

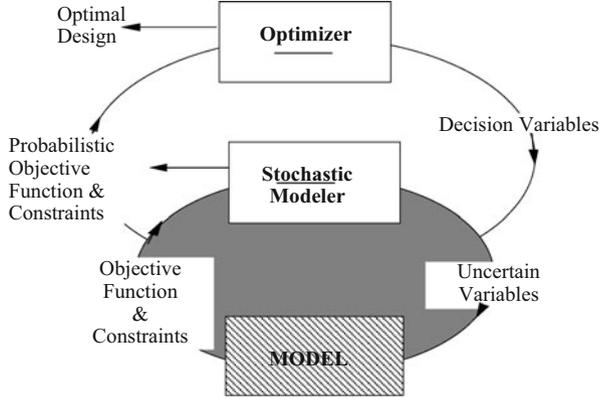
Chapter 4

The BONUS Algorithm

In this chapter we describe the basics of the Better Optimization of Nonlinear Uncertain System (BONUS) algorithm. For better readability, we present the generalized stochastic optimization framework (Fig. 1.4 (from Chap. 1) for stochastic nonlinear programming (NLP) problem below. This chapter is derived from the work by [43].

General techniques for these types of optimization problems determine a statistical representation of the objective such as maximum expected value or minimum variance. Once embedded in an optimization framework, the iterative loop structure emerges where decision variables are determined, a sample set based on these decision variables is generated, the model is evaluated for each of these sample points, and the probabilistic objective function value and constraints are evaluated, as shown in the inner loop of the Fig. 4.1. When one considers that nonlinear optimization techniques rely on an objective function and constraints evaluation for each iteration, along with derivative estimation through perturbation analysis, the sheer number of model evaluations rises significantly rendering this approach ineffective for even moderately complex models. Figure 4.2 shows the general idea behind the BONUS algorithm. BONUS follows the grey arrows. In the stochastic optimization iterations (Fig. 4.1), decision variables values can vary between upper and lower bounds, and in sampling loop various probability distributions are assigned to uncertain variables. In the BONUS approach, initial uniform distributions (between upper and lower bounds) are assumed for decision variables. These uniform distributions together with specified probability distributions of uncertain variables form the base distributions for analysis. BONUS samples the solution space of the objective function at the beginning of the analysis by using the base distributions. As decision variables change, the underlying distributions for the objective function and constraints change, and the proposed algorithm estimates the objective function and constraints values based on the ratios of the probabilities for the current and the base distributions (a reweighting scheme), which are approximated using kernel density estimation (KDE) techniques. Thus, BONUS avoids sample model runs in subsequent iterations.

Fig. 4.1 Pictorial representation of the stochastic programming framework



4.1 Reweighting Schemes

The goal of the reweighting scheme (shown by gray arrows in Fig. 4.2) is to determine changes in output distributions as input distributions change. Hesterberg (1995) presents various reweighting techniques for estimating the expected value of an output distribution cumulative distribution function (CDF), $F[J(u)]$ without evaluating the model for the input distribution probability density function (PDF), $f(u)$ in Fig. 4.2. The ratio of the probability density functions f is used as a weight, which is given as:

$$\omega_i = \frac{f(u_i)}{\hat{f}(u_i^*)}, \quad (4.1)$$

where $\hat{f}(u_i^*)$ is determined for the 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 $P(u)$ for the output of the model is through the product of the weights and the same properties obtained from the base distribution (Eq. 4.2).

$$P(u) = \sum_i \omega_i \cdot P(u_i^*) \quad (4.2)$$

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:

$$\mu[Z(u)] = \sum_i^{N_{samp}} \omega_i \cdot Z(u_i^*), \quad (4.3)$$

where N_{samp} is the sample size.

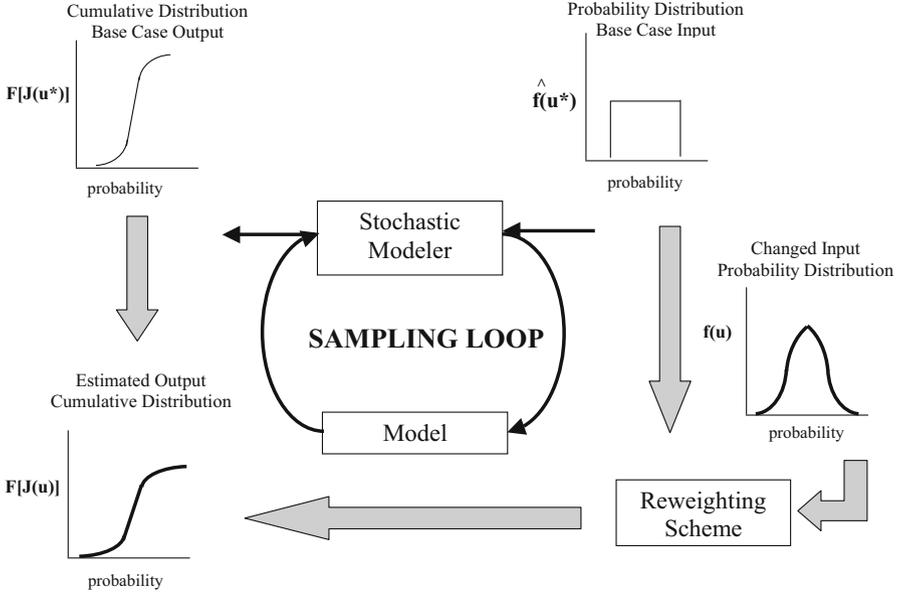


Fig. 4.2 Density estimation approach to optimization under uncertainty

This approach has limitations, as the weights may not sum to 1. This problem is reduced by using normalized weights, as shown in Eq. 4.4. 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) [19]. In the BONUS, instead of using large size of MCS, a more efficient sampling technique as presented in Chap. 2 that provides the same accuracy as MCS in order of magnitude with less number of samples is used.

$$P(u) = \sum_j^{N_{samp}} \frac{\frac{f(u_j)}{\hat{f}(u_j^*)}}{\sum_{i=1}^{N_{samp}} \frac{f(u_i)}{\hat{f}(u_i^*)}} \cdot P(u_j^*) \quad (4.4)$$

As seen in Eq. 4.4 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. Here, the KDE techniques discussed in Chap. 3 are used.

