BOUNDS AND APPROXIMATIONS FOR MULTISTAGE STOCHASTIC PROGRAMS *

FRANCESCA MAGGIONI[†] AND GEORG CH. PFLUG[‡]

Abstract. Consider (typically large) multistage stochastic programs, which are defined on scenario trees as the basic data structure. It is well known that the computational complexity of the solution depends on the size of the tree, which itself increases typically exponentially fast with its height, i.e. the number of decision stages. For this reason approximations which replace the problem by a simpler one and allow bounding the optimal value are of importance.

In this paper we study several methods to obtain lower and upper bounds for multistage stochastic programs and demonstrate their use in a multistage inventory problem.

Key words. multistage stochastic programming, bounds, refinement chain, expected value problem.

AMS subject classifications. 90C15, 90C90, 65K05

1. Introduction. For very large multistage stochastic programs, the solution to optimality may be quite tedious and requires often high performance or parallel software to be carried out. Bounding techniques allow to get quick information about lower and upper bounds for the objective value. Given that the gap between these two bounds is acceptably small, one may even stop and never fully optimize.

The main general principle for bounding minimization problems is relaxation of some constraints to obtain lower bounds and inserting feasible solutions to get upper bounds. Many authors have found bounds along these lines, see e.g. [14], [13], [12]. Additional bounds can be found by methods of convex analysis (Jensen's inequality, Edmundson-Madansky inequality), see e.g. [6], [7], [8], [9]. While the latter approaches consider the approximation of continuous state and/or continuous time models by discretization, we assume in this paper that a discretized (large) tree model is already given and study alternative ways to find easy-to-compute bounds. These bounds are found by solving many much smaller problems instead of one big problem.

The idea of looking at subproblems with only two scenarios goes back to [2] for the two-stage linear case, see also [3] and [18]. An extension to the multistage linear case is in [11].

In this paper we further extend the scope of the bounds by considering multistage convex problems with concave risk functionals as objective and give full sequences of bounds. We construct new refinement chains of lower bounds, where each bound can be computed by solving sets of group subproblems less complex than the original one, and dissecting the probability accordingly. Refinement chains are constructed principally in two ways: by keeping one or several scenarios fixed in all subproblems or choosing them disjoint. A monotonically nondecrasing behavior in the cardinality of scenarios of each subproblem is proved.

Lower bounds for replacing the scenario process by its expectation are also considered. Several upper bounds by inserting feasible solution from smaller problems are presented generalizing the *Multistage Expected Value of the Reference Scenario*, *MEVRS* and the *Multistage Expectation of Pairs Expected Value*, *MEPEV* introduced in [11] with an optimal first-stage solution of a subproblem of smaller size. Relations among them are derived.

Notice that the proposed decomposition gets its justification by the fact that the complexity of a solution method for tree structured problems increase superlinearly with the problem size, even sometimes exponentially. Furthermore, the proposed approach has the advantage to divide a given problem into independent subproblems which may be solved on a parallel machine architecture. Consequently, multistage problems, which are typically computationally complex and often not solvable by commercial solvers, can be handled by the proposed bounding technique. If it turns out that the calculation of lower bounds in the refinement chain is quick, then one may possibly find a good upper bound resulting in a solution gap which is small enough to stop the procedure here and not solve the big problem at all.

In the following, we assume that the stochastic program $Opt(\mathbb{P})$ is of the form

$$Opt(\mathbb{P}): \quad v^*(\mathbb{P}) = \min\{\mathcal{R}_{\mathbb{P}}[H(x,\xi)] \,:\, x \in \mathbb{X}, x \lhd \mathfrak{F}; \mathbb{P} \sim (\Omega, \mathfrak{F}, P, \xi)\},^1 \qquad (1.1)$$

where Ω is a finite, but possibly large probability sample space, $\mathfrak{F}$ is a filtration on Ω, i.e. an increasing sequence of sigma-algebras

$$\mathfrak{F} = (\mathcal{F}_0 = (\Omega, \emptyset), \mathcal{F}_1, \dots, \mathcal{F}_T = 2^{\Omega}), \qquad (1.2)$$

$\xi = (\xi_0, \dots, \xi_T)$ is a stochastic process adapted to the filtration $\mathfrak{F}$ (in symbol $\xi \lhd \mathfrak{F}$), P the probability distribution on $\mathfrak{F}$ and $\mathbb{X}$ the feasible set in which all decisions must lie, i.e. $x(\omega) \in \mathbb{X}$ for all $\omega \in \Omega$. The structure $(\Omega, \mathfrak{F}, P, \xi)$ is summarized by the symbol $\mathbb{P}.^2$ $H(x,\xi)$ is the cost function and $\mathcal{R}$ is a risk functional, like the expectation or the Average Value-at-Risk.

The finite probability space Ω endowed with the filtration $\mathfrak{F}$ of length $T+1$ as in (1.2) defines a finite tree of height T with leaf nodes Ω. The probability P defined on Ω induces probabilities on all nodes $\mathfrak{N}$ of the tree. The distance of every node from the root is its *stage*. The nodes at stage t are denoted by $\mathfrak{N}_t$, where $\mathfrak{N}_0$ is a singleton (the root) and $\mathfrak{N}_T = \Omega$ (the leaves). The stochastic scenario process ξ can be seen as defined on the node set $\mathfrak{N}$.

We exploit the special structure of the problem $Opt(\mathbb{P})$. For a simpler presentation we often omit some components of the model $\mathbb{P} \sim (\Omega, \mathfrak{F}, P, \xi)$ and write for instance just $\mathbb{P} \sim (\mathfrak{F}, P)$, if both the probability space Ω and the random process ξ are fixed and no confusion is possible.

The paper is organized as follows: lower bounds are introduced in Sections 2 and 3 by dissecting the probability measure and filtration and by replacing the scenario process by its expectation respectively. Upper bounds by inserting sub-optimal solutions are analyzed in 4. Lower and upper bounds introduced in the previous sections are then tested on a multistage inventory problem in section 5. Section 6 deals with complexity considerations and 7 concludes the paper.

---

[1] We assume throughout the paper that all optimal values are attained so that we write min and not inf

[2] $\mathbb{P}$ describes the equivalence class of all models which can be bijectively mapped to the structure $(\Omega, \mathfrak{F}, P, \xi)$. This equivalence class is called the nested distribution, see [16]. However, for reading this paper one may always assume that a concrete model $(\Omega, \mathfrak{F}, P, \xi)$ is given.

2. Lower bounds by dissecting the probability measure. Given the optimal value mapping $(\mathfrak{F}, P, \xi) \sim \mathbb{P} \longmapsto v^*(\mathbb{P}) = v^*(\mathfrak{F}, P, \xi)$ we keep the filtration and the process ξ fixed and consider only the mapping

$$P \mapsto v^*(P).$$

One main structural property of some stochastic programs (including those which are of the expectation type) is that this mapping is *concave*.

LEMMA 2.1. *Suppose that in the basic problem (1.1) the functional* $P \longmapsto \mathcal{R}_P(\cdot)$ *is concave (i.e. the mapping* $P \mapsto \mathcal{R}_P(Y)$ *is concave for all random variables* Y *for which* $\mathcal{R}$ *is defined, see [15]). Then the mapping* $P \mapsto v^*(P)$ *is also concave.*

Proof. Let $P = \sum_{i=1}^{k} p_i P_i$ with $\sum p_i = 1$, $p_i > 0$ and let the x^* be a solution of problem $Opt(P) = Opt(\mathfrak{F}, P, \xi)$. Since x^* is feasible for all $Opt(P_i)$ we have that

$$v^*(P) = \mathcal{R}_P[H(x^*, \xi)] \geq \sum p_i \mathcal{R}_{P_i}[H(x^*, \xi)] \geq \sum p_i v^*(P_i).$$

□

Notice that the Lemma holds for any feasible set $\mathbb{X}$. In particular it holds for integer or mixed integer problems.

Numerous consequences result from this simple Lemma 2.1. For instance, let $\mathfrak{N}_T = \Omega = \{\omega_1, \ldots, \omega_k\}$ be the leaf set of the finite tree. Let δ_{ω_i} be the point mass at scenario ω_i. Under δ_{ω_i} the scenario ω_i has probability 1 and all other scenarios have probability zero. Thus solving the problem under δ_{ω_i} leads to a deterministic dynamic optimization problem, with optimal value denoted by $v^*(\delta_{\omega_i})$. Notice that

$$P = \sum_{i=1}^{k} P(\omega_i)\delta_{\omega_i}. \tag{2.1}$$

The value $v_1^* := \sum_i P(\omega_i)\, v^*(\delta_{\omega_i})$ is the *clairvoyant's solution*, which may alternatively be written as

$$v_1^* = v^*(\mathfrak{F}^T, P),$$

where $\mathfrak{F}^T$ is the *clairvoyant's filtration* $\mathfrak{F}^T = (\mathcal{F}_T, \ldots, \mathcal{F}_T)$. By (2.1) and Lemma 2.1 we get the (quite trivial) lower bound by the inequality

$$v_1^* = v^*(\mathfrak{F}^T, P) \leq v^*(\mathfrak{F}, P). \tag{2.2}$$

Some authors would call this inequality

$$\text{wait-and-see} \leq \text{here-and-now}$$

which is a bit misleading, since the essence of the "wait-and-see" solution is here not to wait, but to decide immediately with complete knowledge of the future, an ability which we may attribute only to clairvoyants.

Remark. The inequality

$$v^*(\mathfrak{F}, P) \geq v^*(\mathfrak{F}^T, P).$$

is just one example of the fundamental but quite trivial inequality

$$\mathfrak{F}_1 \subseteq \mathfrak{F}_2 \qquad \text{implies that} \qquad v^*(\mathfrak{F}_1, P) \geq v^*(\mathfrak{F}_2, P). \qquad (2.3)$$

This inequality can produce lower and upper bounds: by refining the filtration, i.e. relaxing the nonanticipativity constraints, one gets lower bounds and by coarsening it, one gets upper bounds.

Refinement Chains. While the dissection of P into its atoms (δ_{ω_i}) is the most extreme dissection, one may also dissect it into a convex combination of probabilities defined on two ore more, but not all leaves. For instance, let ω_1 be a specific leaf node (w.l.o.g. call it ω_1) and set for $i \neq 1$

$$P_i^{(2)} = P(\omega_1)\delta_{\omega_1} + (1 - P(\omega_1))\delta_{\omega_i}. \qquad (2.4)$$

Defining $\pi_i^{(2)} = P(\omega_i)/(1 - P(\omega_1))$ one sees that $P = \sum_{i\neq 1} \pi_i^{(2)} P_i^{(2)}$ and hence

$$v_2^* := \sum_{i\neq 1} \pi_i^{(2)} v^*(P_i^{(2)}) \leq v^*(P). \qquad (2.5)$$

Thus a lower bound for the basic problem can be found by solving $k - 1$ multistage problems with only 2 scenarios each. Since by (2.4) every subproblem with two scenarios is a convex combination of two (clairvoyant's) problems with just one scenario each, i.e. one gets for $i \neq 1$

$$v^*(P_i^{(2)}) \geq P(\omega_1)\,v^*(\delta_{\omega_1}) + (1 - P(\omega_1))\,v^*(\delta_{\omega_i}),$$

and one arrives at a chain of inequalities

$$v_1^* = \sum_{i=1}^{k} P(\omega_i)\,v^*(\delta_{\omega_i}) \leq v_2^* \leq v^*(P).$$

LEMMA 2.2. *If* $\Omega = \{\omega_2, \ldots, \omega_k\}$ *and* $\omega_1 \notin \Omega$ *then* $v_1^* = v_2^*$.
Proof. See [11]. □

Pairs of scenarios have been considered for the first time in [2], also in [3] and [18] for the two-stage linear case and in [11] for the multistage linear case by means of the definition of *Multistage Sum of Pairs Expected Values MSPEV*. Inequality (2.5) has been proven in [11] for the multistage linear case with risk functional being the expectation.

