RC 20993 (09/05/1997)
Subject Area Computer Science/Mathematics

IBM Research Report

Issues in Risk Modeling for Multi-Stage Systems

Alan J. King, Samer Takriti

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

Shabbir Ahmed

Department of Mechanical & Industrial Engineering
University of Illinois at Urbana-Champaign,
120 W.Green Street,
Urbana, Illinois 61801

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

Issues in Risk Modeling for Multi-Stage Systems

Alan J. King * Samer Takriti† Shabbir Ahmed‡

September 30, 1997

Abstract

This paper studies the robust formulation of multi-stage stochastic models. We show that modeling paradigms based on Markowitz's single-stage model lead to non-optimal second-stage decisions. We present a non-convex formulation that enforces optimality. Returning to first principles, we explore a class of two-moment approximations to utility functions for two-stage systems. The resulting mathematical models are convex and can be extended easily to the multi-stage case.

Key Words. Stochastic Programming, Robust Optimization, Decision Making Under Uncertainty, Bilevel Programming

AMS Subject Classifications. 90C15, 90C33, 90B50, 90A09, 90A43

1 Introduction

Applications requiring decision making under uncertainty are often modeled as two-stage stochastic programs. The objective of such programs is to minimize the cost of

*Mathematical Sciences Department, T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598, USA, kingaj@watson.ibm.com

†Mathematical Sciences Department, T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598, USA, takriti@watson.ibm.com

‡Department of Mechanical & Industrial Engineering, University of Illinois at Urbana-Champaign, 120 W.Green Street, Urbana, Illinois 61801, USA, s-ahmed1@uiuc.edu

the first-stage decision and the expected value of the second-stage recourse function. The classical two-stage stochastic program [1, 6] with fixed recourse is

$$\begin{array}{ll} \min_x & c^T x + E_{\tilde{\xi}}\{Q(x,\tilde{\xi})\} \\ \text{s.t.} & Ax = b,\ x \geq 0, \end{array}$$

where

$$Q(x,\xi) = \min\{q^T y | Wy = h(\xi) - T(\xi)x,\ y \geq 0\}.$$

The vector x is the first-stage decision vector with an associated "cost" of c and a feasible set of $\{Ax = b,\ x \geq 0\}$. The "cost" in the second stage, $Q(x,\tilde{\xi})$, is usually called the recourse function. It is a function of x and of the random vector $\tilde{\xi}$ that can vary over a set Ξ with a given probability distribution. In the above formulation, the variable x must be decided before the actual realization of the uncertain parameter, $\tilde{\xi}$. Subsequently, once the decision x has been decided and the random events presented themselves, the variable y is determined as a solution to an optimization problem. For example, assume that we are interested in deciding on the design, x, of a production facility. The design must take into consideration the cost of operating the facility which in turn is a function of x and of random future demands, $h(\tilde{\xi})$. Given a probabilistic distribution for the demand and the physical constraints that govern the production process, we minimize the cost of building the facility, $c^T x$, in addition to the expected cost of operating it, $E_{\tilde{\xi}}\{Q(x,\tilde{\xi})\}$. As a result, we obtain a set of decisions, $y(\xi)$, for operating the facility under different demand values. Later on, the value of the demand becomes known and a corresponding optimal decision, y, is adopted.

If the distribution of the random variable is discrete with a finite support, one can write the two-stage stochastic program in the form

$$\begin{array}{ll} \min_x & c^T x + \sum_{k=1}^K p_k Q_k(x) \\ \text{s.t.} & Ax = b,\ x \geq 0, \end{array} \tag{1}$$

where

$$Q_k(x) = \begin{array}{ll} \min_y & q^T y \\ \text{s.t.} & Dy = h^k - T^k x, \\ & y \geq 0. \end{array}$$

In (1), we assume that there are K possible realizations for the random variable with a probability p_k associated with each one of them. For realization k, the right-hand side of the recourse problem is h^k and the contribution of the first-stage decision is $-T^k x$. The mathematical program (1) can be reformulated as

$$\begin{array}{llrl} \min_x & c^T x & + \sum_{k=1}^K p_k q^T y^k & \\ \text{s.t.} & Ax & & = b, \\ & T^k x & +Dy^k & = h^k, \\ & x \geq 0, & y^k \geq 0. & \end{array} \tag{2}$$

Note that the set of optimal solutions for (1) is the same as that of (2) due to the separability of the second-stage variables, $y^k,\ k = 1,\ldots,K$. As a result, the linear program (2) is called the deterministic equivalent [1, 6].