In order to use the kernel density approach for estimating function values (objective function and constraints), the base sample set u^* has to be generated for model calculations. As stated earlier, we select uniform distributions for the decision variables and specified distributions for uncertain variables for creating the base sample. Once the base sample is obtained, its density can be calculated for each point as:

$$\hat{f}(u_i^*) = \frac{1}{N_{samp} \cdot h} \sum_{j=1}^{N_{samp}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{u_i^* - u_j^*}{h} \right)^2} \quad (4.5)$$

We now want to find the distribution $f(u)$ for the decision variable selected at each optimization iteration. For this purpose a small narrow normal distribution at the decision point for the decision variables is assumed and a new sample set for these normal distributions u is generated. After determining the model output $Z(u_i^*)$ for each u_i^* , the value of output distribution for the decision variables $Z(u)$ is obtained by the reweighting scheme described above using the probability of each new data point u_i , as determined through the kernel density approximation (Eq. 4.6).

$$f(u_i) = \frac{1}{N_{samp} \cdot h} \sum_{j=1}^{N_{samp}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{u_i - u_j^*}{h} \right)^2} \quad (4.6)$$

4.2 Effect of Sampling on Reweighting

The proposed reweighting scheme using KDE has been carried out for case studies up to $d = 10$ dimensions for the following five types of functions [43]. The application of alternative and 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 MCS as shown in this section.

- Function 1: Linear additive: $y = \sum_{m=1}^s u_m \quad s = 2 \dots 10$
- Function 2: Multiplicative: $y = \prod_{m=1}^s u_m \quad s = 2 \dots 10$
- Function 3: Quadratic: $y = \sum_{m=1}^s u_m^2 \quad s = 2 \dots 10$
- Function 4: Exponential: $y = \sum_{m=1}^s u_m \cdot \exp(u_m) \quad s = 2 \dots 10$
- Function 5: Logarithmic: $y = \sum_{m=1}^s \log(u_m) \quad s = 2 \dots 10$

The total analysis includes five functions, with four 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 $N_{samp} = [50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10,000]$. 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 4.1.

As required, the base distributions are uniform distributions of decision variables with bounds given in first three columns of Table 4.2, and the estimated distributions were narrow 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:

$$\Delta u_j = 0.05 \cdot \mu\{u_j\} \quad (4.7)$$

As the model functions are relatively simple, the actual values (analytical) of the mean and variance for both sample sets u and u^Δ are calculated, and compared to

Table 4.1 Calculations for KDE efficiency analysis

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

Table 4.2 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 | 5×10^{-3} | 5×10^{-2} |
| u_{10}^* | 6.0 | 9.0 | u_{10} | 8.0 | 9.0 |

the estimates. Further, the same analysis is conducted for the derivative estimates, allowing for comparison of the errors in the estimates based on the sampling technique that is applied to generate both sample sets u^* and u . The next section provides the results of the preliminary study.

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, only one example is provided here that is both relevant to this analysis as well as representative of the overall behavior of the technique.

The results obtained for the nonlinear function, $y = \sum_{m=1}^3 u_m^2$ are presented here. Variance calculation is more prone to errors than calculation of mean (if sample size

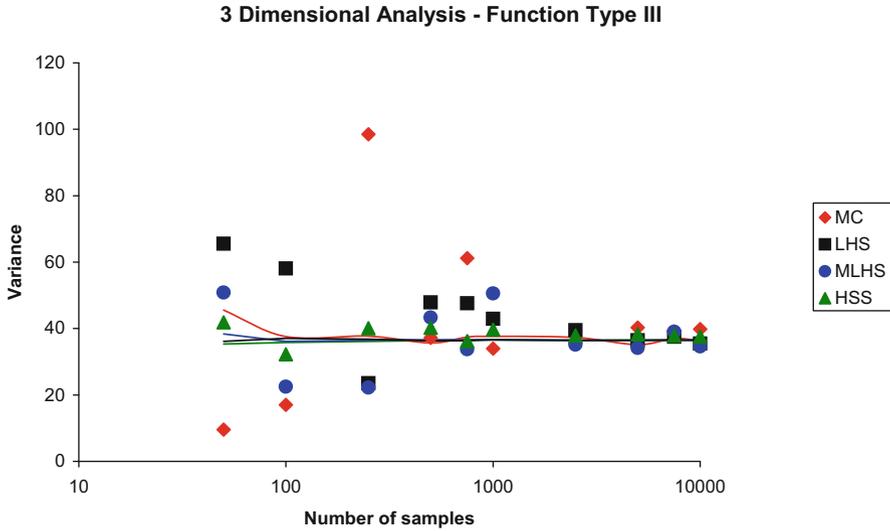


Fig. 4.3 Variance calculation for different sampling techniques

Table 4.3 Percentage error in variance estimation for 3-dimensional analysis using 250 samples

| | MCS | 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 |

is small), and also the case study in the next section aims at calculating the variance of the system at hand that presents the efficiency of the reweighting technique to estimate variance for this function here.

