On the convergence of sampling-based decomposition algorithms for multistage stochastic programs

Karsten Linowsky [*,†] Andy Philpott [‡,§]

March 18, 2004

Abstract

The paper presents a convergence proof for a broad class of sampling algorithms for multi-stage stochastic linear programs in which the uncertain parameters occur only in the constraint right-hand sides. This class includes SDDP, AND, ReSa, and CUPPS. We show under some independence assumptions on the sampling procedure that the algorithms converge with probability 1.

Keywords: Multistage stochastic programming, sampling, almost sure convergence.

1 Introduction

Multistage stochastic linear programming models have many applications but they are notoriously difficult to solve. The most successful approaches in practical applications appear to be the sampling-based methods. The first of these approaches (SDDP) was developed by Pereira and Pinto [6] in the context of hydro-electricity planning. This algorithm has been successfully applied (see [7]) to compute solutions to long-term hydro-thermal reservoir planning models. To the authors' knowledge no convergence result for this method has appeared in the literature. Since Pereira and Pinto's paper, a number of related algorithms have emerged (see e.g. CUPPS [3], AND [4], and ReSa [5]) based on similar ideas.

In this paper we derive a general convergence result for algorithms of this type. A convergence proof specifically aimed at the CUPPS algorithm has already appeared in [3]. The argument we employ in our proof closely resembles that used in [3], in that we use the same induction on stages. However our result is more general, being

*University of St Gallen, Switzerland
†The first author would like to acknowledge support by the Swiss National Science Foundation.
‡University of Auckland, New Zealand
§The second author would like to acknowledge support of NZPGS&T grant UOAX 0203

applicable to SDDP, AND, CUPPS, and ReSa. The main contribution of our work is to identify the crucial conditions that guarantee convergence of sampling-based multi-stage stochastic Benders decomposition methods. These assumptions are made precise in the sequel, but essentially they amount to the following requirements:

1. cuts should eventually be computed in every stage;
2. samples that are used to create scenarios in the forward pass are also used in cut generation.

In the next section we give a general formulation of the multi-stage stochastic programming problem and describe the general algorithmic approach. The convergence proof is then derived in section 3 using a series of lemmas. In section 4 we show how the sampling-based algorithms of [3], [4], [5], and [6] satisfy the conditions of the theorem.

2 Multistage decomposition

Multistage stochastic linear programs with recourse are well known in the stochastic programming community. The general form of these is described in [1]. In this paper we restrict our attention to multistage stochastic programs with the following properties:

A.1 Random quantities appear only on the right-hand side of the linear constraints in each stage.

A.2 The set Ω_t of random outcomes in each stage t is discrete and finite ($\Omega_t = \{\omega_{ti} | \ i = 1, \ldots, q_t < \infty\}$ with probabilities $p_{ti} > 0$, $\forall i$).

A.3 Random quantities in different stages are independent.

A.4 The feasible region of the linear program in each stage is non-empty and bounded.

Under these assumptions, the multi-stage stochastic linear program can be written in the following form:

Solve the problem $[LP_1]$ defined by

$$\begin{aligned} Q_1 &= \min_{x_1} c_1^\top x_1 + \mathcal{Q}_2(x_1) \\ s.t. \quad A_1 x_1 &= b_1 \\ x_1 &\geq 0, \end{aligned}$$

where for all $t = 2, \ldots, T$,

$$\mathcal{Q}_t(x_{t-1}) = \sum_{i=1}^{q_t} p_{ti} Q_t(x_{t-1}, \omega_{ti}),$$

and $[LP_t]$ is the problem

$$\begin{aligned} Q_t(x_{t-1},\omega_t) &= \min_{x_t} c_t^\top x_t + \mathcal{Q}_{t+1}(x_t) \\ s.t. \quad A_t x_t &= \omega_t - B_{t-1}x_{t-1} \\ x_t &\geq 0, \end{aligned}$$

and we set $\mathcal{Q}_{T+1} \equiv 0$.

The problem $[LP_t]$ depends on the choice of ω_t and x_{t-1}, and so we could write $[LP_t(x_{t-1},\omega_t)]$, though we choose to suppress this dependence in the notation. (By assumption A3, $[LP_t]$ is independent of $\omega_{t-1},\omega_{t-2}\ldots$.) Observe that $Q_t(x_{t-1},\omega_t)$ is a polyhedral convex function, and so is continuous in x_{t-1} at all points of its domain.

In the algorithms that are considered in this paper the functions $\mathcal{Q}_t(x_{t-1})$ in each stage are approximated by the maximum of a collection of linear functions, each of which is called a *cut*. This gives rise to a sequence of approximate problems $[AP_t^k]$ for each stage. These are defined for iteration k as follows:

For $t=1$, $[AP_1^k]$ is the linear program

$$\begin{aligned} C_1^k &= \min_{x_1,\theta_2} c_1^\top x_1 + \theta_2 \\ s.t. \quad A_1 x_1 &= b_1 \\ \theta_2 + (\beta_2^j)^\top x_1 &\geq \alpha_{2,j} \qquad (j=0,\ldots,k-1) \\ x_1 &\geq 0, \end{aligned}$$

and, for all $t=2,\ldots,T-1$, $[AP_t^k]$ is the linear program

$$\begin{aligned} C_t^k(x_{t-1},\omega_t) &= \min_{x_t,\theta_{t+1}} c_t^\top x_t + \theta_{t+1} \\ s.t. \quad A_t x_t &= \omega_t - B_{t-1}x_{t-1} \\ \theta_{t+1} + (\beta_{t+1}^j)^\top x_t &\geq \alpha_{t+1,j} \qquad (j=0,\ldots,k-1) \\ x_t &\geq 0. \end{aligned}$$

For all stages, the first cut ($j=0$) is set as the trivial cut $\theta_{t+1} \geq -\infty$. We shall use the notation (π_t,ρ_t) to denote dual variables of the problem $[AP_t^k]$, where π_t corresponds to the equality constraints, and ρ_t corresponds to the cut constraints. We also use the notation $\mathcal{C}_t^k(x_{t-1})$ to denote $\sum_{i=1}^{q_t} p_{ti} C_t^k(x_{t-1},\omega_t)$.

