CONFIDENTIAL

Under Review for Publication

Computation and Visualization of Confidence Regions for Optimal Factor Levels in Constrained Response Surface Problems

Suntara Cahya, Enrique Del Castillo, and John J. Peterson*

Abstract

An improved approach for computing the confidence regions for the optimal factor settings obtained from optimizing a general response surface model is presented. The approach has a better computational efficiency and improved accuracy compared to existing methodology. A three-factor mixture experiment was used for the performance comparison. The coverage rate properties of the resulting confidence regions were assessed through an extensive simulation study. Issues in the visualization of high dimensional confidence regions are discussed and illustrated using a five-factor experimental design.

Key Words: Design of Experiments; Mixture Experiments; Multiresponse Optimization; Nelder-Mead Simplex Method; Response Surface Methodology; Therapeutic Synergysm

1 Introduction

One of the objectives of Response Surface Methodology (RSM) is to find the optimal factor-level configuration of an empirical model obtained through a series of designed experiments. Due to sampling variability inherent in experimentation, various authors (Box and Hunter 1954; Carter et al. 1986; Myers

*Suntara Cahya is a Ph.D. candidate in the Department of Industrial & Manufacturing Engineering, Penn State University, University Park, PA 16802. Enrique Del Castillo is an Associate Professor, Department of Industrial & Manufacturing Engineering, Penn State University, University Park, PA 16802. John J. Peterson is a Principal Statistician, Statistical Sciences Department, UW281A, GlaxoSmithKline Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406.

and Montgomery 1995) suggest to compute the confidence region for the location of the optimal point in order to assess the precision of this point estimate.

Aside from assessing the precision of the optimal point estimate, these confidence regions are useful for a variety of reasons. For example, they provide an answer to the *therapeutic synergysm* problem common in the pharmaceutical industry. Carter et al. (1982) discussed how if the confidence region for the optimal dose combination excludes all individual zero dose combinations (i.e. if the confidence region does not touch any of the axes that define each drug component), then there is evidence that all drug components are beneficial for therapy in a synergistic way. These confidence regions are also useful for handling multiresponse optimization problems. A compromised solution can be found by utilizing nonlinear optimization techniques upon the individual confidence regions for the optimum of each response (Del Castillo 1996; Khun 1999).

An early attempt at computing a confidence region on the location of the optimal point was done by Box and Hunter (1954)(hereafter referred to as BH). They proposed a methodology for computing an exact confidence region for the (unconstrained) stationary point of a response surface. Suppose that the response surface can be modeled as a quadratic polynomial:

$$y = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} + \epsilon, \tag{1}$$

where y is the response variable and ϵ has a normal distribution with mean 0 and variance σ^2. Here, β_0 is the intercept term, $\boldsymbol{x}$ is a $k \times 1$ vector of controllable factors, $\boldsymbol{\beta}$ is a $k \times 1$ vector of linear coefficients β_i, and $\boldsymbol{B}$ is $k \times k$ symmetric matrix of quadratic coefficients with i^{th} diagonal element equals to β_{ii} and $(i,j)^{th}$ off-diagonal element equals to $\beta_{ij}/2$. If $\boldsymbol{x_0}$ is the stationary point of the response surface in (1), then the null hypothesis of zero first derivatives at $\boldsymbol{x_0}$ (i.e. $H_0 : \boldsymbol{\beta} + 2\boldsymbol{B}\boldsymbol{x_0} = \boldsymbol{0}$) is true. BH showed that a $100(1-\alpha)\%$ confidence region for $\boldsymbol{x_0}$ is the set of all $\boldsymbol{x}$ such that

$$\widehat{\boldsymbol{\delta}_{\boldsymbol{x}}}' \widehat{\boldsymbol{V}_{\boldsymbol{x}}}^{-1} \widehat{\boldsymbol{\delta}_{\boldsymbol{x}}} \leq kF(1-\alpha, k, n-p), \tag{2}$$

where $\widehat{\boldsymbol{\delta}_{\boldsymbol{x}}} = \widehat{\boldsymbol{\beta}} + 2\widehat{\boldsymbol{B}}\boldsymbol{x}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{B}}$ are least squares estimates of $\boldsymbol{\beta}$ and $\boldsymbol{B}$ respectively, and $F(1-\alpha, k, n-p)$ is

the upper $100(1-\alpha)$ percentile of the F-distribution with k and $n-p$ degrees of freedom. Here, $\widehat{\boldsymbol{V}_{\boldsymbol{x}}}$ is the estimate of $\boldsymbol{V}_{\boldsymbol{x}}$, the variance of $\widehat{\boldsymbol{\delta}_{\boldsymbol{x}}}$, n is the number of runs, and p is the number of regression coefficients including the intercept. Note that although equation (2) has a closed-form solution, computing the BH confidence regions is a difficult algebraic task (Del Castillo and Cahya 2001). Also, BH discussed in their paper that the methodology can be extended to any other (nonquadratic) response surface model as long as the models are linear in the parameters.

A confidence region for a stationary point is of interest if the stationary point happens to be a feasible optimal point. However, in many instances the stationary point is a saddle point which is not useful from an optimization perspective. Even when the stationary point is indeed an optimal point, this point is sometimes located outside the region of operability hence the optimal solution cannot be implemented. Therefore, in practice, solutions should be limited to be inside of a constrained region of interest. Furthermore, some problems are naturally constrained, such as mixture experiments where the fraction of the mixture ingredients should add up to one. Stablein, Carter, and Wampler (1983) proposed a methodology for computing the confidence regions for constrained optimal points. These authors modified the BH approach by using the Lagrange multiplier approach to incorporate the constraints. However, their method assumes the lagrangian multipliers as constants, hence ignoring the sampling variability of the multipliers. This assumption is not valid since the multipliers are dependent on the model parameter estimates and these estimates are subject to sampling variability (Peterson 1999).

In a recent paper, Peterson, Cahya, and Del Castillo (2001) (hereafter referred to as PCD) proposed a generalized approach that has some advantages over the Stablein et al. (1983) method. This approach does not use lagrangian multipliers, so it avoids the technical difficulty of incorporating the sampling variability of the multipliers . Moreover, the method generalizes the quadratic polynomial model required in Stablein et al. (1983) into any models that are linear in the parameters. This is especially beneficial in mixture experiments where "exotic" functions of the factors sometimes provide a better fit than quadratic polynomials (Khuri and Cornell 1996, chap. 9). However, as it will be discussed in more detail in section 2,

the PCD methodology has some practical drawbacks related to computational speed and the accuracy of the resulting confidence regions. In this paper, we will discuss an approach to overcome these drawbacks. A review of the approach by PCD is given in section 2. Our proposed approach and some performance comparisons are given in sections 3 and 4, respectively. An aspect that cannot be separated from the computation of confidence regions is their visualization. Section 5 discusses some concerns regarding visualizing the resulting confidence regions, and a multi-factor example is used as an illustration. Section 6 gives some conclusions and future work.

2 Review of the Peterson et al. (2001) method

Let us assume that the response surface under consideration is linear in the parameters, i.e., it can be modeled as

$$y = \beta_0 + \boldsymbol{z}(\boldsymbol{x})'\boldsymbol{\theta} + \epsilon, \tag{3}$$

where y is the response variable, β_0 is the intercept term, $\boldsymbol{z}(\boldsymbol{x})$ is a $(p-1)\times 1$ vector valued function of the factor $\boldsymbol{x}$ ($k\times 1$ vector), $\boldsymbol{\theta}$ is a $(p-1)\times 1$ vector of regression coefficients, and ϵ is the error term that follows a normal distribution with mean 0 and variance σ^2.

Without loss of generality, suppose that we are to minimize the response y. Let $\boldsymbol{x}_0$ and $\eta(\boldsymbol{\theta})$ be defined as

$$\boldsymbol{z}(\boldsymbol{x_0})'\boldsymbol{\theta} = \min_{\boldsymbol{x}\in\mathcal{R}} \boldsymbol{z}(\boldsymbol{x})'\boldsymbol{\theta} = \eta(\boldsymbol{\theta}), \tag{4}$$

where $\mathcal{R}$ is the constraining region (for unconstrained optimization $\mathcal{R}$ is simply the space of $\boldsymbol{x}$). Note that $\boldsymbol{x_0}$ is the true minimizer of the response surface and it is unknown. To compute the $100(1-\alpha)\%$ confidence region of $\boldsymbol{x_0}$, one can find the set of all $\boldsymbol{x}$-values such that the null hypothesis $H_0 : \eta(\boldsymbol{\theta}) - \boldsymbol{z}(\boldsymbol{x})'\boldsymbol{\theta} = 0$ is not rejected at level α. PCD proposed a way to compute the $100(1-\alpha)\%$ confidence interval for $\left(\eta(\boldsymbol{\theta}) - \boldsymbol{z}(\boldsymbol{x})'\boldsymbol{\theta}\right)$ and rejecting H_0 simply by checking whether this confidence interval excludes 0. Since $\left(\eta(\boldsymbol{\theta}) - \boldsymbol{z}(\boldsymbol{x})'\boldsymbol{\theta}\right) \leq 0$ for all $\boldsymbol{\theta}$, the lower bound of the confidence interval is always less than zero. Therefore, the confidence interval for $\left(\eta(\boldsymbol{\theta}) - \boldsymbol{z}(\boldsymbol{x})'\boldsymbol{\theta}\right)$ excludes zero if the upper bound is less than zero.