Simultaneous plotting of the actual and estimated values will allow one 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 $N_{samp} = [50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10,000]$. The lines represent the actual values, while the stand-alone points represent the estimated variance values using the four different sampling techniques.

In Fig. 4.3, 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 10,000, with the MCS 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 4.3 provides the percentage error between the estimates and the actual values of the variance for $f(u)$ for all four

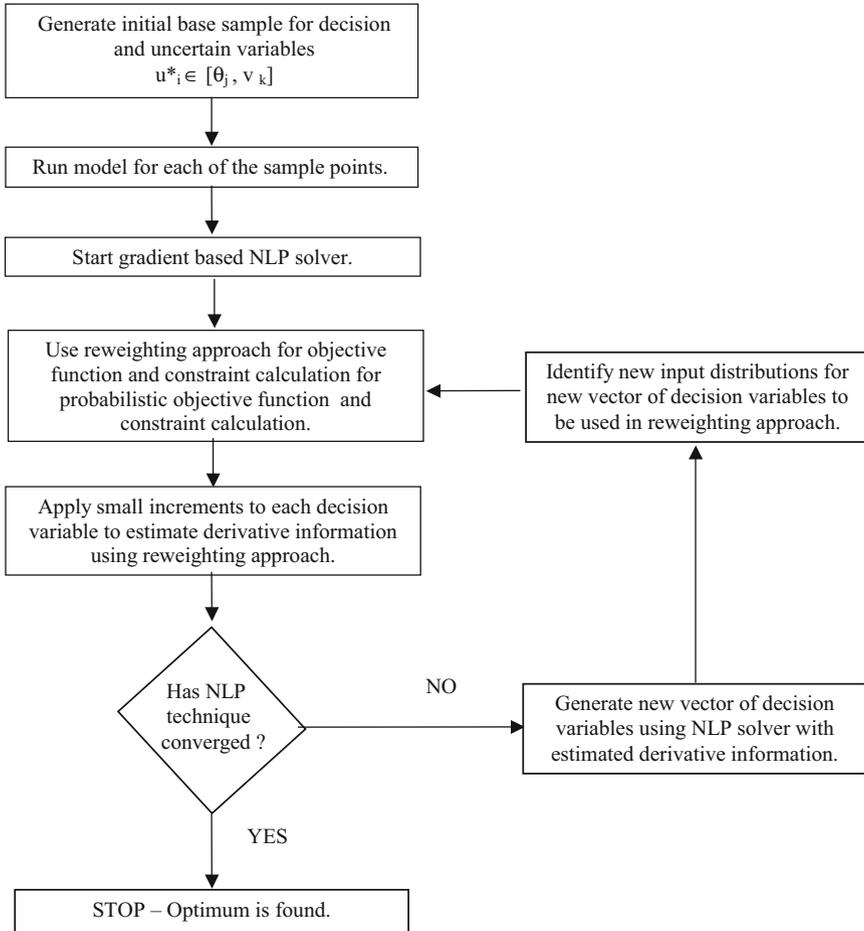


Fig. 4.4 Optimization under uncertainty: The BONUS algorithm

sampling techniques with sample sizes of 250. As seen, HSS yields comparably small percentage errors for all functions.

4.3 BONUS: The Novel SNLP Algorithm

The algorithm for BONUS, given in Fig. 4.4, 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 with respect to

computational time. In this algorithm overview, we denote the D-dimensional vector of deterministic decision variables as $\theta = [\theta_1, \theta_2, \dots, \theta_d, \theta_{d+1}, \dots, x_D]$, while the S-dimensional uncertain variables are defined as $v = [v_1, v_2, \dots, v_s, v_{s+1}, \dots, v_S]$, total S+D-dimensional variable vector $u = [u_1, u_2, \dots, u_{s+d}, u_{s+d+1}, \dots, u_{S+D}]$.

I - Initialization

1. Generate ($i = 1$ to N_{samp}) samples for all decision variables and specified distributions for uncertain variables u_i^* as a base distribution.
2. Run KDE for identifying the probabilities $\hat{f}_s(u_i^*)$.
 - a) Set $s = 1$.
 - i. Set $i = 1$.
 - ii. While $i < N_{samp}$, calculate $\hat{f}_s(u_i^*)$ using Eq. 4.5.
 - iii. $i = i + 1$. Go to step ii.
 - b) $s = s + 1$. If $s < S + D + 1$ return to step I.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, $J = P(\theta^k, v^k)$. Set deterministic decision variable counter $d = 1$.
 - a) Generate ($i = 1$ to N_{samp}) samples (u_i^k) with the appropriate narrow normal distributions at θ_d^k for all decision variables and specified distributions for uncertain variables v_i^k .
 - b) Run KDE for identifying the probabilities $f_s(u_i^k)$ at θ_d^k , similar to step I.2, using Eq. 4.6 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 the probabilistic objective function and constraints values:
 - i. Set $i = 1$, $J^k = 0$.
 - ii. While $i < N_{samp}$, calculate: $J^k = J_i^k * \omega_i / \sum_i \omega_i$.
 - iii. $i = i + 1$. Go to step ii.
 - f) Set $d = d + 1$, return to step II.2.
2. While $d \leq D$, perturb one decision variable θ_d^k to find $\theta_d^{k,\Delta}$. Reset deterministic decision variable counter $d = 1$.
 - a) Generate ($i = 1$ to N_{samp}) samples with the appropriate distributions at $\theta_d^{k,\Delta}$ for all variables u_i^k .
 - b) Run KDE for identifying the probabilities $f_s(u_i^k)$ at $\theta_d^{k,\Delta}$, similar to steps I.2, using Eq. 4.6 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 probabilistic objective function and constraints value:
- i. Set $i = 1$, $J^{k,\Delta} = 0$.
 - ii. While $i < N_{samp}$, calculate: $J^{k,\Delta} = J_i^{k,\Delta} * \omega_i / \sum_i \omega_i$.
 - 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 nonlinear 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-3b 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 nonlinear solver based on sequential quadratic programming (SQP) method. The following examples illustrate the steps involved in BONUS and the efficiency of BONUS for solving SNLP problems.