A limitation of the linear model described above is that it only accounts for the expected values of the second-stage recourse. That is, the variance of the second-stage costs cannot play a role in the decision making process.

The importance of controlling variance as opposed to just optimizing first moments is well recognized in portfolio management applications. In his landmark work, Markowitz [12] uses the expected returns of various securities under consideration and the variance in the return of each security to present an investor with an efficient frontier. To do so, Markowitz assumes that the trade off between an investor's tolerance for risk and expected returns can be modeled using a combination of mean and variance.

In an attempt to generalize the work of [12] to two-stage systems, Mulvey, Vanderbei, and Zenios [13] propose adding a penalty on the variation in the second-stage objective functions of (2). For example, one can use the statistical variance in the second-stage cost functions,

$$\lambda \sum_{k=1}^{K} p_k \left[q^T y^k - \sum_{i=1}^{K} p_i q^T y^i \right]^2, \lambda \geq 0,$$

as a measure of variability and incorporate it into the objective function of (2). The non-negative scalar, λ, models the decision maker's attitude towards risk: large values of λ represent high risk-aversion and small values of λ represent low risk-aversion. Just as in the Markowitz model, Malcolm and Zenios [11] use a similar approach to model a capacity expansion problem that arises in the electric power industry. The goal is to choose the design capacity over a given set of plants—first-stage decision—so that the uncertain future demand for electric power is met at a minimal cost. Given the large amount of capital involved in building a power plant, it is necessary to make a robust decision that minimizes cost without exposing the decision maker to large losses under certain future scenarios.

As with the previous examples, most studies incorporate risk into a two-stage system by following Markowitz's single-stage model; i.e., by adding a penalty term to the objective function. However, by doing so, the vectors $y^k, k = 1, \ldots, K$, may lack optimality for the second-stage optimization problems. This is a consequence of linking the second-stage decisions with each other. In this paper, we first extend the formulation of (1) to include measures of variability of the recourse costs, while maintaining the optimality of the second-stage decisions. The resulting model is a bilevel mathematical program that is difficult to solve if the statistical variance is used as a measure of variability. We then show that using the variation above-target as a measure of risk results in a convex model that is easier to solve.

The rest of this paper is organized as follows. In Section 2, we show that the second-stage decisions of robust optimization may not be optimal and suggest a mathematical formulation that enforces optimality. Section 3 presents an example in which our formulation is applied to a capacity expansion problem. In Section 4, we discuss the difficulties in extending the single-stage model to the two-stage case and suggest maintaining convexity by using an above-target variance as a measure of risk.

2 Enforcing Second-Stage Optimality in Two-Stage Models

Consider the robust optimization model suggested by [11] and [13]:

$$\begin{array}{llll} \min_{x,y} & c^T x & + \sum_{k=1}^{K} q^T y^k & + \lambda \nu(q^T y^1, \ldots, q^T y^K) \\ \text{s.t.} & Ax & & = b, \\ & T^k x & + D y^k & = h^k, \\ & x \geq 0, & y^k \geq 0, & \end{array} \tag{3}$$

where

$$\nu(q^T y^1, \ldots, q^T y^K) = \sum_{k=1}^{K} p_k \left[q^T y^k - \sum_{i=1}^{K} p_i q^T y^i \right]^2 .$$

The parameter λ controls the tradeoff between the mean and variability of the second-stage costs. Note that (3) does not require the second-stage decisions to be optimal for $\{\min q^T y^k | D y^k = h^k - T^k x,\ y \geq 0\}$. As a result, the policies, y^k, obtained from solving (3) are not useful to the decision maker in the second stage. In the case of [11] for example, the amounts of electric power generated under the different realizations of demand, $h^k,\ k = 1, \ldots, K$, need not be optimal and the standard deviation of the real return may be far from that suggested by the solution of (3).

