CONFIDENTIAL

Under Review for Publication

A General Approach to Confidence Regions for Optimal Factor Levels of Response Surfaces

John J. Peterson[1], Suntara Cahya[2], and Enrique del Castillo[2]

[1]Statistical Sciences Department, UW281A, GlaxoSmithKline Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406, U.S.A. John_Peterson@sbphrd.com

[2]Department of Industrial & Manufacturing Engineering, The Pennsylvania State University, University Park, Pennsyvania 19802, U.S.A.

SUMMARY. For a response surface experiment, an approximate hypothesis test and an associated confidence region is proposed for the minimizing (or maximizing) factor-level configuration. Carter at al. (1982) show that confidence regions for optimal conditions provide a way to make decisions about "therapeutic synergism". The response surface may be constrained to be within a specified, bounded region. These constraint regions can be quite general. This allows for more realistic constraint modeling and a wide degree of applicability, including constraints occurring in mixture experiments. The usual assumption of a quadratic model is also generalized to include any regression model that is linear in the model parameters. An intimate connection is established between this confidence region and the Box-Hunter (1954) confidence region for a stationary point. As a byproduct this methodology also provides a way to construct a confidence interval for the difference between the optimal mean response and the mean response at a specified factor level configuration. The application of this confidence region is illustrated with two examples. Extensive simulations indicate that this confidence region has good coverage properties.

KEY WORDS: Combination chemotherapy; Constrained optimization; Controlled-release drug formulation; Mixture experiments; Survival data analysis; Therapeutic synergy.

1. Introduction

In response surface experiments the investigator is often interested in making inferences about a minimizing (or maximizing) factor-level configuration within a specified, bounded region. In this paper we will discuss the minimizing point, the inference for the maximizing point being analogous. Response surface studies play an important role in optimization problems in applied biological and pharmaceutical research. A search of the MEDLINE® reference databases will bring up hundreds of articles involving response surface optimization in diverse areas such as agricultural science, anesthesiology, biotechnology, preclinical cancer chemotherapy, cardiology, dairy science, environmental science, ecology, food science, marine biology, microbiology, neurosciences, nutrition, pharmaceutical sciences, pharmacology, physiology, poultry science, toxicology, virology, and water science. Many of these articles deal with optimization and some make use of previously published methodology to compute confidence regions for the optimal response surface conditions. The book by Carter, Wampler, and Stablein (1983) is entirely devoted to using response surface methodology to find optimal dose combinations for cancer chemotherapy. Stablein et al (1990) give a sobering account of how clear determination of the conditions for the optimal response is important in that effective treatment therapies may be missed by utilization of suboptimal doses or dose combinations. They also point out that confidence regions for constrained optima may provide an effective way to avoid experimentation too close to toxic treatment levels.

An exact confidence region for the (unconstrained) stationary point for a response surface was given by Box and Hunter (1954) (referred to hereafter as BH). However, unconstrained estimation of a stationary point is often not well aligned with the practical needs to make inferences for a minimizing point within the region of the experimentation. In some cases the stationary point corresponds to a saddle point rather than a minimum in the experimental region. Even if stationary and minimizing points agree, it is sometimes the case that this point is outside of the region of experimentation.

Stablein, Carter, and Wampler (1983) (hereafter referred to as SCW) made an important first step in addressing the uncertainty of an optimal point within an experimenter's region, which by practicality must always be constrained. They modified the BH (unconstrained) stationary-point approach to address the stationary point of a Lagrange multiplier problem. For the quadratic response surface model subject to a quadratic constraint function, the work of Myers and Carter (1973) also allows one to identify the constrained minimizing point among multiple stationary points in the Lagrange multiplier system. However, as will be shown in Section 2, the Lagrange multiplier approach has some technical difficulties which impose substantial restrictions upon both modeling of the response surface and the constrained experimental region.

Carter et al. (1982) introduced a key biomedical application of confidence regions by showing how they can be used to statistically assess the notion of "therapeutic synergism" (Venditti, et al.. 1956, Mantel, 1974). In dose combination studies two treatments are said to be therapeutically synergistic if there exists a dose combination of both treatments that is superior to each of the best individual treatment doses. This notion can be extended to three or more treatments. Carter et al. (1982) proposed the idea of using a confidence region for the optimal dose combination as a way to test for therapeutic synergism. If the confidence region for the optimal dose combination excludes all zero dose treatment combinations, then there is statistically significant evidence that all of the treatments are therapeutically synergistic. This idea can be generally applied to situations where the unconstrained optimum is outside of the experimental region, provided a general approach to constructing confidence regions for constrained optima is available.

The area of mixture experiments poses a particular challenge for constructing confidence regions for response surface optima in that such experimental regions are naturally always constrained. Since many mixture experiments involve substance blending for response enhancement, it may often be desirable to test for the above notion of synergism using a confidence region for the constrained optimal factor-level combination. Mixture experiments pose a further challenge in that the models employed sometimes utilize exotic functions of the factor variables as covariates, instead of a standard quadratic model (see for example Cornell, 1981, chap. 6).

This paper reviews the previous methodology on confidence regions associated with response surface optima in the next section. Following that, a more general approach to constructing confidence regions for optimal factor conditions will be introduced in Section 3, with some computational enhancements discussed in Section 4. Two examples of this approach will be given in Section 5, one being a comparison using SCW's preclinical cancer chemotherapy example and one involving a mixture experiment for optimizing a controlled-release drug formulation. In Section 6, the nature of the coverage probabilities associated with this confidence region is discussed. A summary is given in Section 7.

2. Review.

Confidence regions for optimal factor levels in response surface methodology have typically employed the quadratic model

$$Y = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} + e \ , \qquad (2.1)$$

where Y is the response variable and e has a normal distribution with mean 0 and variance σ^2. Here, β_0 is the intercept term, $\boldsymbol{x}$ is a kx1 vector of factor levels, β is a kx1 vector of regression coefficients, and $\boldsymbol{B}$ is a kxk symmetric matrix of regression coefficients with i^{th} diagonal element equal to β_{ii} and the $(i, j)^{th}$ off-diagonal element equal to $½\beta_{ij}$. If $\boldsymbol{x}_0$ is a stationary point of the response surface in (2.1), then H_0: $\beta + 2\boldsymbol{B}\boldsymbol{x}_0 = 0$ is true. BH show that a 100(1-α)% confidence region for $\boldsymbol{x}_0$ is the set of all $\boldsymbol{x}$ such that

$$\widehat{\delta}_{\boldsymbol{x}}'\hat{\boldsymbol{V}}_{\boldsymbol{x}}^{-1}\widehat{\delta}_{\boldsymbol{x}} \le kF(1-\alpha;k,\nu),$$

where $\widehat{\delta}_{\boldsymbol{x}} = \widehat{\beta} + 2\widehat{\boldsymbol{B}}\boldsymbol{x}$, $\widehat{\beta}$ and $\widehat{\boldsymbol{B}}$ are least squares estimates β and $\boldsymbol{B}$ respectively, and $F(1-\alpha,k,\nu)$ is the upper 100(1-α) percentile of the F-distribution with k and $\nu = (n\text{-}p)$df. Here, $\widehat{\boldsymbol{V}}_{\boldsymbol{x}}$ is the estimate of $\boldsymbol{V}_{\boldsymbol{x}}$, the variance of $\widehat{\delta}_{\boldsymbol{x}}$, n is the sample size, and p is the number of regression coefficients (including the intercept term).

This stationary point approach has the inherent drawback that the stationary point is not necessarily the maximum (or minimum) point. If this approach is applied to response surface models other than quadratic, stationary points may not correspond to maxima (or minima) on the response surface. Even if the response surface is restricted to the quadratic form in (2.1), the stationary point may correspond to a saddle point and not an optimal point. A more subtle, and possibly misleading, example can be found in a data set in Box and Draper (1987, chap. 9). If a quadratic model is fit to the data, an adequate fit seems to exist. The test for lack-of-fit gives a p-value of 0.39, while the test for quadratic effects gives a p-value of 0.0003. The estimate of the $\boldsymbol{B}$-matrix is negative definite (with eigenvalues of –10.0, -4.4, and –1.8) indicating a maxium stationary point. A 90% confidence interval, [-3.19, -0.39], for the maximum eigenvalue of $\boldsymbol{B}$, computed using the method in Peterson (1993) indicates that this eigenvalue is statistically significantly less than zero.

Box and Draper (1987, chap.9) state, "On the assumption of the adequacy of the second degree equation, a confidence region for the maximum may be computed. For details, see Box and Hunter (1954)." The associated 90% BH confidence region is given by Figure 1 below. Here, it is clear that this confidence region is composed of two disjoint regions (although in some other cases it is possible that one elongated region could be obtained).