This upper bound is given by (see PCD for the derivation):

$$U_B = \min_{\boldsymbol{w} \in \mathcal{R}} b_x(\boldsymbol{w}), \tag{5}$$

where

$$b_x(\boldsymbol{w}) = \Big(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x})\Big)'\widehat{\boldsymbol{\theta}} + \sqrt{c_\alpha^2 \left[\Big(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x})\Big)'\widehat{\boldsymbol{V}}\Big(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x})\Big)\right]}. \tag{6}$$

Here, $\widehat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{V}}$ is the estimated variance-covariance matrix of $\boldsymbol{z}(\boldsymbol{x})'\boldsymbol{\theta}$, c_α^2 is the $100(1-\alpha)\%$ upper percentile of an F-distribution with k and $n-p$ degrees of freedom.

Equation (5) allows us to construct the $100(1-\alpha)\%$ confidence region for the optimal factor level $\boldsymbol{x_0}$ in a straightforward manner. This confidence region is denoted as $\mathcal{C}_R$ for brevity. A point $\boldsymbol{x} \in \mathcal{R}$ does not belong to $\mathcal{C}_R$ if $U_B < 0$. However, it is not necessary to perform the minimization in (5) and get the minimum value U_B. To discard a point $\boldsymbol{x}$ from being inside the $\mathcal{C}_R$, it is only necessary to find a single other point $\boldsymbol{w} \in \mathcal{R}$ such that the function $b_x(\boldsymbol{w}) < 0$, since $U_B \leq b_x(\boldsymbol{w})$ for all $\boldsymbol{w} \in \mathcal{R}$. This clearly would imply that $U_B < 0$. Furthermore, for a response surface that satisfies the following conditions: (i) the response is quadratic (hence it is differentiable); (ii) the response is strictly convex in a statistical sense, i.e. the matrix of quadratic coefficients $\boldsymbol{B}$ is positive definite (p.d.) for all $\theta \in \mathcal{C}$ where $\mathcal{C} = \{\boldsymbol{\theta} : (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\widehat{\boldsymbol{V_\theta}}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq c_\alpha^2\}$ is the $100(1-\alpha)\%$ confidence region for $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{V_\theta}}$ is an estimate of $\boldsymbol{V_\theta}$, the variance covariance matrix of $\widehat{\boldsymbol{\theta}}$, PCD proposed the following two-step derivative approach:

Step 1 (BH Step): For all points $\boldsymbol{x}$ in the interior of $\mathcal{R}$, select confidence region points according to the BH criterion.

Step 2: For all points $\boldsymbol{x}$ on the boundary of $\mathcal{R}$, search for the first $\boldsymbol{d}$-value on $\mathcal{D}$ such that $b'_x(\boldsymbol{x};\boldsymbol{d}) < 0$, where $\mathcal{D} = \{\boldsymbol{d} : \boldsymbol{d}'\boldsymbol{d} = 1, \boldsymbol{x} + h\boldsymbol{d} \in \mathcal{R} \text{ for small } h > 0\}$ and $b'_x(\boldsymbol{w};\boldsymbol{d}) = \lim_{h \to 0^+}[b_x(\boldsymbol{w}+h\boldsymbol{d}) - b_x(\boldsymbol{w})]/h$.

PCD showed that, for differentiable response models, $b'_x(\boldsymbol{x};\boldsymbol{d}) = \boldsymbol{d}'\boldsymbol{D}(\boldsymbol{x})\widehat{\theta} + c_\alpha[\boldsymbol{d}'\boldsymbol{D}(\boldsymbol{x})\widehat{\boldsymbol{V}}\boldsymbol{D}(\boldsymbol{x})'\boldsymbol{d}]^{1/2}$ where $\boldsymbol{D}(\boldsymbol{x})$ is the $k \times (p-1)$ matrix of partial derivatives of $\boldsymbol{z}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$. For subsequent discussions, the first step of this derivative approach will be referred to as the BH Step.

In general, some response models do not satisfy differentiability nor convexity assumptions (e.g. Becker 1968; Cornel and Gorman 1978). We note that even though a fitted quadratic polynomial is strictly convex (i.e. $\widehat{\boldsymbol{B}}$ is p.d.), $\boldsymbol{B}$ might not be p.d. everywhere on $\mathcal{C}$ (see Peterson 1993 for checking whether $\boldsymbol{B}$ is statistically positive definite). For general response surface models, a two-step derivative-free approach was then proposed:

Step 1: Compute the estimated optimal point $\widehat{\boldsymbol{x}_0}$. Set $\boldsymbol{w} = \widehat{\boldsymbol{x}_0}$ and use $b_x(\widehat{\boldsymbol{x}_0})$ as a criterion to reject as many points in $\mathcal{R}$ as possible.

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

We hasten to point out that for response surfaces that are differentiable on $\mathcal{R}$ but not necessarily convex, an efficient algorithm can be constructed by adding the BH Step (for checking points in the interior of R) in between Steps 1 and 2 of the derivative-free approach.

Problems Attributed to a Low Grid Resolution

In practice, the computation of the confidence region for the optimal factor-level is initialized by first "discretizing" the constrained region $\mathcal{R}$ into grid points (from here onward, what we mean by $\mathcal{R}$ is its discretized version). The second step of the PCD's derivative-free approach described above is a very slow process since to reject $\boldsymbol{x} \in \mathcal{R}$, $b_x(\boldsymbol{w})$ needs to be computed repeatedly until $b_x(\boldsymbol{w}) < 0$. In a worst case scenario, $b_x(\boldsymbol{w})$ needs to be computed for every $\boldsymbol{w} \in \mathcal{R}$. Decreasing the grid resolution reduces the number of grid points and will expedite the computation process at the expense of a lower quality of the confidence region displays.

A more crucial problem of using a low grid resolution is that it reduces the accuracy of the resulting confidence region. To clarify the discussion about this accuracy problem, some definitions will be given. A point $\boldsymbol{w} \in \mathcal{R}$ is defined as a *dominating point* of $\boldsymbol{x} \in \mathcal{R}$, if $b_x(\boldsymbol{w}) < 0$ (next section gives a theorem regarding a special property of this dominance relation). The *domination region* $\mathcal{W}_x$ is defined as the set of $\boldsymbol{w}$ points that dominate $\boldsymbol{x}$, i.e., $\mathcal{W}_x = \{\boldsymbol{w} \in \mathcal{R} \mid b_x(\boldsymbol{w}) < 0\}$. A point $\boldsymbol{x}$ belongs to the confidence

region $\mathcal{C}_R$ if $\mathcal{W}_x$ is empty. The accuracy problem arises when the domination regions $\mathcal{W}_x$ for some points $\boldsymbol{x}$ are too small compared to the grid resolution, i.e., none of the grid points are inside of the domination region (this problem is illustrated in Figure 1). Then, the algorithm is unable to locate points inside these $\mathcal{W}_x$ regions. The algorithm will erroneously conclude that the regions $\mathcal{W}_x$ are empty, and the associated points $\boldsymbol{x}$ will be falsely classified as being inside the confidence region $\mathcal{C}_R$. As a consequence, the resulting confidence region will be bigger than it actually is. Increasing the grid resolution will improve the accuracy, but, then again, we will face the problem of slow computation. Therefore, we need a better algorithm that is fast and more importantly accurate, even if we are using a low grid resolution.

<Figure 1 about here>

The PCD's two-step derivative approach, unlike the derivative-free counterpart, does not utilize grid point comparisons through the $b_x(\boldsymbol{w})$ function. Thus, the derivative approach is independent of grid point resolution. However, since the set $\mathcal{D}$ used in its second step consists of infinite number of directions $\boldsymbol{d}$, this set $\mathcal{D}$ needs to be discretized as well to limit the number of computations. We note that this second step of derivative approach affects boundary points only, thus, in general, the derivative approach is more accurate than the derivative-free approach. The use of this PCD's derivative approach, however, is limited to quadratic responses that are strictly convex in a statistical sense.

3 Proposed Approach

In this section, we discuss some modifications to the PCD's methodology that results in a faster algorithm and improved accuracy. We first state a theorem that underlies the modifications and then the proposed approach will be discussed.

Recall that in order to show that a point $\boldsymbol{x} \in \mathcal{R}$ does not belong to the confidence region $\mathcal{C}_R$, we need to show the existence of a nonempty dominating region $\mathcal{W}_x$, i.e., we need to find a point $\boldsymbol{w} \in \mathcal{R}$

such that $b_x(\boldsymbol{w}) < 0$. Obviously this search process is computationally extensive since the search space is the entire (discretized) $\mathcal{R}$. Fortunately, the $b_x(\boldsymbol{w})$ function has a special property that can be used to improve the computational efficiency and also the accuracy of the algorithm. We state the property in Theorem 1 and its corollaries:

Theorem 1 *Let $\boldsymbol{x}$ and $\boldsymbol{w}$ be distinct points inside the constrained region $\mathcal{R}$. We say $\boldsymbol{w}$ dominates $\boldsymbol{x}$ if $b_x(\boldsymbol{w}) < 0$ and write $\boldsymbol{w} \mapsto \boldsymbol{x}$. The binary relation $\mapsto$ on the set $\mathcal{R}$ is transitive, i.e., if $\boldsymbol{w} \mapsto \boldsymbol{x}$ and $\boldsymbol{x} \mapsto \boldsymbol{y}$ (where $\boldsymbol{y} \in \mathcal{R}$ and $\boldsymbol{x} \neq \boldsymbol{y} \neq \boldsymbol{w}$), then $\boldsymbol{w} \mapsto \boldsymbol{y}$.*

Proof. See the Appendix.

