CONFIDENCE REGIONS FOR OPTIMAL CONTROLLABLE VARIABLES FOR THE ROBUST PARAMETER DESIGN PROBLEM

A Dissertation
Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by
Aili Cheng
January 2013

Examining Committee Members:

Dr. Pallavi Chitturi, Advisory Chair, Statistics
Dr. Damaraju Raghavarao, Statistics
Dr. Marc Sobel, Statistics
Dr. John Peterson, GlaxoSmithKline plc.
Dr. Krupa S. Viswanathan, Risk, Insurance, and Healthcare Management (outside reader)

ABSTRACT

CONFIDENCE REGIONS FOR OPTIMAL CONTROLLABLE VARIABLES FOR THE ROBUST PARAMETER DESIGN PROBLEM

Aili Cheng

DOCTOR OF PHILOSOPHY

Temple University, January 2013

Associate Professor Pallavi Chitturi, Chair

In robust parameter design it is often possible to set the levels of the controllable factors to produce a zero gradient for the transmission of variability from the noise variables. If the number of control variables is greater than the number of noise variables, a continuum of zero-gradient solutions exists. This situation is useful as it provides the experimenter with multiple conditions under which to configure a zero gradient for noise variable transmission. However, this situation requires a confidence region for the multiple-solution factor levels that provides proper simultaneous coverage. This requirement has not been previously recognized in the literature. In the case where the number of control variables is greater than the number of noise variables, we show how to construct critical values needed to maintain the simultaneous coverage rate. Two examples are provided as a demonstration of the practical need to adjust

the critical values for simultaneous coverage.

The zero-gradient confidence region only focuses on the variance, and there are in fact many such situations in which focus is or could be placed entirely on the process variance. In the situation where both mean and variance need to be considered, a general confidence region in control variables is developed by minimizing weighted mean square error. This general method is applicable to many situations including mixture experiments which have an inherit constraint on the control factors. It also gives the user the flexibility to put different weights on the mean and variance parts for simultaneous optimization. It turns out that the same computational algorithm can be used to compute the dual confidence region in both control factors and the response variable.

ACKNOWLEDGEMENTS

I owe my deepest gratitude to Dr. John Peterson. The research would not have been possible without his constant guidance, encouragement, and support. I am also deeply grateful to my advisor Dr. Pallavi Chitturi for her enthusiastic encouragement, advice, and patience. Many of her weekends have been sacrificed for my dissertation. I would like to specially thank Prof. Damaraju Raghavarao for his encouragement and timely guidance on the dissertation requirements.

I would like to express my very great appreciation to Dr. Marc Sobel. His kindness and support helped me passed the difficult theory courses. His advice and suggestions on my dissertation work has been very helpful. I am also indebted to Dr. Krupa S. Viswanathan. Her willingness to give her time and advice so generously has been very much appreciated.

I would like to extend my special thanks to many of my previous and current colleagues: Paul McAllister, Fang Liu, Kai-fen Wang, Ping Li, Jing Chen, Kristie Michaud, Huizi Zhang, Vijayalakshmi Chandrasekaran, and Jie Gao. I would have quitted many times without their encouragement and support. Lastly I would like to thank my family for their support and understanding for years since I joined the program.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Robust Parameter Design (RPD) is also called Robust Design or Parameter Design in the literature (Nair 1992, Robinson, Borror, and Myers 2004). The concept of RPD was introduced in the United States by Genichi Taguchi in the early 1980's. It is a methodology that takes both the mean and the variance into consideration for product or process optimization. Taguchi (1986) divided the variables that affect the performance of a process or system into two categories: control variables and noise variables. Control variables can be held at specific target levels, whereas noise variables are either difficult to control or uncontrollable at a large scale; although they may be set to specific levels at the lab scale. Consider the example of baking a cake from a cake mix. The proportions of the ingredients can be controlled to the desired level by the manufacturer. However, the actual oven temperature could vary from the

prescribed temperature level depending on different types or brands of ovens used by consumers. Therefore, the ingredient amounts can be treated as control variables, but oven temperature should be treated as a noise variable for baking cake optimization. The objective of RPD is to optimize the control variable settings such that the response is least affected by the noise variables. A lot of work has been done to address this issue from different perspectives (see Section 1.1 for a more detailed literature review). One practical solution to this optimization problem is to establish the confidence interval or region of the control variable settings where the variance transmitted to the response by the noise variables is minimized. Myers, Kim, and Griffiths (1997) developed such a confidence region, but it left room for further improvement. In Section 1.2 we will point out the limitations of their approach and explain why further improvement is necessary. Several approaches are proposed in Chapter 2 for the improvement, followed by examples. More work has been done to further generalize the confidence region to optimize both mean and variance, which is shown in Chapter 3. The possibility of applying the same computational algorithm to the joint confidence region in the control factors and response is discussed in Chapter 4.

1.1 Literature Review

Genichi Taguchi is a major contributor to RPD research. He introduced the philosophy of considering the mean and the variance simultaneously, the design to study the effects of control and noise factors, and the Signal-to-Noise Ratio (SNR) as an objective function in analysis. The philosophy of optimizing the mean and the variance is well accepted. However, Taguchi's design and analysis approach generated a lot of interest as well as criticisms. Taguchi proposed using crossed arrays to study the noise-control interaction. Crossed arrays consist of two orthogonal arrays: one for the control factors and the other for the noise factors. The orthogonal array of control factors is called the inner array, which is crossed with the orthogonal array of noise factors called the outer array. (see Figure 1.1 for a crossed array design with two control variables and two noise variables.) One advantage of the crossed array design is that it allows the consideration of all the noise by control variable interactions. However, when the number of control factors and noise factors increases, the number of experiments can increase very quickly. To make the design more efficient, both inner array and outer array are often fractional factorial designs with low resolutions. The consequence of the low resolution design is that very few control-noise variable interactions can be considered. In the past two decades, various designs other than crossed arrays have been proposed to achieve the same goal while reducing the number of

runs. One of them is called the combined array. Welch, Wu, Kang and Sacks (1990), Shoemaker, Tsui and Wu (1991) and Montgomery (1991) proposed the use of combined arrays. A combined array combines control variables and noise variables in a single array. It can reduce the total number of runs while still achieving the same or even higher resolution than the crossed array design. Designs like fractional factorial design, central composite design, Box-Behnken design, and mixed resolution designs (Borror and Montgomery 2002) are common examples of the combined array design. Combined arrays provide the flexibility to consider interactions and even quadratic terms of the control factors if they are of interest. Combined array designs have been the focus of the recent RPD related literature. Borror and Montgomery (2002) have developed a procedure to assess different combined array designs based on the prediction error variance, which provides a useful way to select the appropriate combined array design. Other designs like split-plot designs have also been explored in the RPD field. See the paper by Robinson, Borror, and Myers (2004) for a more detailed review on the research on RPD designs. At the same time, Taguchi's analysis approach of using signal-to-noise ratio (SNR) was criticized and some new analysis methods have been developed and shown superior to Taguchi's method. See Nair's panel discussion (1992) and Robinson et al. (2004) for comprehensive reviews of the criticisms and the research on alternative design and analysis methods up to 2003. The rest of the review

section will focus largely on the more recent work on multivariate RPD that was not covered by Nair (1992) and Robinson et al.(2004).