<Figure 1 about here.>

Insight into this phenomenon can be obtained by way of the Theorem 2.1 below.

THEOREM 2.1: *Suppose we have the response surface model in* (2.1). *Let* C_{BH} *be the BH confidence region and let* $C^*_{BH} = \left\{ \boldsymbol{x} : \beta + \boldsymbol{Bx} = \boldsymbol{0}, \text{ where } \beta, B \in C \right\}$, *where C is the confidence region for the model parameters* β *and* $\boldsymbol{B}$ *which uses the BH critical value, kF(1-*α*, k,* ν*). Then* $C_{BH} = C^*_{BH}$. *This result also holds for any parametrically linear model, not just the model in* (2.1).

Proof. See the Appendix.

Using Theorem 2.1, we have found that points in the right (separated) part of the confidence region are associated with regression parameter values such that the ***B*** matrix is not negative definite on *C*. These points are associated with saddle points on the response surface. The points in the left part of the confidence region are associated with regression parameter values such that the ***B*** matrix is negative definite. As such, these points are associated with maximum points on the response surface. This BH confidence region is misleading in that, given the statistics in the second paragraph of this section, an experimenter would naturally assume that all of the points in the BH confidence region correspond to the response surface maximum. In Section 3, a new method for constructing a confidence region for the maximum (or minimum) point on a response surface, with or without constraints, is proposed. For the example in Figure 1, this new confidence region is composed only of the ***x***-points associated with the maximum of the response surface.

As stated above, SCW incorporated constraints into the BH approach by introducing Lagrange multipliers. Unfortunately, the Lagrange multiplier approach has several technical difficulties associated with it. As SCW point out, the Lagrange multiplier approach in general only addresses a constrained stationary point, not necessarily the minimum point. For a fixed experimental region, the Lagrange multiplier is a function of the estimated model parameters. Usually, this function is not of closed analytic form, requiring as a practical matter that the estimate of the Lagrange multiplier value be treated as a constant. In addition, the Lagrange multiplier approach can be difficult to apply when the experimental region is complex or not easily defined by one differentiable function.

In the next section, a confidence region for optimal factor levels of a response surface is introduced which avoids Lagrange multipliers and allows for more general modeling of both the response surface and the experimental region. This modeling flexibility can in turn help make for better statistical inferences for both constrained and unconstrained experimental regions.

3. A General Approach.

Throughout the rest of this paper the standard quadratic response surface model is generalized to that of a (parametrically) linear model for the response surface. The model in (2.1) is replaced with the more general form,

$$y = \beta_0 + \boldsymbol{z}(\boldsymbol{x})'\theta + e, \tag{3.1}$$

where $\boldsymbol{z}(\boldsymbol{x})$ is a (p-1)x1 vector-valued function of $\boldsymbol{x}$, and θ is a (p-1)x1 vector of regression coefficients. This generality may help to provide for more flexible response surface modeling. A common objection to the use of the model in (2.1) is that it imposes symmetry on the response surface form (Carter and Dornseif, 1990). We further generalize by defining R to be an arbitrary user-defined experimental region (that is a connected subset of the continuous factor space). In the following two subsections we construct a general method for obtaining a confidence region for $\boldsymbol{x}_0$, the point that minimizes $\boldsymbol{z}(\boldsymbol{x})'\theta$ over R, and then establish connections with the Box-Hunter confidence region.

3.1 Derivation of the Confidence Region.

Let $\boldsymbol{x}_0$ and $\eta(\theta)$ be defined as

$$\boldsymbol{z}\left(\boldsymbol{x}_0\right)'\theta = \min_{\boldsymbol{x}\in R} \boldsymbol{z}(\boldsymbol{x})'\theta = \eta(\theta)$$

and consider the null hypothesis H_0: $\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta = 0$ for some $\boldsymbol{x} \in R$. If H_0 is true, then $\boldsymbol{x}$ is an optimal point for the response surface. If a test for H_0 can be constructed, then a 100(1-α)% confidence region for $\boldsymbol{x}_0$ can be created from the set of all $\boldsymbol{x}$-values such that H_0 is not rejected at level α. A test for H_0 can be created if a confidence interval for $(\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta)$ can be found. If the confidence interval for $(\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta)$ does not contain 0, then we can reject H_0; otherwise we cannot reject H_0.

In order to find a confidence interval for $(\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta)$, consider the interval

$$[\min_{\theta\in C} \left(\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta\right),\ \max_{\theta\in C} \left(\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta\right)] \tag{3.2}$$

where C is the usual quadratic-form confidence region for θ,

$$C = \left\{\theta\colon \left(\hat{\theta} - \theta\right)' \hat{\boldsymbol{V}}^{-1}\left(\hat{\theta} - \theta\right) \le c_\alpha^2\right\}.$$

Here, $\hat{\theta}$ is an estimate of θ and $\hat{\boldsymbol{V}}$ is an estimate of $\boldsymbol{V}$, the variance covariance matrix of $\hat{\theta}$. The critical value, c_α^2, is the $100(1-\alpha)$% upper percentile of an F-distribution. The min-max type of confidence interval in (3.2) can be shown to be equivalent to an exact likelihood profile confidence interval (Clarke, 1987). One can obtain a conservative confidence interval by choosing $c_\alpha^2 = (p-1)F(1-\alpha;(p-1),\nu)$ (Rao, 1973, chap 7), but this would be acting very conservatively. The choice of $c_\alpha^2 = F(1-\alpha;1,\nu)$ is recommended by Clarke (1987) as an approximate critical value for estimating a confidence interval for a function of regression model parameters using the approach in (3.2). However, some discussion in Section 5 will indicate that the BH critical value of $c_{\alpha}^{2} = kF(1-\alpha, k, \nu)$ (where k is the dimension of R) is a more appropriate in this inference setting for $\boldsymbol{x}_0$. This is the critical value that will be used in this paper.

For the confidence interval in (3.2) note that it can be expressed as

$$\left[\min_{\theta\in C}\left\{\min_{w\in R}\left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)'\theta\right\},\quad \max_{\theta\in C}\left\{\min_{w\in R}\left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)'\theta\right\}\right] \tag{3.3}$$

Unfortunately, the 'min' term in braces in (3.3) does not in general have a closed functional form. This makes the confidence interval in (3.3) difficult to compute. In this article, the following confidence interval form is considered instead,

$$\left[\min_{w\in R}\left\{\min_{\theta\in C}\left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)'\theta\right\},\quad \min_{w\in R}\left\{\max_{\theta\in C}\left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)'\theta\right\}\right]. \tag{3.4}$$

Since $\left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)'\theta$ is linear in θ it follows that the terms in braces in (3.4) have a closed functional form. If the expression in braces for the upper confidence bound in (3.4) is denoted by $\mathrm{b}_{\boldsymbol{x}}(\boldsymbol{w})$, then

$$\mathrm{b}_{\boldsymbol{x}}(\boldsymbol{w}) = \left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)'\hat{\theta} + c_{\alpha}\left[\left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)'\hat{\boldsymbol{V}}\left(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x})\right)\right]^{1/2} \tag{3.5}$$

(The expression in braces for the lower confidence bound in (3.4) is the same as that in (3.5) but with c_{α} replaced by $-c_{\alpha}$.) The lower confidence bounds in (3.3) and (3.4) are equal by the principal of iterated suprema (Olmstead, 1956, p515). The upper confidence bound in (3.4) is greater than or equal to the corresponding bound in (3.3). This follows directly from a fundamental minimax result . See for example, Zangwill (1969, pp 45-46).

