Working Paper

Scenario Based Stochastic Programs: Strategies for Deleting Scenarios

Jitka Dupačová

WP-95-014
February 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Scenario Based Stochastic Programs: Strategies for Deleting Scenarios

Jitka Dupačová

WP-95-014
February 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Abstract

The proposed strategies for deleting scenarios are based on postoptimality analysis of the optimal value function with respect to probabilities of the included scenarios. These strategies can be used to reduce the size of the large scenario based problems or of the problems constructed in the course of specific numerical procedures, such as stochastic decomposition or scenario aggregation.

A convex nonsmooth optimization problem is replaced by a sequence of line search problems along recursively updated rays. Convergence of the method is proved and applications indicated.

Key words: Two-stage stochastic programs with random recourse, postoptimality, sensitivity, deleting scenarios.

1. THE MAIN IDEAS

In this paper, we shall deal with *stochastic linear programs with relatively complete random recourse* in their generic form

$$\text{(1)} \qquad \text{minimize} \quad E_P\left\{\mathbf{c}(\omega)^\top \mathbf{x} + Q(\mathbf{x},\omega)\right\}$$

$$\text{on the set} \quad \mathcal{X} = \left\{\mathbf{x} \in R_+^{n_1} | \mathbf{A}\mathbf{x} = \mathbf{b}\right\}$$

with the recourse costs $Q(\mathbf{x},\omega)$ defined for a given $\mathbf{x}$ and ω as the optimal value of the auxiliary second stage program

$$\text{(2)} \qquad \text{minimize} \quad \mathbf{q}(\omega)^\top \mathbf{y}$$

$$\text{subject to} \quad \mathbf{y} \in R_+^{n_2} \quad \text{that satisfy} \quad \mathbf{W}(\omega)\mathbf{y} + \mathbf{T}(\omega)\mathbf{x} = \mathbf{h}(\omega)$$

We assume that the set $\mathcal{X}$ is nonempty, the probability distribution P of random coefficients in (1), (2) is carried by a known nonempty closed set Ω, that the matrices $\mathbf{W}(\omega)$ are almost surely of a fixed row rank and that the expectation of the recourse function $Q(\mathbf{x},\omega)$ is finite for all $\mathbf{x} \in \mathcal{X}$. Without any loss of generality, we can use nonrandom coefficients $\mathbf{c}$ in (1).

The numerical techniques designed for solving (1), (2) (see e. g. [7] and [11]) are mostly based on a discrete approximation of the distribution P carried by a finite number of *scenarios*. Such distribution can be obtained as an approximation of the true probability distribution, can be generated in the course of the numerical procedure or by a limited sample information, can be based on a preliminary analysis of the problem or may reflect an ad hoc belief or a subjective opinion of an expert. In this context, two types of procedures can be distinguished: deterministic algorithms that are based on a large and in principle *fixed* set of scenarios, such as large scale linear programming techniques (cf. [15] and references ibid) or the progressive hedging algorithm [13], [16] and stochastic algorithms where a proper sampling procedure and generation of new scenarios becomes a part of the algorithm, e.g., stochastic quasigradient methods [6] or the stochastic decomposition algorithm [11].

Consider now the scenario based form of (1), (2) that coresponds to a given discrete probability distribution P concentrated in a finite number of fixed atoms, called *scenarios* $\omega_1, \ldots, \omega_S$ with prescribed positive probabilities $p_1, \ldots, p_S, \sum_{s=1}^S p_s = 1$. The coefficients of (2) generated by scenario ω_s are denoted $\mathbf{q}_s, \mathbf{W}_s, \mathbf{T}_s, \mathbf{h}_s$. Accordingly, the program (1), (2) takes on the form of the following large linear program

minimize

$$\text{(3)} \qquad \mathbf{c}^\top \mathbf{x} + \sum_{s=1}^{S} p_s \mathbf{q}_s^\top \mathbf{y}_s$$

subject to

$$
(4) \qquad
\begin{array}{llll}
\mathbf{A}\mathbf{x} & & & = \mathbf{b} \\
\mathbf{T}_1\mathbf{x} + \mathbf{W}_1\mathbf{y}_1 & & & = \mathbf{h}_1 \\
\mathbf{T}_2\mathbf{x} + & \mathbf{W}_2\mathbf{y}_2 & & = \mathbf{h}_2 \\
\vdots & & \ddots & \vdots \\
\mathbf{T}_S\mathbf{x} + & \dots & + \; \mathbf{W}_S\mathbf{y}_S & = \mathbf{h}_S \\
\end{array}
$$

$$\mathbf{x} \geq 0, \mathbf{y}_s \geq 0, s = 1, \dots, S$$

where $\omega_s = [\mathbf{q}_s, \mathbf{T}_s, \mathbf{W}_s, \mathbf{h}_s], s = 1, \dots, S$ are scenarios or atoms at which the probability distribution P is concentrated and $p_s \geq 0, s = 1, \dots, S$ are their probabilities, $\sum_s p_s = 1$.

