Annals of Operations Research 132, 47–68, 2004

Better Optimization of Nonlinear Uncertain Systems (BONUS): A New Algorithm for Stochastic Programming Using Reweighting through Kernel Density Estimation

KEMAL H. SAHIN * and URMILA M. DIWEKAR Sahin@cpc-net.com, urmila@uic.edu
Center for Uncertain Systems: Tools for Optimization & Management (CUSTOM), Department of Chemical Engineering, Institute of Environmental Science & Technology, University of Illinois, Chicago, IL, 60514, USA

Abstract. A new nonlinear programming algorithm is proposed for stochastic programming problems. This method relies on sampling to estimate the probabilistic objective function and constraints. The computational burden of excessive model calculations for determining the search direction is bypassed through a reweighting method using Kernel Density Estimation. The improvements accomplished by this algorithm called Better Optimization of Nonlinear Uncertain Systems (BONUS) are presented through two real world case studies involving parameter design for off-line quality control of a chemical reactor, and optimal capacity expansion for electric utilities in uncertain markets.

1. Introduction

Many optimization problems for real-life applications fall under the category of nonlinear programming (NLP) due to the inability of representing the system at hand using only linear formulations. For decades, solution techniques based on quasi-Newton techniques (Davidon, 1991) have been applied to estimate the Hessian for NLPs, bypassing the exact, computationally intensive, calculation of the matrix.

However, these methods are still not fully representative of real life systems, as underlying uncertainties are neglected. Hence, optimization under uncertainty (or stochastic programming) algorithms have to be developed, which can include both the stochastic aspects of the problem as well as their ability to handle nonlinear formulations. From a mathematical perspective, Stochastic Nonlinear Programming (SNLP) problems can be characterized as follows:

$$\text{optimize } J = P_1\big(f_1(\theta, x, u)\big) \tag{1}$$

$$\text{subject to} \quad P_2\big(f_2(\theta, x, u)\big) = 0, \tag{2}$$

$$P_3\big(f_3(\theta, x, u) \leqslant 0\big) \geqslant \alpha, \tag{3}$$

* Current address: CPC Cellular Process Chemistry Systems GmbH, Hanauer Landstr. 526/G58 III D-60343, Frankfurt am Main, Germany.

where θ is the set of decision variables, u is the set of uncertain variables, x is the set of parameters, α is a constant and P_1, P_2, P_3 represent probabilistic functions like mean, variance, or fractiles. Introducing uncertainty within such an optimization framework allows for the determination of more realistic results. However, this also adds complexity, as the uncertainty of the system has to be analyzed. Uncertainties can be represented as discrete random variables in terms of scenarios. However, this generally poses the problem of loss of differentiability and derivative based techniques cannot be applied. The other approach is to use continuous random variables for representing uncertainties and sampling to propagate uncertainties. This approach is more general and is the focus of this paper. However, the sampling approach introduces another problem, with large number of samples or uncertainty realization, traditional nonlinear programming techniques generally break down due to the computational intensity involved in evaluating function and derivative using each sample.

There are two fundamental approaches used to solve stochastic nonlinear programming problems. The first set of techniques identifies problem specific structures and transforms the problem into a deterministic nonlinear programming problem. For instance, chance constrained programming (Charnes and Cooper, 1959) replaces the constraints that include uncertainty with the appropriate probabilities expressed in terms of moments. The major restrictions in applying the chance constrained formulation include that the uncertainty distributions should be stable distribution functions, the uncertain variables should appear in the linear terms in the chance constraint, and that the problem needs to satisfy the general convexity conditions. The advantage of the method is that one can apply the deterministic optimization techniques to solve the problem. Decomposition techniques like L-shaped decomposition (Birge and Louveaux, 1997) divide the problem into stages and generate bounds on the objective function by changing decision variables and solving subproblems that determine the recourse action with respect to the uncertain variables. However, these methods also require convexity conditions and/or dual-block angular structures, and are only applicable to discrete probability distributions. For example, Lagrangian-based approaches have been applied to nonlinear SP formulations. The Lagrangian Dual Ascent method has been proposed by Rockafellar and Wets (Rockafellar and Wets, 1976) for problems with finite outcomes for the uncertain variables. Another technique is Regularized Decomposition, which adds quadratic terms to the objective for improved convergence (Ruszczynski, 1986) of the L-shaped decomposition method. The augmented Lagrangian method adds a quadratic penalty to ensure convexity, yielding more efficient computation (Dempster, 1988). Rockafellar and Wets also developed a similar technique – the Progressive Hedging Algorithm (Rockafellar and Wets, 1991). Clearly, these techniques have significant limitations on either the objective function type or the underlying distributions for the uncertain variables.

As stated earlier, an alternative approach that can be used to capture uncertainty is through a sampling loop that is embedded within the optimization iterations for the decision variables. Figure 1 represents the generalized solution procedure, where the deterministic model is replaced by an iterative stochastic model with sampling loop representing the discretized uncertainty space. This step can be computationally expensive,

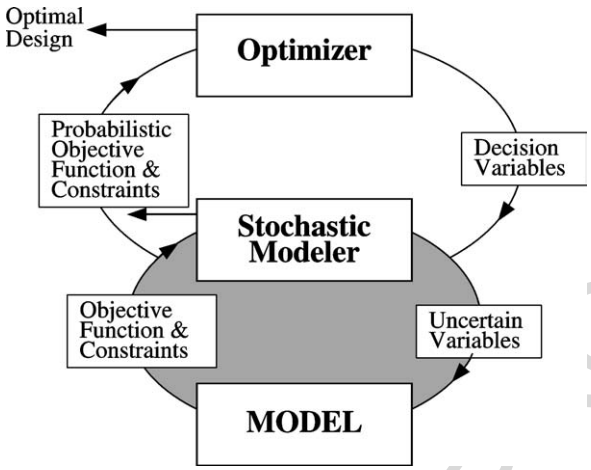

Figure 1. Optimization under uncertainty.

as the model has to be rerun for every sample point. In these cases, one relies on approximations. The approximations in most of the sampling methods takes the form of bounds on actual functions or error estimates of the functions in terms of number of samples. For specific structures where the L-shaped method is applicable, two approaches consider these approximations by embedding sampling within another algorithm without complete optimization. These two approaches are the method of Dantzig (Dantzig and Infanger, 1991), which uses importance sampling to reduce variance in each cut based on a large sample, and the stochastic decomposition method proposed by Higle and Sen (Higle and Sen, 1991), which uses the lower bound of the recourse function based on expectation. Again these methods exploit specific structures of the problem and require convexity conditions. In this paper, we are using the generalized form of the stochastic nonlinear program as is but use a statistical reweighting technique for speeding up function and derivative estimation, which will be used during the optimization algorithm. This approach bypasses the need to evaluate the model at each iteration for each sample point. Our Better Optimization for Nonlinear Uncertain Systems (BONUS) algorithm is schematically explained below.

Optimization under uncertainty involves iteratively improving a probabilistic objective function. Figure 2(a) shows uncertain input variables with an underlying probability density function $pdf_{\text{in}}(\theta^k, u^k)$, shown as the solid triangular distribution. After the model is run, the corresponding output distribution $cdf_{\text{out}}(\theta^k, u^k)$ is generated, shown as the solid line in figure 2(b). As optimization progresses to the next iteration, $k+1$, moments such as mean, variance, and the probability function can change for the uncertain variables, resulting in a new $\widehat{pdf}_{\text{in}}(\theta^{k+1}, u^{k+1})$, indicated by the dashed line in figure 2(a).