Observe that under assumption A4,

$$\{x_t \mid A_t x_t = \omega_t - B_{t-1}x_{t-1}, \quad x_t \geq 0\}$$

is nonempty and bounded so $[AP_t^k]$ always has a nonempty feasible set (with θ_{t+1} chosen large enough) and hence an optimal solution. Thus the dual feasible region of $[AP_t^k]$ is nonempty. Moreover by assumption A1, the dual feasible sets are independent of the outcomes of the random quantities, which allows us to construct a valid cut at each stage based on an assembled collection of dual solutions from different samples.

In the last stage, T, the algorithms solve the actual problem $[LP_T]$, therefore

$$C_T^k(x_{T-1},\omega_T) = Q_T(x_{T-1},\omega_T) \quad \forall k.$$

Since cuts are added from one iteration to the next, and no cuts are taken out, the objective values of the approximated problems form a monotone sequence, i.e.

$$C_t^{k+1}(x_{t-1}, \omega_t) \geq C_t^k(x_{t-1}, \omega_t) \qquad \forall t, \forall k.$$

2.1 A class of sampling-based decomposition algorithm

In this section we define a general class of sampling algorithms for solving $[LP_1]$. We first describe the cut generation of the algorithms.

Definition 1 (Sampled cut)
A sampled cut *at x_{t-1}^k with sample $\Omega_t^k \subseteq \Omega_t$ is computed as follows:*

1. *Solve $[AP_t^k]$ for all $\omega_{ti} \in \Omega_t^k$, and let $(\pi_t^i(x_{t-1}^k), \rho_t^i(x_{t-1}^k))$ be the optimal dual variables (attained at an extreme point). Add them to the set $\mathcal{D}_t^k$.*

2. *For all $\omega_{ti} \notin \Omega_t^k$, set*

$$(\pi_t^i(x_{t-1}^k), \rho_t^i(x_{t-1}^k)) = \arg\max \left\{ \pi_t^\top (\omega_{ti} - B_{t-1} x_{t-1}^k) + \rho_t^\top (\alpha_{t+1}^k) \;\middle|\; (\pi_t, \rho_t) \in \mathcal{D}_t^k \right\}$$

   or if $t = T$:

$$\pi_T^i(x_{T-1}^k) = \arg\max \left\{ \pi_T^\top (\omega_{Ti} - B_{T-1} x_{T-1}^k) \;\middle|\; \pi_T \in \mathcal{D}_T^k \right\}.$$

3. *The cut has formula*

$$\theta_t \geq \alpha_{t,k} - (\beta_t^k)^\top x_{t-1}$$

   where

$$\begin{aligned} \beta_t^k &= \sum_{i=1}^{q_t} p_{ti} \; B_{t-1}^\top \; \pi_t^i(x_{t-1}^k) && \textit{for } 2 \leq t \leq T, \\ \alpha_{t,k} &= \sum_{i=1}^{q_t} p_{ti} \left[\omega_{ti}^\top \; \pi_t^i(x_{t-1}^k) + (\alpha_{t+1}^{k-1})^\top \rho_t^i(x_{t-1}^k) \right] && \textit{for } 2 \leq t \leq T-1, \\ \alpha_{T,k} &= \sum_{i=1}^{q_T} p_{Ti} \; \omega_{Ti}^\top \; \pi_T^i(x_{T-1}^k). \end{aligned}$$

Observe that $\alpha_{t,k}$ is a scalar, whereas α_{t+1}^{k-1} denotes a $(k-1)$-dimensional vector. This means that the dimensions of α_{t+1}^{k-1} and $\rho_t^i(x_{t-1}^k)$ are increasing as the iteration count k increases. Note also that a sampled cut is well defined for $\Omega_t^k = \emptyset$, as long as $\mathcal{D}_t^k \neq \emptyset$. If $\Omega_t^k = \mathcal{D}_t^k = \emptyset$ then we set $\alpha_{t,k} = -\infty$, $\beta_t^k = 0$.

In our convergence proof we shall make use of the fact that $\pi_t^i(x_{t-1}^k)$ lies in a bounded set. In fact $\pi_t^i(x_{t-1}^k)$ can take only a finite number of values in the course of the algorithm. This is a consequence of the fact that π_t and ρ_t are chosen to be extreme-point solutions of the dual of $[AP_t^k]$. We state this result formally as the following lemma.

Lemma 2 *For all t there is some m_t such that $\mathcal{D}_t^k$ has cardinality at most m_t.*

Proof: We use induction on t. First, if $(\pi, \rho) \in \mathcal{D}_T^k$ then $\rho = 0$ and π is an extreme point of $\{\pi \mid A_T^\top \pi \leq c_T\}$ of which there is only a finite number. So $|\mathcal{D}_T^k| \leq m_T$, for some m_T.

