: a decision taken at that node;
- p_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 contains in particular the realizations b_n of b_t, A_n of A_t, and B_n of B_t.
- $\xi_{[n]}$: the history of the realizations of process (ξ_t) from the first stage node n_1 to node n: for a node n of stage t, the i-th component of $\xi_{[n]}$ is $\xi_{\mathcal{P}^{t-i}(n)}$ for $i=1,\ldots,t$, where $\mathcal{P}:\mathcal{N}\to\mathcal{N}$ is the function associating to a node its parent node (the empty set for the root node).

At each iteration of the algorithm, trial points are computed on a sampled scenario and lower bounding affine functions, called cuts in the sequel, are built for convex functions $\mathcal{Q}_t, t=2,\ldots,T+1$, at these trial points. More precisely, at iteration k denoting by x_{t-1}^k the trial point for stage $t-1$, the cut

$$\mathcal{C}_t^k(x_{t-1}) = \theta_t^k + \langle \beta_t^k, x_{t-1}\rangle \tag{3.52}$$

is built for $\mathcal{Q}_t$ with the convention that $\mathcal{C}_{T+1}^k$ is the null function (see below for the computation of θ_t^k, β_t^k). As in SDDP, we end up iteration k with an approximation $\mathcal{Q}_t^k$ of $\mathcal{Q}_t$ which is a maximum of k affine functions: $\mathcal{Q}_t^k(x_{t-1}) = \max_{0\le j\le k}\ \mathcal{C}_t^j(x_{t-1})$.

Additionally, the variant we propose builds cutting plane approximations of convex functions $f_t(\cdot,\cdot,\xi_{tj})$ and $g_{ti}(\cdot,\cdot,\xi_{tj})$, $t=1,\ldots,T, i=1,\ldots,p, j=1,\ldots,M_t$, computing linearizations of these functions. At the end of iteration k, these approximations will be denoted by f_{tj}^k and g_{tij}^k for $f_t(\cdot,\cdot,\xi_{tj})$ and $g_{ti}(\cdot,\cdot,\xi_{tj})$ respectively, and take the form of a maximum of k affine functions. We use the notation

$$\begin{array}{l} f_{tj}^k(x_t,x_{t-1}) = \displaystyle\max_{\ell=0,\ldots,k}\ a_{tj}^\ell x_t + b_{tj}^\ell x_{t-1} + c_{tj}^\ell,\\ g_{tij}^k(x_t,x_{t-1}) = \displaystyle\max_{\ell=0,\ldots,k}\ d_{tij}^\ell x_t + e_{tij}^\ell x_{t-1} + h_{tij}^\ell,\end{array}$$

where $a_{tj}^\ell, b_{tj}^\ell, d_{tij}^\ell$, and e_{tij}^ℓ are n-dimensional row vectors. The trial points of iteration k are computed before updating these functions, therefore using approximations $f_{tj}^{k-1}, g_{tij}^{k-1}$, and $\mathcal{Q}_{t+1}^{k-1}$ of $f_t(\cdot,\cdot,\xi_{tj})$, $g_{ti}(\cdot,\cdot,\xi_{tj})$, and $\mathcal{Q}_{t+1}$ available at the end of iteration $k-1$. These trial points are decisions computed at nodes $(n_1^k, n_2^k, \ldots, n_T^k)$ using these approximations, knowing that $n_1^k = n_1$, and for $t\ge 2$, n_t^k is a node of stage t, child of node n_{t-1}^k, i.e., these nodes correspond to a sample $(\tilde\xi_1^k, \tilde\xi_2^k, \ldots, \tilde\xi_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$. At iteration k, the linearizations for $f_t(\cdot,\cdot,\xi_{tj})$, $g_{ti}(\cdot,\cdot,\xi_{tj})$ (resp. $\mathcal{Q}_t$) are computed at (x_m^k, x_n^k) (resp. x_n^k) where $n = n_{t-1}^k$, and m is the child node of node n such that $\xi_m = \xi_{tj}$. For convenience, for any node m of stage t, we will denote by $j_t(m)$ the unique index $j_t(m)$ such that $\xi_m = \xi_{t j_t(m)}$. Before detailing the steps of StoDCuP, we need more notation: for all $k\ge 1, t=1,\ldots,T, j=1,\ldots,M_t$, let $X_{tj}^k : \mathcal{X}_{t-1} \rightrightarrows \mathcal{X}_t$ be the multifunction given by

$$X_{tj}^k(x_{t-1}) = \{x_t \in \mathcal{X}_t :\ g_{tij}^k(x_t, x_{t-1}) \le 0, i=1,\ldots,p,\ A_{tj}x_t + B_{tj}x_{t-1} = b_{tj}\}, \tag{3.53}$$

where A_{tj}, B_{tj}, b_{tj} are respectively the realizations of A_t, B_t, and b_t in ξ_{tj} and let $\underline{\mathfrak{Q}}_{tj}^k : \mathcal{X}_{t-1} \to \mathbb{R}$ be the function

$$\underline{\mathfrak{Q}}_{tj}^k(x_{t-1}) = \left\{ \begin{array}{l} \displaystyle\min_{x_t}\ f_{tj}^k(x_t, x_{t-1}) + \mathcal{Q}_{t+1}^k(x_t)\\ x_t \in X_{tj}^k(x_{t-1}). \end{array}\right. \tag{3.54}$$

[1]The same notation $\xi_{\texttt{Index}}$ is used to denote the realization of the process at node $\texttt{Index}$ of the scenario tree and the value of the process (ξ_t) for stage $\texttt{Index}$. 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.

Introducing $k \times n$ matrices

$$A_{tj}^k = \begin{bmatrix} a_{tj}^0 \\ a_{tj}^1 \\ \vdots \\ a_{tj}^k \end{bmatrix}, \; B_{tj}^k = \begin{bmatrix} b_{tj}^0 \\ b_{tj}^1 \\ \vdots \\ b_{tj}^k \end{bmatrix}, \; D_{tij}^k = \begin{bmatrix} d_{tij}^0 \\ d_{tij}^1 \\ \vdots \\ d_{tij}^k \end{bmatrix}, E_{tij}^k = \begin{bmatrix} e_{tij}^0 \\ e_{tij}^1 \\ \vdots \\ e_{tij}^k \end{bmatrix}, \; \beta_t^{0:k} = \begin{bmatrix} (\beta_t^0)^\top \\ (\beta_t^1)^\top \\ \vdots \\ (\beta_t^k)^\top \end{bmatrix},$$

k dimensional vectors,

$$C_{tj}^k = \begin{bmatrix} c_{tj}^0 \\ c_{tj}^1 \\ \vdots \\ c_{tj}^k \end{bmatrix}, \; H_{tij}^k = \begin{bmatrix} h_{tij}^0 \\ h_{tij}^1 \\ \vdots \\ h_{tij}^k \end{bmatrix}, \text{ and } \theta_t^{0:k} = \begin{bmatrix} \theta_t^0 \\ \theta_t^1 \\ \vdots \\ \theta_t^k \end{bmatrix},$$

and matrices and vectors

$$D_{tj}^k = \begin{bmatrix} D_{t1j}^k \\ D_{t2j}^k \\ \vdots \\ D_{tpj}^k \end{bmatrix}, \; E_{tj}^k = \begin{bmatrix} E_{t1j}^k \\ E_{t2j}^k \\ \vdots \\ E_{tpj}^k \end{bmatrix}, \; H_{tj}^k = \begin{bmatrix} H_{t1j}^k \\ H_{t2j}^k \\ \vdots \\ H_{tpj}^k \end{bmatrix},$$

if $\mathcal{X}_t = \{x_t : \mathbb{X}_t x_t \geq \bar{x}_t\}$, we can write problem (3.54) as

$$\underline{\mathfrak{Q}}_{tj}^k(x_{t-1}) = \begin{cases} \min\limits_{x_t, f, \theta} \; f + \theta \\ f\mathbf{e} \geq A_{tj}^k x_t + B_{tj}^k x_{t-1} + C_{tj}^k, \\ A_{tj} x_t + B_{tj} x_{t-1} = b_{tj}, \\ D_{tj}^k x_t + E_{tj}^k x_{t-1} + H_{tj}^k \leq 0, \\ \theta \mathbf{e} \geq \theta_{t+1}^{0:k} + \beta_{t+1}^{0:k} x_t, \; \mathbb{X}_t x_t \geq \bar{x}_t. \end{cases} \tag{3.55}$$

Due to Assumption (H1)-Sto-4), for every $x_{t-1} \in \mathcal{X}_{t-1}$ and $j = 1, \ldots, M_t$, there exists $x_t \in \mathcal{X}_t$ such that $A_{tj} x_t + B_{tj} x_{t-1} = b_{tj}$, and $g_{ti}(x_t, x_{t-1}, \xi_{tj}) \leq 0$, $i = 1, \ldots, p$, which implies $g_{tij}^k(x_t, x_{t-1}) \leq 0$, $i = 1, \ldots, p$, $D_{tj}^k x_t + E_{tj}^k x_{t-1} + H_{tj}^k \leq 0$ and therefore the above problem (3.55) is feasible. Recalling (H1)-Sto-1), this linear program also has a bounded feasible set and therefore its optimal value is the optimal value of the dual problem and can be expressed as:

$$\underline{\mathfrak{Q}}_{tj}^k(x_{t-1}) = \begin{cases} \max\limits_{\alpha, \mu, \delta, \nu, \lambda} \; \alpha^\top (B_{tj}^k x_{t-1} + C_{tj}^k) + \mu^\top (E_{tj}^k x_{t-1} + H_{tj}^k) + \delta^\top \theta_{t+1}^{0:k} + \lambda^\top (b_{tj} - B_{tj} x_{t-1}) + \nu^\top \bar{x}_t \\ (A_{tj}^k)^\top \alpha + (D_{tj}^k)^\top \mu + (\beta_{t+1}^{0:k})^\top \delta - \mathbb{X}_t^\top \nu - (A_{tj})^\top \lambda = 0, \\ \mathbf{e}^\top \alpha = 1, \; \mathbf{e}^\top \delta = 1, \alpha, \mu, \delta, \nu \geq 0. \end{cases}$$

The detailed steps of the algorithm are described below.

Forward StoDCuP (Stochastic Dynamic Cutting Plane) with linearizations computed in a forward pass.

Step 1) **Initialization.** For $t = 1, \ldots, T$, $i = 1, \ldots, p$, take $f_{tj}^0, g_{tij}^0 : \mathcal{X}_t \times \mathcal{X}_{t-1} \to \mathbb{R}$ affine functions satisfying $f_{tj}^0 \leq f_t(\cdot, \cdot, \xi_{tj})$, $g_{tij}^0 \leq g_{ti}(\cdot, \cdot, \xi_{tj})$, and for $t = 2, \ldots, T$, $\mathcal{Q}_t^0 : \mathcal{X}_{t-1} \to \mathbb{R}$ is an affine function satisfying $\mathcal{Q}_t^0 \leq \mathcal{Q}_t$. Set $x_{n_0} = x_0$, set the iteration count k to 1, and $\mathcal{Q}_{T+1}^0 \equiv 0$.

