Optimization Problems and Methods in Quality Control and Improvement

AUTHOR:W. MATTHEW CARLYLE, DOUGLAS C. MONTGOMERY, GEORGE C. RUNGER

TITLE:Optimization Problems and Methods in Quality Control and Improvement
SOURCE:Journal of Quality Technology 32 no1 1-31 Ja 2000

ABSTRACT

The connection between optimization methods and statistics dates back at least to the early part of the 19th century and encompasses many aspects of applied and theoretical statistics, including hypothesis testing, parameter estimation, model selection, design of experiments, and process control. This paper is an overview of some of the more frequently encountered optimization problems in statistics, with a focus on quality control and improvement. Descriptions of a variety of optimization procedures are given, including direct search methods, mathematical programming algorithms such as the generalized reduced gradient method, and heuristic approaches such as simulated annealing and genetic algorithms. We hope both to stimulate more interaction between the statistics and optimization methodology communities and to create more awareness of the important role that optimization methods play in quality control and improvement.

Key Words: Control Charts, Design of Experiments, Design Optimality, Optimization, Process Control, Search Methods.

INTRODUCTION

MANY problems in statistical quality control and improvement involve the application of optimization methodology. Both the problem types and the variety of optimization methodology

employed are diverse. Often the context of the problem involves building a model of a system, and then using optimization techniques to determine values for the system parameters that result in the best measures of system effectiveness. Some examples include determining design parameters for a control chart, selecting the appropriate target for a production process where the feasible region for the quality characteristic is constrained by specification limits, and finding the settings for process variables such that one or several quality characteristics are optimized. As an example of the last type of problem, suppose that one is interested in determining the settings for the process variable's reaction time, temperature, and initiator feed rate that maximize the yield of a polymer material while simultaneously keeping the number average molecular weight and the viscosity in ranges that make the product acceptable to the customer.

While the above examples are relatively modern illustrations of the interface between optimization methods and statistics, the connection has been a long and extensive one, as illustrated by the use of least squares and linear regression. The method of least squares, which dates from the early part of the 19th century, involves finding the parameter values in an observational model that minimize the sum of the squared differences between the predicted and observed values. This is an unconstrained optimization problem that is easily solved by calculus methods because the model expectation function is a linear function of the unknown parameters. As in many optimization problems, the difficulty arises not in solving the problem itself but in selecting a model that accurately represents the "real" relationship expressed in the data. Most statisticians are familiar with the problems encountered with over-specified models; if a high-degree polynomial is fit to a handful of data points, predictions based on both interpolation and extrapolation often become unreliable. Indeed, selecting the "best" subset of variables for a linear regression model from a set of candidate regressors is a challenging optimization problem. Furnival and Wilson (1974), in a landmark paper, solved this problem by employing a branch-and-bound approach, a technique widely used in the operations research field. We will give a description of branch-and-bound techniques in a subsequent section.

In general, parameter estimation is often a nice example of the optimization-statistics interface. We discuss this particular problem in the next section because it is an excellent way to introduce the interface of optimization and statistics in a framework that is familiar to a wide audience. In the remaining sections of this paper we give an overview of optimization problems and techniques in process monitoring and control and process characterization and improvement using statistically designed experiments. We also provide a selection of both current and foundational references that we feel represent, but do not exhaust, the existing literature.

The purpose of this paper is to overview many interesting problems that are at the interface of quality improvement and optimization. We do not attempt to be encyclopedic in our coverage, and we realize that many other topics could have been included in this survey. Instead, we have chosen to focus on the problems that are of personal interest and relevant to our professional experience. We hope that this paper encourages discussions of this interface; we are interested in learning more, and we are sure that there are many more interesting topics that we have excluded. One obvious (and deliberate) omission is any coverage of classical mathematical programming techniques in statistics. We feel that, at the level of discussion in this paper, the book by Arthanari and Dodge (1981) does a very good job of introducing these problems and

their formulation to the reader. We also provide descriptive presentations of several of the more useful optimization techniques and give pointers to the relevant literature. Our primary objective is to encourage more interaction between the engineering, statistics, computer science, and operations research communities.

OPTIMIZATION AND PARAMETER ESTIMATION

Parameter estimation by the method of maximum likelihood is another elementary example of the connection between optimization methodology and statistics. Recall that if X is a random variable with probability density f(x; theta), where theta is a vector of the parameters of the distribution, and if X[sub1], X[sub2], ..., X[subn] is a random sample of size n, then the likelihood function of the sample is

L(theta) = f(x[sub1];theta) • f(x[sub2];theta) ... f(x[subn];theta).

The maximum likelihood estimator of theta is the value of theta that maximizes L(theta). Generally, calculus methods are employed to find the maximum likelihood estimator, and for many of the standard distributions this produces acceptable results. However, maximum likelihood estimation in regression and analysis of variance models often leads to complicated optimization problems. For example, in nonlinear regression and time-series applications, the likelihood function may exhibit local as well as global maxima and the solutions obtained may depend critically on the initial estimates selected for the parameters and, to some extent, on the method used. When maximum likelihood estimation is applied to estimation of variance components in random and mixed models, computing methods have to deal with very sparse matrices of high dimension, nonlinear functions, local and global optima, and the possibility of obtaining negative estimates of some of the variance components. A good reference on the computational and optimization aspects of variance component estimation is Searle, Casella, and McCulloch (1992).

PROCESS MONITORING AND CONTROL

Many of the best known connections between statistical quality control and optimization are in the general area of control charts for process monitoring and control. The design of such control charts is an example. To use a control chart, the practitioner must select the sample size, the interval of time between samples, and the control limits. Some charts, such as the exponentially weighted moving average (EWMA) chart, require the determination of additional parameters. Designing control charts to satisfy statistical criteria or performance criteria, such as specified average run lengths, is an optimization problem in the hypothesis testing framework. Early work in this area was done by Weiler (1952). Also, economic design criteria have been considered by many authors starting with Duncan (1956). Review papers by Montgomery (1980), Svoboda (1991), and Ho and Case (1994), the panel discussion contribution by Keats et al. (1997), and the book by Al-Sultan and Rahim (1997) describe much of the key work in this area.

CONTROL CHARTS WITH OPTIMAL ECONOMIC PERFORMANCE

The general approach followed by those working in this field is to make formal use of the fact that operating a process-monitoring scheme has economic consequences. That is, there are costs associated with operating the monitoring procedure, costs associated with investigating out-of-control signals and potentially correcting process upsets, as well as costs associated with

units reaching the customer that either do not conform to requirements or exhibit excessive variability. A formal mathematical model is used to connect these costs to the design parameters of the control chart and other system parameters. It is customary to assume that there is a single out-of-control state and that the time between occurrences of the assignable cause is an exponential random variable. The usual form of the objective function, E(A), is

E(A) = E(C) / E(T), (1)

where T is the length of the cycle (i.e., the length of time the process operates in the in-control state plus the operating period in the out-of-control state until the assignable cause is removed), C is the total cost incurred over a cycle, and E is the expectation operator.

Several authors have studied the objective function in Equation (1) and noted that in many practical situations it is locally convex (in a reasonable region of the parameter space). The first partial derivatives of E(A) with respect to the control chart parameters are typically nonlinear functions and do not have simple direct solutions. Optimization procedures have been developed based on solving numerical approximations to the system of first partial derivatives. However, the usual approach is to employ a direct search procedure.

Direct search methods encompass many approaches to optimization that avoid calculating derivatives. They are usually treated as an alternative to finite-differencing methods, which many believe to be inappropriate for real optimization due to problems with numerical stability. The simplest direct search procedures behave in much the same way as finite differencing, with the main difference being a much larger step size in direct search methods. The intent is to define a "neighborhood" of a given point by searching, perhaps along the coordinate directions, in the hope of finding a direction of significant improvement. For a thorough review of direct search methods, see Powell (1998).

The most popular direct search method is probably the simplex algorithm of Nelder and Mead (1965). In this method, a set of n+1 current solutions and their function values (for a problem in n dimensions) is maintained and updated so as to drop the point with the worst value and add a point in what is hopefully an improving direction. (The name of the algorithm comes from the fact that the convex hull of (n + 1) points in n dimensional space, if they are not in a degenerate configuration, is known as a simplex.) The specific ways in which this is accomplished are numerically complicated, but, basically, the algorithm takes advantage of the local structure of the response surface. It stretches, shrinks, and otherwise manipulates the solution simplex to move quickly when improvements are easy and to be careful when necessary so as to avoid missing a local optimum.

Another approach is the use of grid-based methods. In this approach, a discrete lattice is overlaid on the feasible region; the current solution is then one of the lattice points, and search directions may cycle through the coordinate directions until no improvement can be found. The pattern search method of Hooke and Jeeves (1961) is an extension to this idea whereby the algorithm generates search directions that are not necessarily in the coordinate directions, but are instead based on the last step taken and the improvement achieved along the coordinate directions. The Hooke and Jeeves method may shrink or stretch the current step length as appropriate. Many contemporary optimization texts discuss this method, but for an extremely clear exposition, with classic illustrations, see the text by Wilde and Beightler (1967). One step of

the Hooke and Jeeves method is illustrated in Figure 1.

Most direct search methods do not guarantee convergence to a global optimum; as in most optimization approaches, local optimality is the best that can be guaranteed. Some, however, such as the refining grid-based methods for bounded domains (again, see Powell (1998)), can make this promise, although at the expense of enormous computational effort. Another drawback is that they are only appropriate for very simple constraints of the form

a[subi] [less or equal] x[subi] [less or equal] b[subi], (2)

for each of the decision variables, x[subi].

As an illustration of search methods in designing control charts, Montgomery (1982) gives a procedure that sequentially steps through a range of values of the sample size n, simultaneously both utilizing a one-dimensional search on the width of the control limits and solving a numerical approximation to the derivative of the objective function with respect to the interval between samples. This results in a very effective search. He applies the algorithm to the X control chart. Many other authors have developed optimization procedures for other types of control charts using a similar approach.

Woodall (1986 and 1987) observed that economically-optimal control chart designs often have poor statistical performance; for example, they may exhibit short in-control average run lengths and thus be susceptible to excessive false alarms. Saniga (1989) proposed dealing with this problem by imposing statistical constraints on the economic solution, leading to what is often called economic-statistical design or statistically-constrained economic design approaches. The constraints are typically on average run length and/or average time to signal and take the form of inequalities such as Equation (2). Saniga (1989) originally applied this approach to the X and R control charts. McWilliams (1994) provides a FORTRAN program for this procedure. Subsequently, Saniga, Davis, and McWilliams (1995) considered statistically constrained economic design of attribute control charts and Montgomery et al. (1995) used this approach for the EWMA control chart.

Generally, the addition of constraints to an optimization problem results in a degraded solution, with respect to the optimal objective function value, in comparison to the optimal solution to the unconstrained problem. On the other hand, constraints may be an integral part of the problem. For example, in the economic control chart design problem, constraints ensure adequate statistical performance for the chart, regardless of the economics. This is desirable, since the actual cost parameters are difficult to estimate. Furthermore, many studies have shown that adding statistical performance constraints to an economic model usually results in only a modest increase in system costs.

With the increasing availability of computer software to implement these procedures and the industrial focus on cost reduction, it is perhaps surprising that there has been little industrial use of economic models to design control charts. Possibly many practitioners feel that using a complex optimization procedure to design a relatively simple process monitoring scheme is not worthwhile, or they may be intimidated by the number of cost and other system parameters that must be estimated in order to use these models. One possible solution to the cost estimation issue was suggested by Del Castillo, Mackin, and Montgomery (1996), who proposed a multiple criteria optimal design procedure for control charts. They utilized an interactive, multi-criteria, nonlinear optimization algorithm, along with a model for which the user does not have to specify

the costs of either false alarms or out-of-control performance. Only sampling costs, which are usually easier to estimate, need be specified. Some authors have suggested that because there has been little industrial use of economic models, research in the area should be abandoned. However, Woodall and Montgomery (1999) observe that these authors may have been premature in reaching such a conclusion. For example, much of the theoretical work on alphabetic design optimality in the 1950's has not been extensively used by practitioners until the last 10 years when computer codes that implement the D-criterion (for example) became widely available as features in experimental design software. Furthermore, this was one of the first areas in quality control and applied statistics that attracted the attention of industrial engineers and operations research analysts, and it served to initiate much useful interaction between these disciplines.

CONTROL CHARTS WITH OPTIMAL STATISTICAL PERFORMANCE

The traditional design for an X control chart is subgroups of 3 to 5 objects with 3-sigma control limits. Shewhart (1931) recommended this design; therefore, it can probably be considered a standard for an X chart. These guidelines, along with the critical rational subgrouping principle, have provided an implicit design strategy, at least for Shewhart charts, for decades. However, one might choose to modify one or more of the standard design parameters--subgroup size, control limits, and frequency of sampling--in a specific application. With Shewhart control charts, the concept of statistical errors can be used to assist in the design in a manner nearly identical to hypothesis testing. One trades false alarms against the sensitivity to process upsets. Instead of statistical errors, many prefer to express performance in terms of the run length.

The run length is the number of points plotted until a signal, and the preeminent summary of the run length distribution is the average run length (ARL). A large ARL is desired when no assignable causes are present; a short ARL is desired otherwise. Several authors have commented on the extreme skewness of the run length distribution and the poor summary provided by ARL. Still, ARL is probably the most widely used design criterion.

Weiler (1952) provided an early analysis of control chart performance. He determined the subgroup size that minimizes the average amount of inspection required to detect a change in the process mean for Shewhart X charts. Page (1954) determined the subgroup size and control limits to minimize the average run length of items inspected. However, these analyses did not explicitly consider the frequency of samples. For example, if measurements are recorded every minute, the three-sigma limits might generate false alarms every few hours. The rational subgrouping principle provides guidelines, but often more than one design is still feasible. Keats, Miskulin, and Runger (1995) explicitly considered three design parameters: subgroup size, control limits, and frequency of sampling in the optimization of average production length (APL). APL is the mean number of units produced before a signal from the chart. In the common case of a constant production rate, APL is approximately proportional to the average time to signal (ATS) of a control chart. However, ATS typically assumes that a subgroup is formed instantaneously. Keats, Miskulin, and Runger (1995) provided a simple design strategy for non-economic models that were optimized by straightforward numerical methods.

