A Primer on Two Stage Stochastic Programming through Examples and Proposed Future Work

Alex Dowling

January 9, 2015

1 Introduction and Motivation

The goal of this document is to introduce a few key concepts in stochastic programming (two stage formulation, recourse action, etc.) using a simple unconstrained convex nonlinear optimization problem. This simple example has an analytic solution, making it ideal for algorithm development and debugging.

2 Example with Known Uncertain Parameters

Consider the following unconstrained, convex optimization problem:

$$\begin{aligned} \min_{x_1,x_2,d} \; z := & (x_1 - d + \omega_1)^2 + (1+\omega_1^2)(x_2 - d + \omega_2)^2 + (\omega_1 + \omega_3)x_1 + (\omega_2 + \omega_3)x_2 \\ & + \left[\omega_2^2 + \left(2 + \omega_3^2\right) d\right]^2 \end{aligned} \tag{1}$$

with three decision variables (d, x_1 and x_2) and three uncertain parameters (ω_1, ω_2 and ω_3). These uncertain parameters are random variables from arbitrary probability distributions with nominal values of zero.

2.1 Without Recourse: The Simplest Case

First, consider the case where d, x_1 and x_2 are determined simultaneously and then fixed. In other words, x_1 and x_2 are not allowed to change in response to changes in (i.e. additional information about) the uncertain parameters (ω). Thus this optimization problem is *without recourse*. The analytic solution is determined by setting the gradient of z to zero, which yields the following system of equations:

$$\frac{\partial z}{\partial x_1} = 2(x_1 - d + \omega_1) + \omega_1 + \omega_3 = 0 \tag{2a}$$
$$\frac{\partial z}{\partial x_1} = 2(1 + \omega_1^2)(x_2 - d + \omega_2) + \omega_2 + \omega_3 = 0 \tag{2b}$$
$$\frac{\partial z}{\partial d} = -2(x_1 - d + \omega_1) - 2(1 + \omega_1^2)(x_2 - d + \omega_2) + 2\left[\omega_2^2 + \left(2 + \omega_3^2\right) d\right]\left(2 + \omega_3^2\right) = 0 \tag{2c}$$

This system of equations is linear with respect to the decision variables:

$$\begin{bmatrix} 2 & 0 & -2 \\ 0 & 2\tau_1 & -2\tau_1 \\ -2 & -2\tau_1 & 2 + 2\tau_1 + 2\tau_3^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ d \end{bmatrix} = \begin{bmatrix} -3\omega_1 - \omega_3 \\ -(1 + 2\tau_1)\omega_2 - \omega_3 \\ 2\omega_1 + 2\omega_2\tau_1 - 2\omega_2^2\tau_3 \end{bmatrix} \tag{3}$$

where $\tau_1 = (1 + \omega_1^2)$ and $\tau_3 = (2 + \omega_3^2)$. The solution of (3), shown below, is the global solution of (1).

$$x_1 = -\frac{\omega_1 + \omega_2 + 2\omega_3 + 3\tau_3^2\omega_1 + \tau_3^2\omega_3 + 2\tau_3\omega_2^2}{2\tau_3^2} \tag{4a}$$
$$x_2 = -\frac{\tau_3^2\omega_2 + \tau_3^2\omega_3 + \tau_1\omega_1 + \tau_1\omega_2 + 2\tau_1\omega_3 + 2\tau_1\tau_3\omega_2^2 + 2\tau_3^2\tau_1\omega_2}{2\tau_1\tau_3^2} \tag{4b}$$
$$d = -\frac{\omega_1 + \omega_2 + 2\omega_3 + 2\tau_3\omega_2^2}{2\tau_3^2} \tag{4c}$$

Table 1 shows two scenarios for the uncertainty variables, and the corresponding optimal values if (1) is solved with the values of ω known *a priori*. The final column, marked "Scenario 2*", shows the objective function value if the decision variables from Scenario 1 are applied to Scenario 2. This, this demonstrates the sensitivity of decision variables and objective function values to perturbations in the uncertain parameters.

Table 1: Three Decision Variables

	Scenario 1	Scenario 2	Scenario 2*
ω_1	0	2	2
ω_2	0	2	2
ω_3	0	2	2
x_1	0	-4.7778	0
x_2	0	-3.1778	0
d	0	-0.7778	0
z	0	-26.5778	40

2.2 With Recourse: A Two Stage Approach

In many engineering applications, decisions are made in stages, as additional information regarding the uncertain parameters are realized. For example in a chemical process design application, decisions variables can be partitioned into two categories: design variables and operating variables. Design variables include the size of units, the number of parallel units and connectivity decisions. These variable values must be determined before or while the chemical process is constructed. Operating variables, such as stream flowrates, temperatures, pressures and temperatures, may be changed while the process is under operation. They are typically adjusted in response to uncertainty, such as variable product purity specifications, changing market values of various products and raw materials, fluctuating ambient conditions, etc.