Our goal is to identify rapid and efficient techniques that determine an approximation of the properties of the new output distribution, $\widetilde{cdf}_{\text{out}}(\theta^{k+1}, u^{k+1})$, given as the dashed cumulative distribution function in figure 2(b). Advantage of this approach is its

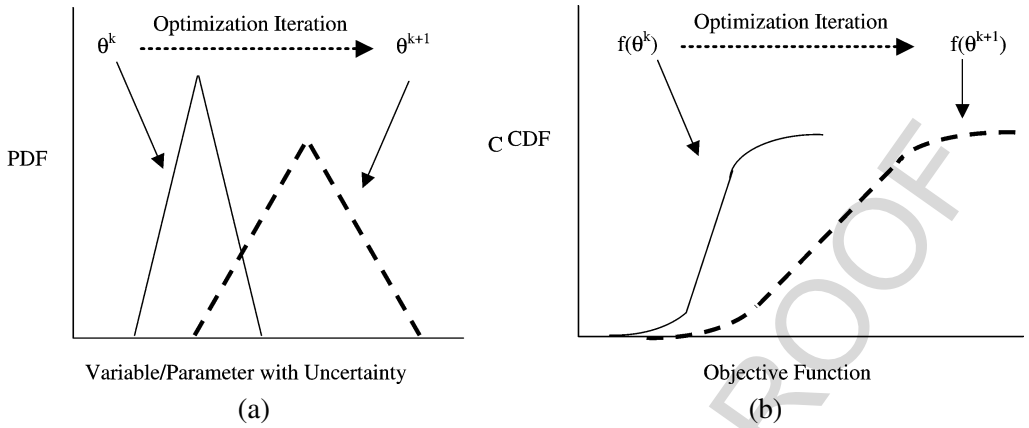

Figure 2. Effect of changes in decision variables.

bypassing of the model evaluations for successive sampling (the inner loop in figure 1), which is computationally the most intensive task for optimization under uncertainty. The following section will introduce the reweighting approach that will be used for estimating the output distribution properties. The BONUS algorithm will be summarized in section 3, and section 4 will introduce an off-line quality control problem aimed at minimizing product variations for a continuously stirred tank reactor (CSTR), one of the most frequently used equipments in chemical industries, followed by the electricity capacity expansion problem popular in operations research literature in section 5. Finally, section 6 will present the conclusions.

2. Nonlinear stochastic programming through reweighting

Figure 3 shows the general idea behind the BONUS algorithm. The traditional SNLP methods, shown in figure 1, rely on developing a sampling loop, and evaluating this loop for every sample that is generated using the input distribution. This step is repeated during the optimization process over and over again, until convergence is observed.

The technique proposed here follows the grey arrows. Instead of running the model for every sample point for every iteration, information about the distributions is used to estimate the output distribution. An initial, uniform base distribution is generated and the model is run to determine the output distribution. The next step involves generating the output distributions for the new sample sets that are encountered during optimization. At this stage, the model is not re-run, instead, a reweighting approach is applied to approximate the probabilistic behavior of the new output distribution. The details on this reweighting approach are provided below.

2.1. Reweighting schemes

As stated above, the goal is to determine changes in output distributions as input distributions change. Hesterberg (1995) presents various reweighting techniques for esti-

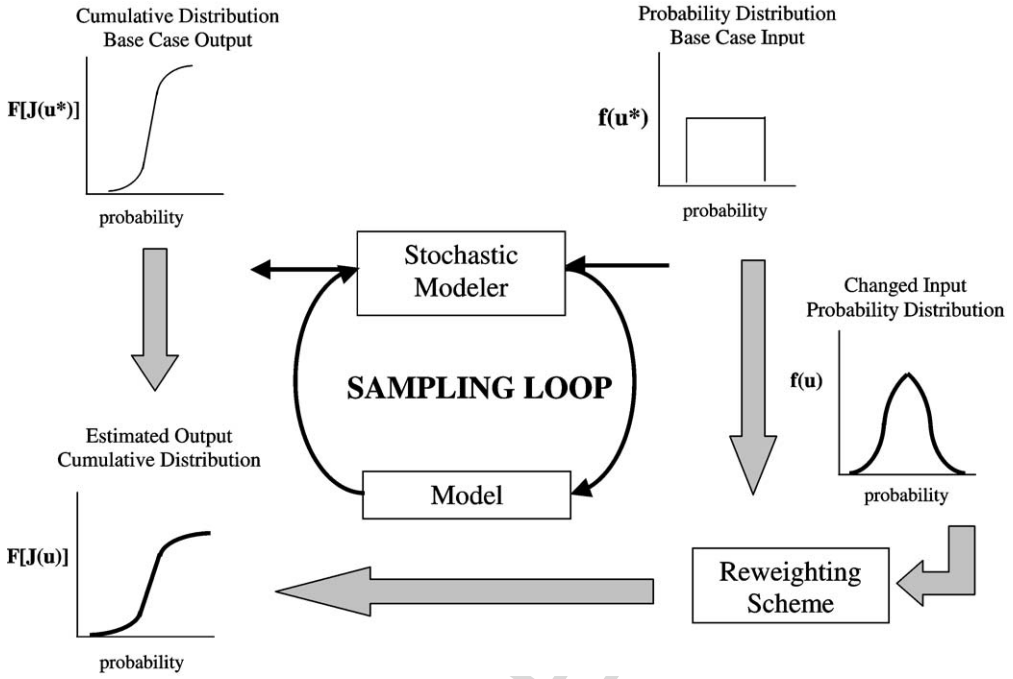

Figure 3. Density estimation approach to optimization under uncertainty.

mating the expected value of an output distribution (CDF, $F[J(u)]$) without evaluating the model for the input distribution (PDF, $f(u)$), as shown in figure 1. The ratio of the probability density functions f is used as a weight, as follows:

$$\omega_i = \frac{f(u_i)}{\hat{f}(u_i^*)}, \tag{4}$$

where $\hat{f}(u_i^*)$ is determined for a base sample set, for which the model response is known, and the probability density $f(u_i)$ is calculated using the sample for which the response has to be estimated. Remember that these two sample sets are not necessarily related. One attempt for estimating statistical properties $P(u)$ of the model is through the product of the weights and the same properties obtained from the base distribution (equation (5)):

$$P(u) = \sum_i \omega_i P\left(u_i^*\right). \tag{5}$$

For instance, to estimate the mean μ of a model response, $Z(u)$, the weight would be multiplied by the individual model responses for the base set:

$$\mu\left[Z(u)\right] = \sum_i^{N_{\text{samp}}} \omega_i Z\left(u_i^*\right), \tag{6}$$

where N_{samp} is the sample size.

This approach has limitations, as the weights may not sum to 1. The work presented in this paper will reduce this problem by using normalized weights, as shown in equation (7). This normalized reweighting (the ratio estimate for weighted average) has another advantage as it provides acceptable performance for a wider range of perturbation, especially for large samples of Monte Carlo simulations (MCS) (Hesterberg, 1996). In this work, instead of using large size of MCS, a more efficient sampling technique that provides the same accuracy as MCS in order of magnitude less number of samples is used:

$$P(u) = \sum_{j}^{N_{\text{samp}}} \frac{f(u_j)/\hat{f}(u_j^*)}{\sum_{i=1}^{N_{\text{samp}}} f(u_i)/\hat{f}(u_i^*)} P\left(u_j^*\right). \tag{7}$$

As seen in equation (7) the mean of the function can be estimated from the ratio of the two input distributions $f(u)$ and $\hat{f}(u^*)$. This requires the determination of the probability distributions from a given sample set of uncertain variables. Given a set of data points, several methods can be used to approximate an underlying distribution. A very common approach is identifying the mean and variance for a normal distribution that is representative of the data at hand. This, and similar parametric approaches assume a specific type of distribution for the PDF, and determine the appropriate constants for the best possible fit.