Step 2) Generate a sample $(\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$ corresponding to a set of nodes $(n_1^k, n_2^k, \ldots, n_T^k)$ where $n_1^k = n_1$, and for $t \geq 2$, n_t^k is a node of stage t, child of node n_{t-1}^k. Set $n_0^k = n_0$.
Do $\theta_{T+1}^k = 0$ and $\beta_{T+1}^k = 0$.
For $t = 1, \ldots, T$,
  Let $n = n_{t-1}^k$.
  For every $m \in C(n)$,

compute an optimal solution x_m^k of

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k) = \left\{ \begin{array}{l} \min_{x_m} f_{tj_t(m)}^{k-1}(x_m, x_n^k) + \mathcal{Q}_{t+1}^{k-1}(x_m) \\ x_m \in X_{tj_t(m)}^{k-1}(x_n^k). \end{array} \right. \quad (3.56)$$

Compute an arbitrary subgradient $[s_1; s_2]$ of convex function $f_t(\cdot, \cdot, \xi_m)$ at (x_m^k, x_n^k) where $s_1, s_2 \in \mathbb{R}^n$ and do $a_{tj_t(m)}^k = s_1^\top$, $b_{tj_t(m)}^k = s_2^\top$. For $i = 1, \ldots, p$, compute an arbitrary subgradient $[s_{1i}; s_{2i}]$ of convex function $g_{ti}(\cdot, \cdot, \xi_m)$ at (x_m^k, x_n^k) where $s_{1i}, s_{2i} \in \mathbb{R}^n$ and do $d_{tij_t(m)}^k = s_{1i}^\top$, $e_{tij_t(m)}^k = s_{2i}^\top$. Compute

$$\begin{array}{l} c_{tj_t(m)}^k = f_t(x_m^k, x_n^k, \xi_m) - a_{tj_t(m)}^k x_m^k - b_{tj_t(m)}^k x_n^k, \\ h_{tij_t(m)}^k = g_{ti}(x_m^k, x_n^k, \xi_m) - d_{tij_t(m)}^k x_m^k - e_{tij_t(m)}^k x_n^k. \end{array}$$

Compute an optimal solution $(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k)$ of the dual problem

$$\begin{array}{l} \max_{\alpha,\mu,\delta,\nu,\lambda} \ \alpha^\top (B_{tj_t(m)}^{k-1} x_n^k + C_{tj_t(m)}^{k-1}) + \mu^\top (E_{tj_t(m)}^{k-1} x_n^k + H_{tj_t(m)}^{k-1}) + \delta^\top \theta_{t+1}^{0:k-1} + \lambda^\top (b_{tj_t(m)} - B_{tj_t(m)} x_n^k) + \nu^\top \bar{x}_t \\ (A_{tj_t(m)}^{k-1})^\top \alpha + (D_{tj_t(m)}^{k-1})^\top \mu + (\beta_{t+1}^{0:k-1})^\top \delta - \mathbb{X}_t^\top \nu - (A_{tj_t(m)})^\top \lambda = 0, \\ \mathbf{e}^\top \alpha = 1, \ \mathbf{e}^\top \delta = 1, \alpha, \mu, \delta, \nu \geq 0. \end{array}$$

End For
If $t \geq 2$ compute:

$$\begin{array}{lcl} \beta_t^k & = & \sum_{m \in C(n)} p_m \Big[(B_{tj_t(m)}^{k-1})^\top \alpha_m^k + (E_{tj_t(m)}^{k-1})^\top \mu_m^k - B_{tj_t(m)}^\top \lambda_m^k \Big], \\ \theta_t^k & = & \sum_{m \in C(n)} p_m \Big[\langle \alpha_m^k, C_{tj_t(m)}^{k-1} \rangle + \langle \mu_m^k, H_{tj_t(m)}^{k-1} \rangle + \langle \delta_m^k, \theta_{t+1}^{0:k-1} \rangle + \langle \lambda_m^k, b_{tj_t(m)} \rangle + \langle \nu_m^k, \bar{x}_t \rangle \Big]. \end{array} \quad (3.57)$$

End If
End For

Step 4) Do $k \leftarrow k+1$ and go to Step 2).

We have for StoDCuP the following analogue of Lemma 2.4 for DCuP (the proof is similar to the proof of Lemma 2.4):

Lemma 3.3. *Let Assumptions (H0) and (H1)-Sto hold. Then, the following statements hold for StoDCuP:*

(a) *For $t = 2, \ldots, T$, the sequence $\{\beta_t^k\}_{k=1}^\infty$ is almost surely bounded.*
(b) *There exists $L \geq 0$ such that for each $t = 2, \ldots, T$, $\mathcal{Q}_t^k$ is L-Lipschitz continuous on $\mathcal{X}_{t-1}$ for every $k \geq 1$.*
(c) *There exists $\hat{L} \geq 0$ such that for each $t = 1, \ldots, T$, $j = 1, \ldots, M_t$, functions f_{tj}^k and g_{tij}^k are $\hat{L}$-Lipschitz continuous on $\mathcal{X}_t \times \mathcal{X}_{t-1}$ for every $k \geq 1$ and $i = 1, \ldots, p$.*

Remark 3.4 (On the cuts and linearizations computed)**.** *Assumption (H0) is fundamental for StoDCuP, due to the following claim:*

(C) *StoDCuP builds a cut for $\mathcal{Q}_t, t = 2, \ldots, T$, on any sampled scenario and a single cut for each of the functions $f_t(\cdot, \cdot, \xi_{tj}), g_{ti}(\cdot, \cdot, \xi_{tj}), t = 1, \ldots, T, j = 1, \ldots, M_t$, $i = 1, \ldots, p$, at each iteration.*

The validity of the formulas of the cuts for $\mathcal{Q}_t$ will be checked in Lemma 3.7. The fact that a single cut is built for functions $f_t(\cdot, \cdot, \xi_{tj}), g_{ti}(\cdot, \cdot, \xi_{tj})$, $i = 1, \ldots, p$, $t = 1, \ldots, T, j = 1, \ldots, M_t$, comes from the fact that at iteration k and stage t a cut is built for each of functions $f_t(\cdot, \cdot, \xi_m), g_{ti}(\cdot, \cdot, \xi_m)$, $i = 1, \ldots, p$, $m \in C(n)$, where $n = n_{t-1}^k$, and due to Assumption (H0), to each $m \in C(n)$, corresponds one and only one index $j = j_t(m)$ such that $\xi_m = \xi_{tj} = \xi_{tj_t(m)}$.

Remark 3.5. *The algorithm can be extended to solve risk-averse problems. It was shown in* [8] *that dynamic programming equations can be written and that SDDP can be applied for multistage stochastic linear optimization problems which minimize some extended polyhedral risk measure of the cost. As a special case, spectral risk measures are considered in* [9] *where analytic formulas for some cut coefficients computed by SDDP are available. Similarly, StoDCuP can be extended to solve multistage nonlinear optimization problems with objective and constraint functions as in* (3.49) *if instead of minimizing the expected cost we minimize an* *extended polyhedral risk measure of the cost, as long as Assumptions (H0) and (H1)-Sto are satisfied. It is also possible to apply StoDCuP to solve risk-averse dynamic programming equations with nested conditional risk measures (see* [19], [20] *for details on conditional risk mappings) and objective and constraint functions as in* (3.49)*, again, as long as Assumptions (H0) and (H1)-Sto are satisfied. Using SDDP in this risk-averse setting was proposed in* [21].

We can simulate the policy obtained after $k-1$ iterations of StoDCuP and define decisions x_n^k at each node n of the scenario tree as follows:

Simulation of StoDCuP after $k-1$ iterations.

Set $x_{n_0}^k = x_0$.
For $t = 1, \ldots, T$,
  For every node $n \in \texttt{Nodes}(t-1)$,
    For every $m \in C(n)$,
      compute an optimal solution x_m^k of

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k) = \left\{ \begin{array}{l} \min\limits_{x_m} f_{tj_t(m)}^{k-1}(x_m, x_n^k) + \mathcal{Q}_{t+1}^{k-1}(x_m) \\ x_m \in X_{tj_t(m)}^{k-1}(x_n^k). \end{array} \right. \tag{3.58}$$

    End For
  End For
End For

3.3. **Convergence analysis.** In what follows, if the stage associated to node n is $\tau(n)$, we use the notation

$$\mathcal{S}_n = \{k \in \mathbb{N}^* : n_{\tau(n)}^k = n\}. \tag{3.59}$$

In other words, $\mathcal{S}_n$ the set of iterations k where the sampled scenario passes through node n.

We show in Lemmas 3.6 and 3.7 below properties of the algorithm useful to prove the convergence of StoDCuP given in Theorem 3.8. We start providing simple relations involving the linearizations of objective and constraint functions:

Lemma 3.6. *Let Assumption (H1)-Sto hold. For every $t = 1, \ldots, T$, $j = 1, \ldots, M_t$, $i = 1, \ldots, p$, we have almost surely*

$$f_t(x_t, x_{t-1}, \xi_{tj}) \geq f_{tj}^k(x_t, x_{t-1}), g_{ti}(x_t, x_{t-1}, \xi_{tj}) \geq g_{tij}^k(x_t, x_{t-1}),\ \forall k \geq 0, \forall x_t \in \mathcal{X}_t, \forall x_{t-1} \in \mathcal{X}_{t-1}, \tag{3.60}$$

and for every $k \geq 1$,

$$X_t(x_{t-1}, \xi_{tj}) \subset X_{tj}^k(x_{t-1}),\ \forall\ x_{t-1} \in \mathcal{X}_{t-1}. \tag{3.61}$$

For all $t = 1, \ldots, T$, $i = 1, \ldots, p$, for all $n \in \texttt{Nodes}(t-1)$, for all $k \in \mathcal{S}_n$, we have for all $m \in C(n)$:

$$f_t(x_m^k, x_n^k, \xi_m) = f_{tj_t(m)}^k(x_m^k, x_n^k) \text{ and } g_{ti}(x_m^k, x_n^k, \xi_m) = g_{tij_t(m)}^k(x_m^k, x_n^k),\ \text{a.s.} \tag{3.62}$$

For all $t = 1, \ldots, T$, $i = 1, \ldots, p$, for all $n \in \texttt{Nodes}(t-1)$, for all $k \geq 1$, for all $m \in C(n)$, we have

$$g_{tij_t(m)}^{k-1}(x_m^k, x_n^k) \leq 0,\ \text{a.s.}, \tag{3.63}$$

$$0 \leq \max(g_{ti}(x_m^k, x_n^k, \xi_m), 0) \leq g_{ti}(x_m^k, x_n^k, \xi_m) - g_{tij_t(m)}^{k-1}(x_m^k, x_n^k),\ \text{a.s.} \tag{3.64}$$

Proof: Let us show (3.60). The relation holds for $k = 0$. Now let us fix $t \in \{1, \ldots, T\}$, $j \in \{1, \ldots, M_t\}$, $k \geq 1$ and $\ell \in \{1, \ldots, k\}$. At iteration ℓ, setting $n = n_{t-1}^\ell$, there exists one and only one node m in the set $C(n)$ such that $\xi_m = \xi_{tj}$ with $j = j_t(m)$ and by the subgradient inequality for every $x_t \in \mathcal{X}_t$, for every $x_{t-1} \in \mathcal{X}_{t-1}$, we have

$$\begin{array}{lll} f_t(x_t, x_{t-1}, \xi_{tj}) = f_t(x_t, x_{t-1}, \xi_m) & \geq & \ell_{f_t(\cdot,\cdot,\xi_m)}(x_t, x_{t-1}; (x_m^\ell, x_n^\ell)) \\ & = & f_t(x_m^\ell, x_n^\ell, \xi_m) + a_{tj_t(m)}^\ell(x_t - x_m^\ell) + b_{tj_t(m)}^\ell(x_{t-1} - x_n^\ell), \\ & = & a_{tj_t(m)}^\ell x_t + b_{tj_t(m)}^\ell x_{t-1} + c_{tj_t(m)}^\ell = a_{tj}^\ell x_t + b_{tj}^\ell x_{t-1} + c_{tj}^\ell, \end{array} \tag{3.65}$$

(3.66)

$$\begin{array}{rcl} g_{ti}(x_t, x_{t-1}, \xi_{tj}) = g_{ti}(x_t, x_{t-1}, \xi_m) & \geq & \ell_{g_{ti}(\cdot,\cdot,\xi_m)}(x_t, x_{t-1}; (x_m^\ell, x_n^\ell)), \\ & = & g_{ti}(x_m^\ell, x_n^\ell, \xi_m) + d^\ell_{tij_t(m)}(x_t - x_m^\ell) + e^\ell_{tij_t(m)}(x_{t-1} - x_n^\ell), \\ & = & d^\ell_{tij_t(m)} x_t + e^\ell_{tij_t(m)} x_{t-1} + h^\ell_{tij_t(m)} = d^\ell_{tij} x_t + e^\ell_{tij} x_{t-1} + h^\ell_{tij}. \end{array}$$

It follows that $f_t(\cdot, \cdot, \xi_{tj})$ and $g_{ti}(\cdot, \cdot, \xi_{tj})$ are above all linearizations built for these functions by StoDCuP and therefore also above f^k_{tj} and g^k_{tij} which is given by the maximum of the first k linearizations. Relation (3.60) follows and clearly inclusion (3.61) is a consequence of (3.60).