Now suppose $|\mathcal{D}_t^k| \leq m_t$. Then the vector

$$\beta_t^j = \sum_{i=1}^{q_t} p_{ti}\ B_{t-1}^\top\ \pi_t^i(x_{t-1}^j)$$

takes at most $(m_t)^{q_t}$ values. This means that $\mathcal{E}_{t-1}^k$, the set of extreme points of

$$\{(\pi_{t-1}, \rho_{t-1}) \mid A_{t-1}^\top \pi_{t-1} + \sum_{j=1}^{k-1} \beta_t^j \rho_{t-1}^j \leq c_{t-1},\ \sum_{j=1}^{k-1} \rho_{t-1}^j = 1\}$$

has cardinality no more than m_{t-1}, say, independent of k. But $\mathcal{E}_{t-1}^k \supseteq \mathcal{D}_{t-1}^k$ which establishes the result. □

Now a general class of sampling-based decomposition algorithms is defined, for which we will show convergence to the optimal solution. The algorithms work in the following way:

Multi-stage Sampled Benders Decomposition (MSBD)

Step 0: (Initialisation) Set iteration counter $k = 1$.

Step 1: (Candidate solutions)
In each iteration k, a complete sample path $\{\omega_t^k\}_{t=2,\ldots,T}$ of the scenario tree is constructed independently of previous iterations. For this path the approximate problems are solved up to stage $T - 1$, to yield the primal solutions (x_t^k, θ_{t+1}^k) of the problem $[AP_t^k]$.

Step 2: (Cut generation)
For each stage $t = 2, \ldots, T$ sampled cuts are generated at x_{t-1}^k with sample Ω_t^k.

Step 3: Set $k = k + 1$ and go to Step 1.

Note that at each stage t, two samples are used in each iteration. Unless $\Omega_t^k = \{\omega_t^k\}$, these samples may be different. Observe also that they need not be independent, in fact one might choose $\Omega_t^k = \{\omega_t^k\}$. However in order to yield a convergence result for MSBD we will require the following two properties of the sampling procedure.

Definition 3 (Cut Sampling Property)
MSBD is said to fulfill the cut-sampling property *(CSP) if for each stage t, $\{k \mid \Omega_t^k = \emptyset\}$ is finite.*

Definition 4 (Sample Intersection Property)
MSBD is said to fulfill the sample-intersection property *(SIP) if for each stage t, and every outcome $\omega_{ti} \in \Omega_t$, $\Pr[(\omega_{ti} \in \Omega_t^k) \cap (\omega_t^k = \omega_{ti})] > 0$ for every k with $\Omega_t^k \neq \emptyset$.*

The cut-sampling property entails that eventually the algorithm will compute $(\pi_t^i(x_{t-1}^k), \rho_t^i(x_{t-1}^k))$ for at least one outcome ω_{ti} at every stage. The sample-intersection property guarantees that the outcomes that are used to compute cuts will (with positive probability) include some information from the outcome in the sample path that is constructed in Step 1. SIP holds for example if ω_t^k and Ω_t^k are sampled independently, or alternatively if Ω_t^k is chosen so as to always include ω_t^k (as in CUPPS).

Lemma 5 *Suppose that MSBD satisfies SIP. Then it fulfills CSP if and only if for any stage t and every infinite subsequence $\{x_{t-1}^k\}_{k\in J}$ generated by the algorithm, for each $i = 1, \ldots, q_t$ the subsequence $\{x_{t-1}^k\}_{k\in J_i}$ with $J_i = J \cap \{k | \omega_{ti} \in \Omega_t^k \text{ and } \omega_t^k = \omega_{ti}\}$ is infinite with probability one (wp1).*

Proof. For some arbitrary stage t, let $\{x_{t-1}^k\}_{k\in J}$ be an infinite subsequence generated by the algorithm and suppose $\{k|\Omega_t^k = \emptyset\}$ is finite (by CSP). Then the intersection $J \cap K$ (where K denotes $\{k|\Omega_t^k \neq \emptyset\}$) is infinite . Furthermore, by SIP for every $k \in J \cap K$ we have $\Pr[(\omega_{ti} \in \Omega_t^k) \cap (\omega_t^k = \omega_{ti})] > 0 \ \forall i$. Due to the Borel-Cantelli Lemma (see e.g. [2]) the set $J_i \subseteq J \cap K$ with $\omega_{ti} \in \Omega_t^k$ and $\omega_t^k = \omega_{ti}$ for all $k \in J_i$ is infinite wp1.

Suppose now that CSP does not hold. Then for some stage t the set $K' = \{k|\Omega_t^k = \emptyset\}$ is infinite. Then for any subsequence J' of K', $\Omega_t^k = \emptyset \ \forall k \in J'$. □

Lemma 6 *The sampled cuts are valid cuts. Furthermore, the following relations hold:*

$$
\begin{aligned}
\mathcal{Q}_t(x_{t-1}^k) &\geq \theta_t^k && ,\forall k \in \mathbb{N}, \forall \mathrm{t} = 2, \ldots, \mathrm{T}, \\
Q_{t-1}(x_{t-2}, \omega_{t-1}) &\geq C_{t-1}^k(x_{t-2}, \omega_{t-1}) && ,\forall x_{t-2}, \omega_{t-1}, \forall k \in \mathbb{N}, \forall \mathrm{t} = 2, \ldots, \mathrm{T}.
\end{aligned}
$$

Proof: The proof of this lemma can be obtained equivalently to Lemmas 4.1 and 4.3 in [3]. □

Lemma 7 *Suppose that MSBD satisfies SIP and CSP. Then for any convergent sequence $\{x_{t-1}^k\}_{k\in J}$ generated by MSBD there exists wp1 a sequence $\{\Delta^k\}_{k\in J}$ and q_t disjoint subsequences of J indexed by $r_i \in J_i$ with:*

1. $\theta_t^k \geq \sum_{i=1}^{q_t} p_{ti}\ C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) + \Delta^k$;
2. $J_i \subseteq J \cap \{k|\ \omega_{ti} \in \Omega_t^k, \omega_t^k = \omega_{ti}\}$;
3. $r_i < k$ *for all but a finite number of* k, *and* $r_i \to \infty$ *as* $k \to \infty$;
4. $\lim_{k\to\infty} |\Delta^k| = 0$.