Corollary 1 *Let $\mathcal{E}$ be the set of points in $\mathcal{R}$ that are already eliminated (i.e. determined to be outside the $\mathcal{C}_R$) at the current iteration process and let $\mathcal{R}\backslash\mathcal{E} = \mathcal{R} - \mathcal{E}$ (i.e. the set of points not eliminated so far). Then, for discarding the remaining $\boldsymbol{x} \in \mathcal{R}\backslash\mathcal{E}$, we just need to consider $\boldsymbol{w} \in \mathcal{R}\backslash\mathcal{E}$ as opposed to considering $\boldsymbol{w} \in \mathcal{R}$.*

Proof. Let $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, $\ldots$ $\boldsymbol{x_j}$ with j a positive integer, be points in the set $\mathcal{E}$ such that $\boldsymbol{x_1} \mapsto \boldsymbol{x_2} \mapsto \cdots \mapsto \boldsymbol{x_j}$ and no other points $\boldsymbol{w} \in \mathcal{E}$ such that $\boldsymbol{w} \mapsto \boldsymbol{x_1}$. Let $\boldsymbol{y}$ be a point in the set $\mathcal{R}\backslash\mathcal{E}$ such that $\boldsymbol{x_j} \mapsto \boldsymbol{y}$. Since $x_1 \in \mathcal{E}$, it implies that there exists a point $\boldsymbol{w} \in \mathcal{R}\backslash\mathcal{E}$ such that $\boldsymbol{w} \mapsto \boldsymbol{x_1}$. However, Theorem 1 implies that $\boldsymbol{w} \mapsto \boldsymbol{y}$. Therefore, the points $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, $\ldots$ $\boldsymbol{x_j}$ are not needed for eliminating $\boldsymbol{y}$.

Corollary 2 *For all $\boldsymbol{x} \notin \mathcal{C}_R$, there exists $\boldsymbol{w} \in \mathcal{C}_R$ such that $\boldsymbol{w} \mapsto \boldsymbol{x}$.*

Proof. Immediate from Corollary 1, since $\mathcal{R}\backslash\mathcal{E}$ converges to $\mathcal{C}_R$.

Corollary 1 has a direct impact on improving the computational speed of the algorithm. Once some points have been determined to be outside of the confidence region, they can be discarded from the search space. In the next iteration, we can concentrate on the points that are not yet eliminated i.e. the set $\mathcal{R}\backslash\mathcal{E}$. As more points are eliminated in each step of the algorithm, the size of the search space $\mathcal{R}\backslash\mathcal{E}$ shrinks reducing the number of $b_x(\boldsymbol{w})$ computations needed. This technique can be incorporated into the second

step of the PCD's two-step derivative-free approach to get a significant improvement in computational speed. However, this technique alone does not improve the accuracy of the algorithm. In subsequent paragraphs, accuracy improvements are discussed.

As defined in the previous section, $\boldsymbol{w}$ is a dominating point of $\boldsymbol{x}$ if $b_x(\boldsymbol{w}) < 0$ and the dominating region of $\boldsymbol{x}$, $\mathcal{W}_x$, is defined as the set of points $\boldsymbol{w}$ that dominate $\boldsymbol{x}$. From (4), note that the value of $b_x(\boldsymbol{w})$ when $\boldsymbol{w} = \boldsymbol{x}$ is zero. If the $b_x(\boldsymbol{w})$ function in a small neighborhood around $\boldsymbol{w} = \boldsymbol{x}$ is a smooth downhill surface, then as we move along downhill, the dominating region $\mathcal{W}_x$ $(= \{\boldsymbol{w} \in \mathcal{R} \mid b_x(\boldsymbol{w}) < 0\})$ will be reached right after passing the point $\boldsymbol{w} = \boldsymbol{x}$. Therefore, to find $\mathcal{W}_x$, it is logical to search initially in the neighborhood of $\boldsymbol{x}$. An example of this property can be seen from Figure 1 where the dominating region $\mathcal{W}_{\tilde{x}}$ is located just next to $\tilde{\boldsymbol{x}}$. In general, however, there are instances where the dominating region $\mathcal{W}_x$ is distant from $\boldsymbol{x}$. To differentiate between these two types of dominating regions, they will be referred as *adjacent dominating regions* and *distant dominating regions*, respectively. Finding distant dominating regions is harder since one does not have any information about its approximate location. In what follows, the methodology for finding adjacent dominating regions is first discussed followed by a methodology for finding distant dominating regions.

3.1 Finding Adjacent Dominating Regions

In order to find an adjacent dominating region $\mathcal{W}_x$, a simple but inefficient way is to "sweep" the neighborhood of $\boldsymbol{x}$ until a dominating point $\boldsymbol{w} \in \mathcal{W}_x$ is found. This can be done by finding a (minimizing) direction $\boldsymbol{d} \in \mathcal{D}$, where $\mathcal{D} = \{\boldsymbol{d} : \boldsymbol{d}'\boldsymbol{d} = 1, \boldsymbol{x} + h\boldsymbol{d} \in \mathcal{R} \text{ for small } h > 0\}$, such that $b_x(\boldsymbol{x} + h\boldsymbol{d}) < 0$. However, Corollary 2 of Theorem 1 suggests that a more efficient method can be proposed. The corollary states that for each point $\boldsymbol{x} \notin \mathcal{C}_R$, there exists an associated dominating point $\boldsymbol{w} \in \mathcal{C}_R$. Therefore, since the dominating region $\mathcal{W}_x$ is adjacent to $\boldsymbol{x}$ (by our assumption) and it is known that $\mathcal{W}_x$ intersects $\mathcal{C}_R$ by Corollary 2, then $\mathcal{W}_x$ will likely reside between $\boldsymbol{x}$ and $\mathcal{C}_R$. Hence, the search can be concentrated in the direction toward the $\mathcal{C}_R$. Figure 1 shows that $\mathcal{W}_{\tilde{x}}$ is located between the current point of interest

$\tilde{\boldsymbol{x}}$ and the true confidence region (solid dots). Of course, the $\mathcal{C}_R$ is not known in advance. However, it is known that the estimated optimal point $\widehat{\boldsymbol{x_0}}$ is inside the $\mathcal{C}_R$ and thus, $\widehat{\boldsymbol{x_0}}$ can be used as a surrogate of the direction where the unknown $\mathcal{C}_R$ is located. In other words, one can use the direction $(\widehat{\boldsymbol{x_0}} - \boldsymbol{x})$ as an initial search direction for finding the adjacent dominating region $\mathcal{W}_x$. Figure 2 illustrates how this search direction from $\tilde{\boldsymbol{x}}$ to $\widehat{\boldsymbol{x_0}}$ passes through the dominating region $\mathcal{W}_{\tilde{x}}$. Any point $\boldsymbol{w}$ in the small neighborhood of $\tilde{\boldsymbol{x}}$ along this search direction can be selected to make $b_{\tilde{x}}(\boldsymbol{w}) < 0$ and hence, the search process will terminate in just one iteration.

<Figure 2 about here>

In general, however, the direction $(\widehat{\boldsymbol{x_0}} - \boldsymbol{x})$ does not always pass through $\mathcal{W}_x$. For example, there are situations when the resulting confidence region is scattered all over the space of $\mathcal{R}$ (see Figure 4 for an example). As a consequence, the estimated optimal point $\widehat{\boldsymbol{x_0}}$ is a poor surrogate of the location of the confidence region. Another example is when the dominating region $\mathcal{W}_x$ is a very narrow ridge and hence, there is only a slight chance that the direction $(\widehat{\boldsymbol{x_0}} - \boldsymbol{x})$ passes through the $\mathcal{W}_x$. In such cases, a more elaborate search algorithm is needed. Since the $b_x(\boldsymbol{w})$ function is allowed to be nonlinear and nondifferentiable, one can resort to derivative-free nonlinear programming techniques available in the optimization literature. One of the techniques is the Downhill Simplex method of Nelder and Mead (1965).

The use of the Nelder-Mead simplex method has generated some debate in the optimization literature (see Wright 1996; McKinnon 1998; Lagarias, Reeds, Wright, and Wright 1998 for recent discussions). The critics have been wary that the convergence of this method is unpredictable. Although there have been numerous successful applications reported in the literature (see Walters, Parker, Morgan, and Deming 1991 for references), it has been shown that the method can converge to a nonminimizing point (Woods 1985; McKinnon 1998). However, for finding adjacent dominating regions $\mathcal{W}_x$, finding a (global) min-

imizing point of the $b_x(\boldsymbol{w})$ function is not of interest. Recall that $b_x(\boldsymbol{w})$ has already a zero value at $\boldsymbol{w} = \boldsymbol{x}$. Thus, a slight improvement in minimizing the $b_x(\boldsymbol{w})$ function is all that is needed for finding the dominating regions. More importantly, a desirable feature of the Nelder-Mead algorithm is that it produces significant improvement for the first few iterations (Lagarias et al. 1998, p. 114). This leads to a very fast algorithm for finding adjacent dominating regions.

There are some modifications needed for the Nelder-Mead simplex method to work for finding adjacent dominating regions. First, in addition to the default termination (convergence) criteria, the method should terminate whenever it evaluates a point $\boldsymbol{w} \in \mathcal{R}$ such that $b_x(\boldsymbol{w}) < 0$. Second, rather than choosing arbitrary points to form a simplex required for initializing the Nelder-Mead search, some points in the direction towards the estimated optimal point $\widehat{\boldsymbol{x_0}}$ should be chosen. As previously discussed, these points are more likely to be inside the dominating region. If one of these initial points is indeed inside the associated dominating region, then search will terminate right away. Care should be taken, however, that the chosen points should form a nondegenerate simplex. Third, the Nelder-Mead method is usually tailored for unconstrained optimizations whereas our interest is on the constrained region $\mathcal{R}$. A common method to convert unconstrained optimizations into constrained ones is to use Penalty Functions (Camp 1955; Fiacco and McCormick 1964). In our experiments, simply assigning a large positive functional value whenever $\boldsymbol{w}$ is outside $\mathcal{R}$ (assuring $\boldsymbol{w}$ will never be in the solution set) worked well and was utilized.

