

Efficiencies from Spatially-Correlated Uncertainty and Sampling in Continuous-Variable Ordinal Optimization

Vicente J. Romero¹

Sandia National Laboratories²

ABSTRACT

A very general and robust approach to solving continuous-variable optimization problems involving uncertainty in the objective function is through the use of ordinal optimization. At each step in the optimization problem, improvement is based only on a relative ranking of the uncertainty effects on local design alternatives, rather than on precise quantification of the effect. One simply asks "Is that alternative better or worse than this one?"—not "HOW MUCH better or worse is that alternative to this one?" The answer to the latter question requires precise characterization of the uncertainty—with the corresponding sampling/integration expense for precise resolution. By looking at things from an ordinal ranking perspective instead, the trade-off between computational expense and vagueness in the uncertainty characterization can be managed to make cost-effective stepping decisions in the design space. This paper demonstrates correct advancement in a continuous-variable probabilistic optimization problem despite extreme vagueness in the statistical characterization of the design options. It is explained and shown how spatial correlation of uncertainty in such design problems can be exploited to dramatically increase the efficiency of ordinal approaches to optimization under uncertainty.

Keywords: Probabilistic ordinal optimization under uncertainty; Correlated uncertainty & sampling, OOU

1. INTRODUCTION

A very general and robust approach to solving optimization problems is via ordinal optimization, where improvement at each step in the optimization problem is based only on a *relative ranking* of response, rather than on precise quantification of response. It is much less difficult and expensive to determine whether one response is greater/less than another, than to accurately quantify their response values ("...think of holding two

objects to determine which one weighs more, versus trying to estimate their actual weights [by holding them]..." [10]); much more complex and accurate machinery is required to get accurate weights. Correspondingly, the ordinal approach appears to be naturally suited to optimization problems involving nonprobabilistic and semiquantitative descriptions of uncertainty. In fact, it can be seen as a formalism of the ordinal selection process employed by decision makers in the real world when they deliberate over competing options whose outcomes are highly uncertain.

For continuously variable design options (continuous-variable design problems), non-gradient-based local optimizers such as simplex and pattern-search methods [9, 12], and global methods such as genetic algorithms and DIRECT ([13]), can be used to search the design space, using ordinal comparison to select improvement steps. This paper considers continuous-variable (C-V) problems of optimization under uncertainty (OOU) where ordinal comparisons are made of alternatives in the design space that have probabilistic information regarding their uncertainty.³ Fundamental concepts of probabilistic ordinal optimization are presented. In particular, efficiencies gained from exploiting spatial correlation of uncertainty and sampling in the design space are demonstrated on a low-dimensional probabilistic optimization problem. More-sophisticated implementational possibilities are discussed, along with merits compared to nonordinal approaches to probabilistic optimization. References [17], [18], and [20] provide more discussion and application case studies regarding C-V probabilistic ordinal optimization.

¹ Contact: vjromer@sandia.gov, 505-844-5890.

² Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

³ C-V OOU problems are principally different from *stochastic optimization problems* where random stochastic noise in the objective function is a primary feature of the deterministic or OOU optimization problem. Specialized optimization techniques exist for this type of random stochastic uncertainty in the objective function (see, e.g., [22]). In C-V OOU problems the uncertainty in the objective function comes principally from parametric uncertainty about the response or behavior of the system being optimized.

2. ENGINEERING OPTIMIZATION PROBLEM MOTIVATING PROBABILISTIC ORDINAL CONCEPTS

2.1 Deterministic Optimization Problem

The probabilistic optimization problem considered here is an outgrowth of a deterministic optimization problem [21] in which heating conditions were sought that put a candidate weapon subsystem design most at risk. As shown in Figure 1, the design problem is parameterized in terms of two key heating variables: 1) the radius r of a circular region of impinging fire on the top of the safing subsystem; and 2) a coordinate x that moves the center of the “blow torch” on the device’s surface along the cut-plane shown. A three-dimensional (3-D) finite-element conduction/radiation thermal model (described in Ref. [21]) is used to calculate the transient temperature response of the device. Figure 1 shows the calculated temperature of the device (cut along its plane of symmetry) at some point in time for a particular radius r and location x of heating.