The method may be refined by considering longer *refinement chains*. Such a chain is of the structure

$$\Omega$$
$$(\Omega_1^{(\ell)}, \Omega_2^{(\ell)}, \ldots, \Omega_{m_\ell}^{(\ell)})$$
$$\vdots$$
$$(\Omega_1^{(2)}, \Omega_2^{(2)}, \ldots, \Omega_{m_2}^{(2)})$$
$$(\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_k\}),$$

where each row is a collection of subsets of the probability space Ω with the property that their union covers the whole space $\Omega = \cup_i \Omega_i^{(j)}$ for all j and that each set $\Omega_i^{(j)}$ is the union of sets from the next more refined collection

$$\Omega_i^{(j)} = \cup_{\Omega_s^{(j-1)} \subseteq \Omega_i^{(j)}} \Omega_s^{(j-1)}.$$

To a refinement chain of the probability space Ω there corresponds a chain of dissections of the probability P into probability measures $P_i^{(j)}$

$$
\begin{gathered}
P \\
(P_1^{(\ell)}, \dots P_{m_\ell}^{(\ell)}) \\
\vdots \\
(P_1^{(2)}, \dots, P_{m_2}^{(2)}) \\
(P_1^{(1)} = \delta_{\omega_1}, \dots, P_k^{(1)} = \delta_{\omega_k}),
\end{gathered}
\tag{2.6}
$$

such that
  (i) $P_i^{(j)}$ has support $\Omega_i^{(j)}$
  (ii) P can be written as $P = \sum_{i=1}^{m_j} \pi_i^{(j)} P_i^{(j)}$
  (iii) each $P_i^{(j)}$ can be written as a convex combination of probabilities from the refined collection $\left\{ P_i^{(j-1)} \right\}$.
It is evident that given such a refinement chain leads to a chain of lower bounds. Denoting

$$v_j^* = \sum_{i=1}^{m_j} \pi_i^{(j)} v^*(P_i^{(j)}),$$

one arrives by Lemma 2.1 as well as (ii) and (iii) at a chain of lower bounds

$$v_1^* \leq v_2^* \leq \cdots \leq v_\ell^* \leq \cdots \leq v^*(P). \tag{2.7}$$

Notice that the higher the index j, the fewer problems have to be solved but with an increasing number of scenarios. Notice also that the clairvoyant solution gives always the smallest lower bound in this chain.

There are many ways to construct refinement chains. One may keep one or several scenarios fixed in all subsets $\Omega_i^{(j)}$ or choose them disjoint. Two examples will illustrate possible choices.

EXAMPLE 1. *One fixed scenario. Suppose that scenario ω_1 is fixed and appears in all subsets. Then the structure of the refinement chain is*

$$
\begin{gathered}
(\{\omega_1, \omega_2, \dots, \omega_k\}) \\
(\{\omega_1, \omega_2, \dots, \omega_s\}, \{\omega_1, \omega_{s+1}, \dots, \omega_k\}) \\
\vdots \\
(\{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \dots, \{\omega_1, \omega_k\}).
\end{gathered}
$$

The situation is illustrated in Figure 2.1. All scenarios which belong to a subset $\Omega_i^{(j)}$ are connected by edges and form a connected subgraph.
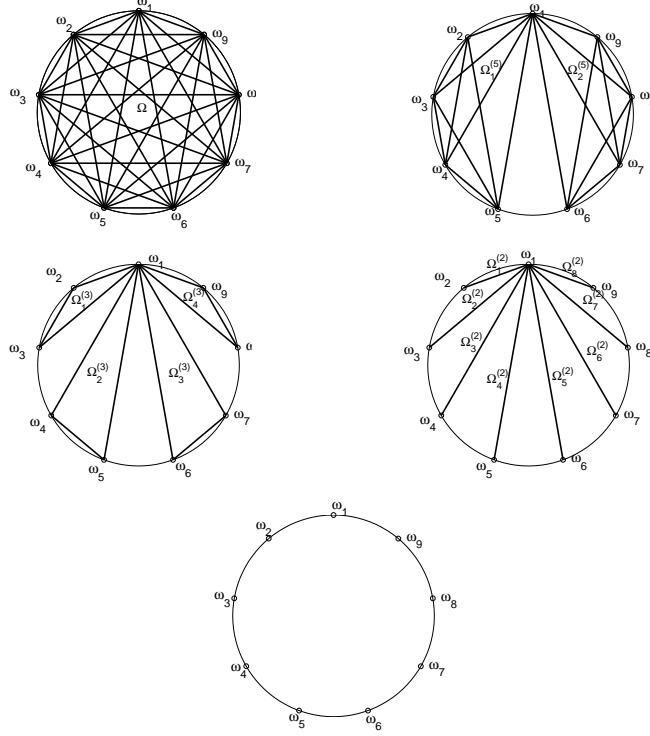
FIG. 2.1. *A graphical representation of a refinement chain with just the scenario ω_1 fixed. The collection $(\Omega_1^{(2)}, \ldots, \Omega_8^{(2)})$ is the "pair of scenarios" collection (with ω_1 fixed). Similarly the collection $(\Omega_1^{(3)}, \ldots, \Omega_4^{(3)})$ is the "triple of scenarios" collection (with ω_1 fixed) and $(\Omega_1^{(5)}, \Omega_2^{(5)})$ for quintuple of scenarios.*

As a generalization, we consider the case that two or more scenarios may be fixed, i.e. they appear in all subsets. In order to calculate the correct weights $\pi_i^{(\cdot)}$ of the dissections, assume w.l.o.g. that the first f scenarios $\{\omega_1, \ldots, \omega_f\}$ are fixed and appear in all subsets. Assume further that each subset contains exactly $j > f$ scenarios, meaning that the total number s of subsets of cardinality j, i.e. $s = (k-f)/(j-f)$ is an integer. Let $x = j - f > 0$. The corresponding s elements of the refinement chain are

$$\Omega_1^{(\cdot)} = \{\omega_1, \ldots, \omega_f, \omega_{f+1}, \ldots, \omega_{f+x}\}, \Omega_2^{(\cdot)} = \{\omega_1, \ldots, \omega_f, \omega_{f+x+1}, \ldots, \omega_{f+2x}\}, \ldots$$

or in general

$$\Omega_i^{(\cdot)} = \{\omega_1, \ldots, \omega_f, \omega_{f+(i-1)\cdot x+1}, \ldots, \omega_{f+i\cdot x}\}, \quad i = 1, \ldots, s.$$

The probabilities $P_i^{(\cdot)}$ as well as the corresponding weights $\pi_i^{(\cdot)}$ can be calculated as follows:

$$P_i^{(\cdot)} = \sum_{m=1}^{f} P(\omega_m) \cdot \delta_{\omega_m} + \frac{(1 - \sum_{m=1}^{f} P(\omega_m))}{\sum_{m=f+(i-1)\cdot x+1}^{f+i\cdot x} P(\omega_m)} \sum_{m=f+(i-1)\cdot x+1}^{f+i\cdot x} P(\omega_m) \cdot \delta_{\omega_m}$$
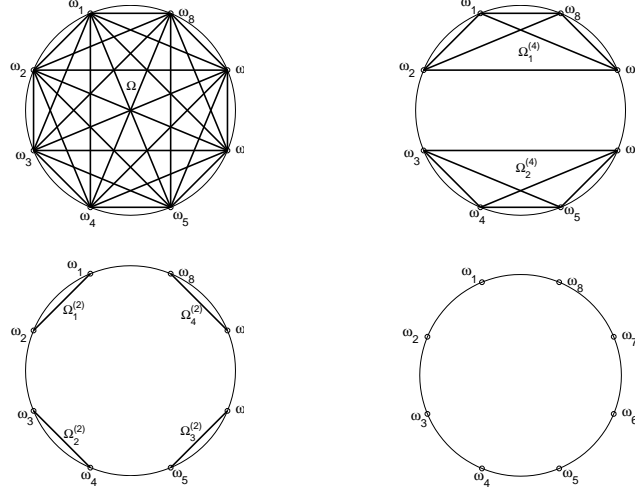
FIG. 2.2. *A refinement chain with disjoint subsets.*

and

$$\pi_i^{(\cdot)} = \frac{\sum_{m=f+(i-1)\cdot x}^{f+i\cdot x} P(\omega_m)}{(1 - \sum_{m=1}^{f} P(\omega_m))}, \quad i = 1, \ldots, s.$$

With this choice one gets $\sum_{m=1}^{s} \pi_i^{(\cdot)} = 1$ and $P = \sum_{i=1}^{s} \pi_i^{(\cdot)} P_i^{(\cdot)}$ and therefore the corresponding lower bound is

$$\sum_{i=1}^{s} \pi_i^{(\cdot)} v^*(P_i^{(\cdot)}) \leq v^*(P).$$

Notice that (2.5) is a special case with $f = 1$ and $j = 2$.

EXAMPLE 2. *Alternatively one may also consider disjoint partitions. In this case, the refinement chain corresponds to a filtration and the probability measures $P_i^{(j)}$ are*

$$P_i^{(j)} = \begin{cases} \frac{1}{\pi_i^{(j)}} P(\omega_s)\delta_{\omega_s} & \text{if } \omega_s \in \Omega_i^{(j)} \\ 0 & \text{otherwise} \end{cases}$$

with

$$\pi_i^{(j)} = \sum_{\omega_s \in \Omega_i^{(j)}} P(\omega_s).$$

Figure 2.2 illustrates such a refinement chain.

3. Lower bounds for replacing the scenario process by its expectation.
A popular choice for simplifying the dynamic stochastic problem is to replace the scenario process $(\xi_t)_{t=1,\ldots,T}$ by its expectation $(\mathbb{E}[\xi_t])_{t=1,\ldots,T}$, if it exists. The basic problem (1.1) is modified to

$$Opt(\mathbb{E}_P): \qquad \min\left\{\mathcal{R}[H(x, \mathbb{E}_P\xi)] \,:\, x \in \mathbb{X}; x \lhd \mathfrak{F}; \mathbb{P} = (\mathfrak{F}, P, \xi)\right\}. \qquad (3.1)$$

We call the problem (3.1) the expectation-reduced problem. Notice that we have replaced the stochastic scenario process by a deterministic one, but kept the possibility for the decisions to depend on the filtration. One might argue that such *random decisions for non-random scenarios* is not appropriate. A closer look reveals in fact that the optimal decision for deterministic compound concave problems can be chosen as deterministic as well, i.e. it is not necessary to consider randomized decisions.