Outside of Shewhart control charts, the lack of the simple geometric distribution for run length has forced a focus on metrics such as the ARL. Both cumulative sum (CUSUM) and EWMA

control charts are often designed for ARL performance. The analysis uses either an approximate solution to an exact integral equation or an exact solution to an approximate Markov chain model. Champ and Rigdon (1991) discussed the relationship between the two methods. Hawkins and Olwell (1997) provided recommendations for CUSUM designs. Prabhu, Runger, and Montgomery (1997) considered the relationship between subgroup sizes and sampling intervals to optimize the time to signal of a CUSUM control chart.

Interesting optimization problems arise from charts designed so that the interval between samples or the sample size is changed during the operation of the chart. These adaptive charts can be designed to achieve a particular objective, and ATS is a convenient metric. Reynolds et al. (1988) and Runger and Pignatiello (1991) considered X charts and many publications extended these results to alternate charts and adaptive strategies. Prabhu, Runger, and Montgomery (1995) provided a design tool. A particularly difficult problem is the optimal adaptive strategy for the important steady-state performance of these charts. Runger and Montgomery (1993) used the Karush-Kuhn-Tucker (KKT) conditions for nonlinear optimization in the context of a discretized version of an X chart to obtain the globally optimal adaptive-sampling rule. Work continues on the optimal rules for other charts.

PROCESS OPTIMIZATION AND DESIGNED EXPERIMENTS

The most obvious connection between designed experiments and optimization is in the area of response surface methodology (RSM). Introduced by Box and Wilson (1951), RSM is a collection of statistical design techniques, empirical model-building procedures, and optimization methods that are intended for use by engineers, statisticians, scientists, and other process specialists in attaining optimal operating conditions for processes. Recent papers by Box and Liu (1999), Box (1999), Myers (1999), and the discussion of those papers highlight many important aspects of the RSM field.

Much of RSM, particularly in the early years, was focused on finding operating conditions for the process variables that resulted in a maximum or minimum response. The canonical analysis of a second-order response surface, in which the model is transformed via a translation and rotation of the original coordinate axes to a form that facilitates interpretation of the fitted surface as either a maximum, a minimum, a saddle point, or some type of ridge system, is equivalent to determining whether the Hessian matrix is positive-definite, negative-definite, or indefinite. Ridge analysis, introduced by Hoerl (1964), is used to find a point of either maximum or minimum response on a saddle system with spherical region constraints. The three texts on RSM by Box and Draper (1987), Khuri and Cornell (1996), and Myers and Montgomery (1995) discuss these and other optimization techniques.

In the early 1980's renewed interest in statistically designed experiments was catalyzed by the robust parameter design problem and by novel design and analysis methods proposed by Genichi Taguchi to solve this problem. In robust parameter design, as viewed by Taguchi, the experimenter is trying to design a system (a product or a process) to minimize variability transmitted to the response by variables that are uncontrollable once the system is deployed in the field. Taguchi referred to these uncontrollable variables as noise variables and assumed that they were at least controllable for purposes of an experiment. The usual formulations of the robust parameter design problem involve maximizing or minimizing the mean of a response

while simultaneously minimizing variability, or driving the mean to a specific target while minimizing variability around the target. Taguchi was really advocating optimization; however, the statistical methods he proposed were inefficient from an experimental design viewpoint, and in many cases they will miss significant information about interactions, which could lead to suboptimal solutions. He also utilized a pick-the-winner approach to optimization that was not based on any system model (empirical or mechanistic). This approach selects one of the original design points (or a non-design corner point in the region) as the optimum and, therefore, does not assure the experimenter that even a local optimum has been achieved. Much of the criticism and debate regarding Taguchi's methods is summarized in Myers and Montgomery (1995, Ch. 10) and in Montgomery (1997, Ch. 14).

Formal recognition of the robust design problem as an optimization problem is due to Vining and Myers (1990), who expressed the problem as maximization or minimization of an objective function subject to a single equality constraint. They assumed both the objective function and the constraint to be complete second-order response surface models, and they employed the dual response optimization algorithm of Myers and Carter (1973) to solve the problem. This algorithm is similar in many respects to ridge analysis, which seeks the maximum or minimum point on a quadratic function subject to the solution lying inside the spherical design region and utilizes a single Lagrange multiplier in solving the optimization problem. Myers and Carter (1973) optimized a quadratic function subject to both a quadratic equality constraint (for instance, a surface of constant variance) and the spherical design region constraint. There have been several other formulations and methods proposed for the dual response problem; for example, see Del Castillo and Montgomery (1993), Lin and Tu (1995), and Vining (1998). Dual responses are a special case of multiple responses, an area in which there is a significant interplay between optimization and statistics.

THE MULTIPLE RESPONSE PROBLEM

Many designed experiments involve more than one response. For example, in determining optimum conditions for yield in a chemical process, we may simultaneously be interested in conditions that bring physical or chemical properties of the material, such as molecular weight, viscosity, or concentration, to levels that satisfy customer requirements. In some industrial settings, such as semiconductor manufacturing, 12 to 15 response variables are not unusual. The robust parameter design problem discussed above is a special case of this problem where there are two responses, the mean and variance of a fundamental response observed during the experiment.

The objective of many designed experiments is response optimization. When there are several responses, this problem is often called the multiple response optimization problem or the simultaneous optimization problem. Historically, in the response surface context, we have solved this problem by first modeling each response individually and graphically superimposing the contour plots from each model. This overlay is then examined to find the appropriate set of operating conditions for the process. Since it is unlikely that a single set of operating conditions produces the optimum for all responses simultaneously, one usually searches for a set of conditions that comes as close as possible to satisfying all response requirements. This set of conditions is often referred to as the "sweet spot" for the process. For an early example of this

approach, see Lind, Goldin, and Hickman (1960).

Figure 2 illustrates this approach using an example from Montgomery (1997). In this example, there are two process variables: x[sub1] = reaction time and x[sub2] = temperature and three responses: y[sub1] = yield, y[sub2] = viscosity, and y[sub3] = molecular weight. The objective is to maximize yield, while simultaneously keeping viscosity and molecular weight in the ranges specified by the customer; namely, y[sub2] = 65+/-3, and y[sub3] < 3400. Figure 2 plots the predicted yield contour 78.5 (this was considered the lowest acceptable yield), predicted viscosity contours 62 and 68, and the predicted molecular weight contour 3400. Clearly, there are several sets of potential operating conditions for time and temperature that will achieve these objectives.

In practice, the graphical approach is limited to two or perhaps three dimensions. Therefore, there is considerable interest in more general approaches. One of these is to formulate the multiple response problem as a constrained optimization problem. In general, suppose that we have m responses and that for each we have a response surface model ^y[subi] = f[subi](x). We select one response as the primary response (or the objective function) and solve

max(min)^y[sub1] = f[sub1](x)
subject to (3)
l[subi] [less or equal] ^y[subi] = f[subi](x) [less or equal] u[subi], i = 2, 3, ... m
x [element of] R,

where we have chosen the first response as the objective function and l[subi] and u[subi] are lower and upper bounds on the remaining responses, i = 2, 3, ..., m. The last constraint in Equation (3) is a "region" constraint that ensures that the solution vector, x[subs], is inside the experimental region, R. For cuboidal designs, this constraint usually takes the form -1 [less or equal] x[subi] [less or equal] 1, i = 1, 2, ..., k (the number of design variables), while for spherical designs the constraint would be x'x [less or equal] r[sup2], where r is the design radius. At least one commercially-available experimental design software package, Design-Expert, will allow the user to formulate the multiple response optimization problem in this manner, although the region constraint is required to be cuboidal. It should be noted that in the special case where the f[subi] are convex functions and there are no lower bounds on the constraint functions, the result is a convex optimization problem (minimizing a convex function over a convex feasible region), which is a well-solved problem with several efficient methods. For a thorough treatment of convex optimization, see, for example, Nesterov and Nemirovskii (1994) and the many references contained therein. The majority of instances do not satisfy these conditions for convexity, either because one or more of the f[subi] are saddle systems and, hence, indefinite, or because there are finite lower bounds. Even so, in some cases the models can be altered slightly (for instance, in the case of a near-ridge saddle) to a convex optimization problem with little impact on the quality of the optimal solution.

There are two broad classes of methods that can be used to solve this problem: direct search methods and mathematical optimization algorithms. The direct search methods usually employed are the pattern search method of Hooke and Jeeves (1961) or the sequential simplex (see Nelder and Mead (1965) or Copeland and Nelson (1996) for a modern presentation in the context of multiple responses). Most commercially available experimental design software that

handles multiple response optimization uses one of these two techniques.

As noted previously, direct search methods are essentially hill-climbing methods. That is, they start from an initial solution and move in the local gradient direction, or an appropriate approximation to it. An obvious disadvantage of this approach is that multiple optima are often present (see, e.g., Figure 2), and there is no guarantee that one will find the absolute or global optimum solution. In fact, we are usually interested in finding all of the optimum solutions because some solutions may be more desirable than others based on practical considerations not taken into account in the mathematical formulation. For example, one of the operating windows in Figure 2 is larger than the other, implying more robust operating conditions for the process. This potential problem can be alleviated somewhat by selecting several starting points and applying the search algorithm from each starting point. Since the volume of the search region increases dramatically with an increase in the number of design variables, the analyst must select enough starting points to ensure that all (or nearly all) of the optimal solutions are found.

Various mathematical optimization algorithms can also be applied to this problem. The generalized reduced gradient (GRG) method (see Lasdon, Fox, and Ratner (1974) and Del Castillo and Montgomery (1993)) is a very popular choice because the method is broadly applicable to many types of mathematical programming problems and because it is widely available in commercial software. The popular spreadsheet package EXCEL employs a version of the GRG, and so it is possible to solve multiple response optimization problems without additional specialized optimization or statistical software. Since the GRG makes explicit use of derivatives, it is considerably more efficient than direct-search methods and may be less likely to be confused by numerous local optima. GRG methods are nice because they can find locally optimal solutions to problems with nonlinear equality constraints, which normally cause problems for other standard approaches. The basic idea is that the gradient (in the case of maximization) is taken to be a desirable search direction. It is projected onto the tangent space of the active constraints (equalities and tight inequalities) to give the reduced gradient. The algorithm then takes a predictor step in this direction (usually moving off of the active constraints) and employs a Newton's (or similar) method to the infeasibilities incurred to correct the step and return to the set of active constraints. The GRG is thus a member of the large class of methods known as predictor-corrector methods. Figure 3 illustrates the predictor-corrector approach.

There are, of course, many possible variations of the optimization problem formulated in Equation (3). Del Castillo (1996) presents a formulation of the multiple response optimization problem that allows one to obtain solutions that simultaneously satisfy confidence region constraints for all responses. His methodology is applicable to both linear and quadratic models, but assumes that a rotatable design has been used.

A third general approach to multiple response optimization is to attempt to simultaneously optimize all m responses. An obvious possibility is to form some function combining the responses, such as

^y = [Graphic Character Omitted] a[subi]^y[subi], (4)

where a[subi] are the weights, and optimize the composite response ^y. In practice, choosing the weights appropriately is usually difficult, and so this approach is not widely used unless there

is some unequivocal way to select the weights (e.g., see Montgomery, Talavage, and Mullins (1972)) or unless the weights can be selected dynamically via some interactive multiple response optimization procedure. Montgomery and Bettencourt (1977) illustrate the use of the Geoffrion-Dyer interactive vector-maximal algorithm (see Geoffrion, Dyer, and Feinberg (1972) and Dyer (1973)) for a response surface system involving five responses.

In general implementations, the weights are chosen based on the relative importance of the various responses, usually obtained through the advice of experts on the process in question. The use of weights to form a composite objective function reduces a multiple-response problem to a simpler, single response problem. It also obviously "loses" some information in the conversion, as one single value of the composite response can correspond to many possible levels of the individual responses. Therefore, when there are multiple responses to be optimized, this method is not expected to provide a robust solution or to even help trace the efficient frontier of solutions.

The most popular approach to simultaneous response optimization is the desirability function approach, as proposed by Derringer and Suich (1980). In this approach, each response is converted into an individual desirability function, say, d[subi] = h(^y[subi]), where d[subi] = 0 if the response is in an unacceptable range, d[subi] = 1 if the response is at the optimum value, and 0 < d[subi] < 1 otherwise. When the response, y[subi], should be at a target value, T, such that L [less or equal] T [less or equal] U, Derringer and Suich use the function

[Formula Omitted]

to convert the response to a desirability. In Equation (5), the exponents s and t are used to determine the shape of the desirability function, thereby controlling how important it is that the response achieve the target in the final optimal solution. Often these exponents are chosen in the range from 0.01 to 10. Similar conversions are made for responses that should be either maximized or minimized. Then the solution, x[subs], is found that maximizes the geometric mean of the individual response desirabilities,

D = ([Graphic Character Omitted] d[subi])[sup1/m].

Direct search methods are used to find the optimum solution because the individual desirability functions are not differentiable. Note that this version of the multiple response problem is still susceptible to the multiple-optima phenomena noted when discussing the mathematical programming formulation. Because direct search methods could easily terminate at a local optimum, when using these methods it is essential to start from several different points in the factor space to ensure that all or at least a reasonable number of the optima have been found. This approach to multiple response optimization has been implemented in several experimental design software packages. For example, see Design-Expert.

Del Castillo, Montgomery, and McCarville (1996) show how to replace the non-differentiable desirability functions of Derringer and Suich with differentiable equivalents that are fourth-order polynomials. The coefficients in these polynomials are treated as tuning constants and set so that they will closely mimic the shape of the original, non-differentiable desirability functions. Then mathematical programming optimization algorithms such as the GRG can be used to optimize the overall desirability function. This allows experimenters to use efficient optimization tools (including spreadsheets) to solve a variety of formulations of the multiple response optimization problem.