Proof: Consider stage $t \in \{2, \ldots, T-1\}$ (Stage T can be treated in a similar way).

Let $\{x_{t-1}^k\}_{k\in J}$ be a convergent sequence and consider iteration $k \in J$. All cuts generated up to this iteration must be satisfied, so for all $r < k$:

$$\begin{aligned}
\theta_t^k &\geq \alpha_{t,r} - (\beta_t^r)^\top x_{t-1}^k \\
&= \sum_{i=1}^{q_t} p_{ti} \ \left[(\pi_t^i(x_{t-1}^r))^\top(\omega_{ti} - B_{t-1}x_{t-1}^k) + (\alpha_{t+1}^r)^\top \rho_t^i(x_{t-1}^r)\right] \\
&= \sum_{i=1}^{q_t} p_{ti} \ \left[(\pi_t^i(x_{t-1}^r))^\top(\omega_{ti} - B_{t-1}x_{t-1}^r) + (\alpha_{t+1}^r)^\top \rho_t^i(x_{t-1}^r)\right] + \Delta_1^k,
\end{aligned}$$

with

$$\Delta_1^k = \sum_{i=1}^{q_t} p_{t,i} \ (\pi_t^i(x_{t-1}^r))^\top B_{t-1}(x_{t-1}^r - x_{t-1}^k).$$

From Lemma 5, for each $i = 1, \ldots, q_t$ the set J has an infinite subset J_i wp1, such that $\omega_{ti} \in \Omega_t^{r_i}$ and $\omega_t^{r_i} = \omega_{ti}$ for all $r_i \in J_i$. Choose r_i as the largest member of the set J_i which is smaller than k. Since the sets J_i are infinite, $r_i \to \infty$ as $k \to \infty$.

Since the dual optimal solution of iteration r_i, $(\pi_t^i(x_{t-1}^{r_i}), \rho_t^i(x_{t-1}^{r_i}))$, is dual feasible in iteration $r = \max\{r_i |\ i = 1, \ldots, q_t\}$, we have

$$\theta_t^k \geq \sum_{i=1}^{q_t} p_{ti} \ \left[(\pi_t^i(x_{t-1}^{r_i}))^\top(\omega_{ti} - B_{t-1}x_{t-1}^r) + (\alpha_{t+1}^r)^\top \rho_t^i(x_{t-1}^{r_i})\right] + \Delta_1^k. \tag{1}$$

Here we adopt the convention that $\rho_t^i(x_{t-1}^{r_i})$ has zero components added to give it dimension r. This means that $(\alpha_{t+1}^r)^\top \rho_t^i(x_{t-1}^{r_i}) = (\alpha_{t+1}^{r_i})^\top \rho_t^i(x_{t-1}^{r_i})$ so the right-hand side of (1) becomes

$$\begin{aligned}
&\sum_{i=1}^{q_t} p_{ti} \ \left[(\pi_t^i(x_{t-1}^{r_i}))^\top(\omega_{ti} - B_{t-1}x_{t-1}^{r_i}) + (\alpha_{t+1}^{r_i})^\top \rho_t^i(x_{t-1}^{r_i})\right] + \Delta_1^k + \Delta_2^k \\
= &\sum_{i=1}^{q_t} p_{ti} \ C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) + \Delta_1^k + \Delta_2^k,
\end{aligned}$$

where

$$\Delta_2^k = \sum_{i=1}^{q_t} p_{t,i}(\pi_t^i(x_{t-1}^{r_i}))^\top B_{t-1}(x_{t-1}^{r_i} - x_{t-1}^r).$$

Now

$$\left|\Delta_1^k\right| \leq \sum_{i=1}^{q_t} p_{t,i} \ \left|\left|\pi_t^i(x_{t-1}^r)\right|\right| \ \left|\left|B_{t-1}(x_{t-1}^r - x_{t-1}^k)\right|\right|,$$

and since the dual extreme points are bounded (because $\mathcal{D}_t^k$ is a bounded set), and the sequence $\{x_{t-1}^k\}_{k\in J}$ is convergent, we have (with $r \to \infty$ as $k \to \infty$):

$$\lim_{k\to\infty} \left|\Delta_1^k\right| = 0.$$

Similarly,

$$\lim_{k\to\infty} \left|\Delta_2^k\right| = 0.$$

This completes the proof. □

3 Convergence of the algorithm

In this section we prove the convergence of algorithms that satisfy SIP and CSP, by induction on the stage t. Following [3], we first prove two lemmas that establish this induction.

Lemma 8 *Suppose MSBD satisfies CSP and SIP. Assume for any given infinite set $K \subseteq \mathbb{N}$ that*

- $\omega_{T-1}^k = \omega_{T-1}^0$ *for some given* ω_{T-1}^0 *for any* $k \in K$*;*
- *the sequence* $\left\{x_{T-2}^k\right\}_{k\in K}$ *converges to some given vector* x_{T-2}^0*.*

Then wp1 there exists an infinite set $J \subseteq K$ such that

(a) the sequence $\left\{x_{T-1}^k\right\}_{k\in J}$ *converges to some vector* x_{T-1}^0*;*

(b) the sequence $\left\{\theta_T^k\right\}_{k\in J}$ *converges to* $\mathcal{Q}_T(x_{T-1}^0)$*;*

(c) the sequence $\left\{C_{T-1}^k(x_{T-2}^k, \omega_{T-1}^0)\right\}_{k\in J}$ *converges to* $Q_{T-1}(x_{T-2}^0, \omega_{T-1}^0)$*.*