3.2 Finding Distant Dominating Regions

Finding distant dominating regions is hard to deal with because the existence of distant dominating regions implies that the $b_x(\boldsymbol{w})$ function has multiple local minima. One local minimum is at $\boldsymbol{w} = \boldsymbol{x}$ since we have $b_x(\boldsymbol{w}) \geq 0$ in the neighborhood of $\boldsymbol{x}$, otherwise we could find an adjacent dominating region. Providing that the size of the dominating region is bigger than the grid resolution, an effective but slow method is to use the "brute-force" grid search as in the second step of PCD's two-step derivative-free approach (as discussed previously, the computation speed of this second step can be improved by selecting

$\boldsymbol{w}$ only from the set $\mathcal{R}\backslash\mathcal{E}$ rather than from the entire $\mathcal{R}$). All distant dominating regions that we observed in numerical examples fell into this category. Therefore, it is our belief that this grid search is sufficient for handling distant dominating regions in practice. However, in theory, a distant dominating region smaller than the grid resolution could happen. One possible scenario is that when a point $\boldsymbol{x} \notin \mathcal{C}_R$ gets closer to the boundary of $\mathcal{C}_R$, its dominating region $\mathcal{W}_x$ gets smaller until it disappears when $\boldsymbol{x}$ belongs to the boundary of $\mathcal{C}_R$. Thus, when $\boldsymbol{x} \notin \mathcal{C}_R$ is sufficiently close to the boundary of $\mathcal{C}_R$, then it is possible that its dominating region is smaller than the grid resolution. For a point $\boldsymbol{x}$ to be close to the boundary of $\mathcal{C}_R$, however, it usually requires that the grid points are placed close to each other (high grid resolution) and, consequently, the associated distant dominating region has to be even smaller to not be detected by the grid search process. For experimental purposes, artificial small distant dominating regions can be created by using a suitably low grid resolution and placing arbitrary points (in between regular grid points) sufficiently close to the boundary of $\mathcal{C}_R$.

As a precaution step in preventing against small distant dominating regions, a "safety check" can be added whenever the brute-force grid search fails to locate distant dominating regions. Once the grid search has been performed, the values of the $b_x(\boldsymbol{w})$ function across the $\mathcal{R}\backslash\mathcal{E}$ region are known. If there exists a small distant dominating region, it is expected that grid points adjacent to this region should have small but positive $b_x(\boldsymbol{w})$ values. Therefore, we can use one of these points as a starting point for the modified Nelder-Mead simplex method in a search similar to that suggested for finding the adjacent dominating regions.

An immediate question is how the starting point for the Nelder-Mead search is chosen. Let $\boldsymbol{d}$ be a direction from $\boldsymbol{x}$ to $\boldsymbol{w}$, i.e. $\boldsymbol{d} = (\boldsymbol{w} - \boldsymbol{x})$, and let

$$b'_x(\boldsymbol{w}; \boldsymbol{d}) = \lim_{h \to 0^+} \frac{b_x(\boldsymbol{w} + h\boldsymbol{d}) - b_x(\boldsymbol{w})}{h}$$

be a directional derivative of $b_x(\boldsymbol{w})$ with respect to $\boldsymbol{d}$. Here, $b'_x(\boldsymbol{w}; \boldsymbol{d})$ can be numerically approximated by using a sufficiently small value of h. Let $\mathcal{D}_x$ be the set of points $\boldsymbol{w} \in \mathcal{R}\backslash\mathcal{E}$ where the surface of $b_x(\boldsymbol{w})$ is descending with respect to $\boldsymbol{d}$, i.e, $\mathcal{D}_x = \{\boldsymbol{w} \in \mathcal{R}\backslash\mathcal{E} : \; b'_x(\boldsymbol{w}; \boldsymbol{d}) < 0, \; \boldsymbol{d} = \boldsymbol{w} - \boldsymbol{x}\}$. A point $\boldsymbol{w}^*$ that is

adjacent to the small distant dominating region can be found by performing the minimization:

$$\boldsymbol{w}^* = \arg \min_{\boldsymbol{w} \in \mathcal{D}_x} b_x(\boldsymbol{w}).$$

We note that, since the $b_x(\boldsymbol{w})$ values are already computed for every $\boldsymbol{w} \in \mathcal{R} \backslash \mathcal{E}$ during the brute-force grid search, $\boldsymbol{w}^*$ can be found by simply sorting the $b_x(\boldsymbol{w})$ values (for $\boldsymbol{w} \in \mathcal{D}_x$) in an ascending order. This process of finding $\boldsymbol{w}^*$ is illustrated in Figure (3) for a one-dimensional case. The modified Nelder-Mead search can then be started from $\boldsymbol{w}^*$ to find the dominating region.

<Figure 3 about here>

To improve the computational efficiency of this safety check step, the search for $\boldsymbol{w}^*$ can be limited to points $\boldsymbol{w} \in \mathcal{R} \backslash \mathcal{E}$ having a small value of $b_x(\boldsymbol{w})$. This can be done by imposing an upperbound $\delta > 0$ such that the search space is limited to the set $\{\boldsymbol{w} \in \mathcal{R} \backslash \mathcal{E} : b_x(\boldsymbol{w}) < \delta\}$. As an example, δ can be chosen as $c\,[\max_{\boldsymbol{w} \in \mathcal{R} \backslash \mathcal{E}} b_x(\boldsymbol{w})]$ where $c \in (0, 1)$. In our experimentations, when small distant dominating regions were artificially generated, a value $c = 0.2$ worked satisfactorily in eliminating all of the small distant dominating regions while the resulting time increment was only a small fraction of the overall computation time. It should be emphasized that in none of our numerical examples, based on actual experiments, was the safety check necessary.

3.3 Proposed Algorithm

Based on the discussion in the previous paragraphs, the proposed algorithm consists of four steps:

Step 1: Compute the estimated optimal point $\widehat{\boldsymbol{x}_0}$. Set $\boldsymbol{w} = \widehat{\boldsymbol{x}_0}$ and use $b_x(\widehat{\boldsymbol{x}_0})$ as a criterion to reject as many points in $\mathcal{R}$ as possible.

Step 2: For any $\boldsymbol{x}$ not rejected in Step 1, perform a local search using the modified Nelder-Mead algorithm.

Step 3: For any $\boldsymbol{x}$ not rejected in the previous steps, search for the first $\boldsymbol{w} \in \mathcal{R} \backslash \mathcal{E}$ such that $b_x(\boldsymbol{w}) < 0$.

Step 4: For any $\boldsymbol{x}$ not rejected in the previous steps, perform a safety check.

The first step is exactly the same as the first step in the PCD's derivative-free algorithm. The second step is the local search for finding adjacent dominating regions. The third step is similar to the second step of PCD, except that Corollary 1 is used for speeding up the process. The final step is a precaution step in case the grid search in the third step fails to detect small distant dominating regions. Next we discuss the performance of our proposed algorithm in comparison to the PCD's algorithm.

4 Performance Comparisons

The goal of the proposed approach is to improve two performance measures in the PCD methodology: computational speed and accuracy. An example from the literature will be used for the comparisons. For assessing accuracy, an extensive simulation to compute the coverage probability was done and will be reported in this section.

4.1 Computational Speed Comparison

For computational speed comparisons a mixture experiment that studies the effect of three chemical substances on the glass transition temperature (Frisbee and McGinity 1994) was used. The objective of the study was to find the optimal factor levels associated with minimizing glass transition temperature. The chemical substances were *Pluronic F68* (x_1), *polyoxyethelene 40 monostearate* (x_2), and *polyoxyethelene sorbitan fatty acid ester NF* (x_3). 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 gave the best fit was an H1 Becker model (Becker 1968):

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} \min(x_1, x_2) + \beta_{13} \min(x_1, x_3) + \beta_{23} \min(x_2, x_3) + \epsilon, \tag{7}$$

where y equals the observed glass transition temperature (^oC). The mean squared 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 is 96.4%. Note that the response model in (7) is not everywhere differentiable and is subject to the mixture constraints. Thus use of the BH (derivative based) Step would not be convenient to implement, requiring careful checking of each $\boldsymbol{x}$-point to make sure that $x_i \neq x_j$ or that $\boldsymbol{x}$ is not on the boundary of mixture experimental region. Furthermore, for any mixture model, use of the BH Step requires that the $x_1 + \ldots + x_k = 1$ constraint be eliminated by reducing the model to $(k-1)$ independent factors.

The confidence regions for the optimal factor levels were computed for three different grid resolutions: 3721, 10201, and 40401 grid points, respectively, inside a unit square (the constrained region for the experiment is a simplex region which is a triangle inside a unit square). The computation times are tabulated in Table 1 and the confidence regions are plotted in Figure 4. All computations were done using a Pentium III based computer (800 MHz with 256 Mb of RAM, Windows 2000 operating system) and both PCD (two-step derivative-free approach) and the proposed approach were implemented in Matlab. For comparison purposes, the computation times in Table 1 are normalized by dividing by the computation time using our proposed approach with the lowest grid resolution (the computation time is 63.2 seconds). As it can be seen from the table, the proposed approach is more than three times faster than the two-step derivative-free approach of PCD. More importantly, the resulting confidence regions provided by the proposed approach have an accuracy that is independent of the grid resolution (see figure 4). Higher accuracy is observed for the two-step derivative-free approach when grid resolution is increased, but higher grid resolution will result in an unnecessarily lengthy computation time.