Problems with highly nonlinear, or multimodal, objective functions are extremely difficult to solve and are further complicated by the presence of multiple objectives. A standard approach in this situation is to use a heuristic search procedure. There are several families of methods that have been proposed for finding global optima for such problems, two of which are simulated annealing and genetic algorithms. Neither is guaranteed to find the global optimum, but they have been designed to use different means of avoiding getting stuck at local optima.

Simulated annealing, although properly described as a Monte-Carlo simulation procedure, (see Metropolis et al. (1953)) is a local search procedure; it maintains a current solution to the problem and defines a neighborhood of "nearby" solutions. Instead of considering all neighbors and only moving to a new solution that improves the objective, a simulated annealing algorithm picks a random neighbor and then has a positive probability of moving to a slightly worse solution. (The algorithm always accepts a neighbor that improves the objective.) This probability is decreasing in the magnitude of the change, and as the algorithm progresses, the probability function is continually lowered for each value (in a process called "cooling") until, in the last stages of the algorithm, only improving changes are made.

Numerical examples give mixed results. Simulated annealing is designed to wander around the feasible region a lot at first, crossing over "hills" in the objective surface in a search for a significantly lower area. As the cooling proceeds, though, it becomes more and more constrained and can become confined to a local "valley." The rate at which the cooling takes place, called the cooling schedule, should therefore be designed with the structure of the response surface in mind. The best results usually require a bit of tuning of the cooling schedule, and for robust implementations, simulated annealing is often coupled with a random starting point and re-started several times. For a more detailed account of the simulated annealing approach, see Bohachevsky, Johnson, and Stein (1986). Haines (1987) applies this approach to the problem of constructing optimal experimental designs.

Genetic algorithms (see Holland (1974)), in contrast to the one-solution-at-a-time approach of most optimization algorithms, maintain a population of hundreds, or thousands, of solutions. They then use the evolutionary concepts of reproduction, mutation, and selection to create successive generations of these solutions in the hope that the population will tend to produce very strong members and, thus, near-optimal solutions.

This is accomplished by encoding each solution as a string of characters, or digits, called a chromosome. This encoding is done in such a way as to allow for a "crossover" operation between two randomly chosen solutions in the current population to produce another solution that combines aspects of each of the originals. For example, in a problem with 10 factors at two levels apiece, the chromosome might be a 10-digit binary string, where the i[supth] digit represents a low or a high level of the i[supth] factor (see Figure 4). Then, given two sets of factor levels, a third solution (and a fourth--as is sometimes done) could be formed by swapping out the first k values of one string for the first k values of the other. This still gives a "feasible" setting of the 10 factors, and the hope is that if the first few factors were set the right way in one and the last few factors were set right in the other, then the new offspring will have most (or all) of its factors set right. Of course, if they aren't, a poor solution won't join the next generation after an objective-based selection procedure is applied. Genetic algorithms also apply "cloning" methods to ensure that particularly strong solutions survive and "mutation" methods to fiddle with

random elements of random solutions to ensure that the entire feasible region eventually gets considered.

The main strength of genetic algorithms is that the objective function doesn't have to have any structure at all. The only structure required is in the encoding of solutions into chromosomes to assist in the definition of the crossover operation. One potential drawback is that chromosomes must be encoded such that every possible chromosome is feasible for the original problem since genetic algorithms do not consider constraints directly. One popular work-around for this shortcoming is to introduce a penalty function for violated constraints, thus shifting all of the work into the objective function. Genetic algorithms also require tuning, much like simulated annealing. The mutation rate, the selection and cloning procedures, the population size, and the number of generations to evaluate (or the convergence criterion to employ) all affect the performance of genetic algorithms.

In the multi-objective case, genetic algorithms are especially appealing because of the ease with which multiple populations can be maintained (one for each objective, perhaps, plus one population for an aggregate objective, such as a desirability function). Then the main algorithm can be designed to look for solutions that are strong with respect to all of the responses present, possibly combining strong solutions from the separate populations in an attempt to produce such solutions. For a small sample of applications involving these approaches, see Horng (1998); Schaffer (1994); Murata, Ishibuchi, and Tanaka (1996); and Cieniawski, Eheart, and Ranjithan (1995).

All of the multiple response optimization techniques we have discussed assume that the responses are independent or uncorrelated. In some situations, this assumption is inappropriate and a method that takes the dependence among responses into account would be appropriate. Khuri and Conlon (1981) and Khuri and Cornell (1996) present an approach in which a generalized distance measure is employed to indicate the weighted distance of each response from its individual optimum value. The variances and covariances of the responses are used in determining the weights. Then the solution is found that minimizes the generalized distance. This is a very nice approach that does not seem to have been utilized extensively in practice probably because it is not available in standard experimental design or general statistical software.

OPTIMAL DESIGNS

By the phrase "optimal" design we usually mean a design that is best with respect to some specific criterion. To fix ideas, suppose that we are interested in fitting the model

y = Xbeta + epsilon, (6)

where y is an n × 1 vector of observations, X is an n × p matrix of the design factor settings and (depending on the order of the model) functions of those settings, beta is a p × 1 vector of unknown parameters, and epsilon is an n × 1 vector of errors usually assumed to be normally and independently distributed with mean zero and common variance sigma[sup2]. The variances and covariances of the least squares estimates of the parameters beta are the elements of sigma[sup2] (X'X)[sup-1]. Furthermore, the joint confidence region for beta is

(^{Begin Greek}b - b)' ({End Greek}X'X) (^{Begin Greek}b - b) {End Greek}[less or equal] p^sigma[sup2] F[subalpha,p,n-p]. (7)

Much of optimal design theory and practice is focused on selecting the settings of the factors that make the elements of (X'X)[sup-1] in some sense small. The fundamental mathematical work in this area was started by Kiefer (1958) and followed up in subsequent papers (Kiefer (1959, 1961, and 1962)). See also Kiefer and Wolfowitz (1960). A very useful recent reference is Pukelsheim (1995), which summarizes many of the results and points out connections between the various optimality criteria.

Because the squared lengths of the principal axes of the confidence ellipsoid in Equation (7) are proportional to the eigenvalues lambda[sub1], lambda[sub2], ..., lambda[subp] of the matrix (X'X)[sup-1], one very popular approach to optimal design involves minimizing these eigenvalues. Specifically, the criterion is to choose a design such that maximizes

|X'X| = [Graphic Character Omitted] lambda[sup-1[sub[subi]. (8)

A design that maximizes this criterion is called a D-optimal design. Other design optimality criterion can be expressed in terms of the eigenvalues of (X'X)[sup-1]. For example, consider:

tr(X'X)[sup-1] = [Graphic Character Omitted] lambda[subi]

lambda[submax] = max(lambda[sub1], lambda[sub2], ..., lambda[subp]).

Designs that minimize the trace of (X'X)[sup-1] or that minimize the largest eigenvalue of (X'X)[sup-1] are called A- and E-optimal designs, respectively. The D-, A-, and E-criteria minimize the volume of the joint confidence region for the regression coefficients (Equation (7)), the sum of the variances of the regression coefficients, and the maximum variance of a regression coefficient, respectively. A very important criterion, particularly for response surface designs, is to find a design that minimizes

G = max [x[sup(m)'] (X'X)[sup-1]x([sup(m)]] for x [element of] R,

where x[sup(m)] is a point of interest in the experimental region R expanded to model form. This is the G-optimal criterion; essentially, it seeks a design that minimizes the maximum prediction variance over the experimental region.

These alphabetic design optimality criteria were originally developed as criteria for situations where the designs are considered a continuous probability measure over the region of interest. These continuous designs are, of course, not realizable but they could be approximated by designs consisting of a discrete arrangement of points with an integer number of trials at each design point.

The D-optimal criterion is probably the most familiar to practitioners and the most widely used in practice, at least in part because it has been implemented in many computer software packages. The usual approach is to have the experimenter define a set of N candidate design points for the model that is to be fit. This candidate set is often a grid of points over either a cuboidal or spherical region. Then an optimization algorithm is employed to select the subset of n < N that maximize the determinant of (X'X). Two general types of algorithms are used.

Exchange algorithms are widely used to generate designs that are nearly D-optimal. These algorithms operate by selecting a set of n points as an initial design and then exchanging points in the design with unused candidate points in an effort to increase the determinant of (X'X). There are various types of exchange algorithms, such as the rank-1 augmentation schemes that choose points to add and remove sequentially (see Wynn (1972) and Mitchell (1974)) and the rank-2 exchange algorithms that choose points to add and remove simultaneously (see Cook and Nachtsheim (1980) and Johnson and Nachtsheim (1983)). In the k-exchange algorithm of

Johnson and Nachtsheim (1983), all one-point-at-a-time exchanges are considered. Then once all these one-point-at-a-time exchanges are completed, two-point-at-a-time exchanges are evaluated. If any two-point exchange is made, all the one-point-at-a-time exchanges are once again evaluated. When all of the two-point exchanges are completed, three-point exchanges are considered. In general, this process continues until all k-point exchanges have been considered. If at any level an exchange is successful, then the process always begins again with one-point-at-a-time exchanges. Many computer codes use k = 5. Li and Wu (1997) report on using exchange algorithms with several different values of k and point out that k = 5 is a good compromise between design efficiency and computing time. Meyer and Nachtsheim (1995) describe a variation of this procedure called the cyclic coordinate exchange algorithm in which individual coordinates of the design points and the candidate points are exchanged.

Often the initial design is chosen at random from the candidate set. In other cases it may be selected according to some other design criterion, such as the distance-based criterion. Since there is no assurance that the final design created by the algorithm is actually a D-optimal design, some computer codes solve the problem several times using different initial designs and select the final design with the largest determinant of (X'X) as the D-optimal design.

Lin and Kernighan (1971) develop an exchange procedure in the context of the Traveling Salesman Problem (TSP), another combinatorial optimization problem of the same complexity as the optimal design problem. It has a bigger neighborhood than the set of simple pairwise (or 2-opt) swaps, but is less computationally intensive than the higher order swapping methods (k-opt). It implements a backtracking algorithm on a sequence of 2-opt swaps at each iteration, resulting in exchanges of different magnitudes at every step. The Lin-Kernighan procedure is one of the most competitive algorithms for the TSP, but as far as we are aware it has not been implemented in the context of optimal designs.

As in the optimal design algorithms, the concept of multiple runs with different starting points can be extremely helpful. Since the local optimum in which a k-opt algorithm eventually gets stuck is completely determined by the starting point, picking several random (or well spread out) starting points offers the hope of avoiding those that are significantly worse than the global optimum. This problem often arises when greedy heuristics are used to construct the initial solution (such as a nearest-neighbor algorithm for the TSP).

The other approach to constructing alphabetically optimal designs is to use a branch and bound algorithm (see Welch (1982)). Generally, a branch and bound algorithm is a divide and conquer approach to combinatorial optimization problems. The two main components of branch and bound algorithms are 1) a way of defining a relaxation of the original problem that is easy to solve optimally and 2) a way of partitioning the relaxation into two (or more) subproblems. The general approach is to solve the original relaxation (in many cases this is a relaxation of an integrality restriction or of a set of "nasty" constraints) to get an upper bound on the value of the optimal (maximal) solution. If the solution to the relaxation is feasible for the original problem, then we have found the optimal solution (since its value must be greater than or equal to any feasible value). Otherwise, the branch and bound algorithm develops two alternative subproblems (which cover all possible feasible solutions) and models each as a relaxed subproblem. These subproblems are then solved, each one providing an upper bound for the respective region. If any subproblem has a feasible solution and it is better than the current best

solution (if any have been seen), then it becomes the new incumbent (best solution found). When all subproblems have been fathomed, the current incumbent must be the optimum because we've ruled out (either explicitly or implicitly) all other possibilities.

This approach obviously leads to an explosion in the number of subproblems. Efforts to mitigate this growth center on the concept of bounding. If a feasible solution is found in a subproblem, then obviously no branching needs to occur in that particular subproblem. The real trick comes in realizing that this feasible solution can impact other subproblems, too. This is because any feasible solution to the original problem automatically places a lower bound on the optimal solution value. If any subproblem relaxation has a value below this lower bound, then that subproblem cannot contain the optimum and therefore no further branching is necessary. Finally, any infeasible subproblem requires no branching. These three methods of halting the branching process are known as fathoming. The order in which subproblems are evaluated can have a large impact on the processing time. Frequently, shifting from a breadth-first search of the subproblem tree to a depth-first, or guided heuristic search, can reduce the total computation time by several orders of magnitude. Finally, the bounding step can be modified by fathoming if the upper bound is within an acceptable margin of the lower bound, say, 5%. When all the subproblems have been fathomed, the remaining feasible solution (if any) will be guaranteed to be within 5% of optimal. This "fudge factor" can also have a drastic effect on the run-time of general branch and bound algorithms.

While alphabetically-optimal designs are useful in certain situations, they have been subjected to a wide variety of criticisms. They are very sensitive to the assumptions made by the experimenter, particularly with regard to the form of the model that is to be fit, and the number of observations that are to be taken. For instance, a design that is D-optimal for a particular model might not support another very similar model (e.g., see Myers and Montgomery (1995, Example 8.11)). There has been some work devoted to making optimal designs less model-dependent. For example, Cook and Nachtscheim (1982) have investigated strategies for constructing optimal designs for a family of models. Another technique for reducing the model dependence is the Bayesian D-optimal approach. In this procedure the experimenter assumes that a model of order d[sub1] is of primary interest, but wishes to protect against the possibility that the true system is best described by a model of order d[sub2], where d[sub2] > d[sub1]. Prior probabilities are assigned to the additional terms in the higher-order model and a design that is D-optimal with respect to the posterior covariance matrix of the model parameters is found. Good references for this procedure are Chaloner (1984), Pilz (1991), and DuMouchel and Jones (1994). Bayesian D-optimal designs for mixture experiments are discussed in Andere-Rendon, Montgomery, and Rollier (1997). Probably the best uses of alphabetically-optimal designs are in irregular design regions (such as found in mixture experiments) or where there are unusually restrictive blocking requirements in the experiment.