THEOREM 3.1. *Suppose that the risk functional $\mathcal{R}_P$ is concave[3] in P. Then the solution of the expectation-reduced problem* (3.1) *can be found for a deterministic decision sequence* $(x_0^*, \ldots, x_{T-1}^*)$.

Proof. Notice that the expectation-reduced problem has a deterministic objective $\bar{H}(x) = H(x, \mathbb{E}\xi)$. Dissect the probability P into the atoms defined on the leaves $P = \sum_{i=1}^k P(\omega_i)\delta_{\omega_i}$. Then, by concavity, for any x, $\mathcal{R}_P[\bar{H}(x)] \geq \sum_{i=1}^k P(\omega_i)\mathcal{R}_{\delta_{\omega_i}}[\bar{H}(x)]$. Thus the optimal decisions can be found by solving the clairvoyant problems. But since all clairvoyant problems are equal, the same solution may be inserted for all scenarios, i.e. the solution may be chosen deterministic as well. □

Let $\mathbb{P}_{\mathbb{E}(\xi)}$ be the degenerated (nested) distribution, which assigns probability one to the sequence $\mathbb{E}\xi_1, \ldots \mathbb{E}\xi_T$. Under the assumption of Theorem 3.1, one may w.l.o.g. assume that the solution of the deterministic problem

$$v^*(\mathbb{P}_{\mathbb{E}(\xi)}) = \min\left\{\mathcal{R}_{\mathbb{P}_{\mathbb{E}(\xi)}}[H(x,\xi)] \; : \; x \in \mathbb{X}\right\} = \min\left\{\mathcal{R}[H(x, \mathbb{E}_P\xi)] \; : \; x \in \mathbb{X}\right\},$$

is deterministic.

The value $v^*(\mathbb{P}_{\mathbb{E}(\xi)})$ in the linear case is known in literature as *Expected Value, EV* (see for instance [2] and many others).

DEFINITION 3.2. *The random variable X_1 is dominated by the random variable X_2 in second-order sense and we write $X_1 \prec_{SSD} X_2$, where SSD stands for second-order stochastic dominance [4], if $\mathbb{E}[U(X_1)] \leq \mathbb{E}[U(X_2)]$ for all monotonically increasing and concave utility functions U.*

LEMMA 3.3. *Suppose that (i) $(x, \xi) \mapsto H(x, \xi)$ is convex, (ii) $P \mapsto \mathcal{R}_P$ is concave and (iii) $\mathcal{R} = \mathcal{R}_P$ has the following property*

$$-Y_1 \prec_{SSD} -Y_2 \quad \text{implies that} \quad \mathcal{R}(Y_1) \geq \mathcal{R}(Y_2).$$

Then, assuming that the solution of the basic problem (1.1) *is integrable and that the set of feasible solutions $\mathbb{X}$ is convex,*

$$v^*(\mathbb{P}_{\mathbb{E}(\xi)}) \leq v^*(\mathbb{P}). \tag{3.2}$$

Proof. Suppose that $\eta = x^*(\xi)$ is the solution of (1.1). For all monotonically increasing and concave utility functions U, the function $V(y) = -U(-y)$ is convex and monotonically increasing. Since $(\eta, \xi) \mapsto V(H(\eta, \xi))$ is convex, by Jensen's inequality $\mathbb{E}[V(H(\eta, \xi)] \geq V(H(\mathbb{E}(\eta), \mathbb{E}\xi))$. Let $x^+ = \mathbb{E}(\eta) = \mathbb{E}(x^*(\xi))$ (a deterministic function), then

$$\mathbb{E}[U(-H(x^*(\xi), \xi)] \leq U(-H(x^+, \mathbb{E}\xi))$$

———

[3]If the functional is expected disutility, i.e. $\mathcal{R}(Y) = \mathbb{E}[V(Y)]$, then it is affine linear and a fortiori concave in P.

and therefore by assumption $\mathcal{R}[H(x^*(\xi), \xi)] \geq \mathcal{R}[H(x^+, \mathbb{E}\xi)]$. By Theorem 1, the solution of the expectation-reduced can be found among deterministic functions and therefore

$$v^*(\mathbb{P}) = \mathcal{R}[H(x^*(\xi), \xi)] \geq \mathcal{R}[H(x^+, \mathbb{E}\xi)] \geq v^*(\mathbb{P}_{\mathbb{E}(\xi)})$$

which is the asserted relation. □

Remark. Notice that the functional $\mathcal{R}(Y) = \mathbb{E}[V(Y)]$ with a monotonically increasing and convex function V satisfies the conditions of the lemma, since $y \mapsto -V(-y)$ is monotonically increasing and concave.

EXAMPLE 3. *The convexity of $\xi \mapsto H(x, \xi)$ is for instance ensured in the following decision model with random costs c_t and random right hand side h_t in the linear constraints. Suppose that the problem is formulated as*

$$\min \{c_0(x_0) + \mathbb{E}_{\xi_1} [\min c_1(x_1, \xi_1) + \mathbb{E}_{\xi_2} [\min c_2(x_2, \xi_2) +$$
$$+ \mathbb{E}_{\xi_T} [\cdots + \mathbb{E} [\min c_T(x_T, \xi_T)] \ldots]]] : x \in \mathbb{X}\},$$

where the feasible set $\mathbb{X}$ is given by

$$W_0 x_0 \geq h_0$$
$$A_1 x_0 + W_1 x_1 \geq h_1(\xi_1)$$
$$A_2 x_1 + W_2 x_2 \geq h_2(\xi_2)$$
$$\vdots$$
$$A_T x_{T-1} + W_T x_T \geq h_T(\xi_T) \qquad (3.3)$$
$$x_1 \lhd \mathcal{F}_1$$
$$\vdots$$
$$x_T \lhd \mathcal{F}_T.$$

Introducing the indicator function

$$\psi_{\mathbb{X}}(x) = \begin{cases} 0 & \text{if } x \in \mathbb{X} \\ \infty & \text{otherwise} \end{cases}$$

one may identify the cost function $H(x, \xi)$ as

$$H(x, \xi) = c_0 x_0 + c_1(x_1, \xi_1) + \cdots + c_T(x_T, \xi_T) + \psi_{W_0 x_0 \geq h_0} + \psi_{A_1 x_0 + W_1 x_1 \geq h_1(\xi_1)}$$
$$+ \cdots + \psi_{A_T x_{T-1} + W_T x_T \geq h_T(\xi_T)}.$$

Notice that $\mathbb{E}[\psi_{A_t x_{t-1} + W_t x_t \geq h_t(\xi_t)}] < \infty$ only if $A_t x_{t-1} + W_t x_t \geq h_t(\xi_1)$ almost surely.

LEMMA 3.4. *If the functions $(x_t, \xi_t) \mapsto c_t(x_t, \xi_t)$ as well as the functions $\xi_t \mapsto h_t(\xi_t)$ are convex, then $(x, \xi) \mapsto H(x, \xi)$ is convex.*

Proof. This follows from the fact that $\{(x_t, x_{t-1}, \xi_t) : A_t x_{t-1} + W_t x_t \geq h_t(\xi_t)\}$ are convex sets. □

EXAMPLE 4. *If $c(x, \xi)$ is not convex in both variables, but just in x and ξ separately, no relation between the original problem and the expectation-reduced problem holds. As a simple example, take the problem*

$$\min\{\xi \cdot x : \xi \leq x \leq 1\}$$

with $\xi \sim Uniform[0,1]$. Then

$$\frac{1}{3} = \mathbb{E}\xi^2 = \mathbb{E}[\min\{\xi \cdot x : \xi \leq x \leq 1\}] < \min\{\mathbb{E}\xi \cdot x : \mathbb{E}\xi \leq x \leq 1\} = \frac{1}{4}$$

On the other hand, for the problem

$$\min\{-\xi \cdot x : -\xi \leq -x \leq 0\}$$

we get

$$-\frac{1}{3} = -\mathbb{E}\xi^2 = \mathbb{E}[\min\{-\xi \cdot x : -\xi \leq -x \leq 0\}] > \min\{-\mathbb{E}\xi \cdot x : -\mathbb{E}\xi \leq -x \leq 0\} = -\frac{1}{4}$$

Notice that $c(\xi) = \xi$ or $c(\xi) = -\xi$ and $h(\xi) = \xi$ or $h(\xi) = -\xi$ are convex in ξ, but $(x, \xi) \mapsto \xi \cdot x$ or $(x, \xi) \mapsto -\xi \cdot x$ are not convex in both variables jointly.

4. Upper bounds by inserting (sub)solutions. Solving a real-world stochastic program is often prohibitive from a computational point of view. Many problems have long planning horizon, which entails serious tractability issues. In such a cases, a typical way out is to solve simpler problems by passing to simpler stochastic tree models, and then fixing the decisions (starting from the root) in the full stochastic program. In these cases it is relevant for decision makers to know how well the solution of a simplified problem performs compared to the corresponding full-size problem. Even in the case the full stochastic program is not solvable, one may try to solve simpler problems for finding upper bounds and proceed to find tighter bounds, if the gap between the lower and the upper bound is too large.

In this section we revise and extend different procedures to derive upper bounds to the optimal objective function. In general upper bounds are obtained by constraining (some or all of) the decision variables to be equal to predetermined fixed values, like the expected value solution, the reference scenario solution or the optimal first stage decision of a group subproblem. From an algorithmic perspective, these approaches require to solve problems with smaller dimension than the original one, a computational advantage.

Denote by $v(\mathbb{P}, x)$ the value of the objective, when the decision x is inserted

$$v(\mathbb{P}, x) = \mathcal{R}_{\mathbb{P}}[H(x, \cdot)].$$

Since any decision x cannot be better than the optimal decision, we have

$$v(\mathbb{P}, x) \geq v^*(\mathbb{P}), \ x \in \mathbb{X}.$$

Inserting a fixed decision $x_{0:t-1}^+$ for times $0, \ldots, t-1$ as fixed and optimizing only the decisions for times $t, \ldots, T$ is denoted by

$$v^*(\mathbb{P}, x_{0:t-1}^+) = \min\left\{\mathcal{R}_{\mathbb{P}}[H([x_{0:t-1}^+, x_{t:T}], \cdot)] \ : \ x_{t:T} \lhd \mathfrak{F}_{t:T}\right\}. \tag{4.1}$$

The following relation holds true:

THEOREM 4.1.

$$v(\mathbb{P}, x_{0:T-1}^+) \geq v^*(\mathbb{P}, x_{0:t-1}^+) \geq v^*(\mathbb{P}). \tag{4.2}$$

Proof. Since the feasible region of $v(\mathbb{P}, x_{0:T-1}^+)$ is obviously smaller than the the one of $v^*(\mathbb{P}, x_{0:t-1}^+)$, $t = 0, \dots, T-2$, which in turn is smaller than the one of $v^*(\mathbb{P})$, the relation is obviously satisfied. □

Special cases of the concept discussed above are represented by the following class of measures as already introduced in [11] in the linear case.