<Table 1 about here>

<Figure 4 about here>

Table 2 gives a breakdown of the computation times for the proposed approach. The times are divided into two parts: the time for performing the first two steps of the algorithm and the time for performing the last two steps. Recall that the last two steps are needed for eliminating points that have distant dominating regions. As can be observed from the table, most of the time is spent performing the last two steps which eliminate only a small, if not null, portion of points. Thus, if it is known that the problem at hand does not have any distant dominating regions, one can stop the algorithm after the first two steps. To determine whether the problem at hand does not involve distant dominating regions, one can try the following simple approach. Using a suitably low grid resolution, compute the confidence region of interest twice using the first two steps and then the full four steps. If the resulting confidence regions are similar, then there is a good indication that the last two steps are not necessary. Then, recompute the confidence regions using only the first two steps, but with a higher grid resolution, to produce the confidence regions with desired display quality. In the mixture example, the last two steps are necessary since the resulting confidence regions are different (see figure 5). This technique is useful especially for high dimensional problems where the total number of grid points is large even though a low grid resolution is used.

<Table 2 about here>

<Figure 5 about here>

4.2 Coverage Rate Simulation

It was shown in the previous discussion that the proposed approach is accurate and its accuracy is independent of the grid resolution. To assess the accuracy in more detail, an extensive simulation similar to that in PCD was performed. A standard quadratic model as in (1) was used with two factors along with spherical constraint region $\mathcal{R} = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} \leq 1\}$. The response surface was varied by modifying the eigenvalues of the quadratic coefficient matrix $\boldsymbol{B}$. Five eigenvalue pairs were used for the simulation: (0.2 0.2), (0.2 7.0), (2.0 2.0), (2.0 7.0), and (13.0 15.0). A higher eigenvalue characterizes a more peaked

response surface and asymmetry in an eigenvalue pair characterizes surface elongation. The locus of the true optimum x_0 was varied from the center of $\mathcal{R}$, to the boundary of $\mathcal{R}$, and then outside of $\mathcal{R}$ in order to assess the effects on the coverage rates. For each of the simulation combinations, 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 coverage rates, with a confidence level of 95%, were computed for four methods: (i) PCD's two-step derivative-free approach (denoted by *PCD2*); (ii) the proposed approached described in section 3 (denoted by *Sec.3*); (iii) PCD's two-step derivative approach (denoted by *PCD1*); (iv) PCD's two-step derivative approach with an addition of BH Step in between steps 1 and 2 (denoted by *PCD2+*). A low grid resolution of 400 points within a unit square was used for methods Sec.3 and PCD2. This grid resolution corresponds to, approximately, a total of 1250 grid points inside the spherical region $\mathcal{R}$. Our conjecture was that, since a low grid resolution is used, the coverage rates of PCD2 would be higher than that of Sec.3 because the resulting confidence regions of PCD2 are larger. Since the response is quadratic (hence differentiable), the derivative-based methods, PCD1 and PCD2+, can also be used. Note that, however, although the true eigenvalues used in the simulation are all positive, the eigenvalue of 0.2 is very close to zero. Thus, the resulting $\boldsymbol{B}$ is not p.d. everywhere at 95% confidence level. However, the resulting $\boldsymbol{B}$ from the last three eigenvalue pairs of the five pairs given above are p.d. everywhere at 95% confidence level. Thus, PCD1 would be the most accurate method for these cases. For the other cases involving an eigenvalue equal to 0.2, high accuracy can be obtained by adding the BH Step to PCD2 (i.e. PCD2+). This BH Step guarantees that points having a small adjacent dominating region are properly eliminated. To protect against small distant dominating regions, a higher grid resolution of 2500 grid points within a unit square was used for PCD2+. Thus, this PCD2+ method is expected to have a high accuracy across the simulation combinations.

<Table 3 about here>

The resulting coverage rates are given in Table 3. As it was expected, PCD's two-step derivative-free approach (PCD2) has higher coverage rates compared to the proposed approach (Sec.3). The discrepancy is even more pronounced when the response surface is more peaked (this also can be verified by looking at the average coverage rates across different locus of optimal points given in the last column of Table 3). This behavior can be explained as follows. When the underlying response is peaked, the $b_x(\boldsymbol{w})$ function tends to be peaked as well. Consequently, when the $b_x(\boldsymbol{w})$ function is peaked, the occurrence of small dominating regions is more likely. PCD1, PCD2+, and Sec.3 methods, on the other hand, are almost identical across the board. Looking at the averages, PCD2+ has the lowest coverage rates for the most "flat" surface (eigenvalues of 0.2 and 0.2), while Sec.3 is a close second and PCD1 is noticeably larger than both PCD2+ and Sec.3. This larger average for PCD1 was expected since the resulting $\boldsymbol{B}$ is not p.d. everywhere. When the eigenvalues are the largest, PCD1 has the lowest average coverage since for this case (significantly convex surface) PCD1 is supposed to give accurate confidence regions. In overall, the proposed method (Sec.3) has the lowest or a close second average coverage rates. Therefore, in summary, this coverage rate simulation gives another strong justification that the proposed method described in Section 3 is indeed accurate. More importantly, this accuracy is obtained even when a low grid resolution is used and, unlike the two-step derivative approach, the proposed approach works for any response models that are linear in the parameters.

5 Visualization of the Confidence Regions

An ubiquitous problem when computing confidence regions is how the resulting confidence regions should be displayed, especially for high dimensional problems. By nature, we can only grasp a one-dimensional (1D) or two-dimensional (2D) object as a whole. We can rotate a 3D object to get "snapshots" from different angles and then reconstruct the snapshots to create an imaginary representation that can be

discerned by our mind. It is hard, if not possible, to create an imaginary representation of kD object where $k > 3$. The usual way is to create lower dimensional snapshots by projecting the object into a 2D or 3D space. Let us call such projections the *unconditional* confidence regions. As it is well-known, these projections do not convey entirely the multivariate nature of the k-dimensional confidence region. For example, for the sake of a simple illustration, suppose the confidence region is a hollow 3D ball with an empty interior (imagine a soccer ball filled with air). If we project this confidence region into any two-dimensional subspaces, what we see is a solid circle. Any two-dimensional projections will never indicate that the confidence region is actually hollow. But if we "slice" this confidence region, we will see hollow circles when we slice at certain levels. Let us call these slices the *conditional* confidence regions in the sense that they display the confidence region when some of the factors are fixed (conditioned) at certain levels. Therefore, when viewing high dimensional confidence regions, the capability of displaying conditional confidence regions is a valuable tool in addition to the capability of displaying unconditional confidence regions.

In order to illustrate unconditional and conditional confidence regions, we use a five factor example from an experimental study by Richert, Morr, and Cooney (1974). The study was to investigate the effect of *heat treatments* (x_1), *pH levels* (x_2), *calcium concentration* (x_3), *redox potential* (x_4), and *sodium lauryl sulfate* (x_5) upon foaming properties of whey protein concentrates (WPC). There are seven responses that were measured in the study. For illustration purposes, one of the responses was arbitrarily selected, the percentage of undenatured protein. The experimental design used was a central composite design which consisted of a one-half fraction of a 2^5 factorial design, 10 axial points, and five center points. A quadratic response was fitted with some insignificant terms omitted from the model. The omitted terms are $x_4^2,\ x_5^2,\ x_1x_5,\ x_2x_4,\ x_2x_5,\ x_3x_4,\ x_4x_5$. The full quadratic model has a p-value of 0.0459 for lack-of-fit test and *adjusted-*R^2 of 89.8%, while the reduced quadratic model has a p-value of 0.1297 for lack-of-fit test and *adjusted-*R^2 of 92.8%. Thus, the reduced model is a better model. The estimated stationary point of the system is $\{x_1$=0.40, x_2=-1.02, x_3=-0.91, x_4=5.43, x_5=-4.60$\}$. It is a saddle point

located far from the experimental region. Thus, this stationary point is useless from an optimization point of view. Suppose that our goal was to find a minimizing factor level located inside a unit spherical region. The estimated constrained optimal factor level was found to be $\{x_1$=0.73, x_2=0.69, x_3=-0.03, x_4=-0.02, x_5=-0.05$\}$. The unconditional and conditional confidence regions are given in figures 6 and 7, respectively. For the conditional confidence regions, the undisplayed factors are fixed at the estimated optimal point.

<Figure 6 about here>

<Figure 7 about here>

6 Conclusions and Further Work

The methodology proposed by Peterson et al. (2001) has some advantages over the existing methodology. The method does not use Lagrangian multipliers as in Stablein et al. (1983) methodology. It works for a wide range of response models as long as they are linear in the parameters, an attractive feature particularly for mixture experiments where nonquadratic models often provide an improved fit over quadratic ones. Another benefit of the method is that there is no limitation on the physical form of the constrained region $\mathcal{R}$. It can be a spherical, a cuboidal, a simplex, a convex hull that encloses the experimental runs, or even an irregular shaped region. This allows the method to work for the more general problems found in practice.

Given the applicability of the Peterson et al. (2001) approach for computing regions, it is necessary that the methodology works for practical implementations. However, as it is discussed in section 2, the methodology has some limitations due to computational efficiency and more importantly the accuracy of the resulting confidence regions. We proposed some modifications to the methodology that result in an improved algorithm. Discussion in section 4 shows that our proposed approach is not only faster but more importantly it is accurate and the accuracy is independent of the grid resolution.