Another area in which there has been a useful connection between experimental design and optimization methodology is construction of supersaturated designs. These designs were introduced by Booth and Cox (1962) and employ n observations to investigate k factors, where k > n - 1. Since even in the relatively simple case of two-level factors a supersaturated design cannot be orthogonal, the experimenter would like to find a design that is as "close" to orthogonal as possible. Booth and Cox (1962) used a computer search procedure to generate

their designs. Lin (1993) renewed interest in both the theory and application of these designs, presenting a method of design construction based on half-fractions of Hadamard matrices. Lin (1995) and Li and Wu (1997) describe other algorithms for constructing supersaturated designs. Their methods are similar to the local search algorithms employed in other related combinatorial optimization problems.

MULTIVARIATE ANALYSIS

Optimization is widely used in multivariate statistical analysis. In particular, the likelihood principle and maximum likelihood estimates are common approaches that provide benchmarks for alternate multivariate methods. Because the likelihood methods are similar to those used in univariate statistical analyses, we prefer to mention some alternative uses of optimization.

An interesting optimization problem results from modern methods to detect multivariate outliers. Outliers are much nastier in high-dimensional data because graphical detection is more difficult. Consequently, numerical methods that are not sensitive to the outliers--but, instead, detect the outliers are valuable. Rousseeuw (1985) proposed the minimum volume ellipse (MVE) as a high-breakdown estimate for the mean vector and covariance matrix of a multivariate dataset. The MVE is the smallest ellipsoid covering just over half of the data. The robust estimates of the mean and covariance matrix come from classical calculation of these quantities only using the subset of observations that are contained in the MVE. Although simply described, the numerical methods to determine the MVE are not trivial. The original algorithm used random resampling to find the subset of observations with the smallest volume of the ellipsoid covering them.

Hawkins (1993) improved the algorithm using steepest descent with random restarts rather than the random sampling method. Woodruff and Rocke (1993) proposed a heuristic search optimization procedure. Currently, the software S-plus 4.5 uses genetic algorithms.

Unfortunately, the MVE is inefficient with asymptotic efficiency of zero (Davies (1992)). The implementation in S-Plus goes an additional step and adds all remaining observations apart from the MVE back in for the final estimate of the mean and covariance matrix if their Mahalanobis distances (calculated with the MVE estimates) are less than a cutoff value from the chi-squared distribution. This significantly increases the efficiency of the estimates.

An alternative estimator is the Minimum Covariance Determinant (MCD) that was also introduced by Rousseeuw (1985). The MCD searches for the sample of size $q < n$ that has the minimum value among all samples evaluated by the determinant of its covariance matrix. The estimator is most often a 50% breakdown estimator so q is set to the integer part of $(n+p+1)/2$. The algorithm has evolved from random resampling much like the MVE. The improvements proposed by Hawkins (1994), Woodruff and Rocke (1994) and the genetic algorithms parallel those of the MVE. Butler, Davies, and Jhun (1993) proved that the MCD has much better statistical properties--notably, efficiency--than the MVE. Several other authors recommended the MCD over the MVE (Simpson and Chang (1997)) simulations in Rocke and Woodruff (1996)). However, Rocke and Woodruff (1997) did not recommend either as a stand-alone procedure because of the computational complexity in high dimensions. It is not known how the genetic algorithms perform as stand-alone procedures.

An operations research approach to classification provides some interesting optimization

problems. Fisher's linear discriminant function and the quadratic discriminant extension are well-known statistical methods for classification (Mardia, Kent, and Bibby (1979)). The problem scenario is to assign each case to one of several classes based on vectors with continuous variables. A training sample of cases that consist of the vectors and the known class membership is used to develop the assignment rule. It is not well known among the statistical community that multiple programming methods for these same problems have been considered for some time (Freed and Glover (1981)). In fact, there is an extensive literature for both linear and quadratic methods, with a good summary of references provided by Silva and Stam (1994).

For example, consider a classification into one of two classes based on a two-dimensional vector of independent variables. The problem is illustrated in Figure 5. One multiple programming formulation seeks the line that minimizes the sum of the absolute deviations of the incorrectly classified points. This is the sum of the line segments shown in Figure 5. Alternative formulations seek to minimize the maximum deviation, to minimize the number of incorrectly classified points, and to minimize the deviations for incorrectly classified points as well as maximize the deviations for correctly classified points. These formulations provide reasonable objectives for classification algorithms that extend the traditional statistical objectives. They can typically be solved by a standard linear programming package.

CONCLUDING REMARKS

We have provided a broad overview of optimization problems arising in statistics and quality control with a goal of stimulating more interaction between researchers in these two communities. We hope that this survey helps present some of the frontiers of optimization techniques in statistics. There are possibilities for methodological breakthroughs in areas as diverse as optimal experimental design, economic process monitoring and control, and response surface optimization, to name just a few. New techniques for multi-objective optimization, effective heuristics for combinatorial optimization, and cutting-edge interior point methods for nonlinear programming, seem to offer as many research opportunities as there are combinations of ideas and techniques from the two areas.

ADDED MATERIAL

W. MATTHEW CARLYLE, DOUGLAS C. MONTGOMERY, and GEORGE C. RUNGER

Arizona State University, Tempe, AZ 85287-5906

Dr. Carlyle is an Assistant Professor of Industrial Engineering.

Dr. Montgomery is a Professor of Industrial Engineering. He is a Fellow and Shewhart Medalist of ASQ.

Dr. Runger is an Associate Professor of Industrial Engineering. He is a Senior Member of ASQ.

FIGURE 1. One Step in Hooke and Jeeves' Direct Search Method. Dotted Vector Indicates Previous Step, Dashed Vector Indicates Extension of Previous Step, White Circles Indicate Points Considered, and Solid Vector Indicates Resulting (Best) Step.

FIGURE 2. Design Expert Plot of Non-Convex Region Determined by Three Response Constraints.

FIGURE 3. Predictor-Corrector Step on an Equality Constraint (bold curve) with Objective Contours Shown. Dashed Arrow Indicates True Gradient.

FIGURE 4. Illustration of a Crossover Operation Using Two Binary Chromosomes. The Offspring is Formed from the Head of the First Parent and the Tail of the Second.

FIGURE 5. An Illustration of Classification by Discriminant Function Chosen to Minimize the Sum of the Distances from Incorrectly Classified Points. Two Predictor Variables and Two Classes are Shown.

REFERENCES

AL-SULTAN, K. S. and RAHIM, M. A. (1997). Optimization in Quality Control. Kluwer Academic, Norwell, MA.

ANDERE-RENDON, J.; MONTGOMERY, D. C.; and ROLLIER, D. A. (1997). "Design of Mixture Experiments Using Bayesian D-Optimality". Journal of Quality Technology 29, pp. 451-463.

BOHACHEVSKY, I. O.; JOHNSON, M. E.; and STEIN, M. L. (1986). "Generalized Simulated Annealing for Function Optimization". Technometrics 28, pp. 209-217.

BOOTH, K. H. V. and COX, D. R. (1962). "Some Systematic Supersaturated Designs". Technometrics 4, pp. 489-495.

BOX, G. E. P. (1999). "Statistics as a Catalyst to Learning by Scientific Method Part I--A Discussion" (with discussion). Journal of Quality Technology 31, pp. 16-29.

BOX, G. E. P. and DRAPER, N. R. (1987). Empirical Model Building and Response Surfaces. John Wiley & Sons, New York, NY.

BOX, G. E. P. and LIU, P. Y. T. (1999). "Statistics as a Catalyst to Learning by Scientific Method Part I--An Example" (with discussion). Journal of Quality Technology 31, pp. 1-15.

BOX, G. E. P. and WILSON, K. B. (1951). "On the Experimental Attainment of Optimum Conditions". Journal of the Royal Statistical Society B 13, pp. 1-45.

BUTLER, R. W.; DAVIES, P. L.; and JHUN, M. (1993). "Asymptotics for the Minimum Covariance Determinant Estimator". Annals of Statistics 21, pp. 1385-1400.

CHALONER, K. (1984). "Optimum Bayesian Experimental Designs for Linear Models" Annals of Statistics 12, pp. 283-300. Corrigenda: 1985, 13, p. 836.

CHAMP, C. W. and RIGDON S. E. (1991). "A Comparison of the Markov Chain and the Integral Equation Approaches for Evaluating the Run Length Distribution of Quality Control Charts". Communications in Statistics--Simulation and Computation 20, pp. 191-204.

CIENIAWSKI, S. E.; EHEART, J. W.; and RANJITHAN, S. (1995). "Using Genetic Algorithms to Solve a Multiobjective Groundwater Monitoring Problem". Water Resources Research 31, pp. 399-409.

COOK, R. D. and NACHTSHEIM, C. J. (1980). "A Comparison of Algorithms for Constructing Exact D-optimal Designs". Technometrics 22, pp. 315-324.

COOK, R. D. and NACHTSHEIM, C. J. (1982). "Model Robust, Linear Optimal Designs". Technometrics 24, pp. 49-54.

COPELAND, K. A. F. and NELSON, P. R. (1996). "Dual Response Optimization via Direct Function Minimization". Journal of Quality Technology 28, pp. 331-336.

DAVIES, P. L. (1992). "Asymptotics of Rousseeuw's Minimum Volume Ellipsoid Estimator". Annals of Statistics 15, pp. 1269-1292.

DEL CASTILLO, E. (1996). "Multiresponse Process Optimization via Constrained Confidence

Regions". Journal of Quality Technology 28, pp. 61-70

DEL CASTILLO, E.; MACKIN, P.; and MONTGOMERY, D. C. (1996). "Multiple Criteria Optimal Design of Control Charts". IIE Transactions 28, pp. 467-474.

DEL CASTILLO, E. and MONTGOMERY, D. C. (1993). "A Non-linear Programming Solution to the Dual Response Problem". Journal of Quality Technology 25, pp. 199-204.

DEL CASTILLO, E.; MONTGOMERY, D.C.; and McCARVILLE, D. R. (1996). "Modified Desirability Functions". Journal of Quality Technology 28, pp. 337-345.

DERRINGER, G. and SUICH, R (1980). "Simultaneous Optimization of Several Response Variables". Journal of Quality Technology 12, pp. 214-219.

DUMOUCHEL, W. and JONES, B. (1994). "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model". Techmometrics 36, pp. 37-47.

DUNCAN, A. J. (1956). "The Economic Design of Control Charts Used to Maintain Current Control of a Process". Journal of the American Statistical Association 51, pp. 228-242.

DYER, J. S. (1973). "A Time Sharing Computer Program for the Solution of the Multiple Criterion Problem". Management Science 19, pp. 1379-1383.

FREED, N. and GLOVER, F. (1981). "A Linear Programming Approach to the Discriminant Problem". Decision Sciences 12, pp. 687-74.

FURNIVAL, G. M. and WILSON, R. W. (1974). "Regression by Leaps and Bounds". Technometrics 16, pp. 499-511.

GEOFFRION, A. M.; DYER, J. S.; and FEINBERG, A. (1972). "An Interactive Approach for Multi-Criterion Optimization with An Application to the Operation of an Academic Department". Management Science 19, pp. 357-368.

HAINES, L. M. (1987). "The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear-Regression Models". Technometrics 29, pp. 439-447.

HAWKINS, D. M. (1993). "A Feasible Solution Algorithm for the Minimum Volume Ellipsoid Estimator". Computational Statistics 9, pp. 95-107.

HAWKINS, D. M. (1994). "The Feasible Solution Algorithm for the Minimum Covariance Determinant Estimator in Multivariate Data". Computational Statistics and Data Analysis 17, pp. 197-210.

HAWKINS, D. M. and OLWELL, D. H. (1997). Cumulative Sum Charts and Charting for Quality Improvement. Springer, New York, NY.

HO, C. and CASE, K. E. (1994). "Economic Design of Control Charts: A Literature Review for 1981-1991". Journal of Quality Technology 26, pp. 39-53.

HOERL, A. E. (1964). "Ridge Analysis". Chemical Engineering Symposium Series 60, pp. 67-77.

HOLLAND, J. (1974). "Adaptation in Natural and Artificial Systems". The University of Michigan Press, Ann Arbor, MI.

HOOKE, R. and JEEVES, T. A. (1961). "Direct Search Solution of Numerical and Statistical Problems". Journal of the Association of Computing Machinery 8 2, pp. 212-229.

HORNG, S.-M. (1998). The Use of Genetic Algorithms to Solve Multi-Objective Scheduling Problems on Parallel Machines". Ph.D. Dissertation, Arizona State University, Tempe, AZ.

JOHNSON, M. E. and NACHTSHEIM, C. J. (1983). "Some guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces". Technometrics 25, pp. 271-277.

KEATS, J. B.; DEL CASTILLO, E.; VON COLLANI, E.; and SANIGA, E. M. (1997). "Economic Modeling for Statistical Process Control". Journal of Quality Technology 29, pp. 144-147.

KEATS, J. B.; MISKULIN, J. D.; and RUNGER, G. C. (1995). "Statistical Process Control Scheme Design". Journal of Quality Technology 27, pp. 214-225.

KHURI, A. I. and CONLON, M. (1981). "Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions". Technometrics 23, pp. 363-375.

KHURI, A. I. and CORNELL, J. A. (1996). Response Surfaces: Designs and Analyses, 2nd ed. Marcel Dekker, New York, NY.

KIEFER, J. (1958). "On the Nonrandomized Optimality and the Randomized Nonoptimality of Symmetrical Designs". Annals of Mathematical Statistics 29, pp. 675-699.

KIEFER, J. (1959). "Optimum Experimental Designs" (with discussion). Journal of the Royal Statistical Society B 21, pp. 272-319.

KIEFER, J. (1961). "Optimum Designs in Regression Problems II". Annals of Mathematical Statistics 32, pp. 298-325.

KIEFER, J. (1962). "Two More Criteria Equivalent to D-Optimality of Designs". Annals of Mathematical Statistics 33, pp. 792-796.

KIEFER, J. and WOLFOWITZ, J. (1960). "The Equivalence of Two Extremum Problems". Canadian Journal of Mathematics 12, pp. 363-366.

LASDON, L. S.; FOX, R. L.; and RATNER, M. W. (1974). "Nonlinear Optimization Using the Generalized Reduced Gradient Method". Revue Francaise d'Automatique Informatique Recherche Operationnelle 8, pp. 73-103.