Let $C_{\boldsymbol{x}_0}$ be the confidence region for $\boldsymbol{x}_0$. A point in R will be included in $C_{\boldsymbol{x}_0}$ if and only if the confidence interval for $(\eta(\theta)-\boldsymbol{z}(\boldsymbol{x})'\theta)$ contains zero. Since $(\eta(\theta)-\boldsymbol{z}(\boldsymbol{x})'\theta)\le 0$ for all θ, and the lower bounds in (3.3) and (3.4) coincide, it follows directly that the lower bound in (3.4) is always less than or equal to zero. Therefore, a point in R will be not be included in $C_{\boldsymbol{x}_0}$ if and only if the upper bound in (3.4) is less than zero, or equivalently

$$\min_{\boldsymbol{w}\in R} b_{\boldsymbol{x}}(\boldsymbol{w}) \tag{3.6}$$

is less than zero. The confidence bound in (3.6) is intuitively appealing in that it is derived from a confidence band for a response surface for $(\boldsymbol{z}(\boldsymbol{w})-\boldsymbol{z}(\boldsymbol{x}))'\theta$ (for each fixed $\boldsymbol{x}$). This form can also be modified to provide some insight from a hypothesis testing perspective. Note that if the bound in (3.6) is less than zero, then for some $\boldsymbol{w}\in R$ we have

$$\frac{(\boldsymbol{z}(\boldsymbol{x})-\boldsymbol{z}(\boldsymbol{w}))'\hat{\theta}}{\left[\boldsymbol{z}(\boldsymbol{x})-\boldsymbol{z}(\boldsymbol{w})'\hat{\boldsymbol{V}}(\boldsymbol{z}(\boldsymbol{x})-\boldsymbol{z}(\boldsymbol{w}))\right]^{1/2}} > c_{\alpha} \tag{3.7}$$

So (3.7) implies that if we can find a point in R which corresponds to a point on the response surface that is statistically significantly less than the point which corresponds to $\boldsymbol{x}$, then we reject $\boldsymbol{x}$, i.e. reject H_0. Therefore $C_{\boldsymbol{x}_0}$ is simply the set of all $\boldsymbol{x}$-points in R for which we can find no other $\boldsymbol{w}$-points in R where $\boldsymbol{z}(\boldsymbol{w})'\theta$ statistically significantly less than $\boldsymbol{z}(\boldsymbol{x})'\theta$.

If the constraint region is formed by an equality constraint (e.g. $R=\{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x}=r^2\}$) then $C_{\boldsymbol{x}_0}$ has the following advantage over the Lagrange multiplier approach. Using polar coordinates we can reformulate the optimization problem as $\min_{\boldsymbol{a}\in A} \boldsymbol{z}^*(\boldsymbol{a})'\theta$ where A is the region of polar coordinate angles and $\boldsymbol{x} = r\boldsymbol{t}(\boldsymbol{a})$ (the polar coordinate transformation) and $\boldsymbol{z}(\boldsymbol{x}) = \mathrm{z}(r\boldsymbol{t}(\boldsymbol{a})) = \boldsymbol{z}^*(\boldsymbol{a})$. Since the dimension of A is k-1, it follows that the (smaller) critical value used would be (k-

1)$F(1\text{-}\alpha, (k\text{-}1), \nu)$. However, this dimension reducing property holds even if we cannot identify the transformation needed. As long as the constraint region, R, is of dimension k' (less than k) we can use the critical value $k'F(1\text{-}\alpha, k', \nu)$.

This confidence interval approach to constructing a confidence region for $\boldsymbol{x}_0$ is useful in that it provides as a byproduct a way (see (3.2)) to construct a confidence interval for $\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta$ where $\boldsymbol{x}$ is perhaps the current operating factor-level configuration. As such one can then make an inference as to how much improvement we can expect at the optimal point.

If a confidence interval for $\eta(\theta)$ is desired, a similar approach as in (3.4) can be applied to $\boldsymbol{z}(\boldsymbol{w})'\theta$. Peterson (1993) addresses the constrained estimation case for $\eta(\theta)$ in ridge analysis ($\boldsymbol{x}$ on a circle of radius r) but the results apply to a general experimental region using $c_\alpha^2 = F(1-\alpha;1,\nu)$. Asymptotic (delta method) approaches have also been proposed by Peterson (1989) and by Chinchilli et al. (1991). Chinchilli et al. (1991) assume an (unconstrained) quadratic model. Peterson (1989) assumes a general response surface model, and considers both constrained and certain unconstrained situations. Note that constructing a confidence interval for $\eta(\theta)$ or $\left(\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta\right)$ is much more direct than finding a confidence region for $\boldsymbol{x}_0$. (For the confidence region discussed above, $C_{\boldsymbol{x}_0}$, a confidence interval is used to test a null hypothesis which is then inverted in such a way as to create a confidence region for $\boldsymbol{x}_0$.)

Interestingly, the (linear approximation) delta method applied to the confidence region problem of this paper does not work as the linear term in the Taylor series approximation of $\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta$ collapses to zero under H_0. This happens because the gradient vector of $\eta(\theta)$ equals $\boldsymbol{z}(\boldsymbol{x}_0(\theta))$ where $\boldsymbol{x}_0(\theta)$ is the optimal value of $\boldsymbol{x}$ for a given θ (Peterson, 1989).

3.2 Connections with the Box-Hunter Region.

It turns out that for any response surface model of the form in (3.1), there is an intimate connection between the function $b_{\boldsymbol{x}}(\boldsymbol{w})$ and the BH criterion for determining whether or not a point should be in the confidence region. This connection can be established by examining the local behavior of $b_{\boldsymbol{x}}(\boldsymbol{w})$ in a $\boldsymbol{w}$-neighborhood of $\boldsymbol{x}$. However, the local behavior of $b_{\boldsymbol{x}}(\boldsymbol{w})$ in a $\boldsymbol{w}$-neighborhood of $\boldsymbol{x}$ cannot be assessed by computing the gradient vector of $b_{\boldsymbol{x}}(\boldsymbol{w})$ evaluated at $\boldsymbol{x}$ since $b_{\boldsymbol{x}}(\boldsymbol{w})$ is not differentiable at $\boldsymbol{x}$. Nonetheless, the directional derivatives of $b_{\boldsymbol{x}}(\boldsymbol{w})$ can be easily computed if $\boldsymbol{z}(\boldsymbol{x})$ is differentiable with respect to $\boldsymbol{x}$. The directional derivatives can be used to determine if there exists a $\boldsymbol{w}$-point (local to $\boldsymbol{x}$ and inside of $\boldsymbol{R}$) where $b_{\boldsymbol{x}}(\boldsymbol{w})$ becomes negative. The directional derivative of $b_{\boldsymbol{x}}(\boldsymbol{w})$ associated with the direction vector, $\boldsymbol{d}$, is defined as

$$b'_{\boldsymbol{x}}(\boldsymbol{w};\boldsymbol{d}) = \lim_{h\to 0^+} \frac{\left(b_{\boldsymbol{x}}(\boldsymbol{w}+h\boldsymbol{d}) - b_{\boldsymbol{x}}(\boldsymbol{w})\right)}{h}.$$

If $b'_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{d}) < 0$ then there exists a sufficiently small, positive h such that $b_{\boldsymbol{x}}(\boldsymbol{x}+h\boldsymbol{d}) < 0$. If $\boldsymbol{z}(\boldsymbol{x})$ is differentiable, then it is straightforward to show, taking limits, that

$$b'_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{d}) = \boldsymbol{d}'\boldsymbol{D}(x)\hat{\theta} + c_{\alpha}\left[\boldsymbol{d}'\boldsymbol{D}(x)\hat{\boldsymbol{V}}\boldsymbol{D}(x)'\boldsymbol{d}\right]^{1/2}, \tag{3.7}$$

where $\boldsymbol{D}(\boldsymbol{x})$ is the kxp matrix of derivatives of $\boldsymbol{z}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$. Computing $b'_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{d})$ over the set $\mathcal{D} = \{\boldsymbol{d} : \boldsymbol{d}'\boldsymbol{d} = 1, \boldsymbol{x} + \mathrm{h}\boldsymbol{d} \in R \text{ for small } \mathrm{h} > 0\}$ provides a local check of $b_{\boldsymbol{x}}(\boldsymbol{w}) < 0$ for some $\boldsymbol{w}$ in a neighborhood of $\boldsymbol{x}$. Hence if $b'_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{d}) < 0$ for some direction vector $\boldsymbol{d}$ then $\boldsymbol{x}$ would be rejected from the $C_{\boldsymbol{x}_0}$ confidence region. The theorem below establishes a close connection between $b'_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{d})$ and C_{BH} for $\boldsymbol{x}$-points in int$R$, the interior of R.

THEOREM 3.1: *Suppose **z(x)** is differentiable. If **x** is in intR, then* $b'_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{d})<0$ *for some **d** if and only if*

$$\hat{\theta}'\boldsymbol{D}(\boldsymbol{x})'\left[\boldsymbol{D}(\boldsymbol{x})\hat{\boldsymbol{V}}\boldsymbol{D}(\boldsymbol{x})'\right]^{-1}\boldsymbol{D}(\boldsymbol{x})\hat{\theta} > c_{\alpha}^{2}. \tag{3.8}$$

Furthermore:

(i) $\left(C_{\boldsymbol{x}_0}\cap \operatorname{int}R\right)\subset\left(C_{BH}\cap \operatorname{int}R\right)$ *and (ii) For unconstrained optimization,* $C_{\boldsymbol{x}_0}\subset C_{BH}$.

