



## A simple heuristic for reducing the number of scenarios in two-stage stochastic programming

Ramkumar Karuppiah, Mariano Martín, Ignacio E. Grossmann\*

Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

### ARTICLE INFO

#### Article history:

Received 10 April 2009

Received in revised form

22 September 2009

Accepted 13 October 2009

Available online 21 October 2009

#### Keywords:

Two-stage stochastic programming

Multiscenario model

Scenario reduction

Stochastic approximation

### ABSTRACT

In this work we address the problem of solving multiscenario optimization models that are deterministic equivalents of two-stage stochastic programs. We present a heuristic approximation strategy where we reduce the number of scenarios and obtain an approximation of the original multiscenario optimization problem. In this strategy, a subset of the given set of scenarios is selected based on a proposed criterion, and probabilities are assigned to the occurrence of scenarios in the reduced set. The original stochastic programming model is converted into a deterministic equivalent using the reduced set of scenarios. A mixed-integer linear program (MILP) is proposed for the reduced scenario selection. We apply this practical heuristic strategy to four numerical examples and show that reformulating and solving the stochastic program with the reduced set of scenarios yields an objective value close to the optimum of the original multiscenario problem.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

Optimization under uncertainty is a major issue in solving real world problems. Uncertainty is a common feature that presents itself during the operation or design of any system. There is an abundance of literature in the area of optimization under uncertainty involving several applications. Some of these include: production planning (Cheng, Subrahmanian, & Westerberg, 2003; Clay & Grossmann, 1997), scheduling (Balasubramanian & Grossmann, 2002; Birge & Dempster, 1996), optimal chemical process synthesis (Acevedo & Pistikopoulos, 1998; Liu & Sahinidis, 1996; Rooney & Biegler, 2003), electricity production (Nowak, Schultz, & Westphalen, 2005; Takriti, Birge, & Long, 1996). Usually problems with uncertainty are represented as stochastic programming problems (Birge & Louveaux, 1997) or as deterministic flexibility problems (Grossmann, Halemane, & Swaney, 1983). The focus of this work is on solving two-stage stochastic programs with recourse, where we have some uncertain parameters that either follow a continuous distribution or take on a finite set of values. The aim in such problems is to determine the 1st stage decision variables such that the sum of the 1st stage costs and the expected value of the 2nd stage costs is minimized. Other approaches for solving problems under uncertainty include robust optimization, probabilistic programming, fuzzy optimization, and dynamic programming. Sahinidis (2004) presents a recent review of problems

under uncertainty along with the approaches used to solve such problems.

There are different methods for solving stochastic programming problems. Algorithms for stochastic integer programs have been presented by Ahmed, Tawarmalani, and Sahinidis (2004), Carøe and Tind (1997), Carøe and Schultz (1999), Klein Haneveld, Stougie, and van der Vlerk (1995, 1996) among other authors. Norkin, Pflug, and Ruszczyński (1998) developed a branch and bound technique for global optimization of nonconvex nonlinear stochastic programs, where stochastic lower and upper bounds are made to converge with some confidence levels. In two-stage stochastic programming with recourse, a common approach is to discretize the uncertain parameter space and formulate a deterministic equivalent of the stochastic program, which leads to a multiscenario optimization problem (Dantzig, 1963). A single combination of the values of the uncertain parameters leads to a particular scenario. In this discretization approach there are usually a number of uncertain parameters in a system and these are assumed to take on a finite set of values. All possible combinations of these values lead to an explosion in the number of scenarios. This greatly increases the size of the optimization problem, making it very hard to solve.

The multiscenario problems represent an approximation of the original stochastic problems and they are usually difficult to solve directly due to their large dimension. In this work we address the problem of reducing the number of scenarios in multiscenario optimization problems. To deal with this problem, two approaches can be used: (i) optimization methods, which exploit the structure of the problem, or (ii) aggregation techniques, which reduce the dimensions of the original scenario tree (Barro & Canestrelli, 2005).

\* Corresponding author. Tel.: +1 412 268 3642; fax: +1 412 268 7139.  
E-mail address: [grossmann@cmu.edu](mailto:grossmann@cmu.edu) (I.E. Grossmann).

Scenario aggregation techniques follow different principles such as random or Quasi Monte-Carlo sampling (Shapiro, 2003). An aggregation technique presented by Hoyland, Kaut, and Wallace (2003) is based on moment matching that uses nonconvex optimization to generate scenarios, which match a specified set of statistical properties.

Another alternative is the use of probability metrics (see Dupačová, Gröwe-Kuska, & Römisch, 2003; Heitsch & Römisch, 2003; Henrion & Römisch, 1999; Rachev & Römisch, 2002). Some of these scenario reduction algorithms determine a scenario subset (of prescribed cardinality or accuracy) and assign new probabilities to the preserved scenarios. This is such that the corresponding reduced probability measure (the probability distribution of the approximation process) is the closest to the original measure (the probability distribution of the original empirical process) in terms of a certain probability distance between the original and the approximation probability distribution.

There are also scenario tree construction algorithms that successively reduce the number of nodes of a subtree of individual scenarios by modifying the tree structure and by bundling similar scenarios. It makes use of the Fortet-Mourier metric and the duality theory in linear programming (LP) to compute the distance between two probability measures, one corresponding to a huge scenario tree with all the sequences, and another to a smaller scenario tree of fixed cardinality.

Some heuristic approximation methods have been developed to solve the stochastic program with fewer scenarios and still obtain a close to optimal solution. Novak and Kravanja (1999) have presented a reduced dimensional stochastic optimization technique where they determine a subset of the vertices of the feasible polyhedral space of the uncertain parameters and their corresponding weights to approximate the expected value of the objective function of the original problem.

In this paper, we use a similar idea as given in Novak and Kravanja (1999), and Dupačová et al. (2003) to select a subset of scenarios from a given larger set, for solving the stochastic program. The goal is that the optimal objective of the full scenario problem is closely approximated by the optimal objective value of the reduced problem. A mixed-integer linear programming (MILP) model is presented for the selection of the subset of scenarios. The remainder of the paper is organized as follows. Section 2 presents the problem statement, while the heuristic strategy to approximate the original multiscenario problem is given in Section 3. Numerical examples on which the approach was applied are given in Section 4, and finally Section 5 summarizes the conclusions.

## 2. Problem statement

We are given a two-stage stochastic program whose deterministic equivalent has  $|S|$  separate scenarios with different realizations of uncertain parameters. Each of these scenarios has a certain probability of occurrence. The uncertain parameters that make up these scenarios take on a finite set of values. The probabilities of this finite set of values for each uncertain parameter add up to 1. This discrete finite set is either given, or else it can be computed from a continuous distribution (see Luceno, 1999).

The goal of this paper is to develop an approach where we can select a subset  $S'$  of scenarios from the original set of scenarios ( $S$ ) with new probabilities given to each of the  $|S'|$  scenarios, and approximate the optimal objective value of original multiscenario problem as closely as possible with the reduced number of scenarios. This means that on solving the reduced dimensional problem (with fewer scenarios), we get an objective value close to one of the original multiscenario problem with  $|S|$  scenarios. The problem at hand is to devise an MILP (or a linear programming) formulation that allows us to select a subset of the scenarios, and gives us their

associated probabilities that would help in approximating the original optimization problem. We are also interested in getting some bounds on the theoretical error estimates.