The original optimization problem, (1), can be reformulated into the classic two stage formulation

$$\min_{d} \min_{x_1,x_2} z(d, x_1, x_2, \omega_1, \omega_2, \omega_3) \tag{5}$$

where d is a design variable, and x_1 and x_2 are operating variables. The outer problem represents the first stage of decisions (design variables) and the inner problem represents the second stage (operating variables). The optimal choice for x_1 and x_2 in the inner problem depends on the realized values for the uncertain parameters (ω) and the selected value for the design variable (d).

Consider the inner problem, where d and ω are fixed. Setting the gradients of z (the objective) to zero yields the following system of equations:

$$\begin{bmatrix} 2 & 0 \\ 0 & 2\tau_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3\omega_1 - \omega_3 + 2d \\ -(1+2\tau_1)\omega_2 - \omega_3 + 2\tau_1 d \end{bmatrix} \tag{6}$$

with the solution

$$x_1 = d - \frac{3\omega_1 + \omega_3}{2} \tag{7a}$$

$$x_2 = d - \omega_2 - \frac{\omega_2 + \omega_3}{2\tau_1} \tag{7b}$$

Table 2 returns to the two scenarios previously discussed, but assumed d if fixed at the optimal value for the nominal case (Scenario 1). Thus, the problems in Table 2 have one less degree of freedom, and as expected the objective function value is larger (worse), $z = -4.8$ versus $z = -26.6$, than when d is also considered in the optimization problem (Table 1). However, the objective function value is much better than the no recourse case ("Scenario 2*" in Table 1), highlighting the importance of recourse actions.

3 Stochastic Formulation

Ultimately, uncertainty information should be utilized for the solution procedure for the outer problem of (5). The consequences of not considering uncertain information when selecting d are shown later. A straightforward approach for this is the stochastic two stage mathematical program with recourse, which is shown below:

Table 2: Two Decision Variables

	Scenario 1	Scenario 2
ω_1	0	2
ω_2	0	2
ω_3	0	2
d	0	0
x_1	0	-4
x_2	0	-2.4
z	0	-4.8

$$\min_{d} \sum_{i \in \{I\}} p_i \left[\min_{\mathbf{x_i}} \; z(d, \mathbf{x_i}, \omega_{\mathbf{i}}) \right] \tag{8}$$

with probability information for the uncertain parameters reflected in the collection of scenarios, $\{I\}$. Scenario i contains a single value for each uncertain parameter $(\omega_{1,i}, ..., \omega_{3,i})$ and the probability of occurrence for the scenario (p_i). Thus the set of scenarios $(\{I\})$ is a discrete multivariate probability distribution for the uncertain parameters. The two stage formulation shown in (8) may be flattened, resulting in

$$\min_{x_1,x_2,d} \quad \phi = \sum_{i \in \{I\}} p_i z_i \tag{9a}$$

$$\begin{aligned} \text{s.t.} \quad z_i = {} & (x_{1,i} - d + \omega_{1,i})^2 + (1 + (\omega_{1,i})^2)(x_{2,i} - d + \omega_{2,i})^2 \\ & + (\omega_{1,i} + \omega_{3,i})x_{1,i} + (\omega_{2,i} + \omega_{3,i})x_{2,i} \\ & + \left[\omega_{2,i}^2 + \left(2 + \omega_{3,i}^2\right) d\right]^2 \qquad \forall i \in \{I\} \end{aligned} \tag{9b}$$

which allows for d along with $\mathbf{x_1}$ and $\mathbf{x_2}$ to be determined simultaneously. This is the preferred approach for many stochastic programs that leverage "off the shelf" optimization solvers, such as CPLEX. Note that the operating variables may have unique values for each scenario, which allows for recourse actions. As consequence, the operating variables are indexed over each scenario, notated $x_{1,i}$ and $x_{2,i}$. This means the number of decision variables in the stochastic program grows rapidly, as each operating variable is copied into each scenario.