The optimal value φ and the set $\mathcal{X}^*$ of optimal solutions of (3), (4) are optimal *with respect to the choice of scenarios and of their probabilities.* For stability and postoptimality of the optimal value φ and of the set of optimal solutions $\mathcal{X}^*$ of the large scale linear program (3),(4) with respect to changes in probabilities p_i one can rely on the well known results for linear programs with linearly perturbed objective function, see e.g. [14]: inter alia, the optimal value function φ is concave, piecewise linear on its domain with the mapping $\mathcal{X}^*$ upper semicontinuous. Continuity of the optimal solution with respect to probabilities of the considered scenarios was discussed also in [12] as an application of more general stability results, a postoptimality procedure is suggested in [16] in the context of the progressive hedging algorithm, resistance of the output with respect to additional scenarios is treated in [4]. The obtained results suggest that even for deterministic algorithms for (1), (2), specific techniques related to the special structure of stochastic programs with recourse can help to get postoptimality results valid under more general circumstances than those based solely on linear programming techniques.

In this paper we shall concentrate on *designing strategies for deleting scenarios.* The starting point will be the postoptimality analysis with respect to probabilities of the included scenarios. This fact seem to suggest that rules for deleting scenarios ex post, when the problem has been already solved, are not of a great interest. To see a reason for designing these rules consider a large multiperiod two-stage stochastic program for financial planning described in [10], [17] in which, due to its size, just a few short term interest rate scenarios can be used. Additional simulation studies [10] and bounds based on the contamination technique (cf. [3], [4]) can be used to provide information about the behavior of the obtained solution for other out-of-sample scenarios. In case of a bad performance one should include additional scenarios into the model and repeat the computations. Before doing it, one is definitely interested in deleting "noninfluential" scenarios to decrease the computational effort. Another reason appears in connection with algorithms based on solution of *sequences* of growing scenario based optimization problems, e.g., stochastic decomposition [9]: exploitation of properly designed rules for deleting scenarios in individual iterations of the algorithm will save the computing time and may contribute essentially to numerical tractability of the underlying problem; see Section 5 for the first ideas.

In this Section, we shall explain the essence of the suggested scheme for deleting scenarios for the case of a deterministic algorithm based on solution of the large scale linear program

(3), (4) for a priori chosen scenarios. Notice, that in this case, randomness of all coefficients of the second-stage problem (2) is allowed. Assume thus that the program (3), (4) has ben solved for the given set of scenarios $\omega_1, \dots, \omega_S \in \Omega$ with probabilities $p_i^* > 0 \forall i, \sum_i p_i^* = 1$.

Influence of *deleting scenario* ω_s can be evidently formulated as a special problem of stability or postoptimality w.r.t. the probabilities p_i^* under condition that p_s^* is changed to 0. The already mentioned continuity properties [12], [14] suggest to choose the new probabilities $\hat{p}_i, i = 1, \dots, S$ as the projection of $p_i^*, i = 1, \dots, S$ on the facet of the simplex $\mathcal{P}_S = \{\mathbf{p} \in R_+^S | \sum_{i=1}^S p_i = 1\}$ that corresponds to the requirement $p_s = 0$. Accordingly, we get *the first redistribution rule*

$$(5) \qquad \hat{p}_i = p_i^* + \frac{1}{S-1} p_s^* \quad \text{for} \quad i \neq s \quad \text{and} \quad \hat{p}_s = 0$$

and the minimal distance (given the choice of ω_s) equals $p_s^* \sqrt{(\frac{S}{S-1})}$.

Hence, the first heuristic rule for deleting scenarios: *Delete the scenario whose probability is minimal.* However, this simple rule does not help in the case of eqiprobable scenarios and even for unequal probabilities, it should be supported by an additional analysis.

Assume that the set $\mathcal{X}^*(\mathbf{p}^*)$ of optimal solutions of (3), (4) for $\mathbf{p} = \mathbf{p}^*$ is nonempty and bounded; then the directional derivative of the optimal value function φ at $\mathbf{p}^*$ exists in an arbitrary direction [8] and equals

$$(6) \quad \varphi'(\mathbf{p}^*; \hat{\mathbf{p}} - \mathbf{p}^*) = \min_{\mathbf{x} \in \mathcal{X}^*(\mathbf{p}^*)} f(\mathbf{x}, \hat{\mathbf{p}}) - \varphi(\mathbf{p}^*) =$$
$$\min_{\mathbf{x} \in \mathcal{X}^*(\mathbf{p}^*)} \left\{ \sum_{i=1}^S \hat{p}_i [\mathbf{c}^\top \mathbf{x} + Q(\mathbf{x}, \omega_i)] - \sum_{i=1}^S p_i^* [\mathbf{c}^\top \mathbf{x} + Q(\mathbf{x}, \omega_i)] \right\} =$$
$$p_s^* \min_{\mathbf{x} \in \mathcal{X}^*(\mathbf{p}^*)} \left[\frac{1}{S-1} \sum_{i \neq s} Q(\mathbf{x}, \omega_i) - Q(\mathbf{x}, \omega_s) \right]$$