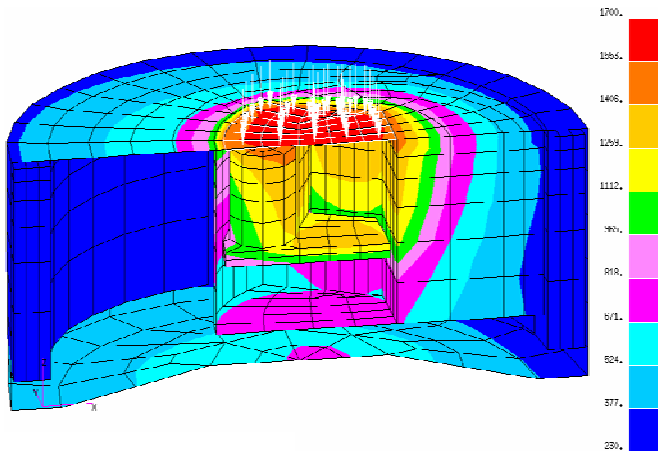


Figure 1: Safing device exposed to circular region of heating.

Figure 2 shows temperature histories for two safety-critical components within the safing device: a ‘strong link’ and a ‘weak link’. The components are termed as such because the weak link must be thermally weaker and fail at a lower temperature than the thermally stronger strong link in order to prevent inadvertent operation of the weapon.

For now focusing just on the *nominal* failure temperatures of the components (nominal $T_{fail_{WL}} = 250^{\circ}\text{C}$ and nominal $T_{fail_{SL}} = 600^{\circ}\text{C}$ —the means of the uncertainty distributions depicted in Fig. 2), an associated ‘safety margin’ of time can be defined:

$$S = t_{fail_{SL}} - t_{fail_{WL}}, \quad (1)$$

where the subscripts SL and WL denote the strong link and weak link. Here $t_{fail_{SL}}$ and $t_{fail_{WL}}$ are the elapsed times (from time zero at the beginning of the thermal

simulation) required for the links to reach their nominal failure temperatures. In Figure 2, the safety margin associated with the nominal failure temperatures of the components is positive and has a value of about $S = +9$ minutes (right-pointing arrow in the figure). A negative or zero safety margin indicates that the strong link is not failing after the weak link as desired, so the safing device is experiencing a vulnerability for that set of heating conditions and component failure temperatures.

The safety margin is a deterministic indicator of the safety of the device. A probabilistic indicator of the safety of the device will later be defined that takes into account the depicted uncertainty in the component failure temperatures.

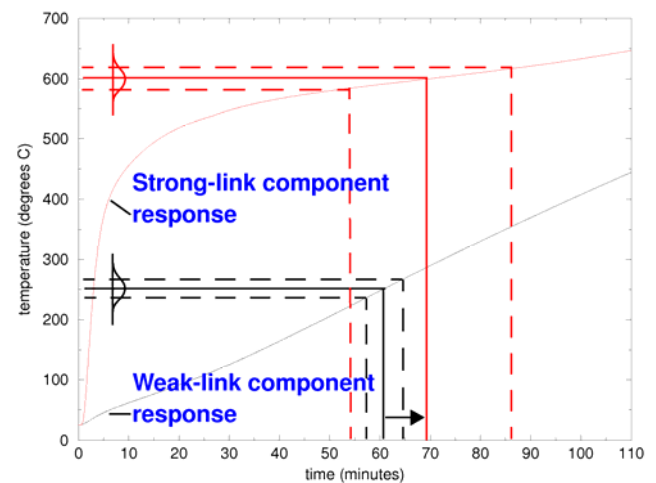


Figure 2: A mean realization of each component’s failure temperature (solid line) reflects through component temperature response curve into a failure time. The difference in component failure times, $t_{fail_{SL}} - t_{fail_{WL}}$, is the corresponding *safety margin*, shown by right-pointing arrow for the positive margin here.

In the optimization problem, the intent is to find the heating conditions that minimize the indicated safety of the device. This minimum corresponds to the heating conditions that most threaten the intended function of the safing device. If the vulnerability is deemed unacceptable, the design can be modified to sufficiently harden the device to these worst-case heating conditions.