Take $t \in \{1, \ldots, T\}$, $i \in \{1, \ldots, p\}$, take a node $n \in \texttt{Nodes}(t-1)$ and $k \in \mathcal{S}_n$. Then for any $m \in C(n)$, a linearization is built for $f_t(\cdot, \cdot, \xi_m)$ and $g_{ti}(\cdot, \cdot, \xi_m)$ at (x_m^k, x_n^k). Therefore,

$$\begin{array}{rcl} f_t(x_m^k, x_n^k, \xi_m) & \overset{(3.60)}{\geq} & f^k_{tj_t(m)}(x_m^k, x_n^k) \\ & \geq & a^k_{tj_t(m)} x_m^k + b^k_{tj_t(m)} x_n^k + c^k_{tj_t(m)}, \\ & \overset{(3.65)}{=} & \ell_{f_t(\cdot,\cdot,\xi_m)}(x_m^k, x_n^k; (x_m^k, x_n^k)) = f_t(x_m^k, x_n^k, \xi_m) \text{ since } n^k_{t-1} = n, \\ g_{ti}(x_m^k, x_n^k, \xi_m) & \overset{(3.60)}{\geq} & g^k_{tij_t(m)}(x_m^k, x_n^k) \\ & \geq & d^k_{tij_t(m)} x_m^k + e^k_{tij_t(m)} x_n^k + h^k_{tij_t(m)}, \\ & \overset{(3.66)}{=} & \ell_{g_{ti}(\cdot,\cdot,\xi_m)}(x_m^k, x_n^k; (x_m^k, x_n^k)) = g_{ti}(x_m^k, x_n^k, \xi_m), \text{ since } n^k_{t-1} = n, \end{array}$$

and (3.62) follows.

Relation (3.63) comes from the fact that $x_m^k \in X^{k-1}_{tj_t(m)}(x_n^k)$ by definition of x_m^k (see the simulation of StoDCuP).

Finally take a realization ω of StoDCuP. We show that

$$0 \leq \max(g_{ti}(x_m^k(\omega), x_n^k(\omega), \xi_m), 0) \leq g_{ti}(x_m^k(\omega), x_n^k(\omega), \xi_m) - g^{k-1}_{tij_t(m)}(\omega)(x_m^k(\omega), x_n^k(\omega)). \tag{3.67}$$

If $g_{ti}(x_m^k(\omega), x_n^k(\omega), \xi_m) \leq 0$ then (3.67) holds because $g_{ti}(\cdot, \cdot, \xi_m) \geq g^{k-1}_{tij_t(m)}(\omega)$ and if $g_{ti}(x_m^k(\omega), x_n^k(\omega), \xi_m) > 0$ then (3.67) holds too because of inequality (3.63). Therefore, (3.67) holds. ■

Lemma 3.7 shows the validity of the cuts computed for $\mathcal{Q}_t$:

Lemma 3.7. *Let Assumptions (H0) and (H1)-Sto hold. For every $t = 2, \ldots, T+1$, for every $k \geq 1$, we have almost surely*

$$\mathcal{Q}_t(x_{t-1}) \geq \mathcal{C}^k_t(x_{t-1}) \text{ and } \mathcal{Q}_t(x_{t-1}) \geq \mathcal{Q}^k_t(x_{t-1}),\ \forall x_{t-1} \in \mathcal{X}_{t-1}. \tag{3.68}$$

For all $t = 1, \ldots, T$, $j = 1, \ldots, M_t$, for every $k \geq 1$, we have almost surely

$$\underline{\mathfrak{Q}}^k_{tj}(x_{t-1}) \leq \mathfrak{Q}_t(x_{t-1}, \xi_{tj}) \text{ for all } x_{t-1} \in \mathcal{X}_{t-1}. \tag{3.69}$$

For all $t = 2, \ldots, T$, for every $k \geq 1$, defining $\underline{\mathcal{Q}}^{k-1}_t(x_n^k) = \sum_{j=1}^{M_t} p_{tj} \underline{\mathfrak{Q}}^{k-1}_{tj}(x_n^k)$, we have for every $n \in$ $\texttt{Nodes}(t-1)$ and for all $k \in \mathcal{S}_n$:

$$\underline{\mathcal{Q}}^{k-1}_t(x_n^k) = \mathcal{C}^k_t(x_n^k),\ a.s. \tag{3.70}$$

Proof: Let us show (3.68)-(3.69) by backward induction on t. Relation (3.68) clearly holds for $t = T+1$. Now assume that for some $t \in \{1, \ldots, T\}$, we have $\mathcal{Q}_{t+1}(x_t) \geq \mathcal{Q}^k_{t+1}(x_t)$ for all $x_t \in \mathcal{X}_t$ and all $k \geq 1$. Using Lemma 3.6, we have for all $k \geq 1$, for all $j = 1, \ldots, M_t$, for all $x_t \in \mathcal{X}_t$, $x_{t-1} \in \mathcal{X}_{t-1}$, that $f^k_{tj}(x_t, x_{t-1}) \leq f_t(x_t, x_{t-1}, \xi_{tj})$ and $X_t(x_{t-1}, \xi_{tj}) \subset X^k_{tj}(x_{t-1})$, which, together with the induction hypothesis $\mathcal{Q}^k_{t+1} \leq \mathcal{Q}_{t+1}$, implies

$$\underline{\mathfrak{Q}}^k_{tj}(x_{t-1}) \leq \mathfrak{Q}_t(x_{t-1}, \xi_{tj}) \text{ for all } x_{t-1} \in \mathcal{X}_{t-1}, \tag{3.71}$$

i.e., (3.69). Now observe that due to Assumption (H1)-Sto, for every $x_{t-1} \in \mathcal{X}_{t-1}$, the optimization problem

$$\underline{\mathfrak{Q}}^{k-1}_{tj}(x_{t-1}) = \left\{ \begin{array}{l} \min_{x_t} f^{k-1}_{tj}(x_t, x_{t-1}) + \mathcal{Q}^{k-1}_{t+1}(x_t) \\ x_t \in X^{k-1}_{tj}(x_{t-1}), \end{array} \right.$$

is a linear program with feasible set that is bounded (since $\mathcal{X}_t$ is compact) and nonempty (it contains the nonempty set $X_t(x_{t-1})$). Therefore it has a finite optimal value which is also the optimal value of the dual problem given by

$$\underline{\mathfrak{Q}}_{tj}^{k-1}(x_{t-1}) = \left\{ \begin{array}{l} \max\limits_{\alpha,\mu,\delta,\nu,\lambda} \ \mathcal{D}_{tj}^{k-1}(\alpha,\mu,\delta,\nu,\lambda;x_{t-1}) \\ (A_{tj}^{k-1})^\top \alpha + (D_{tj}^{k-1})^\top \mu + (\beta_{t+1}^{0:k-1})^\top \delta - \mathbb{X}_t^\top \nu - (A_{tj})^\top \lambda = 0, \\ \mathbf{e}^\top \alpha = 1,\ \mathbf{e}^\top \delta = 1, \alpha,\mu,\delta,\nu \geq 0, \end{array}\right. \tag{3.72}$$

where

$$\mathcal{D}_{tj}^{k-1}(\alpha,\mu,\delta,\nu,\lambda;x_{t-1}) = \alpha^\top (B_{tj}^{k-1} x_{t-1} + C_{tj}^{k-1}) + \mu^\top (E_{tj}^{k-1} x_{t-1} + H_{tj}^{k-1}) + \delta^\top \theta_{t+1}^{0:k-1} + \lambda^\top (b_{tj} - B_{tj} x_{t-1}) + \nu^\top \bar{x}_t.$$

Now assume that $t \geq 2$. Let us take $m \in C(n_{t-1}^k)$. Recall that $j_t(m)$ is the unique index j such that $\xi_{tj} = \xi_m$. Clearly $(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k)$ is feasible for dual problem (3.72) written for $j = j_t(m)$ and therefore for any $x_{t-1} \in \mathcal{X}_{t-1}$ we have

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_{t-1}) \geq \mathcal{D}_{tj_t(m)}^{k-1}(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k; x_{t-1}), \tag{3.73}$$

which gives

$$\begin{array}{lcl} \mathcal{Q}_t(x_{t-1}) & = & \displaystyle\sum_{j=1}^{M_t} p_{tj} \mathfrak{Q}_t(x_{t-1}, \xi_{tj}) \\ & \stackrel{(H0)}{=} & \displaystyle\sum_{m \in C(n_{t-1}^k)} p_m \mathfrak{Q}_t(x_{t-1}, \xi_m) \\ & = & \displaystyle\sum_{m \in C(n_{t-1}^k)} p_m \mathfrak{Q}_t(x_{t-1}, \xi_{tj_t(m)}) \\ & \stackrel{(3.71)}{\geq} & \displaystyle\sum_{m \in C(n_{t-1}^k)} p_m \underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_{t-1}) \\ & \stackrel{(3.73)}{\geq} & \displaystyle\sum_{m \in C(n_{t-1}^k)} p_m \mathcal{D}_{tj_t(m)}^{k-1}(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k; x_{t-1}) \\ & = & \mathcal{C}_t^k(x_{t-1}), \end{array}$$

for every $x_{t-1} \in \mathcal{X}_{t-1}$, where for the last equality, we have used (3.52) and (3.57). Therefore we have shown (3.68).

Now take $n \in \texttt{Nodes}(t-1)$ and $k \in \mathcal{S}_n$. Then by definition of $(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k)$ and of $\mathcal{C}_t^k$, we get for any $m \in C(n)$:

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k) = \mathcal{D}_{tj_t(m)}^{k-1}(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k; x_n^k) \tag{3.74}$$

and

$$\mathcal{C}_t^k(x_n^k) = \sum_{m \in C(n)} p_m \mathcal{D}_{tj_t(m)}^{k-1}(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k; x_n^k) = \sum_{m \in C(n)} p_m \underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k) = \underline{\mathcal{Q}}_t^{k-1}(x_n^k). \tag{3.75}$$

■

To prove the convergence of StoDCuP, we need the following assumption:

(H2) The samples of (ξ_t) generated in StoDCuP are independent: $(\tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ is a realization of $\xi^k = (\xi_2^k, \ldots, \xi_T^k) \sim (\xi_2, \ldots, \xi_T)$ and $\xi^k, k \geq 1$, are independent.

Theorem 3.8 (Convergence of StoDCuP)**.** *Let Assumption (H0), (H1)-Sto, and (H2) hold. Then*

(i) for every $t = 1, \ldots, T$, $i = 1, \ldots, p$, almost surely

$$\lim_{k \to +\infty} \max(g_{ti}(x_m^k, x_n^k, \xi_m), 0) = 0,\ \forall m \in \texttt{Nodes}(t), n = \mathcal{P}(m). \tag{3.76}$$

For all $t = 2, \ldots, T+1$, for all node $n \in \texttt{Nodes}(t-1)$, we have almost surely

$$\mathcal{H}(t) : \lim_{k \to +\infty} \mathcal{Q}_t(x_n^k) - \mathcal{Q}_t^k(x_n^k) = 0. \tag{3.77}$$

(ii) The limit of the sequence of first stage problems optimal values $(f_{11}^{k-1}(x_{n_1}^k, x_0) + \mathcal{Q}_2^{k-1}(x_{n_1}^k))_{k \geq 1}$ is the optimal value $\mathcal{Q}_1(x_0)$ of (3.50) and any accumulation point of the sequence $(x_{n_1}^k)$ is an optimal solution to the first stage problem (3.50).

Proof: We first show (3.76). Let us fix $t \in \{1, \ldots, T\}$, $i \in \{1, \ldots, p\}$, $m \in \texttt{Nodes}(t)$, $n = \mathcal{P}(m)$. Recall from Lemma 3.6 that

$$0 \leq \max(g_{ti}(x_m^k, x_n^k, \xi_m), 0) \leq g_{ti}(x_m^k, x_n^k, \xi_m) - g_{tij_t(m)}^{k-1}(x_m^k, x_n^k). \tag{3.78}$$

We now show that

$$\lim_{k \to +\infty} g_{ti}(x_m^k, x_n^k, \xi_m) - g_{tij_t(m)}^{k-1}(x_m^k, x_n^k) = 0, \tag{3.79}$$

which will show (3.76) due to relation (3.78).