Note that (3.8) is the condition for rejecting a stationary ***x***-point with respect to $\boldsymbol{z}(\boldsymbol{x})'\theta$ for the BH confidence region. This makes sense in that rejecting a point in int*R* as stationary is equivalent to rejecting that point as optimal. Additionally, Theorem 3.1 shows that for points in int*R* $C_{\boldsymbol{x}_0}$ can never be any larger than any portion of C_{BH} that may be in int*R*.

The local check in (3.8) becomes a global check under the conditions of the theorem below.

THEOREM 3.2: *Suppose* $\boldsymbol{z}(\boldsymbol{x})'\theta=\beta'\boldsymbol{x}+\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$. *If **x** is in intR and* B *is positive definite (p.d.). for all* $\theta\in C$, *then* (3.8*) holds if and only if* $b_{\boldsymbol{x}}(\boldsymbol{w})<0$ *for some **w** in R.*

Furthermore: (i) $C_{\boldsymbol{x}_0}\cap \operatorname{int}R=C_{BH}\cap \operatorname{int}R$ *and (ii) For unconstrained optimization,* $C_{\boldsymbol{x}_0}=C_{BH}$.

So under the conditions of Theorem 3.2, for unconstrained optimization, the confidence region introduced in Section 3.1 and BH confidence region are equivalent. It is possible to check to see if ***B*** is p.d. on *C* by doing a straightforward computation which is related to a lower confidence bound (using the BH critical value) for the minimum eigenvalue of ***B***. (See Peterson, 1993 for details.) It is important to check to see if ***B*** is p.d. on *C*, because if it is not it follows by Theorem 2.1 that there will be points in C_{BH} associated with saddle points.

4. Computational Efficiency

For testing H_0, an immediate computational simplification is possible. For a given $\boldsymbol{x}$, we need only find one $\boldsymbol{w}$ for which $b_X(\boldsymbol{w}) < 0$ in order to be able to reject H_0. Some additional modifications are possible to further increase computational efficiency. If a quadratic response surface model satisfies the conditions of Theorem 3, then the following two step approach provides a procedure that is almost as quick as the BH approach for computing $C_{\boldsymbol{x}_0}$.

A quadratic model approach where $\boldsymbol{B}$ is p.d. for all $\theta \in C$ is given as follows:

'*B*-matrix p.d.' method for computing $C_{\boldsymbol{x}_0}$:

Step 1: For all $\boldsymbol{x}$-values in the int*R*, select confidence region points according to the BH criterion.

Step 2: -For $\boldsymbol{x}$ values on the boundary of *R*, search for the first direction (into *R*) such that the directional derivative in (3.7) is less than zero.

A more general, derivative-free, approach can be obtained using the following two-step procedure. Here, some additional computational efficiency is obtained by comparing points in *R* to the minimizing $\boldsymbol{x}$-point to initially reject as many points as possible in *R*.

Derivative-free Method for computing $C_{\boldsymbol{x}_0}$:

Step 1: Compute the minimizing $\boldsymbol{x}$-value on *R*, and call it $\hat{\boldsymbol{x}}_0$. Set $\boldsymbol{w} = \hat{\boldsymbol{x}}_0$ and use $b_{\boldsymbol{x}}(\hat{\boldsymbol{x}}_0)$ as a criterion to reject as many points in *R* as possible. (This will reject most of the points that need to be rejected.)

Step 2: For any $\boldsymbol{x}$-value not rejected by Step 1, search for the first $\boldsymbol{w}$-value in *R* such that $b_{\boldsymbol{x}}(\boldsymbol{w}) < 0$.

5. Two Examples.

In this section, we apply the methodology in Section 3 to two examples from the literature. For purposes of comparison with the Lagrange multiplier method, the first example applies the methodology developed in Section 3 to the preclinical cancer chemotherapy study illustrated in SCW. Here, it is sought to estimate the optimal dose combination of two cytotoxic chemicals, 5-Fluorouracil (5-FU) and Teniposide (VM26) for treatment of leukemia in the B62DF1 mouse animal-model. It is also of keen importance from a practical treatment perspective to clearly determine whether or not there exists a combination therapy that is superior to the monotherapies used in this study. Therefore a confidence region for the constrained optimal combination dose was constructed. Sixteen different treatment combinations were used. The study end-point was survival time in days. The data was analyzed using the proportional hazards regression model. The standard quadratic model was fit to the data. (However, for some similar experiments listed in Carter et al (1983) a third-order polynomial provides a better fit.) The unconstrained minimum (stationary point) of the fitted quadratic response surface yields a 5FU value of 428.9 mg/kg and a VM26 value of 36.1 mg/kg. As SCW point out, this shows that the unconstrained optimal combination estimate is far from the experimental region. As such, it would be unwise to try to make statistical inferences about optima so far from the experimental region. Using the method in Peterson (1993), a 95% confidence interval for $\lambda_{\min}(\boldsymbol{B})$ is [0.026, 1.09]. Hence the response surface is statistically significantly convex. However, $\boldsymbol{B}$ is not p.d. everywhere on C (for $\alpha \le 0.1$) with asymptotic BH critical value, $c_\alpha^2 = \chi^2(1-\alpha; k), k = 2$. Therefore the general (non-derivative) approach was used to compute $C_{\boldsymbol{x}_0}$. For a direct comparison with the SCW region, the circular constraint regions, $R_1 = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} \le 1\}$ and $R_2 = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} = 1\}$, in the coded factor space were initially used. Figures 2 displays the confidence regions from SCW and Section 3 respectively for the constraint region, R_1. Here, it is clear that $C_{\boldsymbol{x}_0}$ has smaller area than the Lagrange multiplier confidence region. Figure 3 displays the confidence regions from SCW and Section 3 respectively for the constraint region, R_2. Here, also the size of $C_{\boldsymbol{x}_0}$ is smaller than for the SCW region, a 15% reduction.

<Figure 2 about here.>

<Figure 3 about here.>

In order to use a more natural constrained experimental region, R was redefined to be the set of points within the convex hull of the design points (illustrated by the symbol '■') shown in Figure 4. Using this constraint region the resulting confidence region is represented below as the gray area in Figure 4. The confidence region indicates that 5FU and VM26 are synergistic with respect to the dose-region defined within the convex hull of the experimental treatment combinations. Computing the SCW confidence region for this example would be extremely difficult given the complexity of the constraint region.

<Figure 4 about here>

The second example involves a mixture experiment to study, in part, an important response variable involved in the formulation of a controlled-release drug substance to aid in obtaining more uniform blood levels (Frisbee and McGinity, 1994). Lower glass transition temperature is associated with enhanced film coating performance so it is desired to find factor levels associated with minimizing glass transition temperature. As with the preclinical cancer chemotherapy example, in this example it is also important to determine if all of the mixture components are necessary for the optimal blend. The factors chosen were: x_1="% of Pluronic® F68", x_2="% of polyoxyethlene 40 monostearate", x_3="% of polyoxyethylene sorbitan fatty acid ester NF". The experimental design used was a modified McLean- Anderson design (McLean and Anderson,

1966) with two centroid points, resulting in a sample size of eleven. The response surface model that seemed to give the best fit was an H1 Becker model (Becker, 1968),

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} \min(x_1, x_2) + \beta_{13} \min(x_1, x_3) + \beta_{23} \min(x_2, x_3) + e, \qquad (4.1)$$

where Y equals the observed glass transition temperature (^{o}C). The mean square error (mse) associated with this model is 1.71 which is a 53% reduction in mse from the standard quadratic model. The adjusted R^2 for the model in (4.1) is 96.4%.

For mixture experiments it is important to note that the intercept form of the model should be used in order to execute the computations needed to construct the confidence region for $\boldsymbol{x}_0$ or to compute a confidence interval for $\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta$ for some $\boldsymbol{x}$. This is because *with the intercept form of the model* $\eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta$ depends only upon the other (non-intercept) regression coefficients. In other words, θ properly represents only the (non-intercept) regression coefficients in the intercept form of the model. The no-intercept form of the mixture model, on the other hand, has the underlying intercept term embedded in at least some of its coefficients.