To further clarify this point, consider the following two-stage stochastic program:

$$\begin{array}{lllll} \min & 2x & +0.5y_1 & +0.5y_2 & \\ \text{s.t.} & x & +y_1 & & \geq 3, \\ & x & & +y_2 & \geq 2, \\ & x \geq 0, & y_1 \geq 0, & y_2 \geq 0. & \end{array} \tag{4}$$

The previous program has two scenarios each of which has a probability of 0.5. The objective function coefficient of the first-stage decision, $x \in \Re$, is $c = 2$. For the second-stage variables, the objective function coefficient is $q = 1$. The optimal solution is $x = 0$, $y_1 = 3$, $y_2 = 2$ with an objective function value of 2.5. To use the model of (3), we add the term $\lambda[0.5(y_1 - (0.5y_1 + 0.5y_2))^2 + 0.5(y_2 - (0.5y_1 + 0.5y_2))^2]$, $\lambda \geq 0$. If λ is large enough, the problem becomes that of minimizing the variance of the second-stage cost. The optimal solution in this case is $x = 0$, $y_1 = 3$, and $y_2 = 3$. That is, the decision maker tends to believe that the variance in his/her second-stage cost is 0. However, after observing the value of the random right-hand side, there is no point in not adopting an optimal second-stage policy. As a result, the actual variance in the second-stage cost corresponding to $x = 0$ is 0.5.

Conceptually, the loss of optimality in (3) is a result of not using a well-defined dis-utility function in the formulation. Mathematically, the lack of optimality is a result of coupling the second-stage variables in (3). Of course, one can incorporate risk into (1) and force optimality by considering the following formulation

$$\left\{ \min_x c^T x + \sum_{k=1}^{K} p_k Q_k(x) + \lambda \nu \left(Q_1(x), \ldots, Q_K(x) \right) | Ax = b,\ x \geq 0, \right\} \tag{5}$$

where $Q_k(x)$ is the optimal second-stage cost under scenario k. The formulation (5) captures the two-stage nature of the decision making process because we determine the second-stage decisions by solving an optimization problem. Note that if we apply (5) to the numerical example (4), then the optimal solution is $x = 3$, $y_1 = 0$, and $y_2 = 0$ with an objective function value of 6 and a variance of zero.

One can construct the deterministic equivalent of (5) by incorporating the second-stage optimality conditions. The result is a quadratic program with equilibrium constraints [14]

$$\begin{array}{lllllll} \min_{x,y} & c^T x & +\sum_{k=1}^K q^T y^k & +\lambda\nu(q^T y^1, \ldots, q^T y^K) & & \\ \text{s.t.} & Ax & & & & = b, \\ & T^k x & +Dy^k & & & = h^k, \\ & & & & D^T\pi^k & \leq q, \\ & & (q - D^T\pi^k)^T y^k & & & = 0, \\ & x \geq 0, & y^k \geq 0, & & & \end{array} \tag{6}$$

where π^k is the dual vector associated with the second-stage constraints of problem k. In (6), the constraints $D^T\pi^k \leq q$ and $(q - D^T\pi^k)^T y^k = 0$ enforce dual feasibility and complementarity slackness. It is clear that (3) is a relaxation of (6) in which the complementarity conditions are removed.

The formulation (5) is a bilevel program. In fact, our formulation is an instance of the Stackelberg game [15]. In this game, there is a leader—first stage—who makes a decision and followers—second stage—who make their decisions based on the leader's decision. The cost function of the leader depends on both his/her decision and on the followers' strategies. Each of the followers adopts the best strategy given the decision of the leader. It is well known that these models are not convex and that finding an optimal solution is difficult. One way to solve such problems is by reformulating them as mixed-integer programs. Let M be a large number and $(z^1, \ldots, z^K)$ be a set of binary vectors. The last constraint in (6) can be written as

$$q - D^T\pi^k \leq Mz^k \text{ and } y^k \leq M(e - z^k)$$

where e is a vector of ones. On the other hand, if one is interested in finding a stationary point for (6), then such a point can be reached by staring from the solution of (3) and improving it using an appropriate technique [5, 14, 16].