Let $k(1), k(2), \ldots,$ be the iterations in $\mathcal{S}_n$ with $k(i) < k(i+1)$: $\mathcal{S}_n = \{k(1), k(2), k(3), \ldots\}$. Let us first show that we have

$$\lim_{k \to +\infty, k \in \mathcal{S}_n} \max(g_{ti}(x_m^k, x_n^k, \xi_m), 0) = 0. \tag{3.80}$$

For all $\ell \geq 1$, relation (3.62) gives

$$g_{ti}(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) = g_{tij_t(m)}^{k(\ell)}(x_m^{k(\ell)}, x_n^{k(\ell)}). \tag{3.81}$$

Let us now apply Lemma 2.5 to $y^\ell = (x_m^{k(\ell)}, x_n^{k(\ell)})$, sequence $f^\ell = g_{tij_t(m)}^{k(\ell)}$, and $f = g_{ti}(\cdot, \cdot, \xi_m)$ (observe that the assumptions of the lemma are satisfied with $k_0 = 1$). Since

$$\lim_{\ell \to +\infty} f(y^\ell) - f^\ell(y^\ell) = 0,$$

we deduce that

$$\lim_{\ell \to +\infty} f(y^\ell) - f^{\ell-1}(y^\ell) = \lim_{\ell \to +\infty} g_{ti}(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) - g_{tij_t(m)}^{k(\ell-1)}(x_m^{k(\ell)}, x_n^{k(\ell)}) = 0. \tag{3.82}$$

Since $k(\ell) \geq 1 + k(\ell-1)$, we have $0 \leq g_{ti}(\cdot, \cdot, \xi_m) - g_{tij_t(m)}^{k(\ell)-1}(\cdot, \cdot) \leq g_{ti}(\cdot, \cdot, \xi_m) - g_{tij_t(m)}^{k(\ell-1)}(\cdot, \cdot)$ and therefore (3.82) implies

$$\lim_{\ell \to +\infty} g_{ti}(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) - g_{tij_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) = \lim_{k \to +\infty, k \in \mathcal{S}_n} g_{ti}(x_m^k, x_n^k, \xi_m) - g_{tij_t(m)}^{k-1}(x_m^k, x_n^k) = 0. \tag{3.83}$$

Finally, we show in the Appendix that

$$\lim_{k \to +\infty, k \notin \mathcal{S}_n} g_{ti}(x_m^k, x_n^k) - g_{tij_t(m)}^{k-1}(x_m^k, x_n^k) = 0, \tag{3.84}$$

which achieves the proof of (3.79) and therefore of (3.76).

Let us now show $\mathcal{H}(t)$ by backward induction on t. $\mathcal{H}(T+1)$ holds since $\mathcal{Q}_{T+1} = \mathcal{Q}_{T+1}^k$. Assume now that $\mathcal{H}(t+1)$ holds for some $t \in \{2, \ldots, T\}$ and let us show that $\mathcal{H}(t)$ holds. Take a node $n \in \texttt{Nodes}(t-1)$ and let us denote again by $k(1), k(2), \ldots,$ the iterations in $\mathcal{S}_n$ with $k(i) < k(i+1)$: $\mathcal{S}_n = \{k(1), k(2), k(3), \ldots\}$. Let us first show that

$$\lim_{k \to +\infty, k \in \mathcal{S}_n} \mathcal{Q}_t(x_n^k) - \mathcal{Q}_t^k(x_n^k) = \lim_{\ell \to +\infty} \mathcal{Q}_t(x_n^{k(\ell)}) - \mathcal{Q}_t^{k(\ell)}(x_n^{k(\ell)}) = 0. \tag{3.85}$$

By definition of $\mathcal{Q}_t^{k(\ell)}$, we have $\mathcal{Q}_t^{k(\ell)}(x_n^{k(\ell)}) \geq \mathcal{C}_t^{k(\ell)}(x_n^{k(\ell)})$ and therefore for all $\ell \geq 1$ we get:

$$\begin{array}{rcl} 0 \leq \mathcal{Q}_t(x_n^{k(\ell)}) - \mathcal{Q}_t^{k(\ell)}(x_n^{k(\ell)}) & \leq & \mathcal{Q}_t(x_n^{k(\ell)}) - \mathcal{C}_t^{k(\ell)}(x_n^{k(\ell)}) \\ & = & \mathcal{Q}_t(x_n^{k(\ell)}) - \underline{\mathcal{Q}}_t^{k(\ell)-1}(x_n^{k(\ell)}), \\ & = & \displaystyle\sum_{m \in C(n_{t-1}^{k(\ell)})} p_m \Big[\mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) - \underline{\mathfrak{Q}}_{tj_t(m)}^{k(\ell)-1}(x_n^{k(\ell)}) \Big]. \end{array} \tag{3.86}$$

By definiton of x_m^k, we have

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k(\ell)-1}(x_n^{k(\ell)}) = f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) + \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}), \tag{3.87}$$

which, plugged into (3.86), gives

$$0 \le \mathcal{Q}_t(x_n^{k(\ell)}) - \mathcal{Q}_t^{k(\ell)}(x_n^{k(\ell)}) \le \sum_{m \in C(n_{t-1}^{k(\ell)})} p_m \Big[\mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) - f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) - \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) \Big]. \quad (3.88)$$

Let us apply Lemma 2.5 to $y^\ell = (x_m^{k(\ell)}, x_n^{k(\ell)})$, sequence $f^\ell = f_{tj_t(m)}^{k(\ell)}$, and $f = f_t(\cdot,\cdot,\xi_m)$ (observe that the assumptions of the lemma are satisfied). Due to (3.62), we have

$$\lim_{\ell \to +\infty} f(y^\ell) - f^\ell(y^\ell) = 0$$

and therefore

$$\lim_{\ell \to +\infty} f(y^\ell) - f^{\ell-1}(y^\ell) = \lim_{\ell \to +\infty} f_t(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) - f_{tj_t(m)}^{k(\ell-1)}(x_m^{k(\ell)}, x_n^{k(\ell)}) = 0. \quad (3.89)$$

Since $k(\ell) \ge k(\ell-1)+1$, we have $0 \le f_t(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) - f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) \le f_t(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) - f_{tj_t(m)}^{k(\ell-1)}(x_m^{k(\ell)}, x_n^{k(\ell)})$ which combined with (3.89) gives

$$\lim_{\ell \to +\infty} f_t(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) - f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) = 0. \quad (3.90)$$

Using (3.87) and (3.69), we get

$$f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) + \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) = \underline{\mathfrak{Q}}_{tj_t(m)}^{k(\ell)-1}(x_n^{k(\ell)}) \le \mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m).$$

Therefore the sequence $(f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) + \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) - \mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m))_{\ell \ge 1}$ is bounded and has a finite limit sup which satisfies

$$\varlimsup_{\ell \to +\infty} f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) + \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) - \mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) \le 0. \quad (3.91)$$

Applying Lemma 2.5 to $y^\ell = x_m^{k(\ell)}$, sequence $f^\ell = \mathcal{Q}_{t+1}^{k(\ell)}$, and $f = \mathcal{Q}_{t+1}$ (observe that the assumptions of the lemma are satisfied), since from the induction hypothesis we know that

$$\lim_{\ell \to +\infty} f(y^\ell) - f^\ell(y^\ell) = 0$$

we deduce that

$$\lim_{\ell \to +\infty} f(y^\ell) - f^{\ell-1}(y^\ell) = \lim_{\ell \to +\infty} \mathcal{Q}_{t+1}(x_m^{k(\ell)}) - \mathcal{Q}_{t+1}^{k(\ell-1)}(x_m^{k(\ell)}) = 0. \quad (3.92)$$

Since $k(\ell) \ge k(\ell-1)+1$, we have $0 \le \mathcal{Q}_{t+1}(x_m^{k(\ell)}) - \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) \le \mathcal{Q}_{t+1}(x_m^{k(\ell)}) - \mathcal{Q}_{t+1}^{k(\ell-1)}(x_m^{k(\ell)})$, which combines with (3.92) to give

$$\lim_{\ell \to +\infty} \mathcal{Q}_{t+1}(x_m^{k(\ell)}) - \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) = 0. \quad (3.93)$$

Combining (3.90), (3.91), and (3.93), we obtain

$$\varlimsup_{\ell \to +\infty} f_t(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) + \mathcal{Q}_{t+1}(x_m^{k(\ell)}) - \mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) \le 0. \quad (3.94)$$

Let us now show by contradiction that

$$\varliminf_{k \to +\infty} f_t(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) + \mathcal{Q}_{t+1}(x_n^{k(\ell)}) - \mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) \ge 0. \quad (3.95)$$

Assume that (3.95) does not hold. Using the fact that sequence $(x_m^k, x_n^k)_{k \in \mathcal{S}_n}$ belongs to the compact set $\mathcal{X}_t \times \mathcal{X}_{t-1}$, and the lower semicontinuity of $f_t(\cdot,\cdot,\xi_m)$, $g_t(\cdot,\cdot,\xi_m)$, $\mathcal{Q}_t$, $\mathfrak{Q}_t(\cdot,\xi_m)$, there is a subsequence $(x_m^k, x_n^k)_{k \in K}$ with $K \subset \mathcal{S}_n$ converging to some $(\bar{x}_m, \bar{x}_n) \in \mathcal{X}_t \times \mathcal{X}_{t-1}$ such that

$$f_t(\bar{x}_m, \bar{x}_n, \xi_m) + \mathcal{Q}_{t+1}(\bar{x}_n) - \mathfrak{Q}_t(\bar{x}_n, \xi_m) < 0$$

and $\bar{x}_m \in X_t(\bar{x}_n, \xi_m)$. This is in contradiction with the definition of $\mathfrak{Q}_t$. Therefore we must have

$$\begin{array}{rcl} 0 & = & \lim_{\ell \to +\infty} f_t(x_m^{k(\ell)}, x_n^{k(\ell)}, \xi_m) + \mathcal{Q}_{t+1}(x_m^{k(\ell)}) - \mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) \\ & = & \lim_{\ell \to +\infty} f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) + \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) - \mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) \end{array}$$

which, plugged into (3.88) gives

$$\lim_{k\to+\infty, k\in\mathcal{S}_n} \mathcal{Q}_t(x_n^k) - \mathcal{Q}_t^k(x_n^k) = 0. \tag{3.96}$$

Finally, we show in the Appendix that

$$\lim_{k\to+\infty, k\notin\mathcal{S}_n} \mathcal{Q}_t(x_n^k) - \mathcal{Q}_t^k(x_n^k) = 0, \tag{3.97}$$

which achieves the proof of $\mathcal{H}(t)$.

(ii) The proof of (ii) can easily be obtained from (i), see Theorem 4.1-(ii) in [3] for details. ∎

Remark 3.9 (Stopping criterion)**.** *The stopping criterion is similar to SDDP. We can stop the algorithm when the gap $\frac{Ub-Lb}{Ub}$ is less than a threshold, for instance 5%, where Ub and Lb are upper and lower bounds, respectively, defined as follows. Due to Lemma 3.7, we can take as a lower bound on the optimal value of problem (3.49) the value $Lb = \underline{\mathfrak{Q}}_{11}^{k-1}(x_0)$. The upper bound Ub corresponds to the upper end of a 100(1-α)%-one-sided confidence interval (with for instance $\alpha = 0.05$) on the optimal value for N policy realizations (using the costs of decisions taken on N independent sampled scenarios).*

4. Variants of StoDCuP

4.1. **Forward-backward StoDCuP.** Similarly to DCuP, we can extend forward StoDCuP presented in the previous section to forward-backward StoDCuP. In this variant, at iteration k, we still compute in a forward pass trial points x_n^k for nodes $n \in \{n_1, n_2^k, \ldots, n_T^k\}$ with n_t^k a child node of node n_{t-1}^k. For each $t = 1, \ldots, T$, and $j = 1, \ldots, M_t$, $i = 1, \ldots, p$, a linearization is also computed in the forward pass for $f_t(\cdot,\cdot,\xi_{tj})$ and $g_{ti}(\cdot,\cdot,\xi_{tj})$ at these trial points. However, cuts for $\mathcal{Q}_t$ are computed in a backward pass and in this backward pass an additional linearization is also built for $f_t(\cdot,\cdot,\xi_{tj})$ and $g_{ti}(\cdot,\cdot,\xi_{tj})$ using points computed in both the backward and forward pass. For the cut computed at iteration k for $\mathcal{Q}_t$, we will still use the notation:

$$\mathcal{C}_t^k(x_{t-1}) = \theta_t^k + \langle \beta_t^k, x_{t-1}\rangle$$

with the convention that $\mathcal{C}_{T+1}^k$ is the null function (see below for the computation of θ_t^k, β_t^k). We end up iteration k with approximation $\mathcal{Q}_t^k(x_{t-1}) = \max_{0\le j\le k} \mathcal{C}_t^j(x_{t-1})$ of $\mathcal{Q}_t$.