If, in addition, the optimal solution $\mathbf{x}^* = \mathbf{x}(\mathbf{p}^*)$ is unique, we have

$$(7) \qquad \varphi'(\mathbf{p}^*; \hat{\mathbf{p}} - \mathbf{p}^*) = p_s^* \left[\frac{1}{S-1} \sum_{i \neq s} Q(\mathbf{x}^*, \omega_i) - Q(\mathbf{x}^*, \omega_s) \right]$$

Inspection of formula (7) leads to the following straightforward conclusions:

(i) Deleting scenario ω_s can cause both *local* increase and *local* decrease of the optimal value and the criterion based on the sign of the marginal value $\varphi'(\mathbf{p}^*; \hat{\mathbf{p}} - \mathbf{p}^*)$ depends on the initial probabilities $\mathbf{p}^*$ only via the corresponding optimal solution $\mathbf{x}^* = \mathbf{x}(\mathbf{p}^*)$.

(ii) Concavity of φ implies that deleting scenario ω_s for which $\varphi'(\mathbf{p}^*; \hat{\mathbf{p}} - \mathbf{p}^*) \leq 0$ causes *decrease* of the optimal value whereas deleting scenario ω_s for which $\varphi'(\mathbf{p}^*; \hat{\mathbf{p}} - \mathbf{p}^*) > 0$ can lead both to *the increase and to the decrease* of the optimal value; the reason is that deleting scenario ω_s corresponds to the step of the length 1 in the direction of $\hat{\mathbf{p}} - \mathbf{p}^*$ with $\hat{\mathbf{p}}$ given by (5).

(iii) The locally "least influential scenario" ω_s is characterized not only by a small probability p_s but also by the the minimal possible absolute value of the difference between the average recourse costs $Q(\mathbf{x}^*, \omega_i)$ for $i \neq s$ and the recourse costs $Q(\mathbf{x}^*, \omega_s)$.

(iv) To identify the locally least influential scenario means to select such scenario ω_s for which

$$Q(\mathbf{x}^*, \omega_s) \doteq \frac{1}{S} \sum_{i=1}^{S} Q(\mathbf{x}^*, \omega_i) \tag{8}$$

Deleting scenario ω_s according to (8) implies that the value $f(\mathbf{x}(\mathbf{p}^*), \mathbf{p})$ is not very sensitive to small changes in $\mathbf{p}$ described by $\mathbf{p} = \mathbf{p}^* + \lambda(\hat{\mathbf{p}} - \mathbf{p}^*)$ with $\hat{\mathbf{p}}$ defined according to (5), i. e., that $f(\mathbf{x}(\mathbf{p}^*), \mathbf{p}^* + \lambda(\hat{\mathbf{p}} - \mathbf{p}^*)) \doteq \varphi(\mathbf{p}^*)$ for λ small enough. In the terms of the optimal value function, (7) means that the function φ attains its maximum on the straight line $\mathbf{p}^* + \lambda(\hat{\mathbf{p}} - \mathbf{p}^*)$ at the point $\mathbf{p}^*$, so that its function values for $\lambda > 0$ including that for $\hat{\mathbf{p}}$ (i. e., for deleted scenario ω_s according to (7) and its probability mass redistributed according to (5)) are not greater than $\varphi(\mathbf{p}^*)$.

The heuristic procedure for deleting one scenario based on the above observations consists of selection scenario ω_s according to (8) and of redistribution of its probability p_s according to (5) . If there is no scenario for which (8) holds true with a sufficient precision one considers scenarios for which

$$Q(\mathbf{x}^*, \omega_s) \geq \frac{1}{S} \sum_{i=1}^{S} Q(\mathbf{x}^*, \omega_i) \tag{9}$$

and selects the one for which *the product*

$$p_s^*[Q(\mathbf{x}^*, \omega_s) - \frac{1}{S} \sum_{i=1}^{S} Q(\mathbf{x}^*, \omega_i)] \tag{10}$$

is minimal. For this choice of scenario to be deleted and for its probability mass redistributed according to (5), the optimal value function decreases at $\mathbf{p}^*$ in the direction $\hat{\mathbf{p}} - \mathbf{p}^*$ so that deleting this scenario means $\varphi(\hat{\mathbf{p}}) < \varphi(\mathbf{p}^*)$.

The last conclusion can be easily modified for the case of *multiple optimal solutions* of the initial program: Whenewer inequality

$$Q(\mathbf{x}, \omega_s) \geq \frac{1}{S} \sum_{i=1}^{S} Q(\mathbf{x}, \omega_i) \tag{9'}$$

holds true for an optimal solution $\mathbf{x} \in \mathcal{X}^*(\mathbf{p}^*)$ and for a scenario ω_s, then deleting ω_s and redistribution of its probability p_s according to (5) leads to decrease of the optimal value.