In the *deterministic* optimization problem, the optimal values of the variables r and x are sought that minimize the safety margin of the device. As the $\{r, x\}$ variables are changed, the strong-link and weak-link temperature responses change, and the value of the safety margin changes accordingly (assuming fixed nominal failure temperatures for the components). Thus, the deterministic safety-margin objective function $S(r, x)$ is navigated (minimized) in the deterministic optimization problem.

The deterministic optimization problem is complicated by numerical noise resulting from discrete time and space representation in the model (see Refs. [5] and [21]). On

a more global scale, the deterministic optimization problem contains mathematical difficulties associated with navigating to a global minimum at $\{r, x\} = \{1.62, 0.782\}$ on a nonconvex design surface having a fold and several local minima as described in Refs. [5] and [21].

Toward solving the deterministic local optimization problem, Newton-based nonlinear programming (NLP) methods (sequential quadratic programming, BFGS quasi-Newton, and quasi-Newton) from several different research and commercial optimization packages were tried with little success as explained in Ref. [5]. Nonconvexity and noise resulted respectively in ill-conditioning and inaccuracy of the Hessian matrix of finite-differenced second-order derivatives.

To avoid these problems with the second-order information, first-order NLP conjugate-gradient methods were employed in Ref. [5] with more success, but still exhibited some noise-induced nonrobustness. The choice of finite-difference step size for computation of sufficiently accurate gradients proved to be important. The initial starting point also proved to be important because of the noise.

The relatively high degree of model-discretization resolution required for sufficiently smooth computed response for reliable gradient-based navigation of the design space added large cost compared to the more noise-tolerant derivative-free optimization approaches applied in Ref. [6]. There, a simple Coordinate Pattern Search (CPS) method was shown to be successful with lower model resolution/cost requirements than the conjugate-gradient method. The CPS method was both more robust and less costly than the conjugate-gradient method.

More recently, a global-to-local optimizer devised with moving low-order polynomial local response surfaces to be efficient in the presence of small-scale noise was applied to this problem in Ref. [15]. The method performed considerably better than all those previously tried in terms of number of function evaluations required, cost of function evaluations (model discretization level required for accuracy), and robustness to noise level. The method was later recognized to be a type of surrogate-based *trust-region* optimization approach. Successful approaches in this vein had already been introduced elsewhere (e.g., Refs. [1] and [14]). The general advantages of surrogate-based trust-region approaches for handling noisy optimization problems are now well recognized.

2.2 Probabilistic Optimization Problem

The associated probabilistic optimization problem is now considered. A glance at Figure 2 indicates that strong-link and weak-link failure temperatures trace out to relatively flat portions of the temperature response curves. This is more true for the strong link than for the weak link, but regardless, it can be seen that the component failure times are very sensitive to the failure temperature values. Thus, moderate uncertainties in the component

failure temperatures $T_{fail_{WL}}$ and $T_{fail_{SL}}$ can have a substantial impact on the uncertainty of the safety margin. In fact, when reasonable uncertainty bands 5% above and below the nominal failure temperatures are considered in Figure 2, the corresponding bands in failure times indicate that the safety margin could vary from about 28 minutes at best to about -11 minutes at worst. Thus, it is apparent that the effects of uncertainty in the component failure thresholds are very important in this problem.

2.2.1 Safety-margin distribution due to uncertain component failure thresholds. For the purposes here, the strong-link and weak-link failure temperatures $T_{fail_{SL}}$ and $T_{fail_{WL}}$ are each assumed to be described by truncated normal distributions with means μ equal to the respective nominal failure temperatures of 600°C and 250°C, and standard deviations σ equal to 3% of the means, or 18°C and 7.5°C respectively. (The distributions are truncated at 3σ above and below their mean values, and then renormalized to integrate to unity.)

Multiple sets of weak-link and strong-link failure temperatures can be generated by standard Monte Carlo sampling, and then a safety margin can be computed for each set by using the time histories of the weak and strong links obtained from the thermal model run at the $\{r, x\}$ heating conditions. The resulting population of safety-margin realizations will be distributed with some probability density, as exemplified in Figure 3. Associated statistics, such as the mean, the standard deviation, and the probability of attaining a zero or negative safety margin, can be calculated. For example, the mean safety margin denoted in the figure is +4.3 minutes for temperature response curves of the model run at the deterministically worst case set of heating conditions $\{r, x\}_{opt-det} = \{1.62, 0.782\}$.