Therefore, at iteration k, two approximations of functions $f_t(\cdot,\cdot,\xi_{tj})$ and $g_{ti}(\cdot,\cdot,\xi_{tj})$ are computed which will be denoted by f_{tj}^{2k-1} and g_{tij}^{2k-1}, respectively, in the end of the forward pass, and by f_{tj}^{2k} and g_{tj}^{2k}, respectively, in the end of the backward pass.

The detailed steps of forward-backward StoDCuP are described below.

Forward-Backward StoDCuP (Stochastic Dynamic Cutting Plane) with linearizations computed in forward and backward passes.

Step 1) **Initialization.** For $t = 1, \ldots, T$, take $f_{tj}^0, g_{tij}^0 : \mathcal{X}_t \times \mathcal{X}_{t-1} \to \mathbb{R}$ affine functions satisfying $f_{tj}^0 \le f_t(\cdot,\cdot,\xi_{tj})$, $g_{tij}^0 \le g_{ti}(\cdot,\cdot,\xi_{tj})$, and for $t = 2, \ldots, T$, $\mathcal{Q}_t^0 : \mathcal{X}_{t-1} \to \mathbb{R}$ is an affine function satisfying $\mathcal{Q}_t^0 \le \mathcal{Q}_t$. Set the iteration count k to 1 and $\mathcal{Q}_{T+1}^0 \equiv 0$.

Step 2) **Forward pass.**
Generate a sample $(\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$.
For $t = 1, \ldots, T$,
  For $j = 1, \ldots, M_t$,
    If $\xi_{tj} = \tilde{\xi}_t^k$ **then** compute an optimal solution x_t^k of

$$\underline{\mathfrak{Q}}_{tj}^{2k-2}(x_{t-1}^k) = \left\{ \begin{array}{l} \inf_{x_t} f_{tj}^{2k-2}(x_t, x_{t-1}^k) + \mathcal{Q}_{t+1}^{k-1}(x_t) \\ x_t \in X_{tj}^{2k-2}(x_{t-1}^k), \end{array}\right. \tag{4.98}$$

where $x_0^k = x_0$ and where for all $k \ge 1$,

$$X_{tj}^{2k-2}(x_{t-1}^k) = \{x_t \in \mathcal{X}_t : \; g_{tij}^{2k-2}(x_t, x_{t-1}^k) \le 0, i = 1, \ldots, p, \; A_{tj}x_t + B_{tj}x_{t-1}^k = b_{tj}\}. \tag{4.99}$$

Compute $f_t(x_t^k, x_{t-1}^k, \xi_{tj})$, $g_{ti}(x_t^k, x_{t-1}^k, \xi_{tj})$, and subgradients of $f_t(\cdot,\cdot,\xi_{tj})$, $g_{ti}(\cdot,\cdot,\xi_{tj})$ at (x_t^k, x_{t-1}^k) with corresponding linearizations $\ell_{f_t(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_t^k, x_{t-1}^k))$ and $\ell_{g_{ti}(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_t^k, x_{t-1}^k))$. Compute

$$\begin{array}{l} f_{tj}^{2k-1}(\cdot,\cdot) \leftarrow \max\Big(f_{tj}^{2k-2}(\cdot,\cdot), \ell_{f_t(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_t^k, x_{t-1}^k))\Big),\\ g_{tij}^{2k-1}(\cdot,\cdot) \leftarrow \max\Big(g_{tij}^{2k-2}(\cdot,\cdot), \ell_{g_{ti}(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_t^k, x_{t-1}^k))\Big).\end{array}$$

Else

$$f_{tj}^{2k-1} = f_{tj}^{2k-2},\ g_{tij}^{2k-1} = g_{tij}^{2k-2}.$$

End If

End For

End For

Step 3) **Backward pass.**

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

For $t = T, \ldots, 2$,

For $j = 1, \ldots, M_t$,

Compute an optimal solution x_{tj}^{Bk} of

$$\underline{\mathfrak{Q}}_{tj}^{2k-1}(x_{t-1}^k) = \left\{ \begin{array}{l} \inf\limits_{x_t} \ f_{tj}^{2k-1}(x_t, x_{t-1}^k) + \mathcal{Q}_{t+1}^k(x_t) \\ x_t \in X_{tj}^{2k-1}(x_{t-1}^k). \end{array}\right. \tag{4.100}$$

Compute $f_t(x_{tj}^{Bk}, x_{t-1}^k, \xi_{tj})$, $g_{ti}(x_{tj}^{Bk}, x_{t-1}^k, \xi_{tj})$ and subgradients of $f_t(\cdot,\cdot,\xi_{tj})$ and $g_{ti}(\cdot,\cdot,\xi_{tj})$ at (x_{tj}^{Bk}, x_{t-1}^k) with corresponding linearizations $\ell_{f_t(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_{tj}^{Bk}, x_{t-1}^k))$ and $\ell_{g_{ti}(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_{tj}^{Bk}, x_{t-1}^k))$.
Compute

$$\begin{array}{l} f_{tj}^{2k}(\cdot,\cdot) \leftarrow \max\Big(f_{tj}^{2k-1}(\cdot,\cdot), \ell_{f_t(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_{tj}^{Bk}, x_{t-1}^k))\Big),\\ g_{tij}^{2k}(\cdot,\cdot) \leftarrow \max\Big(g_{tij}^{2k-1}(\cdot,\cdot), \ell_{g_{ti}(\cdot,\cdot,\xi_{tj})}(\cdot,\cdot;(x_{tj}^{Bk}, x_{t-1}^k))\Big).\end{array}$$

Compute (for instance using Lemma 2.1 in [3]) a subgradient β_t^{kj} of $\underline{\mathfrak{Q}}_{tj}^{2k-1}$ at x_{t-1}^k and the cut coefficients:

$$\theta_t^k \ = \ \sum_{j=1}^{M_t} p_{tj}(\underline{\mathfrak{Q}}_{tj}^{2k-1}(x_{t-1}^k) - \langle \beta_t^{kj}, x_{t-1}^k\rangle) \text{ and } \beta_t^k = \sum_{j=1}^{M_t} p_{tj}\beta_t^{kj}.$$

End For

End For

Compute an optimal solution x_1^{Bk} of

$$\left\{ \begin{array}{l} \inf\limits_{x_1} \ f_{11}^{2k-1}(x_1, x_0) + \mathcal{Q}_2^k(x_1) \\ x_1 \in X_{11}^{2k-1}(x_0). \end{array}\right. \tag{4.101}$$

Compute $f_1(x_1^{Bk}, x_0, \xi_1)$, $g_{1i}(x_1^{Bk}, x_0, \xi_1)$, and subgradients of $f_1(\cdot,\cdot,\xi_1)$, $g_{1i}(\cdot,\cdot,\xi_1)$ at (x_1^{Bk}, x_0) with corresponding linearizations $\ell_{f_1(\cdot,\cdot,\xi_1)}(\cdot,\cdot;(x_1^{Bk}, x_0))$ and $\ell_{g_{1i}(\cdot,\cdot,\xi_1)}(\cdot,\cdot;(x_1^{Bk}, x_0))$.
Compute

$$\begin{array}{l} f_{11}^{2k}(\cdot,\cdot) \leftarrow \max\Big(f_{11}^{2k-1}(\cdot,\cdot), \ell_{f_1(\cdot,\cdot,\xi_1)}(\cdot,\cdot;(x_1^{Bk}, x_0))\Big),\\ g_{1i1}^{2k}(\cdot,\cdot) \leftarrow \max\Big(g_{1i1}^{2k-1}(\cdot,\cdot), \ell_{g_{1i}(\cdot,\cdot,\xi_1)}(\cdot,\cdot;(x_1^{Bk}, x_0))\Big).\end{array}$$

Step 4) Do $k \leftarrow k+1$ and go to Step 2).

4.2. **Inexact cuts in StoDCuP.** In this section, we present an extension of StoDCuP to solve problem (3.49). Since all subproblems of forward StoDCuP presented in Section 3 are linear programs, it is easy to derive an inexact variant of StoDCuP that computes ε_t^k-optimal solutions (instead of optimal solutions in StoDCuP) of the subproblems solved for iteration k and stage t. We show in Lemma 4.1 below that the cuts computed by this variant are still valid and that the distance between the cuts and $\underline{\mathcal{Q}}_t^{k-1}(\cdot) = \sum_{j=1}^{M_t} p_{tj} \underline{\mathfrak{Q}}_{tj}^{k-1}(\cdot)$ at the trial point x_n^k for stage t and iteration k is at most ε_t^k. This variant of StoDCuP, called inexact StoDCuP, is given below and the convergence of the method is proved in Theorem 4.3:

Inexact StoDCuP.

Step 1) **Initialization.** For $t = 1, \ldots, T$, take $f_{tj}^0, g_{tij}^0 : \mathcal{X}_t \times \mathcal{X}_{t-1} \to \mathbb{R}$ affine functions satisfying $f_{tj}^0 \leq f_t(\cdot, \cdot, \xi_{tj})$, $g_{tij}^0 \leq g_{ti}(\cdot, \cdot, \xi_{tj})$, and for $t = 2, \ldots, T$, $\mathcal{Q}_t^0 : \mathcal{X}_{t-1} \to \mathbb{R}$ is an affine function satisfying $\mathcal{Q}_t^0 \leq \mathcal{Q}_t$. Set $x_{n_0} = x_0$, set the iteration count k to 1, and $\mathcal{Q}_{T+1}^0 \equiv 0$.

Step 2) Generate a sample $(\tilde{\xi}_1^k, \tilde{\xi}_2^k, \ldots, \tilde{\xi}_T^k)$ of $(\xi_1, \xi_2, \ldots, \xi_T)$ corresponding to a set of nodes $(n_1^k, n_2^k, \ldots, n_T^k)$ where $n_1^k = n_1$, and for $t \geq 2$, n_t^k is a node of stage t, child of node n_{t-1}^k. Set $n_0^k = n_0$.
Do $\theta_{T+1}^k = 0$ and $\beta_{T+1}^k = 0$.
For $t = 1, \ldots, T$,
Let $n = n_{t-1}^k$.
For every $m \in C(n)$,
compute an ε_t^k-optimal feasible solution x_m^k of

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k) = \left\{ \begin{array}{l} \min\limits_{x_m} \; f_{tj_t(m)}^{k-1}(x_m, x_n^k) + \mathcal{Q}_{t+1}^{k-1}(x_m) \\ x_m \in X_{tj_t(m)}^{k-1}(x_n^k). \end{array} \right. \tag{4.102}$$

Compute an arbitrary subgradient $[s_1; s_2]$ of convex function $f_t(\cdot, \cdot, \xi_m)$ at (x_m^k, x_n^k) where $s_1, s_2 \in \mathbb{R}^n$ and do $a_{tj_t(m)}^k = s_1^\top$, $b_{tj_t(m)}^k = s_2^\top$. For $i = 1, \ldots, p$, compute an arbitrary subgradient $[s_{1i}; s_{2i}]$ of convex function $g_{ti}(\cdot, \cdot, \xi_m)$ at (x_m^k, x_n^k) where $s_{1i}, s_{2i} \in \mathbb{R}^n$ and do $d_{tij_t(m)}^k = s_{1i}^\top$, $e_{tij_t(m)}^k = s_{2i}^\top$. Compute

$$\begin{array}{l} c_{tj_t(m)}^k = f_t(x_m^k, x_n^k, \xi_m) - a_{tj_t(m)}^k x_m^k - b_{tj_t(m)}^k x_n^k, \\ h_{tij_t(m)}^k = g_{ti}(x_m^k, x_n^k, \xi_m) - d_{tij_t(m)}^k x_m^k - e_{tij_t(m)}^k x_n^k. \end{array}$$