The above ideas can be extended to the case of *deleting more than one scenario*: Assume, for instance, that a subset of K scenarios, say, $\omega_{s_k}, k = 1, \ldots, K$ should be kept and the remaining $D = S - K$ scenarios should be deleted. Let the total probability mass of the deleted scenarios be $p_0^* = \sum_{k=K+1}^{S} p_{s_k}$. According to the minimum L_2 distance criterion, we get probabilities

$$\hat{p}_{s_k} = p_{s_k}^* + \frac{1}{K} p_0^*, \quad k = 1, \ldots, K \quad \text{and} \quad \hat{p}_s = 0 \quad \text{otherwise} \tag{5'}$$

Under assumption of unique optimal solution $\mathbf{x}^*$ of the initial program for $\mathbf{p} = \mathbf{p}^*$, the directional derivative

$$(7') \qquad \varphi'(\mathbf{p}^*; \hat{\mathbf{p}} - \mathbf{p}^*) = \frac{p_0^*}{K} \sum_{k=1}^{K} Q(\mathbf{x}^*, \omega_{s_k}) - \sum_{k=K+1}^{S} p_{s_k}^* Q(\mathbf{x}^*, \omega_{s_k})$$

Again, it is possible to write decision rules similar to (9), (10). In general, however, it is not that easy to identify in advance the scenarios whose deleting leads to zero value of (7').

Remarks and generalizations.

(1) First of all notice that exploitation of *direct postoptimality techniques for linear program* (3), (4) with respect to coefficients p_s in the objective function means to check if the optimality conditions hold true for $\hat{\mathbf{p}}$ with the same optimal basis, or equivalently, to check the dual feasibility of the dual variables obtained for the original program under changed $\mathbf{p}$. Robust behavior with optimal value differentiable with respect to $\mathbf{p}$ occurs when there is a unique optimal solution $\mathbf{x}^*, \mathbf{y}_s^*, s = 1, \ldots, S$ for the initial vector of probabilities $\mathbf{p}^*$. In this case, our formula (7) holds true again. Notice, however, that an additional assumption of *unique* optimal second stage decisions for all scenarios is needed to get it and that such approach would be hardly applicable for stochastic algorithms.

(2) *Concerning the suggested method for deleting scenarios*, instead of L_2, another norm can be used to redistribute the probability mass of the deleted scenarios. For instance, L_1 norm criterion for one deleted scenario leads to multiple solutions $\hat{\mathbf{p}}$ and (5) describes one of them.

(3) For equal probabilities $p_s = 1/S \quad \forall s$, the rule (5) gives equal probabilities for scenarios of the reduced problem and (8) means to delete the scenario for which $\mathbf{c}^\top \mathbf{x}^* + Q(\mathbf{x}^*, \omega_s)$ equals approximately the optimal value $\varphi(\mathbf{p}^*)$; for details see Section 2.

(4) Another rule can be used for redistribution, for instance, to keep proportionality of the remaining probabilities or to get the directional derivative equal to 0; see Section 3.

(5) The required precision in (8) will depend on the type of the numerical technique used to solve the two-stage stochastic program. I guess that setting a nonzero level of discrepancy in (8) as well as the intuitive use of local arguments for global conclusions will be acceptable, namely, in connection with numerical techniques that in principle allow for return of deleted scenarios; an example is the stochastic decomposition algorithm [9], see Section 5.

(6) In a straightforward way, a similar approach can be designed for general scenario based *expected utility problems* that leave the second stage hidden, cf. [10], [17] or [5].

(7) The redistribution rule (5) can be extended to *multistage stochastic programs*; however, it is not yet clear how to design a rule for deleting scenarios in this case. A technique for elimination of inessential scenarios for multistage stochastic programs was suggested in [1] and applied in [2]. It is based on optimal Lagrange multipliers, interpreted as marginal EVPI, that are associated with the nonanticipativity constraints and the rule is to delete scenarios for which the value of the

multiplier is low. An extension of the approach developed in this paper to the multistage problems and its comparison with EVPI based reduction of [1], [2] will be a subject of subsequent studies.

In the sequel we shall detail some of these problems and extensions.

2. The case of equal probabilities

Assume now that the probabilities of scenarios are equal to $1/S$ for all scenarios. Then the redistribution formulas (5), (5') give probabilities $1/(S-1)$ and $1/K$ to all kept scenarios in the case of one deleted scenario and $D = S - K$ deleted scenarios, respectively. The corresponding minimal distances are $\sqrt{\frac{1}{(S-1)S}}$ and $\sqrt{\frac{D}{KS}}$.