Although we limited our previous discussion to two-stage stochastic programs with recourse, our idea of incorporating the second-stage optimality conditions explicitly in the model is general. That is, we can handle multi-stage stochastic programs, stochastic programs with random recourse, and non-linear stochastic programs. Note that (6) allows penalizing the second-stage optimal decisions directly. In other words, one can apply penalties of the form $\lambda\nu(f_1(y^1), \ldots, f_k(y^K))$ where f_k is the contribution of y^k to the variability measure. In addition, our formulation allows adding restrictive constraints on the variability of the second-stage decisions as suggested by [17] without sacrificing optimality. Of course, as the model becomes more sophisticated, the resulting mathematical program (with equilibrium constraints) becomes harder to solve.

3 A Numerical Example

We adopt the capacity expansion problem from Malcolm and Zenios [11] as our test case. Malcolm and Zenios allow violations of the problem constraints by including penalties on the infeasibilities. Since we are only interested in observing the relationship between the expected cost and its variance using (3) and (5), we do not consider such infeasibilities in our formulation. We use GAMS to model a problem with two plants and three scenarios then solve it using DICOPT [2].

To compare (3) and (5), we determine the optimal second-stage decisions, y_*^k, by solving $\{\min q^T y | Dy = h^k - T^k x,\ y \geq 0\}$, where x is the optimal solution of (3) or (5). The resulting y_*^k is used to compute the true expected costs and variances for both models. Table 1 presents the results of our test. The columns labeled μ and σ contain the expectations and the standard deviations of the second-stage costs. The second-stage decisions used in calculating the values of μ and σ are those obtained from solving (3) and (5). The columns labeled μ_* and σ_* are the true expectations and standard deviations of the second-stage costs. That is, μ_* and σ_* are the expectation and standard deviation when y_*^k is used in the calculations. The last column, σ_*^+, is the square root of the semi-variance. That is,

$$\sigma_*^{+^2} = \sum_{k=1}^{K} p_k [q^T y_*^k - \sum_{i=1}^{K} q^T y_*^i]_+^2.$$

The importance of the semi-variance will be clear in Section 4. It is included in Table 1 for the sake of completeness.

As expected, the solutions and the cost statistics for the two models are identical for $\lambda = 0$. Also, $\mu = \mu_*$ and $\sigma = \sigma_*$ because both formulations reduce to (2). As robustness is progressively enforced through increasing λ, the difference between the true expectation, μ_*, and the expectation of the second-stage cost of (3) increases. A similar remark holds when comparing σ_* and σ. In fact, although the first stage solution from (3) remains the same, the expected costs and the standard deviations are changing. This is a result of (3) selecting sub-optimal second-stage decisions to control the variance. On the other hand, $\mu = \mu_*$ and $\sigma = \sigma_*$ for (5) for any value of λ. Furthermore, the expectation and standard deviation are independent of λ if x is the same.

4 Extending Markowitz's Approach to Multi-Stage Models—The Mean/Target Frontier

In Section 2, we showed that it was necessary to force second-stage optimality in the robust optimization formulation of a two-stage stochastic program. The resulting problem (5) turns out to be non-convex which renders this formulation impractical. In this section, we go back to first principles and use utility functions to model