Computing SCW confidence region would be problematic since there are several factor constraints involved and the model in (4.1) is not everywhere differentiable. The derivative-free approach was used to compute the confidence region in Section 3 for the above mixture experiment. The 95% confidence region for the minimizing factor level combination is illustrated in Figure 5 below. It is clear from the confidence region in Figure 5 that we do not have statistically significant evidence of factor synergy (as defined in Section 1). For a follow-up study, Frisbee and McGinity(1994) used only Pluronic® F68 (factor x_1) as their surfactant as the minimal estimated response surface was found at $x_1 = 1$, with low mean response values near this point. Use of only one surfactant makes for simpler future formulation studies. However, form Figure 5 it is evident that they do not have statistically significant evidence that using only Pluronic® F68 is optimal. If use of only factor x_1 is optimal then further experimentation should reveal this by use of the confidence region method described in Section 3. For sufficient data, the confidence region would become reduced to a point ($\boldsymbol{x} = (1,0,0)$) thereby providing statistically significant evidence that use of only Pluronic® F68 is optimal. This "reduction to a point" phenomenon can even occur with modest sample sizes as evidenced by the (strawberry mite) pesticide mixture experiment of Cornell and Gorman (1978). Here a 95% confidence region constructed using the method of Section 3 produces a single point (only one pesticide optimal) using a 3-component mixture design with 15 data points.

<Figure 5 about here>

6. Coverage Rates

Using the notation of Section 3, suppose that $\boldsymbol{z}(\boldsymbol{x})'\theta$ is a differentiable, unimodal function of $\boldsymbol{x}$ for all θ in *C*. It follows then that the null hypothesis $H_0^{(1)}: \eta(\theta) - \boldsymbol{z}(\boldsymbol{x})'\theta = 0$ is equivalent to $H_0^{(2)}: D(\boldsymbol{x})\theta = \mathbf{0}$. However, testing $H_0^{(2)}$ requires the use of the BH critical value, $c_\alpha^2 = kF\left(1-\alpha; k, \nu\right)$. As discussed below, use of the BH critical value for C_{x_0} will produce confidence regions with coverage probabilities that are competitive with those of SCW but have smaller area.

It is worth noting that in many cases, where the underlying optimal conditions are constrained to lie on or near the boundary of *R*, that the associated coverage rate may be considerably larger than the nominal 100(1-α)% level. This does not mean that the confidence region is very conservative in terms of being large or inefficient. In fact, in many such cases the confidence region will tend to be small as it is forced up against the boundary of *R*. Furthermore, Theorem

3.1 implies that any $\boldsymbol{x}$-point rejected by the BH criterion will also be rejected by the method in Section 3. As such, the size of the confidence region in the interior of *R* can never be larger than the part of the unconstrained BH confidence region that may cover the interior of *R*.

The property of increasing coverage rate and decreasing confidence region area can be most simply illustrated with the following example. Consider the simple quadratic regression model, $E(Y \mid x) = \beta_0 + \beta_1 x + x^2$, where $R = \{x : -1 \le x \le 1\}$. Suppose that $\beta_1 < -2$ so that the true constrained minimizing point is $x_0 = 1$. Let $\hat{\sigma}_1$ be an estimate of the variance of $\hat{\beta}_1$. If $\left(\hat{\beta}_1 - (-2)\right) / \hat{\sigma}_1 < -c_\alpha$ then it can be shown that $C_{x_0} = \{1\}$. Hence if $\hat{\beta}_1$ is less than –2 in a statistically significant fashion (relative to c_α) then C_{x_0} will be as small as possible, i.e. a point. Furthermore, if $\beta_1 < -2$ then $\Pr\left(x_0 \in C_{x_0}\right) = \Pr\left(\left(\hat{\beta}_1 - (-2)\right) / \hat{\sigma}_1 < -c_\alpha\right)$, which could be considerably larger than 1-α . In general, if the response surface over *R* is decreasing with high probability, then C_{x_0} will tend to be forced up against the boundary of *R*. This also makes sense with respect to Theorem 3.1, in that if the BH criterion rejects all points in the interior of *R* then also C_{x_0} will have no points in the interior of *R*.

For the examples in the previous section, simulations were done to assess the underlying coverage rates. The first example employs the asymptotic sampling theory for Cox-regression. One-thousand θ-estimates were simulated from a multivariate normal distribution with mean equal to the $\hat{\theta}$-value and covariance matrix equal to the $\hat{\boldsymbol{V}}$-value from the maximized likelihood. The constraint regions R_1 and R_2, corresponding to those used in Figures 2 and 3 respectively, were employed in simulations to compare the coverage rates for the SCW and $C_{\boldsymbol{x}_0}$ confidence regions. For the R_1 region, the coverage rates of the SCW and $C_{\boldsymbol{x}_0}$ 95% confidence regions were 98.3% and 98.5% respectively. For the R_2 region, the coverage rates of the SCW and $C_{\boldsymbol{x}_0}$ 95% confidence regions were 97.7% and 95.3% respectively. In addition, several simulation examples were printed out to examine the relative sizes of the resulting confidence regions. In all cases, for both the R_1 and R_2 regions the $C_{\boldsymbol{x}_0}$ regions were smaller. For the convex hull experimental region corresponding to Figure 3, nominal coverage rates of 90% and 95% yielded empirical coverage rates of 99.6% and 99.9% respectively. However, as we saw in Section 4 this confidence region was rather small.

For the second example, 1000 data sets each of size 11 were simulated from the same design as in example 2 with the same experimental region. For nominal coverage rates of 90% and 95% empirical coverage rates of 98.7% and 99.6% were obtained respectively.

Further simulations were done to provide a detailed comparison of the Lagrange multiplier and $C_{\boldsymbol{x}_0}$ confidence regions across several different response surface properties. Here, the quadratic model in (2.1) was used with two factors, along with the constraint region, $R = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} \leq 1\}$. One-thousand samples of size 11 were simulated from a central composite design with axial points at a distance of $\sqrt{2}$ from the origin and three center points. The simulated errors were from a normal distribution with mean 0 and variance 1. The true optimum $\boldsymbol{x}_0$-point was varied from the center of *R*, to the boundary of *R*, and then outside of *R* in order to assess the effects on the coverage rates. The locus of true optimal points is give in Table 1 below. The nominal coverage rate used was 95%. Five different eigenvalue combinations were used for the ***B*** matrix. Larger eigenvalues are associated with more curvature in the direction of the corresponding eigenvector associated with ***B***.

<Table 1 about here.>

While the coverage rates for $C_{\boldsymbol{x}_0}$ and the Lagrange multiplier confidence regions were similar, the $C_{\boldsymbol{x}_0}$ region has smaller area as evidenced by confidence region printouts from many simulated data sets. In all cases the $C_{\boldsymbol{x}_0}$ was smaller than the associated Lagrange multiplier confidence region.

Table 1 indicates that the method of Section 3 may be just slightly liberal in situations where the optimal point is inside the experimental region *and* the response surface is not well determined (e.g. one or more eigenvalues that are close to zero relative to their standard error). See, for example, the eigenvalues corresponding to (0.2, 0.2) or (0.2, 0.7). However, in these cases we still recommend using the method of Section 3 over the BH method for two reasons.

Firstly, note that with regard to the simulations for Table 1, the method of Section 3 will only reject a true minimizing point, $\boldsymbol{x}_0$, when a simulated data set is such that the fitted response surface has a point in *R* associated with a response that is statistically significantly (according to the BH critical value) less than the response associated with the true minimizing $\boldsymbol{x}_0$. Such a response surface (fitted to the simulated data) will not be convex; it will be a saddle surface or possibly even concave. It is possible, however, that the BH method may not reject $\boldsymbol{x}_0$, as $\boldsymbol{x}_0$ may not significantly differ from a non-optimal stationary point on the fitted response surface. In practice, however, if one were to observe such a non-convex fitted surface one would use a constrained confidence region procedure, as the minimum of the fitted surface must be on the boundary of *R*. Secondly note that use of the BH method with eigenvalues close to zero is not recommended as the BH confidence region may admit stationary points associated with non-optimal points (e.g. saddle points) even if the fitted response surface has a stationary optimal point. See the example in Section 2 and Figure 1. This BH confidence region would be very hard to interpret in practice.

For situations where the (unconstrained) optimum point is outside of the experimental region, we recommend the method of Section 3 over the SCW method. While the SCW and Section 3 confidence regions have similar coverage rates, we have observed that the Section 3 confidence regions are smaller than the SCW regions. As mentioned previously the high coverage rates are no problem as the confidence region will tend to be small as it is forced up against the boundary of *R*.