If there is a unique optimal solution $\mathbf{x}^*$ of the initial problem for probabilities $p_s^* = 1/S, \quad s = 1, \dots, S$ the directional derivative of the optimal value function for deleting $D = S - K$ scenarios $\omega_{s_i}, i = K+1, \dots, S$ equals

$$\varphi'(\mathbf{p}^*; \hat{\mathbf{p}} - \mathbf{p}^*) = \frac{1}{KS}\left[D\sum\nolimits_{i=1}^{K} Q(\mathbf{x}^*, \omega_{s_i}) - K\sum\nolimits_{i=K+1}^{S} Q(\mathbf{x}^*, \omega_{s_i})\right]$$

$$(11) \qquad = \frac{D}{K}\left[\frac{1}{S}\sum\nolimits_{i=1}^{S} Q(\mathbf{x}^*, \omega_i) - \frac{1}{D}\sum\nolimits_{i=K+1}^{S} Q(\mathbf{x}^*, \omega_{s_i})\right]$$

hence, the rule for deleting scenarios

$$(12) \qquad \frac{1}{S}\sum\nolimits_{i=1}^{S} Q(\mathbf{x}^*, \omega_i) - \frac{1}{D}\sum\nolimits_{i=K+1}^{S} Q(\mathbf{x}^*, \omega_{s_i}) \doteq 0$$

In this special case, it is possible to design a simple procedure for detecting scenarios whose elimination is locally inessential:

• If

$$(13) \qquad Q(\mathbf{x}^*, \omega_s) \doteq \varphi(\mathbf{p}^*) - \mathbf{c}^\top \mathbf{x}^*$$

delete scenario ω_s. Indeed, (13) is equivalent to $Q(\mathbf{x}^*, \omega_s) \doteq \frac{1}{S}\sum_{i=1}^{S} Q(\mathbf{x}^*, \omega_i)$, i.e., to (12) for $D = 1$.

If the reduction according the previous rule is not possible, use *pairwise comparisons*:

• For all pairs of scenarios $\omega_i, \omega_j, i, j = 1, \dots, S$ compute averages

$$Q_{ij} = 1/2\,[Q(\mathbf{x}^*, \omega_i) + Q(\mathbf{x}^*, \omega_j)]$$

If

$$(14) \qquad Q_{ij} \doteq \varphi(\mathbf{p}^*) - \mathbf{c}^\top \mathbf{x}^*$$

delete scenarios ω_i and ω_j.

Again, (14) can be written as

$$1/2\,[Q(\mathbf{x}^*,\omega_i)+Q(\mathbf{x}^*,\omega_j)] = \frac{1}{S}\sum\nolimits_{s=1}^{S} Q(\mathbf{x}^*,\omega_s)$$

i.e., (12) for deleting two scenarios, ω_i, ω_j.

Theoretically, one can compute in this way average recourse costs for sets of more than two scenarios, compare them with the average recourse costs for the complete set of S scenarios and to decide on deleting these scenarios if the difference is negligible. For reasons of numerical efficiency, however, this possibility is evidently limited to sets of deleted scenarios of a small cardinality.

If there are multiple optimal solutions of the initial problem the requirement of $\varphi'(\mathbf{p}^*;\hat{\mathbf{p}}-\mathbf{p}^*) \doteq 0$ suggests to delete the scenario ω_s for which

$$\text{(13')} \qquad \max_{\mathbf{x}\in\mathcal{X}^*(\mathbf{p}^*)} [\mathbf{c}^\top\mathbf{x} + Q(\mathbf{x},\omega_s)] \doteq \varphi(\mathbf{p}^*)$$

Accordingly, one eliminates such scenario whose worst performance over the set of original optimal solutions equals approximately the original optimal value $\varphi(\mathbf{p}^*)$. Of course, it is a question how to detect such scenario. Once more, deleting scenario ω_s for which

$$\text{(15)} \qquad \mathbf{c}^\top\mathbf{x} + Q(\mathbf{x},\omega_s) \geq \varphi(\mathbf{p}^*)$$

holds true for an optimal solution $\mathbf{x} \in \mathcal{X}^*(\mathbf{p}^*)$ means decreasing the optimal value; compare with (9').

3. Another rule for redistribution

The previous rules for deleting scenarios have initiated from stability results according to which one should try to fix new probabilities $\hat{\mathbf{p}}$ as close as possible to the original ones. Imagine now another situation: As a result of a sampling strategy or of another rule, a set of D scenarios to be deleted has been already fixed. The problem is how to redistribute their original probability mass p_0^* to the kept $K = S - D$ scenarios taking into account the goal: to keep the value of the objective function at the original optimal solution unchanged as much as possible.

Inspired by the previous results we assume that the probability mass p_0^* of the deleted scenarios $\omega_{s_i}, i = K+1,\ldots,S$ is added to the probabilities $p_{s_i}^*, i = 1,\ldots,K$ of the remaining scenarios so that the new probabilities of the kept scenarios become

$$\hat{p}_{s_i} = p_{s_i}^* + \lambda_i, \quad i = 1,\ldots,K \quad \text{with} \quad 0 \leq \lambda_i \leq 1 - p_{s_i}^* \quad \forall i, \sum\nolimits_i \lambda_i = p_0^*$$

and $\hat{p}_{s_i} = 0$ for deleted scenarios. Our goal is to fix λ_i to get

$$\varphi'(\mathbf{p}^*;\hat{\mathbf{p}}-\mathbf{p}^*) = \sum\nolimits_i \hat{p}_i Q(\mathbf{x}^*,\omega_i) - \sum\nolimits_i p_i^* Q(\mathbf{x}^*,\omega_i) =$$
$$\sum\nolimits_{i=1}^{K} \lambda_i Q(\mathbf{x}^*,\omega_i) - \sum\nolimits_{i=K+1}^{S} p_s^* Q(\mathbf{x}^*,\omega_s) = 0$$