Example 4.1 Consider the optimization problem presented in Example 3.1 again. Illustrate the reweighting scheme and solve the problem using BONUS.

$$\min E[Z] = E[(\tilde{x}_1 - 7)^2 + (\tilde{x}_2 - 4)^2] \quad (4.8)$$

$$\text{s.t. } \tilde{x}_1 \in N[\mu = x_1^*, \sigma = 0.033 \cdot x_1^*] \quad (4.9)$$

$$\tilde{x}_2 \in U[0.9 \cdot x_2^*, 1.2 \cdot x_2^*] \quad (4.10)$$

$$4 \leq x_1 \leq 10 \quad (4.11)$$

$$0 \leq x_2 \leq 5 \quad (4.12)$$

Here, E represents the expected value, and the goal is to minimize the mean of the objective function calculated for two uncertain decision variables, x_1 and x_2 . The optimizer determines the value x_1^* , which has an underlying normal distribution with $\pm 10\%$ of the nominal value of x_1^* as the upper and lower 0.1% quantiles. Similarly, \tilde{x}_2 is uniformly distributed around x_2^* , with cutoff ranges at $[-10\%, +20\%]$.

Solution The following steps illustrate the steps of BONUS algorithm to solve this problem.

Step 1 The first step in BONUS is determining the base distributions for the decision variables and uncertain variables, followed by generating the output values for this model. Since in this case decision variable and uncertain variables are merged, we use the entire possible range for the two variables as these base distributions have to cover the entire range, including variations. For instance, for x_2 , the range extends to $(0 \times 0.9) \leq x_2 \leq (5 \times 1.2)$ to account for the uniformly distributed uncertainty. Due to space limitations, the illustrative presentation of the kernel density and reweighting

Table 4.4 Base sample

| Sample no. | x_1 | x_2 | Z |
|------------|--------|--------|---------|
| 1 | 5.6091 | 0.3573 | 15.2035 |
| 2 | 3.7217 | 1.9974 | 14.7576 |
| 3 | 6.2927 | 4.2713 | 0.5738 |
| 4 | 7.2671 | 3.3062 | 0.5527 |
| 5 | 4.1182 | 1.3274 | 15.4478 |
| 6 | 7.7831 | 1.5233 | 6.7472 |
| 7 | 6.9578 | 1.1575 | 8.0818 |
| 8 | 5.4475 | 3.6813 | 2.5119 |
| 9 | 8.8302 | 2.9210 | 4.5137 |
| 10 | 6.9428 | 3.7507 | 0.0654 |
| Mean | 6.2970 | 2.4293 | – |
| Std. Dev | 1.5984 | 1.3271 | – |

approach is performed for a sample size of 10, while the remainder of the work uses $N = 100$ samples. A sample realization using MCS is given in Table 4.4.

After this sample is generated, KDE for the base sample is applied to determine the probability of each sample point with respect to the sample set. This is performed for each decision variable separately by approximating each point through a Gaussian kernel, and adding these kernels to generate the probability distribution for each point, as given in Eq. 4.13 [52].

$$\hat{f}(x_i(k)) = \frac{1}{N \cdot h} \sum_{j=1}^N \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{x_i(k) - x_i(j)}{h} \right)^2}. \quad (4.13)$$

Here, h is the width for the Gaussian kernel and depends on the variance σ and sample size N of the data set and is given as follows:

$$h = 1.06 \times \sigma \times N^{-\frac{1}{5}}. \quad (4.14)$$

For our example, $h(x_1) = 1.06 \times 1.5984 \times 10^{-0.2} = 1.0690$ and $h(x_2) = 1.06 \times 1.3271 \times 10^{-0.2} = 0.8876$. Using the first value, one can calculate $\hat{f}(x_1(1)) = \frac{1}{10 \times 1.0690} \sum_{j=1}^{10} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{5.6091 - x_1(j)}{1.0690} \right)^2} = 0.1769$. This step is repeated for every point, resulting in the KDE provided in Table 4.5.

Step 2 All these steps were preparations for the optimization algorithm, where repeated calculations of the objective function will be bypassed through the reweighting scheme.

Step 2a For the first iteration, assume that the initial value for the decision variables is $x_1 = 5$ and $x_2 = 5$. For these values, another sample set is generated, as shown in Table 4.6, accounting for the uncertainties described in Eqs. 4.9 and 4.10.

Table 4.5 Base sample kernel density estimates

| x_1 | $\hat{f}(x_1)$ | x_2 | $\hat{f}(x_2)$ |
|--------|----------------|--------|----------------|
| 5.6091 | 0.1769 | 0.3573 | 0.1277 |
| 3.7217 | 0.0932 | 1.9974 | 0.2114 |
| 6.2927 | 0.2046 | 4.2713 | 0.1602 |
| 7.2671 | 0.2000 | 3.3062 | 0.2190 |
| 4.1182 | 0.1110 | 1.3274 | 0.2068 |
| 7.7831 | 0.1711 | 1.5233 | 0.2117 |
| 6.9578 | 0.2090 | 1.1575 | 0.1992 |
| 5.4475 | 0.1691 | 3.6813 | 0.2100 |
| 8.8302 | 0.0920 | 2.9210 | 0.2152 |
| 6.9428 | 0.2092 | 3.7507 | 0.2063 |