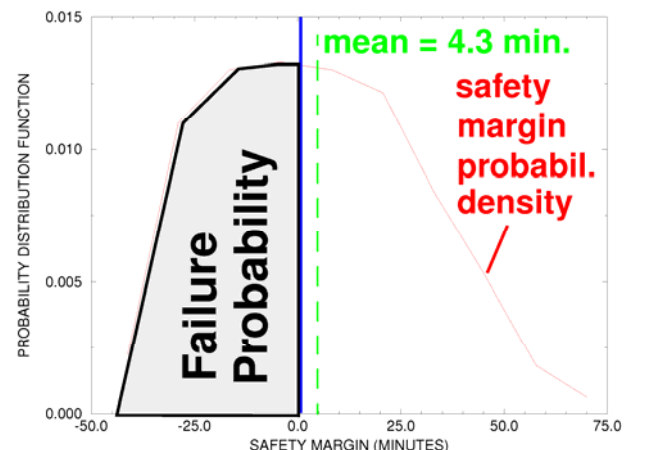


Figure 3: Safety-Margin probability density function at a particular set $\{r, x\}$ of heating conditions.

The deterministic safety margin for this set of heating conditions (and nominal failure temperatures of $T_{fail_{WL}} = 250^\circ\text{C}$ and $T_{fail_{SL}} = 600^\circ\text{C}$) is about 2.5 minutes, or 40% less than the 4.3-minute mean. This problem is

clearly nonlinear in the uncertain parameters $T_{fail_{WL}}$ and $T_{fail_{SL}}$; otherwise, the mean safety margin would equal the deterministic safety margin at the mean values of the uncertain failure temperatures. This nonlinearity is expected to result in different optimal (worst-case) heating conditions $\{r, x\}_{opt}$ if component failure uncertainty is taken into account rather than being ignored as in the deterministic optimization problem.

Figure 4 shows a 3×3 grid of points over a small subset of the design space, centered about the deterministic optima $\{r, x\}_{opt-det} = \{1.62, 0.782\}$. Figure 5 shows magnitude bars for failure probabilities calculated at the nine points of the grid by Monte Carlo sampling over the strong-link and weak-link failure temperature uncertainties. Thus, safety-margin distributions and corresponding failure-probability magnitudes (like in Figure 3) are obtained at each of the 9 design points. The optimization problem now is to *maximize* the *probabilistic* objective function for failure probability in order to determine the worst-case heating values of r and x .

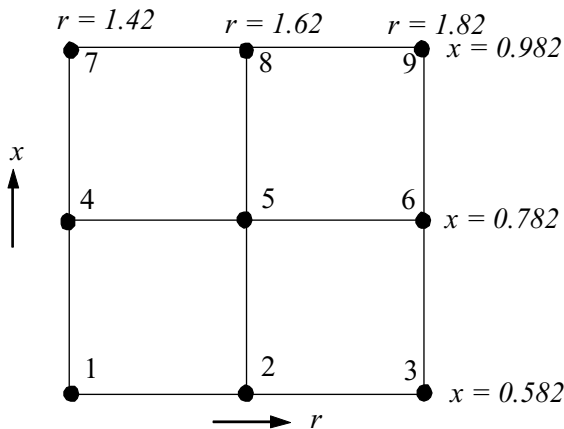


Figure 4: 9-point grid over important subset of 2-D optimization space, centered at point where the deterministic optimum for worst-case heating occurs.