Proof:
(a)
By assumption the primal feasible sets are bounded. But every infinite bounded sequence has a convergent subsequence. Denote x_{T-1}^0 as the corresponding limit and J as the corresponding index set.
(b)
From Lemma 6 and Lemma 7 we have wp1 a sequence $\{\Delta^k\}_{k\in J}$ and subsequences of J indexed by r_i with

$$\begin{aligned}
\mathcal{Q}_T(x_{T-1}^k) \;\geq\; \theta_T^k \;&\geq\; \sum_{i=1}^{q_T} p_{Ti}\; C_T^{r_i}(x_{T-1}^{r_i}, \omega_{Ti}) + \Delta^k \\
&=\; \mathcal{Q}_T(x_{T-1}^k) + \sum_{i=1}^{q_T} p_{Ti}\; \left[C_T^{r_i}(x_{T-1}^{r_i}, \omega_{Ti}) - Q_T(x_{T-1}^k, \omega_{Ti})\right] + \Delta^k \\
&=\; \mathcal{Q}_T(x_{T-1}^k) + \sum_{i=1}^{q_T} p_{Ti}\; \left[Q_T(x_{T-1}^{r_i}, \omega_{Ti}) - Q_T(x_{T-1}^k, \omega_{Ti})\right] + \Delta^k,
\end{aligned}$$

which yields

$$\left|\theta_T^k - \mathcal{Q}_T(x_{T-1}^k)\right| \leq \left|\Delta_1^k\right|,$$

with

$$\Delta_1^k = \sum_{i=1}^{q_T} p_{Ti}\; \left[Q_T(x_{T-1}^{r_i}, \omega_{Ti}) - Q_T(x_{T-1}^k, \omega_{Ti})\right] + \Delta^k.$$

Now

$$\begin{aligned}
\left|\Delta_1^k\right| \;&\leq\; \sum_{i=1}^{q_T} p_{Ti}\; \left|Q_T(x_{T-1}^{r_i}, \omega_{Ti}) - Q_T(x_{T-1}^k, \omega_{Ti})\right| + \left|\Delta^k\right| \\
&\longrightarrow\; 0 \quad (\text{as } k \to \infty),
\end{aligned}$$

since the function $Q_T(x_{T-1}, \omega_T)$ is continuous in x_{T-1}, $\{x_{T-1}^k\}_{k\in J}$ is a convergent sequence, and $r_i \to \infty$ as $k \to \infty$. This leads to the following intermediate result:

$$\left|\theta_T^k - \mathcal{Q}_T(x_{T-1}^k)\right| \le \left|\Delta_1^k\right| \quad \text{with} \lim_{k\to\infty} \left|\Delta_1^k\right| = 0.$$

Furthermore, due to continuity of $\mathcal{Q}_T(x_{T-1})$ in x_{T-1}, we have

$$\lim_{k\to\infty} \left|\mathcal{Q}_T(x_{T-1}^k) - \mathcal{Q}_T(x_{T-1}^0)\right| = 0.$$

Therefore

$$\begin{aligned}
\left|\theta_T^k - \mathcal{Q}_T(x_{T-1}^0)\right| &\le \left|\theta_T^k - \mathcal{Q}_T(x_{T-1}^k)\right| + \left|\mathcal{Q}_T(x_{T-1}^k) - \mathcal{Q}_T(x_{T-1}^0)\right| \\
&\le \left|\Delta_1^k\right| + \left|\mathcal{Q}_T(x_{T-1}^k) - \mathcal{Q}_T(x_{T-1}^0)\right| \\
&\longrightarrow 0 \quad (\text{as } k \to \infty).
\end{aligned}$$

Hence the sequence $\{\theta_T^k\}_{k\in J}$ converges to $\mathcal{Q}_T(x_{T-1}^0)$ with probabilty 1 which shows part (b).

(c)

Considering Lemma 6 and Lemma 7 again we have

$$\begin{aligned}
Q_{T-1}(x_{T-2}^k, \omega_{T-1}^0) &\ge C_{T-1}^k(x_{T-2}^k, \omega_{T-1}^0) \\
&= c_{T-1}^\top x_{T-1}^k + \theta_T^k \\
&\ge c_{T-1}^\top x_{T-1}^k + \mathcal{Q}_T(x_{T-1}^k) + \Delta_1^k \\
&\ge Q_{T-1}(x_{T-2}^k, \omega_{T-1}^0) + \Delta_1^k,
\end{aligned}$$

where the last inequality comes from the fact that x_{T-1}^k is also feasible for the problem $[LP_{T-1}]$ with $x_{T-2} = x_{T-2}^k$ and $\omega_{T-1} = \omega_{T-1}^0$. This implies that

$$\left|C_{T-1}^k(x_{T-2}^k, \omega_{T-1}^0) - Q_{T-1}(x_{T-2}^k, \omega_{T-1}^0)\right| \le \left|\Delta_1^k\right|.$$

Since the function $Q_{T-1}(x_{T-2}, \omega_{T-1})$ is continuous in x_{T-2}, and the sequence $\{x_{T-2}^k\}_{k\in K}$ is convergent in K (hence also in J):

$$\lim_{k\to\infty} \left|Q_{T-1}(x_{T-2}^k, \omega_{T-1}^0) - Q_{T-1}(x_{T-2}^0, \omega_{T-1}^0)\right| = 0.$$

Therefore

$$\begin{aligned}
&\left|C_{T-1}^k(x_{T-2}^k, \omega_{T-1}^0) - Q_{T-1}(x_{T-2}^0, \omega_{T-1}^0)\right| \\
\le\ &\left|C_{T-1}^k(x_{T-2}^k, \omega_{T-1}^0) - Q_{T-1}(x_{T-2}^k, \omega_{T-1}^0)\right| + \left|Q_{T-1}(x_{T-2}^k, \omega_{T-1}^0) - Q_{T-1}(x_{T-2}^0, \omega_{T-1}^0)\right| \\
\le\ &\left|\Delta_1^k\right| + \left|Q_{T-1}(x_{T-2}^k, \omega_{T-1}^0) - Q_{T-1}(x_{T-2}^0, \omega_{T-1}^0)\right| \\
\longrightarrow\ &0 \quad (\text{as } k \to \infty).
\end{aligned}$$