Table 4.6 Sample-optimization iteration 1

| Sample no. | \tilde{x}_1 | \tilde{x}_2 |
|------------|---------------|---------------|
| 1 | 4.7790 | 5.7625 |
| 2 | 4.9029 | 5.5740 |
| 3 | 5.0347 | 5.9199 |
| 4 | 4.9686 | 5.8697 |
| 5 | 4.9001 | 5.9967 |
| 6 | 4.9819 | 5.1281 |
| 7 | 5.0316 | 5.4877 |
| 8 | 5.0403 | 5.4841 |
| 9 | 4.9447 | 5.7557 |
| 10 | 5.0344 | 4.7531 |
| Mean | 4.9618 | 5.5731 |
| Std. Dev | 0.0836 | 0.3862 |

The expected value of Z is estimated using the reweighting approach, given in Steps 2b and 2c.

Step 2b Now, the KDE for the sample ($f(x_i)$) generated around the decision variables has to be calculated. The Gaussian kernel width $h(\tilde{x}_1) = 1.06 \times 0.0837 \times 10^{-0.2} = 5.598 \times 10^{-2}$. Using this value, one can calculate $f(x_1(1)) = \frac{1}{10 \times 5.598 \times 10^{-2}} \sum_{j=1}^{10} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{5.609 - \tilde{x}_1(j)}{5.598 \times 10^{-2}} \right)^2} = 5.125 \times 10^{-23}$. Again, this step is repeated for every point of the sample with respect to the base distribution data resulting in the KDE provided in Table 4.7.

Step 2c Using these and the base KDE values, weights are calculated for each sample point j as

$$\omega_j = \frac{f(x_1(j))}{\hat{f}(x_1(j))} \times \frac{f(x_2(j))}{\hat{f}(x_2(j))}, \quad j = 1, \dots, N \quad (4.15)$$

Table 4.7 Optimization iteration 1-KDE

| No | x_1 | $f(x_1)$ | x_2 | $f(x_2)$ |
|----|--------|-------------------------|--------|-------------------------|
| 1 | 5.6091 | 5.125×10^{-23} | 0.3573 | 0 |
| 2 | 3.7217 | 0 | 1.9974 | 2.989×10^{-26} |
| 3 | 6.2927 | 0 | 4.2713 | 2.777×10^{-2} |
| 4 | 7.2671 | 0 | 3.3062 | 2.376×10^{-8} |
| 5 | 4.1182 | 3.918×10^{-31} | 1.3274 | 9.958×10^{-40} |
| 6 | 7.7831 | 0 | 1.5233 | 1.745×10^{-35} |
| 7 | 6.9578 | 0 | 1.1575 | 1.303×10^{-43} |
| 8 | 5.4475 | 5.218×10^{-12} | 3.6813 | 2.826×10^{-5} |
| 9 | 8.8302 | 0 | 2.9210 | 1.844×10^{-12} |
| 10 | 6.9428 | 0 | 3.7507 | 8.311×10^{-5} |

Table 4.8 Optimization progress at $N = 100$

| Iteration | x_1 | x_2 | $E^{est}[Z]$ |
|-----------|-------|-------|--------------|
| 0 | 5.000 | 5.000 | 5.958 |
| 1 | 9.610 | 2.353 | 9.238 |
| 2 | 7.065 | 3.814 | 0.258 |

In our illustrative example, the only two nonzero weights are $\omega_5 = 1.699 \times 10^{-68}$ and $\omega_8 = 4.152 \times 10^{-15}$. These weights are normalized and multiplied with the output of the base distribution to estimate the objective function value:

$$E^{est}[Z] = \sum_j^N \overline{\omega}_j \cdot Z(j). \quad (4.16)$$

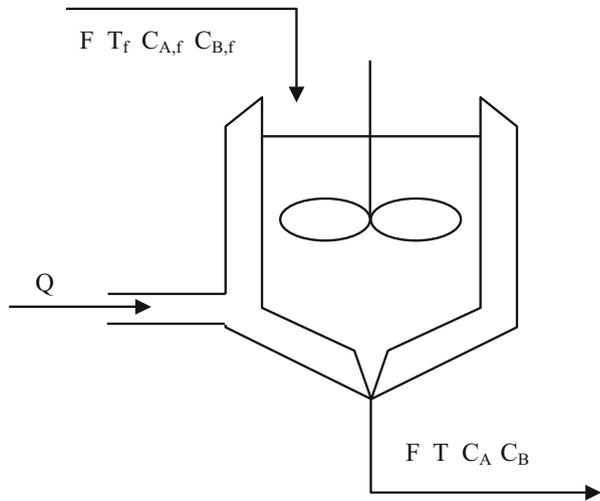
For our illustrative example, this reduces to

$$E^{est}[Z] = \overline{\omega}_8 \cdot Z(8) = 1.0000 \times 2.5119 = 2.5119, \quad (4.17)$$

as the normalization eliminates all but one weight. Note that this illustrative example was developed with an unrealistically small sample size. Hence, the accuracy of the estimation technique cannot be judged from this example. Further, due to the inaccuracy of the estimate resulting from the small sample size, we will not present results for Steps 2d and 2e for just 10 samples, but use 100 samples (note that estimated value of expected value of Z is different in Table 4.8, and is different than that of 10 samples). Also note that Steps 2d and 2e basically repeat the procedures in Steps 2a through 2c for a new sample set around a perturbed point, for instance $x_1 + \Delta x_1 = 5 + 0.001 \times 5 = 5.005$.