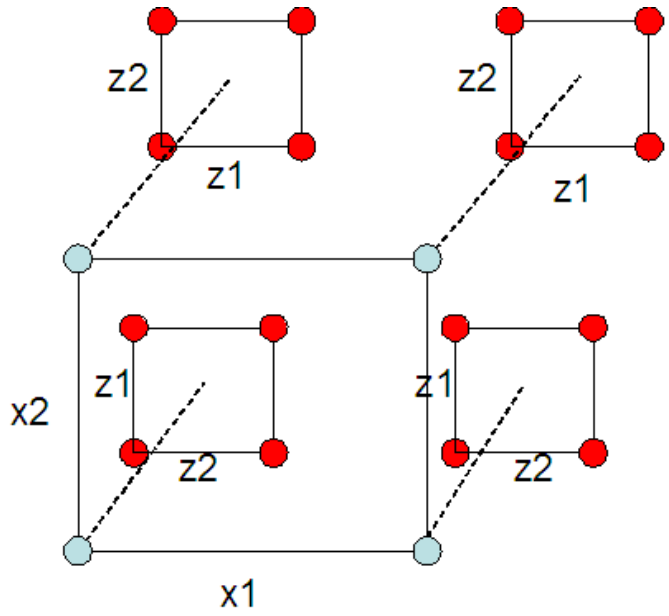

Figure 1.1: The crossed array design with two control variables ($X1$ and $X2$), and two noise variables ($Z1$ and $Z2$)

So far, very limited work has been done on multivariate RPD problems. Chiao and Hamada (2001) assumed that the multiple responses follow a normal distribution where the mean vector and the variance-covariance matrix are functions of the control variables and noise variables. They proposed a simple methodology to find the optimized setting of the control variables that maximizes the probability that all the responses are in some pre-specified ranges simultaneously. However, this approach relies on crossed array designs to model the mean vector and the variance-covariance matrix. In 1990, Vin-

Table 1.1: The comparison between the univariate and multivariate approach where x is the vector of control variables, a, b, $l_{q\times 1}$, $u_{q\times 1}$ are the corresponding target ranges

Univariate Case	Multivariate Case
Find x to minimize V_Y subject to $a \leq E(Y) \leq b$	Find x to minimize $f(V_Y)$ subject to $l_{q\times 1} \leq E(Y) \leq u_{q\times 1}$

ing and Myers applied the dual response approach (Myers and Carter, 1973) to analyze RPD data where the two responses are the mean and variance of a single response variable. Miro-Quesada, del Castillo and Peterson (2004) extended this approach to multiple response variables. Corresponding to the goal of minimizing the variance of a response vector Y, V_Y, Miro-Quesada et al. (2004) proposed to minimize a scalar function of the multivariate variance-covariance matrix subject to the mean vector being close to target. The scalar function can be either the determinant or the trace of the matrix. This approach considers the correlation of the multiple responses. However, the scalar function is hard to understand intuitively. Let $E(\bullet)$ be the expectation operator. X is the experimental region of the control variables. $f(\bullet)$ is a scalar function. The comparison between the univariate and multivariate approach in terms of the final objectives can be summarized as shown in Table 1.1.

Miro-Quesada et al. (2004) expanded the Multivariate Bayesian predictive approach developed by Peterson (2000, 2002) to RPD where they calculated the posterior probability that all the responses simultaneously fall within the specification region at any given condition, determined by control variables in

the presence of noise variables.

1.2 Myers' Confidence Region

Most of the RPD literature has been focused on finding the point estimate of the control variable setting where 1) variance caused by the change of noise variables is minimized and 2) the mean response is close to target. However, a more practical approach is to find a range of control variables to reach these two goals. A two-step approach to reach these goals simultaneously was proposed by Myers, et al. (1997). The first step is to find a confidence region in control variables where the variance due to noise variables is zero. The second step is to find within this confidence region the control variable setting where the mean response is closest to the target. Myers, et al. (1997) constructed a confidence region in control variables based upon the standard response surface model (1.1) for incorporating noise variables (Box and Jones, 1990, Lucas, 1994, Myers, Khuri, and Vining, 1992).

$$y(x, z) = \beta_0 + x'\beta + x'Bx + z'\gamma + x'\Delta z + \varepsilon \tag{1.1}$$

where x is a vector of k control variables, z is a vector of h noise variables, β_0 is the intercept, β is a $k \times 1$ vector of coefficients for the main effects of control variables, γ is a $h \times 1$ vector of coefficients for the main effects of the

noise variables, B is a $k \times k$ matrix whose diagonals are the coefficients for the second order of control variables and whose off-diagonals are one-half of the control variable interaction effects, and Δ is a $k \times h$ matrix of control by noise variable interaction effects. ε is the random error term. It is assumed that $\varepsilon \sim N\,(0,\ \sigma^2)$.

Assuming that noise variables have mean zero and variance-covariance matrix V_z, the variance of the response in (1.1) is $\sigma^2[y(x,z)] = (\gamma + \Delta' x)' V_z (\gamma + \Delta' x) + \sigma^2$. Myers, et al. (1997) then proposed that "minimum process variance" can be reached by setting the slope of the noise variables, $\gamma + \Delta' x$, equal to zero. A confidence region for such control variable values can be constructed by inverting a hypothesis test of the form

$$H_0 : \gamma + \Delta' x = 0 \tag{1.2}$$

To simplify the notation, let ψ be the $p \times 1$ vector that contains all the elements of γ and Δ, i.e. $\psi = (\gamma_1, \delta_{11}, \delta_{12}, \cdots, \delta_{1k}, \gamma_2, \delta_{21}, \delta_{22}, \cdots, \delta_{2k}, \cdots,$ $\gamma_h, \delta_{h1}, \delta_{h2}, \cdots, \delta_{hk})'$, where $\gamma_i, i = 1, 2, 3, \cdots, h$, is the ith element of vector γ and $\delta_{ij}, i = 1, 2, 3, \cdots, h, j = 1, 2, 3, \cdots, k$ is the element of Δ' in the ith row and jth column. Let $M(x) = I_h \otimes (1, x')$, where I_h is an identity matrix of dimension h, x is a $k \times 1$ vector of the corresponding control variables that interact with noise variables, and $\otimes$ denotes the Kronecker product, then the

null hypothesis can be written as:

$$H_0 : M(x)\psi = 0 \tag{1.3}$$

where $M(x)$ is a $h \times p$ matrix and $M(x)\psi = \gamma + \Delta' x$. Now the $100(1-\alpha)\%$ confidence region in control variables for zero variance due to noise variables can be defined as:

$$\{x : M(x)\psi = 0 \text{ is not rejected at level } \alpha\} \tag{1.4}$$

Furthermore, let $Q(x; \hat{\psi})$ denote the test statistic for H_0 which is

$$(M(x)\hat{\psi})'[M(x)\hat{V}_{\hat{\psi}}M(x)']^{-1}(M(x)\hat{\psi}) \tag{1.5}$$

where $\hat{V}_{\hat{\psi}}$ is the usual unbiased estimator of the variance of $\hat{\psi}$, $V_{\hat{\psi}}$.