This means that the sequence $\{C_{T-1}^k(x_{T-2}^k, \omega_{T-1}^0)\}_{k\in J}$ converges to $Q_{T-1}(x_{T-2}^0, \omega_{T-1}^0)$ wp1. □

Lemma 9 *Suppose MSBD satisfies CSP and SIP. For any given t* $(1 \leq t \leq T-1)$, *and any given set* $K \subseteq \mathbb{N}$, *suppose*

- $\omega_t^k = \omega_t^0$ *for some given* ω_t^0 *for any* $k \in K$;
- *the sequence* $\{x_{t-1}^k\}_{k \in K}$ *converges to some given vector* x_{t-1}^0.

Then there exists an infinite set $J \subseteq K$ *such that*

(a) the sequence $\{x_t^k\}_{k \in J}$ *converges to some vector* x_t^0 *wp1;*

(b) the sequence $\{\theta_{t+1}^k\}_{k \in J}$ *converges to* $\mathcal{Q}_{t+1}(x_t^0)$ *wp1;*

(c) the sequence $\{C_t^k(x_{t-1}^k, \omega_t^0)\}_{k \in J}$ *converges to* $Q_t(x_{t-1}^0, \omega_t^0)$ *wp1.*

Proof: The lemma is proved by induction on t. When $t = T-1$, this lemma is exactly Lemma 8 and hence holds. Suppose the lemma holds for t, then we need to prove it for $t-1$. Therefore, assume now that for a given set $K \subseteq \mathbb{N}$ $\omega_{t-1}^k = \omega_{t-1}^0 \ \forall k \in K$, and $\{x_{t-2}^k\}_{k \in K} \to x_{t-2}^0$.

(a)

In iteration $k \in K$, the algorithm solves the problem $[AP_{t-1}^k]$ with $x_{t-2} = x_{t-2}^k$ and $\omega_{t-1} = \omega_{t-1}^0$, and gets the solution (x_{t-1}^k, θ_t^k). Since the feasible set is bounded, the sequence $\{x_{t-1}^k\}_{k \in K}$ has a convergent subsequence. Denote the corresponding limit as x_{t-1}^0 and the corresponding index set as L.

Now the set J is constructed in a way that the induction hypothesis can be applied. As shown in Lemma 5 for each $i = 1, \ldots, q_t$, the set L has wp1 an infinite subsequence L_i, such that for $\omega_t^k = \omega_{ti}$ and $\omega_{ti} \in \Omega_t^k$ for all $k \in L_i$.

For each $i = 1, \ldots, q_t$, by the induction assumption that the lemma holds for stage t and by the facts that $\omega_t^k = \omega_{ti}$ for all $k \in L_i$, and that the sequence $\{x_{t-1}^k\}_{k \in L}$ (and hence the sequence $\{x_{t-1}^k\}_{k \in L_i}$) converges to some vector x_{t-1}^0, there must exist an infinite subset J_i of L_i, for each $i = 1, \ldots, q_t$, such that the sequence $\{C_t^k(x_{t-1}^k, \omega_{ti})\}_{k \in J_i}$ converges to $Q_t(x_{t-1}^0, \omega_{ti})$ wp1. Therefore with $k \in J_i$:

$$\lim_{k \to \infty} \left| C_t^k(x_{t-1}^k, \omega_{ti}) - Q_t(x_{t-1}^0, \omega_{ti}) \right| = 0. \tag{2}$$

Define $J = \cup_{i=1}^{q_t} J_i$. Clearly, $J \subseteq L$, hence the sequence $\{x_{t-1}^k\}_{k \in J}$ converges to x_{t-1}^0 wp1.

(b) and (c)

From Lemma 6 and Lemma 7 we have for $k \in J$:

$$\mathcal{Q}_t(x_{t-1}^k) \geq \theta_t^k \geq \sum_{i=1}^{q_t} p_{it}\ C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) + \Delta^k \quad \text{with } \lim_{k \to \infty} \left| \Delta^k \right| = 0,$$

where the r_i are elements of $J_i = J \cap \{k | \ \omega_{ti} \in \Omega_t^k, \omega_t^k = \omega_{ti}\}$. This is equivalent to

$$\mathcal{Q}_t(x_{t-1}^k) \ \geq \ \theta_t^k \ \geq \ \mathcal{Q}_t(x_{t-1}^k) + \sum_{i=1}^{q_t} p_{it} \ \left[C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) - Q_t(x_{t-1}^k, \omega_{ti}) \right] + \Delta^k,$$

whence

$$\left|\theta_t^k - \mathcal{Q}_t(x_{t-1}^k)\right| \leq \left|\Delta_1^k\right|,$$

with

$$\Delta_1^k = \sum_{i=1}^{q_t} p_{it} \left[C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) - Q_t(x_{t-1}^k, \omega_{ti})\right] + \Delta^k.$$

Now

$$\begin{aligned}
\left|\Delta_1^k\right| &\leq \sum_{i=1}^{q_t} p_{it} \left|C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) - Q_t(x_{t-1}^k, \omega_{ti})\right| + \left|\Delta^k\right| \\
&\leq \sum_{i=1}^{q_t} p_{it} \left\{\left|C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) - Q_t(x_{t-1}^0, \omega_{ti})\right| + \left|Q_t(x_{t-1}^0, \omega_{ti}) - Q_t(x_{t-1}^k, \omega_{ti})\right|\right\} + \left|\Delta^k\right|.
\end{aligned}$$