In this particular example, it is easiest to consider the formulation of (8), as the analytic solution for the inner problem is given in (7). Thus the inner problem may be replaced with a simple formula for z using substitution:

$$\begin{aligned} z = {} & \frac{(\omega_1 + \omega_3)^2}{4} + \frac{(\omega_2 + \omega_3)^2}{4(1 + \omega_1^2)} + (\omega_1 + \omega_3)(d - \frac{3\omega_1 + \omega_3}{2}) \\ & + (\omega_2 + \omega_3)(d - \omega_2 - \frac{\omega_2 + \omega_3}{2(1 + \omega_1^2)}) + \left[\omega_2^2 + \left(2 + \omega_3^2\right) d\right]^2 \end{aligned} \tag{10}$$

3.1 Numeric Example: Uniform Distributions for Uncertainty

Consider uniform distributions from -3 to 3 for all three uncertain parameters (ω_1, ω_2, ω_3). These continuous probability distributions can be approximated with a discrete probability distribution, in the form of many scenarios. In this example, consider seven equally spaced levels for each uncertain parameter (-3, -2, -1, 0, 1, 2, 3). A full factorial design results in $7^3 = 343$ scenarios. $\{I\}$ represents the full set of scenarios.

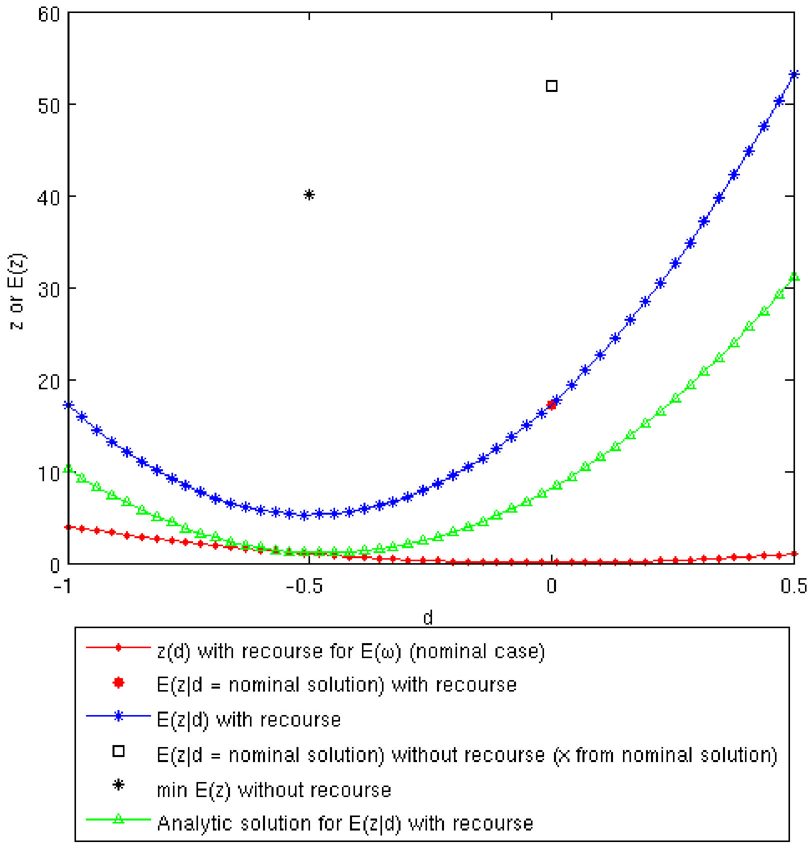

Figure 1: Comparison of the objective function for the nominal and stochastic cases. Note: all of the stochastic cases, except the analytic solution, use 343 scenarios in place of a continious multivariable probability distribution.

Figure 1 shows the objective as a function of d, the design parameter, for both the nominal case ($\omega_1 = \omega_2 = \omega_3$) and the stochastic case approximated with the 343 scenarios. In the nominal case, $d = 0$ coincides with the best objective function ($z = 0$). However, when uncertainty is taken into account, the optimal value for d moves to -0.5 with $E(z) = 5.2571$. If one simply optimizes d for the nominal case and then considers uncertainty, $E(z) = 17.2571$ (which is shown by the red dot on the blue line). Thus considering uncertainty while optimizing d results in a $(17.2571 - 5.2571)/17.2571 = 69.5\%$ savings for this synthetic test problem. The unfilled square and black * show the expected value of E(z) without recourse (i.e. x_1 and x_2 are the same for all scenarios) for the nominal case ($d = 0$) and the stochastic optimal solution ($d = -0.5$), respectively. These points have substantially higher objective function values, which highlights the importance of recourse action.