Let $\hat{\Psi} \sim N(\psi, V_{\hat{\psi}})$. Myers, et al. (1997) have shown that $Q(x; \hat{\Psi}) \sim hF(h, v)$, where $\hat{\Psi}$ is a vector of random variables (i.e. $\hat{\Psi}$ is an estimator of ψ, whereas $\hat{\psi}$ is an estimate from the actual data), v is the residual degrees of freedom (df) and $F(h, v)$ is F distribution with numerator df equal to h and denominator df equal to v. They then concluded that the $100(1-\alpha)\%$ confidence region in (1.4) is:

$$C_x(c_\alpha) = \left\{ x : Q(x; \hat{\psi}) \leq c_\alpha \right\} \tag{1.6}$$

where the critical value $c_\alpha = hF(1-\alpha, h, v)$.

Note that the confidence region in (1.6) (called the "MKG confidence region" from now on) and the critical value $c_\alpha = hF(1-\alpha, h, v)$ were derived based on two critical assumptions: 1. the minimum of the variance due to noise variables is zero; 2. the solution to the zero-gradient equation in (1.3) is unique. The first assumption is not completely correct in that the minimum noise variance could be a positive number instead of zero. Even if the minimum noise variance is zero, the second assumption is only true when $h \geq k$. In other words, the MKG confidence region provides the correct critical value for the zero-gradient solution (if it exists) only when there are at least as many noise variables as control variables. However, when the number of noise variables is less than the number of control variables, multiple solutions could exist to the zero-gradient equation in (1.3). As such, a confidence region which covers all the solutions simultaneously needs to be developed. Statistical inference for the multiple solution problems is important as this gives the experimenter more options with regard to finding the zero-gradient factor settings. The objective of the next chapter is to generalize the MKG confidence region such that it will provide the right coverage for both the single solution case (where $h \geq k$) and multiple solution case (where $h < k$).

A more general confidence region in control variables to optimize both mean and noise variances is developed in Chapter 3. The possibility of applying the idea to develop a joint confidence region in both control variables x and

response y is discussed in Chapter 4.

CHAPTER 2

A CONFIDENCE REGION FOR ZERO-GRADIENT SOLUTIONS FOR ROBUST PARAMETER DESIGN EXPERIMENTS

Chapter 2 is organized as follows: Section 2.1 will focus on the derivation of a generalized confidence region that covers all the solutions simultaneously and the corresponding critical values to ensure the right coverage rate. We will show in Section 2.2 how to compute the confidence region once the critical

values are known. In Section 2.3 we give two examples to demonstrate the difference in simultaneous coverage between the MKG confidence region and our proposed confidence region when the number of control factors exceeds the number of noise variables. Section 2.4 provides a summary of the results. The content of this chapter was published in International Journal of Quality, Statistics, and Reliability in 2011.

2.1 A Generalized Confidence Region Approach

2.1.1 The Multiple Zero-Gradient Solution Problem

To address the multiple solution situation, the hypothesis in (1.3) is generalized to $H_0 : M(x)\psi = 0, \forall x \in \mathcal{L}$, where $\mathcal{L}$ is the linear subspace representing either a unique single solution (i.e. a point) or multiple solutions (i.e. a line, or a hyperplane). In other words, the confidence region could be a collection of either points or linear subspaces (of dimension ≥ 1) depending on whether the solution to the equation $M(x)\psi = 0$ is unique or not. Therefore, we propose to generalize the MKG confidence region to:

$$C_{\mathcal{L}}(c_\alpha) = \{\mathcal{L} : Q(x; \hat{\psi}) \leq c_\alpha, \forall x \in \mathcal{L}\} \tag{2.1}$$

where $\mathcal{L}$ represents the linear subspace of the space defined by the elements of x, which are the solutions to $\gamma + \Delta' x = 0$. Here x is a $k \times 1$ vector. $\mathcal{L}$ has dimension d, where d is defined as:

$$d = \left\{ \begin{array}{ll} k-h & \text{if } h < k \\ \\ 0 & \text{otherwise} \end{array} \right\} \tag{2.2}$$

Therefore, when the solution to $\gamma + \Delta' x = 0$ is a single point, $d = 0$; otherwise, $d > 0$. We call the confidence region given in (2.1) the generalized zero gradient (GZG) confidence region. As indicated by the definition, the MKG confidence region is a special case of the GZG confidence region where $\mathcal{L}$ is a point and $d = 0$. We know that the MKG confidence region is correct and the critical value is $c_\alpha = hF(1-\alpha, h, v)$ for $d = 0$ or $h \geq k$, if a solution exists. (If $h \geq k$ and the MKG region is the null set, then there is statistically significant evidence that a solution does not exist.) The next question is: what value should c_α take when $d > 0$? For $d > 0$, note that a $100(1-\alpha)\%$ GZG confidence region should contain the zero gradient solution set $\mathcal{L}$, with probability $1-\alpha$ before the experiment is performed.

It is worth pointing out that when $h < k$, the GZG confidence region in (2.1) is a simultaneous confidence region problem in that a line or hyperplane will be included in the confidence region only if all the points on the line or hyperplane satisfy the criteria. Therefore the GZG confidence region in (2.1) can also be expressed as:

$$\{\mathcal{L} : Q_{\mathcal{L}} \leq c_\alpha\} \tag{2.3}$$

where $Q_{\mathcal{L}} = \max_{x \in \mathcal{L}} Q\left(x; \hat{\psi}\right)$ and $\mathcal{L} = \{x : M(x)\psi = 0\}$. To find the critical

value c_α for $h < k$, we need to first understand the distribution of the test statistic $Q_{\mathcal{L}}$ when $H_0 : M(x)\psi = 0$ true.

We will explore the distribution of $Q_{\mathcal{L}}$ in special cases first and then the general case.