Compute an ε_t^k-optimal feasible solution $(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k)$ of the dual problem

$$\begin{array}{l} \max\limits_{\alpha,\mu,\delta,\nu,\lambda} \; \alpha^\top (B_{tj_t(m)}^{k-1} x_n^k + C_{tj_t(m)}^{k-1}) + \mu^\top (E_{tj_t(m)}^{k-1} x_n^k + H_{tj_t(m)}^{k-1}) + \delta^\top \theta_{t+1}^{0:k-1} + \lambda^\top (b_{tj_t(m)} - B_{tj_t(m)} x_n^k) + \nu^\top \bar{x}_t \\ (A_{tj_t(m)}^{k-1})^\top \alpha + (D_{tj_t(m)}^{k-1})^\top \mu + (\beta_{t+1}^{0:k-1})^\top \delta - \mathbb{X}_t^\top \nu - (A_{tj_t(m)})^\top \lambda = 0, \\ \mathbf{e}^\top \alpha = 1, \; \mathbf{e}^\top \delta = 1, \alpha, \mu, \delta, \nu \geq 0. \end{array}$$

End For
If $t \geq 2$ compute:

$$\begin{array}{lcl} \beta_t^k & = & \sum\limits_{m \in C(n)} p_m \Big[(B_{tj_t(m)}^k)^\top \alpha_m^{k-1} + (E_{tj_t(m)}^{k-1})^\top \mu_m^k - B_{tj_t(m)}^\top \lambda_m^k \Big], \\ \theta_t^k & = & \sum\limits_{m \in C(n)} p_m \Big[\langle \alpha_m^k, C_{tj_t(m)}^{k-1} \rangle + \langle \mu_m^k, H_{tj_t(m)}^{k-1} \rangle + \langle \delta_m^k, \theta_{t+1}^{0:k-1} \rangle + \langle \lambda_m^k, b_{tj_t(m)} \rangle + \langle \nu_m^k, \bar{x}_t \rangle \Big]. \end{array} \tag{4.103}$$

End If
End For

Step 4) Do $k \leftarrow k + 1$ and go to Step 2).

Clearly Lemma 3.6 still holds for Inexact StoDCuP. The quality of the cuts computed for $\mathcal{Q}_t$ by Inexact StoDCuP is given in Lemma 4.1:

Lemma 4.1 (Validity and quality of cuts computed by Inexact StoDCuP). *Let Assumptions (H0) and (H1)-Sto hold. For every $t = 2, \ldots, T+1$, for every $k \geq 1$, we have*

$$\mathcal{Q}_t(x_{t-1}) \geq \mathcal{C}_t^k(x_{t-1}) \text{ and } \mathcal{Q}_t(x_{t-1}) \geq \mathcal{Q}_t^k(x_{t-1}),\ \forall x_{t-1} \in \mathcal{X}_{t-1}. \tag{4.104}$$

For all $t = 1, \ldots, T$, $j = 1, \ldots, M_t$, for every $k \geq 1$, we have

$$\underline{\mathfrak{Q}}_{tj}^k(x_{t-1}) \leq \mathfrak{Q}_t(x_{t-1}, \xi_{tj}) \text{ for all } x_{t-1} \in \mathcal{X}_{t-1}. \tag{4.105}$$

For all $t = 2, \ldots, T$, for every $k \geq 1$, defining $\underline{\mathcal{Q}}_t^{k-1}(x_n^k) = \sum_{j=1}^{M_t} p_{tj} \underline{\mathfrak{Q}}_{tj}^{k-1}(x_n^k)$, we have for every $n \in$ $\texttt{Nodes}(t-1)$ and for all $k \in \mathcal{S}_n$:

$$0 \leq \underline{\mathcal{Q}}_t^{k-1}(x_n^k) - \mathcal{C}_t^k(x_n^k) \leq \varepsilon_t^k. \tag{4.106}$$

Proof: The proofs of (3.68) and (3.69) in Lemma 3.7 can be used to prove (4.104) and (4.105) for Inexact StoDCuP, observing that only feasibility and not optimality of the primal and dual solutions computed as well as Lemma 3.6 (which, as we have already observed, holds) are needed in these proofs.

Now take $n \in \texttt{Nodes}(t-1)$ and $k \in \mathcal{S}_n$. Then recalling that

$$\mathcal{D}_{tj}^{k-1}(\alpha, \mu, \delta, \nu, \lambda; x_{t-1}) = \alpha^\top (B_{tj}^{k-1} x_{t-1} + C_{tj}^{k-1}) + \mu^\top (E_{tj}^{k-1} x_{t-1} + H_{tj}^{k-1}) + \delta^\top \theta_{t+1}^{0:k-1} + \lambda^\top (b_{tj} - B_{tj} x_{t-1}) + \nu^\top \bar{x}_t,$$

by definition of $(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k)$ and of $\mathcal{C}_t^k$, we get

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k) - \varepsilon_t^k \leq \mathcal{D}_{tj_t(m)}^{k-1}(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k; x_n^k) \leq \underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k) \tag{4.107}$$

and

$$\mathcal{C}_t^k(x_n^k) = \sum_{m \in C(n)} p_m \mathcal{D}_{tj_t(m)}^{k-1}(\alpha_m^k, \mu_m^k, \delta_m^k, \nu_m^k, \lambda_m^k; x_n^k). \tag{4.108}$$

Since $\underline{\mathcal{Q}}_t^{k-1}(x_n^k) = \sum_{m \in C(n_{t-1}^k)} p_m \underline{\mathfrak{Q}}_{tj_t(m)}^{k-1}(x_n^k)$, $p_m \geq 0$, and $\sum_{m \in C(n)} p_m = 1$, relations (4.107) and (4.108) imply (4.106). ■

Lemma 4.2 below is the analogue of Lemma 3.3:

Lemma 4.2. *Let Assumptions (H0) and (H1)-Sto hold and assume that sequences ε_t^k are bounded: $|\varepsilon_t^k| \leq \hat{\varepsilon}$ for all t, k, for some $0 \leq \hat{\varepsilon} < +\infty$. Then, the following statements hold for Inexact StoDCuP:*

(a) *For $t = 2, \ldots, T$, the sequences $\{\theta_t^k\}_{k=1}^\infty$ and $\{\beta_t^k\}_{k=1}^\infty$ are almost surely bounded.*
(b) *There exists $L \geq 0$ such that for each $t = 2, \ldots, T$, $\mathcal{Q}_t^k$ is L-Lipschitz continuous on $\mathcal{X}_{t-1}$ for every $k \geq 1$.*
(c) *There exists $\hat{L} \geq 0$ such that for each $t = 1, \ldots, T$, $j = 1, \ldots, M_t$, functions f_{tj}^k and g_{tij}^k are $\hat{L}$-Lipschitz continuous on $\mathcal{X}_t \times \mathcal{X}_{t-1}$ for every $k \geq 1$ and $i = 1, \ldots, p$.*

Proof: (a) Using (H1)-Sto, there is $\varepsilon > 0$ such that for every $t \in \{2, \ldots, T\}$, every $x_{t-1} \in \mathcal{X}_{t-1} + \bar{B}(0; \varepsilon)$, and every $j = 1, \ldots, M_t$, the set $X_{tj}^0(x_{t-1})$ is nonempty and $f_{tj}^0(\cdot, x_{t-1}) + \mathcal{Q}_{t+1}^0(\cdot)$ is continuous on this set. Therefore $\underline{\mathfrak{Q}}_{tj}^0$ is convex and finite on $\mathcal{X}_{t-1} + \bar{B}(0; \varepsilon)$, implying that $\underline{\mathfrak{Q}}_{tj}^0$ is Lipschitz continuous on $\mathcal{X}_{t-1}$. It follows that $\underline{\mathcal{Q}}_t^0$ is also Lipschitz continuous on $\mathcal{X}_{t-1}$ and we can define $\min\limits_{x_{t-1} \in \mathcal{X}_{t-1}} \underline{\mathcal{Q}}_t^0(x_{t-1}) \in \mathbb{R}$. Similarly to DCuP, due to (H1)-Sto, we can also choose $\varepsilon > 0$ in such a way that $\mathcal{Q}_t$ is Lipschitz continuous on $\mathcal{X}_{t-1} + \bar{B}(0; \varepsilon)$, implying that we can define $\max_{x_{t-1} \in \mathcal{X}_{t-1} + \bar{B}(0;\varepsilon)} \mathcal{Q}_t(x_{t-1}) < +\infty$. We can now easily extend the proof of Lemma 3.3: for every $x_{t-1} \in \mathcal{X}_{t-1} + \bar{B}(0; \varepsilon)$, denoting $n = n_{t-1}^k$, we have for $k \geq 2$:

$$\begin{array}{rcl}
\max\limits_{x_{t-1} \in \mathcal{X}_{t-1} + \bar{B}(0;\varepsilon)} \mathcal{Q}_t(x_{t-1}) \geq \mathcal{Q}_t(x_{t-1}) & \overset{(4.104)}{\geq} & \mathcal{C}_t^k(x_{t-1}) \\
& = & \mathcal{C}_t^k(x_n^k) + \langle \beta_t^k, x_{t-1} - x_n^k \rangle \quad [\mathcal{C}_t^k \text{ is affine}], \\
& \overset{(4.106)}{\geq} & \underline{\mathcal{Q}}_t^{k-1}(x_n^k) - \varepsilon_t^k + \langle \beta_t^k, x_{t-1} - x_n^k \rangle, \\
& \geq & \min\limits_{x_{t-1} \in \mathcal{X}_{t-1}} \underline{\mathcal{Q}}_t^0(x_{t-1}) - \hat{\varepsilon} + \langle \beta_t^k, x_{t-1} - x_n^k \rangle.
\end{array}$$

For $\beta_t^k \neq 0$, take $x_{t-1} = x_n^k + \frac{\varepsilon}{2} \frac{\beta_t^k}{\|\beta_t^k\|}$ to obtain

$$\|\beta_t^k\| \leq L := \frac{2}{\varepsilon} \left(\hat{\varepsilon} + \max_{x_{t-1} \in \mathcal{X}_{t-1} + \bar{B}(0;\varepsilon)} \mathcal{Q}_t(x_{t-1}) - \min_{x_{t-1} \in \mathcal{X}_{t-1}} \underline{\mathcal{Q}}_t^0(x_{t-1}) \right).$$

Using (4.106), we also have for $n = n_{t-1}^k$:

$$-\hat{\varepsilon} + \min_{x_{t-1} \in \mathcal{X}_{t-1}} \underline{\mathcal{Q}}_t^0(x_{t-1}) \leq \theta_t^k = \mathcal{C}_t^k(x_n^k) \leq \max_{x_{t-1} \in \mathcal{X}_{t-1}} \mathcal{Q}_t(x_{t-1}).$$

(b) immediately follows from (a) and (c) from (H1)-Sto. ∎

Theorem 4.3 (Convergence of Inexact StoDCuP)**.** *Let Assumptions (H0), (H1)-Sto, and (H2) hold and assume that* $\lim_{k \to +\infty} \varepsilon_t^k = 0$ *for* $t = 1, \ldots, T$. *Then the conclusions of Theorem 3.8 hold: for every* $t = 1, \ldots, T, i = 1, \ldots, p$, *almost surely* (3.76) *and* (3.77) *hold and the limit of the sequence of first stage problems optimal values* $(f_{11}^{k-1}(x_{n_1}^k, x_0) + \mathcal{Q}_2^{k-1}(x_{n_1}^k))_{k \geq 1}$ *is the optimal value* $\mathcal{Q}_1(x_0)$ *of* (3.50) *and any accumulation point of the sequence* $(x_{n_1}^k)$ *is an optimal solution to the first stage problem* (3.50).