It is an easy task with multiple solutions unless the average recourse costs of the deleted scenarios are extremal in the sense that

$$\frac{1}{p_0^*}\sum_{i=K+1}^{S} p_i^* Q(\mathbf{x}^*, \omega_{s_i}) \notin \operatorname{conv}\{Q(\mathbf{x}^*, \omega_{s_i}), i = 1, \ldots, K\}$$

in which case, there is no redistribution of the required properties. Otherwise, it is sufficient to find two scenarios, say, ω_l, ω_u among the kept scenarios such that

$$Q(\mathbf{x}^*, \omega_l) < \frac{1}{p_0^*}\sum_{i=K+1}^{S} p_i^* Q(\mathbf{x}^*, \omega_{s_i}) < Q(\mathbf{x}^*, \omega_u)$$

and to put

$$\begin{aligned} \lambda_i &= 0 \quad \text{for} \quad i \neq l, u \\ \lambda_u &= \frac{p_0^* Q(\mathbf{x}^*, \omega_l) - \sum_{i=K+1}^{S} p_i^* Q(\mathbf{x}^*, \omega_{s_i})}{Q(\mathbf{x}^*, \omega_u) - Q(\mathbf{x}^*, \omega_l)} \\ \lambda_l &= p_0^* - \lambda_u \end{aligned}$$

The best choice of ω_l, ω_u (in the sense of the minimal distance between $\mathbf{p}^*$ and $\hat{\mathbf{p}}$) is to reach symmetry, i.e., $\lambda_l \doteq \lambda_u$.

4. Deleting scenarios in the progressive hedging algorithm

We shall apply now the explained ideas for designing a criterion for *deleting scenarios for the progressive hedging algorithm*. The postoptimality procedure suggested in [16] allows for changing the probabilities p_s^* but it keeps all of them positive.

In the progressive hedging algorithm one uses individual scenario solutions to get an averaged solution that hedges against all possible scenarios. The original objective functions $f(\mathbf{x}, \omega_s)$ for individual scenarios are augmented by additional terms that are updated in the course of computations:

$$f^\nu(\mathbf{x}, \omega_s) := f(\mathbf{x}, \omega_s) + \mathbf{w}^{\nu-1}(\omega_s)^\top \mathbf{x} + \rho/2\|\mathbf{x} - \hat{\mathbf{x}}^{\nu-1}\|^2$$

The algorithm for the simplest variant of the method as described in [16] consists of the following steps:

Step 0. Initialization: $\mathbf{w}^0(\omega_s) = 0, \hat{\mathbf{x}}^0 = 0, f^0(\mathbf{x}, \omega_s) = 0 \forall s, \rho > 0, \nu = 1$.

Step 1. For $s = 1, \ldots, S$ get $\mathbf{x}^\nu(\omega_s) \in \arg\min\{f^\nu(\mathbf{x}, \omega_s) | \mathbf{x} \in \mathcal{X}_s\}$.

Step 2. Calculate the averaged solution $\hat{\mathbf{x}}^\nu = \sum_s p_s^* \mathbf{x}^\nu(\omega_s)$, update

$$\mathbf{w}^\nu(\omega_s) = \mathbf{w}^{\nu-1}(\omega_s) + \rho[\mathbf{x}^\nu(\omega_s) - \hat{\mathbf{x}}^\nu]$$

so that in all iterations

$$\sum_s p_s^* \mathbf{w}^\nu(\omega_s) = 0$$

and return to the Step 1 with $\nu = \nu + 1$.

The optimality criteria for the problem solved by this algorithm for probabilities $\mathbf{p}^*$ with optimal solution $\mathbf{x}^*(\omega_s) = \mathbf{x}^* \quad \forall s$ and with final weights $\mathbf{w}^*(\omega_s)$ imply (cf. [16]):

$$(i)\mathbf{x}^* \in \arg\min\{f^*(\mathbf{x},\omega_s)|\mathbf{x} \in \mathcal{X}_s\} \quad \forall s$$

$$(ii) f^*(\mathbf{x},\omega_s) := f(\mathbf{x},\omega_s) + \mathbf{x}^\top \mathbf{w}^*(\omega_s) + \rho/2\|\mathbf{x} - \mathbf{x}^*\|^2$$

$$(iii) \sum_s p_s^* \mathbf{w}^*(\omega_s) = 0$$

Deleting scenario ω_s means again postoptimality with respect to a change in the vector of probabilities $\mathbf{p}^*$ that results in $\hat{p}_s = 0$. Redistribution according to (5) together with requirement that condition (iii) remains valid results into the following rule:

Delete the scenario ω_s for which

$$\mathbf{w}^*(\omega_s) \doteq \frac{1}{S-1}\sum_{i\neq s} \mathbf{w}^*(\omega_i) \tag{16}$$

If this is possible, the optimality conditions (i)–(iii) hold true and optimality of the obtained solution $\mathbf{x}^*$ is retained. Notice that for *equal probabilities* the rule (16) reduces to

$$\mathbf{w}^*(\omega_s) \doteq 0 \tag{16'}$$

A natural question is: Could we benefit from a similar rule for deleting scenarios also in the course of the algorithmic solution, i.e., not only at its termination? Indeed, this seems to be possible at least in the considered class of scenario based stochastic linear programs with recourse for which

$$f(\mathbf{x},\omega_s) = \mathbf{c}^\top \mathbf{x} + Q(\mathbf{x},\omega_s)$$

with $Q(\mathbf{x},\omega)$ defined according to (2) and with $\mathcal{X}_s = \mathcal{X} \quad \forall s$. This conclusion is supported by the convergence results (cf. [13], [16]) and by the obvious fact that any scenario solution, say, $\mathbf{x}^\nu(\omega_s)$ is admissible for other scenario subproblems and that it is also implementable from the point of view of the original problem based on all scenarios $\omega_i, i = 1,\ldots,S$.

We suggest to delete scenario ω_s according to the rule

$$\mathbf{w}^\nu(\omega_s) \doteq \frac{1}{S-1}\sum_{i\neq s} \mathbf{w}^\nu(\omega_i) \tag{17}$$

in the instant when $\theta^\nu := \sum_{i=1}^S p_s^* \|\mathbf{x}^\nu(\omega_i) - \hat{\mathbf{x}}^\nu\|^2$ is small enough, i.e., in vicinity of the sought optimal implementable solution. The next step is to redistribute the probability p_s according to (5) and to restart the algorithm with

$$\hat{\mathbf{x}}^\nu = \sum_{i\neq s} \hat{p}_i \mathbf{x}^\nu(\omega_i)$$

keeping the multipliers $\mathbf{w}^\nu(\omega_i)$ for $i \neq s$ unchanged.

The averages $\hat{\mathbf{x}}^{\nu+k}$ obtained in the continuation of the algorithm will converge to the optimal solution of the problem based on the scenarios $\omega_i, i = 1,\ldots,S, i \neq s$, this solution will be admissible for the sth scenario problem and thanks to the mentioned convergence properties, it will make a good suboptimal solution of the original problem based on all scenarios $\omega_i, i = 1,\ldots,S$.

There are many open questions, namely the level of θ^ν that will be sufficient for the convergence based conclusions.

5. Deleting scenarios in the stochastic decomposition algorithm

This Section summarizes the first observations concerning applicability of the main ideas for deleting scenarios for the stochastic decomposition algorithm that are based on discussions with J. Higle and S. Sen.

For the stochastic decomposition algorithm, it is necessary to assume fixed recourse and fixed recourse costs. We shall limit our discussion here to *random right-hand sides*, but this limitation can be apparently removed. The additional assumptions introduced in [9] are compactness of the set of the first-stage solutions $\mathcal{X}$, of the support Ω of the distribution P of random right-hand sides ω and of the set Π of feasible solutions of the dual to the second-stage program. It means that there is a finite set V of vertices of Π such that

$$\Pi = \{\pi | \mathbf{W}^\top \pi \leq \mathbf{q}\} = \text{conv} V$$

All of these sets are assumed to be nonempty and the recourse costs $Q(\mathbf{x}, \omega) \geq 0 \quad \forall \mathbf{x}, \omega$.

The simplified version of the algorithm (without incubement and stoping rules) consists of repeated solution of the second stage subproblems for the already obtained iterate $\mathbf{x}^k$ of the optimal solution and for a new sample point ω_k to get the recourse costs

$$Q(\mathbf{x}^k, \omega_k) = \min \{\mathbf{q}^\top \mathbf{y} | \mathbf{W}\mathbf{y} = \omega_k - \mathbf{T}\mathbf{x}^k, y \geq 0\}$$

and of the master problems

$$\min_{\mathbf{x} \in \mathcal{X}} \left\{ \mathbf{c}^\top \mathbf{x} + \max_{t=1,\dots,k} \left[\alpha_t^k + \mathbf{x}^\top \beta_t^k\right] \right\}$$

It can be summarized as follows:

Step 0. Initialization. Get $\mathbf{x}^1 \in \arg\min_{\mathbf{x} \in \mathcal{X}} \{\mathbf{c}^\top \mathbf{x} + Q(\mathbf{x}, E\omega)\}$

Step 1. Randomly generate ω_k according to probability distribution P and independently on the previously generated scenarios.

Step 2. Evaluate $Q(\mathbf{x}^k, \omega_k)$, add the corresponding optimal vertex π^k of Π to the set V^{k-1} of the vertices recorded in the previous iterations; put $V^k = V^{k-1} \cup \{\pi^k\}$.