## 3. Approximation strategy

Two-stage stochastic programs are often converted to deterministic multiscenario optimization problems, by discretizing the uncertain parameters with a finite set of scenarios. Such problems grow larger with the number of scenarios. A multiscenario model with a scenario set  $S$  can be expressed as follows:

$$\begin{aligned} \min_{d, x_s} \quad & f^0(d) + \sum_s p_s f_s(x_s, \bar{\theta}_s) \\ \text{s.t.} \quad & \left. \begin{aligned} h_s(d, x_s, \bar{\theta}_s) &= 0 \\ g_s(d, x_s, \bar{\theta}_s) &\leq 0 \end{aligned} \right\} s \in S \\ & d \in D, x_s \in X, \bar{\theta}_s \in \Theta \end{aligned} \quad (P)$$

where  $s \in S$  is a single scenario in the multiscenario problem.  $d$  is the set of 1st stage decision variables, while  $x_s$  is the set of 2nd stage variables in scenario  $s$ .  $\bar{\theta}_s$  is the vector of uncertain parameters in scenario  $s$ .  $h(\cdot) = 0$  and  $g(\cdot) \leq 0$  include the first and second stage constraints. Our goal is to approximate the original set with  $|S|$  scenarios with a set with  $|S'|$  scenarios so that we have a smaller multiscenario problem that yields close to the optimal objective value.

### 3.1. Selection of subset of scenarios

In a multiscenario problem let  $\theta = \{\theta_i\}_{i=1, \dots, J}$  be the vector of uncertain parameters. Let the uncertain parameter  $\theta_i$  take on a finite set of values given by  $\{\theta_i^j\}_{j=1, \dots, J_i}$ . The probability associated with the uncertain parameter  $\theta_i$  taking on a value  $\theta_i^j$  is  $p_i^j$ . With multiple uncertain parameters, these can be combined together by considering the corresponding Cartesian product of all the values of the uncertain parameters to yield the set with  $|S|$  scenarios. The scenario  $s$  involves the following vector of uncertain parameters  $\bar{\theta}_s = \{\theta_i^j\}_{i=1, \dots, J}$ , and there are a total of  $|S| = \prod_{i=1}^J J_i$  scenarios. Assuming independent distributions, the probability associated with a scenario  $s$  in the original set of scenarios is given by  $p_s = p_{1j_1, 2j_2, \dots, Jj_J} = \prod_{i=1}^J p_i^{j_i}$ .

In order to select a minimum subset from the original set of scenarios, we propose the following heuristic criterion:

*Criterion:* The sum of the probabilities of the new scenarios in which the uncertain parameter value  $\theta_i^j$  appears is equal to  $p_i^j$ .

The above criterion should hold for all of the values of each of the uncertain parameters. Furthermore, the sum of the probabilities of the reduced set of scenarios should be equal to 1. We want to re-arrange the scenario probabilities in such a way that the overall probability of occurrence of a particular value of an uncertain parameter across different scenarios matches the probability of occurrence of that value for the given uncertain parameter. The motivation behind such an approximation is to heuristically reduce the approximation error. If the objective function of a multiscenario optimization formulation can be approximated as a sum of the functions of the individual uncertain parameter values multiplied by the respective probabilities of the scenarios in which they occur, its value will be close to  $\sum_s p_s \cdot \sum_{i=1}^I f_i(\theta_i^s) = \sum_{i=1}^I \sum_{j=1}^{J_i} p_i^j \cdot f_i(\theta_i^j)$ . The re-assignment of probabilities to selected scenarios based on the above-mentioned criterion and solving the optimization problem will then give an objective function that can be approximated by  $\sum_{s'} p_{s'} \cdot \sum_{i=1}^I f_i(\theta_i^{s'}) = \sum_{i=1}^I \sum_{j=1}^{J_i} p_i^j \cdot f_i(\theta_i^j)$ . Since the objective functions of the original and the reduced scenario problem can be approximated by the same expression, the difference between their values is also expected to be small.

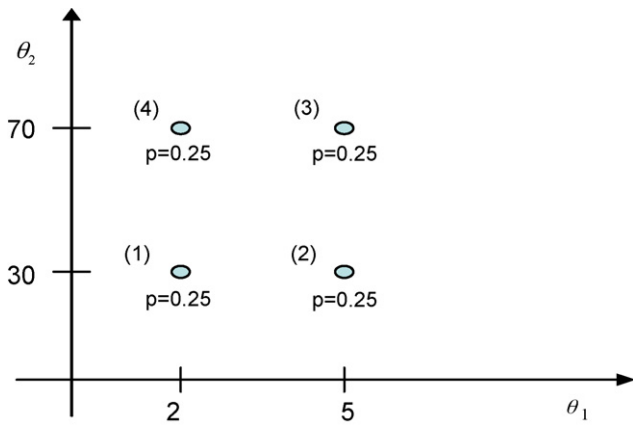


Fig. 1. Scenarios in illustrative example.

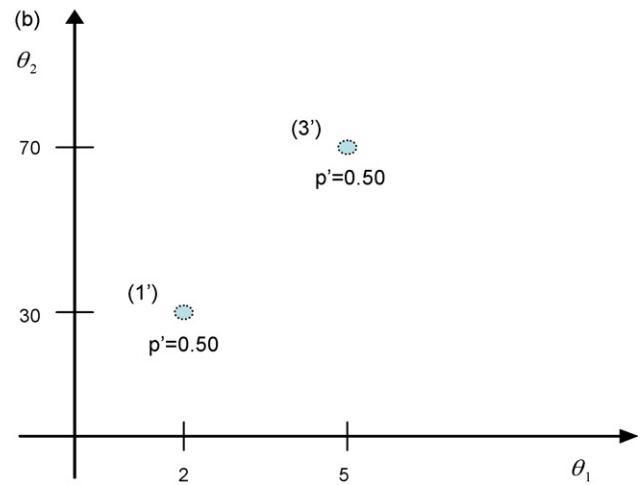
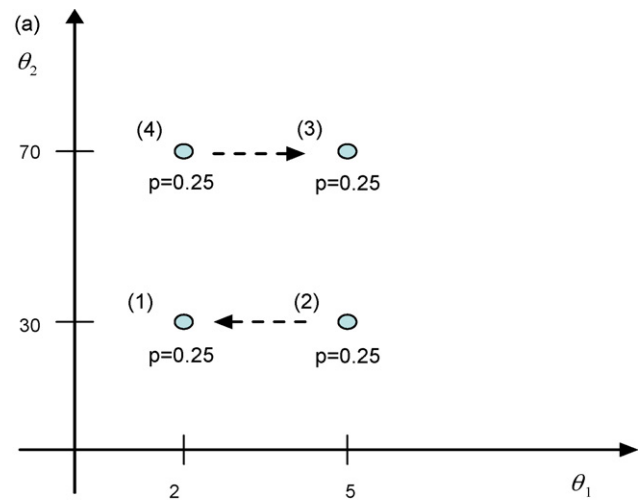


Fig. 2. (a) Re-combination of scenarios in illustrative example; (b) scenarios with modified probabilities in illustrative example.

To illustrate the scenario reduction approach with a small example, consider two uncertain parameters  $\theta_1$ , and  $\theta_2$ , where each parameter can take on two values. Let  $\theta_1$  take the two values  $\{2,5\}$  each occurring with a probability of 0.5, and let parameter  $\theta_2$  take a value of 30 with a probability of 0.5 and a value of 70 with a probability of 0.5. We obtain a set of four scenarios  $\{(2,30), (2,70), (5,30), (5,70)\}$  that result from the Cartesian product of  $\{2,5\}$  and  $\{30,70\}$ . These are shown in Fig. 1 and denoted by (1), (2), (3) and (4). The probability of occurrence of each of these scenarios is 0.25, which is obtained by multiplying the probabilities of the uncertain parameter values in each scenario.

Looking at the  $\theta_1$  axis, we find that  $\theta_1$  takes a value of 2 in scenarios (1) and (4), where each of these scenarios occurs with a probability of 0.25, thus making the overall probability of occurrence of the value 2 for  $\theta_1$  to be 0.5. The creation of scenarios has separated the value of 2 taken by  $\theta_1$  into different scenarios. However, the creation of scenarios has ensured that the sum of the probabilities of the scenarios in which  $\theta_1$  takes a value of 2 is the same as the occurrence probability of  $\theta_1 = 2$ , which is 0.5. The same analysis is true for the value of 5 taken on by  $\theta_1$ . When looking at the  $\theta_2$  axis, we can find an identical analysis for the values taken on by  $\theta_2$ .