Alternatively, non-parametric density estimation does not impose a strict form on the data, allowing for a more accurate approximation. The following section describes the fundamental properties for the non-parametric density estimation method used in this work.

2.2. *Estimating distributions using kernel density functions*

Silverman (1986) reviews several techniques for nonparametric estimation of the probability density function. The easiest approach to non-parametric density estimation is to use histograms. However, histograms are not suitable for the optimization approach due to their discontinuous nature. Among the methods described in this review by Silverman (1986), the Gaussian Kernel Density Estimation technique has been used in this work due to the advantage of having continuous derivatives for all orders.

To estimate the probability density function for a random variable u, the Gaussian Kernel Estimator relies on a kernel function (equation (8)) that generates a normal distribution for data point of the known set u_i. By adding up these normal distributions with mean u_i and variance h^2 over the entire data set (equation (9)), an estimate of the underlying probability density function $\hat{f}(u)$ is obtained. Figure 4 is a graphical representation, where the data points are given as the crossed boxes on the x-axis. The individual Gaussian distributions are the dashed lines, and their sum is given as the solid line for the probability distribution of the system:

$$K\left(\frac{u-u_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(1/2)((u-u_i)/h)^2}, \tag{8}$$

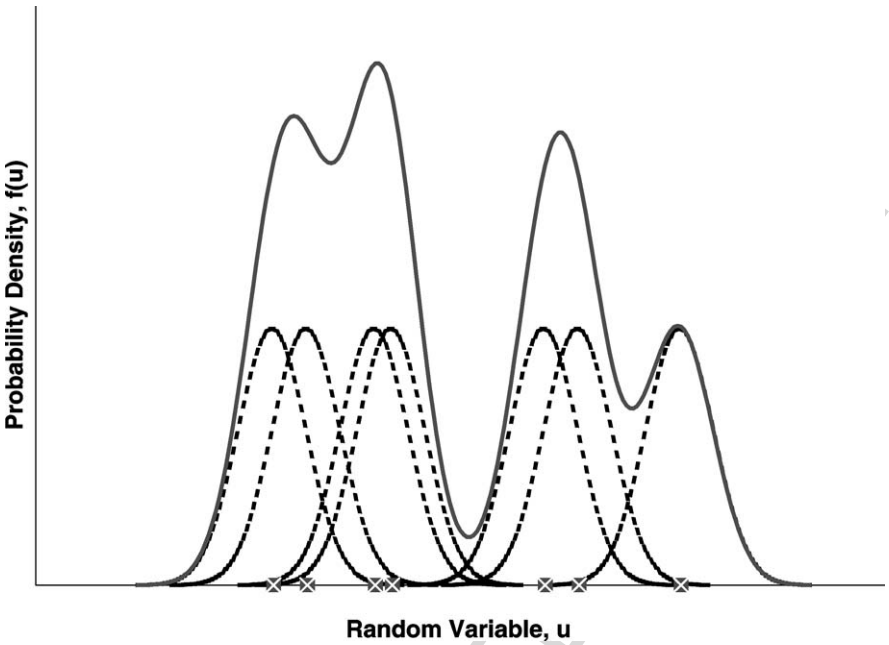

Figure 4. Approximating probability distributions using kernel density approach.

$$f(u) = \frac{1}{N_{\text{samp}} h} \sum_{i=1}^{N_{\text{samp}}} K\left(\frac{u - u_i}{h}\right), \tag{9}$$

where N_{samp} is sample size for kernel estimation.

In order to use the kernel density approach for estimating function values (objective function and constraints), a base sample set u^* has to be generated for model calculations. We are selecting base distribution to be uniform, to avoid emphasis on any particular regions. The base distribution has to encompass the entire region of possible levels of uncertainties that may be encountered in the following iterations. For instance, if a decision variable has uncertainty of $\pm 10\%$ with lower and upper bounds of 100 and 200, respectively, the uniformly distributed sample should be generated across the range [90, 220]. Once the base distribution is obtained, its density can be calculated for each point as

$$\hat{f}\left(u_i^*\right) = \frac{1}{N_{\text{samp}} h} \sum_{j=1}^{N_{\text{samp}}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(1/2)((u_i^* - u_j^*)/h)^2}. \tag{10}$$

After determining the model output $Z(u_i^*)$ for each u_i^*, the optimization algorithm generates new decision variables. This is followed by the generation of a new sample for uncertainty analysis. Instead of re-running the model for the new sample, the probability of each new data point u_i is determined through the kernel density approximation

(equation (11)):

$$\hat{f}(u_i) = \frac{1}{N_{\text{samp}} h} \sum_{j=1}^{N_{\text{samp}}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(1/2)((u_i - u_j^*)/h)^2}. \tag{11}$$

2.3. Numerical examples for reweighting

The proposed reweighting scheme using kernel density estimation has been carried out for case studies up to $d = 10$ dimensions. The application of alternative, more efficient sampling techniques, such as Latin Hypercube Sampling (LHS), Median Latin Hypercube Sampling (MLHS), and Hammersley Sequence Sampling (HSS) have resulted in significant reductions of computational requirements compared to Monte Carlo Sampling (Kalagnanam and Diwekar, 1997). The effects and computational efficiencies of these sampling techniques will be compared for the following five types of functions:

- *Function 1.* Linear additive: $y = \sum_{m=1}^{s} u_m$, $s = 2, \ldots, 10$;
- *Function 2.* Multiplicative: $y = \Pi_{m=1}^{s} u_m$, $s = 2, \ldots, 10$;
- *Function 3.* Quadratic: $y = \sum_{m=1}^{s} u_m^2$, $s = 2, \ldots, 10$;
- *Function 4.* Exponential: $y = \sum_{m=1}^{s} u_m \exp(u_m)$, $s = 2, \ldots, 10$;
- *Function 5.* Logarithmic: $y = \sum_{m=1}^{s} \log(u_m)$, $s = 2, \ldots, 10$.

The total analysis includes 5 function, with 4 sampling techniques being compared for each of these functions. The number of sample points for each sample is also analyzed, by selecting sample sizes as $N_{\text{samp}} = [50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000]$. This results in a total of 200 runs for which the proposed reweighting approach has been tested. For each run, the means and variances are both calculated and estimated, as are the derivatives of each of these with respect to each u. Further, the percentage error between the actual and estimated values is determined as well, as shown in table 1.

As required, the base distributions are uniform distributions with bounds given in first three columns of table 2, and the estimated distributions were normal, with the upper and lower bounds in the last three columns of the table indicating the region enclosing the 99.999 percentile.

For the generation of the shifted sample set u^{Δ} and for derivative calculations, the step size Δu_j was selected as

$$\Delta u_j = 0.05\mu\{u_j\}. \tag{12}$$

As the model functions are relatively simple, the actual values of the mean and variance for both sample sets u and u^{Δ} are calculated, and compared to the estimates. Further, the same analysis is conducted for the derivative estimates, allowing to compare the error in the estimates based on the sampling technique that is applied to generate both sample sets u^* and u. The next subsection provides the results of the preliminary study.

Table 1
Calculations for KDE efficiency analysis.

n-dimensional calculations	2–10 dimensions $= 10! = 3628800$
Functions	5
Sampling techniques	4
Sample sizes	10
Total runs	$5 \cdot 4 \cdot 10 = 200$
Moment calculations/run	4
Derivative calculations/run	$10! \cdot 2$
Moment estimations/run	4
Derivative estimations/run	$10! \cdot 2$
% Error calculations moments/run	2
% Error calculations derivatives/run	$10! \cdot 2$
Total calculations	$(4 + 4 + 2 + 6 \cdot 10!) \cdot 200 \simeq 4.32 \cdot 10^9$

Table 2
Bounds for base (uniform) and estimated (normal) distributions.