2.1.2 $h = 1, h < k$

If $h = 1$, $M(x) = (1, x')$ (assuming the intercept case is the default, i.e. the main effect of a noise variable has to exist if the noise variable interacts with at least one control variable), which is a $1 \times p$ vector. Based on Theorem 2.1 below, the critical value should be $(d+1)F(1-\alpha, (d+1), v)$, where d is the dimension of the solution set for the equation $M(x)\psi = 0$ and is defined in equation (2.2). Here, v is the degrees of freedom of the residuals.

Theorem 2.1. *When $h = 1, h < k$, $Q_{\mathcal{L}} \sim (d+1)F((d+1), v)$, where v is the degrees of freedom of the residuals.*

Theorem 2.1 is a direct result from Miller's Theorem(Miller 1981, p65, p113). Detailed proof is also available in Appendix A.1.

Note that when $h = k = 1$, $d = 0$. It means that the solution is a point, which is a linear space of dimension zero. The critical value $(d+1)F(1-\alpha, (d+1), v)$ then becomes into MKG critical value $F(1-\alpha, 1, v)$ because $d = 0$.

For the $k > h > 1$ case, the distribution of $Q_{\mathcal{L}}$ is more complex. Section 2.1.3 addresses the full interaction model case in (1.1) where the experimental

design is completely orthogonal, or partially orthogonal so that $V_{\hat{\psi}} = c\sigma^2 I$ for some positive constant, c, residual variance σ^2 and an identity matrix I, of dimension of p. Here an exact simultaneous confidence region is derived. For the general cases, Section 2.1.4 proposes a simulation method based upon the multivariate t-distribution to find approximate critical values with which to construct the confidence region.

2.1.3 $k > h > 1$ and Full Interaction Model Orthogonal Design Case

Here, we assume that the data are generated from an orthogonal design or partially orthogonal design such that $V_{\hat{\psi}} = c\sigma^2 I$ for $c > 0$. Furthermore, it is assumed that we have a full noise-control variable interaction model (henceforth referred to as "full model"), meaning that each noise variable interacts with the same set of control variables. i.e. each element of the interaction matrix Δ is nonzero. Corollary 2.1 below shows that the critical value for the $k > h > 1$ case can be obtained via a Monte Carlo simulation of a chi-square variable and an independent Wishart random matrix.

Theorem 2.2. *Let the variance-covariance matrix $V_{\hat{\psi}} = \Omega\sigma^2$, where Ω is a $p \times p$ matrix. Assume that each element of the interaction matrix Δ is nonzero. And further assume that $\Omega = V_1 \otimes V_2$ where V_1 is a $h \times h$ variance-covariance matrix and V_2 is $(k+1) \times (k+1)$ variance-covariance matrix. For $h < k$ and*

$h > 1$, the test statistic $Q_{\mathcal{L}}$ has the same distribution as a function of a chi-square random variable and a random Wishart matrix (which are stochastically independent) as shown below:

$$Q_{\mathcal{L}} \sim \frac{\lambda_{max}(A)}{U/v}$$

where $U \sim \chi^2(v)$ and $A \sim Wishart_{d+1}(I_{d+1}, h)$. $Wishart_{d+1}(I_{d+1}, h)$ represents a Wishart distribution with degrees of freedom equal to h and scale matrix equal to a $(d+1) \times (d+1)$ identity matrix, I_{d+1}. v is the residual degrees of freedom. d is same as defined before. $\lambda_{max}(\bullet)$ is the maximum eigenvalue of the matrix.

See proof in Appendix A.2.

Remark: Theorem 2.2 assumes that the main effects of noise variables γ always exist. The result is also valid when $\gamma = 0$ (see the proof in Appendix A.2).However, the dimesion of V_2 will be reduced to $k \times k$ when $\gamma = 0$.

Corollary 2.1. *Theorem 2.2 will hold when the design is orthogonal or partially orthogonal such that $V_{\hat{\psi}} = c\sigma^2 I$ for $c > 0$.*

See proof in Appendix A.2.

Based on Corollary 2.1, the test statistic $Q_{\mathcal{L}}$ is a function of a random χ^2 variable and the maximum eigenvalue of a random Wishart matrix whose distribution only depends on the three parameters: h, d, and degrees freedom of the error v. In other words, the distribution of $Q_{\mathcal{L}}$ only depends on the

degrees of freedom of the error and the number of noise and control variables because d is a function of the number of noise and control variables. The critical value c_α is then the $100(1-\alpha)^{th}$ percentile of the distribution of $Q_{\mathcal{L}}$, which can be obtained by simple Monte Carlo simulation from χ^2 and Wishart distributions. See Appendix B.1 for the tabulated critical values based on Corollary 2.1.

Note that the critical value determined by Corollary 2.1 becomes the MKG critical value, $hF(1-\alpha, h, v)$, when $d = 0$, i.e. $h = k$. This is true because when $d = 0$, A follows $Wishart(1, 1, h)$, which is $\chi^2(h)$. In other words, A is NOT a matrix anymore. Instead, it becomes a $\chi^2(h)$ variable. Hence $\lambda_{max}(A) = A$. $Q_{\mathcal{L}}$ can then be written as:

$$Q_{\mathcal{L}} = \frac{A}{U/v} = h\frac{A/h}{U/v} \sim hF(h, v)$$

2.1.4 Two general computation approaches

Although orthogonality and full model are ideal properties to have, sometimes it is not practical to implement an orthogonal design. And it is not guaranteed that the noise variables always interact with the same set of control variables. In some cases, the experimental design may be such that $V_{\hat{\psi}}$ does not have the orthogonal $c\sigma^2 I$ form or we may wish to use a model with some 'control $\times$ noise variable' interaction terms deleted, i.e.,the interaction matrix Δ has some elements equal to zero. In such situations, when $k > h > 1$,

the distribution of $Q_{\mathcal{L}}$ does not have a simple form, and may depend upon ψ even under H_0. Nonetheless, in such situations it is still possible to obtain approximately conservative simultaneous confidence regions for control variables associated with zero-gradient solutions. We provide such a construction as follows. Recall that $Q_{\mathcal{L}}$ is a function of ψ and consider

$$Q_{max} = \max_{\psi \in C_\psi(b_\alpha)} Q_{\mathcal{L}} = \max_{\psi \in C_\psi(b_\alpha)} \max_{M(x)\psi = 0} Q(x; \hat{\Psi}) \tag{2.4}$$

where $C_\psi(b_\alpha) = \{\psi : (\psi - \hat{\psi})' \hat{V}_{\hat{\psi}} (\psi - \hat{\psi}) \leq b_\alpha\}$ and $b_\alpha = F\,(1-\alpha, 1, n-p)$. Note that by using $b_\alpha = F\,(1-\alpha, 1, n-p)$, $Q_{\max}$ is an approximate upper confidence bound for the scalar-valued quantity, $Q_{\mathcal{L}}$ (See Clarke, 1987, for a discussion of confidence bounds on nonlinear functions of model parameters constructed from confidence regions.) Let c^*_α denote the $100(1-\alpha)$ percentile of the distribution of Q_{max} under H_0. Consider the confidence region defined by

$$C_x^{\max} = \left\{ x : \; Q(x; \hat{\psi}) \leq c^*_\alpha \right\} \tag{2.5}$$

This confidence region should provide (at least approximately) a conservative simultaneous confidence region for the zero-gradient solutions. However, computation of c^*_α (using (2.4)) and the associated confidence region is numerically difficult due to the complex constraints associated with the definition of Q_{max}.