Extensions of this methodology to multiresponse optimizations are possible, at least for two cases: (i) computing the confidence region for the location of the optimal point obtained using desirability functions (Harrington 1965; Derringer and Suich 1980); (ii) obtaining the compromised solution through a methodology described in Del Castillo (1996). For the first case, a desirability function of m responses, $d = D(y_1, \ldots, y_m; \boldsymbol{x})$, can be fit using a flexible regression model (Myers and Montgomery 1995, p. 252). The proposed confidence region approach can then be applied on the resulting regression model. This type of extension is possible since the proposed approach gives confidence regions for optima of arbitrary response functions as long as they are linear in the parameters. For the second case, Del Castillo (1996) proposed that the confidence region of the optimal point be used to "relax" the optimal point estimate into a set of points that are statistically optimal at the specified confidence level. Therefore, if the confidence regions for the optimal point of each response intersect, then the intersection gives a set of points that are (statistically) optimal with respect to all of the conflicting responses. A slight modification to this approach is a "nested" approach, where the confidence regions for each response is computed in a sequential manner rather than simultaneous as in Del Castillo (1996). The confidence region for the primary response is computed first. The resulting confidence region is then used as a constrained region for computing the confidence region for the second response. This procedure is repeated for the remaining responses. Note that this nested approach is possible since the proposed methodology for computing the confidence regions can be applied to irregular constrained regions (since the resulting confidence regions of the prior step can take any shapes). However, this approach is only meaningful when there are priorities among the responses.

An easy to use Matlab program implementing the approach in this paper is available from the second author. The current version of the program works with spherical and simplex constrained regions.

Acknowledgement

This work was partially funded by NSF grant DMI 9988563.

Appendix: Proof of Theorem 1

Let $\boldsymbol{w}$, $\boldsymbol{x}$, and $\boldsymbol{y}$ be distinct points in $\mathcal{R}$. We need to proof that the dominance relation $\mapsto$ is transitive, i.e.,

$$\boldsymbol{w} \mapsto \boldsymbol{x} \; and \; \boldsymbol{x} \mapsto \boldsymbol{y} \;\Rightarrow\; \boldsymbol{w} \mapsto \boldsymbol{y} \,. \tag{8}$$

Let

$$Q_{wx} = [(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x}))' \widehat{\boldsymbol{V}} (\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x}))]^{1/2},$$

then from (6),

$$b_x(\boldsymbol{w}) = (\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x}))' \widehat{\boldsymbol{\theta}} + c_\alpha Q_{wx} \,.$$

Note that $Q_{wx} > 0$ since $\widehat{\boldsymbol{V}}$ is positive definite. The left-hand side of (8) implies that

$$(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x}))' \widehat{\boldsymbol{\theta}} + c_\alpha Q_{wx} < 0$$

and that

$$(\boldsymbol{z}(\boldsymbol{x}) - \boldsymbol{z}(\boldsymbol{y}))' \widehat{\boldsymbol{\theta}} + c_\alpha Q_{xy} < 0 \,.$$

Adding these two inequalities we have that

$$(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{y}))' \widehat{\boldsymbol{\theta}} + c_\alpha (Q_{wx} + Q_{xy}) < 0 \,. \tag{9}$$

By contrapositive, let us assume that the right-hand side of (8) does not hold, i.e., $\boldsymbol{w} \not\mapsto \boldsymbol{y}$. This implies that

$$(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{y}))' \widehat{\boldsymbol{\theta}} \geq -c_\alpha Q_{wy} \,, \tag{10}$$

and the proof is completed if we can show that (9) does not hold. This can be done by showing that

$$-c_\alpha Q_{wy} > -c_\alpha (Q_{wx} + Q_{xy}) \,,$$

or $Q_{wy} < Q_{wx} + Q_{xy}$. Since Q_{wy}, Q_{wx}, and Q_{xy} are all positive quantities , the last inequality holds if and only if

$$Q_{wy}^2 < (Q_{wx} + Q_{xy})^2 \,.$$

The right-hand side equals to:

$$\text{RHS} = (\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x}))'\widehat{\boldsymbol{V}}(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{x})) + 2Q_{wx}Q_{xy} + (\boldsymbol{z}(\boldsymbol{x}) - \boldsymbol{z}(\boldsymbol{y}))'\widehat{\boldsymbol{V}}(\boldsymbol{z}(\boldsymbol{x}) - \boldsymbol{z}(\boldsymbol{y})).$$

Rearranging terms we have that

$$\begin{aligned} \text{RHS} &= (\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{y}))'\widehat{\boldsymbol{V}}(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{y}) + 2\boldsymbol{z}(\boldsymbol{x})'\widehat{\boldsymbol{V}}\boldsymbol{z}(\boldsymbol{x}) + 2Q_{wx}Q_{xy} \\ &> (\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{y})'\widehat{\boldsymbol{V}}(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{y})) = Q_{wy}^2. \end{aligned}$$

Thus, we have that $Q_{wy} < Q_{wx} + Q_{xy}$ and therefore, from (10) we have

$$(\boldsymbol{z}(\boldsymbol{w}) - \boldsymbol{z}(\boldsymbol{y}))'\widehat{\boldsymbol{\theta}} \geq -c_\alpha Q_{wy} > -c_\alpha(Q_{wx} + Q_{xy}),$$

which implies that (9) does not hold and the proof is completed.

References

Becker, N. G. (1968), "Models for the Response of a Mixture," *Journal of the Royal Statistical Society, Series B (Methodological)*, 30, pp. 349-358.

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, pp. 190-199.

Camp, G. D. (1955), "Inequality-Constrained Stationary-Value Problems," *Operation Research*, 3, pp. 548-550.

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, pp. 2963-2971.

——— , Chinchilli, V. M., Myers, R. H., and Campbell, E. D. (1986), "Confidence Intervals and an Improved Ridge Analysis of Response Surfaces," *Technometrics*, 28, 4, pp. 339-346.

Cornell, J. A., and Gorman J. W. (1978), "On the Detection of An Additive Blending Component in Multicomponent Mixtures," *Biometrics*, 34, pp. 251-263.

Del Castillo, E. (1996), "Multiresponse Optimization via Constrained Confidence Regions," *Journal of Quality Technology*, 28, pp. 61-70.

——— , Cahya S. (2001), "A Tool for Computing Confidence Regions on the Stationary Point of a Response Surface," to appear in The American Statistician.

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

Fiacco A. V., and McCormick, G. P. (1964), "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming," *Management Science*, 10, pp. 360-366.

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, pp. 355-363.

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

Khun, A. M. (1999), "Confidence Regions for Constrained Optima in Response Surface Experiments with Noise Factors," American Statistical Association Proceedings of the Section on Physical and Engineering Sciences, pp. 61-66.

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

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998), "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization*, 9, pp. 112-147.

McKinnon, K. I. M. (1998), "Convergence of the Nelder-Mead simplex method to a nonstationary point," *SIAM Journal of Optimization*, 9, 148-158.

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

Myers, R. H., and Montgomery, D. C. (1995), *Response Surface Methodology: Process and Product Improvement with Designed Experiments*, New York: John Wiley & Sons.

Nelder, J. A., and Mead, R. (1965), "A Simplex Method for Function Minimization," *Computer Journal*, 7, p.308-313.

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

——— (1999), "A General Approach to Inference for Optimal Conditions Subject to Constraints," American Statistical Association Proceedings of the Section on Physical and Engineering Sciences, pp. 67-72.

——— , Cahya, S., and Del Castillo, E. (2001), "A General Approach to Confidence Regions for Optimal Factor Levels of Response Surfaces," Submitted to Biometrics.

Richert, S. H., Morr, C. V., and Cooney, C. M. (1974), "Effect of Heat and Other Factors Upon Foaming Properties of Whey Protein Concentrates," *Journal of Food Science*, 39, pp. 42-48.

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

Walters, F. H., Parker, L. R., Morgan, S. L., and Deming S. N. (1991), *Sequential Simplex Optimization*, Boca Raton, FL: CRC Press.

Wright, M. H. (1996), "Direct Search Method: Once Scorned, Now Respectable," in Numerical Analysis 1995: Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis, D. F. Griffiths and G. A. Watson, eds., Addison Wesley Longman, Harlow, UK, 191-208.

Woods, D. J. (1985), *An Interactive Approach for Solving Multi-Objective Optimization Problems*, Ph.D. thesis, Tech. Report TR85-5, Department of Computational and Applied Mathematics, Rice University, Houston, TX.

Table 1: Computational speed comparison. PCD2 refers to the two-step derivative-free approach by PCD and Sec.3 refers to our proposed approach described in section 3. The computation times are normalized by dividing with the computation time using the proposed approach with low grid resolution.

	Relative Computation Time	
Grid Resolution[1]	PCD2	Sec.3
3271 pts	3.0	1.0
10201 pts	21.0	5.7
40401 pts	317.0	90.1

[1]The number of grid points inside a unit square.

Table 2: Computation times breakdowns when the confidence regions are computed using the proposed approach described in section 3 . The ratios to the total time are given in parantheses

	Computation Time (seconds)	
Grid Resolution[1]	Step 1 + Step 2	Step 3 + Step 4
3271 pts	11.9 (18.8%)	51.3 (81.2%)
10201 pts	30.4 (8.4%)	331.5 (91.6%)
40401 pts	121.9 (2.1%)	5573.7 (97.9%)

[1]The number of grid points inside a unit square.