DEFINITION 4.2. *By denoting $\bar{x}_{0:t-1}$, for times $0, \dots, t-1$, the optimal solution of the expectation-reduced problem (3.1), the Expected result at stage t of the Expected Value solution , EEV^t, $(t = 1, \dots, T)$ is defined as:*

$$EEV^t = v^*(\mathbb{P}, \bar{x}_{0:t-1}) = \min\left\{\mathcal{R}_{\mathbb{P}}[H([\bar{x}_{0:t-1}, x_{t:T}], \cdot)] \ : \ x_{t:T} \lhd \mathfrak{F}_{t:T}\right\}. \tag{4.3}$$

See in [5] for an alternative definition.

For the same reason of relation (4.2), we have

$$EEV^{t+1} \geq EEV^t \geq v^*(\mathbb{P}), \quad t = 1, \dots, T-1. \tag{4.4}$$

Similarly:

DEFINITION 4.3. *The Multistage Expected Value of the Reference Scenario, $MEVRS^t$, $(t = 1, \dots, T)$ defined as:*

$$MEVRS^t = v^*(\mathbb{P}, x_{0:t-1}^{+\omega_i}) = \min\left\{\mathcal{R}_{\mathbb{P}}[H([x_{0:t-1}^{+\omega_i}, x_{t:T}], \cdot)] \ : \ x_{t:T} \lhd \mathfrak{F}_{t:T}\right\}, \tag{4.5}$$

is obtained by inserting the optimal solution $x_{0:t-1}^{+\omega_i}$ until stage $t-1$ of the deterministic problem under a reference scenario $\omega_i \in \mathfrak{N}_T$.

It is worth to point out that the problems EEV^t and $MEVRS^t$, $t = 1, \dots, T$ could be infeasible since too many variables are fixed to their deterministic solution values. In this case no valid upper bounds can be obtained. However, even when the difference $v^*(\mathbb{P}, \bar{x}_{0:t-1}) - v^*(\mathbb{P})$ is large or in the extreme case infinite, the deterministic solution could still carries useful information for the stochastic one. This is evaluated by the *Mulstistage Expected Skeleton Solution Value $MESSV^t$, $t = 1, \dots, T$,* from [10] and [11] based on the following procedure:

1. fix at zero (or at the lower bound) all the variables which are at zero (or at the lower bound) in the expected value model;
2. solve the related stochastic program.

$MESSV^t$ is defined as follows:

DEFINITION 4.4. *By denoting $\bar{x}_{0:t-1}^0$, for stages $0, \dots, t-1$, the optimal solution variables which are at zero or at their lower bound in the expectation-reduced problem (3.1), the Multistage Expected Skeleton solution Value at stage t, $MESSV^t$, $(t = 1, \dots, T)$ is given by*

$$MESSV^t = v^*(\mathbb{P}, \bar{x}_{0:t-1}^0) = \min\left\{\mathcal{R}_{\mathbb{P}}[H([\bar{x}_{0:t-1}^0, x_{t:T}], \cdot)] \ : \ x_{t:T} \lhd \mathfrak{F}_{t:T}\right\}. \tag{4.6}$$

Again

$$MESSV^{t+1} \geq MESSV^t , \quad t = 1, \dots, T-1. \tag{4.7}$$

Such sequences of upper bounds allow to understand why deterministic solutions do not perform well. Sometimes the deterministic model selects the wrong positive variables and sometimes the choice of out-of-basis variables is fine, but the values are off.

First stage solutions can be also fixed from other types of simplified problems, like pair subproblems described before. This is the case of the *Multistage Expectation of Pairs Expected Value, $MEPEV$* as already introduced in [11] in the linear case.

DEFINITION 4.5. *Let* $x_{0:0}^{+i^{(2)}}$ *be the optimal first stage decision to the pair sub-problem* $v^*(P_i^{(2)})$, $i \neq 1$, *then MEPEV is defined as follows:*

$$
\begin{aligned}
MEPEV &:= \min_{i \neq 1} v^*(\mathbb{P}, x_{0:0}^{+i^{(2)}}) \\
&= \min_{i \neq 1} \min \left\{ \mathcal{R}_\mathbb{P}[H([x_{0:0}^{+i^{(2)}}, x_{1:T}], \cdot)] \; : \; x_{1:T} \lhd \mathfrak{F}_{1:T} \right\}.
\end{aligned}
\tag{4.8}
$$

MEPEV represents the minimum optimal value among those obtained by solving the original stochastic program (1.1), using the optimal first stage solution of each pair subproblem $v^*(P_i^{(2)})$, $i \neq 1$. As a consequence of (4.2), the following chain of inequalities hold true:

THEOREM 4.6.

$$
v^*(\mathbb{P}) \leq MEPEV \leq MEVRS^1 \leq \cdots \leq MEVRS^T.
\tag{4.9}
$$

Proof. See Theorem 4.1. □

MEPEV definition can be generalized by considering the optimal first stage decision to other subproblems $v^*(P_i^{(j)})$, $i \neq 1$, introduced in Section 2.

DEFINITION 4.7. *Let* $x_{0:0}^{+i^{(j)}}$ *be the optimal first stage decision to the subproblem* $v^*(P_i^{(j)})$, *then the Multistage Expectation of Subproblem Expected Value, MESEV^j is defined as follows:*

$$
\begin{aligned}
MESEV^j &:= \min_i v^*(\mathbb{P}, x_{0:0}^{+i^{(j)}}) \\
&= \min_i \min \left\{ \mathcal{R}_\mathbb{P}[H([x_{0:0}^{+i^{(j)}}, x_{1:T}], \cdot)] \; : \; x_{1:T} \lhd \mathfrak{F}_{1:T} \right\}.
\end{aligned}
\tag{4.10}
$$

MESEV^j represents the minimum optimal value among those obtained by solving the original stochastic program (1.1), using the optimal first stage solution of each subproblem $v^*(P_i^{(j)})$, $j \neq i$. For all these subproblems

$$
v^*(\mathbb{P}) \leq MESEV^j.
$$

Finally, we study the quality of the upper bound, if first stage decisions are fixed at the first stage solutions of a smaller subproblem defined below. To this end, we refer to the following well known Theorem.

THEOREM 4.8 (Pflug, Pichler, 2012).

Let $\mathbb{P} \sim (\Omega, \mathfrak{F}, P, \xi)$ *be a tree model with filtration* $\mathfrak{F}$ *and* $\tilde{\mathbb{P}} \sim (\tilde{\Omega}, \tilde{\mathfrak{F}}, \tilde{P}, \tilde{\xi})$ *be an alternative tree model of same height with filtration* $\tilde{\mathfrak{F}}$. *The nested distance between* $\mathbb{P}$ *and* $\tilde{\mathbb{P}}$ *is defined as*

$$
d(\mathbb{P}, \tilde{\mathbb{P}}) = \inf_\pi \iint \|\xi(\omega) - \tilde{\xi}(\tilde{\omega})\| \, \pi(\mathrm{d}\omega, \mathrm{d}\tilde{\omega})
\tag{4.11}
$$

where π *is a probability measure with (conditional) marginals* P *and* $\tilde{P}$, *i.e.,*

$$
\begin{aligned}
\pi\big(A \times \tilde{\Omega} \big| \mathfrak{F}_t \otimes \tilde{\mathfrak{F}}_t\big) &= P\big(A \big| \mathfrak{F}_t\big) \text{ and} \\
\pi\big(\Omega \times B \big| \mathfrak{F}_t \otimes \tilde{\mathfrak{F}}_t\big) &= \tilde{P}\big(B \big| \tilde{\mathfrak{F}}_t\big), \quad t = 0, \dots T,
\end{aligned}
$$

where $A \in \mathfrak{F}_T$ and $B \in \tilde{\mathfrak{F}}_T$ and $\mathfrak{F}_t \otimes \tilde{\mathfrak{F}}_t$ represents the product of sigma algebras. Consider the multistage stochastic optimization problem

$$v^*(\mathbb{P}) := \inf \left\{ \mathbb{E}_P H(\xi, x) : x \lhd \mathfrak{F}, x \in \mathbb{X} \right\}, \qquad (4.12)$$

where $\mathbb{X}$ is convex and L is convex in x for any ξ fixed, and Lipschitz with constant L in ξ for any x fixed. Let $\tilde{x}^*$ be a solution of

$$v^*(\tilde{\mathbb{P}}) := \inf \left\{ \mathbb{E}_{\tilde{P}} H(\xi, x) : x \lhd \tilde{\mathfrak{F}}, x \in \mathbb{X} \right\}, \qquad (4.13)$$

and let $x^+(\xi) = \mathbb{E}_\pi(\tilde{x}^*|\xi)$, where π is the optimizing measure is (4.11). Then x^+ is an approximate solution of (4.12) in the sense that

$$0 \le v(\mathbb{P}, x^+) - v^*(\mathbb{P}) \le L \cdot d(\mathbb{P}, \tilde{\mathbb{P}}).$$

Proof. See [17]. □

Based on this Theorem we may formulate the following Corollary.

COROLLARY 4.9. *Let the assumptions of the previous Theorem be fulfilled. Let $\tilde{x}_0^*$ be the first stage decision of the simpler problem (4.13). Fixing this solution as the first stage decision of the original problem (4.12) we get*

$$v^*(\mathbb{P}) \le v(\mathbb{P}, \tilde{x}_0^*) \le v^*(\mathbb{P}) + L \cdot d(\mathbb{P}, \tilde{\mathbb{P}}).$$

Proof. Since the first stage decisions are always deterministic, we have that $x_0^+ = \tilde{x}_0^*$, and therefore $v(\mathbb{P}, \tilde{x}_0^*) = v(\mathbb{P}, x_0^+) \le v(\mathbb{P}, x^+) \le v^*(\mathbb{P}) + L \cdot d(\mathbb{P}, \tilde{\mathbb{P}})$. □

5. Case study: a multistage inventory problem. This section presents a simple multistage inventory problem adopted to test the bounds introduced before. The problem can be summarized as follows: random demands have to be satisfied from an inventory. If the demand exceeds the stock, it will be satisfied by rapid orders from a different source, which come at a higher price. At each time step (stage), orders can be placed, and they will be delivered one period later. The objective is to minimize the expected disutility of the total costs where profits are considered as negative costs. Demands are the only random quantities in the model and all financial quantities are assumed to be already discounted to the present.

Let assume the following notation.

Deterministic parameters:

b_t	the buying price at time $t = 0, \dots, T-1$
s_t	the selling price at time $t = 1, \dots, T$
h_t	the inventory holding costs from time t to time $t+1$
c_t	the procurement costs for extra stock from another retailer at $t = 1, \dots, T$
d	the final value of the inventory
ζ_0	the initial stock value.

Stochastic scenario process:

ξ_t	the demand at time $t = 1, \dots, T$ (the random scenario process)

Stochastic decision variables:

$x_t \geq 0$ the order sizes at time $t = 0, \ldots, T-1$

Derived variables:

ζ_t the stock after all sales are effectuated at $t = 1, \ldots, T$.