LI, W. W. and WU, C. F. J. (1997). "Columnwise-Pairwise Algorithms with Applications to the Construction of Supersaturated Designs". Technometrics 39, pp. 171-179.

LIN, D. K. J. (1993). "A New Class of Supersaturated Designs". Technometrics 35, pp. 28-31.

LIN, D. K. J. (1995). "Generating Systematic Supersaturated Designs". Technometrics 37, pp. 213-225.

LIN, D. K. J. and TU, W. (1995). "Dual Response Surface Optimization". Journal of Quality Technology 27, pp. 34-39.

LIN, S. and KERNIGHAN, B. W. (1973). "An Effective Heuristic Algorithm for the Traveling-Salesman Problem". Operations Research 21, pp. 498-516.

LIND, E. E.; GOLDIN, J.; and HICKMAN, J. B. (1960). "Fitting Yield and Cost Response Surfaces". Chemical Engineering Progress 56, p. 62.

MARDIA, K. V.; KENT, J. T.; and BIBBY, J. M. (1979). Multivariate Analysis. Academic Press, New York, NY.

McWILLIAMS, T. P. (1994). "Economic, Statistical, and Economic-Statistical Chart Designs". Journal of Quality Technology 26, pp. 227-238.

METROPOLIS, N.; ROSENBLUTH, A.; ROSENBLUTH, M.; TELLER, A.; and TELLER, E. (1953). "Equations of State Calculations by Fast Computing Machines". Journal of Chemical Physics 21, pp. 1087-1092.

MEYER, R. K. and NACHTSHEIM, C. J. (1995). "The Coordinate Exchange Algorithm for Constructing Exact Optimal Experimental Designs". Technometrics 37, pp. 60-69.

MITCHELL, T. J. (1974). "An Algorithm for Construction of D-Optimal Experimental Designs".

Technometrics 16, pp. 203-210.

MONTGOMERY, D. C. (1980). "The Economic Design of Control Charts: A Review and Literature Survey". Journal of Quality Technology 12, pp. 75-87.

MONTGOMERY, D. C. (1982). "The Economic Design of an X Control Chart". Journal of Quality Technology 14, pp. 40-43.

MONTGOMERY, D. C. (1997). Design and Analysis of Experiments, 4th ed. John Wiley & Sons, New York, NY.

MONTGOMERY, D. C. and BETTENCOURT, V. M., JR. (1977). "Multiple Response Surface Methods in Computer Simulation". Simulation 29, pp. 113-121.

MONTGOMERY, D. C.; TALAVAGE, J. J.; and MULLEN, C. J. (1972). "A Response Surface Approach to Improving Traffic Signal Settings in a Street Network". Transportation Research 6, pp. 69-80.

MONTGOMERY, D. C.; TORNG, J. C. C.; COCHRAN, J. K.; and LAWRENCE, F. P. (1995). "Statistically Constrained Economic Design of the EWMA Control Chart". Journal of Quality Technology 27, pp. 250-256.

MURATA, T.; ISHIBUCHI, H.; and TANAKA, H. (1996). "Multi-Objective Genetic Algorithm and its Applications to Flowshop Scheduling". Computers in Industrial Engineering 30, pp. 957-968.

MYERS, R. H. (1999). "Response Surface Methodology--Current Status and Future Directions" (with discussion). Journal of Quality Technology 31, pp. 30-44.

MYERS, R. H. and CARTER, W. H., JR. (1973). "Response Surface Techniques for Dual Response Systems". Technometrics 15, pp. 301-317.

MYERS, R. H. and MONTGOMERY, D. C. (1995). Response Surface Methodology: Process and Product Optimization using Designed Experiments. John Wiley & Sons, New York, NY.

NELDER, J. A., and R. MEAD (1965). "A Simplex Method for Function Minimization". Computing Journal 7, pp. 308-313

NESTEROV, Y. and NEMIROVSKII, A. (1994). Interior-Point Polynomial Algorithms in Convex Programming. Society for Industrial and Applied Mathematics, Philadelphia, PA.

PAGE, E. S. (1954). "Control Charts for the Mean of a Normal Population". Journal of the Royal Statistical Society B 16, pp. 131-135.

PILZ, J. (1991). Bayesian Estimation and Experimental Designs in Linear Regression Models. John Wiley & Sons, New York, NY.

POWELL, M. J. D. (1998). "Direct Search Algorithms for Optimization Calculations". Acta Numerica 7, pp. 287-336

PRABHU, S. S.; RUNGER, G. C.; and MONTGOMERY, D. C. (1995). "A Design Tool to Evaluate Average Time to Signal Properties of Adaptive Control X Charts". Journal of Quality Technology 27, pp. 74-83.

PRABHU, S. S.; RUNGER, G. C.; and MONTGOMERY, D. C. (1997). "Selection of a Subgroup Size and a Sampling Interval for a CUSUM Control Chart". IIE Transactions 29, pp. 451-457.

PUKELSHEIM, F. (1995). Optimum Design of Experiments. Chapman & Hall, New York, NY.

REYNOLDS, M. T.. JR.; AMIN, R. W.; ARNOLD, J. C.; and NACHLAS, J. A. (1988). "Xbar-Charts with Variable Sampling Intervals". Technometrics 8, pp. 51-77.

ROCKE, D. M. and WOODRUFF, D. L. (1996). "Identification of Outliers in Multivariate Data". Journal of the American Statistical Association 91, pp. 1047-1061.

ROCKE, D. M. and WOODRUFF, D. L. (1997). "Robust Estimation of Multivariate Shape and Location". Journal of Statistical Planning and Inference 57, pp. 245-255.

ROUSSEEUW, P. J. (1985). "Multivariate Estimation with High Breakdown Point" in Mathematical Statistics and Applications, Vol. B edited by W. Grossmann, G. Pflug, T. Vincze, and W. Werz. Reidel, Dordect, The Netherlands. pp. 283-297.

ROUSSEEUW P. J. and VAN ZOMEREN, B. C. (1991). "Robust Distances: Simulations and Cutoff Values" in Directions in Robust Statistics and Diagnostics, Part II edited by W. Stahel and S. Weisberg. Springer-Verlag, New York, NY.

RUNGER, G. C. and MONTGOMERY, D. C. (1993). "Adaptive Sampling Enhancements for Shewhart Control Charts". IIE Transactions 25, pp. 41-51.

RUNGER, G. C. and PIGNATIELLO, J. J., JR. (1991). "Adaptive Sampling for Process Control" Journal of Quality Technology 23, pp. 135-155.

SANIGA, E. M. (1989). "Economic Statistical Control Chart Design with an Application to X and R Control Charts". Technometrics 31, pp. 313-320.

SANIGA, E. M.; DAVIS, D. J.; and McWILLIAMS, T. P. (1995). "Economic, Statistical, and Economic-Statistical Design of Attribute Charts". Journal of Quality Technology 27, pp. 56-75.

SCHAFFER, D. J. (1985). "Multiple Objective Optimization with Vector Evaluated Genetic Algorithms". Proceedings of the First International Conference on Genetic Algorithms, pp. 93-100.

SEARLE, S. R.; CASELLA, G.; and McCULLOCH, C. E. (1992). Variance Components. John Wiley & Sons, New York, NY.

SHEWHART, W. E. (1931). Economic Control of Quality of Manufactured Product. Van Nostrand, New York, NY.

SILVA, A. P. D. and STAM, A. (1994). "Second Order Mathematical Programming Formulations for Discriminant Analysis". European Journal of Operational Research 72, pp. 4-22.

SVOBODA, L. (1991). "Economic Design of Control Charts: A Review and Literature Survey (1979-1989)" in Statistical Process Control in Manufacturing edited by J. B. Keats and D. C. Montgomery. Marcel Dekker, New York, NY.

VINING, G. G. and MYERS, R. H. (1990). "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach". Journal of Quality Technology 22, pp. 38-45.

VINING, G. G. (1998). "A Compromise Approach to Multiresponse Optimization". Journal of Quality Technology 30, pp. 309-313.

WEILER, H. (1952). "On the Most Economical Sample Size for Controlling the Mean of a Population". Annals of Mathematical Statistics 23, pp. 247-254.

WELCH, W. J. (1982). "Branch-and-Bound Search for Experimental Designs Based on D-Optimality and Other Criteria". Technometrics 26, pp. 217-224.

WILDE, D. J. and BEIGHTLER, C. S. (1967). Foundations of Optimization. Prentice-Hall, Englewood Cliffs, NJ.

WOODALL, W. H. (1986). "Weakness of the Economic Design of Control Charts". Technometrics 28, pp. 408-409.

WOODALL, W. H. (1987). "Conflicts Between Deming's Philosophy and the Economic Design of Control Charts". Frontiers in Statistical Quality Control 3, pp. 242-248.

WOODALL, W. H. and MONTGOMERY, D. C. (1999). "Research Issues and Ideas in Statistical Process Control". Journal of Quality Technology 31, pp. 376-386.

WOODRUFF, D. L. and ROCKE, D. M. (1993). "Heuristic Search Algorithms for the Minimum Volume Ellipsoid". Journal of Computational Graphical Statistics 2, pp. 69-95.

WOODRUFF, D. L. and ROCKE, D. M. (1994). "Computable Robust Estimation of Multivariate Shape in High Dimension Using Compound Estimators". Journal of the American Statistical Association 89, pp. 888-896.

WYNN, H. P. (1972). "Results in the Theory and Construction of D-Optimum Experimental Designs". Journal of the Royal Statistical Society B 34, pp. 133-147.

DISCUSSION

ERWIN M. SANIGA

University of Delaware, Newark, DE 19716

CARLYLE, Montgomery, and Runger's (CMR) coverage of optimization problems and methods in the field of quality provides a good overview of most of the primary problems and solution methods explored in the literature to this point.

In this discussion, I briefly address three issues related to their paper: 1. finding a solution to the control chart design problem, 2. finding a solution to optimization problems in quality in general, and 3. building models for practical use.

Control chart design is a problem addressed in the literature for more than thirty years (see, e.g., Keats et al. (1997) and related discussion by Box et al. (1997)). In this problem, the objective is to determine the parameters of a control chart such that cost is minimized or profit is maximized according to an economic model, desired average run length (ARL) or average time to signal (ATS) are achieved, or both are achieved simultaneously. These problems are called economic design, statistical design, and economic statistical design, respectively.

Characteristics of this problem include nonlinear cost or profit functions and nonlinear constraints with functions that sometimes do not exist in closed form. An example of the latter is the sampling distribution of the sample range in an R chart. Since sample size is an integer and sampling frequency is a real number (while the width of the control limit can be expressed as either depending upon the type of chart) the economic and economic-statistical design problem is a mixed-integer nonlinear programming problem. The statistical design problem is one where one wishes to find a feasible solution to a set of nonlinear constraints. The resulting constraint satisfaction problem can also be solved using mixed-integer nlp methods.

As hard as it sounds, it really is not too difficult to find very good locally optimal solutions to any of the problems above. Couple that with the fact that since real time solutions are not an issue there has been very little work along the lines of exploring alternate solution methodologies. In our own work we have effectively used such methods as two-stage grid search (Saniga (1977, 1979) and Saniga and Montgomery (1981)), generalized reduced gradient methods (Saniga (1989) and Collani, Saniga, and Weigand (1994)) and enumeration (Saniga (1991) and Saniga, Davis, and McWilliams (1995)). The considerable experimentation we have done in those papers point to the fact that all of the optimization techniques we have

used seem to work very well and it is hard to make a point that one method is preferred to another.

Such is not the case with some of the combinatorial optimization problems, such as some mathematical programming applications to discrimination and classification discussed by Stam (1997) and optimal design, two problems mentioned by CMR. Problems of a specific combinatorial nature sometimes require solution methods built upon a knowledge of the particular problem structure. Combinatorial optimization problems can be an ugly beast and one needs to take advantage of all knowledge when solving them. This requires one to keep abreast of the literature and the new solution breakthroughs as they occur. And while one can argue that these problems are not generally of a real-time nature, it is also true that time is a problem when solutions to even moderate sized problems cannot be found.

Most of the problems CMR mention that cannot be solved analytically or are not of a combinatorial nature can be solved in a number of ways. As we mention above, we have solved a number of different control chart design problems using a variety of methods including writing programs to use methods addressed by CMR. Users not interested in writing their own programs can sometimes find special purpose programs published in the Journal of Quality Technology (these and others are also available on-line on the STATLIB website managed by Vlachos at Carnegie Mellon University (http://www.lib.stat.cmu.edu/jqt/). In addition, many commercial packages are available to solve most of the problems that CMR mention. A review by Fourer (1999) of linear programming packages is broader than that because many of these packages will solve nonlinear problems and integer problems. It is an excellent reference for those interested in using existing programs. Also, those interested in learning more about the advances in mathematical programming regarding solvers, modeling systems, and internet optimization services should read Fourer (1998).

In closing, I wish to bring up a point mentioned by CMR in the context of optimal design. They mention that alphabetically-optimal designs are subject to criticism because they are only optimal in the context of the specified form of the model. That is, it may be better to use a different model than the one specified. I believe the criticism about context can be expanded to many of the other applications of optimization methodology CMR discuss. Moreover, I think there are additional issues one must consider when building models for practical use. Five that come to mind are listed below.

1. As we note, the context must be appropriate. An optimal design for a completely randomized experiment is not really optimal when a split plot experiment is better. Finding an economic control chart design when an organizations' quality suffers in comparison to a competitors may not be as good as finding and implementing a statistical design which guarantees quick signaling of assignable causes.
2. The objective criterion must be appropriately selected.
3. The coefficients must be accurately measured.
4. A good solution must be obtainable, something which is not always true for some quality problems of a combinatorial nature.
5. The model should be robust which means the solution should be good even if the conditions under which the model was built changes.

Careful consideration should be given to all of those issues when using any optimization

method or any model in general for the purpose of quality control and improvement.

Dr. Saniga is a Professor in the Department of Business Administration. His email address is saniga@college.be.udel.edu.