To reduce the number of scenarios, we use the idea of reversing this disaggregation of uncertain parameter values and their probabilities, and combining back the scenarios so that the probabilities of occurrence of the individual uncertain parameter values remain intact. One possible re-combination is shown in Fig. 2a, where scenario (2) is combined with scenario (1), while scenario (4) is combined with scenario (3), leading to the new scenarios (1') and (3') in Fig. 2b.

In Fig. 2b, the individual uncertain parameter values have the same probability of occurrence through the scenarios (1') and (3'), as when the scenarios had not even been created. For instance, looking at the  $\theta_1$  axis, we see that  $\theta_1$  takes a value of 2 only in scenario (1'), which has a probability of 0.5 ensuring the probability of occurrence of the value 2 for  $\theta_1$  to be 0.5. Similarly for  $\theta_1 = 5$ , this value now occurs only in scenario (3') whose probability is 0.5, meaning that  $\theta_1 = 5$  occurs with a probability of 0.5 in the overall system, which is the probability of occurrence of the value 5 for  $\theta_1$ . A similar analysis holds for the values on the  $\theta_2$  axis.

The minimum number of reduced scenarios that we can obtain depends also on the individual probabilities of the values taken on by the uncertain parameters. To illustrate this, let  $\theta_1$  now take two values  $\{2,5\}$  each occurring with corresponding probabilities of 0.3 and 0.7, respectively. The parameter  $\theta_2$  takes a value of 30 with a probability of 0.6 and a value of 70 with a probability of 0.4. Com-

Table 1  
Scenarios for two uncertain parameters in illustrative example.

Scenario, $s$	$\theta_1$	$\theta_2$	Probability of scenario, $p_s$
1	2	30	0.18
2	2	70	0.12
3	5	30	0.42
4	5	70	0.28

binning these two uncertain parameters, we obtain four scenarios as shown in Table 1.

Now on selecting a subset of scenarios from the given set in Table 1, we obtain a reduced set of scenarios  $S'$  (see Table 2) that can be used for approximating the original multisenario problem (P).

The aforementioned criterion is satisfied by the scenarios  $s' \in S'$ . For instance, in the original problem,  $\theta_1$  takes a value of 5 with a probability 0.7. In the reduced set of scenarios  $S'$ ,  $\theta_1$  takes a value of 5

Table 2  
Reduced set of scenarios for two uncertain parameters in illustrative example.

Scenario, $s'$	$\theta_1$	$\theta_2$	Probability of scenario, $p_{s'}$
1'	2	30	0.3
2'	5	30	0.3
3'	5	70	0.4

in  $s' = 2'$  and in  $s' = 3'$ , and the sum of the probabilities of occurrence of  $s' = 2'$  and  $s' = 3'$  is 0.7 ( $=0.3+0.4$ ), ensuring that the proposed criterion holds. The same logic holds for each of the values of both the uncertain parameters.

To model the selection of scenarios, let the new probability assigned to a scenario  $s$  be  $\hat{p}_{1j_1, 2j_2, \dots, lj_l}$ , a continuous variable. The binary variable  $w_{1j_1, 2j_2, \dots, lj_l}$  corresponds to the existence of the scenario  $s$  in the new set of scenarios. The MILP formulation to determine the minimum number of scenarios satisfying the proposed criterion is as follows:

$$\begin{aligned} \min f &= \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_l=1}^{J_l} w_{1j_1, 2j_2, \dots, lj_l} \\ \text{s.t.} & \\ & \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} \cdots \sum_{j_l=1}^{J_l} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = p_1^1, \quad j_1 = 1, \dots, J_1 \\ & \sum_{j_1=1}^{J_1} \sum_{j_3=1}^{J_3} \cdots \sum_{j_l=1}^{J_l} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = p_2^2, \quad j_2 = 1, \dots, J_2 \\ & \vdots \\ & \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_{l-1}=1}^{J_{l-1}} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = p_l^l, \quad j_l = 1, \dots, J_l \\ & \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_l=1}^{J_l} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = 1 \\ & \hat{p}_{1j_1, 2j_2, \dots, lj_l} \leq w_{1j_1, 2j_2, \dots, lj_l}, \quad \forall j_1, j_2, \dots, j_l \\ & 0 \leq \hat{p}_{1j_1, 2j_2, \dots, lj_l} \leq 1, \quad \forall j_1, j_2, \dots, j_l \\ & w_{1j_1, 2j_2, \dots, lj_l} \in \{0, 1\}, \quad \forall j_1, j_2, \dots, j_l \end{aligned} \tag{SG}$$

On solving the MILP model (SG) we obtain the minimum set of scenarios and their associated probabilities. The numerical value of the probability corresponding to a scenario  $s$  with the uncertain parameters  $\{\theta_1^1, \theta_2^2, \dots, \theta_l^l\}$  in the reduced set of scenarios is  $\hat{p}_{1j_1, 2j_2, \dots, lj_l}^*$ .

**Remarks**

1. Since the problem (SG) can yield a very large MILP problem, we can consider instead a linear programming relaxation (SG-L) to obtain a set of scenarios satisfying the proposed criterion. This can be done by eliminating the binary variables and modifying the objective function as given in the formulation below:

$$\begin{aligned} \min f &= \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_l=1}^{J_l} (1 - p_1^1, p_2^2, \dots, p_l^l) \cdot \hat{p}_{1j_1, 2j_2, \dots, lj_l} \\ \text{s.t.} & \\ & \sum_{j_2=1}^{J_2} \sum_{j_3=1}^{J_3} \cdots \sum_{j_l=1}^{J_l} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = p_1^1, \quad j_1 = 1, \dots, J_1 \\ & \sum_{j_1=1}^{J_1} \sum_{j_3=1}^{J_3} \cdots \sum_{j_l=1}^{J_l} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = p_2^2, \quad j_2 = 1, \dots, J_2 \\ & \vdots \\ & \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_{l-1}=1}^{J_{l-1}} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = p_l^l, \quad j_l = 1, \dots, J_l \\ & \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \cdots \sum_{j_l=1}^{J_l} \hat{p}_{1j_1, 2j_2, \dots, lj_l} = 1 \\ & 0 \leq \hat{p}_{1j_1, 2j_2, \dots, lj_l} \leq 1, \quad \forall j_1, j_2, \dots, j_l \end{aligned} \tag{SG-L}$$

The weights in the new objective function involve the known probabilities of the existing set of scenarios and are present to drive the optimization to reduce the number of scenarios while trying to keep the original set of scenarios that had relatively larger probabilities. The solution will be a subset of the initial set of  $|S|$  scenarios, although it is not guaranteed to be the minimum number of scenarios since the problems (SG) and (SG-L) are not equivalent. It is observed through the numerical examples that the solution times for (SG) and (SG-L) are negligible.

2. It is also possible to assign weights to the individual terms in (SG) to help select scenarios with new probabilities close to their original probabilities.
3. In this method for determining a smaller number of scenarios, it may be possible to identify the worst-case scenarios from among the given discrete set of scenarios (that guarantee feasibility of design for all the given discrete scenarios). If such scenarios exist, and are easily identified, they can be included in the reduced set of scenarios. The more general approach would be to use a method for flexibility analysis to rigorously identify such a worst point (Grossmann et al., 1983). In the MILP formulation (SG), this would mean assigning a lower bound on the probability of the worst-case scenarios (if known) as  $p_{\text{worst-case scenarios}} \geq \varepsilon$ , where  $\varepsilon$  is a small positive number less than or equal to 1. The binary variables  $w_{\text{worst-case scenarios}}$  are fixed to a value of 1.
4. The theoretical minimum number of scenarios that can be obtained using this method is the maximum of the number of independent values that each uncertain parameter can take. The other limiting case is that if no value of an uncertain parameter occurs in more than a single scenario in the set  $S$ , the number of scenarios cannot be reduced with this method.