If $k \to \infty$, then $r_i \to \infty$, and from (2) we have for $r_i \in J_i$:

$$\lim_{k\to\infty} \left|C_t^{r_i}(x_{t-1}^{r_i}, \omega_{ti}) - Q_t(x_{t-1}^0, \omega_{ti})\right| = 0 \quad (wp1).$$

Furthermore, due to continuity of $Q_t(x_{t-1}, \omega_t)$ in x_{t-1}, and convergence of the sequence $\left\{x_{t-1}^k\right\}_{k\in J} \to x_{t-1}^0$:

$$\lim_{k\to\infty} \left|Q_t(x_{t-1}^0, \omega_{ti}) - Q_t(x_{t-1}^k, \omega_{ti})\right| = 0.$$

Therefore $\lim_{k\to\infty} \left|\Delta_1^k\right| = 0 \quad (wp1)$, and so

$$\left|\theta_t^k - \mathcal{Q}_t(x_{t-1}^k)\right| \to 0 \text{ as } k \to \infty \quad (wp1).$$

Continuity of $\mathcal{Q}_t$ gives

$$\lim_{k\to\infty} \left|\mathcal{Q}_t(x_{t-1}^k) - \mathcal{Q}_t(x_{t-1}^0)\right| = 0,$$

and therefore wp1:

$$\left|\theta_t^k - \mathcal{Q}_t(x_{t-1}^0)\right| \longrightarrow 0 \quad (\text{as } k \to \infty).$$

Hence the sequence $\left\{\theta_t^k\right\}_{k\in J}$ converges to $\mathcal{Q}_t(x_{t-1}^0)$ wp1, which shows part (b).

Now, for (c), using the same argument as in the proof of Lemma 8, we obtain

$$\left|C_{t-1}^k(x_{t-2}^k, \omega_{t-1}^0) - Q_{t-1}(x_{t-2}^k, \omega_{t-1}^0)\right| \leq \left|\Delta_1^k\right|,$$

and by continuity

$$\lim_{k\to\infty} \left|Q_{t-1}(x_{t-2}^k, \omega_{t-1}^0) - Q_{t-1}(x_{t-2}^0, \omega_{t-1}^0)\right| = 0.$$

This yields wp1:

$$\left|C_{t-1}^k(x_{t-2}^k, \omega_{t-1}^0) - Q_{t-1}(x_{t-2}^0, \omega_{t-1}^0)\right| \longrightarrow 0 \quad (\text{as } k \to \infty),$$

which means that the sequence $\left\{C_{t-1}^k(x_{t-2}^k, \omega_{t-1}^0)\right\}_{k\in J}$ converges to $Q_{t-1}(x_{t-2}^0, \omega_{t-1}^0)$ wp1. □

Theorem 10 *Suppose MSBD satisfies CSP and SIP. The sequence of the solution values* $\left\{C_1^k\right\}_{k\in\mathbb{N}}$ *of the problem* $[AP_1^k]$ *converges to* Q_1 *wp1.*

Proof: In the approximated first-stage problem $[AP_1^k]$ the constraint $A_1x_1 = b_1$ can be formulated as $A_1x_1 = \omega_1 - B_0x_0$ with $\omega_1 \equiv b_1$, $x_0 \equiv 0$ and any given B_0. The value C_1^k can be seen as a trivial function $C_1^k(x_0, \omega_1)$. The result then follows from Lemma 9 equivalently to [3], Theorem 5.1. □

Theorem 11 *Suppose MSBD satisfies CSP and SIP. Then any accumulation point of the sequence* $\left\{x_1^k\right\}_{k\in\mathbb{N}}$ *is an optimal solution of the problem* $[LP_1]$ *wp1.*

Proof: See [3], Theorem 5.2. □

4 Convergent sampling algorithms

The class of algorithms described in the previous sections is quite general and includes a wide range of approaches. One subclass (which includes CUPPS) follows a forward pass for both getting candidate solutions and generating cuts for the next iteration. An alternative (such as SDDP) generates cuts in a backward pass. That means that while generating the cuts for stage t in iteration k, the cuts that were generated for stage $t+1$ in iteration k are already taken into account (which may lead to an improvement in the speed of convergence). Under CSP and SIP, the convergence result above remains the same for the backward-pass algorithms, if one observes that the sampled cuts should be modified so that one cut constraint additional to $[AP_t^k]$ is considered when obtaining the dual variables $(\pi_t^i(x_{t-1}^k), \rho_t^i(x_{t-1}^k))$.

The analysis of the previous section considers algorithms which use only one path of the tree per iteration. The class of algorithms can be extended to the *multipath* case of n_k sample paths, whereby in iteration k there are n_k paths sampled. If, say, the last of the paths is sampled independently from stage to stage, then this can be thought of as a single iteration of the algorithm with one path with (possibly) extra cuts added on the backward pass.

In fact the following *general multipath* scheme is possible: in iteration k a candidate solution for the first stage is determined and a cut for stage 1 is generated. In the second stage n_2^k scenarios are sampled. Then of the n_2^k scenarios s_2^k are chosen, and candidate solutions x_2^k are determined. Of these s_2^k candidate solutions, c_2^k are chosen at which the algorithm will generate cuts for stage 2. For each of the s_2^k candidate solutions, n_3^k samples are considered at stage 3. Then of the $s_2^k \cdot n_3^k$ samples s_3^k samples are chosen, and of these c_3^k cuts are generated for stage 3 etc.