Proof: The proof is an adaptation of the proof of Theorem 3.8 and uses Lemmas 3.6, 4.1, and 4.2. We highlight these adaptations below.

Using Lemma 4.1, for Inexact StoDCuP relation (3.86) becomes

$$\begin{array}{lll} 0 \leq \mathcal{Q}_t(x_n^{k(\ell)}) - \mathcal{Q}_t^{k(\ell)}(x_n^{k(\ell)}) & \leq & \mathcal{Q}_t(x_n^{k(\ell)}) - \mathcal{C}_t^{k(\ell)}(x_n^{k(\ell)}) \\ & \overset{(3.70)}{\leq} & \varepsilon_t^{k(\ell)} + \mathcal{Q}_t(x_n^{k(\ell)}) - \underline{\mathcal{Q}}_t^{k(\ell)-1}(x_n^{k(\ell)}), \\ & = & \varepsilon_t^{k(\ell)} + \displaystyle\sum_{m \in C(n_{t-1}^{k(\ell)})} p_m \Big[\mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) - \underline{\mathfrak{Q}}_{tj_t(m)}^{k(\ell)-1}(x_n^{k(\ell)}) \Big]. \end{array} \tag{4.109}$$

Also, by definiton of x_m^k, we now have

$$\underline{\mathfrak{Q}}_{tj_t(m)}^{k(\ell)-1}(x_n^{k(\ell)}) \leq f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) + \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) \leq \underline{\mathfrak{Q}}_{tj_t(m)}^{k(\ell)-1}(x_n^{k(\ell)}) + \varepsilon_t^{k(\ell)}, \tag{4.110}$$

which, plugged into (4.109) gives

$$0 \leq \mathcal{Q}_t(x_n^{k(\ell)}) - \mathcal{Q}_t^{k(\ell)}(x_n^{k(\ell)}) \leq 2\varepsilon_t^{k(\ell)} + \sum_{m \in C(n_{t-1}^{k(\ell)})} p_m \Big[\mathfrak{Q}_t(x_n^{k(\ell)}, \xi_m) - f_{tj_t(m)}^{k(\ell)-1}(x_m^{k(\ell)}, x_n^{k(\ell)}) - \mathcal{Q}_{t+1}^{k(\ell)-1}(x_m^{k(\ell)}) \Big]. \tag{4.111}$$

The remaining relations and arguments used in the convergence proof of StoDCuP apply to prove the theorem. ∎

4.3. **Other variants.** It is also easy to incorporate in StoDCuP regularization as in [10], to apply multicut variants as in [7], [1], and cut selection strategies for the bundles of cuts of $\mathcal{Q}_t$, for instance along the lines of [14], [4], [7]. Observe, however, that all linearizations for $f_t(\cdot, \cdot, \xi_{tj})$ and $g_{ti}(\cdot, \cdot, \xi_{tj})$ are tight and therefore no cut selection is needed for these linearizations.

5. Conclusion

We introduced the exact and inexact StoDCuP (Stochastic Dynamic Cutting Plane) methods which are extensions of the SDDP method to solve MSPs. As a future work, it would be interesting to compare for several MSPs, for instance on real-life applications modelled by MSPs, the performances of SDDP, Inexact SDDP from [5], and StoDCuP and its variants presented in this paper.

Acknowledgments

The research of the first author was partially supported by an FGV grant, CNPq grants 401371/2014-0, 311289/2016-9, 204872/2018-9, and FAPERJ grant E-26/201.599/2014. Research of the second author was partially supported by CNPq grant 401371/2014-0.

References

[1] J.R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Oper. Res.*, 33:989–1007, 1985.

[2] 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.

[3] V. Guigues. Convergence analysis of sampling-based decomposition methods for risk-averse multistage stochastic convex programs. *SIAM Journal on Optimization*, 26:2468–2494, 2016.

[4] V. Guigues. Dual dynamic programing with cut selection: Convergence proof and numerical experiments. *European Journal of Operational Research*, 258:47–57, 2017.
[5] V. Guigues. Inexact cuts in Stochastic Dual Dynamic Programming. *Siam Journal on Optimization, to appear*, 2018.
[6] V. Guigues. Inexact Stochastic Mirror Descent for two-stage nonlinear stochastic programs. *arXiv*, 2018. `https://arxiv.org/pdf/1805.11732.pdf`.
[7] V. Guigues and M. Bandarra. Single cut and multicut SDDP with cut selection for multistage stochastic linear programs: convergence proof and numerical experiments. *arXiv*, 2019. `https://arxiv.org/abs/1902.06757`.
[8] V. Guigues and W. Römisch. Sampling-based decomposition methods for multistage stochastic programs based on extended polyhedral risk measures. *SIAM J. Optim.*, 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] V. Guigues, W. Tekaya, and M. Lejeune. Regularized decomposition methods for deterministic and stochastic convex optimization and application to portfolio selection with direct transaction and market impact costs. *Optimization OnLine*, 2017.
[11] R. P. Liu and A. Shapiro. Risk neutral reformulation approach to risk averse stochastc programming. *arXiv*, 2018. `https://arxiv.org/abs/1901.01302`.
[12] M.V.F. Pereira and L.M.V.G Pinto. Multi-stage stochastic optimization applied to energy planning. *Math. Program.*, 52:359–375, 1991.
[13] 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.
[14] A. Philpott, V. de Matos, and E. Finardi. Improving the performance of stochastic dual dynamic programming. *Journal of Computational and Applied Mathematics*, 290:196–208, 2012.
[15] A. B. Philpott and Z. Guan. On the convergence of stochastic dual dynamic programming and related methods. *Oper. Res. Lett.*, 36:450–455, 2008.
[16] W.P. Powell. *Approximate Dynamic Programming.* John Wiley and Sons, 2nd edition, 2011.
[17] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.
[18] T. Rockafellar. Conjugate Duality and Optimization. *No 16 in Conference Board of Math. Sciences Series, SIAM Publications*, pages 1–79, 1974.
[19] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31:544–561, 2006.
[20] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.
[21] A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209:63–72, 2011.
[22] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* SIAM, Philadelphia, 2009.

Appendix

Proof of formula (2.48)**.** We prove (2.48) adapting the proof of Lemma 2.1 in [6] to the special case of value function $\underline{\mathcal{Q}}_t^{k-1}$. Defining

$$\mathcal{S}_k = \mathcal{X}_t \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n \cap C_k \cap D,$$

where

$$\begin{array}{lcl} C_k & = & \left\{ (x_t, f, \theta, x_{t-1}) : \left\{ \begin{array}{l} A_t^{k-1} x_t + B_t^{k-1} x_{t-1} + C_t^{k-1} \leq f \mathbf{e}, \\ \theta_{t+1}^{0:k-1} + \beta_{t+1}^{0:k-1} x_t \leq \theta \mathbf{e}, \\ D_t^{k-1} x_t + E_t^{k-1} x_{t-1} + H_t^{k-1} \leq 0 \end{array} \right. \right\} \\ D & = & \{ (x_t, f, \theta, x_{t-1}) : A_t x_t + B_t x_{t-1} = b_t \}, \end{array}$$

we have

$$\underline{\mathcal{Q}}_t^k (x_{t-1}^k) = \left\{ \begin{array}{l} \inf \ f + \theta + \mathbb{I}_{\mathcal{S}_k}(x_t, f, \theta, x_{t-1}^k) \\ x_t \in \mathbb{R}^n, f, \theta \in \mathbb{R}. \end{array} \right. \tag{5.112}$$

Using Theorem 24(a) in Rockafellar [18], we have

$$\begin{array}{lcl} \beta_t^k \in \partial \underline{\mathcal{Q}}_t^k (x_{t-1}^k) & \Leftrightarrow & (0,0,0,\beta_t^k) \in \partial (f + \theta + \mathbb{I}_{\mathcal{S}_k})(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k) \\ & \Leftrightarrow & (0,0,0,\beta_t^k) \in [0;1;1;0] + \mathcal{N}_{\mathcal{S}_k}(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k), \quad (a) \end{array} \tag{5.113}$$

where f_{tk} and θ_{tk} are the optimal values of respectively f and θ in (2.45). For equivalence (5.113)-(a), we have used the fact that $(x_t, f, \theta, x_{t-1}) \to f + \theta$ and $\mathbb{I}_{\mathcal{S}_k}$ are proper, finite at $(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k)$, and the intersection of the relative interior of the domain of these functions, i.e., set $\mathrm{ri}(\mathcal{S}_k)$, is nonempty. Next,

$$\mathcal{N}_{\mathcal{S}_k}(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k) = \mathcal{N}_{C_k}(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k) + \mathcal{N}_D(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k) + \mathcal{N}_{\mathcal{X}_t \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n}(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k), \tag{5.114}$$

and standard calculus on normal cones gives

$$\begin{array}{lcl} \mathcal{N}_{\mathcal{X}_t\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}^n}(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k) & = & \mathcal{N}_{\mathcal{X}_t}(x_t^k)\times\{0\}\times\{0\}\times\{0\}, \\ \mathcal{N}_D(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k) & = & \Big\{[A_t^\top; 0; 0; B_t^\top]\lambda \;:\; \lambda\in\mathbb{R}^q\Big\}, \end{array} \tag{5.115}$$

and $\mathcal{N}_{C_k}(x_t^k, f_{tk}, \theta_{tk}, x_{t-1}^k)$ is the set of points of form

$$\left(\begin{array}{c} (A_t^{k-1})^\top\alpha + (\beta_{t+1}^{0:k-1})^\top\delta + (D_t^{k-1})^\top\mu \\ -\mathbf{e}^\top\alpha \\ -\mathbf{e}^\top\delta \\ (B_t^{k-1})^\top\alpha + (E_t^{k-1})^\top\mu \end{array}\right) \tag{5.116}$$

where α, δ, μ satisfy

$$\begin{array}{c} \alpha,\delta,\mu\geq 0, \\ \left(\begin{array}{c}\alpha\\ \delta\\ \mu\end{array}\right)^\top \left(\begin{array}{l} A_t^{k-1}x_t^k + B_t^{k-1}x_{t-1}^k + C_t^{k-1} - f_{tk}\mathbf{e} \\ \theta_{t+1}^{0:k-1} + \beta_{t+1}^{0:k-1}x_t^k - \theta_{tk}\mathbf{e} \\ D_t^{k-1}x_t^k + E_t^{k-1}x_{t-1}^k + H_t^{k-1} \end{array}\right) = 0. \end{array} \tag{5.117}$$

Combining (5.113), (5.114), (5.115), (5.116), we see that $\beta_t^k \in \partial\underline{\mathcal{Q}}_t^k(x_{t-1}^k)$ if and only if β_t^k is of form

$$B_t^\top\lambda + (B_t^{k-1})^\top\alpha + (E_t^{k-1})^\top\mu \tag{5.118}$$

where α, λ, μ satisfies (5.117) and

$$\begin{array}{lcl} 0 & \in & \mathcal{N}_{\mathcal{X}_t}(x_t^k) + A_t^\top\lambda + (A_t^{k-1})^\top\alpha + (\beta_{t+1}^{0:k-1})\delta + (D_t^{k-1})^\top\mu, \\ 0 & = & 1-\mathbf{e}^\top\alpha, \\ 0 & = & 1-\mathbf{e}^\top\delta. \end{array} \tag{5.119}$$

Finally, it suffices to observe that α, λ, μ satisfies (5.117) and (5.119) if and only if $\alpha, \lambda, \mu, \delta$ is an optimal solution of dual problem (2.46). Therefore $\partial\underline{\mathcal{Q}}_t^k(x_{t-1}^k)$ is the set of points of form (5.118) where $\alpha, \lambda, \mu, \delta$ is an optimal solution of dual problem (2.46). □

To prove (3.84) and (3.97), we will need the following lemma (the proof of (ii) of this lemma was given in [2] for a more general sampling scheme and the proof of (i), that we detail, is similar to the proof of (ii)):

Lemma 5.1. *Assume that Assumptions (H0), (H1)-Sto, and (H2) hold for StoDCuP. Define random variables* $y_n^k = 1(k\in\mathcal{S}_n)$.