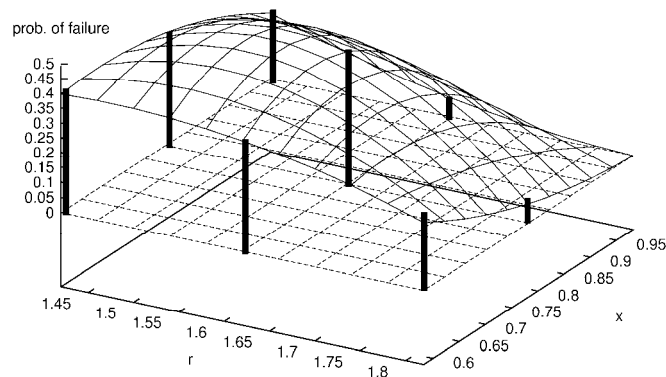


Figure 5: Probabilistic objective function (biquadratic response surface built from Monte Carlo point estimates of failure probabilities at the nine grid points).

2.2.2 Motivation for probabilistic ordinal optimization in this problem. A biquadratic response surface is

shown in Figure 5 that can be seen to not go exactly through the end points of the probability bars because the response surface was created from probabilities obtained from 500 Monte Carlo samples, rather than from the 1000 samples that the bars are based on. The mismatch between bar height and surface height varies somewhat over the design space. The mismatch reflects the fact that MC estimates from a given number of samples have confidence intervals (CI) as depicted in Figure 6 that reflect uncertainty (potential error) in the point estimates of probability.

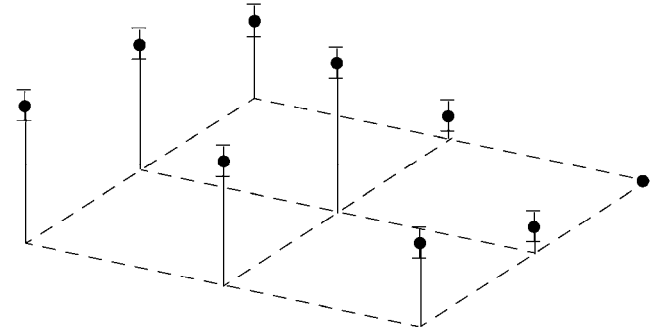


Figure 6: Monte Carlo point estimates of failure probabilities with associated confidence intervals shown.

The following questions then arise: What effect does the number of Monte Carlo samples have on the accuracy with which a probabilistic optimum can be identified in the design space? If one wants to identify the heating variables $\{r, x\}$ that correspond to the highest probability of device failure, how can the impact of Monte Carlo sampling errors be controlled? How many samples need to be taken at each point in the design space to shrink the confidence intervals small enough so that it is unambiguous as to which design point corresponds to the highest failure probability?

3. FUNDAMENTAL CONCEPTS OF PROBABILISTIC ORDINAL OPTIMIZATION

Toward addressing the above questions, it is noted that probabilistic ordinal optimization has been studied for some time in the operations research field (e.g. Ref. [11]), to compare discrete design alternatives (as opposed to continuous design alternatives like in this paper). In particular, the work has addressed apportioning Monte Carlo sampling amongst multiple uncertain or stochastic discrete systems to most efficiently resolve their statistical behavior for the purpose of selecting the best option or several best options. The concepts and procedures apply whether it is desired to: A) minimize the total number of samples required to reach a desired probability of correct selection of the best or several best options; or B) maximize the probability of correct selection for a given budget of total samples N_T , to be optimally apportioned among the various alternatives. In either case, the odds that the current identification of the best alternative(s) is correct can be estimated at every stage of the sampling.

The concepts of probabilistic ordinal optimization can also be applied to continuous-variable probabilistic optimization problems. For such problems, improvement steps in the design space can be taken based on ordinal ranking of candidate alternatives according to relative probabilistic merit, rather than by attempting to resolve the actual merit value of each alternative as nonordinal approaches do. (See e.g. [2] and [7] for brief overviews of some nonordinal approaches to OOU, such as surrogate-based and Reliability-based methods.)

The relaxed ranking conditions for progress in ordinal optimization allow it to be more efficient than nonordinal OOU methods for certain problems, as established later. These latter approaches can introduce sources of noise and approximation error into the probabilistic objective function that can be avoided with probabilistic ordinal approaches (see [18] for a discussion). Furthermore, for continuous-variable problems the efficiency of sampling operations for ordinal ranking can be greatly increased by considering local spatial correlation in the design space as explored in Section 3.2. Additionally, advanced continuous-variable ordinal optimizers can be used, such as genetic algorithms, evolutionary simplex and pattern search methods ([9], [12]), and the simultaneous global/local ordinal method DIRECT[13].