If ζ_t is positive (i.e. $\zeta_t = [\zeta_t]_+$), then $[\zeta_t]_+$ is the amount of stock which will be carried to the next time step. If ζ_t is negative (i.e. $\zeta_t = -[\zeta_t]_-$) then the extra procurement costs of $c_t[\zeta_t]_-$ occur at time t and no stock has to be stored until the next time step. The initial stock ζ_0 is fixed at a given value. The final stock is valuated with the value $d[\zeta_T]_+$.

The cost function is

$$\sum_{t=0}^{T-1} b_t \cdot x_t + \sum_{t=1}^{T} c_t \cdot [\zeta_t]_- + \sum_{t=0}^{T-1} h_t \cdot [\zeta_t]_+ - \sum_{t=1}^{T} s_t \cdot \xi_t - d \cdot [\zeta_T]_+, \tag{5.1}$$

and the constraints are

$$x_{t-1} + [\zeta_{t-1}]_+ - \xi_t = \zeta_t = [\zeta_t]_+ - [\zeta_t]_- \qquad t = 1, \ldots, T \tag{5.2}$$

$$[\zeta_t]_+ \geq 0 \qquad t = 1, \ldots, T \tag{5.3}$$

$$[\zeta_t]_- \geq 0 \qquad t = 1, \ldots, T. \tag{5.4}$$

The first sum in the cost function (5.1) denotes the expected buying cost, while the second sum represents the procurement costs for extra stock from another retailer, the third term is the inventory holding cost between two consecutive periods and the last two terms the profits respectively from selling and for the final value of the inventory. Finally the constraints (5.2), (5.3) and (5.4) define the decision variables of the problem and the relation among them.

The objective function is

$$\min \left\{ \mathbb{E}\left[V\left(\sum_{t=0}^{T-1} b_t \cdot x_t + \sum_{t=1}^{T} c_t \cdot [\zeta_t]_- + \sum_{t=0}^{T-1} h_t \cdot [\zeta_t]_+ - \sum_{t=1}^{T} s_t \cdot \xi_t - d \cdot [\zeta_T]_+ \right) \right] \right\},$$

where V is a convex disutility function, which we choose as

$$V(y) = \begin{cases} y^{1+\delta} & \text{if } y \geq 0 \\ y & \text{if } y < 0 \end{cases}$$

for a $\delta > 0$.

5.1. Computation of bounds for a multistage inventory problem. This section presents some computational tests on the multistage inventory problem.

For this purpose, we consider the scenario tree structure presented in Figure 5.1, with demand process shown in Figure 5.2: this is a six-stage tree with 5 branches from the root, 4 from each of the second-stage nodes, 3 from each of the third-stage nodes, 3 from each of the fourth-stage nodes and 3 from each of the fifth-stage nodes, resulting in $k = 5 \times 4 \times 3 \times 3 \times 3 = 540$ scenarios and 806 nodes. We adopt it as a benchmark to evaluate the cost of optimal solutions obtained using the other reduced scenario trees.

The scenario tree has been generated under *MATLAB* environment on the basis of a time-inhomogeneous exponential auto-regressive $AR(1)$ model for the demand process ξ_t, $t = 1, \ldots, 6$:

$$\xi_t = e^{c + \varphi \xi_{t-1} + \varepsilon_t}, \tag{5.5}$$
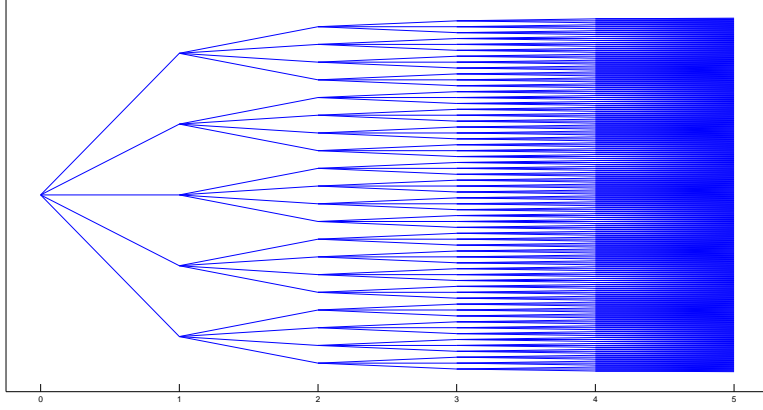
FIG. 5.1. *Scenario tree structure adopted for the full multistage stochastic inventory problem with $k = 540$ scenarios.*
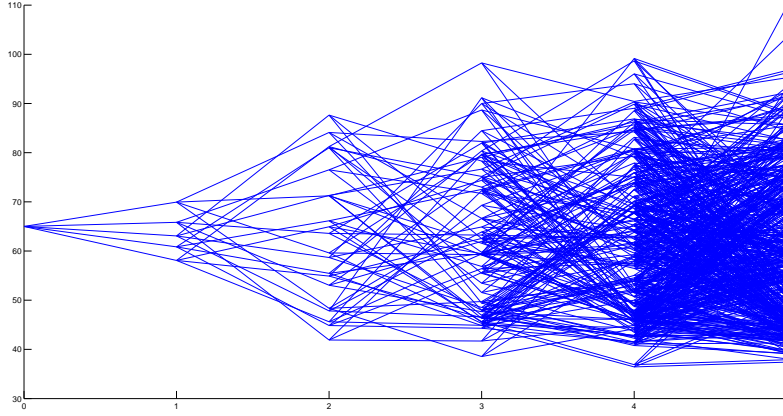

FIG. 5.2. *Scenario tree structure with $k = 540$ scenarios, generated on the basis of an time-inhomogeneous exponential auto-regressive $AR(1)$ model, with demand values ξ_t, $t = 0, \ldots, 5$ represented as y-values.*

with $c = 0$, $\varphi = \frac{0.8}{\sqrt{t}+1}$, $\varepsilon_t = \frac{1-0.8}{\sqrt{t}+1} \mathcal{N}(4, 0.3^2)$ and $\mathcal{N}(4, 0.3^2)$ represents a normally distributed random variable with mean parameter $\mu = 4$ and standard deviation parameter $\sigma = 0.3$.

The value of the process at the root is $\xi_0 = 65$. The final value of the inventory is $d = 2$ per unit, the initial stock value is $\zeta_0 = 2$. $V(y)$ was chosen as y^2 and values of buying price b_t, selling price s_t, inventory holding cost h_t and procurement cost c_t at time period t are presented in Table 5.1.

The non-linear problems derived from our case study have been solved under *AMPL* environment along *MOSEK* solver by interior-point algorithm. All the com-

TABLE 5.1
Buying price b_t, selling price s_t, inventory holding cost h_t from time t to time $t + 1$ and procurement cost c_t for extra stock from another retailer at time period t.

t	b_t	s_t	h_t	c_t
0	3.5	-	2	-
1	3.6	10.7	1.9	8
2	2.3	10.5	2.1	8.1
3	2.8	10.9	2.2	7.9
4	3	10.6	2.1	7
5	-	10	-	7.5

putations have been performed on a 64-bit machine with 12 GB of RAM and a 2.90 GHz processor.

Summary statistics on the size of the optimization problems that *AMPL* generates, are reported in Tables 5.2, 5.3, 5.4 and 5.5 for the six-stage case.

The full stochastic problem is composed by 4304 scalar variables and 2693 constraints.

We construct refinement chains keeping one or several scenarios fixed in all subsets $\Omega_i^{(j)}$ or choosing them disjoint as described in Section 2.

We arbitrarily choose the first f scenarios in the set of available scenarios as fixed for all instances. For any given group size j, choosing alternative fixed scenarios can potentially change the sequence of refinements. Knowledge of special structure in a particular instance may lead to better choices of the f fixed scenarios and it is desirable.

TABLE 5.2
Refinement chain of lower bounds made with disjoint subsets for the "multistage inventory problem" with $k = 540$ scenarios.

j	# subproblems s	Objective v.			CPU s. per subproblem
1	540	v_1^*	=	-2198.81	0.066
5	108	v_5^*	=	-2193.31	0.0705
20	27	v_{20}^*	=	-2162.65	0.0825
60	9	v_{60}^*	=	-2138.17	0.108
180	3	v_{180}^*	=	-2114.67	0.176
540	1	v^*	=	-2096.66	0.31

Tables 5.3-5.4 provide detailed results obtained respectively keeping fixed the first and the first eight scenarios in all subsets $\Omega_i^{(j)}$, $i = 1, \ldots, s$.

Figure 5.3 refers to results of Table 5.2 for increasing values of complexity of the calculation measured in CPU seconds. The refinement chain has been constructed choosing disjoint subsets $\Omega_i^{(j)}$, $i = 1, \ldots, s$ and following the structure of the scenario tree: 5 branches from the root, 4 from each of the second-stage nodes, 3 from each of the third-stage nodes, 3 from each of the fourth-stage nodes and 3 from each of the fifth-stage nodes, resulting in $k = 5 \times 4 \times 3 \times 3 \times 3 = 540$ scenarios.

The first bound v_1^* in the chain has been obtained by solving all 540 scenario subproblems independently and taking their mean, v_5^* by taking 180 subproblems made by 5 equidistant scenarios from the five branches in the root, v_{20}^* by taking 27

TABLE 5.3

Refinement chain of lower bounds by keeping the first scenario fixed in all subsets $\Omega_i^{(j)}$ for the "multistage inventory problem" with $k = 540$ scenarios.

j	f	# subproblems s		Objective v.	CPU s. per subproblem
2	1	539	$v_2^* =$	-2198.77	0.068
8	1	77	$v_8^* =$	-2165.61	0.072
12	1	49	$v_{12}^* =$	-2157.53	0.077
50	1	11	$v_{50}^* =$	-2125.75	0.091
78	1	7	$v_{78}^* =$	-2120.92	0.135
540	1	1	$v^* =$	-2096.66	0.31

TABLE 5.4

Refinement chain of lower bounds by keeping the first eight scenarios fixed in all subsets $\Omega_i^{(j)}$ for the "multistage inventory problem" with $k = 540$ scenarios.

j	f	# subproblems s		Objective v.	CPU s. per subproblem
9	1-8	532	$v_9^* =$	-2198.37	0.065
10	1-8	266	$v_{10}^* =$	-2189.3	0.071
12	1-8	133	$v_{12}^* =$	-2176.85	0.073
15	1-8	76	$v_{15}^* =$	-2165.47	0.079
22	1-8	38	$v_{22}^* =$	-2152.49	0.075
27	1-8	28	$v_{27}^* =$	-2145.58	0.080
36	1-8	19	$v_{36}^* =$	-2139.86	0.09
46	1-8	14	$v_{46}^* =$	-2132.18	0.104
84	1-8	7	$v_{84}^* =$	-2119.82	0.135
141	1-8	4	$v_{141}^* =$	-2114.75	0.195
274	1-8	2	$v_{274}^* =$	-2104.86	0.202
540	1-8	1	$v^* =$	-2096.66	0.31

subproblems made by 20 equidistant scenarios from the branches at the second-stage, v_{60}^* by taking 9 subproblems made by 60 equidistant scenarios from the branches at the third-stage, v_{180}^* by taking 3 subproblems made by 180 equidistant scenarios from the branches at the fourth-stage, and finally the 540 scenarios all together. In this case the refinement chain corresponds to a filtration.