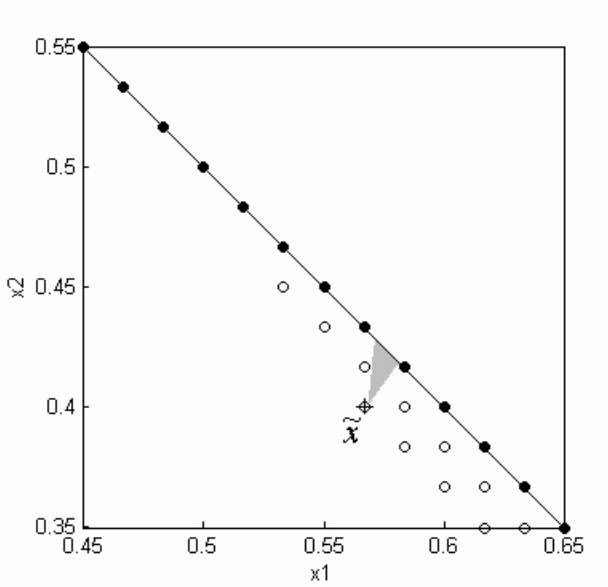

Figure 1: Inaccuracy due to inadequate grid resolution. The gray area $\mathcal{W}_{\tilde{x}}$ is the dominating region of $\tilde{\boldsymbol{x}}$. None of the grid points (both solid and hollow dots) reside inside of $\mathcal{W}_{\tilde{x}}$. Hence, $\tilde{\boldsymbol{x}}$ is falsely accepted as being inside the confidence region. Solid dots are truly part of the confidence region, while the hollow dots are not. Consequently, the resulting confidence region is bigger than it actually is.

Table 3: 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. PCD1 and PCD2 refer to the PCD's two-step derivative and derivative-free approaches, respectively. The plus sign in PCD2+ refers to the addition of BH Step and the fact that higher grid resolution was used. Sec.3 refers to our proposed approach described in section 3. The average coverage rates across different locus of optimal points are given in the last column AVG.

		Locus of Optimal Points										
Eigenvalues	Methods	0.0	0.7	0.9	1.0	1.1	1.3	1.6	2.0	2.5	3.0	AVG
(0.2 0.2)	PCD2	94.3	94.9	93.1	95.5	95.9	95.7	97.6	96.4	98.3	98.4	96.01
	Sec.3	93.8	94.5	92.8	95.0	95.6	95.4	97.2	96.0	97.7	97.4	95.54
	PCD1	94.8	95.3	93.1	95.2	96.1	95.9	97.6	96.1	97.7	97.5	95.93
	PCD2+	93.8	94.5	92.8	95.1	95.6	95.3	97.0	96.1	97.8	97.3	95.53
(0.2 7.0)	PCD2	97.8	96.1	97.6	97.7	98.1	99.2	99.9	100.0	100.0	100.0	98.64
	Sec.3	95.5	93.4	95.9	96.7	97.2	98.2	98.9	98.7	98.5	98.8	97.18
	PCD1	95.8	93.6	95.9	96.8	97.3	98.3	99.0	99.0	98.6	99.0	97.33
	PCD2+	95.2	93.4	95.9	96.8	97.4	98.4	98.9	98.9	98.6	99.0	97.25
(2.0 2.0)	PCD2	97.2	98.0	97.2	98.0	98.6	99.4	99.6	100.0	99.7	99.9	98.76
	Sec.3	94.0	95.8	95.8	97.0	96.9	97.6	98.1	99.4	99.0	98.7	97.23
	PCD1	94.0	95.7	95.8	96.9	96.9	97.6	98.2	99.4	99.1	98.8	97.24
	PCD2+	94.0	95.7	95.8	97.2	97.5	97.9	98.2	99.4	99.1	98.8	97.36
(2.0 7.0)	PCD2	98.3	97.8	97.3	98.9	99.1	98.7	99.9	100.0	100.0	100.0	99.00
	Sec.3	93.9	95.2	95.2	97.3	97.6	97.8	98.9	98.4	98.8	99.1	97.22
	PCD1	93.9	95.2	95.2	97.3	97.5	97.8	98.9	98.4	98.8	99.1	97.21
	PCD2+	93.9	95.2	95.2	97.1	97.8	97.9	98.9	98.6	98.8	99.0	97.24
(13.0 15.0)	PCD2	100.0	99.3	99.0	99.2	100.0	100.0	100.0	100.0	100.0	100.0	99.75
	Sec.3	95.5	95.6	94.5	97.5	98.7	98.2	98.5	98.8	98.8	99.1	97.52
	PCD1	95.1	95.6	94.5	97.3	98.7	98.1	98.4	98.9	99.0	99.1	97.47
	PCD2+	95.1	95.6	94.5	97.6	99.0	98.5	98.8	99.1	99.1	99.7	97.70

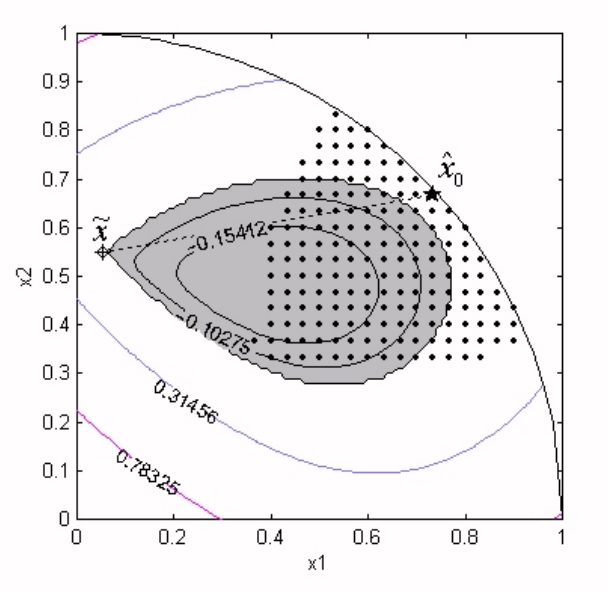

Figure 2: The direction from $\tilde{\boldsymbol{x}}$ to $\widehat{\boldsymbol{x_0}}$ (denoted by star symbol) is used as an initial search direction. Solid dots are the confidence region. $\tilde{\boldsymbol{x}}$ is the current point of interest. The gray area is the dominating region of $\tilde{\boldsymbol{x}}$. Note that the search direction passes through the dominating region $\mathcal{W}_{\tilde{x}}$. Thus, any point in the neighborhood of $\tilde{\boldsymbol{x}}$ along this search direction can be selected for discarding $\tilde{\boldsymbol{x}}$ from being inside the confidence region.

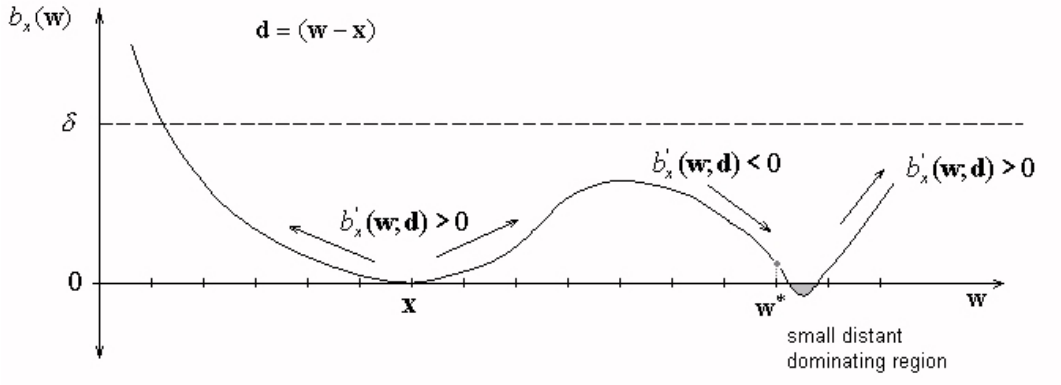

Figure 3: Locating $\boldsymbol{w}^*$, the starting point for the Nelder-Mead search. Note that points $\boldsymbol{w}$ in the vicinity of $\boldsymbol{x}$ are likely to have small $b_x(\boldsymbol{w})$ values. However, by searching only points $\boldsymbol{w}$ that have negative $b'_x(\boldsymbol{w};\boldsymbol{d})$, no points in the vicinity of $\boldsymbol{x}$ will get chosen. To improve computational efficiency, an upperbound $\delta > 0$ can be chosen such that the search space is limited to the set $\{\boldsymbol{w} \in \mathcal{R} \backslash \mathcal{E} : b_x(\boldsymbol{w}) < \delta\}$.