ADDITIONAL REFERENCES

BOX, G. E. P.; COLEMAN, D. E.; FALTIN, F. W.; RYAN, T. P.; SANIGA, E. M.; COLLANI E. V.; WHEELER, D. J.; and WOODALL, W. H. (1997). "A Discussion on Statistically-Based Process Monitoring and Control". Journal of Quality Technology 29 pp. 148-156.

COLLANI, E. V.; SANIGA, E. M.; and WEIGAND, C. (1994). "Economic Adjustment Designs for X-bar Charts". IIE Transactions 26 6, pp. 37-43,

FOURER, R. (1998). "A Survey of Recent Trends in Mathematical Programming Systems". OR/MS Today December.

FOURER, R. (1999). "Linear Programming". OR/MS Today August.

SANIGA, E. M. (1977). "Joint Economically Optimal Design of X-bar and R Control Charts". Management Science 24, pp. 420-431.

SANIGA, E. M. (1979). "Joint Economic Design of X-bar/R Control Charts with Alternate Process Models". AIIE Transactions 11, pp. 254-260.

SANIGA, E. M. (1991). "Statistical Design of Fraction Defective Control Charts and Some Economic Implications". Presented at the 13th Irish Mathematical and Computation Society World Congress on Computation and Applied Mathematics, Dublin, Ireland, July 22-26.

SANIGA, E. M. and Montgomery, D. C. (1981). "Economical Quality Control Policies for a Single Cause System". AIIE Transactions 13, pp. 258-264.

STAM, A. (1997) "Nontraditional Approaches to Statistical Classification: Some perspectives on L[subp]-norm Methods". Annals of Operations Research 74, pp. 1-36.

ENRIQUE DEL CASTILLO

The Pennsylvania State University, University Park, PA 16802

JOHN SEMPLE

Southern Methodist University, Dallas, TX 75275-0333

WE would like to begin by thanking the authors for a very informative paper. They have made a successful attempt to summarize a very broad technical area. As they mention in their paper, optimization problems lie at the core not only of Applied or "Industrial" Statistics, but of statistical methodology in general. Following the authors line of attack to such a formidable topic, we would like to confine our comments to some quite specific optimization problems existing in the areas of quality control and improvement that we have studied in the last few years.

OPTIMIZATION TECHNIQUES IN RESPONSE SURFACE METHODOLOGY

The authors mention optimization problems arising in the "Phase II" or "end game" of response surface methodology (RSM) once curvature has been detected and quadratic responses have been fitted to data. However, important optimization issues also exist in "Phase I" of RSM. In particular, the multiple response problem also exists when operating a process far from its (multiple) optima, when fitted responses are likely to be first order. In this case, the question is how to find a direction of improvement in the controllable factors space for all responses of interest. Del Castillo (1996) proposes to use confidence cones on the direction of

steepest ascent/descent as a guide for finding directions where all or at least the most important responses improve. We would like to ask the authors about their views on this problem, and if they have encountered it in their consulting practice.

"Phase II" RSM optimization problems are the more challenging. Consider the authors' Problem (3) when m = 1. Suppose, without loss of generality, that we wish to minimize the primary response and assume the region of interest is a sphere of radius rho. Thus, we wish to minimize

^Y[subp] (x) = b[sub0] + x'b + x'Bx,

subject to

^Y[subs] (x) = c[sub0] + x'c + x'Cx = T (1)

x'x [less or equal] rho[sup2].

This is the so-called "Dual Response" (DR) problem that has received considerable attention in the literature due in part to its relation to Taguchi's parameter design problems (Vining and Myers (1990)).

General purpose nonlinear programming codes (e.g., GRG2) have been used to solve this type of problem (Del Castillo and Montgomery (1993)), but for indefinite B and C matrices these methods may settle on local optima (Semple (1997)). It is our experience in RSM that for more than three controllable factors, the fitted quadratic responses are usually indefinite; that is, they are saddle functions. Our work has concentrated on methods for the global optimization of such multiple quadratic-response systems, and our current work is considering the case m > 1.

In a series of papers (Semple (1997) and Del Castillo, Fan, and Semple (1997, 1999)), we have developed algorithms based on the so-called trust region methods (Sorensen (1982)) that practically guarantee the location of a global optimum for dual response systems (m = 1). In the course of this investigation two types of problems were identified: degenerate problems and nondegenerate problems. The former are much harder to solve than the latter, but occur less frequently. To understand degeneracy in DR systems, let theta[sub1] and theta[subrho] denote the Lagrange multipliers associated with the two constraints. Semple (1997) showed that a solution generated from the stationarity equation

(B - theta[sub1]C + theta[subrho]I)x = 1/2(theta[sub1]c - b) (2)

for (theta[sub1], theta[subrho]) [element of] Gamma [triple bond] [(theta[sub1], theta[subrho]) : (B - theta[sub1]C + theta[subrho]I) is positive definite and theta[subrho] [greater or equal] 0] with ^Y[subs](x) = T and x'x [less or equal] rho[sup2] is in fact a global minimum of problem. The search in Gamma was originally suggested by Myers and Carter (1973) as an extension of Ridge Analysis techniques. A systematic search of the Lagrange multipliers can be done by initially fixing theta[sub1] and varying theta[subrho], a search related to the parametric trust region problem:

min ^Y[subp](x) - theta[sub1]^Y[subs](x) s. t. x'x [less or equal] rho[sup2].

To find theta[subrho], the following auxiliary function needs to be solved for a given theta[sub1]:

Ø(theta[subrho]) = 1/4 (theta[sub1]c - b)'(B - theta[sub1]C + theta[subrho]I)[sup-2] (theta[sub1]c - b) = [Graphic Character Omitted] gamma[sup2[sub[subi] / (theta[subrho] + lambda[subi])[sup2] = rho[sup2],

where gamma = Q'(theta[sub1]c - b)/2, Q is an orthonormal basis of eigenvectors of B -

theta[sub1]C and the lambda[subi] are their corresponding eigenvalues. Since Newton's method applied to this function may diverge, a transformed equation is actually solved. An existing subroutine, DRSALG, solves the DR problem using a nested search in theta[sub1] until ^Y[subs](x) = T. If the problem is nondegenerate, then the global optimum (theta[sub1], theta[subrho]) [element of] Gamma and DRSALG will converge to it. However, the target is not always achieved on Gamma. This occurs when gamma[sub1] = 0, so the first pole of Ø(theta[subrho]) is missing, and ^Y[subs](x* (theta[sub1])) may therefore exhibit discontinuities. Thus, there may be target values T that will be unattainable and fall inside the discontinuity gap if the search is restricted to Gamma. For this degenerate case, a rotation method called AXIS was proposed and shown to be effective in locating global optima in more than 99% of degenerate cases (Del Castillo, Fan, and Semple (1999)). The DRSALG/AXIS procedure, DR2, is available at http://www.ie.psu/people/faculty/castillo/research.htm for downloading. We point out that the case of an inequality constraint can be solved (and global optimal solutions found with very high probability) by solving the equality constrained case parametrically on T, the target value for the secondary response. Inequality constraints result in solutions very close to the minimum mean square error (MSE) solutions (Del Castillo and Montgomery (1993)) without having to resort to an MSE objective function, which is fourth order if the responses are quadratic.

EXAMPLE

Consider the example in Lin and Tu (1995); they minimize an MSE objective applied to a printing process (a 3-factor problem in Vining and Myers (1990)). The mean and standard deviation functions are quadratic polynomials; thus, the MSE criterion is quartic. Instead of minimizing such a complicated function (without any guarantee of global optimality, a particular concern if the number of factors is larger than in this example), consider solving, instead,

min ^sigma(x) s.t. ^μ(x) = T, x'x [less or equal] 3

for different values of T around the desired value of 500. Using DR2, the results are as in Table 1.

The minimum MSE solution, from Lin and Tu (1995), is x' = (1.5651, -0.7373, 0.0883), which gives ^μ = 495.68 and MSE = 1634.57. The results in Table 1 indicate that the on-target solution (^μ = T = 500) is very close to the minimum MSE solution (contrary to what Lin and Tu state). The advantage of using DR2 in this parametric form is that we are certain of obtaining the global optimum MSE solution.

OPTIMAL DESIGN OF CONTROL CHARTS

There is certainly a large body of literature already existing in economic design of control charts that considers modifications to Duncan's (1956) original model in many different ways. Duncan's model has been extended to several different types of control charts. In contrast, there are almost no real-life applications of economic models for control chart design. The only exception we are aware of is the work of Lorenzen and Vance (1986) who claimed to have applied economically-designed charts at General Motors. At first, the reasons for the lack of applications seem obvious: control charts are relatively easy to use, but economic models are very complicated non-linear optimization models that require hard-to-estimate cost parameters. Keats et al. (1997) elaborate on these and other barriers to implementation of economic

designs.

Let us propose another plausible explanation for the lack of practical applications of economic chart design models. Simply put: Are these models solving a real problem for industry? Examples abound in the operations research literature (where research on optimization theory is usually published) in which invented, hard to solve problems are first proposed and then solved--exactly or heuristically--by elegantly developed techniques. Although we do not claim that the underlying chart design problem is invented or unreal, it seems that the proposed models have failed to capture the actual economics. After more than 40 years of intense research in the area, it is hard to believe that industry has failed to recognize the benefits of using economically-designed control charts.

Perhaps there is a less drastic explanation for the lack of interest in economic designs in practice. As described by the authors, pure statistically-based designs are usually sufficient in practice. Consider designing an X chart where it is necessary to decide a sample size (n), a control limit width (k), and a time between samples (h). In principle, n and k can be uniquely determined based solely on statistical (and not economic) grounds, unless the cost of sampling is so high to warrant very small samples (regardless of h). In such high-cost scenarios, we probably should not be using an X chart to begin with. Constraining the risks of Type I and II errors (or constraining the average run lengths (ARL's)) is equivalent to, and easier than, defining costs associated with these errors, given the shadow price interpretation of Lagrange multipliers associated with the constraints.

The time between samples needs to be decided not only on economic grounds but also in the context of the underlying production system. Besides sampling costs, production specifics (such as the production schedule, the length of the production runs, the production rate, the maintenance schedule, the probability of bad setup operations) enter into the practical determination of h. Some of these issues have been studied in the literature (Weigand (1994); Del Castillo and Montgomery (1996); Keats, Miskulin, and Runger (1997)), but these considerations add complexity to an already complex model and do not consider all the possible production aspects simultaneously.

In our experience sampling costs are the only costs essential to finding a complete chart design. This reasoning motivated Del Castillo, Mackin, and Montgomery (1996) to develop a multicriteria method for designing X charts (the program can be downloaded from the same web address as mentioned above). The inputs to the program are shift size, rate of assignable causes, sampling costs, maximum alpha risk and minimum desired power. The model simultaneously minimizes, subject to statistical constraints,

* the expected number of false alarms (f[sub1]),
* the average time to signal (f[sub2]), and
* the sampling cost per cycle (f[sub3]).

The program presents different solutions (f[sub1], f[sub2], f[sub3]) to the user who enters his or her preferences into the program in an interactive manner, similarly to the description given by Carlyle, Montgomery, and Runger when discussing their Equation (4).

THE GEOMETRIC DISTRIBUTION AND ARL'S

The reason why the ARL is sufficient to compare the performance of charts with geometric-

distributed run lengths is that, if RL[sub1] ~ Geom[sub1] and RL[sub2] ~ Geom[sub2], with ARL[sub1] [less or equal] ARL[sub2], then this implies that RL[sub1] is stochastically larger than RL[sub2]; that is,

P(RL[sub2] [less or equal] r) [less or equal] P(RL[sub1] [less or equal] r)

for all integer r > 0. Thus we are safe in choosing the first chart (if we are talking about the out-of-control performance) or the second chart (if we are talking about the in-control performance) based solely on ARL performance. Given that the geometric run length distribution is very skewed, studying the performance of a single chart based on the ARL criterion can be misleading, but for comparing two or more charts, ARL comparisons are adequate if the run length distributions are geometric. We note that the run length distribution is not geometric when parameters are estimated, even for the relatively simple case of an X chart. For a non-geometric random variable, ordering of means does not necessarily imply stochastic ordering (although the reverse is always true, see Ross (1983, Ch. 8)); thus ARL comparisons should be complemented with more run length distribution information for chart design comparison purposes. From an optimization point of view, when the run lengths are not geometric, we could end up with a multicriteria problem in which we simultaneously look at higher moments of the run length distribution.

Dr. Del Castillo is an Associate Professor in the Industrial & Manufacturing Engineering Department. He is a Member of ASQ. His email address is exd13@psu.edu.

Dr. Semple is an Assistant Professor in the Department of Information Systems and Operations Management.

ACKNOWLEDGMENTS

Dr. Castillo's research was partially funded by NSF grant DMI-9996031 (9623669).

TABLE 1. The DR2 Results for the Vining and Myers' (1990) 3-Factor Problem Using Different T Values Around the Desired Value of 500.

T	x'	^μ(x)	^sigma[sup2](x)	MSE
500	(1.5719, -0.7220, -0.0874)	500	1652	1652
495	(1.5648, -0.7373, -0.0805)	495	1610	1635
496	(1.5663, -0.7343, -0.0863)	496	1618.67	1634.67
497	(1.5676, -0.7313, -0.0866)	497	1626.96	1635.96

ADDITIONAL REFERENCES

DEL CASTILLO, E.; FAN, S. K.; and SEMPLE, J. (1997). "The Computation of Global Optima in Dual Response Systems". Journal of Quality Technology 29 pp. 347-353.

DEL CASTILLO, E.; FAN, S. K.; and SEMPLE, J. (1999). "Optimization of Dual Response Systems: A Comprehensive Procedure for Degenerate and Nondegenerate Problems". European Journal of Operations Research 112, pp. 174-186.

DEL CASTILLO, E. and MONTGOMERY, D. C. (1996). "A General Model for the Economic Optimal Design of X-bar Charts Used to Control Short or Long Run Processes". IIE Transactions 28, pp. 193-201.

Ross, S. M. (1983) Stochastic Processes. John Wiley & Sons, New York, NY.

SEMPLE, J. (1997). "Optimality Conditions and Solutions Procedures for Nondegenerate

Dual Response Systems". IIE Transactions 29, pp. 743-752.