The results obtained using the BONUS algorithm for optimization converge to the same optimal solution as obtained using a brute force analysis normally used in stochastic NLPs where the objective is calculated for each iteration by calculating the

Fig. 4.5 Nonisothermal CSTR



objective function value for each generated sample point. In this example, BONUS used only 100 model runs, while the brute force optimization evaluated the model 600 times for the two iterations.

The following example is based on Taguchi's approach to off-line quality control [55] applied to output of a chemical reactor system.

Example 4.2, Taguchi's Quality Control Problem Consider the following problem of off-line quality control of a continuous stirred tank reactor (CSTR) derived from [23].

The system to be investigated consists of a first-order sequential reaction, $A \rightarrow B \rightarrow C$, taking place in a nonisothermal continuous CSTR. The process and the associated variables are illustrated in Fig. 4.5. We are interested in designing and operating this process such that the rate of production of species B (R_B) is 60 moles/min. 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 off-line 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 as $\tau = V/F$, where V is the reactor volume and F is the feed flow rate.

Table 4.9 Parameters and their values in CSTR study

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

$$Q = F\rho C_p(T - T_f) + V(r_A H_{RA} + r_B H_{RB}) \quad (4.18)$$

$$C_A = \frac{C_{A_f}}{1 + k_A^0 e^{-\frac{E_A}{RT}} \tau} \quad (4.19)$$

$$C_B = \frac{C_{B_f} + k_A^0 e^{-\frac{E_A}{RT}} \tau C_A}{1 + k_B^0 e^{-\frac{E_B}{RT}} \tau} \quad (4.20)$$

$$-r_A = k_A^0 e^{-\frac{E_A}{RT}} C_A \quad (4.21)$$

$$-r_B = k_B^0 e^{-\frac{E_B}{RT}} C_B - k_A^0 e^{-\frac{E_A}{RT}} C_A \quad (4.22)$$

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 preexponential 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 C_p 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, Eq. 4.18 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 concentration for C_{A_f} and C_{B_f} , the bulk CSTR concentrations C_A and C_B can be estimated using Eqs. 4.19 and 4.20. The production rates r_A and r_B can now be calculated from Eqs. 4.21 and 4.22. The system parameters are summarized in Table 4.9. 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 also given in Table 4.9.

Table 4.10 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 4.10. 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 , which needs to be minimized. Solve this problem using traditional SNLP and BONUS, and compare the results.

Solution The goal is to determine process parameters for a nonisothermal CSTR (Fig. 4.5) that result in minimum variance in product properties when fluctuations are encountered [23]. The mathematical representation for the problem is given as:

$$\min \sigma_{R_B}^2 = \int_0^1 (R_B - \overline{R_B})^2 dF \quad (4.23)$$

$$\text{s.t. } \overline{R_B} = \int_0^1 R_B(\theta, x, u) dF \quad (4.24)$$

$$C_A = \frac{C_{A_f}}{1 + k_A^0 \cdot e^{-E_A/RT} \cdot \tau} \quad (4.25)$$

$$C_B = \frac{C_{B_f} + k_A^0 \cdot e^{-E_A/RT} \cdot \tau \cdot C_A}{1 + k_B^0 \cdot e^{-E_B/RT} \cdot \tau} \quad (4.26)$$

$$-r_A = k_A^0 \cdot e^{-E_A/RT} \quad (4.27)$$

$$-r_B = k_B^0 \cdot e^{-E_B/RT} - k_A^0 \cdot e^{-E_A/RT} \quad (4.28)$$

$$Q = F\rho C_p \cdot (T - T_f) + V \cdot (r_A H_{RA} + r_B H_{RB}) \quad (4.29)$$

$$\tau = V/F \quad (4.30)$$

$$R_B = r_B \cdot V \quad (4.31)$$

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 4.10 is determined as $\sigma_{R_B,init}^2 = 1034$.

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 traditional SNLP. Using this traditional approach, the algorithm converged to the optimal