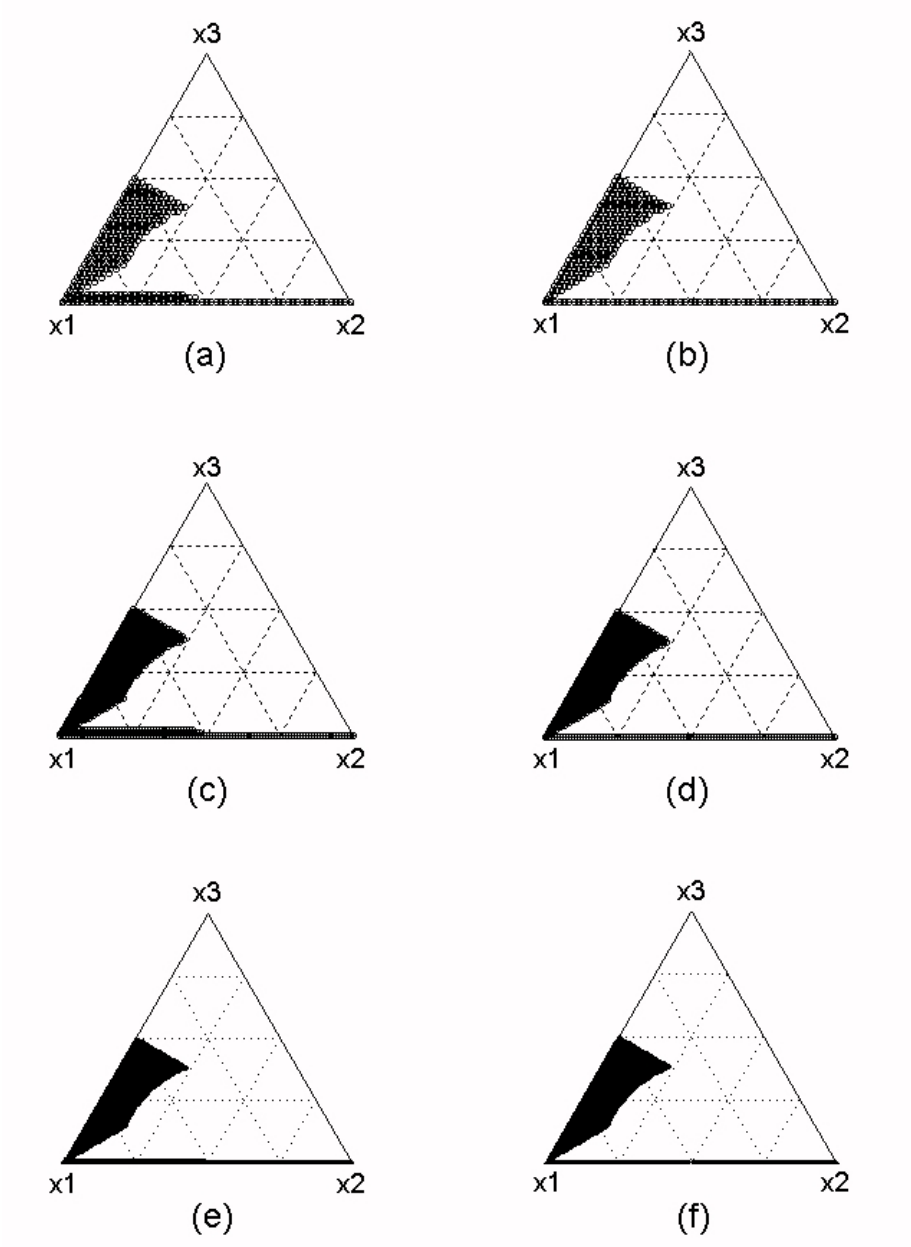

Figure 4: 95% Confidence Regions for Frisbee and McGinity (1994) example. The confidence regions were obtained using (a) PCD approach (3271); (b) Proposed approach (3271); (c) PCD approach (10201); (d) Proposed approach (10201); (e) PCD approach (40401); (f) Proposed approach (40401). The numbers in the parentheses are the total number of grid points inside a unit square. Note that the resulting confidence regions of the PCD derivative-free approach (a,c,e) are inaccurate (bigger than they actually are). This inaccuracy is decreased as the grid resolution is increased.

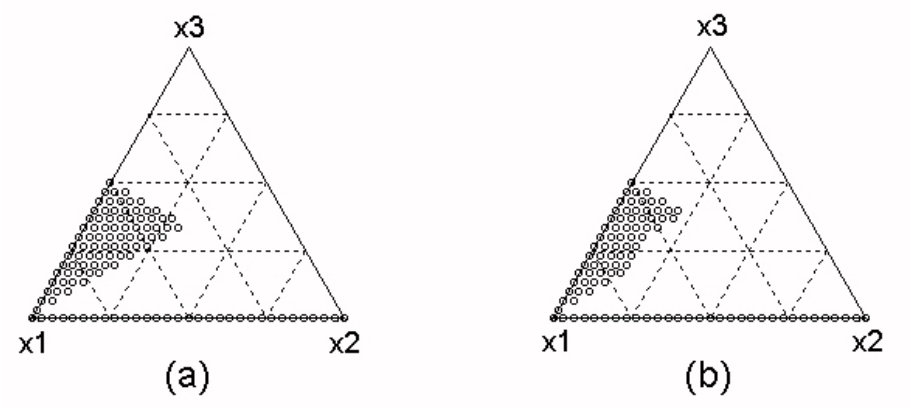

Figure 5: 95% Confidence Regions for the Frisbee and McGinity (1994) example computed using: (a) only the first two steps; (b) the full four steps of the proposed approach; both with a suitably low grid resolution. Note that the confidence regions are different, indicating that all four steps are needed.

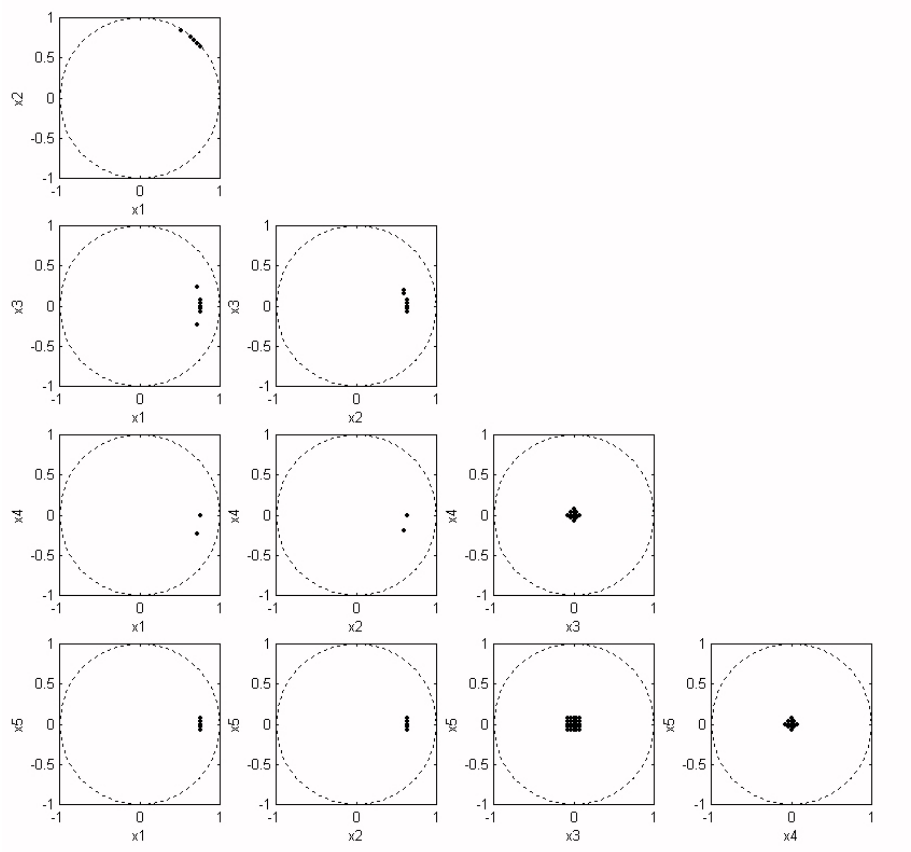

Figure 6: Unconditional 95% Confidence Regions for the Richert et al. (1974) example. The estimated optimal point is $\widehat{\boldsymbol{x_0}} = \{0.73, 0.69, -0.03, -0.02, -0.05\}$.

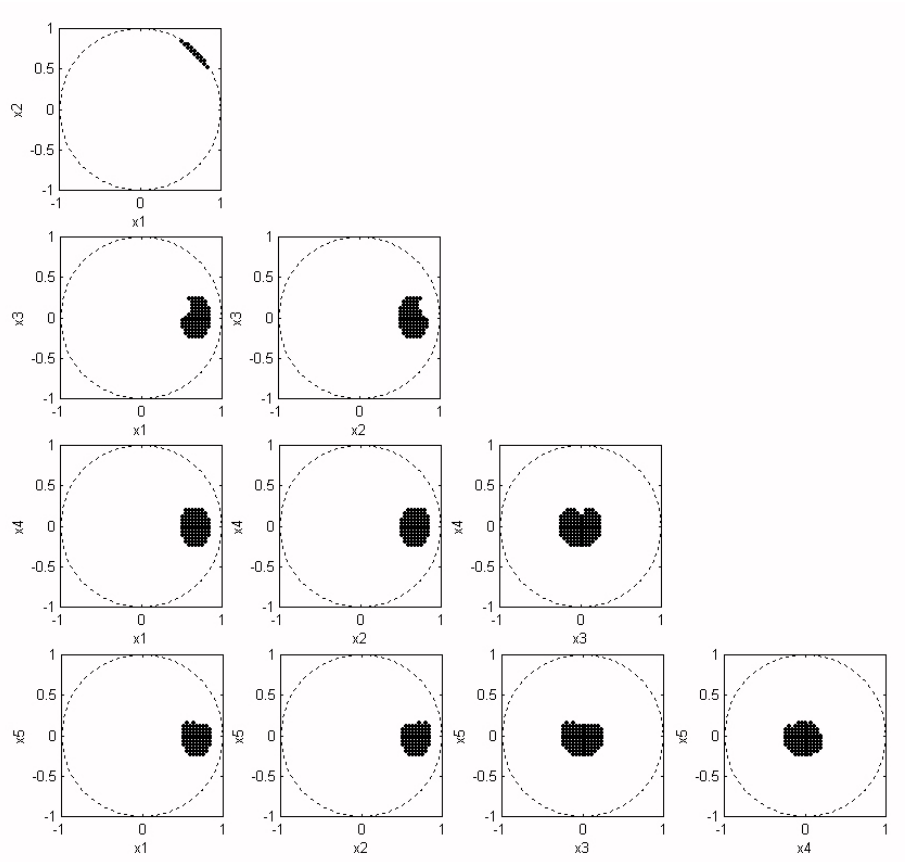

Figure 7: Conditional 95% Confidence Regions for the Richert et al. (1974) example. Factors not shown on each plane are at their optimal factor levels.