	Base distribution			Estimated distribution	
	Lower bound	Upper bound		Lower bound	Upper bound
u_1^*	1.0	6.0	u_1	3.0	5.0
u_2^*	3.0	7.0	u_2	4.0	7.0
u_3^*	1.0	5.0	u_3	3.0	4.0
u_4^*	8.0	12.0	u_4	9.5	10.0
u_5^*	10.0	17.0	u_5	11.5	14.0
u_6^*	2.0	9.0	u_6	4.0	6.0
u_7^*	3.0	7.0	u_7	4.5	6.5
u_8^*	0.0	7.5	u_8	1.0	6.0
u_9^*	10^{-5}	10^{-1}	u_9	$5 \cdot 10^{-3}$	$5 \cdot 10^{-2}$
u_{10}^*	6.0	9.0	u_{10}	8.0	9.0

2.4. *Results and discussion for reweighting approach*

As indicated above, 200 different runs have been used to verify the applicability of the technique. For each run, means, variances, and derivatives have been calculated and estimated using the reweighting scheme, and percentage errors between each of these have been determined. Due to the extensive nature of this analysis, we provide only one example here, that is both relevant to our analysis as well as representative of the overall behavior of the technique.

We present the results obtained for the nonlinear function, $y = \sum_{m=1}^{3} u_m^2$. As variance calculation is more prone to errors than calculation of mean (if sample size is less), and also the case study in section 4 aims at calculating the variance of the system at hand, we present the efficiency of the reweighting technique to estimate variance for this function here.

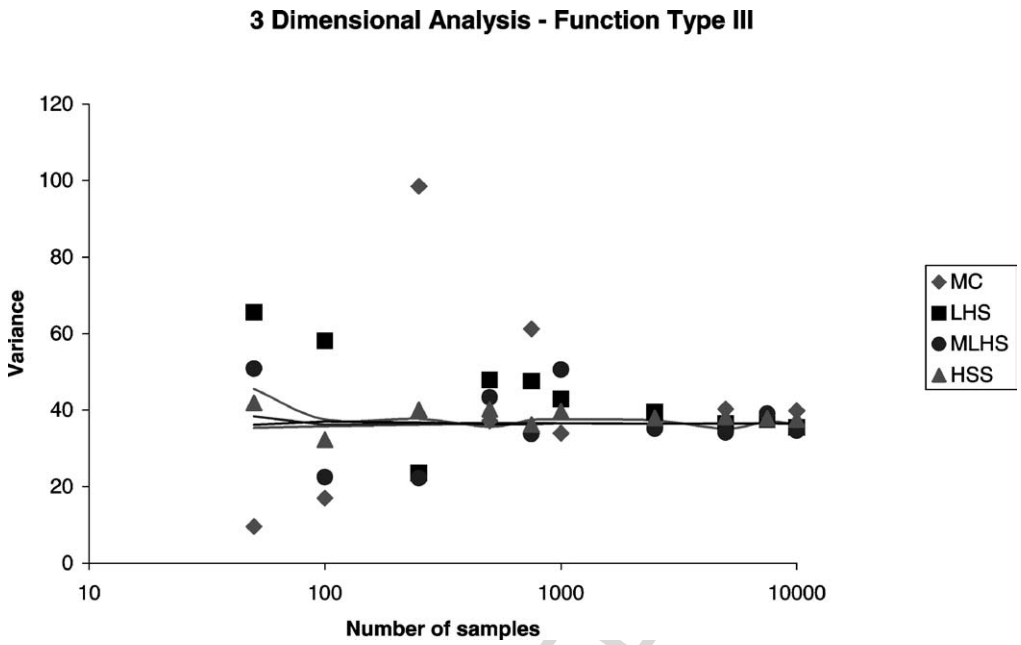

Figure 5. Variance calculation for different sampling techniques.

Table 3
Percentage error in variance estimation for 3 dimensional analysis using 250 samples.

	MC	LHS	MLHS	HSS
Function 1	178.5688	34.6615	50.0478	7.3945
Function 2	179.7385	30.1337	54.5286	11.1636
Function 3	161.1127	36.0205	39.1106	10.9293
Function 4	140.1933	9.2476	13.0681	4.1226
Function 5	183.3928	30.1835	54.2601	8.9386

Simultaneous plotting of the actual and estimated values will allow us to identify how accurate each technique is. Note that the x-axis is in log scale to capture the change of the sample sizes through $N_{\text{samp}} = [50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000]$. The lines present the actual values, while the stand-alone points represent the estimated variance values using the four different sampling techniques.

In figure 5, the variance of function 3 is plotted with respect to the number of samples. As seen, all four sampling techniques converge to the same value as N_{samp} approaches 10000, with the MC sampling technique showing the highest variations. While most approaches over- or underestimate the mean at low sample sizes, HSS provides a rather accurate estimate in this region. Table 3 provides the percentage error between the estimate and the actual values of the variance for $f(u)$ for all four sampling techniques with sample sizes of 250. As seen, HSS yields comparably small percentage errors for all functions.

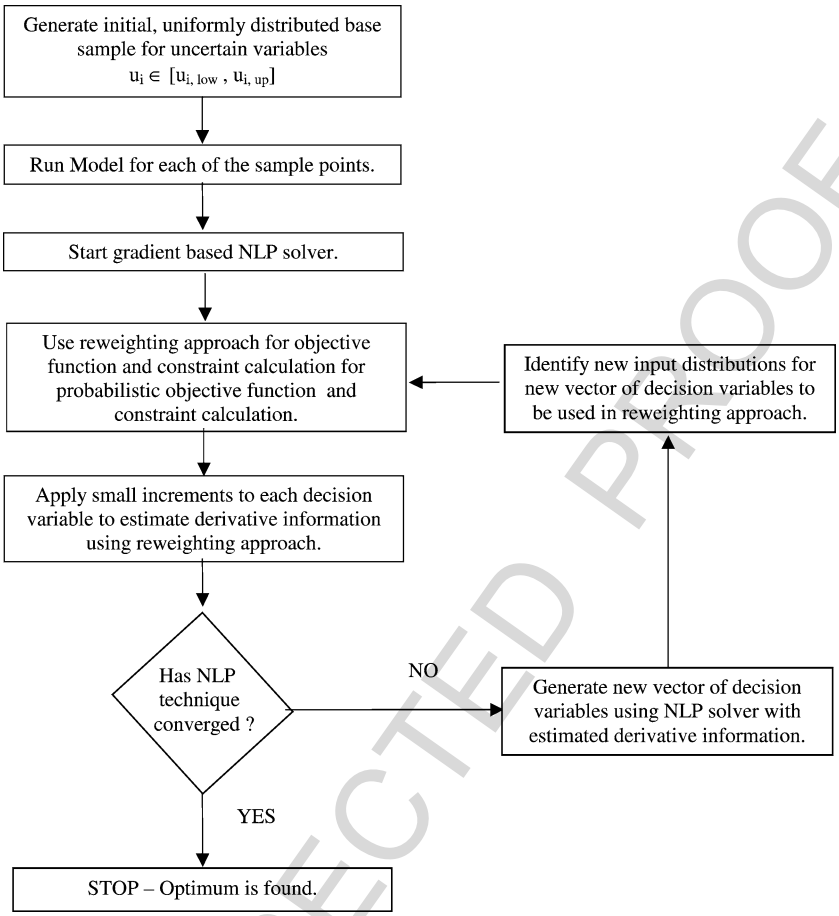

Figure 6. Optimization under uncertainty: the BONUS algorithm.