Figures 5.4-5.5 show the percentage deviation of refinement chains reported in Tables 5.3-5.4 for increasing values of complexity of calculation measured in CPU seconds. Figure 5.4 refers to the case of one scenario ($f = 1$) fixed in all subsets $\Omega_i^{(j)}$, and Figure 5.5 to the case of eight scenarios ($f = 8$) fixed in all subsets $\Omega_i^{(j)}$. The cardinality j of each subset in the chain, has been chosen to have the ratio $\frac{540-f}{j-f} = s$ an integer. For instance with $f = 1$ the cardinality of each subproblem in the chain can take only the values $j = 2, 8, 12, 50, 78, 540$ or with $f = 8$, $j = 9, 10, 12, 15, 22, 27, 36, 46, 84, 141, 274, 540$.

From the results we observe that v_j^* improves monotonically with the number of scenarios j of each subproblem (for a fix number f of fixed scenarios), as proved in (2.7). As expected the worst lower bound in the refinement chains is given by v_1 which underestimates the optimal value by 4.87% but requires the lowest CPU time per subproblem (0.066 CPU seconds over 30 runs).
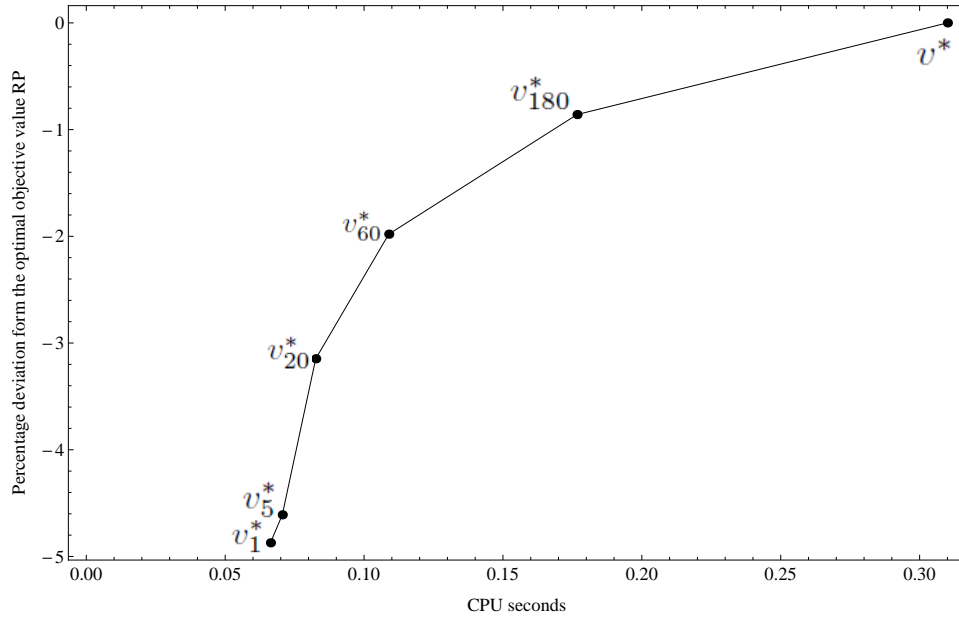
FIG. 5.3. *Percentage deviation from the optimal objective value* v^* *of refinement chain reported in Table 5.2 for increasing values of complexity of calculation measured in CPU seconds.*
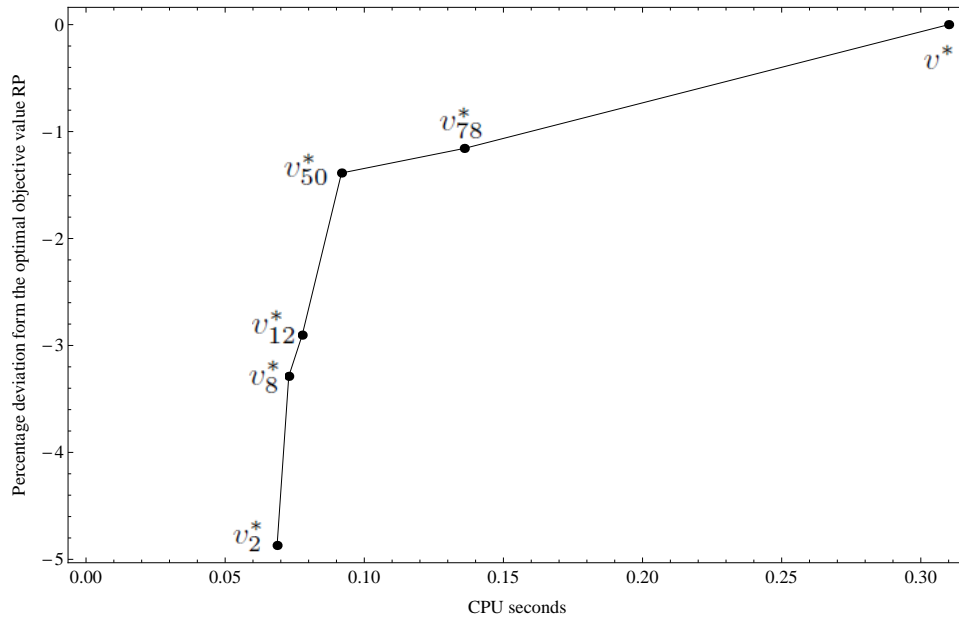

FIG. 5.4. *Percentage deviation from the optimal objective value* v^* *of refinement chain reported in Table 5.3 for increasing values of complexity of calculation measured in CPU seconds.*

Increasing the group size j in the refinement chain and keeping f fixed, signifi-
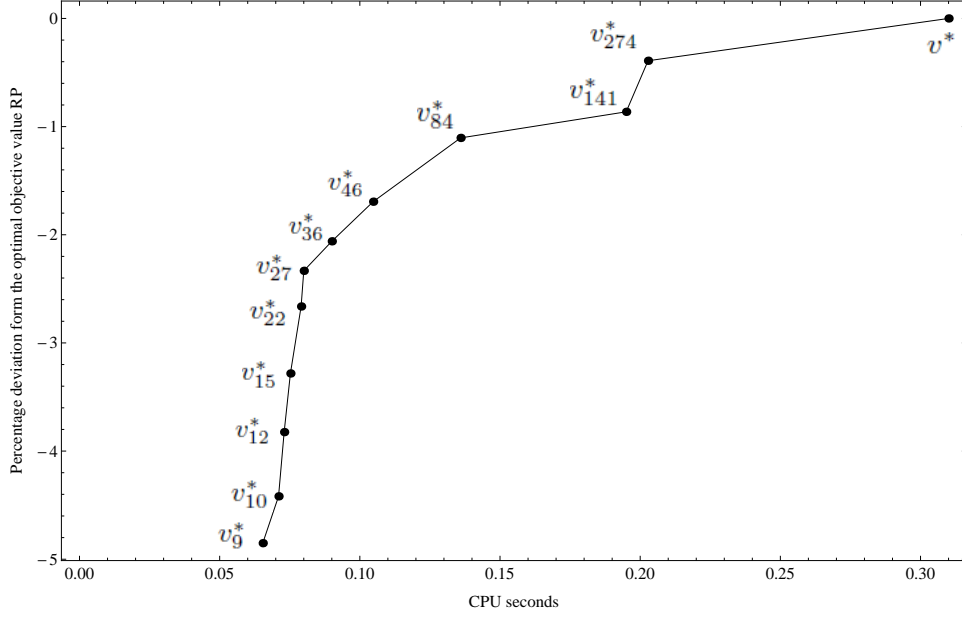
FIG. 5.5. *Percentage deviation from the optimal objective value v^* of refinement chain reported in Table 5.4 for increasing values of complexity of calculation measured in CPU seconds.*

cantly improves the bounds in both construction types, monotonically reaching lower values of percentage deviation. The time required to solve the subproblems in the chain monotonically increases with the dimension of each subproblem reaching the highest value for the biggest scenario tree considered ($j = 540$).

Notice that the same lower bound can be obtained in several ways, keeping fixed a different number of scenarios. This is the case, for example, of lower bound v_{12}^* which can be obtained with one fixed scenario and eleven free (see Table 5.3) or with eight fixed scenarios and four free (see Table 5.4). In the case of one fixed scenario, v_{12}^* is obtained by solving 49 subproblems with a negative percentage deviation of only 2.9033. In the case of eight fixed scenarios, the negative percentage deviation increases to 3.8246 and is obtained by solving a much larger number of subproblems, i.e. 133. Thus we can deduce that better lower bounds are obtained by keeping low the number f of fixed scenarios in all subsets.

Table 5.5 shows the comparison between the stochastic, the expected value solution and upper bounds by inserting (sub)solutions for the multistage inventory problem, in terms of first stage decision variables, total cost and CPU seconds. Inequality (3.2) is verified as well as (4.4) where EEV^t, $t = 1, \ldots, 5$ is obtained by fixing the stochastic variables until stage t to be equal to the expected value solution (see Figure 5.6). Results shows that the best upper bound in terms of percentage deviation versus CPU time is given by EEV^1. The worst upper bound is given by EEV^5 overestimating the optimal value by 1.69% but with the lowest CPU time since the dimension of the problem is smaller (0.19 CPU seconds over 30 runs). All EEV^t computations take less time compared to the full stochastic problem.

In this example upper bounds defined by formulas (4.8) and (4.10) are sharp, since they correspond exactly to v^*, but very expensive in terms of running time

(56.59 CPU seconds for $MEPEV$).

$MEPEV = MESEV_j = v^*$ means that among all the subproblems considered for the computation of v_j^*, there exists at least one with optimal first-stage solution equal to the optimal first-stage stochastic solution. The high computational cost of such approaches is due to the comparison of the objective function of the full stochastic problem with first-stage solution fixed from each of the subproblems considered. $MEPEV$ should then be considered as an alternative to EEV^t upper bounds only in case their distance in terms of total cost from v^* remain high, that is not the case of the example here analyzed.

TABLE 5.5

Comparison between the stochastic, the expected value solution and upper bounds by inserting (sub)solutions for the multistage inventory problem ($k = 540$ scenarios).