**3.2. Reduced scenario optimization**

The stochastic optimization problem (P') that uses a reduced set of scenarios  $|S'|$  is as follows:

$$\begin{aligned} \min_{d, x_{s'}} z' &= f^0(d) + \sum_{s'} p_{s'} f_{s'}(x_{s'}, \bar{\theta}_{s'}) \\ \text{s.t.} & \left. \begin{aligned} h_{s'}(d, x_{s'}, \bar{\theta}_{s'}) &= 0 \\ g_{s'}(d, x_{s'}, \bar{\theta}_{s'}) &\leq 0 \end{aligned} \right\} s' \in S' \\ & d \in D, x_{s'} \in X, \bar{\theta}_{s'} \in \Theta \end{aligned} \tag{P'}$$

All the functions in (P') have the exact same form as the corresponding functions in (P).  $p_{s'}$  is the probability of a selected scenario  $s'$  obtained by solving (SG) or (SG-L). The optimal value of the design variable vector obtained by solving (P') is denoted by  $\hat{d}$ , and the optimal expected objective value by  $z^*$ . In case the worst-case scenarios from the set  $S$  are included in the set  $S'$ , then  $\hat{d}$  will be feasible for every scenario in the original scenario set  $S$  (Grossmann et al., 1983). We can also solve the original problem (P) by fixing the design variables in (P) to the value  $\hat{d}$ . Note that this makes the model (P) decomposable into  $|S|$  separate optimization subproblems with each subproblem corresponding to a single scenario. Solving (P) by fixing the design variables to the value  $\hat{d}$  gives us a locally optimal solution to the original problem where the optimal objective value obtained using this method is  $\hat{z}^*$ .

For practical purposes, we can obtain a bound on the error in such an approximation as follows:

- (a) The expected objective value is computed by solving each scenario (SP<sub>s</sub>) separately (e.g. wait-and-see approach) or may be approximated by taking a subset of scenarios with larger probabilities. The optimal objective values of all considered scenarios  $s \in S$  are summed to obtain  $z^{B*} = \sum_{s \in S} z^{s*}$ .
- (b) The objective value of the best found feasible solution so far is  $\hat{z}^*$  (this value is either the global minimum or higher than it).

**Table 3**  
Reduced number of scenarios for example 1.

Scenario, $s'$	$\theta_1$	$\theta_2$	Probability of scenario, $p_{s'}$
1'	1	1	0.1
2'	1.5	1.5	0.6
3'	2	2	0.2
4'	1	1.5	0.1

The value of  $EB = \tilde{z}^* - z^{B*}$  is calculated, which is an upper bound on the error using the approximation technique. Note that this bound may be loose.

**4. Numerical examples**

The proposed scenario reduction and approximation approach is applied to four examples. The optimization problems are formulated using GAMS (Brooke, Kendrick, Meeraus, & Raman, 1998) and solved on an Intel Centrino duo Windows machine with 2 GB memory. The LP and MILP problems are solved using GAMS/CPLEX 11.2 while GAMS/CONOPT 3.0 is used for the nonlinear programming (NLP) problems.

*Example 1.* This small example is taken from Clay and Grossmann (1997) and corresponds to problem (EX1) in that paper. It is a stochastic program with 2 uncertain parameters  $\{\theta_1, \theta_2\}$  and one 1st stage decision variable ( $d$ ). In order to convert this problem into its deterministic equivalent, the uncertain parameters are assigned three values each, which leads to the creation of the LP deterministic equivalent with 9 scenarios. The two uncertain parameters are given by  $\theta_1$  and  $\theta_2$ , each with 3 values and associated probabilities as follows:

$$\theta_1^i = \{1, 1.5, 2\}, \quad p_1^i = \{0.2, 0.6, 0.2\}$$

$$\theta_2^j = \{1, 1.5, 2\}, \quad p_2^j = \{0.1, 0.7, 0.2\}$$

The LP deterministic equivalent is as follows:

$$\begin{aligned} \min z &= d + \sum_{j_1=1}^3 \sum_{j_2=1}^3 p_{1j_1,2j_2} (x_{1j_1j_2} + x_{2j_1j_2}) \\ \text{s.t.} \quad &x_{1j_1j_2} \geq 3\theta_1^i - \theta_2^j, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \\ &x_{2j_1j_2} \geq -\theta_1^i + 2\theta_2^j, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \\ &x_{1j_1j_2} + x_{2j_1j_2} - d \leq \theta_1^i + \theta_2^j, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \\ &d \in \mathfrak{R}_+^1 \\ &x_{1j_1j_2} \in \mathfrak{R}_+^1, x_{2j_1j_2} \in \mathfrak{R}_+^1, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \end{aligned} \tag{E1}$$

In the above formulation,  $x_{1j_1j_2}$  and  $x_{2j_1j_2}$  are the continuous 2nd stage variables and  $p_{1j_1,2j_2} = p_1^i \cdot p_2^j$ .

The LP formulation for the deterministic equivalent (E1) with 9 scenarios has 19 continuous variables and 27 constraints. Solving this model yields an optimal objective value of 10.1, where the optimal value of the 1st stage variables is 4.0. We apply the scenario reduction technique to this problem and obtain the 4 scenarios in Table 3.

The corresponding reduced LP problem has 4 scenarios, 9 continuous variables and 12 constraints. It is to be noted that an inspection of the values of the uncertain parameters and their corresponding probabilities is used in determining the reduced set of scenarios, such that the criterion given in Section 3.1 is satisfied. The MILP formulation (SG) is not used to select the scenarios in this example. On solving the reduced scenario problem, we obtain the optimum value of 10.1, and the value of the 1st stage variable is again 4.0. This means that we have a zero approximation error

**Table 4**  
Reduced number of scenarios for example 2 obtained from solving model (SG-E2).

Scenario, $s'$	$\theta_1$	$\theta_2$	Probability of scenario, $p_{s'}$
1'	1	3	0.2
2'	2	2	0.5
3'	3	1	0.2
4'	3	2	0.1

in this case. Note however that the reduced set of scenarios is not unique. Solving the model (P') with different scenarios with different probabilities could potentially lead to a value of the design variable that is infeasible for the original problem (P). A design variable obtained by solving the approximate model (P') will be feasible for the original problem only if the worst-case scenarios from the original set of 9 scenarios are included in the reduced set of scenarios used in formulating (P').

*Example 2.* We solve the model (EX2P) taken from Clay and Grossmann (1997) as a next example. This is an LP with 10 continuous variables and 18 constraints. It has 2 uncertain parameters that are assumed to take on 3 values, each leading to a total of 9 scenarios. The two uncertain parameters and their distributions are given below:

$$\theta_1^i = \{1, 2, 3\}, \quad p_1^i = \{0.2, 0.5, 0.3\}$$

$$\theta_2^j = \{1, 2, 3\}, \quad p_2^j = \{0.2, 0.6, 0.2\}$$

The formulation corresponding to this example is as follows,

$$\begin{aligned} \min z &= d + \sum_{j_1=1}^3 \sum_{j_2=1}^3 p_{1j_1,2j_2} \cdot x_{2j_1j_2} \\ \text{s.t.} \quad &d + x_{2j_1j_2} \geq \theta_1^i, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \\ &2d + x_{2j_1j_2} \geq \theta_2^j, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \\ &d \in \mathfrak{R}_+^1 \\ &x_{2j_1j_2} \in \mathfrak{R}_+^1, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \end{aligned} \tag{E2}$$

where  $d$  and  $x_{2j_1j_2}$  are the continuous 1st and 2nd stage variables, respectively, and the probability  $p_{1j_1,2j_2} = p_1^i \cdot p_2^j$ . Solving this model, we obtain an optimum expected value of 1.6333, and the optimal value of the 1st stage variable is 0.6667. Using the proposed scenario selection approach, we can obtain a minimum of 4 scenarios satisfying the proposed probability criterion in Section 3.1. The formulation (SG) corresponding to this example is given below,