SORENSEN, D. C. (1982). "Newton's Method with A Model Trust Region Modification". SIAM Journal of Numerical Analysis 19, pp. 409-426.

WEIGAND, C. (1994). "Optimal Production Schedule for a K Products Single Machine Problem". Economic Quality Control 9, pp. 91-101.

PETER R. NELSON and MARGARET M. WIECEK

Clemson University, Clemson, SC 29634-0975

WE would like to thank Carlyle, Montgomery, and Runger (CMR) for providing an interesting overview of some of the problems at the interface of quality improvement and optimization. In our discussion we will comment on a few of the ideas presented in their paper that either one or both of us found confusing or on which we have a slightly different perspective. We will also provide some additional references relevant to the interface of statistics and optimization.

DEFINITIONS

Since the vast majority of the people reading this paper will have some expertise in either statistics or optimization, but probably not both, we feel it is important to carefully define any terminology specific to one of these areas. In particular, we think the following definitions may be useful.

The Karush-Kuhn-Tucker conditions identify optimal solutions of constrained minimization (maximization) problems whose objective function and constraint functions are continuous and differentiable. These are necessary conditions for general problems and sufficient conditions for convex (concave) problems. For more details about the conditions see Bazaraa, Sherali, and Shetty (1993).

The efficient frontier is the set of efficient outcomes to a (un)constrained optimization problem with multiple objective functions. A feasible solution of a multiple-objective optimization problem is said to be efficient (Pareto-optimal, noninferior) if trading one efficient solution for another in order to improve an objective results in the deterioration of at least one other objective. Efficient (nondominated) outcomes are images of efficient solutions.

A k-opt heuristic is an algorithm to improve an existing tour of the traveling salesman problem, or it can apply more generally to a problem that seeks a subgraph with all nodes. All subsets of the edge-set of a tour of size k are considered. Each subset is removed in turn to see whether the resulting paths can be recombined to form a tour of lesser cost. Since the number of subsets grows exponentially with k, k-opt for $k > 3$ is seldom used.

A greedy heuristic is one that is based on a "greedy" principle: at each step make the cheapest choice having maximum immediate improvement with little if any regard to future decisions or consequences.

The integrality restriction requires that all data and/or variables of a problem are integral. The restriction is not limiting for most applications because one can always transform rational data to integer data by multiplying the rational data by a suitably large number.

A breadth-first search is a graph searching strategy based on branching to generate all the nodes at the current depth of the graph before generating any that are further down.

A depth-first search is a graph searching strategy based on generating nodes as far down as

possible and backing up one node to initiate a new probe when no new nodes can be found at deeper levels.

The breakdown point of an estimator is the smallest fraction of the observations that one would have to replace in order to cause the estimator to exceed any bound. For example, the sample median is a 50% breakdown estimator (the best possible breakdown point) since one would have to replace at least half the observations in order to ensure the middle value was among them.

A rotatable experimental design is one such that the variance of the predicted response (using the design) depends only on the distance from the center point and is, therefore, unchanged if the design is rotated about its center point.

It was unclear to us what was meant by "an operations research approach to classification" and by a "multiple programming formulation". Also, we did not find the statement that Prabhu, Runger, and Montgomery (1995) provide a design tool to be particularly informative. Similarly, the description of the methods of Lin (1995) and Li and Wu (1997) as being similar to the local search algorithms employed in other related combinatorial optimization problems provided little insight.

ECONOMIC DESIGN AND ALPHABETIC OPTIMALITY

With regard to the economic design of control charts, CMR suggest that one possible reason they have found little industrial use is that practitioners may be intimidated by the number of cost and other system parameters that must be estimated. An alternative possibility is that such information is virtually never attainable. While it does not necessarily follow that lack of industrial use should suggest abandoning a research area, their analogy with alphabetic optimality does not seem appropriate since lack of computer implementation is not the issue with economic design of control charts. CMR also suggest that alphabetic optimality of experimental designs was one of the first areas in quality control and applied statistics to attract the attention of industrial engineers and operations research analysts, but we have been unable to verify this. Rather it seems that design of control charts and acceptance sampling plans provided the initial problems of mutual interest.

THE EM ALGORITHM

While CMR do mention maximum likelihood estimation in regression and analysis of variance leading to complicated optimization problems, they do not mention the interesting problems that arise when looking for maximum likelihood estimates (MLE's) of the parameters of a mixture of normal distributions. That problem is often solved using the EM algorithm proposed by Demster, Laird, and Rubin (1977). The algorithm was named for its two steps at each iteration: the E step replaces the missing values in the incomplete data set with their expectations, and the M step derives the estimates by maximizing the likelihood function of the "pseudo complete" data. In the mixture-of-normals problem the "incompleteness" in the data is due to not knowing which component of the mixture each observation came from. While the EM algorithm has the disadvantage of converging slowly, it has the advantage that at each iteration the likelihood function is guaranteed to be nondecreasing. Modifications to the EM algorithm to improve its performance were suggested by Meng and Rubin (1993) and Rai and Matthews (1993). In the

context of searching for MLE's for the parameters of a mixture of two normal distributions, Finch, Mendell, and Thode (1989) discussed the general problem of assessing the adequacy of a numerical search for a global optimum.

MULTI-OBJECTIVE PROBLEMS

In their discussion of optimization techniques applicable in statistics and quality control, CMR focus on single-objective nonlinear or combinatorial optimization and multi-objective optimization. The latter is of particular interest in multiple-response optimization of designed experiments, which gives rise to problems with highly nonlinear, or multimodal, multiple objective functions. As CMR mention, the weighted-sum scalarization, which converts the original problem with a vector-objective function to a single-objective (scalar) problem, is a classical approach to multi-objective problems. This approach, however, works much better than CMR imply. It works very well for problems with concave (convex) objective functions maximized (minimized) over a convex feasible set, and it can be used to trace almost the whole efficient frontier. Geoffrion (1968) showed that a feasible point is properly efficient if and only if it is an optimal solution of the weighted-sum problem for some positive weights. Thus, except for those efficient points that are not properly efficient, that is, for which finite trade-offs do not exist, the weighted-sum approach can be used to obtain the efficient frontier.

Due to their complexity, many multi-objective problems are treated with heuristic algorithms, and CMR discuss this class of methods. Their references on the use of genetic algorithms, however, are problem oriented (scheduling, groundwater monitoring). Two references offering general insight into this subject are Srinivas and Deb (1994) and Fonseca and Fleming (1995). Since CMR discuss simulated annealing only in the single-objective context, it is worth adding that Serafini (1992); Ulungu, Teghem, and Fortemps (1995); and Czyzak and Jaszkiewicz (1998) developed simulated annealing approaches to multi-objective combinatorial problems.

Methods other than those mentioned by CMR can be employed in multi-response optimization for generating the efficient frontier. The weighted-Tchebycheff approach (see Steuer (1986)) can successfully trace the efficient frontier of a very general multi-objective problem, and the method proposed by Kostreva, Ordoyne, and Wiecek (1992) was specially developed for problems with polynomial objective and constraint functions. More recent approaches attempt to approximate the efficient frontier rather than generate its individual points. Approximations based on exact algorithms have been developed for bi-objective problems (see Jahn and Merkel (1992) and Payne (1993)). However, due to the complexity of the frontier in higher dimensions, approximations for higher-dimensional problems have been based on heuristics (see Gandibleux, Mezdaoui, and Freville (1997) as well as the references on genetic algorithms and simulated annealing).

ADDITIONAL SOFTWARE AND OTHER REFERENCES

A good discussion of Taguchi's robust parameter design ideas can be found in Kackar (1985), and a particularly enlightening discussion of some of the statistical shortcomings is given by Hunter (1985). Wynn (1970) is another reference associated with exchange algorithms. While almost everyone is familiar with EXCEL, specific references for Design-Expert (1996) and S-Plus (1997) are probably useful. More information on simulated annealing can be found in

Volume 8, Nos. 3 & 4 (1988) of the American Journal of Mathematical and Management Sciences. Of particular interest are the papers by Meyer and Nachtsheim (1988) and Collins, Eglese, and Golden (1988). The first of these discusses constructing exact D-optimal designs using simulated annealing, and the second one provides an annotated bibliography on simulated annealing of almost 100 pages. The book by Goldberg (1989) discusses genetic algorithms. Fourer (1999) and Nash (1998) provide surveys of software available for linear programming and nonlinear programming, respectively. The February 1997 issue of OR/MS Today (a publication of the the Institute for Operations Research and the Management Sciences) provides a survey on add-in software for spreadsheets, including routines for optimization, linear programming, nonlinear programming, and genetic algorithms.

CONCLUSION

Recent scholarly events give evidence of the interest in statistics within the operations research and optimization community. Conference talks on multi-objective formulations of some statistical problems were given by Wellington (1999) and Narula (1999), who claimed that useful insights can be obtained when classical statistical topics such as point and interval estimation, hypothesis testing, acceptance sampling, regression analysis, and discriminant analysis are posed as multi-objective problems.

Finally, CMR have limited their discussion to certain aspects of optimization techniques in statistics. The reverse is also possible; that is, statistical techniques can be used to improve optimization procedures. See, for example, Parsons and Johnson (1997).

Dr. Nelson is a Professor in the Department of Mathematical Sciences. He is a Fellow of ASQ. His email address is npeter@math.clemson.edu.

Dr. Wiecek is an Associate Professor in the Department of Mathematical Sciences.

ADDITIONAL REFERENCES

BAZARAA, M. S.; SHERALI, H. D.; and SHETTY, C. M. (1993). Nonlinear Programming, Theory and Algorithms. John Wiley & Sons, New York, NY.

COLLINS, N. E.; EGLESE, R. W.;and GOLDEN, B. L. (1988). "Simulated Annealing--An Annotated Bibliography". American Journal of Mathematical and Management Sciences 8, pp. 209-307.

CZYZAK, P. and JASZKIEWICZ, A. (1998). "Pareto Simulated Annealing--A Metaheuristic Technique for Multiple-Objective Combinatorial Optimization". Journal of Multi-Criteria Decision Analysis 7, pp. 34-47.

DEMSTER, A. P.; LAIRD, N. M.; and RUBIN, D. B. (1977). "Maximum Likelihood From Incomplete Data Via the EM Algorithm". Journal of the Royal Statistical Society B 39, pp. 1-38.

DESIGN-EXPERT (1996). Version 5.0.7, Stat-Ease, Minneapolis, MN (www.statease.com).

FINCH, S. J.; MENDELL, N. R.; and THODE, H. C., JR. (1989). "Probabilistic Measures of Adequacy of a Numerical Search for a Global Maximum". Journal of the American Statistical Association 84, pp. 1020-1023.

FONSECA, C. M. and FLEMING, P. J. (1995). "An Overview of Evolutionary Algorithms in Multiobjective Optimization". Evolutionary Computation 3, pp. 1-16.

FORREST, S. (1993). "Genetic Algorithms: Principles of Natural Selection Applied to

Computation". Science 261, pp. 872-878.

FOURER, R. (1999). "Linear Programming". OR/MS Today 26 (4), pp. 64-71.

GANDIBLEUX, X.; MEZDAOUI, N.; and FREVILLE, A. (1997). "A Tabu Search Procedure for MultiObjective Combinatorial Optimization Problems" in Advances in Multiple Objective and Goal Programming edited by R. Caballero, F. Ruiz, and R. E. Steuer, Springer-Verlag, Berlin. pp. 291-300.

GEOFFRION, A. M. (1968). "Proper Efficiency and the Theory of Vector Optimization". Journal of Mathematical Analysis and Applications 3, pp. 618-630.

GOLDBERG, D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA.

HUNTER, J. S. (1985). "Statistical Design Applied to Product Design". Journal of Quality Technology 17, pp. 210-221.

JAHN, J. and MERKEL, A. (1992). "Reference Point Approximation Method for the Solution of Bicriteria Nonlinear Optimization Problems". Journal of Optimization Theory and Applications 74, pp. 87-103.

KACKAR, R. N. (1985). "Off-Line Quality Control, Parameter Design, and the Taguchi Method". Journal of Quality Technology 17, pp. 176-188.

KOSTREVA, M. M.; ORDOYNE, T. J.; and WIECEK, M. (1992). "Multiple-objective Programming with Polynomial Objectives and Constraints". European Journal of Operational Research 57, pp. 381-394.

MEYER, R. K. and NACHTSHEIM, C. J. (1988). "Constructing D-Optimal Experimental Designs by Simulated Annealing". American Journal of Mathematical and Management Sciences 8, pp. 329-359.

MENG, X.-L. and RUBIN, D. B. (1993). "Maximum Likelihood Estimation Via the EMC Algorithm: A General Framework". Biometrika 80, pp. 267-278.

NARULA, S. C. (1999). "Multicriteria and Statistics?". Presented at the 15th IFORS Conference, Beijing, China, August 16-20.

NASH, S. G. (1998). "Nonlinear Programming". OR/MS Today 26 (3), pp. 36-45.

PARSONS, R. and JOHNSON, M. E. (1997). "A Case Study in Experimental Design Applied to Genetic Algorithms with Applications to DNA Sequence Assembly". American Journal of Mathematical and Management Sciences 17, pp. 369-396.

PAYNE, A. N. (1993). "Efficient Approximate Representation of Bi-objective Trade-off Sets". Journal of the Franklin Institute 330, pp. 1219-1233.

RAI, S. N. and MATTHEWS, D. E. (1993). "Improving the EM Algorithm". Biometrics 49, pp. 587-591.

S-PLUS (1997). MathSoft, Version 4.5, Mathsoft, Inc., Cambridge, MA (www.mathsoft.com).

SERAFINI, P. (1992). "Simulated Annealing for Multiple Objective Optimization Problems" in Multiple Criteria Decision Making: Expand and Enrich the Domains of Thinking and Application edited by G. H. Tzeng, H. F. Wang, V. P. Wen, and P. L. Yu. Springer, New York, NY. pp. 283-292.

SRINIVAS, N. and DEB, K. (1994). "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms". Evolutionary Computation 2, pp. 221-248.

STEUER, R. E. (1986). Multiple Criteria Optimization: Theory, Computation, and Application.

John Wiley & Sons, New York, NY.