Fortunately, it can be shown that

$$Q_{max} = \max_{x \in C_x(b_\alpha)} Q(x; \hat{\Psi}) \tag{2.6}$$

where $C_x(b_\alpha) = \{x : Q(x; \hat{\psi}) \leq b_\alpha\}$. A proof is given in Appendix A.3. The expression for Q_{max} in (2.6) allows for much easier computation of the c^*_α critical value. The actual construction of the GZG confidence region from the relevant critical value will be outlined in Section 2.2.

The critical value computation for the general case

Note that under $H_0 : M(x)\psi = 0$ we can express $Q(x; \hat{\Psi})$ as

$$(M(x)\mathbf{t})'(M(x)\Omega M(x)')^{-1}(M(x)\mathbf{t}) \tag{2.7}$$

where $\hat{V}_{\hat{\psi}} = s^2\Omega$, s^2 is the mean squared error, Ω is a known matrix computed from the design matrix, and $\mathbf{t} = (\hat{\Psi} - \psi)/s$. Here the vector $\mathbf{t}$ follows the multivariate t distribution with location parameter equal to zero, scale matrix Ω, and degrees of freedom v. Using equation (2.7) we can then compute the critical value, c^*_α, using Monte Carlo simulations as follows.

Step 0. Compute $C_x(b_\alpha) = \{x : Q(x; \hat{\psi}) \leq b_\alpha\}$ where $b_\alpha = F(1-\alpha, 1, n-p)$.

Step 1. Simulate a multivariate t random vector (rv) with scale matrix Ω and v df. (This can be done by simulation of a multivariate normal random variable with mean vector 0 and variance-covariance matrix Ω and a chi-square random variable with v df. See Kotz and Johnson, 1982, for details.)

Step 2. Compute Q_{max} using the expressions in (2.6) and (2.7). (Computation of Q_{max} can be done by maximization of $Q(x;\hat{\Psi})$ over $C_x(b_\alpha)\cap R$ instead, where R is a pre-specified bounded region. This will calibrate the coverage to be simultaneous only over $\mathcal{L}_0 \cap R$, where $\mathcal{L}_0$ is the true linear subspace such that $M(x)\psi = 0$.)

Step 3. Do Steps 1-2 a large number of times to estimate the $100(1-\alpha)^{th}$ percentile of the Monte Carlo distribution of Q_{max} . This $100(1-\alpha)^{th}$ percentile is then a Monte Carlo estimate of c^*_α.

The coverage rate of the critical value

In order to check the accuracy of the above critical values c^*_α, we have done Monte Carlo simulations of the above four step procedure using three different noise variable models in conjunction with both orthogonal and non-orthogonal designs. The statistical models used are summarized in Table 2.1 below. Three partially orthogonal face-centered central composite experimental designs were assessed, with associated statistical models 1, 2, and 3, respectively. These designs employed a coded factor space with factor levels equal to ± 1 (except for the center points). The axial points in noise variables are deleted to maintain partial orthogonality. The factorial part of the designs is either full factorial (e.g. model 1 and model 2) or half factorial (e.g. model 3). The non-orthogonal designs are constructed by changing the factorial point (comprised of all -1's)

Table 2.1: The Reduced Interaction Models for Simulation

Model no.	k	h	Model Form
1	3	2	$y = 1 + x_1 + x_2 + x_3 + x_1^2 + x_2^2 + x_1 \times x_2 + z_1 + x_1 \times z_1 + x_2 \times z_1 + z_2 - x_3 \times z_2 + e$
2	4	2	$y = 1 + x_1 + x_2 + x_3 + x_4 + x_1^2 + x_2^2 + x_2 \times x_3 + z_1 + x_1 \times z_1 + x_2 \times z_1 + z_2 - x_3 \times z_2 + x_4 \times z_2 + e$
3	4	3	$y = 1 + x_1 + x_2 + x_3 + x_4 + x_1^2 + x_2^2 + x_2 \times x_3 + z_1 + x_1 \times z_1 + x_2 \times z_1 + z_2 + x_3 \times z_2 + z_3 + x_4 \times z_3 + e$

from $(-1, -1, \cdots, -1, -1)$ to $(-1, -1, \cdots, -1, 0)$. For demonstration purposes, we simply chose model parameters to be either 1 or -1, with error variance equal to 1.

The results of these coverage rate simulations are summarized in Table 2.2 below. The coverage rates were computed as follows. The models in Table 2.1 were used to compute the true $\mathcal{L}$ space, $\mathcal{L}_0$, and to simulate associated data sets. For each of the three models the region, R, was a hyper-cube constructed from the Cartesian product of intervals of the form $[-10, 10]$. For each simulation, a check was done to see if the event

$$\max_{x \in \mathcal{L}_0 \cap E} Q(x; \hat{\Psi}) \leq c_\alpha^* \tag{2.8}$$

occurred, where E is the convex hull formed by the factor levels. In an attempt to reduce the conservatism of the above approach for computing c_α^*, we also considered the approximate approach obtained by maximizing $Q(x; \hat{\Psi})$ over $\mathcal{L}_0(\hat{\psi}) \cap R$ where $\mathcal{L}_0(\hat{\psi}) = \{x : M(x)\hat{\psi} = 0\} = C_x(0)$. We denote this

approximate critical value by $\tilde{c}_\alpha$.

Remark 1. Because $C_x(b_\alpha)$ or $\mathcal{L}_0(\hat{\psi})$ is a function of the data, the relatively large region, R, was chosen for these simulations so that $C_x(b_\alpha) \cap R$ or $\mathcal{L}_0(\hat{\psi}) \cap R$ would be very unlikely to be empty for any simulated data set. In addition, we did not want to rule out situations where the confidence region was outside of the experimental region. While in practice, such extrapolated inferences must be treated with caution, nonetheless it may be desired to compute such a confidence region. Such a confidence region outside of the experimental region suggests that it may not be possible to obtain a "zero gradient" solution for noise transmission, at least within the current experimental region. However, such a confidence region just outside of the experimental region may offer hope that resetting process control conditions may allow for a more robust process. Of course, additional experiments outside of the current experimental region would be needed to confirm this.