3. BONUS: the new SNLP algorithm

The algorithm for BONUS, given in figure 6, can be divided into two sections. The first section, Initialization, starts with generating the base distribution that will be used as the source for all estimations throughout the optimization. After the base distribution is generated, the second section starts, which includes the estimation technique that results in the improvements associated with BONUS.

In this algorithm overview, we denote the D-dimensional vector of deterministic decision variables as $\theta = [\theta_1, \theta_2, \ldots, \theta_d, \theta_{d+1}, \ldots, \theta_D]$, while the S-dimensional uncertain variables are defined as $u = [u_1, u_2, \ldots, u_s, u_{s+1}, \ldots, u_S]$.

I. Initialization

1. Generate ($i = 1, \ldots, N_{\text{samp}}$) uniformly distributed samples for all uncertain variables u_i^* as a base distribution.

2. Run Kernel density estimation for identifying the probabilities $\hat{f}_s(u_i^*)$.
   (a) Set $s = 1$.
      (i) Set $i = 1$.
      (ii) While $i < N_{\text{samp}}$, calculate $\hat{f}_s(u_i^*)$ using equation (10).
      (iii) $i = i + 1$. Go to step (ii).
   (b) $s = s + 1$. If $s < S + 1$ return to step 2(a,i).
3. Run the model for each sample point to find the corresponding model output, store value Z_i.

II. SNLP optimization

1. Set $k = 1$. Determine objective function value for starting point, $J = P(\theta^k, u^k, \theta)$. Set deterministic decision variable counter $d = 1$.
   (a) Generate ($i = 1, \ldots, N_{\text{samp}}$) samples with the appropriate distributions at θ_d^k for all uncertain variables u_i^k.
   (b) Run Kernel density estimation for identifying the probabilities $f_s(u_i^k)$ at θ_d^k, similar to step I.2, using equation (11) in step (ii) instead.
   (c) Determine the weights ω_i from the product of ratios, $\prod_s f_s(u_i^k)/\hat{f}_s(u_i^*)$.
   (d) Calculate $\sum_i \omega_i$.
   (e) Estimate objective function value:
      (i) Set $i = 1$, $J^k = 0$.
      (ii) While $i < N_{\text{samp}}$, calculate: $J^k = J^k \cdot \omega_i / \sum_j \omega_j$.
      (iii) $i = i + 1$. Go to step (ii).
   (f) Set $d = d + 1$, return to step II.2.
2. While $d \leqslant D$, perturb one decision variable θ_d^k to find $\theta_d^{k,\Delta}$. Reset deterministic decision variable counter $d = 1$.
   (a) Generate ($i = 1, \ldots, N_{\text{samp}}$) samples with the appropriate distributions at $x_d^{k,\Delta}$ for all uncertain variables u_i^k.
   (b) Run Kernel density estimation for identifying the probabilities $f_s(u_i^k)$ at $x_d^{k,\Delta}$, similar to step I.2, using equation (11) in step (ii) instead.
   (c) Determine the weights ω_i from the product of ratios, $\prod_s f_s(u_i^k)/\hat{f}_s(u_i^*)$.
   (d) Calculate $\sum_i \omega_i$.
   (e) Estimate objective function value:
      (i) Set $i = 1$, $J^{k,\Delta} = 0$.

(ii) While $i < N_{\text{samp}}$, calculate: $J^{k,\Delta} = J^{k,\Delta} \cdot \omega_i / \sum_j \omega_j$.

(iii) $i = i + 1$. Go to step (ii).

(f) Set $d = d + 1$, return to step II.2.

3. Calculate gradient information obtained from II.1 and II.3.

4. Check convergence criteria for NLP solver (KKT conditions), if satisfied, STOP – optimum found. Otherwise, identify new vector of decision variables through gradients obtained from objective function value estimation via reweighting. Set $k = k + 1$. Return to step II.2.

Note that traditional techniques rely on repeated model runs for steps II.3(b) in the algorithm. For computationally complex nonlinear models, this task can become the critical bottleneck for solving the SNLP. BONUS, on the other hand, bypasses these by estimating the objective function values via reweighting. The BONUS algorithm is implemented using the NLP solver based on Sequential Quadratic Programming (SQP) method.[1] In the following sections we introduce two case studies to illustrate the algorithm.

4. Case study: offline quality control for parameter design of a CSTR

This study is based on Taguchi's approach to off-line quality (Taguchi, 1986) control to determine optimal operating conditions of a chemical reactor. The system investigated consists of a first-order sequential reaction, $A \rightarrow B \rightarrow C$, taking place in a nonisothermal continuous stirred tank reactor (CSTR). The process and the associated variables are illustrated in figure 7. We are interested in designing and operating this process such that the rate of production of species B (R_B) is 60 moles/minute. However, as is apparent from the reaction pathway, species B degrades to species C if the conditions in the CSTR such as the temperature T and heat removal Q are conducive. The objective of parameter design is to produce species B at target levels with minimal fluctuations around the target in spite of continuous variation in the inputs. The inlet concentration of A (C_{A_f}), the inlet temperature T_f, the volumetric flow rate F, and the reactor temperature T are considered prone to continuous variations. The objective of offline parameter design is to choose parameter settings for the design variables such that the variation in the production rate of r_B around the set point is kept at a minimum.

The five design equations that govern the production of species B (and the steady state values of other variables) in the CSTR are given below. The average residence time τ of each species in the reactor is given $\tau = V/F$, where V is the reactor volume and F is the feed flow rate.

$$Q = F\rho Cp(T - T_f) + V(r_A H_{RA} + r_B H_{RB}), \tag{13}$$

[1] In future, this code for BONUS will be made available through Sandia National Laboratories and CUSTOM.

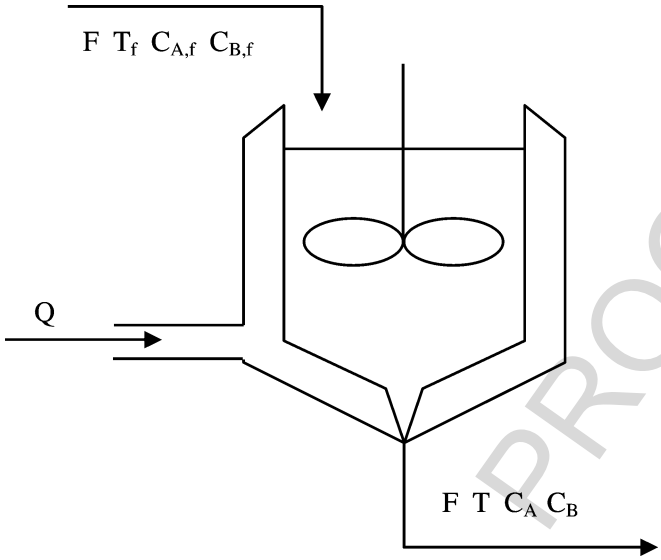

Figure 7. Non-isothermal CSTR.