3.1 Independent Designs

Consider two different designs or processes that each have some variability in their behavior or outputs. To compare the relative merit of the two designs, statistics of their relative merit can be compared. Thus, one can ask whether design A or design B has the higher mean output, or the lower probability of meeting or not meeting some acceptable threshold of behavior, or the smallest variance in the product produced.

The behavior or output of these independent designs can be sampled to generate statistics of their tendencies. For mean and probability, it is known from classical confidence interval theory that the values $\hat{S}_A$ and $\hat{S}_B$ of a statistic calculated from finite sampling will be realizations from normal distributions with means of the true statistical values S_A and S_B . The normal distributions have variances $\sigma_{\hat{S}_A}^2$ and $\sigma_{\hat{S}_B}^2$ that decrease as the number of samples increases. Figure 7 depicts normal distributions of the calculated statistics from two designs being compared.

Under limited sampling, the two normal distributions each have nonzero variance, which means that the distributions overlap to some degree as illustrated in Figure 7. Hence, it is possible for a calculated statistic $\hat{S}_A$ of design A to have a value greater than the calculated statistic $\hat{S}_B$ of design B, even though this is not the case for the true values S_A and S_B . If the optimization goal is to pick the design that maximizes the critical statistic (say, maximizes the mean output rate or probability of accep-

tance), then there is a nonzero probability that a misleading indication would be given that design A is better than design B. The more the error distributions overlap, the less certain it is that the chosen design is indeed the better design.

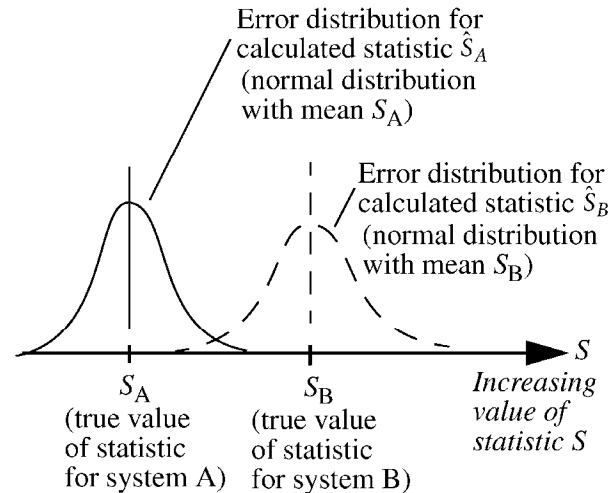


Figure 7: Normally distributed errors in Monte Carlo estimates of statistical behaviors of two different systems.

As the number of Monte Carlo samples of each design increases, the critical statistics become better resolved. That is, their confidence intervals decrease such that it begins to become more and more apparent which alternative has the better statistical behavior. Hence, the probability of making a correct selection between the two alternatives (i.e., the 'probability of correct selection', $P\{CS\}$), increases with more Monte Carlo sampling.

The $P\{CS\}$ value has the property that it converges to 100% at an exponentially fast rate as the number of samples increases ([4]). This is a much faster rate than for resolving the individual alternatives' behavioral statistics. That is, the standard deviations of the individual distributions in Figure 7 only converge at a rate of $1/\sqrt{\text{number of samples of that alternative}}$. However, by looking at the figure it is intuitive that as the "width" (standard deviation) of either or both PDFs decrease at a given rate, the question of which alternative's value is greater is resolved at a much faster rate. In fact, such differentiation occurs at an exponential rate with the total number of samples expended, N_T , whether only one alternative is sampled or both are sampled: $P\{CS\} \geq 1 - \alpha e^{-\beta N_T}$ (see Ref. [4]).

Ref. [18] describes a methodology for calculating the probability of correct selection when only two design options are considered at a time. A requirement can then be posed to determine to a given level of statistical assurance $P\{CS\}$ that the option picked has the better actual figure of merit. Meeting this requirement necessitates sampling the alternatives' behaviors a sufficient number of times. The question then arises: What is an adequate number of samples?