Remark 2. Maximization of $Q(x; \hat{\Psi})$ over $C_x(b_\alpha) \cap R$ to compute the Monte Carlo critical value, c^*_α , was accomplished by using the SAS/IML Nelder-Mead simplex algorithm, nlpnms. This was done to make the Monte Carlo simulations of this Monte Carlo procedure tractable. Some limited simulations were also done whereby the maximization of $Q(x; \hat{\Psi})$ over $C_x(b_\alpha) \cap R$ was computed by gridding instead. This was done to make sure that the Nelder-Mead algorithm did not stop its maximization prematurely. In all cases, each critical

Table 2.2: The coverage rates for the GZG confidence region using the approximate and conservative critical values, $\tilde{c}_\alpha$ and c^*_α, respectively. The coverage rates from $\tilde{c}_\alpha$ are in parentheses. The nominal coverage rate here is 95%. (Here k = no. of control factors, h =no. of noise variables, n = sample size, and v = residual df.)

Model No.	k	h	n	v	The coverage rate using c^*_α ($\tilde{c}_\alpha$)	
					Orthogonal	Non-orthogonal
1	3	2	40	28	96.9% (96.1%)	96.9% (96.1%)
2	4	2	74	60	98.0% (96.3%)	98.0% (96.3%)
3	4	3	72	57	97.2% (96.1%)	97.2% (96.1%)

value, c^*_α , computed using nlpnms was larger than that obtained using gridding. (Random number seeds were aligned to avoid Monte Carlo differences in the comparisons between gridding and the use of the Nelder-Mead simplex algorithm.) Maximization over $\mathcal{L}_0(\hat{\psi}) \cap R$ was done by gridding as this was simple to accomplish.

Table 2.2 below displays the percent of times the event in (2.8) occurred for each of the three models with and without an orthogonal design. If the event occurs then that portion of the true linear subspace, $\mathcal{L}_0$ (within E), is entirely covered by the GZG confidence region; otherwise it is not. For each simulated data set, the critical value c^*_α or $\tilde{c}_\alpha$ was computed using 1000 Monte Carlo simulations. For each critical value, 5000 simulations were done to assess the simultaneous coverage of the GZG confidence region for the set $\mathcal{L}_0 \cap E$.

Table 2.2 indicates that the simultaneous coverage rate of the GZG confidence region using the conservative critical value, c^*_α, produces reasonably

conservative results, while the approximate approach (that maximizes over $\mathcal{L}_0(\hat{\psi}) \cap R$, instead) achieves closer to nominal (yet slightly conservative) coverage rates. It is interesting to note that for each approach the coverage rate appears to be insensitive to the minor departure from orthogonality that was induced by changing the (one) factorial point (comprised of all -1's) from $(-1, -1, \cdots, -1, -1)$ to $(-1, -1, \cdots, -1, 0)$. Such a departure from design orthogonality could happen due to a design execution error or a process restriction.

Full model non-orthogonal design case

Because computation of $\tilde{c}_\alpha$ and c^*_α requires maximization within a Monte Carlo calculation, it would be useful to assess if this can be eliminated when a full model is employed. We therefore conduct another simulation study to see if the critical value based upon Corollary 2.1 can be used as an approximate critical value for mild departures from orthogonality. We use the same non-orthogonal designs as used in Table 2.2. The corresponding full interaction models are listed in Table 2.3. This time the Corollary 2.1 critical value was used with these non-orthogonal designs to assess the simultaneous coverage rate. The results are shown in Table 2.4 below. In order to assess the coverage rate gridding had to be done over a subset of $\mathcal{L}_0$. As a more fair comparison with the theoretical c_α critical value, gridding was done over $\mathcal{L}_0 \cap R$, (where

Table 2.3: The Full Interaction Model Equations for Simulation

Model no.	Model Equation
4	$y = 1 + x_1 + x_2 + x_3 + x_1^2 + x_2^2 + x_3^2 + x_2 \times x_3 + z_1 + x_1 \times z_1 + x_2 \times z_1 + x_3 \times z_1 + z_2 - x_1 \times z_2 + x_2 \times z_2 - x_3 \times z_2$
5	$y = 1+x_1+x_2+x_3+x_4+x_1^2+x_2^2+x_2\times x_3+z_1+x_1\times z_1+x_2\times z_1+x_3\times z_1+x_4\times z_1+z_2-x_1\times z_2+x_2\times z_2-x_3\times z_2+x_4\times z_2$
6	$y = 1 + x_1 + x_2 + x_3 + x_4 + x_1^2 + x_2^2 + x_2 \times x_3 + z_1 + x_1 \times z_1 + x_2 \times z_1 + x_3 \times z_1 + x_4 \times z_1 + z_2 - x_1 \times z_2 + x_2 \times z_2 - x_3 \times z_2 + x_4 \times z_2 + z_3 + x_1 \times z_3 + x_2 \times z_3 + x_3 \times z_3 - x_4 \times z_3$

as before R is a hypercube region composed of the Cartesian product of the intervals $[-10, 10]$, instead of $[-1, 1]$). This is because the Corollary 2.1 critical value is computed by maximization over all of the linear subspace, $\mathcal{L}$.

Table 2.4 indicates that this minor departure from orthogonality has virtually no effect on the coverage rate of the GZG confidence region when the more convenient Corollary 2.1 critical value is used. For more radical departures from orthogonality, it may possibly be safer to use the conservative c_α^* critical value. But further robustness studies are needed to ascertain how well the more convenient Corollary 2.1 critical value works under departures from its assumptions.

2.2 Computation of Confidence Region

For the $k > h$ case, once we have computed the critical value, the confidence region can be computed by searching linear subspaces, $\mathcal{L}$, that satisfy

Table 2.4: The coverage rate of Corollary 2.1 critical values for the full interaction model non-orthogonal Case. Simulation size=100,000 and nominal coverage rate is 95%. Corollary 2.1 critical values are generated from 1,000,000 Monte Carlo Simulations of Chi-square and Wishart distributions

Model no.	True Solution Set	k	h	v	Critical Value	Coverage Rate (%)
4	A line	3	2	24	10.02	95.2
5	A plane	4	2	56	11.58	95.0
6	A line	4	3	49	11.70	95.1

the condition as defined in (2.3). However, searching over various lines or hyperplanes that span an experimental region is more computationally difficult than searching the same experimental region in a pointwise fashion. Fortunately, it can be shown that for any given critical value, the GZG confidence region can be computed by pointwise gridding. This is because for $C_x(c_\alpha)$ in (1.6) and $C_{\mathcal{L}}(c_\alpha)$ in (2.1), with the same critical value, $C_x(c_\alpha) = C_{\mathcal{L}}(c_\alpha)$. A proof is provided in Appendix A.4. This equivalency shows that one can construct the GZG confidence region by simply gridding over the experimental region in a pointwise fashion.