$$C_A = \frac{C_{A_f}}{1 + k_A^0 e^{-E_A/(RT)}\tau}, \tag{14}$$

$$C_B = \frac{C_{B_f} + k_A^0 e^{-E_A/(RT)}\tau C_A}{1 + K_B^0 e^{-E_B/(RT)}\tau}, \tag{15}$$

$$-r_A = k_A^0 e^{-E_A/(RT)} C_A, \tag{16}$$

$$-r_B = k_B^0 e^{-E_B/(RT)} C_B - k_A^0 e^{-E_A/(RT)} C_A, \tag{17}$$

where C_A and C_B are the bulk concentrations of A and B, T is the bulk temperature of the material in the CSTR, subscript f denotes initial feed, and the rate of consumption of A and B are given by $-r_A$ and $-r_B$. These five variables are the state variables of the CSTR and can be estimated for a given set of values for the input variables (C_{A_f}, C_{B_f}, T_f, T, F, and V) and the following physical constants: k_A^0, k_B^0 and E_A, E_B the pre-exponential Arrhenius constants and activation energies, respectively; H_{RA} and H_{RB}, the molar heats of the reactions which are assumed to be independent of temperature; ρ and Cp the density and specific heats of the system which are assumed to be same for all processing streams. Once input variables T and T_f are specified, equation (13) can be numerically solved to estimate Q, the heat added to or removed from the CSTR. The average residence time can be calculated from the input variables F and V. Subsequently, for a given input concentrations for C_{A_f} and C_{B_f}, the bulk CSTR concentrations C_A and C_B can estimated using equations (14), (15). The production rates r_A and r_B can now be calculated from equations (16), (17). The system parameters are summarized in table 4. Note that this analysis fixes the set-point for both the feed concentration of B, C_{B_f}, and the CSTR temperature T. Both values are given in table 4 as well.

Table 4
Parameters and their values in CSTR study.

Parameter	Value	Units
k_A^0	$8.4 \cdot 10^5$	min^{-1}
k_B^0	$7.6 \cdot 10^4$	min^{-1}
H_{RA}	$-2.12 \cdot 10^4$	J/mol
H_{RB}	$-6.36 \cdot 10^4$	J/mol
E_A	$3.64 \cdot 10^4$	J/mol
E_B	$3.46 \cdot 10^4$	J/mol
R	8.314	J/(mol K)
C_p	$3.2 \cdot 10^3$	J/(kg K)
ρ	1, 180	kg/m^3
C_{B_f}	328	mol/m^3
T	314	K

Table 5
Decision variables for optimization.

	Lower bound	Upper bound	Initial value	Optimal value
C_{A_f}	3000 mol/m^3	4000 mol/m^3	3118 mol/m^3	3125.1 mol/m^3
T_f	300 K	350 K	314 K	328.93 K
F	0.01 m^3/min	0.1 m^3/min	0.070 m^3/min	0.057 m^3/min
V	0.02 m^3	0.05 m^3	0.0391 m^3	0.0500 m^3

The design objective is to produce 60 mol/min of component B, i.e., $R_B = 60$. The initial nominal set points for the decision variables are provided in table 5. However, the continuous variations in the variables (C_{A_f}, T_f, F, and T) result in continuous variations of the production rate, R_B. The variations in the input are described using two parameter (mean and variance) normal distributions.

The BONUS algorithm presented earlier will be used for optimization of this nonlinear, uncertain systems for off-line quality control. The goal is to determine process parameters for a nonisothermal CSTR (figure 7) that result in minimum variance in product properties when fluctuations are encountered (Kalagnanam and Diwekar, 1997). The mathematical representation for the problem is given as:

$$\min \sigma_{R_B}^2 = \int_0^1 \left(R_B - \overline{R_B}\right)^2 \mathrm{d}F, \tag{18}$$

$$\text{s.t.} \quad \overline{R_B} = \int_0^1 R_B(\theta, x, u)\, \mathrm{d}F, \tag{19}$$

$$C_A = \frac{C_{A_f}}{1 + k_A^0 \mathrm{e}^{-E_A/RT}\tau}, \tag{20}$$

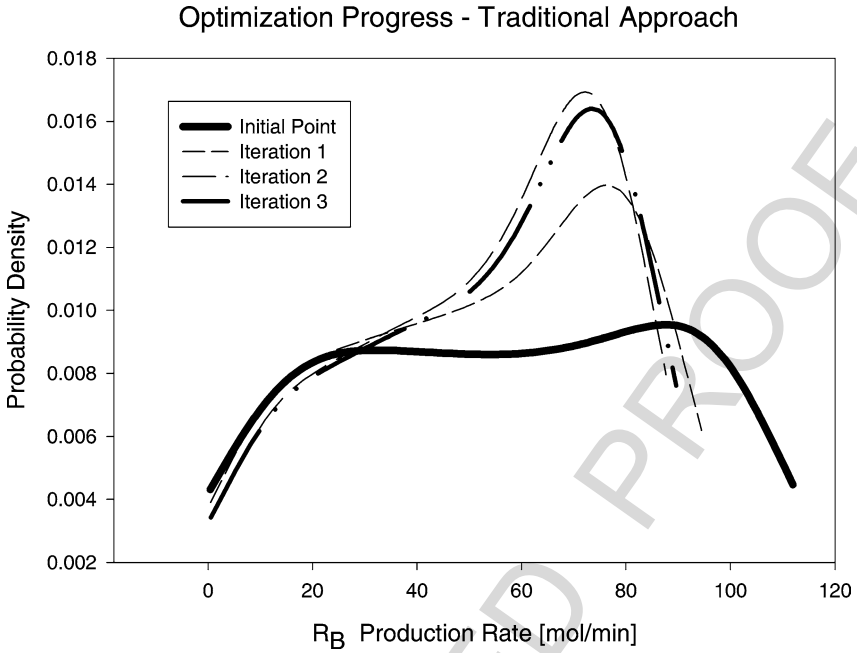

Figure 8. Optimization progress for traditional SNLP approach.

$$C_B = \frac{C_{B_f} + k_A^0 \mathrm{e}^{-E_A/RT}\tau C_A}{1 + k_B^0 \mathrm{e}^{-E_B/RT}\tau}, \tag{21}$$

$$-r_A = k_A^0 \mathrm{e}^{-E_A/RT}, \tag{22}$$

$$-r_B = k_B^0 \mathrm{e}^{-E_B/RT} - k_A^0 \mathrm{e}^{-E_A/RT}, \tag{23}$$

$$Q = F\rho Cp(T - T_f) + V(r_A H_{RA} + r_B H_{RB}), \tag{24}$$

$$\tau = \frac{V}{F}, \tag{25}$$

$$R_B = r_B V. \tag{26}$$

Uncertain variables are $[C_A, T_f, F, T]$, and the range of uncertainty for these variables is normally distributed with means at $[C_{A_f}, T_f^1, F^1, T^1]$. For the first three uncertain variables, the fluctuations 0.001th fractiles are at $\pm 10\%$. However, for T, several factors can contribute to fluctuations and the level of fluctuation around the reactor temperature T is set at $\pm 30\%$ around T^1. Based on these values, the initial variance at the starting point given in table 5 is determined as $\sigma^2_{R_B,\mathrm{init}} = 1034$.

To compare the performance of bypassing the model and using the estimation technique through kernel densities, the model was run first for the case with model evaluation only. Using this traditional approach, the algorithm converged to the optimal solution of $[C_{A_f} = 3124.7\ \mathrm{mol/m^3}, T_f^1 = 350\ \mathrm{K}, F^1 = 0.0557\ \mathrm{m^3/min}, V = 0.0500\ \mathrm{m^3}]$ after 3 iterations, for a sample size $N_{\mathrm{samp}} = 150$. This reactor design has a variance of $\sigma^2_{R_B} = 608.16$. Here, the model is run for every sample point during each iteration step.