$$\begin{aligned} \min f &= \sum_{j_1=1}^3 \sum_{j_2=1}^3 w_{1j_1,2j_2} \\ \text{s.t.} \quad &\hat{p}_{11,21} + \hat{p}_{11,22} + \hat{p}_{11,23} = 0.2 \\ &\hat{p}_{12,21} + \hat{p}_{12,22} + \hat{p}_{12,23} = 0.5 \\ &\hat{p}_{13,21} + \hat{p}_{13,22} + \hat{p}_{13,23} = 0.3 \\ &\hat{p}_{11,21} + \hat{p}_{12,21} + \hat{p}_{13,21} = 0.2 \\ &\hat{p}_{11,22} + \hat{p}_{12,22} + \hat{p}_{13,22} = 0.6 \\ &\hat{p}_{11,23} + \hat{p}_{12,23} + \hat{p}_{13,23} = 0.2 \\ &\sum_{j_1=1}^3 \sum_{j_2=1}^3 \hat{p}_{1j_1,2j_2} = 1 \\ &\hat{p}_{1j_1,2j_2} \leq w_{1j_1,2j_2}, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \\ &0 \leq \hat{p}_{1j_1,2j_2} \leq 1, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \\ &w_{1j_1,2j_2} \in \{0, 1\}, \quad j_1 = 1, \dots, 3, j_2 = 1, \dots, 3 \end{aligned} \tag{SG-E2}$$

**Table 5**  
Values and probabilities for uncertain parameters  $T_3, T_5, T_9$ .

$j$	$T_3^j$ (°C)	$p(T_3^j)$	$T_5^j$ (°C)	$p(T_5^j)$	$T_9^j$ (°C)	$p(T_9^j)$
1	378.9	0.007	573.9	0.007	293.5	0.007
2	382.6	0.1545	577.6	0.1545	295.3	0.1545
3	388.0	0.677	583.0	0.677	298.0	0.677
4	393.4	0.1545	588.4	0.1545	300.7	0.1545
5	397.1	0.007	592.1	0.007	302.5	0.007

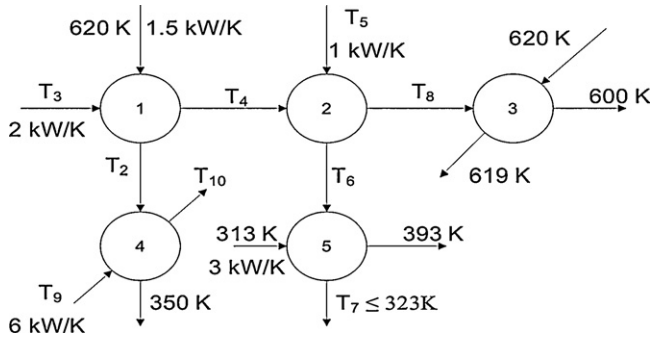


Fig. 3. Heat exchanger network for example 3.

Table 4 shows the scenarios obtained by solving the above formulation (SG-E2).

On solving the reduced dimensional model (P') with the four scenarios shown in Table 4 obtained by solving model (SG-E2), we obtain an optimal value of  $z^* = 1.6833$ , and the optimal value of the design variable is found to be 0.667. Fixing the value of the design variable  $d$  to 0.667 in model (E2) and re-solving it we obtain the

optimal objective value of  $\tilde{z}^* = 1.6333$ , which is the same as the optimum of model (P). We also see if we can refine the solution by generating the scenarios using model (SG-L). We find that we obtain the same set of 4 scenarios, as shown in Table 4, by solving (SG-L) corresponding to this example.

Example 3. The third example is a larger case study and is taken from Novak and Kravanja (1999) with some modifications. This problem corresponds to the design of a heat exchanger network with 5 heat exchangers, 2 hot streams, 2 cold streams, and 2 utilities. The network structure is given in Fig. 3.

The three temperatures  $T_3, T_5$  and  $T_9$  are uncertain parameters that change during network operation. Each of these uncertain parameters is assumed to take on 5 values with probabilities ( $p$ ) given in Table 5.

The optimization problem is formulated as a two-stage stochastic program, which is converted to its multiscenario equivalent. There are a total of 125 scenarios in this problem. The goal of the design problem is to minimize the expected total cost that includes the capital cost of the heat exchangers and the expected utility cost. The heat exchanger areas are the 1st stage design variables, while the heat loads and the temperatures that are not fixed are the 2nd stage variables. The multiscenario model (E3) is as

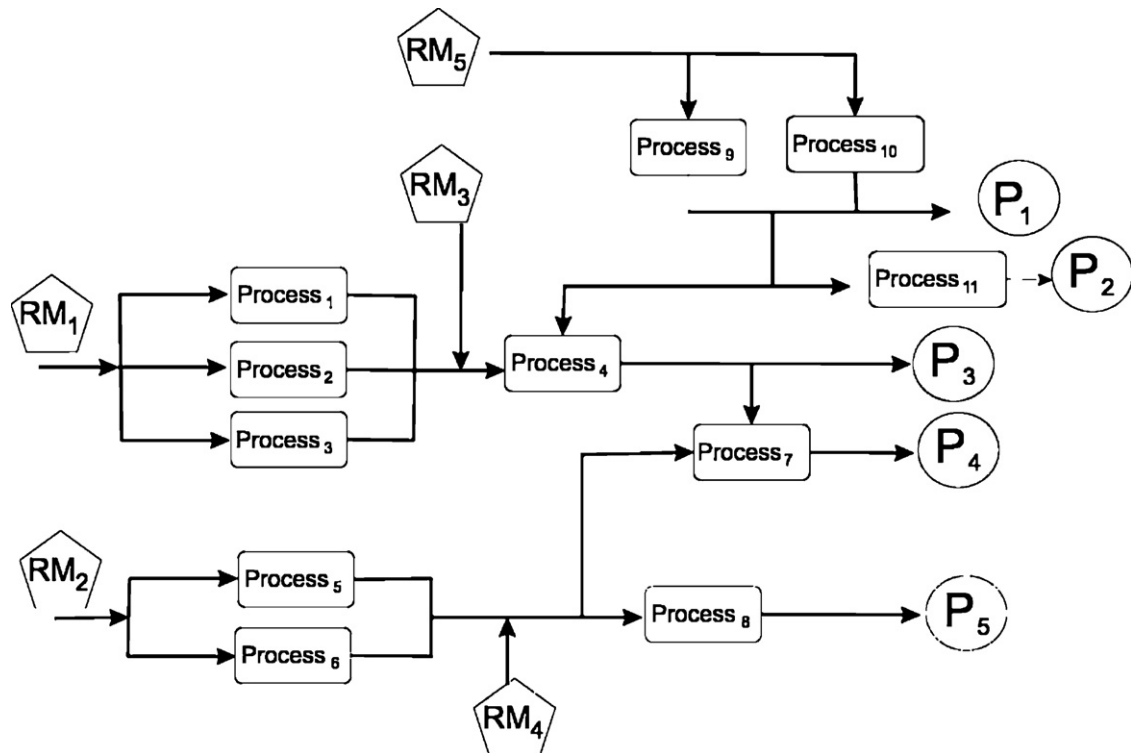


Fig. 4. Process network for example 4.

follows:

$$\min z = \sum_{s=1}^{125} \left[ p_s \left( 1846 \sum_{r=1}^4 A_r^{0.65} + 2350 A_5^{0.65} + 0.02 \phi_{4,s} + 0.23 \phi_{5,s} \right) \right]$$

s.t.