While the confidence region forced to the boundary of *R* may be relatively small, one may still wish to consider ways of reducing the conservatism. One possibility is to use the very conservative Scheffe' critical value, $c_{\alpha}^{2} = (p-1)F\left(1-\alpha;(p-1),\nu\right)$. If this confidence region lies entirely on the boundary of *R* this is very strong evidence that the minimum point must be somewhere on the boundary of *R*. As such we can be sure that we can go back and restrict our optimization region from *R* to ∂*R*, the boundary of *R*. This in turn will justify reducing c_{α}^{2} to $lF\left(1-\alpha;l,\nu\right)$ where $l = \max\{\dim \partial R, 1\}$. This inference is, of course, conditional upon the assumption that the minimizing point is somewhere on the boundary of *R*.

7. Discussion.

As can be seen from the example in the previous section, the approach in Section 3 for estimating and testing optimal factor combinations is attractive in that general experimental regions and a wide variety of (parametrically) linear models can be employed. This allows for a much improved and wider application of Carter et al.'s (1982) important idea of using confidence regions to test for factor synergism. The generality in Section 3 is also useful for constructing confidence regions for optima in mixture experiments. These constrained confidence regions also have the interesting and pleasing property that as the optimal (unconstrained) factor-level configuration moves beyond the boundary of *R* their coverage rate becomes higher than nominal *but* simultaneously their area shrinks (as they become forced up against the boundary of *R*).

In addition, this approach can easily be generalized to other families of response surface models in cases where the sampling distribution of the regression model parameters can be approximated by a multivariate normal distribution. For large (or possibly modest) sample situations then, $C_{\boldsymbol{x}_0}$ could be computed for (parametrically) nonlinear or generalized linear regression models. Some examples of useful nonlinear response-surface models can be found in Nelder (1966), Mead and Pike (1975), Box and Draper (1987, chap. 12), and Khuri and Cornell (1996, chap. 10). Since the confidence region inference is based upon confidence bounds for the mean response function, the parameter-effects nonlinearity associated with such inferences will be zero (Ratkowsky, 1983, chap. 9). The use of generalized linear models for response surface inference has been espoused by Myers (1999). The inference methods in this paper are generalizable to such models as the form in (3.5) has an analogous approximate form for generalized linear models (Myers and Montgomery, 1997). Lewis, Montgomery, and Myers (2001) show that for generalized linear models using standard response surface designs with modest sample sizes the confidence interval coverage for the mean response surface is reasonably accurate. Hence the methodology of Section 3 ought to be applicable to generalized linear models as well. A proportional hazards regression framework can also be used as was shown in example 1 above.

Extensions of this methodology to multiresponse surface optimization may be possible using desirability functions (Harrington, 1965, Derringer and Suich, 1980). Here special functions for *m* response variables can be used to build a single "desirability" response surface (Myers and Montgomery, 1995, p. 252). The methodology of Section 3 can then be applied. The more general form in (3.1) may prove useful for developing such a model.

An easy to use MATLAB® program is available from the third author for computing the $C_{\boldsymbol{x}_0}$ confidence region using the derivative-free, two-step algorithm shown in Section 4. The program graphs $C_{\boldsymbol{x}_0}$ for $R = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} \leq 1\}$ or for a simplex region associated with a mixture experiment. Any response surface model of the form $\beta_0 + \boldsymbol{z}(\boldsymbol{x})'\theta$ can be used provided that the sampling distribution of estimate, $\hat{\theta}$, has an approximate multivariate normal distribution.

ACKNOWLEDGEMENT

We are grateful to the referees for their comments which have helped to improve this paper. The last two authors were partially funded by NSF grant DMI-9988563.

REFERENCES

Becker, N. G. (1968), "Models for the response of a mixture", *Journal of the Royal Statistics Society – Series B*, 30, 349-358.

Box, G. E. P. and Draper, N. R. (1987) *Empirical Model Building and Response Surfaces*, New York: John Wiley.

Box, G. E. P. and Hunter, J. S. (1954) "A Confidence Region for the Solution of a Set of Simultaneous Equations with an Application to Experimental Design", *Biometrika*, 41, 190-199

Carter, W.H., Wampler, G.L., Stablein, D. M. and Campbell, E.D. (1982) "Drug Activity and Therapeutic Synergy in Cancer Treatment" *Cancer Research* **42**, 2963-2971.

Carter, W. H., Wampler, G.L., Stablein, D. M. (1983), *Regression Analysis of Survival data in Cancer Chemotherapy*, New York: Marcel Dekker.

Carter, W. H. and Dornseif, B. E. (1990), Maximizing Drug Benefit in Combination Therapy (Dose-Response Estimation), *Drug Information Journal*, 24, 351-359.

Chinchilli, V. M., Carter, W. H., Breen, T. J., and Campbell, E. D. (1991) "An Asymptotic Confidence Interval for the Response at the Stationary Point of a Response Surface", *Journal of Biopharmaceutical Statistics*, **1**, 287-301.

Clarke, G. P. Y. (1987) "Approximate Confidence Limits for a Parameter Function in Nonlinear Regression", *Journal of the American Statistical Association*, **82**, 221-230.

Cornell, J. A. and Gorman, J. W. (1978) "On the detection of an additive blending component in multicomponent mixtures", *Biometrics,* **34**, 251-263.

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

Frisbee, S.E. and McGinity, J.W. (1994) "Influence of Nonionic Surfactants on the Physical and Chemical Properties of Biodegradable Pseudolatex", *European Journal of Pharmaceuticals and Biopharmaceuticals* **40**, 355-363. ‘

Harrington, E. C. (1965), "The Desirability Function", *Industrial Quality Control*, 21 pp. 494-498

Khuri A. I. and Cornell, J. A. (1996) *Response Surfaces: Designs and Analyses* (2nd ed.), New York: Marcel Dekker.

Lewis, S.L., Montgomery, D. C., and Myers, R.H. (2001), "Confidence Interval Coverage for Designed Experiments Analyzed with GLMs", *Journal of Quality Technology*, 33, pp 279-292.

Mantel, N. (1974) "Therapeutic Synergism", *Cancer Chemotherapy Reports Part II* **4**, 147-149.

MATLAB (1999) The MathWorks Inc., Natick, MA.

McLean, R.A. and Anderson, V.L. (1966) "Extreme Verticies Design of Mixture Experiments", *Technometrics*, 8, 447-454.

Mead R. and Pike D. J.(1975), "A Review of Response Surface Methodology from a Biometric Viewpoint", *Biometrics*, **22**, 803-851.

Myers R. H. and Carter, W. H., (1973) "Response Surface Techniques for Dual Response Systems", *Technometrics*, **15**, 301-317

Myers, R.H. and Montgomery, D.C. (1995) *Response Surface Methodology* , New York: John Wiley.

Myers, R.H. and Montgomery, D.C. (1997) "A Tutorial on Generalized Linear Models", *Journal of Quality Technology*, **29**, 274-291.

Myers, R.H. (1999), "Response Surface Methodology – Current Status and Future Directions", *Journal of Quality Technology,* 31, 30-44.

Nelder, J. A. (1966) "Inverse Polynomials, a Useful Group of Multifactor Response Functions", *Biometrics*, **22**, 128-141.

Olmsted, J. (1956) *Real Variables*, New York: Appleton-Century-Crofts

Peterson, J. J. (1989) "First and Second Order Derivatives Having Applications to Estimation of Response Surface Optima", *Statistics and Probability Letters*, **8**, 29-34.

Peterson, J. J. (1993) "A General Approach to Ridge Analysis with Confidence Intervals", *Technometrics*, **35**, 204-214.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, New York, John Wiley.

Ratkowsky, D. A. (1983), *Nonlinear Regression Modeling: A Unified Approach*, New York: Marcel Dekker.

Stablein, D. L., Carter, W. H., and Wampler, G. L., (1983) "Confidence Regions for Constrained Optima in Response Surface Experiments", *Biometrics*, **39**, 759-763.

Stablein, D.M., Novak, J.W., Peace, K.E., Laska, E.M., and Meisner , M. J. (1990) "Optimization in Clinical Trials and Combination Drug Development", in *Statistical Issues in Drug Research and Development*, New York: Marcel Dekker, 263-303.

Venditti, J.M., Humphreys, S.R., Mantel, N. and Goldin, A. (1956) "Combined Treatment of Advanced Leukemia (L1210) in Mice with Amethopterin and 6-mercaptopurine", *Journal of the National Cancer Institute*, **17**, 631-638.

Zangwill, W. I. (1969) *Nonlinear Programming: A Unified Approach*, Englewood Cliffs, NJ, Prentice-Hall.

APPENDIX

Proof of Theorem 2.1.