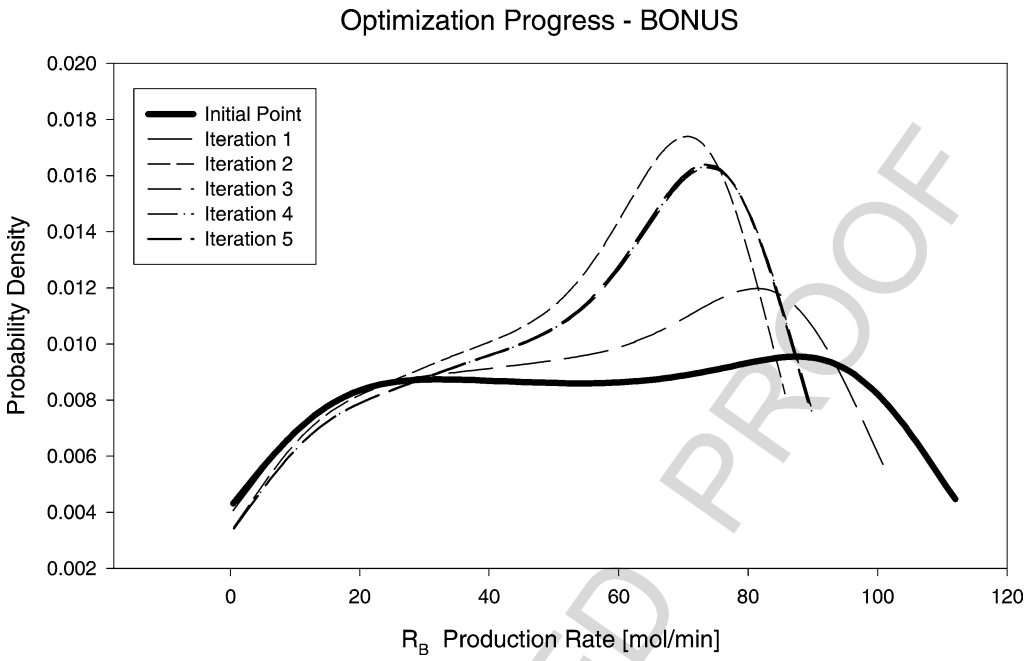

Figure 9. Optimization progress in reducing product variance through BONUS.

Further, the derivatives used for SQP are estimated by running the model an additional 4 times for shifted sample sets of each variable. This requires a total of

$$150\frac{\text{model calls}}{\text{derivative calc.}} \cdot (4+1)\frac{\text{derivative calc.}}{\text{iterations}} \cdot 3 \text{ iteration} = 2250 \text{ model calls.}$$

Optimization progress is presented in figure 8 for the traditional approach. The initial point is shown as the thick line covering variations up to 120 mol/min. As optimization progresses, the probability around the desired rate of $R_B = 60$ increases, as seen in the optimal solution presented as the bold dashed/dotted line.

The analysis for the BONUS algorithm using model bypass converges after 5 iterations to the optimum given in table 5. This solution shows almost identical behavior to the optimum found using the traditional approach and even has a slightly lower variance of $\sigma^2_{R_B} = 607.11$. However, the real advantage of using BONUS is that this analysis called the model just 150 times, only for the determination of the initial base distribution $F[R^*_B]$, in contrast to a total of 2250 model evaluations for the traditional approach.

5. Case study: optimal capacity expansion for electricity utilities in uncertain markets

Capacity expansion for electricity utilities has been an active area of research, having been analyzed using a multitude of methods, including optimization, simulation, and decision analysis (Ku, 1995). The nature of the problem is inherently uncertain, as it is

impossible to determine exact values for future cost levels, the demand for electricity, the development of alternative, more efficient technologies, and many more factors. Hence, the capacity planning example has been analyzed by various researchers in the stochastic programming (SP) community (Birge and Louveaux, 1997).

Due to the limitations of conventional algorithms for optimization under uncertainty, several assumptions have been made, converting the capacity expansion SP into a linear problem through estimations and approximations in order to solve these problems. Among these simplifications, the load curve, which identifies the probability of electricity demand levels, is generally discretized into linear sections, allowing the use of decomposition techniques that require a finite number of realizations of the uncertain variables (Louveaux and Smeers, 1988).

The ability of BONUS to handle nonlinear formulations will be exploited in the analysis presented below. Further, the model presented here will account for the possibility of capacity shortcomings and possibilities of supplemental acquisition of electricity from external sources to satisfy demand, approximating the behavior of free market electricity distribution.

The mathematical representation for the problem is given below. The objective is to minimize the expected cost of capacity expansion subject to uncertain demand and cost factors, while ensuring that no shortages are present. Note that the objective function (27) is the expected value for total cost calculated for $n = 1, \ldots, N_{\text{samp}}$ samples. In the formulation given below, capital nomenclature is used for decision variables, while the uncertain variables are indicated through a tilde symbol:

$$\min E[cost], \tag{27}$$

$$\text{s.t.} \quad cost = \sum_t cost_t^{\text{op}} + cost_t^{\text{cap}} + cost_t^{\text{buy}}, \tag{28}$$

$$cost_t^{\text{op}} = \sum_i P_t^i \widetilde{oc}_t^i, \tag{29}$$

$$cost_t^{\text{cap}} = \sum_i \alpha^i \left(AC_t^i\right)^{\beta^i}, \tag{30}$$

$$cost_t^{\text{buy}} = \widetilde{\kappa}_t \left(\widetilde{d}_t - tp_t\right)^{\gamma}, \tag{31}$$

$$c_t^i = c_{t-1}^i + AC_t^i, \tag{32}$$

$$tp_t = \sum_i P_t^i, \tag{33}$$

$$P_t^i \leqslant c_t^i, \tag{34}$$

$$i \in Technology_1, Technology_2, \ldots, Technology_I, \tag{35}$$

$$t \in Period_1, Period_2, \ldots, Period_T. \tag{36}$$

Equation (28) sums up the respective costs for operation, capacity expansion, and the option to purchase electricity for meeting demand in case the total available capacity is below demand. The operating costs are calculated using equation (29), where

Table 6
Uncertain variables in capacity expansion case.

Parameter	Lower bound	Upper bound
Demand Period I	400 MWh	500 MWh
Demand growth rate Period II	0.75	1.50
Technology I generated cost increase for P-II	0.95	1.12
Technology II generated cost for Period I	0.17 k$/MWh	0.37 k$/MWh
Technology II generated cost for Period II	0.17 k$/MWh	0.50 k$/MWh

oc_t^i is a cost parameter for electricity generation of technology i in time period t, and P_t^i are decision variables determining how much electricity should be produced using technology/power-plant i at time t.

Equation (30) determines the cost of capacity expansion. Traditional models use a linear relationship between the cost of expansion C_{cap}^t and the added capacity ac_t^i. We have used data and models from Integrated Environmental Control Model (ICEM) (Rubin et al., 1997), a computational tool developed for the Department of Energy. This resulted in the power law model for more accurate cost estimation (equation (30)). In this formula, α_i is a proportionality factor for capacity expansion, while β_i provides the exponential factor that allows capital expansion cost to follow economies of scale.

Another nonlinear expression, equation (31), is used to determine the cost of electricity purchased, C_{buy}, when demand D_t exceeds capacity. The power factor γ must be greater than 1 to ensure that relying on external sources is not used as the sole option when increases in demand are expected. This is accomplished as C_{buy} increases exponentially when capacity is significantly below possible demand levels. The primary goal of this approach is to account for the common market practice of purchasing electricity in a deregulated environment when demand reaches peak levels, surpassing available capacity in a given location.

Finally, equation (32) is used to calculate the available capacity at each time step following expansion, equation (33) calculates the total electricty produced, tp_t and equation (34) ensures that no power plant can produce more energy than its installed capacity.