λ	Model	x	μ	σ	μ_*	σ_*	σ_*^+
0.0	(3)	(3.33,8.67)	7500.00	129.61	7500.00	129.61	77.45
	(5)	(3.33,8.67)	7500.00	129.61	7500.00	129.61	77.45
0.1	(3)	(3.17,8.83)	7587.00	3.54	7510.00	113.14	65.48
	(5)	(3.00,9.00)	7520.00	101.98	7520.00	101.98	62.18
0.2	(3)	(3.17,8.83)	7588.75	1.77	7510.00	113.14	65.48
	(5)	(3.00,9.00)	7510.00	101.98	7510.00	101.98	62.18
2.0	(3)	(3.17,8.83)	7589.87	0.017	7510.00	113.14	65.48
	(5)	(3.00,9.00)	7520.00	101.98	7520.00	101.98	62.18
5.0	(3)	(3.17,8.83)	7589.87	0.00	7510.00	113.14	65.48
	(5)	(0.42,11.58)	8083.94	40.83	8083.94	40.83	28.79

Table 1: Results for the power systems capacity expansion problem

risk aversion. The resulting model is a convex optimization problem whose solution $(x_*, y_*^1, \ldots, y_*^K)$ satisfies $q^T y_*^k = Q_k(x_*)$ for all k.

In the single-stage setting, the deterministic equivalent model with expected value objective is

$$\begin{array}{ll} \min & c^T x + \sum_{k=1}^K p_k r_k^T x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array}$$

The terms $r_1, \ldots, r_K$ represent random costs as a result of making decision x. The general robust formulation of the single-stage problem is

$$\begin{array}{ll} \min & \sum_{k=1}^K p_k \mathcal{U}(c^T x + r_k^T x) \\ \text{s.t.} & Ax = b, \\ & x \geq 0, \end{array} \tag{7}$$

where $\mathcal{U}(.)$ is a convex non-increasing dis-utility function chosen to represent risk preferences of the decision maker. Let

$$R_* = c^T x_* + \sum_{k=1}^K p_k r_k^T x_* \tag{8}$$

where x_* is an optimal solution to (7). The second-order approximation of $\mathcal{U}$ about R_* is

$$\mathcal{U}(v) = \mathcal{U}(R_*) + [v - R_*]\mathcal{U}'(R_*) + [v - R_*]^2 \mathcal{U}''(R_*)/2. \tag{9}$$

Now, we can use (9) in approximating (7) while enforcing (8) for all x. The resulting problem is

$$\begin{array}{ll} \min & \sum_{k=1}^K p_k [c^T x + r_k^T x - R_*]^2 \\ \text{s.t.} & c^T x + \sum_{k=1}^K p_k r_K^T x = R_*, \\ & Ax = b, \\ & x \geq 0. \end{array} \tag{10}$$

Note that R_* is arbitrary because $\mathcal{U}(.)$ is arbitrary. It follows that any dis-utility minimizing planner is willing to accept the approximation (9) and will choose the solution form (10) for some value of R_*. Introducing the parameter λ^{-1} as the dual multiplier to the constraint $c^T x + \sum_{k=1}^K p_k r_k^T x = R_*$ leads to

$$\begin{array}{ll} \min & c^T x + \sum_{k=1}^K p_k r_k^T x + \lambda \sum_{k=1}^K p_k [r_k^T x - R_*]^2 \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{array}$$

The previous formulation justifies the mean/variance formulation as the second order approximation to the utility formulation.

In the two-stage case, the analogous utility formulation is

$$\begin{array}{llll} \min & \sum_{k=1}^K p_k \mathcal{U}(c^T x & + q_k^T y^k) & \\ \text{s.t.} & Ax & & = b, \\ & T^k x & + D y^k & = h^k, \\ & x \geq 0, & y^k \geq 0. & \end{array}$$

Simply reproducing the above argument leads to the robust optimization model, which has serious problems as we have seen. The source of the difficulty is enforcing the constraint $c^T x + \sum_{k=1}^K p_k q^T y^k = R_*$. If we relax this constraint and use the semi-variance as a measure of risk, we obtain the following formulation

$$\begin{array}{lllll} \min & c^T x & + \sum_{k=1}^K p_k q_k^T y^k & + \lambda \sum_{k=1}^K p_k [q_k^T y^k - R_*]_+^2 \\ \text{s.t.} & Ax & & = b, \\ & T^k x & + D y^k & = h^k, \\ & x \geq 0, & y^k \geq 0, & \end{array} \tag{11}$$

which we call the mean/target formulation. Solutions to (11) have the desired property that the second-stage solutions are optimal. Furthermore, this formulation is a convex quadratic program and can be extended easily to the multi-stage case. Mean/target returns in the single-stage case are studied by Fishburn [4]. In contrast to the single-stage case, however, we must develop an efficient surface in the two parameters R_* and λ. Fortunately, many decision-makers have quite a precise idea of a reasonable target R_*. For example, one can vary R_* around the expected value of the second-stage cost of (1).