Step 3. Construct coefficients of the new cut: Get

$$\pi_t^k \in \arg\max_{\pi \in V^k} \pi^\top (\omega_t - \mathbf{T}\mathbf{x}^k)$$

and put

$$\alpha_k^k + \mathbf{x}^\top \beta_k^k = \frac{1}{k} \sum_{t=1}^k (\omega_t - \mathbf{T}\mathbf{x})^\top \pi_t^k$$

Update the cuts constructed in the preceding iteration:

$$\alpha_t^k := \frac{k-1}{k} \alpha_t^{k-1}, \quad \beta_t^k := \frac{k-1}{k} \beta_t^{k-1}, \quad t = 1, \dots, k-1$$

Step 4. Solve the kth master problem to obtain $\mathbf{x}^{k+1}$. Repeat from Step 1 with $k \leftarrow k+1$.

In the course of this algorithm, the last updates of the coefficients α_t^k and $\beta_t^k \quad \forall t$, the set of vertices V^k and all sample values ω_t have to be stored. Following the previous arguments we suggest to *delete in the kth iteration the scenario* ω_s for which

$$\omega_s^\top \pi_s^k \doteq \alpha_k^k \tag{18}$$

The coefficients β_k^k obtained in Step 3 do not depend on the sample values ω_t so that the new cut will not be changed when deleting ω_s according to (18). The set of vertices V^k and the former cuts are kept and updated according to Step 3 and in the extended version of the algorithm, the cut connected with the current incubement will be updated regardless the deleted sample point ω_s. The rule can be extended to deleting more than one scenario in the given iteration similar to (14): Delete the pair of scenarios ω_i, ω_j for which

$$1/2 \left[\omega_i^\top \pi_i^k + \omega_j^\top \pi_j^k\right] \doteq \alpha_k^k \tag{19}$$

and so on.

Acknowledgement. When preparing this paper, the author benefited from discussions with Julia Higle and Suvrajeet Sen during her visit of University of Arizona in Tucson in August 1994 and with Andrzej Ruszczynski during her visits of IIASA in September 1994 and in February 1995.

References

(1) M. A. H. Dempster: On stochastic programming II: Dynamic problems under risk. Stochastics 25 (1988) 15–42.

(2) M. A. H. Dempster and A. M. Ireland: A financial expert decision system. In: G. Mitra (ed.), Mathematical Models for Decision Support, NATO ASI Ser. Vol 48 (Springer, Berlin, 1988) p. 415–440.

(3) J. Dupačová: Postoptimality for multistage stochastic programs. To appear in Annals of Oper. Res. (Udine 1992).

(4) J. Dupačová: Scenario based stochastic programs: Resistance with respect to sample. Submitted for Annals of Oper. Res. (Lillehammer 1994).

(5) J. Dupačová and M. Bertocchi: Management of bond portfolios via stochastic programming – Postoptimality and sensitivity analysis. Paper submitted for the IFIP Conference on System Modelling and Optimization, Prague, July 1995.

(6) Yu. Ermoliev: Stochastic quasigradient methods and their application to systems optimization. Stochastics 9 (1983) 1 –36.

(7) Yu. Ermoliev and R. J.-B. Wets (eds.): Numerical Techniques for Stochastic Optimization (Springer, Berlin, 1988).

(8) E. G. Gol'shtein: Theory of Convex Programming, Translations of Mathematical Monographs 36 (American Mathematical Society, Providence RI, 1972).

(9) J. L. Higle and S. Sen: Stochastic decomposition: An algorithm for two stage linear programs with recourse. Mathematics of Oper. Res. 16 (1991) 650–669.

(10) M. Holmer et al.: Dynamic models for fixed-income portfolio management under uncertainty. RR 93–11–01, Operations & Information Management Dept., The Wharton School, Univ. of Pennsylvania, Philadelphia, 1993.

(11) P. Kall and S. W. Wallace: Stochastic Programming (Wiley-Interscience Series in Systems and Optimization, Chichester, 1994).

(12) S. M. Robinson and R. J.-B. Wets: Stability in two-stage stochastic programming. SIAM J. Control and Optimization 25 (1987) 1409–1416.

(13) T. R. Rockafellar and R. J.-B. Wets: Scenarios and policy aggregation in optimization under uncertainty. Math. of Oper. Res. 16 (1991) 119–147.

(14) R. J.-B. Wets: On the continuity of the value of a linear program and of related polyhedral-valued multifunction. Math. Programming Study 24 (1985) 14–29.

(15) R. J.-B. Wets: Large scale linear programming techniques in stochastic programming. In: [7] p. 65–93.

(16) R. J.-B. Wets: The aggregation principle in scenario analysis and stochastic optimization. In: S. W. Wallace (ed.), Algorithms and Model Formulations in Mathematical Programming (Springer, Berlin, 1989) p. 91–113.

(17) S. A. Zenios: Massively parallel computations for financial planning under uncertainty. In: J. Mesirov (ed.), Very Large Scale Computing in the 21-st Century (SIAM, Philadelphia, 1991) p. 273–294.