$$\left. \begin{aligned} \phi_{1,s} &= 1500(620 - T_{2,s}) = 2000(T_{4,s} - T_{3,s}) \\ \phi_{2,s} &= 1000(T_{5,s} - T_{6,s}) = 2000(T_{8,s} - T_{4,s}) \\ \phi_{3,s} &= 1000(T_{6,s} - T_{7,s}) = 3000(393 - 313) \\ \phi_{4,s} &= 1500(T_{2,s} - 350) \\ T_{10,s} &= \frac{\phi_{4,s}}{6000} + T_{9,s} \\ \phi_{5,s} &= 2000(600 - T_{8,s}) \\ T_{6,s} &\geq T_{4,s} + 1 \\ A_r &\geq \frac{\phi_{r,s}}{U_r \Delta \ln T_{r,s}}, \quad r = 1, \dots, 5 \\ \Delta \ln T_{1,s} \cdot \ln \left( \frac{620 - T_{4,s}}{T_{2,s} - T_{3,s}} \right) &= (620 - T_{4,s}) - (T_{2,s} - T_{3,s}) \\ \Delta \ln T_{2,s} \cdot \ln \left( \frac{T_{5,s} - T_{8,s}}{T_{6,s} - T_{4,s}} \right) &= (T_{5,s} - T_{8,s}) - (T_{6,s} - T_{4,s}) \\ \Delta \ln T_{3,s} \cdot \ln \left( \frac{T_{6,s} - 393}{T_{7,s} - 313} \right) &= (T_{6,s} - 393) - (T_{7,s} - 313) \\ \Delta \ln T_{4,s} \cdot \ln \left( \frac{T_{2,s} - T_{10,s}}{350 - T_{9,s}} \right) &= (T_{2,s} - T_{10,s}) - (350 - T_{9,s}) \\ \Delta \ln T_{5,s} \cdot \ln \left( \frac{620 - 600}{619 - T_{8,s}} \right) &= (620 - 600) - (619 - T_{8,s}) \end{aligned} \right\} s = 1, \dots, 125 \tag{E3}$$

$$\begin{aligned} T_{3,s} + 1 &\leq T_{2,s} \leq 620 \\ T_{3,s} &\leq T_{4,s} \leq 620 \\ \max(394, T_{3,s} + 1) &\leq T_{6,s} \leq T_{5,s} \\ 314 &\leq T_{7,s} \leq 323 \\ T_{3,s} &\leq T_{8,s} \leq T_{5,s} - 1 \\ T_{9,s} &\leq T_{10,s} \leq 323 \\ T_{1,s} &\geq 0 \\ \phi_{r,s} &\geq 0, \quad \forall r \\ U_1 = U_2 = U_3 = U_4 &= 700 \text{ W/(m}^2 \text{ K)}, U_5 = 1000 \text{ W/(m}^2 \text{ K)} \\ A_r &\geq 0, \quad \forall r \end{aligned}$$

In the above model, the subscript *s* corresponds to a particular scenario.  $\phi_{r,s}$  pertains to the heat load in heat exchanger *r* in scenario *s*. The 1st stage design variable  $A_r$  pertains to the area of heat exchanger *r*.  $T_{3,s}, T_{5,s}, T_{9,s}$  are the values of the respective uncertain parameters,  $T_3, T_5$  and  $T_9$ , in scenario *s*. The coefficient  $p_s$  denotes the probability of occurrence of scenario *s*, and is calculated by multiplying the individual probabilities of the values of the uncertain parameters which occur in that scenario *s*. The model (E3) with 125 scenarios is a nonconvex nonlinear program with 2005 continuous variables and 2375 constraints. On solving this model, we obtain the optimal solution of \$45,223.07 with the following optimal values of the design variables:

$$A_1 = 15.34 \text{ m}^2, A_2 = 2.37 \text{ m}^2, A_3 = 6.32 \text{ m}^2, \\ A_4 = 1.99 \text{ m}^2, A_5 = 2.31 \text{ m}^2.$$

Applying our scenario reduction approach to this example, where we first solve the MILP model (SG) for this problem, we obtain 5 scenarios (see Table 6).

**Table 6**  
Reduced number of scenarios for example 3.

Scenario, <i>s'</i>	$T_{3,s'}$ (°C)	$T_{5,s'}$ (°C)	$T_{9,s'}$ (°C)	$p_{s'}$
1'	378.9	573.9	293.5	0.007
2'	382.6	577.6	295.3	0.1545
3'	388.0	583.0	298.0	0.677
4'	393.4	588.4	300.7	0.1545
5'	397.1	592.1	302.5	0.007

On using the above 5 scenarios in problem (E3), reformulating it and solving it, we obtain a nonconvex NLP model with 86 variables and 96 constraints. The optimal objective of this reduced problem is \$45,310.08, which is 0.2% higher than the optimum of the original multisenario problem with 125 scenarios. The optimal values of the design variables so obtained are  $\hat{A}_1 = 15.28 \text{ m}^2, \hat{A}_2 = 2.43 \text{ m}^2, \hat{A}_3 = 6.38 \text{ m}^2, \hat{A}_4 = 1.98 \text{ m}^2, \hat{A}_5 = 2.312 \text{ m}^2$ . On solving the 125 scenario model (E3) by fixing the design variables to the optimal values obtained by solving the reduced model, we obtain an expected cost of  $\bar{z}^* = \$45,228.05$  which is very close to the exact solution of the original stochastic program. In terms of the computational times for solving the optimization problems, it takes 4.6 CPUs to solve (E3), while the reduced dimensional problem with 5 scenarios is solved in 0.33 CPUs.

**Example 4.** The last example is a modified version of the one used by Acevedo and Pistikopoulos (1998) and Wei and Realf (2004). The original problem involves the production of 5 products from 5 raw materials using 11 different processes (Fig. 4). In this problem, the uncertain parameters are the maximum availabilities of raw materials, and the demands for products. The continuous decision variables are the capacities for the processes, whereas the binary variables denote the selection of the required processes.

The deterministic model for this example is based on basic mass balances and is described as follows.

In Fig. 4, the nodes are either splitters or mixers.  $F(\text{unit}, \text{unit}_1)$  is the mass flow rate from a source 'unit' to a destination 'unit<sub>1</sub>'. For a splitter 'split' connected to a source 'unit' and destinations 'unit<sub>q</sub>', the mass balance is given by,

$$F(\text{unit}, \text{split}) = \sum_q F(\text{split}, \text{unit}_q)$$

For a mixer 'mix', with input connections from 'unit<sub>q</sub>' and an output to 'unit', the mass balances are,

$$\sum_q F(\text{unit}_q, \text{mix}) = F(\text{mix}, \text{unit})$$

A raw material *j* with flowrate 'RM<sub>*j*</sub>' is assumed to come from an inlet 'source<sub>*j*</sub>',

$$F(\text{source}_j, \text{unit}) = \text{RM}_j, \quad \forall j$$

A product *i* with mass flow rate 'P<sub>*i*</sub>' is assumed to go out to a destination 'out<sub>*i*</sub>',

$$F(\text{unit}, \text{out}_i) = P_i, \quad \forall i$$

The sum of the mass flows to a process *k* from inlet sources 'unit<sub>q</sub>' is equal to 'IS<sub>*k*</sub>',

$$\sum_q F(\text{unit}_q, \text{process}_k) = \text{IS}_k, \quad \forall k$$

The mass flow from a process *k* to a destination 'unit' is equal to 'OS<sub>*k*</sub>',

$$F(\text{process}_k, \text{unit}) = \text{OS}_k, \quad \forall k$$

Other balances include:

Yield relations	$OS_k = PC_k IS_k, \forall k$
Desired production	$P_i \leq D_i, \forall i$
Availability of raw material	$RM_j \leq \max \text{RM}_j, \forall j$
Logic constraints	$IS_k - MI_k Q_k \leq 0, \forall k$ $Q_k - \max Q_k y_k \leq 0, \forall k$