Using the notation of section 3, we can express the BH confidence region as $C_{BH} = \left\{ \boldsymbol{x} : \hat{\theta}' \boldsymbol{D}(\boldsymbol{x}) \left(\boldsymbol{D}(\boldsymbol{x}) \hat{\boldsymbol{V}} \boldsymbol{D}(\boldsymbol{x})' \right)^{-1} \boldsymbol{D}(\boldsymbol{x}) \hat{\theta} \le c_{\alpha}^2 \right\}$ and C_{BH}^* as $C_{BH}^* = \left\{ \boldsymbol{x} : \boldsymbol{D}(\boldsymbol{x}) \theta = \boldsymbol{0}, \theta \in C \right\}$. First suppose that $\boldsymbol{x} \in C_{BH}^*$. Then $\boldsymbol{x}$ is such that the θ-hyperplane, $\boldsymbol{D}(\boldsymbol{x}) \theta = \boldsymbol{0}$, intersects *C*. Since $Q(\theta) = \left(\theta - \hat{\theta}\right)' \hat{\boldsymbol{V}} \left(\theta - \hat{\theta}\right)$ is a convex function with a minimum at $\hat{\theta}$, it follows that the value of θ_* that minimizes $Q(\theta)$ on the hyperplane, $\boldsymbol{D}(\boldsymbol{x}) \theta = \boldsymbol{0}$, must be contained in *C*. It follows then that

$$2\boldsymbol{V}^{-1} \left(\theta_* - \hat{\theta}\right) = \boldsymbol{D}(\boldsymbol{x})' \lambda, \tag{A.1}$$

where λ is a *k*x1 vector of Lagrange multipliers. Since $\theta_* \in C$, we must have $\left(\theta_* - \hat{\theta}\right)' \hat{\boldsymbol{V}} \left(\theta_* - \hat{\theta}\right) \le c_{\alpha}^2$. Thus by pre-multipling (A.1) by $\left(\theta_* - \hat{\theta}\right)'$ it follows that

$$2c_{\alpha}^2 \ge \left(\theta_* - \hat{\theta}\right)' \boldsymbol{D}(\boldsymbol{x})' \lambda. \tag{A.2}$$

It also follows from (A.1) that $2\left(\boldsymbol{D}(\boldsymbol{x})'\hat{\boldsymbol{V}}\boldsymbol{D}(\boldsymbol{x})\right)^{-1}\boldsymbol{D}(\boldsymbol{x})\left(\theta_* - \hat{\theta}\right) = \lambda$. Substituting this expression for λ into (A.2) one can show that

$$c_\alpha^2 \geq \left(\theta_* - \hat{\theta}\right)' \boldsymbol{D}(\boldsymbol{x})' \left(\boldsymbol{D}(\boldsymbol{x})\hat{\boldsymbol{V}}\boldsymbol{D}(\boldsymbol{x})'\right)^{-1} \boldsymbol{D}(\boldsymbol{x})\left(\theta_* - \hat{\theta}\right) \tag{A.3}$$

Now since $\boldsymbol{D}(\boldsymbol{x})\theta_* = \boldsymbol{0}$, it follows from (A.3) that $c_\alpha^2 \geq \hat{\theta}'\boldsymbol{D}(\boldsymbol{x})'\left(\boldsymbol{D}(\boldsymbol{x})\hat{\boldsymbol{V}}\boldsymbol{D}(\boldsymbol{x})'\right)^{-1}\boldsymbol{D}(\boldsymbol{x})\hat{\theta}$. But this implies that $\boldsymbol{x} \in C_{BH}$. Arguing from the contrapositive, one can show, using analogous reasoning as above, that $\boldsymbol{x} \notin C_{BH}^* \Rightarrow \boldsymbol{x} \notin C_{BH}$, thereby completing the proof. .

Proof of Theorem 3.1.

There exists a $\boldsymbol{d}$ such that $b_{\boldsymbol{x}}'(\boldsymbol{x};\boldsymbol{d}) < 0$ if and only if

$$\max_{\boldsymbol{d}} \frac{\left(\boldsymbol{d}'\boldsymbol{D}(\boldsymbol{x})\theta\right)^2}{\boldsymbol{d}'\boldsymbol{D}(\boldsymbol{x})\hat{\boldsymbol{V}}\boldsymbol{D}(\boldsymbol{x})\boldsymbol{d}} > c_\alpha^2 \tag{A.4}$$

However the expression on the left-hand side of the inequality in (A.4) is equal to

$$\hat{\theta}'\boldsymbol{D}(\boldsymbol{x})'\left\{\boldsymbol{D}(\boldsymbol{x})\hat{\boldsymbol{V}}\boldsymbol{D}(\boldsymbol{x})'\right\}^{-1}\boldsymbol{D}(\boldsymbol{x})\hat{\theta}$$

(Rao, 1973, p60). Hence the result follows.

Proof of Theorem 3.2.

Since $\boldsymbol{B}$ is p.d. on *C*, $\boldsymbol{z}(\boldsymbol{x})'\theta$ is convex for all $\theta \in C$. Hence $\left(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x})\right)'\theta$ is convex in $\boldsymbol{w}$ for each fixed $\boldsymbol{x}$ and all $\theta \in C$. Note that

$$\max_{\theta \in C}\left(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x})\right)'\theta = b_{\boldsymbol{x}}(\boldsymbol{w}).$$

By the convexity property of maximization it follows that $b_{\boldsymbol{x}}(\boldsymbol{w})$ is also convex in $\boldsymbol{w}$. But convexity of $b_{\boldsymbol{x}}(\boldsymbol{w})$ means that there exists a $\boldsymbol{d}$ such that $b'_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{d})<0$ if and only if $b_{\boldsymbol{x}}(\boldsymbol{w})<0$ for some $\boldsymbol{w}$ in R. Theorem 3.2 now follows by Theorem 3.1.

Eigenvalues		(0.2 , 0.2)		(7.0 , 0.2)		(1.0 , 1.0)		(7.0 , 1.0)		(13 , 15)	
Methods		**SCW**	**Sec. 3**	**SCW**	**Sec. 3**	**SCW**	**Sec. 3**	**SCW**	**Sec. 3**	**SCW**	**Sec. 3**
Locus of Optimal Points	***0.0***	*94.6*	*94.2*	*94.8*	*94.4*	*94.2*	*94.2*	*94.6*	*94.6*	*94.6*	*94.6*
	0.7	*93.8*	*94.1*	*93.2*	*93.6*	*95.4*	*94.5*	*95.3*	*94.1*	*95.6*	*95.6*
	0.9	*92.9*	*93.5*	*92.4*	*93.7*	*96.0*	*95.0*	*95.1*	*94.2*	*95.2*	*95.2*
	1.0	*92.8*	*95.2*	*93.9*	*97.3*	*96.1*	*97.5*	*97.1*	*98.5*	*97.2*	*97.6*
	1.1	*92.9*	*96.0*	*93.5*	*97.7*	*96.4*	*97.7*	*96.9*	*98.0*	*98.7*	*99.0*
	1.3	*93.1*	*95.7*	*96.4*	*98.8*	*97.1*	*97.9*	*97.1*	*98.5*	*97.9*	*99.4*
	1.6	*95.3*	*97.6*	*98.4*	*99.0*	*97.4*	*98.0*	*97.3*	*99.4*	*97.4*	*99.5*
	2.0	*94.7*	*96.3*	*98.6*	*99.2*	*99.1*	*99.6*	*97.7*	*99.2*	*98.6*	*100.0*
	2.5	*95.4*	*98.2*	*98.1*	*99.0*	*98.8*	*99.4*	*97.6*	*98.8*	*97.9*	*100.0*
	3.0	*96.6*	*98.5*	*98.9*	*99.6*	*98.4*	*99.1*	*98.0*	*98.7*	*98.2*	*100.0*

Table 1. Coverage Rates for 95% confidence regions based upon 1000 simulations. The response surface is varied by modifying the eigenvalues of the quadratic coefficients matrix. The locus of optimal points is the distance of optimal point from the origin in the coded factor space.