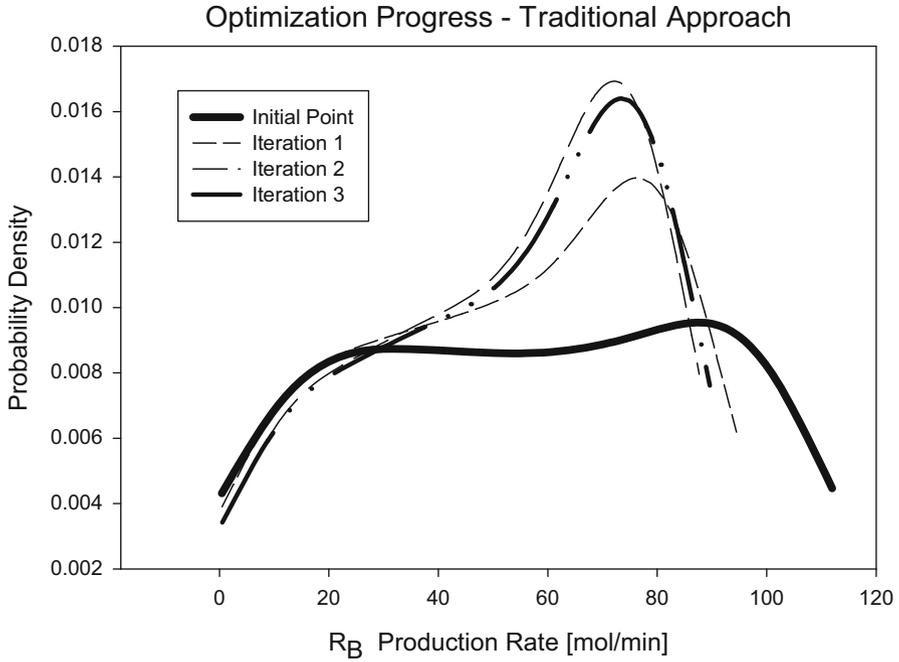


Fig. 4.6 Optimization progress for traditional SNLP approach

solution of $[C_{A_f} = 3124.7 \text{ mol/m}^3, T_f^1 = 350 \text{ K}, F^1 = 0.0557 \text{ m}^3/\text{min}, V = 0.0500 \text{ m}^3]$ after three iterations, for a sample size $N_{\text{samp}} = 150$. This reactor design has a variance of $\sigma_{R_B}^2 = 608.16$. Here, the model is run for every sample point during each iteration step. Further, the derivatives used for SQP are estimated by running the model an additional four 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 Fig. 4.6 for the traditional approach and in Fig. 4.7 for BONUS. 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 five iterations to the same optimum values of decision variables $C_{A_f} = 3125.1 \text{ mol/m}^3$, $T_f = 328.93 \text{ K}$, $F = 0.057 \text{ m}^3/\text{min}$, and $V = 0.0500 \text{ m}^3$. This solution shows almost identical behavior to the optimum found using the traditional approach and even has a slightly lower variance of $\sigma_{R_B}^2 = 607.11$. However, the real advantage of using BONUS is that this analysis called the model just 150 times, only for the

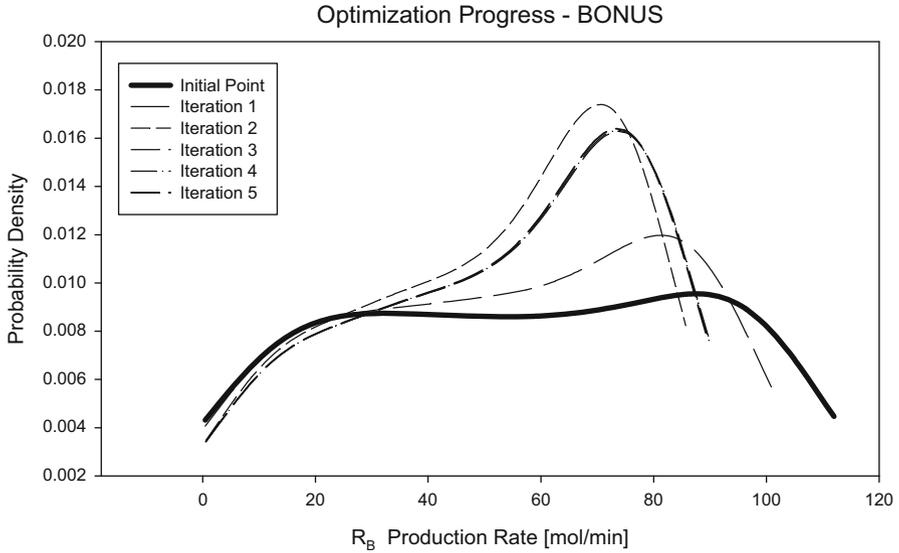


Fig. 4.7 Optimization progress in reducing product variance using BONUS

determination of the initial base distribution $F[R_B^*]$, in contrast to a total of 2250 model evaluations for the traditional approach.

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 [27]. 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 and more efficient technologies, and many more factors. Hence, the capacity planning example has been analyzed by various researchers in the stochastic programming (SP) community [2].

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 [30]. The ability of BONUS to handle nonlinearity allows this problem to be handled without these limitations; this is presented in the following examples.

Example 4.3 Capacity Expansion for Electric Utilities The mathematical representation of the problem is given below. The objective is to minimize the expected cost of capacity expansion subject to uncertain demands and cost factors, while ensuring that no shortages are present. Note that the objective function 4.32 is the expected value for the total cost calculated for $n = 1, \dots, N_{samp}$ samples. In the formulation given below, capital nomenclature is used for decision variables, while the uncertain

variables are indicated through a tilde symbol.

$$\min E[\text{cost}] \quad (4.32)$$

$$\text{s.t. } \text{cost} = \sum_t \text{cost}_t^{\text{op}} + \text{cost}_t^{\text{cap}} + \text{cost}_t^{\text{buy}} \quad (4.33)$$

$$\text{cost}_t^{\text{op}} = \sum_i P_t^i \cdot \widetilde{oc}_t^i \quad (4.34)$$

$$\text{cost}_t^{\text{cap}} = \sum_i \alpha^i \cdot (AC_t^i)^{\beta^i} \quad (4.35)$$

$$\text{cost}_t^{\text{buy}} = \widetilde{\kappa}_t \cdot (\widetilde{d}_t - tp_t)^\gamma \quad (4.36)$$

$$c_t^i = c_{t-1}^i + AC_t^i \quad (4.37)$$

$$tp_t = \sum_i P_t^i \quad (4.38)$$

$$P_t^i \leq c_t^i \quad (4.39)$$

$$i \in \text{Technology}_1, \text{Technology}_2, \dots, \text{Technology}_I \quad (4.40)$$

$$t \in \text{Period}_1, \text{Period}_2, \dots, \text{Period}_T \quad (4.41)$$

Equation 4.33 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 Eq. 4.34, where 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 4.35 determines the cost of capacity expansion. Traditional models use a linear relationship between the cost of expansion $\text{Cost}_t^{\text{cap}}$ and the added capacity ac_t^i . Use the data and models from the Integrated Environmental Control Model (ICEM), a computational tool developed for the Department of Energy. This will provide the power law model for more accurate cost estimation (Eq. 4.35). 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, Eq. 4.36, will be used to determine the cost of electricity purchased, Cost^{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 increase in demand is expected. This is accomplished as Cost^{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, use Eq. 4.37 to calculate the available capacity at each time step following expansion, Eq. 4.38 calculates the total electricity produced, tp_t , and Eq. 4.39 ensures that no power plant can produce more energy than its installed capacity.