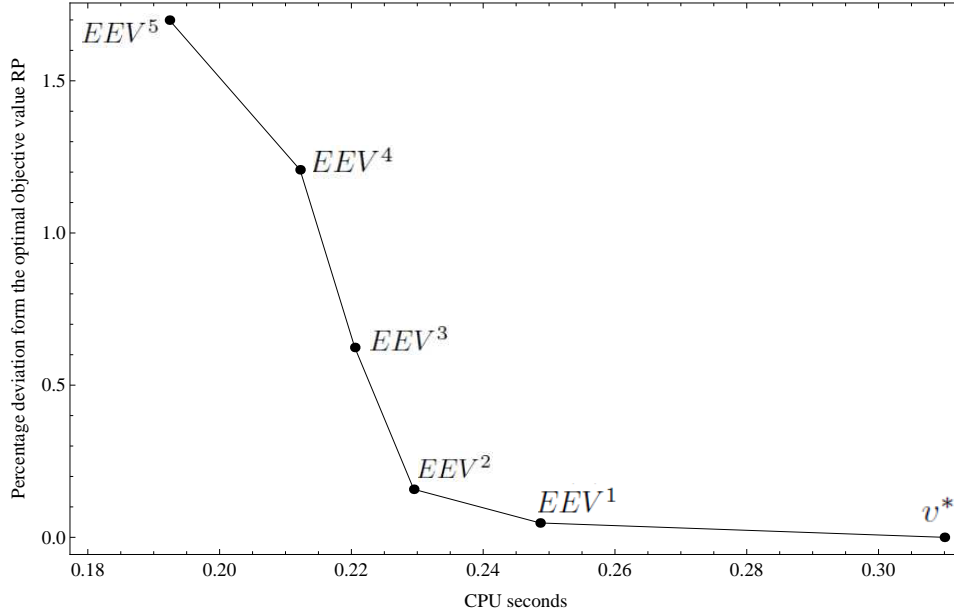| problem type	x_0	$[\zeta_0]_+$	$[\zeta_0]_-$	ζ_0	Objective v.	CPU s.
$v^*(\mathbb{P}_{\mathbb{E}(\xi)})$	61.4188	2	0	2	-2199.70	0.02
$v^*(\mathbb{P})$	63.8405	2	0	2	-2096.66	0.31
$EEV^1 = v^*(\mathbb{P}, \bar{x}_{0:0})$	61.4188	2	0	2	-2095.67	0.2486
$EEV^2 = v^*(\mathbb{P}, \bar{x}_{0:1})$	61.4188	2	0	2	-2093.35	0.2294
$EEV^3 = v^*(\mathbb{P}, \bar{x}_{0:2})$	61.4188	2	0	2	-2083.58	0.2205
$EEV^4 = v^*(\mathbb{P}, \bar{x}_{0:3})$	61.4188	2	0	2	-2071.34	0.2122
$EEV^5 = v^*(\mathbb{P}, \bar{x}_{0:4})$	61.4188	2	0	2	-2061.03	0.1923
$MEPEV$	63.8405	2	0	2	-2096.66	56.69
$MESEV_j$	63.8405	2	0	2	-2096.66	

FIG. 5.6. *Percentage deviation from the optimal objective value v^* of upper bounds EEV^t, $t = 1, \ldots, 5$ reported in Table 5.5 for increasing values of complexity of calculation measured in CPU seconds in the case of a six-stage scenario tree with $k = 540$ scenarios.*

Now we consider the case of a problem that is too large to be solved exactly. This is the typical situation in which most modelers would actually resort to the bounding methods described in this paper.

For this purpose, we consider a larger scenario tree structure given by a ten-stages tree with 8 branches from the root, 7 from each of the second-stage nodes, 6 from each of the third-stage nodes, 6 from each of the fourth-stage nodes, 5 from each of the fifth-stage nodes, 4 from each of the sixth-stage nodes, 3 from each of the seventh-stage nodes, 3 from each of the eight-stage nodes and 2 from each of the ninth-stage nodes resulting in $k = 8 \times 7 \times 6 \times 6 \times 5 \times 4 \times 3 \times 3 \times 2 = 725760$ scenarios and 1262417 nodes. We use it as a benchmark to evaluate the cost of optimal solutions obtained using the other reduced scenario trees.

The scenario tree has been generated as in the previous case with exception of the following values: the value the process at the root was given by $\xi_0 = 60$, and δ was chosen as 3. The values of buying price b_t, selling price s_t, inventory holding cost h_t and procurement cost c_t at time period t are presented in Table 5.6.

TABLE 5.6
Buying price b_t, selling price s_t, inventory holding cost h_t from time t to time $t + 1$ and procurement cost c_t for extra stock from another retailer at time period t in case of a scenario tree with 10 stages, 725760 scenarios and 1262417 nodes.

t	b_t	s_t	h_t	c_t
0	3.5	-	2	-
1	3.6	10.7	1.9	8
2	2.3	10.5	2.1	8.1
3	2.8	10.9	2.2	7.9
4	3	10.6	2.1	7
5	3.1	10	2.3	7.5
6	3.2	10.4	1.8	7.3
7	3.7	10.3	1.9	7.7
8	4	10.8	2.1	8.1
9	-	10.9	-	7.5

We again use *AMPL* environment along *MOSEK* solver on a 64-bit machine with 12 GB of RAM and a 2.90 GHz processor.

Summary statistics on the size of the optimization problems that *AMPL* generates, are reported in Tables 5.7 and Figure 5.7 for the ten-stage case.

The full stochastic problem is composed by 6501188 scalar variables and 3976355 constraints.

We construct refinement chains choosing them disjoint as described in Section 2.

Figure 5.7 refers to results of Table 5.7 for increasing values of complexity of calculation measured in CPU seconds. The refinement chain have been constructed choosing disjoint subsets $\Omega_i^{(j)}$, $i = 1, \ldots, s$ and following the structure of the scenario tree: 8 branches from the root, 7 from each of the second-stage nodes, 6 from each of the third-stage nodes, 6 from each of the fourth-stage nodes, 5 from each of the fifth-stage nodes, 4 from each of the sixth-stage nodes, 3 from each of the seventh-stage nodes, 3 from each of the eight-stage nodes and 2 from each of the ninth-stage nodes resulting in $k = 8 \times 7 \times 6 \times 6 \times 5 \times 4 \times 3 \times 3 \times 2 = 725760$ scenarios.

The first measure v_8^* in the chain has been obtained by solving 90720 scenario subproblems independently and taking the expectation, v_{56}^* by taking 12960 subproblems made by 56 equidistant scenarios from the eight branches in the root, v_{336}^* by taking

TABLE 5.7
Refinement chain of lower bounds made with disjoint subsets for the "multistage inventory problem" with k = 725760 scenarios.

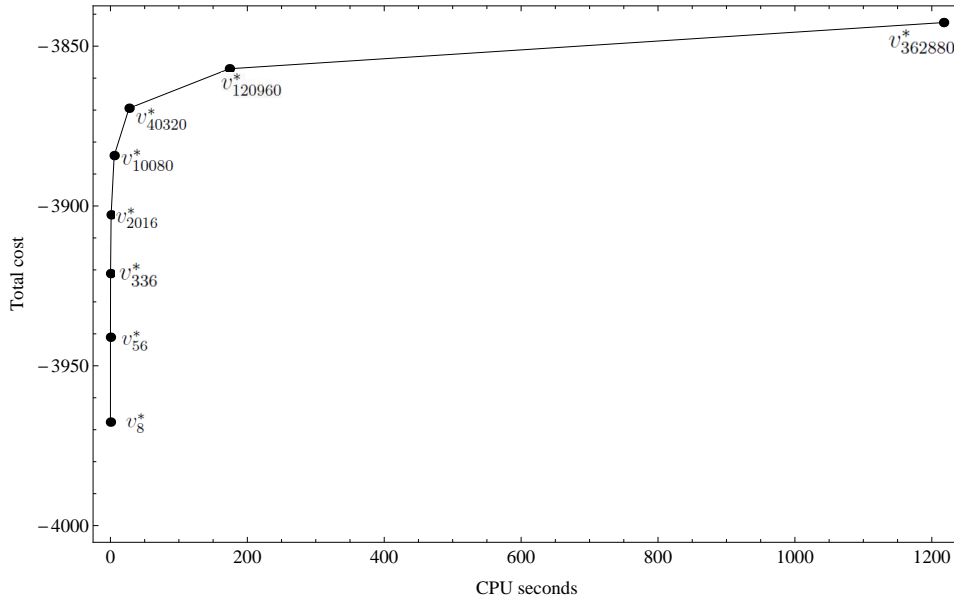| j	# subprob. s		Objective v.	CPU s. subprob.	CPU s. procedure
8	90720	v_8^*	$= -3967.61$	0.0287	2612.29
56	12960	v_{56}^*	$= -3941.06$	0.0539	698.71
336	2160	v_{336}^*	$= -3921.15$	0.1990	430.011
2016	360	v_{2016}^*	$= -3902.77$	0.9385	337.89
10080	72	v_{10080}^*	$= -3884.23$	5.4852	394.94
40320	18	v_{40320}^*	$= -3869.42$	27.4512	494.123
120960	6	v_{120960}^*	$= -3857.06$	166.5516	999.31
362880	2	v_{362880}^*	$= -3842.61$	1274.424	2548.848

FIG. 5.7. *Total costs of refinement chain of lower bounds reported in Table 5.7 for increasing values of complexity of calculation measured in CPU seconds in the case of the 10-stage scenario tree with k = 725760 scenarios.*

2160 subproblems made by 336 equidistant scenarios from the branches at the second-stage, v_{2016}^* by taking 360 subproblems made by 2016 equidistant scenarios from the branches at the third-stage and so on until v_{362880}^*, obtained by taking 2 subproblems made by 362880 equidistant scenarios from the branches at the ninth-stage. The full problem was unsolvable with the above specified computing environment.

From the results we observe that v_j^* improves monotonically with the number of scenarios j of each subproblem, as proved in (2.7). As expected the worst lower bound in the refinement chains is given by v_8 but requires the lowest CPU time per subproblem (0.0287 CPU seconds).

Increasing the group size j in the refinement chain, as expected the bounds improve, with a difference between the first and last bounds in the chain $v_{362880}^* - v_8^* =$

125 ($= 3.25\%$ v_{362880}^*). Since the exact value of the full stochastic problem is not available, values of percentage deviations with respect to it cannot be explicitly computed. However the best lower bound obtained is given by v_{362880}^* but with the largest CPU time per subproblem.

Last column of Table 5.7 reports the CPU seconds for the computation of the bounds sequentially. From the numerical results of the case study considered is more convenient to solve $s = 360$ subproblems with $j = 2016$ and a time of 337.89 CPU seconds than the bounds obtained with a lower subproblems cardinality ($j < 2016$) which are also further away from the optimum. Increasing the subproblem cardinality j, the bounds are tighter and tighter but require higher CPU times.

Another lower bound can be obtained from the expected value problem $v^*(\mathbb{P}_{\mathbb{E}(\xi)}) = -3986.71$, (see Table 5.8) still worse than v_8^* but solved with a computational time of only 0.09 CPU seconds.

Table 5.8 shows also upper bounds for the multistage inventory problem with $k = 725760$ scenarios, in terms of total cost and CPU seconds. The only upper bounds that we were able to solve were EEV^t, $t = 8, 9$ obtained by fixing all the variables until stage t at the expected value solution, reducing the number of free nodes to 1088640 for $t = 8$ and to the number of scenarios for $t = 9$, i.e. 725760. All the other proposed upper bounds EEV^t, $t = 1, \ldots, 7$, $MEPEV$ and $MESEV^j$ go out of memory because the number of fixed variables from the expected value solution is not enough to reduce the dimension of the scenario tree. The difference between the best upper and lower bounds obtained, EEV^8 and v_{362880}^* respectively, is 299.45 (7.79% v_{362880}^*) and gives information about the range where we should expect to find the total cost of full stochastic problem that we are not able to compute.