The analytic solution using the assumed continuous probability distributions may be calculated as follows,

$$\mathbb{E}(z|d) = \int_{\Omega} z(d,\omega)\rho(\omega)\,\mathrm{d}\omega \tag{11}$$

where $z(d,\omega)$ is given in (10) and $\rho(\omega)$ is the joint continuous probability density function for the vector of uncertain parameters, ω. For this example, independent uniform probability distribution are assumed and $\rho(\omega) = (1/6)^3$. Thus,

$$\mathbb{E}(z|d) = \frac{1}{6^3}\int_{-3}^{3}\int_{-3}^{3}\int_{-3}^{3} z(d,\omega_1,\omega_2,\omega_3)\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2\,\mathrm{d}\omega_3 \tag{12a}$$

$$= \frac{161}{5}d^2 + 30d + \frac{87}{10} - \frac{\arctan(3)}{2} \tag{12b}$$

The minimum of (12) occurs when $\frac{\mathrm{d}\mathbb{E}(z|d)}{\mathrm{d}d} = 0$, i.e. $d = -75/161 = -0.4658...$ and $\mathbb{E}(z) = 2757/1610 - \arctan(3)/2 = 1.0880...$, which suggests the discretization with 373 scenarios results in noticeable error. At the nominal solution, $\mathbb{E}(z|d = 0) = 8.0755...$ and at the previously found stochastic optimum $\mathbb{E}(z|d = -0.5) = 1.1255$. This shows that although the stochastic optimization procedure with 343 scenarios does not identify the correct $\mathbb{E}(z)$, the choice for d is reasonable.

4 Challenges and Proposed Work

The most significant downside of the two stage stochastic program with recourse formulation is problem size explosion, which is a consequence of two factors. First, the number of scenarios is combinatorial with the number of uncertain parameters considered. For example, ten uncertain parameters with five levels each results in $5^{10} = 9.77$ million scenarios. Furthermore, considering recourse requires each operating variable to be duplicated for each scenario, dramatically increasing the number of variables considered.

The remainder of this section proposed work to mitigate problem size explosion, along with a demonstration problem.

4.1 Uncertainty Team

Scenarios are essential for two reasons: first, scenarios ensure the inner problem is feasible first, they allow for the impact of recourse actions to be quantified. Without recourse action, design decisions tend to be too conservative and the stochastic objective function is overestimated. Too many scenarios result in problem size explosion, whereas too few scenarios either poorly represent the underling probability distribution, or don't enable enough recourse actions (resulting in too conservative of decisions). The later may happen, as the uncertain parameter probability distributions are typically continous in nature, but the set of scenarios are a discrete probability distribution. Multi-modal probability distributions may require a large number of levels at achieve an accurate discretiation.

There is extensive room for theoretical and algorithmic contributions regarding scenario generation/reduction. For example, not all of the scenarios are significant in many applications. With a fine discretization, many scenarios may have near zero probabilities of occurring. Thus, it may be possible to aggregate a number of similar low probability scenarios together into a few scenarios with non-negligible probabilities but little impact on accuracy. Furthermore, the objective function of the inner problem (with recourse) may be insensitive to perturbations in certain dimensions of the uncertain parameter space. Reducing the number of scenarios in these dimensions may allow for a smaller stochastic program with minimal additional error. Finally, it may be possible to show ideas using both continuous and discrete probability distributions and integral approximations presented by Brenda and Charles are mathematical equivalent to a scenario reduction strategy.

Possible Tasks for the Uncertainty Team:

1. Study existing scenario reduction techniques and literature, especially the work of Werner Römisch (very mathematical)

2. Develop a methodology to identify the most important uncertain dimensions by considering both the available probability information and sensitivity to the inner problem (with recourse) objective with respect to uncertain parameters. Principle component analysis was previous discussed during a teleconference.

3. Develop a scenario reduction strategy that leverages the available continuous probability information. This may be closely related to the previously proposed ideas with integral approximations and surrogate models.

4. Demonstrate these methods on both synthetic examples and an application process design problem

5. Implement the final versions of the methods in FOQUS

4.2 Optimization Team

Optimization problem size explosion of formulation (9) may be mitigated by instead considering the two stage formulation, (8), and using built-in optimization tools in commercial flowsheet simulators to solve the inner problem. The outer problem may then be solved using derivative-free optimization algorithms. Future work may involve using a derivative-based algorithm for the outer problem, although this requires some theoretical developments (the inner problem may be non-smooth).

Possible Tasks for the Optimization Team:

1. Assess the feasibility of using the SQP solver in Aspen for the inner problem

2. Assess alternate DFO solvers, such as BOBYQA, and determine the best settings for the selected solver

3. Demonstrate the proposed method on both synthetic examples and an application process design problem

4. Implement the final versions of the methods in FOQUS

4.3 Demonstration Problem

One unique advantage of the CCSI project is an established methodology to propagate uncertain information from bench scale experiments into equipment models. In contrast, many works focused on optimization under uncertainty use generated uncertainty information in synthetic academic problems. Building off this advantage, the bubbling fluidized bed with uncertain kinetic and transport model parameters is a logical demonstration system.