(i) Let $\varepsilon>0$, $t\in\{1,\ldots,T\}$, $n\in\texttt{Nodes(t}-\texttt{1)}$, $m\in C(n)$, $i\in\{1,\ldots,p\}$ *and set*

$$K_{\varepsilon,m,i} = \Big\{k\geq 1 : g_{ti}(x_m^k, x_n^k, \xi_m) - g_{tij_t(m)}^{k-1}(x_m^k, x_n^k) \geq \varepsilon\Big\}.$$

Let

$$\Omega_0(\varepsilon) = \{\omega\in\Omega : |K_{\varepsilon,m,i}(\omega)| \text{ is infinite}\}$$

and assume that $\Omega_0(\varepsilon)\neq\emptyset$. *Define on the sample space* $\Omega_0(\varepsilon)$ *the random variables* $\mathcal{I}_{\varepsilon,m,i}(j), j\geq 1$, *where* $\mathcal{I}_{\varepsilon,m,i}(1) = \min\{k\geq 1 : k\in K_{\varepsilon,m,i}(\omega)\}$ *and for* $j\geq 2$

$$\mathcal{I}_{\varepsilon,m,i}(j) = \min\{k > \mathcal{I}_{\varepsilon,m,i}(j-1) : k\in K_{\varepsilon,m,i}(\omega)\},$$

i.e., $\mathcal{I}_{\varepsilon,m,i}(j)(\omega)$ *is the index of* j*th iteration* k *such that* $g_{ti}(x_m^k, x_n^k, \xi_m) - g_{tij_t(m)}^{k-1}(x_m^k, x_n^k) \geq \varepsilon$. *Then random variables* $(y_n^{\mathcal{I}_{\varepsilon,m,i}(j)})_{j\geq 1}$ *defined on sample space* $\Omega_0(\varepsilon)$ *are independent, have the distribution of* y_n^1 *and therefore by the Strong Law of Large numbers we have*

$$\mathbb{P}\left(\lim_{N\to+\infty}\frac{1}{N}\sum_{j=1}^N y_n^{\mathcal{I}_{\varepsilon,m,i}(j)} = \mathbb{E}[y_n^1]\right) = 1. \tag{5.120}$$

(ii) Let $\varepsilon>0$, $t\in\{1,\ldots,T\}$, $n\in\texttt{Nodes(t}-\texttt{1)}$, *and set*

$$K_{\varepsilon,n} = \left\{k\geq 1 : \mathcal{Q}_t(x_n^k) - \mathcal{Q}_t^k(x_n^k)\geq\varepsilon\right\}.$$

Let

$$\Omega_1(\varepsilon) = \{\omega\in\Omega : |K_{\varepsilon,n}(\omega)| \text{ is infinite}\}$$

and assume that $\Omega_1(\varepsilon) \neq \emptyset$. Define on the sample space $\Omega_1(\varepsilon)$ the random variables $\mathcal{I}_{\varepsilon,n}(j), j \geq 1$, where $\mathcal{I}_{\varepsilon,n}(1) = \min\{k \geq 1 : k \in K_{\varepsilon,n}(\omega)\}$ and for $j \geq 2$

$$\mathcal{I}_{\varepsilon,n}(j) = \min\{k > \mathcal{I}_{\varepsilon,n}(j-1) : k \in K_{\varepsilon,n}(\omega)\},$$

i.e., $\mathcal{I}_{\varepsilon,n}(j)(\omega)$ is the index of jth iteration k such that $\mathcal{Q}_t(x_n^k) - \mathcal{Q}_t^k(x_n^k) \geq \varepsilon$. Then random variables $(y_n^{\mathcal{I}_{\varepsilon,n}(j)})_{j\geq 1}$ defined on sample space $\Omega_1(\varepsilon)$ are independent, have the distribution of y_n^1 and therefore by the Strong Law of Large numbers we have

$$\mathbb{P}\left(\lim_{N\to+\infty} \frac{1}{N}\sum_{j=1}^{N} y_n^{\mathcal{I}_{\varepsilon,n}(j)} = \mathbb{E}[y_n^1]\right) = 1. \tag{5.121}$$

Proof: (i) Define on the sample space $\Omega_0(\varepsilon)$ the random variables $(w_{\varepsilon,m,i}^k)_k$ by

$$w_{\varepsilon,m,i}^k(\omega) = \begin{cases} 1 & \text{if } k \in K_{\varepsilon,m,i}(\omega) \\ 0 & \text{otherwise.} \end{cases}$$

To alleviate notation (ε, m, n, i being fixed), let us put $w^k := w_{\varepsilon,m,i}^k$, $\mathcal{I}(j) := \mathcal{I}_{\varepsilon,m,i}(j)$, For $\overline{y}_j \in \{0,1\}$, we have

$$\mathbb{P}\Big(y_n^{\mathcal{I}(j)} = \overline{y}_j\Big) = \sum_{\overline{\mathcal{I}}_j=1}^{\infty} \mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_j} = \overline{y}_j; \mathcal{I}(j) = \overline{\mathcal{I}}_j\Big). \tag{5.122}$$

Observe that the event $\mathcal{I}(j) = \overline{\mathcal{I}}_j$ can be written as the union $\bigcup_{1\leq\overline{\mathcal{I}}_1<\overline{\mathcal{I}}_2<\ldots<\overline{\mathcal{I}}_j} E(\overline{\mathcal{I}}_1,\ldots,\overline{\mathcal{I}}_j)$ of events

$$E(\overline{\mathcal{I}}_1,\ldots,\overline{\mathcal{I}}_j) := \left\{ \begin{array}{l} w^{\overline{\mathcal{I}}_1} = \ldots = w^{\overline{\mathcal{I}}_j} = 1, \\ w^{\ell} = 0, 1 \leq \ell < \overline{\mathcal{I}}_j, \ell \notin \{\overline{\mathcal{I}}_1,\ldots,\overline{\mathcal{I}}_j\} \end{array} \right\}.$$

Due to Assumption (H2) observe that random variable $y_n^{\overline{\mathcal{I}}_j}$ is independent of random variables $w^i, i = 1,\ldots,\overline{\mathcal{I}}_j$, and therefore events $\{y_n^{\overline{\mathcal{I}}_j} = \overline{y}_j\}$ and $\{\mathcal{I}(j) = \overline{\mathcal{I}}_j\}$ are independent which gives

(5.123)

$$\begin{array}{lll} \mathbb{P}\Big(y_n^{\mathcal{I}(j)} = \overline{y}_j\Big) & = & \displaystyle\sum_{\overline{\mathcal{I}}_j=1}^{\infty} \mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_j} = \overline{y}_j; \mathcal{I}(j) = \overline{\mathcal{I}}_j\Big) \\ & = & \displaystyle\sum_{\overline{\mathcal{I}}_j=1}^{\infty} \mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_j} = \overline{y}_j\Big)\mathbb{P}\Big(\mathcal{I}(j) = \overline{\mathcal{I}}_j\Big) = \mathbb{P}\Big(y_n^1 = \overline{y}_j\Big) \displaystyle\sum_{\overline{\mathcal{I}}_j=1}^{\infty} \mathbb{P}\Big(\mathcal{I}(j) = \overline{\mathcal{I}}_j\Big) = \mathbb{P}\Big(y_n^1 = \overline{y}_j\Big) \end{array}$$

where we have used the fact that y_n^1 and $y_n^{\overline{\mathcal{I}}_j}$ have the same distribution (from (H2)).

Next for $\overline{y}_1,\ldots,\overline{y}_p \in \{0,1\}$, we have

$$\mathbb{P}\Big(y_n^{\mathcal{I}(1)} = \overline{y}_1,\ldots,y_n^{\mathcal{I}(p)} = \overline{y}_p\Big) = \sum_{1\leq\overline{\mathcal{I}}_1<\overline{\mathcal{I}}_2<\ldots<\overline{\mathcal{I}}_p}^{\infty} \mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_1} = \overline{y}_1;\ldots;y_n^{\overline{\mathcal{I}}_p} = \overline{y}_p; \mathcal{I}(1) = \overline{\mathcal{I}}_1;\ldots;\mathcal{I}(p) = \overline{\mathcal{I}}_p\Big).$$

By the same reasoning as above, the event

$$\Big\{y_n^{\overline{\mathcal{I}}_1} = \overline{y}_1;\ldots;y_n^{\overline{\mathcal{I}}_{p-1}} = \overline{y}_{p-1}; \mathcal{I}(1) = \overline{\mathcal{I}}_1;\ldots;\mathcal{I}(p) = \overline{\mathcal{I}}_p\Big\}$$

can be expressed in terms of random variables $y_n^{\overline{\mathcal{I}}_1}, \ldots, y_n^{\overline{\mathcal{I}}_{p-1}}, w_n^{\overline{\mathcal{I}}_1}, \ldots, w_n^{\overline{\mathcal{I}}_p}$, and is therefore independent of event $\{y_n^{\overline{\mathcal{I}}_p} = \overline{y}_p\}$. It follows that
(5.124)

$$\mathbb{P}\Big(y_n^{\mathcal{I}(j)} = \overline{y}_j, 1 \le j \le p\Big) = \sum_{1 \le \overline{\mathcal{I}}_1 < \overline{\mathcal{I}}_2 < \ldots < \overline{\mathcal{I}}_p}^{\infty} \mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_p} = \overline{y}_p\Big)\mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_j} = \overline{y}_j, 1 \le j \le p-1; \mathcal{I}(j) = \overline{\mathcal{I}}_j, 1 \le j \le p\Big)$$

$$= \mathbb{P}\Big(y_n^1 = \overline{y}_p\Big) \sum_{1 \le \overline{\mathcal{I}}_1 < \overline{\mathcal{I}}_2 < \ldots < \overline{\mathcal{I}}_p}^{\infty} \mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_j} = \overline{y}_j, 1 \le j \le p-1; \mathcal{I}(j) = \overline{\mathcal{I}}_j, 1 \le j \le p\Big)$$

$$= \mathbb{P}\Big(y_n^1 = \overline{y}_p\Big) \sum_{1 \le \overline{\mathcal{I}}_1 < \overline{\mathcal{I}}_2 < \ldots < \overline{\mathcal{I}}_{p-1}}^{\infty} \mathbb{P}\Big(y_n^{\overline{\mathcal{I}}_j} = \overline{y}_j, 1 \le j \le p-1; \mathcal{I}(j) = \overline{\mathcal{I}}_j, 1 \le j \le p-1\Big)$$

$$= \mathbb{P}\Big(y_n^1 = \overline{y}_p\Big)\mathbb{P}\Big(y_n^{\mathcal{I}(j)} = \overline{y}_j, 1 \le j \le p-1\Big).$$

By induction this implies

$$\mathbb{P}\Big(y_n^{\mathcal{I}(j)} = \overline{y}_j, 1 \le j \le p\Big) \quad = \quad \prod_{j=1}^{p} \mathbb{P}\Big(y_n^1 = \overline{y}_j\Big) \overset{(5.123)}{=} \prod_{j=1}^{p} \mathbb{P}\Big(y_n^{\mathcal{I}(j)} = \overline{y}_j\Big) \tag{5.125}$$

which shows that random variables $(y_n^{\mathcal{I}(j)})_{j \ge 1}$ are independent.

The proof of (ii) is similar to the proof of (i). ■

Proof of (3.84) **and** (3.97)**.** As in [2], we can now use the previous lemma to prove (3.84) and (3.97). Let us prove (3.84). By contradiction, assume that (3.84) does not hold. Then there is $\varepsilon > 0$ such that the set $\Omega_0(\varepsilon)$ defined in Lemma 5.1 is nonempty. By Lemma 5.1, this implies that (5.120) holds. But due to (3.83), only a finite number of indices $\mathcal{I}_{\varepsilon,m,i}(j)$ can be in $\mathcal{S}_n$ (with corresponding variable $y_n^{\mathcal{I}_{\varepsilon,m,i}(j)}$ being one) and therefore $\mathbb{P}\left(\lim_{N \to +\infty} \frac{1}{N} \sum_{j=1}^{N} y_n^{\mathcal{I}_{\varepsilon,m,i}(j)} = 0\right) = 1$, which is a contradiction with (5.120).

The proof of (3.97) is similar to the proof of (3.84), by contradiction and using (3.96) and Lemma 5.1-(ii) (see also [2], [3]). □