TABLE 5.8
Expected value solution and upper bounds EEV^t, $t = 8, 9$ by inserting (sub)solutions for the multistage inventory problem until stage t, in the case of the 10-stage scenario tree with $k = 725760$ scenarios.

problem type	Objective v.	CPU s.
$v^*(\mathbb{P}_{\mathbb{E}(\xi)})$	-3986.71	0.09
$EEV^8 = v^*(\mathbb{P}, \bar{x}_{0:8})$	-3543.16	2412.51
$EEV^9 = v^*(\mathbb{P}, \bar{x}_{0:8})$	-3483.56	4254.97

6. Complexity considerations. In the previous section it has been demonstrated that the calculation of lower bounds requires to solve independently many problems of smaller size. Since the complexity of a stochastic program grows superlinearly with the number of nodes of the scenario tree, this may result in a considerable reduction of complexity compared to the solution of the full problem, even if the subproblems are all executed sequentially.

To illustrate this, assume that $c(n)$ denotes the worst case execution complexity of a tree structured stochastic program with n nodes, denoted from now on with $\mathcal{P}$. Suppose that the tree has branching factor b, i.e. $b = 2$ for binary trees, $b = 3$ for ternary trees and so on. For a tree with height T, the number of scenarios is $|\Omega| = k = b^T$ and the number of nodes is $|\mathfrak{N}| = (b^{T+1} - 1)/(b-1)$. If the subproblems

$$\Omega_i^{(j)} = \{\omega_1, \ldots, \omega_f, \omega_{f+(i-1)\cdot x+1}, \ldots, \omega_{f+i\cdot x}\}, \quad i = 1, \ldots, s,$$

are based on j leaf nodes, out of which f are kept fixed, with $x = j - f > 0$, one has to solve $s = (b^T - f)/(j - f)$ of them, where each subproblem $\Omega_i^{(j)}$ has at most $|\mathfrak{N}_T^{(j)}| = jT + 1$ nodes. Let $\mathcal{P}^{(j)}$ the problem obtained by solving $\Omega^{(j)} = \left\{\Omega_i^{(j)} | i = 1, \ldots, s\right\}$, with associated lower bound v_j^*. Its complexity is then given by

$$c(\mathcal{P}^{(j)}) = c(jT + 1) \cdot \frac{b^T - f}{j - f}$$

and the ratio between the worst case complexities of the lower bound calculation $\mathcal{P}^{(j)}$ and the full stochastic problem $\mathcal{P}$ is

$$\frac{c(\mathcal{P}^{(j)})}{c(\mathcal{P})} = \frac{c(jT + 1)}{c((b^{T+1} - 1)/(b-1))} \cdot \frac{b^T - f}{j - f}. \tag{6.1}$$

To insert a concrete function for $c(\cdot)$ assume for simplicity that the stochastic program is linear with just one decision variable and one linking constraint per node. If this program is solved by interior point method, then one may use the complexity function

$$c(n) = O(L \cdot n^3/log(n)) \tag{6.2}$$

given by Anstreicher [1], with L being the data bit size. Inserting (6.2) into (6.1) (with $L = 1$), one gets the ratio behaviours shown in Figures 6.1, 6.2 and 6.3, which go fast to zero for large T, showing the advantage of the proposed procedure. Figures 6.2 and 6.3 exhibit the fact that for small b and T, this ratio may even be larger than one.

7. Conclusions. The paper develops lower and upper bounds for multistage stochastic programs both in the linear and non-linear case.

Lower bounds to the optimal solution have been found by relaxation of some constraints and upper bound by inserting feasible solutions derived by simpler problems.

We introduced lower bounds to the optimal solution by changing the probability measure P (see Section 2) or by replacing the scenario process by its expectation (see Section 3). In the first case each lower bound has been built solving group subproblems choosing them disjoint or keeping one or several scenarios fixed in all subsets. The monotonicity of the chain of lower bounds in terms of the cardinality of subproblems has been proved as a consequence of the concavity of the probability mapping $P \mapsto v^*(P)$ (see Lemma 2.1).
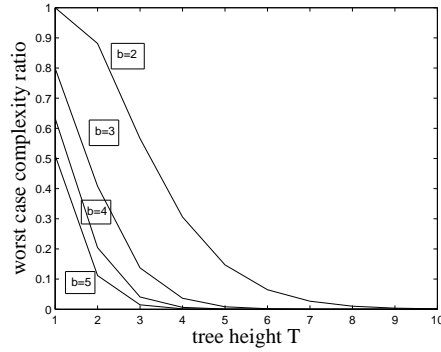
FIG. 6.1. *The complexity ratio given by (6.1) using (6.2) for various values of the branching factor b and the tree height T. Here the number of leaf nodes for each subproblem is $j = 2$ and there are no fixed scenarios among the subproblems, $f = 0$.*
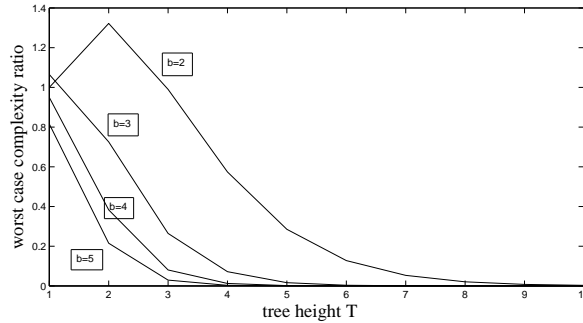

FIG. 6.2. *A similar picture as Figure 6.1, but with the choice of ten leaf nodes for each subproblem, $j = 10$, and one fixed scenario $f = 1$.*

Different upper bounds by inserting (sub)solutions are also analyzed and compared in Section 4.

The proposed approach has the important advantage to divide a given problem into independent subproblems which may take advantage of parallel computations, if parallel hardware is available. Notice that lower bounds typically do not lead to feasible solutions of the basic problem, but once only a small gap is detected between an upper bound with feasible solution and a lower bound, one may stop and state that a nearly optimal solution is found in an efficient way.

For illustration, numerical results on a large scale non linear multistage inventory problem have been presented both in the case of solvability and not of the full stochastic problem. Complexity considerations for the computation of the proposed lower bounds as function of tree depth and branching factor are also discussed.

REFERENCES

[1] K. M. ANSTREICHER, *Linear programming in $O(n^3/ln\,n \cdot L)$ operations*, SIAM J. Optimization 9 (4) (1999) pp. 803–812.
[2] J. R. BIRGE, *The value of the stochastic solution in stochastic linear programs with fixed recourse*, Math. Prog., 24, (1982) pp. 314–325.
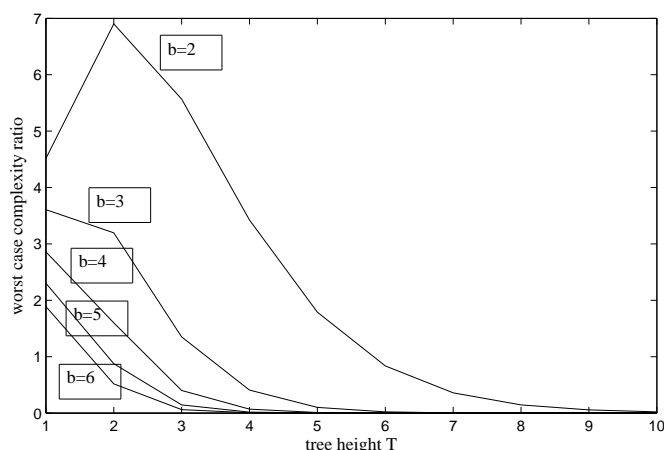
FIG. 6.3. *The complexity ratio given by (6.1) using (6.2) for various values of the branching factor b and the tree height T. Here the number of leaf nodes for each subproblem is $j = 10$ and there are no fixed scenarios among the subproblems, $f = 0$.*

[3] J. R. BIRGE, AND F. LOUVEAUX, *Introduction to stochastic programming*, Springer-Verlag, New York, 1997.

[4] D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Optimization with Stochastic Dominance Constraints*, SIAM Journal on Optimization, 14, (2003), pp. 548–566.

[5] L. F. ESCUDERO, A. GARÍN, M. MERINO, AND G. PÉREZ, *The value of the stochastic solution in multistage problems*, Top, 15, (2007) pp. 48–64.

[6] K. FRAUENDORFER, AND M. SCHÜRLE, *Multistage stochastic programming: Barycentric approximation*. P. Pardalos, and C. A. Floudas eds., Encyclopedia of Optimization, Vol. 3 Dordrecht, NL: Kluwer Academic Publishers, ISBN 0-7923-6932-7, 2001, pp. 576–580.

[7] D. KUHN , *Generalized Bounds for Convex Multistage Stochastic Programs*. Lecture Notes in Economics and Mathematical Systems 548, Spinger Verlag, 2005.

[8] D. KUHN , *Aggregation and discretization in multistage stochastic programming* Math. Programming, Ser. A, 113 (2008) pp. 61–94.

[9] D. KUHN , *An Information-Based Approximation Scheme for Stochastic Optimization Problems in Continuous Time*. Mathematics of Operations Research, 34 (2), (2009) pp. 428–444.

[10] F. MAGGIONI, AND W. S. WALLACE, *Analyzing the quality of the expected value solution in stochastic programming*, Annals of Operations Research, 200, (2012) pp. 37–54.

[11] F. MAGGIONI, E. ALLEVI, AND M. BERTOCCHI, *Bounds in Multistage Linear Stochastic Programming*, Journal of Optimization Theory and Applications, 163 (1) (2014) pp. 200–229.

[12] W.K. MAK, D.P. MORTON, AND R.K. WOOD, *Monte Carlo bounding techniques for determining solution quality in stochastic programs*, Operations Research Letters 24, (1999) pp. 47-56.

[13] D.P. MORTON, AND R.K. WOOD, *Restricted-recourse bounds for stochastic linear programming*, Operations Research 47, (1999) pp. 943-956.

[14] G. CH. PFLUG, A. RUSZCZYNSKI, AND R. SCHULTZ, *On the Glivenko Cantelli Problem in Stochastic Programming: Linear Recourse*. Mathematics of OR 23 (1) (1998) pp. 204–220.

[15] G. CH. PFLUG, AND W. RÖMISCH *Modeling, measuring and managing risk*, World Scientific, 2007.

[16] G. CH. PFLUG , *Version-independence and nested distributions in multistage stochastic optimization*. SIAM J. on Optimization 20(3), (2010) pp. 1406-1420.

[17] G. CH. PFLUG, A. PICHLER. *A distance for multi-stage stochastic optimization models*. SIAM J. Optimization 22 (1), (2012) pp. 1-23.

[18] B. SANDIKÇI, N. KONG, A.J. SCHAEFER, *A hierarchy of bounds for stochastic mixed-integer programs*, Math. Program., Ser. A, 138 (1-2), (2012) pp. 253–272.