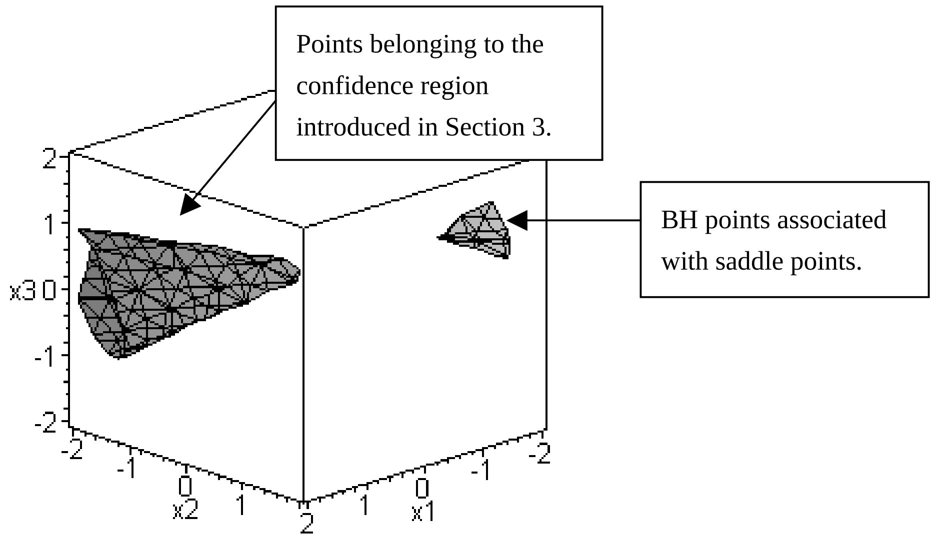
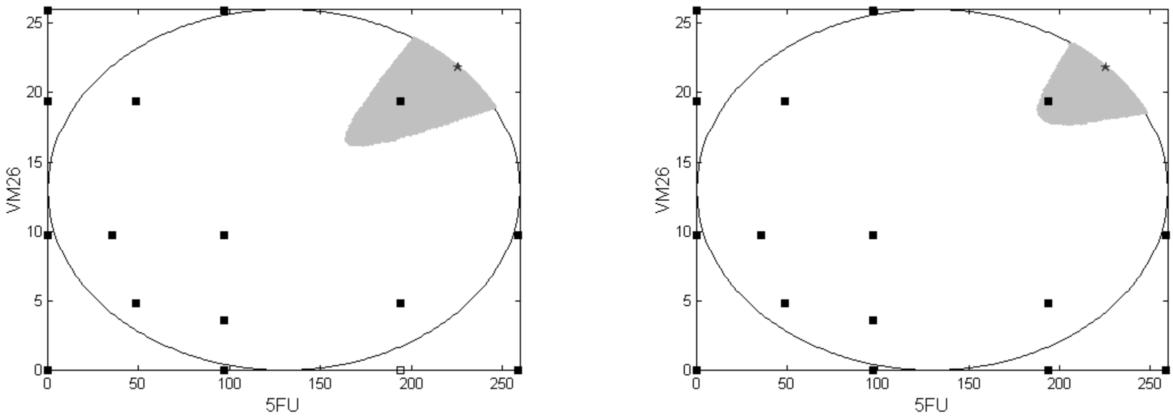
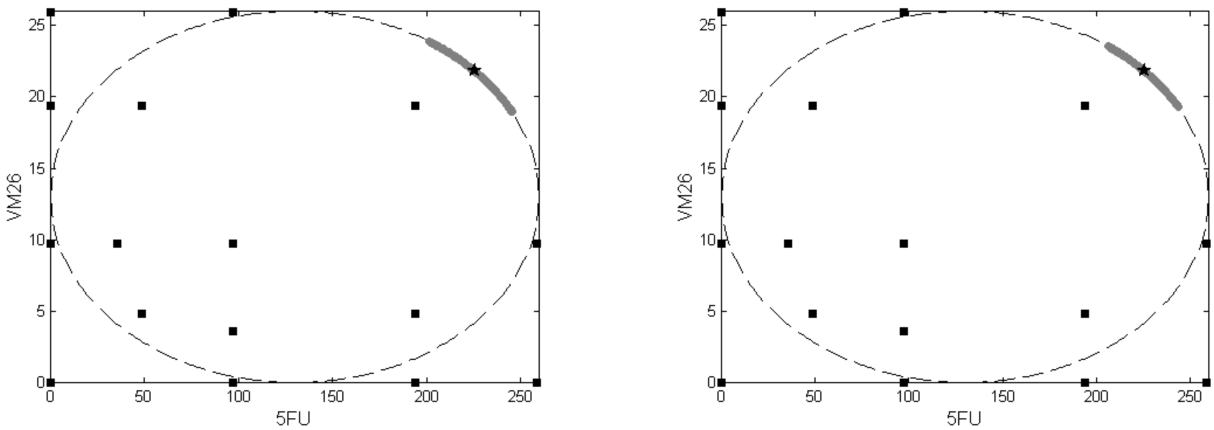


Figure 1. The disjoint 90% Box-Hunter confidence region for the maximizing factor levels in the example in Box and Draper (1987, Chap. 9). Points in the left part of the confidence region are associated with maximizing points, but points in the right part are associated with saddle points. The confidence region introduced in Section 3 coincides only with the maximizing points.



SCW Confidence Region *Section 3 Confidence Region*

Figure 2. The Grey Areas Represent the 95% Constrained Confidence Regions for the Minimizing Combination of Chemotherapy Dose Levels of 5 FU and VM 26. The experimental region is $R_1 = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} \le 1\}$ *in the coded factor space. The experimental runs are denoted by ('■').*



SCW Confidence Region *Section 3 Confidence Region*

Figure 3. The Solid Grey Lines Represent the 95% Constrained Confidence Regions for the Minimizing Combination of Chemotherapy Dose Levels of 5 FU and VM 26. The experimental region is $R_1 = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} = 1\}$ *in the coded factor space. The experimental runs are denoted by ('■').*

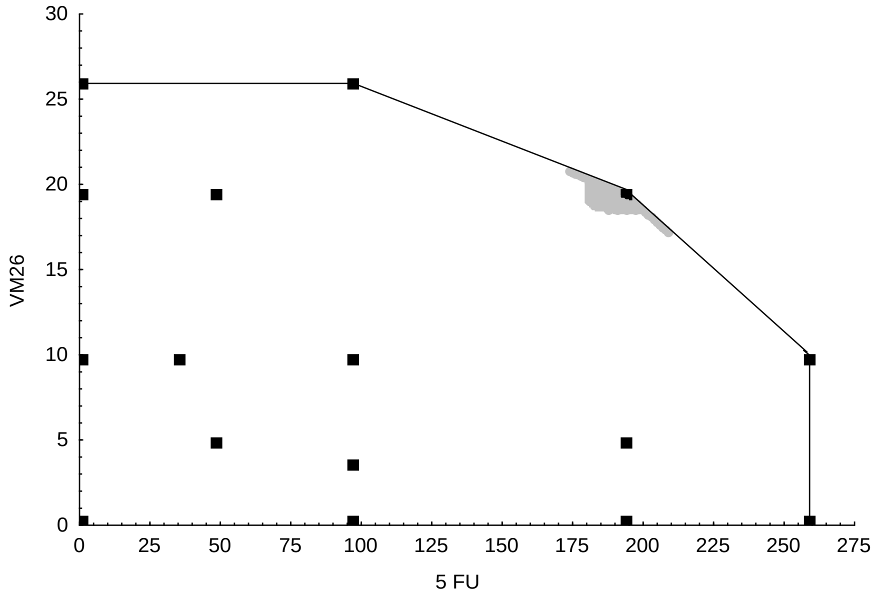


Figure 4. The Grey Area Represents a 95% Constrained Confidence Region for the Minimum Combination of Chemotherapy Dose Levels of 5 FU and VM 26. The experimental region is the convex hull of the points ('■') in the graph.

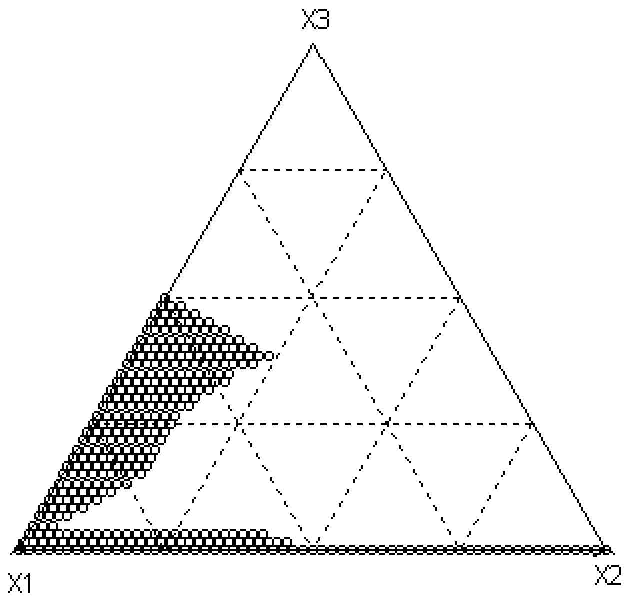


Figure 5. The 95% confidence region for the factor blend that minimizes the glass transition temperature is denoted by the collection of o-*symbols.*