2.3 Examples

2.3.1 One Noise Variable

This example is taken from Myers, Kim, Griffiths (1997). It was originally taken from Montgomery (2009, page 231). The data was generated from a 2^4 factorial experiment from a pilot plant to explore the factors that could affect

the filtration rate of a chemical bonding substance. The goal is to maximize the filtration rate, y. There are three control variables: x_1 = pressure, x_2 = concentration, x_3 = stirring rate, and one noise variable: z = temperature. The response y is filtration rate. The model is

$$y = 70.06 + 10.81z + 4.94x_2 + 7.31x_3 - 9.06x_2z + 8.31x_3z - 0.56x_2x_3$$

with mean square error equal to 21.12 and residual df equal to 9. x_1 is not in the predictive model equation because the effect of x_1 is not statistically significant. The estimated slope of noise variable is:

$$\hat{\gamma} + \hat{\Delta}x = 10.81 - 9.06x_2 + 8.31x_3$$

Therefore, $k = 2$ and $h = 1$. The solution to the null hypothesis $H_0 : \gamma + \Delta x = 0$ is a line. Then the general critical value $2F(1 - \alpha, 2, v)$ should be used to calculate the confidence region in control variables for the zero slope in the noise variable. For comparison purposes, both the MKG confidence region and the GZG confidence region are computed and shown in Figure 2.1. The MKG confidence region is same as what has been published in Myers et al. (1997, Figure 2). But the GZG confidence region is much wider than MKG region.

Next, we do some simulations to compare the coverage rates of the GZG and MKG confidence regions. Since the true optima is not known in practice, we calculate the coverage rate for the solution to $\hat{\gamma} + \hat{\Delta}x = 10.81 - 9.06x_2 + 8.31x_3 = 0$ using a simulation model equal to the fitted model with $\sigma^2 = 21.12$.

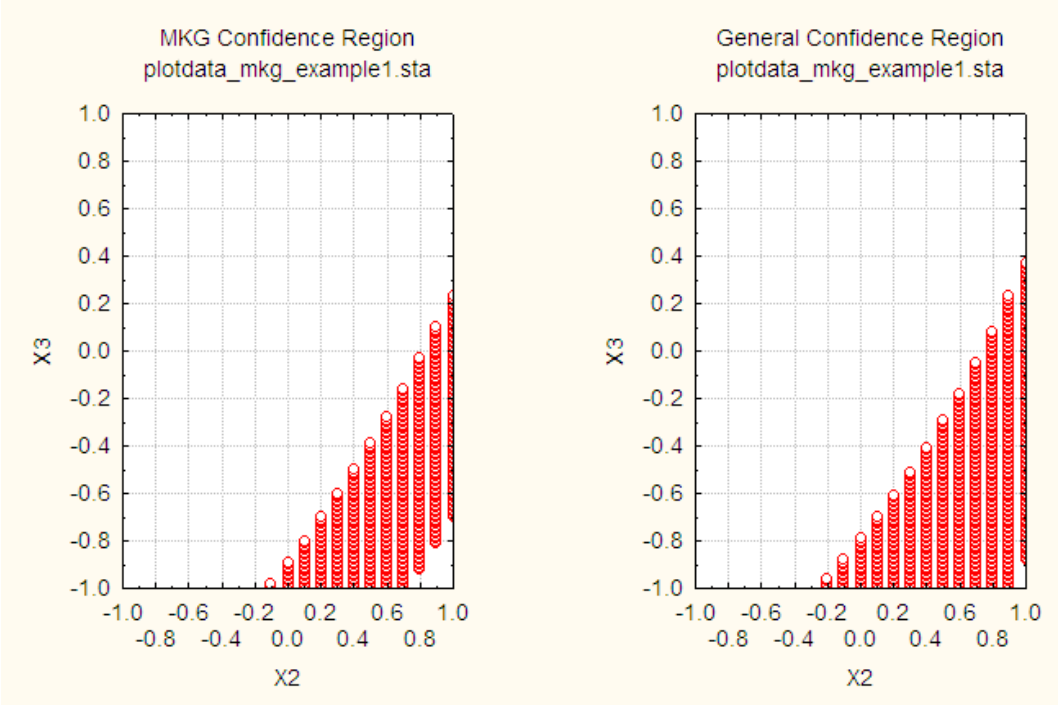

Figure 2.1: The comparison between MKG (left) and GZG confidence region (right)

Please note that the true solution in this example is a line with infinite length. But the simulation is done only for the line within the experimental region, i.e. $[-1, 1]$ of the control variables. Using 100,000 Monte Carlo simulations, the simultaneous coverage rate of the GZG confidence region for all of the zero-gradient solutions in the experimental region is 97% while MKG confidence region only has 91% coverage rate. MKG confidence region has lower coverage rate because it was designed to contain the true optima only when the optima is a point. Although the GZG confidence region is designed to contain all the true solutions (which could be a point, a line, or a hyperplane), the simulated coverage rate tends to exceed the nominal coverage rate because the simulation is done to cover a finite range of the $x's$ while the line or hyperplane has infinite range in theory.

2.3.2 Two Noise Variables

This example comes from a face-centered central composite design with the factorial part being a half fractional factorial design (see details in Kristensen, Xu, and Mu, 2005). The objective of this study is to find the optimized condition that maximizes the yield of diacylglycerol oil, which is a natural component of various edible oils and has been shown some beneficial effects as compared to the traditional triacylglycerol oil. Five factors were studied in this experiment: reaction time (RTIME), enzyme load (ENZL), reaction

temperature (RTEMP), water content (WATC), substrate molar ratio (SUBR). Water content (WATC) is difficult to control at large scale (Kristensen, et al. 2005) and therefore treated as a noise variable. For illustration purposes, substrate molar ratio is also treated as a noise variable and the axial points corresponding to the noise variables are excluded from the analysis to obtain partial design orthogonality with respect to the noise variables (i.e. to ensure $V_{\hat{\psi}} = c\sigma^2 I$). The final model in coded factor value is as follows:

$$\begin{aligned}\hat{y} &= 57.58 + 9.12x_1 + 4.78x_2 + 11.01x_3 - 4.69x_1^2 - 9.47x_2^2 - 7.37x_3^2 - 1.61x_2x_3 \\ &- 2.05z_1 + 4.83z_2 - 2.92z_1x_1 - 2.07z_1x_2 - 2.12z_1x_3 \\ &- 3.17z_2x_1 + 4.90z_2x_2 - 2.41z_2x_3\end{aligned}$$

where x_1 =RTIME, x_2 =ENZL, x_3 =RTEMP, z_1 =WATC, z_2 =SUBR. Here, the residual mean squared error is equal to 2.56 with 25 observations and residual df equal to 9.