Therefore n_t^k is the number of samples in stage t for each sample s_{t-1}^k of stage $t-1$, s_t^k the number of samples of which to proceed to the next stage, and c_t^k the number of cuts generated for stage t. The following relations hold:

$$\begin{aligned} s_{t-1}^k n_t^k \;\geq\; s_t^k \;\geq\; c_t^k &\;\geq\; 0 \qquad \forall t \geq 2, \forall k, \\ n_t^k &\;\geq\; 1 \qquad \forall t, \forall k, \\ s_1^k &\;=\; 1 \qquad \forall k. \end{aligned}$$

Observe that not all possible choices of these parameters satisfy the condition CSP. For example, if for some $t \geq 2$, $c_t^k = 0 \ \forall k$, then there is no guarantee of convergence.

We conclude this section by showing how MSBD algorithms from the literature fit into this framework. The results are summarized in Table 1.

Example 1 ***SDDP***
Stochastic Dual Dynamic Programming (SDDP) was introduced in [6]. In SDDP n scenario paths are sampled in each iteration. In a forward pass, for each stage in each scenario a candidate solution is calculated by solving $[AP_t^k]$. Then, in a backward pass, in each stage t the entire single-period subtree ($\Omega_t^k = \Omega_t$) is solved and a cut is generated for stage $t-1$. Thus SDDP is a multipath scheme with $n_k = n \quad \forall k$.

Example 2 ***CUPPS***
The Convergent Cutting-Plane and Partial Sampling (CUPPS) algorithm is given in [3]. In each iteration it samples only one scenario ($\Omega_t^k = \{\omega_t^k\}$). Calculating candidate solutions and generating cuts are both performed in the forward pass.

Example 3 ***AND***
The Abridged Nested Decomposition (AND) is described in [4]. Like SDDP $\Omega_t^k = \Omega_t$, and like SDDP it involves sampling in the forward pass, but the main difference is that it does not proceed forward from all solutions of the realizations sampled in each stage. Instead, in each stage n_t^k successors are sampled as in the general multipath scheme. Of these nodes, $s_t^k \leq n_t^k$ nodes are sampled from which to proceed, and $c_t^k = s_t^k \quad \forall k \ \forall t$.

Example 4 ***ReSa***
The Reduced Sampling method (ReSa) was developed in [5]. The basic structure is the same as in SDDP. First some scenarios of the tree are sampled. The difference from SDDP lies in the backward pass. In ReSa, in each stage the subtrees are solved to generate a cut only for some (randomly chosen) scenarios. Therefore fewer subproblems have to be solved than for SDDP, but one also gets fewer cuts per iteration. ReSa is a general multipath scheme, with $n_t^k = s_t^k \geq c_t^k$. The cuts are generated by solving entire single-period subtrees ($\Omega_t^k = \Omega_t$). If $c_t^k > 0$ for all t and all k sufficiently large then ReSa satisfies the cut-sampling property and so converges wp1.

Algorithm	Forward / Backward	Sampled Cut	Multipath scheme
SDDP	B	$\Omega_t^k = \Omega_t$	$n_k = n \quad \forall k$
CUPPS	F	$\Omega_t^k = \{\omega_t^k\}$	–
AND	B	$\Omega_t^k = \Omega_t$	$s_t^k = c_t^k \quad \forall k \ \forall t$
ReSa	B	$\Omega_t^k = \Omega_t$	$n_t^k = s_t^k \quad \forall k \ \forall t$

Table 1: Examples of convergent algorithms

5 Conclusions

This paper presents a general convergence result for multi-stage stochastic Benders decomposition codes that use sampling. Although we make no assertions in this paper about the rate of convergence of these algorithms, this theory provides some guidance for researchers who select parameters to tune these algorithms. To ensure convergence wp1 in ReSa, for example, one should ensure in accordance with the cut-sampling property that the algorithm (eventually) computes at least one cut for each stage in the backward pass.

A key restriction on the algorithms we study is the sample-intersection property, which guarantees that some proportion of random outcomes obtained by sampling moving forwards in time are chosen for cut calculation. It is not hard to see why such a condition might be needed. Certainly one can conceive of (perverse) algorithms that only compute cuts when the x_t^k values lie in certain subsets of the feasible region of $[AP_1^k]$. Since the optimal solution might have x_t lying outside this subset, there is no reason to suppose that the algorithm would converge wp1 even if the subsets where cuts are computed are visited infinitely often.

The convergence proof above uses a bound on the optimal dual variables for $[AP_t^k]$, that comes from their construction as extreme-point solutions. A possible extension is to allow the calculation of cuts for $\omega_{ti} \in \Omega_t^k$ to be inexact, in the sense that the dual variables (π_t^k, ρ_t^k) are computed to be within ϵ_t^k of optimality (see [8]). Under the assumption of a bounded dual feasible region for each problem $[AP_t^k]$, it is easy to extend our results to show convergence under the sampling property wp1 if $\epsilon_t^k \to 0$ for each t.

Acknowledgement The authors are grateful to the anonymous referees for their comments that have improved the exposition of this paper.

References

[1] J.R. Birge and F. Louveaux, Introduction to Stochastic Programming, Springer Verlag, New York, 1997.

[2] G.R. Grimmett and D.R. Stirzaker, Probability and Random Processes, Oxford University Press, 1992.

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

[4] C. J. Donohue, Stochastic Network Programming and the Dynamic Vehicle Allocation Problem, *Ph.D. Dissertation, University of Michigan*, 1996.

[5] M. Hindsberger and A.B. Philpott, ReSa: A Method for Solving Multi-stage Stochastic Linear Programs, *submitted to EJOR*, 2003.

[6] M.V.F. Pereira, L.M.V.G. Pinto, Multi-Stage Stochastic Optimization Applied to Energy Planning, *Mathematical Programming 52*, 359-375, 1991.

[7] Power Systems Research homepage, http://www.psr-inc.com.br/sddp.asp

[8] G. Zakeri, A.B. Philpott, and D.M. Ryan, Inexact Cuts in Stochastic Benders Decomposition, *SIAM J. on Optimization*, 10,3, 643-657, 2000.