**Table 7**  
Parameters used in the model for example 4.

Process <i>k</i>	1	2	3	4	5	6	7	8	9	10	11	
$PC_k$	13	15	17	14	10	15	16	11	13	15	17	
$MI_k$	18	20	15	20	20	21	15	15	25	15	20	
$OC_k$	400	400	400	400	400	400	400	400	400	400	400	
$DC_k$	2500	2500	2500	2500	2500	2500	2500	2500	2500	2500	2500	
$FC_k$	400	2500	3500	300	4500	2500	300	2200	2800	2700	2500	
$\max Q_k$	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	
Product <i>i</i>	1			2			3			4		5
$D_i$	28	32	27	31	29	32	28	31	30			
$p(D_i)$	0.3	0.7	0.35	0.65	0.65	0.35	0.5	0.5	1			
Raw material <i>j</i>	1			2			3			4		6
$\max RM_j$	33	36	34	37	32	35	35	36	35			
$p(\max RM_j)$	0.4	0.6	0.45	0.55	0.55	0.45	0.5	0.5	1			

**Table 8**  
Design variable values obtained from solution of full multiscenario problem.

<i>k</i>	4	7	8	10	11
$Q_k$	0.121	0.129	0.182	0.142	0.091

The objective function is given by,

$$\min z = - \left\{ \sum_{i=1}^5 \beta_i P_i - \sum_{j=1}^5 \alpha_j RM_j - \sum_{k=1}^{11} OC_k IS_k - \sum_{k=1}^{11} [DC_k Q_k + FC_k y_k] \right\}$$

The symbols in the previous equations are summarized as follows:

- $D_i$  is the uncertain demand for product *i* (parameter)
- $DC_k$  is cost for process *k* (parameter)
- $F(\text{unit}, \text{unit}_1)$  is the mass flow rate in the stream between *unit* and *unit*<sub>1</sub> (variable)
- $FC_k$  is the fixed cost of process *k* (parameter)
- $IS_k$  is the mass flow in input stream to process *k* (variable)
- $\max Q_k$  is the maximum volume capacity of process *k* (parameter)
- $\max RM_j$  is the uncertain maximum availability of raw material *j* (parameter)
- $MI_k$  is the mass flow to volume relationship constant for process *k* (parameter)
- $OC_k$  is the operating cost of process *k* (parameter)

**Table 9**  
Reduced set of scenarios for example 4.

Scenario, <i>s'</i>	$\max RM_1$	$\max RM_2$	$\max RM_3$	$\max RM_4$	$D_1$	$D_2$	$D_3$	$D_4$	$p_{s'}$
1	33	34	35	35	28	27	32	28	0.3
2	33	34	35	35	32	31	29	28	0.1
3	36	34	35	35	32	31	29	28	0.05
4	36	37	32	35	32	27	32	28	0.05
5	36	37	32	36	32	31	29	31	0.5

**Table 10**  
Uncertain parameter data for extended example 4.

Product <i>i</i>	1			2			3			4		5
$D_i$	28	30	32	27	29	31	29	31	32	28	30	30
$p(D_i)$	0.2	0.55	0.25	0.3	0.4	0.3	0.3	0.45	0.25	0.15	0.65	0.2
Raw material <i>j</i>	1			2			3			4		5
$\max RM_j$	33	35	36	34	36	37	32	34	35	35	35.5	36
$p(\max RM_j)$	0.1	0.5	0.4	0.3	0.4	0.3	0.2	0.35	0.45	0.25	0.6	0.15

**Table 11**  
Non-zero design variable values for extended example 4.

<i>k</i>	4	7	8	10	11
$Q_k$	0.121	0.129	0.182	0.141	0.091
$y_k$	1	1	1	1	1

- $OS_k$  is the mass flow in output stream to process *k* (variable)
- $P_i$  is the mass flow of product *i* (variable)
- $PC_k$  is the yield constant for process *k* (parameter)
- $Q_k$  is the capacity of process *k* (variable)
- $RM_j$  is the mass flow of raw material *j* (variable)
- $y_k$  is the binary variable for selection of process *k* (binary variable)
- $\alpha_j$  is the cost of raw material *j* (parameter)
- $\beta_i$  is the price of product *i* (parameter)

Since  $D_i$  and  $\max RM_j$  are uncertain parameters, the above model is converted into a two-stage stochastic program, which is then reformulated as a deterministic multiscenario model by discretizing the uncertain parameters.  $Q_k$  and  $y_k$  are the first stage decision variables, while the flows in the system, raw material consumptions, and product flows are the second stage variables. The objective is to minimize the negative of the profit function,

In the multiscenario formulation, the uncertain parameters are,  $D_i, i = 1, \dots, 4$ , and  $\max RM_j, j = 1, \dots, 4$ , and each of these is assumed to take two values.  $D_5$  and  $\max RM_5$  are assumed to be known and constant. The values for all the parameters used in this example 4 can be seen in Table 7 where the two levels of the eight uncertain parameters and their probabilities can also be found.



**Table 12**  
Reduced set of scenarios for extended example 4.

Scenario, $s'$	max $RM_1$	max $RM_2$	max $RM_3$	max $RM_4$	$D_1$	$D_2$	$D_3$	$D_4$	$p_s$
1	33	34	32	36	28	27	32	28	0.075
2	33	37	32	36	28	31	32	28	0.025
3	35	36	35	35.5	30	29	31	30	0.4
4	35	37	34	35.5	30	27	29	30	0.05
5	35	37	35	35.5	30	27	31	30	0.05
6	36	34	32	35	28	31	32	31	0.05
7	36	34	34	35	32	31	29	31	0.1
8	36	34	34	35.5	30	31	29	30	0.025
9	36	34	34	35.5	32	31	29	30	0.05
10	36	37	32	36	28	31	32	28	0.05
11	36	37	34	35	32	27	29	30	0.05
12	36	37	34	35	32	27	32	31	0.05
13	36	37	34	35.5	30	27	29	30	0.025

In this example we obtain an exact solution considering all the scenarios, and compare that optimal value of the objective function with those obtained using the proposed approach and by using Sample Average Approximation (SAA).

The multiscenario problem with 256 scenarios consists of 11 binary variables, 17,165 continuous variables, and 19,212 constraints. By solving the full multiscenario problem, we find that only processes 4, 7, 8, 10 and 11 are in operation ( $y_k = 1$ ;  $k = 4, 7, 8, 10, 11$ ) and the optimal objective function value is  $-63,677.5$ . This MILP model solves in only 2.04 CPUs. Table 8 shows the values for the design capacity variable  $Q_k$ .

Using the proposed method in the paper, a reduced set of scenarios along with their probabilities is obtained by solving model (SG) corresponding to this example and the results are shown in Table 9. The reduced scenario problem has only 5 scenarios and 11 binary variables, 348 continuous variables, and 387 constraints.

On using the reduced set of scenarios, reformulating and solving the problem, the values for  $Q_k$  (except  $Q_{10}$ ;  $Q_{10} = 0.141$ ) and  $y_k$  are the same as in Table 8, and the optimal objective value is found to be  $-63,754.5$ . The solution time for the reduced scenario model is 0.22 CPU s. If we fix the values of  $Q_k$  and  $y_k$  in the full multiscenario model to those obtained by solving the reduced scenario problem, we obtain an objective function value of  $-63,643.8$  which is within 0.05% of actual optimal objective function value. We should note that though the solution times for the original and the reduced scenario problems are very small, they show the potential of the proposed approach to reduce computational times for much larger problems.

Finally, we find the solution provided by the SAA method and an interval for the solution with a confidence limit of 95% (see Wei & Realff, 2004). A statistical lower limit on this interval ( $-63,715.4$ ) is found by solving the stochastic program 10 times each with 10 randomly selected samples (scenarios). The statistical upper limit ( $-62,947.8$ ) is found by formulating a multiscenario problem with 50 scenarios randomly selected from the given set of 256 scenarios, and solving it with fixed values of the first stage decision variables obtained during calculation of the lower statistical limit on the confidence interval for the solution. Since we sampled scenarios from a finite population, adjustments were made in the calculation of the statistical limits. The optimal objective value using the proposed approach also lies between the statistical limits computed by the SAA method.