Table 2 presents the results for the capacity expansion problem of Section 3 when (11) is used. As expected, the variables y^k of the optimal solution of (11) are optimal for the second-stage problems.

5 Conclusion

We showed that extending the single-stage mean/variance formulation into the two-stage case may lead to erroneous results. To fix this problem, we incorporated the optimality conditions of the second-stage problems into the model. The result is a

R_*	λ	x	μ	σ^+
7320	0.0	(3.33, 8.67)	7500.00	221.81
	0.2	(3.23,8.77)	7506.39	220.91
	≥ 0.4	(3.22, 8.78)	7506.53	220.91
7420	0.0	(3.33, 8.67)	7500.00	140.95
	≥ 0.2	(3.17, 8.83)	7509.58	138.81
7520	0.0	(3.33, 8.67)	7500.00	62.18
	≥ 0.2	(3.17, 8.83)	7509.58	57.16

Table 2: Using the semi-variance as a measure of risk

quadratic program with equilibrium constraints. To find a global optimal solution, we formulated the problem as a mixed-integer program and solved it using branch-and-bound. We suggested using the semi-variance as a measure of variability for the second-stage cost. This mean/target formulation yields optimal solutions for the second-stage problems without losing convexity.

References

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

[2] A. Brooke, D. Kendrick, and A. Murows. *GAMS: A User's Guide.* Scientific Press, 1992.

[3] G. D. Eppen, R. K. Martin, and L. Schrage. A scenario approach to capacity planning *Operations Research*, 37(1989), 517–527.

[4] P. C. Fishburn. Mean-risk analysis with Risk associated with below-target returns. *American Economic Review*, 67(2), 116–125.

[5] M. Fukushima, Z. Q. Luo, and J. S. Pang. A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. Working paper.

[6] P. Kall and S. W. Wallace. *Stochastic Programming.* John Wiley & Sons, 1994.

[7] A. J. King. Stochastic Programming Problems: Examples from the Literature. *Numerical Techniques for Stochastic Optimization*, Y. Ermoliev and R. J. B. Wets (ed.), Springer Verlag, 1988.

[8] A. J. King. Asymmetric risk measures and tracking models for portfolio optimization under uncertainty. *Annals of Operations Research*, 45(1993), 165–177.

[9] H. Levy and H. M. Markowitz. Approximating expected utility by a function of mean and variance. *American Economic Review*, June 1979, 308–317.

[10] F. V. Louveaux and Y. Smeers. Optimal Investments for Electricity Generation: A Stochastic Model and A Test-Problem. *Numerical Techniques for Stochastic Optimization*, Y. Ermoliev and R. J. B. Wets (ed.), Springer Verlag, 1988.

[11] S. A. Malcolm and S. A. Zenios. Robust optimization of power systems capacity expansion under uncertainty. *Journal of the operational research society*, 45:1040–1049, 1994.

[12] H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments.* John Wiley & Sons, 1959.

[13] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios. Robust optimization of large-scale systems. *Operations Research*, 43:264–281, 1995.

[14] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, 1997.

[15] H. Van Stackelberg. *The Theory of Market Economy.* Oxford University Press, 1952.

[16] L. Vicente, G. Savard, and J. Júdice. Descent approaches for quadratic bilevel programming. *Journal of Optimization Theory and Applications*, 81:379–398, May 1994.

[17] H. Vladimirou and S. A. Zenios. Stochastic linear programs with restricted recourse. *European Journal of Operational Research*, 1996.