The above model is applied to a case study using two alternative reactor types for a 2-stage capacity expansion problem. Using the ICEM model, Technology I is selected as a Cyclone type coal power plant, while Technology II is a Tangential plant.

Optimization is performed using BONUS, while the uncertainty is captured through HSS with a sample size of 100, including 5 uncertain variables (table 6) and 8 decision variables that determine capacity expansion and electricity generation for each technology at each time step.

Here, demand growth rate for Period II implies that the total demand in Period I is multiplied by a normally distributed uncertain factor changing between 0.75 and 1.50, while the unit cost of electricity generated through Technology I can change between -5% or $+12\%$ for the second period. Table 7 provides the constants and initial values used for this case study.

Table 7
Constants for capacity expansion case.

Parameter	Value
Initial Capacity Technology I	250 MW
Initial Capacity Technology II	150 MW
β^1	1.25 k\$/MW$^{0.7472}$
β^2	0.95 k\$/MW$^{0.7856}$
α^1	0.7472
α^2	0.7856
γ	1.75
oc_1^1	0.25 k\$/MWh

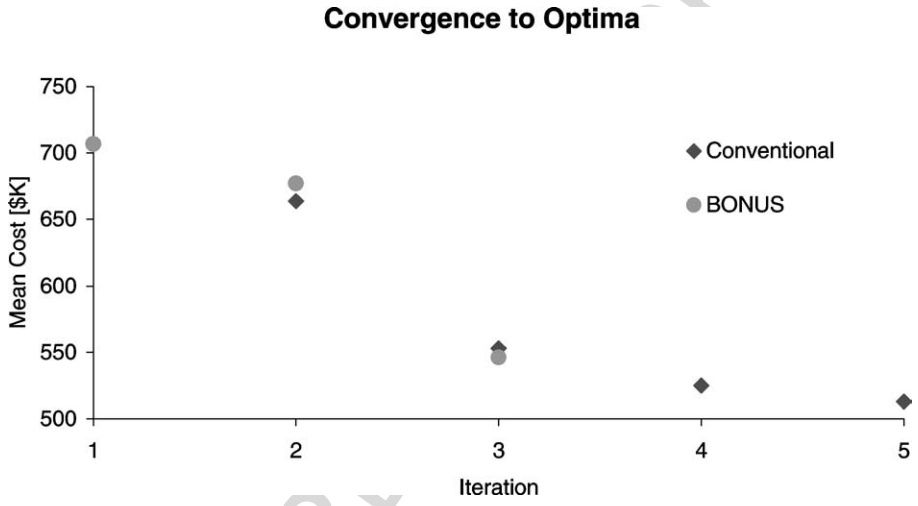

Figure 10. Comparison of optimization progress.

Finally, the pre-exponential factor for the cost of purchasing electricity, κ_t is determined as the greater value between the two per unit electricity generation costs for the different technologies, oc_t^1 and oc_t^2.

Starting from a system with initial annualized cost of the capacity expansion at \$760.9K, the system is optimized both via BONUS, as well as exhaustive model runs for derivative estimation through objective function value calculation. The conventional approach converges after 5 iterations, requiring a total of

$$100\frac{\text{model calls}}{\text{derivative calc.}} \cdot (8+1)\frac{\text{derivative calc.}}{\text{iterations}} \cdot 5 \text{ iteration} = 4{,}500 \text{ model calls}$$

compared to only 100 models and just 3 iterations runs for the BONUS algorithm (figure 10). Table 8 presents the decision variables and their optimal values found by

Table 8
Decision variables in capacity expansion case.

Variable	Initial value (MW)	Optimal value (MW)
Capacity Addition Technology I Period I	100	93
Capacity Addition Technology I Period II	200	197
Capacity Addition Technology II Period I	100	154
Capacity Addition Technology II Period II	200	197
Electricity Production Technology I Period I	250 h	257 h
Electricity Production Technology I Period II	250 h	247 h
Electricity Production Technology II Period I	250 h	291 h
Electricity Production Technology II Period II	250 h	330 h

BONUS.

$$\epsilon_{\text{BONUS}} = \frac{(\%\ \text{Mean reduction})_{\text{BONUS}}}{(\%\ \text{Mean reduction})_{\text{Modelruns}}} = 0.867.$$

6. Conclusions

In this paper, we have introduced BONUS, a reweighting approach for estimating derivative information needed during optimization of nonlinear stochastic problems. The technique relies on kernel density estimation of a base distribution and the sample space encountered during optimization. Two real world case studies: (1) off-line quality control problem from chemical engineering, and (2) the electricity expansion problem from operations research literature illustrate, efficiency of the technique to determine derivatives, and hence the search directions during optimization loop. Further, by selecting efficient sampling techniques like HSS allows for significant computational improvement, as the repetitive nature of model evaluations is avoided by using kernel density estimation.

Acknowledgments

The authors would like to acknowledge funding from Sandia National Laboratories for this research project.

References

Birge, J.R. and F. Louveaux. (1997). *Introduction to Stochastic Programming*. New York: Springer.
Charnes, A. and W.W. Cooper. (1959). "Chance-Constrained Programming." *Management Science* 5, 73.
Dantzig, G.B. and G. Infanger. (1991). "Large Scale Stochastic Linear Programs – Importance Sampling and Bender Decomposition." In Brezinski and U. Kulisch (eds.), *Computational and Applied Mathematics*, p. 111.
Davidon, W.C. (1991). "Variable Metric Method for Minimization." *SIAM Journal of Optimization* 1, 1.

Dempster, M.A.H. (1988). "On Stochastic Programming II: Dynamic Problems under Risk." *Stochastics* 25, 15.

Hesterberg, T. (1995). "Weighted Average Importance Sampling Defensive Mixture Distributions." *Technometrics* 37, 185.

Hesterberg, T. (1996). "Estimate and Confidence Intervals for Importance Sampling Sensitivity Analysis." *Mathematical and Computer Modeling* 23(8–9), 79–86.

Higle, J.L. and S. Sen. (1991). "Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse." *Mathematics of Operations Research* 16, 650.

Kalagnanam, J.R. and U.M. Diwekar. (1997). "An Efficient Sampling Technique for Off-Line Quality Control." *Technometrics* 39, 308.

Ku, A. (1995). "Modelling Uncertainty in Electricity Capacity Planning." Ph.D. Thesis, University of London, London Business School, 445 p.

Louveaux, F. and Y. Smeers. (1988). "Optimal Investments for Electricity Generation: A Stochastic Model and Test-Problem." In *Numerical Techniques for Stochastic Optimization.* Springer Series in Computational Mathematics, Vol. 445, pp. 33–64. Berlin: Springer.

Rockafellar, R.T. and R.J.-B. Wets. (1976). "Stochastic Convex Programming: Basic Duality." *Pacific Journal of Mathematics* 63, 173.

Rockafellar, R.T. and R.J.-B. Wets. (1991). "Scenarios and Policy Aggregation in Optimization under Uncertainty." *Mathematics of Operations Research* 16, 119.

Rubin, E.S., J.R. Kalagnanam, H.C. Frey and M.B. Berkenpas. (1997). "Integrated Environmental Control Modeling (IECM) of Coal-Fired Power Systems." *Journal of Air and Waste Management Association* 47, 1180.

Ruszczynski, A. (1986). "A Regularized Decompotision for Minimizing a Sum of Polyhedral Functions." *Mathematical Programming* 35, 309.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Boca Raton: Chapman & Hall (CRC Reprint, 1998).

Taguchi G. (1986). *Introduction to Quality Engineering*. Tokyo: Asian Productivity Center.