Table 4.11 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 |
| Tech. I Gen. cost increase for P-II | 0.95 | 1.12 |
| Tech. II Gen. cost for Period I | 0.17 k\$/MW h | 0.37 k\$/MW h |
| Tech. II Gen. cost for Period II | 0.17 k\$/MW h | 0.50 k\$/MW h |

Table 4.12 Constants for capacity expansion case

| Parameter | Value |
|---------------------------|-------------------------------|
| Initial capacity Tech. I | 250 MW |
| Initial capacity Tech. 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 |

In this problem, Technology I is selected as a Cyclone type coal power plant, while Technology II is a Tangential plant. Again, data for these technologies can be obtained using IECM [41] model.

There are five uncertain variables (Table 4.11) and eight 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 varying between 0.75 and 1.50, while the unit cost of electricity generated through Technology I can vary between -5 and +12% for the second period. Table 4.12 provides the constants and initial values used for this case study.

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

Solve this problem using traditional SNLP and BONUS.

Solution 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 five iterations, requiring a total of

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

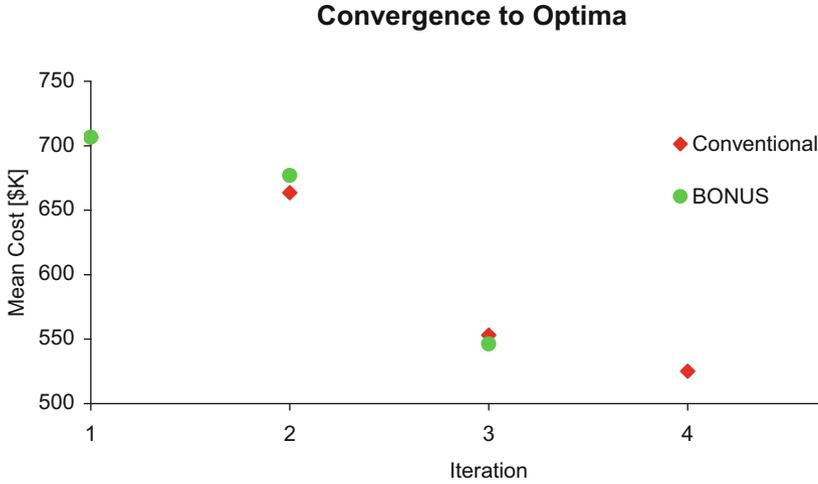


Fig. 4.8 Comparison of optimization progress

Table 4.13 Decision variables in capacity expansion case

| Variable | Initial value (MW) | Optimal Value (MW) |
|--|--------------------|--------------------|
| Capacity addition Tech I Period I | 100 | 93 |
| Capacity addition Tech I Period II | 200 | 197 |
| Capacity addition Tech II Period I | 100 | 154 |
| Capacity addition Tech II Period II | 200 | 197 |
| Electricity production Tech I Period I | 250 h | 257 h |
| Electricity production Tech I Period II | 250 h | 247 h |
| Electricity production Tech II Period I | 250 h | 291 h |
| Electricity production Tech II Period II | 250 h | 330 h |

compared to only 100 models and just three iterations run for the BONUS algorithm (Fig. 4.8). Table 4.13 presents the decision variables and their optimal values found by BONUS.

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

4.4 Summary

In this chapter, we have introduced the BONUS based on the reweighting approach for estimating derivative information needed during optimization of nonlinear stochastic problems. The technique relies on KDE of a base distribution and the sample

space encountered during optimization. Two real world case studies; (1) an off-line quality control problem from chemical engineering, and (2) the electricity expansion problem from operations research literature, illustrates efficiency of the technique in determining derivatives, and hence the search directions during optimization loop. Further, by selection of efficient sampling techniques like HSS allows for significant computational improvement, as the repetitive nature of model evaluations is avoided by using the reweighting scheme. The BONUS algorithm is very useful for solving large-scale real-world problems of significance (e.g., for black-box models) is illustrated in the following three chapters.

Notations

| | |
|----------------|---|
| AC_t^i | capacity of technology i in time period t |
| C_i | concentration of component i in mol/m^3 |
| C_{if} | inlet concentration of component i in mol/m^3 |
| $cost_t^{op}$ | operating cost |
| $cost_t^{cap}$ | capacity expansion cost |
| $cost_t^{buy}$ | cost of buying electricity |
| d_t | electricity demand |
| D | number of decision variables |
| E | expected value function |
| f | probability density function |
| F | volumetric flowrate, m^3/min |
| oc_t^i | a cost parameter for electricity generation of technology i in time period t |
| P | probability function |
| P_t^i | decision variables determining how much electricity should be produced using technology/power plant i at time t |
| Q | heat removal J |
| N_{samp} | number of samples |
| R_B | rate of production of B |
| \bar{R}_B | average rate of production of B |
| S | number of uncertain variables |
| T | temperature, 0K |
| T_f | inlet temperature, 0K |
| u | uncertain and decision variable from input distributions |
| u^* | base uncertain and decision variable from input distributions |
| v | uncertain variable |
| V | reactor volume, m^3 |
| Z | output variable |

Greek letters

| | |
|------------|---|
| α_i | a proportionality factor for capacity expansion |
|------------|---|

| | |
|-----------|---|
| β_i | an exponential factor that allows capital expansion cost to follow economies of scale |
| γ | power factor |
| θ | decision variable |
| μ | mean |
| σ | standard deviation |
| τ | residence time, <i>min</i> |
| ω | weighting function |