ULUNGU, E. L. B.; TEGHEM, J.; and FORTEMPS, P. (1995). "Heuristics for Multiobjective Combinatorial Optimization Problems by Simulated Annealing" in MCDM: Theory and Applications edited by J. Gu, G. C. Q. Wei, and Sh. Wang. Scitech Information Services, pp. 228-238.

WELLINGTON, J. F. (1999). "Multiple Criteria Formulations of Some Statistical Problems". Presented at the Spring 1999 INFORMS Meeting, Cincinnati, OH, May 2-5.

WYNN, H. P. (1970). "The Sequential Generation of D-Optimal Experimental Designs". Annals of Mathematical Statistics 41, pp. 1655-1664.

MARK E. JOHNSON

University of Central Florida Orlando, FL 32816-2370

ANYON who is a bona fide practitioner in the area of quality control and process improvement (or more generally, applied statistics) would have faced the sometimes daunting task of solving optimization problems. Carlyle, Montgomery, and Runger (CMR) have provided us a virtual travelogue of their experiences with optimization methods on their voyages through many venues populated with real problems involving industrial applications. Their journeys appear seamless--identifying the appropriate off-the-shelf technique effortlessly at every destination. This may be a testament to their insight in selecting the appropriate scheme or perhaps they are overly modest in downplaying any difficulties that they have encountered. In contrast, my own experiences in optimization/statistical applications would be best described as rocky, with endless potholes, speed bumps, and detours in an odyssey to obtain or to refine an optimization scheme that consistently works. The advantage, of course, of the more tortured journey is the challenge in overcoming the obstacles in hopes of eventually relating the lessons-learned to future travelers.

The analysis of light water reactor failure data (Booker, Easterling, and Johnson (1981)) provided an area of application early in my career that led to a detailed consideration of optimization methods. By this time, the Cox proportional hazards model (Cox (1972)) had blossomed as the method for observational data in epidemiological studies, and it was natural to try to apply it to reactor reliability problems for which failures (at least at that time) were exceedingly rare. To cut to the chase, in many data set situations parameter estimation algorithms would lead to the dreaded outcome "Newton-Raphson did not converge." No amount of refinement as to the starting values corrected the problem (grid search, preliminary Nelder-Mead, etc.), which was also very discouraging. Eventually, working with Maurice Bryson (Bryson and Johnson (1981)), we realized that there were many circumstances in optimizing the partial likelihood associated with the Cox model that led to non-convergence owing to optimal parameter estimates with limiting infinite values! Hence, Newton-Raphson should not converge! This greatly complicated a subsequent Monte Carlo study of the small sample performance of the Cox model (Johnson et al. (1982)), which in turn called into question the suitability of it in some circumstances. Not too surprisingly, the biostatistics community has not been enthusiastic about this work which raises practical issues with the implementation of the proportional hazards model. CMR do not explicitly mention reliability as an area involving their personal experiences with optimization schemes. Have they had such encounters and if so, would they comment upon them?

CMR cite our work with simulated annealing (Bohachevsky, Johnson, and Stein (1986)). Bohachevsky originally called my attention to this method following the landmark paper by Kirkpatrick, Gelatt, and Vecchi (1983). This Science article inspired numerous forays into consideration of simulated annealing. Our efforts first focussed on a marvelous and challenging optimization problem posed by Bates (1985). Bates went to Herculean efforts to find a global optimum to a problem using the fundamental approach of Box and Wilson (1951) applied to a nonlinear response function. Our variant of simulated annealing (with an empirical cooling schedule) led to what appears to be the global optimum for this problem. Extensive sensitivity analyses reported in the paper offered additional compelling evidence that our variant of simulated annealing had in fact meandered to the optimal solution. We have also been encouraged by the reactions of cohorts at other universities, such as in a global optimization seminar in 1999 at the University of Colorado at Denver arranged by Dr. Karen Kafadar. It can be somewhat surprising that such a "crude" or "simplistic" scheme such as simulated annealing can work, as it is little more than a biased random walk. Bohachevsky noted the possibility of couching simulated annealing in a gambler's ruin framework to gain insight into its improbably good performance (Bohachevsky, Johnson, and Stein (1992, 1995)). Do CMR have any specific impressions of simulated annealing either based upon personal experience or observations of industrial applications?

In their comprehensive review of various optimization schemes applied to industrial applications, CMR mention genetic algorithms which offer a rather unusual approach to optimization by maintaining a population of candidate solutions and letting it evolve through reproduction, mutation, and ultimately survival-of-the-fittest solutions. The applications cited tend to involve classical operations research applications such as scheduling problems. Recently, I collaborated with Rebecca Parsons (Parsons and Johnson (1997)) to apply RSM techniques to optimize the performance of genetic algorithms in a DNA sequence assembly problem. This approach led to the setting of tuning parameters far afield of the "conventional wisdom" values in molecular biology and led to excellent solutions of previously intractable problems. In particular, we found that the population size could be reduced drastically with a corresponding reduction in execution time. Along these lines, I am again interested, if CMR would comment on the use of optimization methods (especially RSM) in the general area of software engineering. This seems to be an area that is begging for attention, but is difficult since the response functions tend not to emulate physical measurements of manufactured projects.

Obviously, the CMR article provided me a timely excuse to revisit my own forays into optimization in the context of statistical applications. I would like to close by addressing a concern as to the likelihood that the CMR article will attain the goal of stimulating more interaction between the statistics and operations research communities. I would conjecture that the operations research (OR) crowd reads JQT about as much as we read the journals Operations Research or Management Science. Given that the OR community does discover this article, what is in it for them? A very positive feature of the CMR article is the reporting of the widespread use of optimization methods. What exactly are the open questions at the statistics/operations interface that could attract collaborations with our friends in OR, management science, or industrial engineering departments? Both communities have their pet problems that continue to get attention. The traveling salesman problem and Rosenbrock's

function continue to draw interest by optimizers. The analysis of Fisher's iris data and Brownlee's stack loss data may never be completely exhausted. I wonder if there is an interesting problem at the interface of optimization and quality control/process improvement that could draw both groups together. Given CMR's track record, I am confident that they can identify such a problem (or problems), thereby establishing a track toward their stated goal.

Dr. Johnson is Professor in the Department of Statistics. His email address is mejohnso@mail.ucf.edu.

ADDITIONAL REFERENCES

BATES, D. (1983). "The Derivative of X'X and Its Uses". Technometrics 25, pp. 373-376.

BOHACHEVSKY, I. O.; JOHNSON, M. E.; and STEIN, M. L. (1992). "Stochastic Optimization and the Gambler's Ruin Problem". Journal of Computational and Graphical Statistics 1, pp. 367-384.

BOHACHEVSKY, I. O.; JOHNSON, M. E.; and STEIN, M. L. (1995). "Generalized Simulated Annealing with Applications to Chemistry" in Simulated Annealing Adapted to Chemistry edited by J. Kalivas, pp. 3-24.

BOOKER, J.; EASTERLING, R. G.; and JOHNSON, M. E. (1981). "Proportional Hazards Models for Improving the Analysis of Light Water Reactor Component Failure Data". International ANS/ENS Topical Meeting on Probabilistic Risk Assessment, Port Chester, NY.

BRYSON, M. C., and JOHNSON, M. E. (1981). "The Incidence of Monotone Likelihood in the Cox Model". Technometrics 23, pp. 381-383.

COX, D. R. (1972). "Regression Models and Life Tables (with Discussion)". Journal of the Royal Statistical Society, B 34, pp. 187-220.

JOHNSON, M. E.; TOLLEY, H. D.; BRYSON, M. C.; and GOLDMAN, A. S. (1982). "Covariate Analysis of Survival Data: A Small-Sample Study of Cox's Model". Biometrics 38, pp. 685-698.

KIRKPATRICK, S.; GELATT, C. D., JR.; and VECCHI, M. P. (1983). "Optimization by Simulated Annealing". Science 220, pp. 671-680.

PARSONS, R. and JOHNSON, M. E. (1997). "A Case Study in Experimental Design Applied to Genetic Algorithms with Applications to DNA Sequence Assembly". American Journal of Mathematical and Management Sciences 17, pp. 369-396.

RESPONSE

W. MATTHEW CARLYLE, DOUGLAS C. MONTGOMERY, and GEORGE C. RUNGER

IN our paper, we have attempted to outline some of the problems on the boundaries of the fields of quality control/improvement and optimization. Our intent was to initiate discussion and provide a starting point for further research. We feel as if we have had some success towards attaining this goal, and we would like to thank all of the discussants for their participation. Their insightful remarks significantly improved the presentation of the material, and provided exactly the kind of valuable discussion we were hoping for. Several of the discussants were also referees for the original version of the manuscript, and we are appreciative of their helpful criticism on the previous version of the paper.

Of course, any omissions are due to our preferences, and we encourage our colleagues to

initiate conversations regarding their favorite areas, as well. For example, Johnson reminds us that the field of reliability provides many interesting problems, and also provides examples from his own experience of successful uses of contemporary heuristics. Saniga provides valuable references to surveys and reviews of optimization software, (notably absent from our paper), as well as pointers to his experiences with several co-authors with difficult optimization problems. We would like to take this opportunity to answer some of the points raised in the responses to our article and to expand briefly upon some of our experiences.

Nelson and Wiecek mention that the weighted-sum approach for multiple response problems works well in many instances and can be used to trace out the efficient frontier. While this is true for problems where each response (or objective) is to be extremized (for instance, in the case of economic or time-based objectives), in many interesting problems in quality control some of the responses are to be restricted to certain operating ranges. The multiple response optimization problem in response surfaces is a classic example.

Our main intent in this section was not to malign the weighted-sum approach, but instead to point out the versatility of the desirability function approach, which can (with suitable choices for the ranges, midpoints, and exponents) subsume many of the older techniques. Indeed, we feel that this is a clear example of a problem that has caused considerable debate among optimization enthusiasts for quite some time and for which the statistics community has developed a very general, powerful tool.

Johnson poses a question regarding our experiences with simulated annealing (SA), and we enlarge our reply to include genetic algorithms (GA) as well. Our principal experiences with both of these methods are very similar; the concepts are easy to explain in the classroom, or to a research group, so that home-brewed algorithms are solving complex problems in a matter of days or weeks. After several experiments, the students or the researchers find that they spend a lot of time "tuning" the parameters, and their frustration grows as they encounter two difficulties. First, "obvious" choices for the parameters don't work. In GA, the most common mistakes are to build too much intelligence in the initial population, avoiding entire regions of the solution space and to use too small of an initial population in the interests of computation time. Second, good parameter choices for one problem are almost worthless for another. In SA implementations, one cooling schedule can work well for one scheduling problem, but get stuck in (very poor) local optima for problems in a very similar type of problem. (Or even for a different instance of the same type of problem!)

The solution, of course, is the one that Johnson himself proposes. Our students all take courses on experimental design, and most have sufficient exposure to response-surface methodology to enable them to set up a sequence of experiments to find good parameter values for their heuristics and, therefore, find good solution methodologies for their problems. Of course, as they perform their experiments, they end up learning rules of thumb that work for their problem domain, reducing the number of completely unknown parameters and speeding up their search.

Del Castillo and Semple observe that most of our comments on the multiple response optimization problem are confined to the second phase of RSM, where the experimenter is focusing on finding optimum solutions within a region of interest. They (quite correctly) note that the problem also occurs in phase one, where the experimenter is trying to move from the

starting conditions for the system towards the region of the eventual optimum. We have found that the desirability function approach can also be applied in conjunction with steepest ascent in this situation, so that the experimenter is essentially finding a path of steepest ascent in desirability space. An important issue in this problem is dealing with two, say, responses that have paths of steepest ascent that are at large angles to each other, so that the best direction for one response is very nearly the worst direction for the other. There is often little that can be done about this situation from an optimization perspective. The solution is usually of an engineering or scientific nature, requiring a change in some systems aspect of the problem, such as alternative materials, different processing or new design/formulation of the product. Since steepest ascent is frequently performed in development work, this is often a feasible approach. On the other hand, finding out that there is a problem of this type once the full-scale process is in operation can potentially be catastrophic. This is another argument for starting statistically designed experiments very early in product and process design.

The comment that Johnson made about the effort-lessness implied by our paper brought a smile to our faces. Of course, we have committed the classic error of describing clean results without elaborating on the torturous path to which Johnson alludes. We have certainly encountered our share of problems along the way that did not yield to the obvious method(s), and if we didn't enjoy such challenges we would probably not be in this field. In the paper we hardly mention one of the biggest difficulties, and it is one that plagues us constantly. Especially in the field of combinatorial optimization, the importance of building a good model cannot be overemphasized. We can't begin to count the number of models we have built, and worked at for days, only to find that a reformulation yields a rapid solution. Space restrictions and the lack of a complete answer do not permit us to describe the "right" way to build models, and we feel that a complete guide will probably never be written.

Finally, we close with a response to Johnson's question as to whether the operations research community reads the Journal of Quality Technology or will get anything out of our paper. This is a very good point; this paper was intentionally targeted towards that part of the statistics community interested in using optimization methods for their applications. To this end, we have given brief descriptions of problems already familiar to practitioners and researchers alike, and then elaborated on the general ideas behind some of the most popular optimization techniques applicable to these problems.

For such a survey to be attractive to the optimization community we would need to shift the focus somewhat and elaborate on the details of the problems, and then use less space describing the methods that have already been employed along with some indications of their success. Problem details could include the specific characteristics of the functions involved, the forms of the constraints, the "standard" version of the problems encountered in practice, and so forth. In this way, each community would have their side of a bridge to the other, and if these two documents were reasonably parallel in their construction, they could serve as a starting point for many valuable collaborations. We appreciate the fact that the biggest obstacle would probably be the differing vocabularies of the two communities, but given our common mathematical background we feel that reasonable compromises could be made.

The first author, whose academic roots lie firmly planted in optimization, presented our paper at the Fall Technical Conference. Perhaps one of the other authors should make a modified

version of this presentation to the optimization community, and thereby complete the circuit.

Source: Journal of Quality Technology, Jan2000, Vol. 32 Issue 1, p32, 2p
Item: 2729419