We extend example 4 by increasing the number of values taken on by each uncertain parameters from two to three. The new uncertain parameter realizations and associated probabilities are given in Table 10.

The extended example has 6561 scenarios, which leads to a multiscenario MILP with 11 binary variables, 439,600 continuous variables, and 492,087 constraints. This model solves in 472 CPUs

and yields an objective function value of  $-63,799.2$ . The non-zero design variable values are shown in Table 11.

In order to obtain a reduced set of scenarios (see Table 12), we solve model (SG-L) for the extension of example 4.

The reduced scenario model for the extended example 4 is considerably smaller than the original formulation as it has only 13 scenarios, yielding an MILP with 11 binary variables, 884 continuous variables, and 987 constraints. The reduced model optimal objective value is  $-63,856.2$  and solves in just 0.24 CPUs. The design variable values obtained by solving the reduced scenario model are the same as those obtained by solving the full multiscenario model corresponding to the extended example 4 (except for a small difference with  $Q_{10} = 0.142$  in the solution of the reduced scenario model). We then fix the values of the design variables  $Q_k$  and  $y_k$  in the full multiscenario model corresponding to the extended example 4 to those obtained by solving the reduced scenario model. On solving the large model we find that it solves in 20.6 CPUs and yields an optimal objective value of  $-63,784.1$ , which is very close to the actual objective function value of  $-63,799.4$ .

## 5. Conclusions

This work has presented a new practical heuristic strategy for solving two-stage stochastic programming problems formulated as deterministic multiscenario optimization problems. The idea consists of replacing a given set of scenarios, obtained by discretization of the uncertain parameter space, by a smaller set of scenarios and thus approximating the optimization problem in a reduced space. The proposed criterion for selecting a subset of given set of scenarios is that the overall probability of occurrence of a particular realization of any uncertain parameter in the final set of scenarios should be equal to the probability of the uncertain parameter taking on that particular value. This criterion has to be satisfied for each uncertain parameter in the model. We presented a MILP formulation, as well as a relaxed LP model for determining a minimum subset of scenarios from a given scenario set such that this criterion is satisfied. The stochastic programs were reformulated with the smaller set of scenarios in order to obtain approximate models. The application of this heuristic technique on numerical examples has shown that we obtain close to optimal solutions using the approximate model with the smaller number of scenarios. It is also our observation that when the objective value of the reduced function matches closely with that of the original problem, the design variable values match very closely as well. This method would also complement other sampling based optimization methods as this heuristic can be applied to the samples collected from an infinite space to further simplify the problem.

## Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation under Grant CTS-0521769. We are also grateful to Dr. Kevin Furman at ExxonMobil Research and Engineering for his helpful comments and suggestions. Mariano Martín also wishes to acknowledge a MICINN–Fulbright postdoctoral fellowship.

## References

- Acevedo, J., & Pistikopoulos, E. N. (1998). Stochastic optimization based algorithms for process synthesis under uncertainty. *Computers and Chemical Engineering*, 22, 647–671.
- Ahmed, S., Tawarmalani, M., & Sahinidis, N. (2004). A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming*, 100, 355–377.
- Balasubramanian, J., & Grossmann, I. E. (2002). A novel branch and bound algorithm for scheduling flowshop plants with uncertain processing times. *Computers and Chemical Engineering*, 26, 41–57.
- Barro, D., & Canestrelli, E. (2005). Dynamic portfolio optimization: Time decomposition using the maximum principle with a scenario approach. *European Journal of Operational Research*, 163, 217–229.
- Birge, J. R., & Dempster, M. A. H. (1996). Stochastic programming approaches to stochastic scheduling. *Journal of Global Optimization*, 9, 417–451.
- Birge, J. R., & Louveaux, F. V. (1997). *Introduction to stochastic programming*. New York: Springer.
- Brooke, A., Kendrick, D., Meeraus, A., & Raman, R. (1998). *GAMS: A user's guide, release 2.50*. GAMS Development Corporation.
- Carøe, C. C., & Schultz, R. (1999). Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24, 37–45.
- Carøe, C. C., & Tind, J. (1997). A cutting-plane approach to mixed 0-1 stochastic integer programs. *European Journal of Operational Research*, 101, 306–316.
- Cheng, L., Subrahmanian, E., & Westerberg, A. W. (2003). Design and planning under uncertainty: Issues on problem formulation and solution. *Computers and Chemical Engineering*, 27, 781–801.
- Clay, R., & Grossmann, I. E. (1997). A disaggregation algorithm for the optimization of stochastic planning models. *Computers and Chemical Engineering*, 21, 751–774.
- Dantzig, G. B. (1963). *Linear programming and extensions*. Princeton, NJ: Princeton University Press.
- Dupačová, J., Gröwe-Kuska, N., & Römisch, W. (2003). Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming*, 95, 493–511.
- Grossmann, I. E., Halemane, K. P., & Swaney, R. E. (1983). Optimization strategies for flexible chemical processes. *Computers and Chemical Engineering*, 7, 439–462.
- Henrion, R., & Römisch, W. (1999). Metric regularity and quantitative stability in stochastic programs with probabilities constraints. *Mathematical Programming*, 84, 55–88.
- Heitsch, H., & Römisch, W. (2003). Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, 24, 187–206.
- Hoyland, K., Kaut, M., & Wallace, S. W. (2003). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24, 169–185.
- Klein Haneveld, W. K., Stougie, L., & van der Vlerk, M. H. (1995). On the convex hull of the simple integer recourse objective function. *Annals of Operations Research*, 56, 209–224.
- Klein Haneveld, W. K., Stougie, L., & van der Vlerk, M. H. (1996). An algorithm for the construction of convex hulls in simple integer recourse programming. *Annals of Operations Research*, 64, 67–81.
- Liu, M. L., & Sahinidis, N. V. (1996). Optimization in process planning under uncertainty. *Industrial and Engineering Chemistry Research*, 35, 4154–4165.
- Luceno, A. (1999). Discrete approximations to continuous univariate distributions—An alternative simulation. *Journal of the Royal Statistical Society (Series B)*, 61, 345–352.
- Norkin, V. I., Pflug, G. Ch., & Ruszczyński, A. (1998). A branch and bound method for stochastic global optimization. *Mathematical Programming*, 83, 425–450.
- Nowak, M. P., Schultz, R., & Westphalen, M. (2005). A stochastic integer programming model for incorporating day-ahead trading of electricity into hydro-thermal unit commitment. *Optimization and Engineering*, 6, 163–176.
- Novak, Z., & Kravanja, Z. (1999). Mixed-integer nonlinear programming problem process synthesis under uncertainty by reduced dimensional stochastic optimization. *Industrial and Engineering Chemistry Research*, 38, 2690–2698.
- Rachev, S. T., & Römisch, W. (2002). Quantitative stability in stochastic programming. The methods of probability metrics. *Mathematics of Operations Research*, 27, 792–818.
- Rooney, W. C., & Biegler, L. T. (2003). Optimal process design with model parameter uncertainty and process variability. *AIChE Journal*, 49, 438–449.
- Sahinidis, N. V. (2004). Optimization under uncertainty: State-of-the-art and opportunities. *Computers and Chemical Engineering*, 28, 971–983.
- Shapiro, A. (2003). Monte Carlo Sampling Methods, Chapter 6. In A. Ruszczyński, & A. Shapiro (Eds.), *Handbooks in OR & MS* (pp. 353–425). Elsevier Science.
- Takriti, S., Birge, J. R., & Long, E. (1996). A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11, 1497–1506.
- Wei, J., & Realf, M. J. (2004). Sample average approximation methods for stochastic MINLPs. *Computers and Chemical Engineering*, 28, 333–346.