Since $h = 2, k = 3$, the solution to the null hypothesis is a line in a 3-dimensional space determined by control variables x_1, x_2, and x_3. Therefore, the confidence region for this line is a tube in this 3-dimensional space. A 95% GZG confidence region is shown in Figure 2.2. Based on Corollary 2.1, the GZG critical value is obtained via chi-square and Wishart distribution. As with the previous example, we compare the coverage rates for the GZG and MKG confidence regions using the fitted model as the true population

model. Using 100,000 Monte Carlo simulations (based upon the fitted model with $\sigma^2 = 2.56$) the simultaneous coverage rate for GZG confidence region is 96% while it is only 90% for the nominal 95% MKG confidence region. (Here, gridding was done over the cube formed by the Cartesian product of $[-1, 1]$ associated with each x_i.)

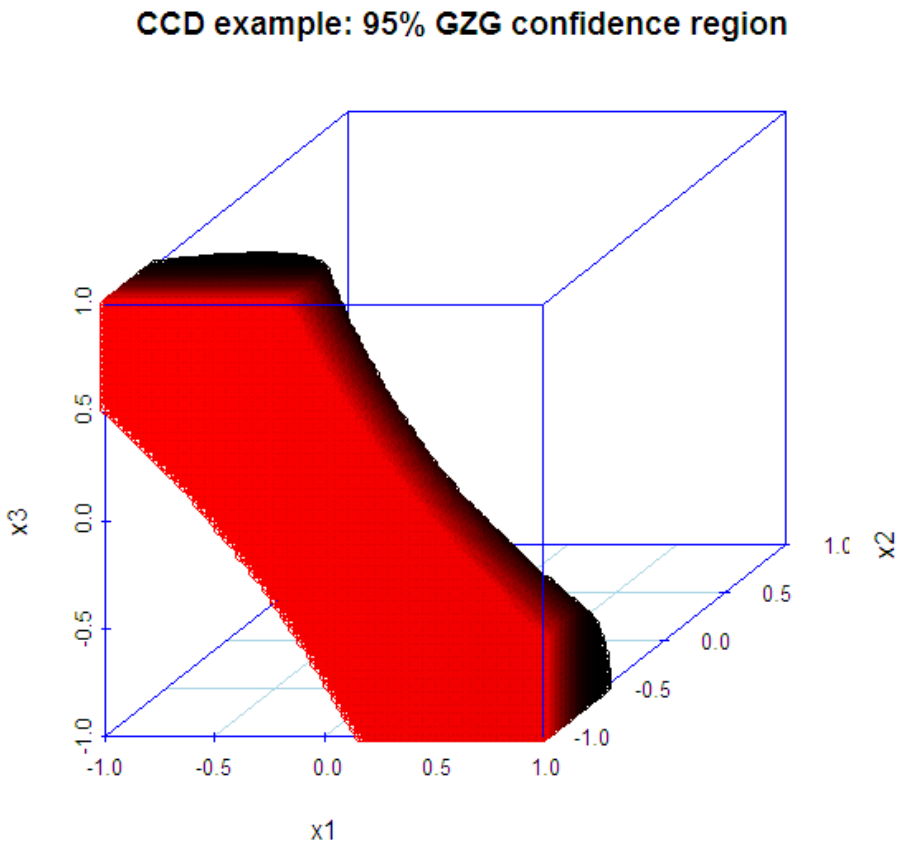

Figure 2.2: The 95% GZG confidence region for the two-noise variable case

2.4 Summary

This chapter shows that when the number of control variables does not exceed the number of noise variables, the MKG approach provides a confidence region for control variables associated with a zero-gradient for noise transmission. Otherwise, the MKG approach results in a confidence region that is too small for simultaneous coverage of the linear subspace of zero-gradient solutions. It is important to know that the true optimal condition represented by control variables is either a line or a hyperplane instead of a single point when $h < k$. In this situation, constructing a simultaneous confidence region about the linear subspace solution is desirable in that a subspace of solutions provides the investigator with many options for setting the zero-gradient control level. Of course a confidence region also provides the experimenter with a measure of uncertainty for the optimal solution. If the confidence region is too large, further experimental runs may be needed to make more accurate inferences. If the current manufacturing set-point is outside of the confidence region, this provides statistically significant evidence that re-configuration of the set point should help improve process variability by lowering the transmission of noise through the system. The GZG confidence region for the zero gradient conditions is proposed and is shown to provide nominal or reasonably conservative coverage rates for many noise variable experiments that occur in practice.

CHAPTER 3

GENERALIZED CONFIDENCE REGION FOR THE OPTIMAL SETTINGS FOR BOTH THE VARIANCE AND THE MEAN

The GZG confidence region provides inferences about the optimal control point or points that yield a zero-gradient for the transmission of variability from the noise variables. However, practitioners do not always have the fortune of having the optimal response mean and small response variance in the

same control variable area. There are also situations where the control points corresponding to the zero noise variance are either outside the experimental region or simply do not exist. For instance, let $h = 3$ and $k = 2$ and further assume that $\gamma + \Delta' x = 0$ takes the following form:

$$1 + x_1 + x_2 = 0$$

$$-1 - x_1 + x_2 = 0$$

$$1 - x_1 + x_2 = 0$$

Then the above equations have no solutions. However, the noise variance is $V^z_{y|x} = (\Delta' x + \gamma)' V_z (\Delta' x + \gamma)$. To minimize the noise variance, we can take a derivative of the noise variance with respect to x, which is $\Delta(\gamma + \Delta' x)$. If we let the derivative equal to zero, then we get a series of two equations with two control variables, which leads to a unique solution of $x_1 = x_2 = -\frac{1}{2}$. In other words, there are situations where the solution to zero-gradient does not exist, but it does exist for minimum noise variance which is not zero. Also, $\gamma + \Delta' x = 0$ is a sufficient condition to Equation $\Delta(\gamma + \Delta' x) = 0$, but not a necessary condition. If Δ is $k \times k$ and nonsingular or if $\Delta'\Delta$ is nonsingular, then $2\Delta(\gamma + \Delta' x) = 0$ implies $\gamma + \Delta' x = 0$, which is the case in Chapter 2. Therefore, the x that satisfies $\Delta(\gamma + \Delta' x) = 0$ is the actual solution to minimize the noise variance.

A more systematic methodology is needed to address the need for response

mean as well as the variance. Kuhn (2003) developed the confidence region to optimize the mean and variance simultaneously using the constraint optimization approach. The users have the flexibility to choose if they want to optimize the mean under the constraint of small variance value or if they want to minimize the variance under the constraint of certain target value of mean. However, this is an approximate method. It is highly dependent upon the value of a Lagrange Multiplier. In this method, the Lagrange multiplier is simply the maximum likelihood estimates of the model parameters, which creates additional uncertainty. Also, it is hard to apply this method to more complicated situations like mixture experiments because of additional constraints.

We developed a more general method to compute the confidence region that optimizes the mean as well as the variance. This general method is applicable to most of the situations including mixture experiments which have inherit constraints on the control factors. It also gives users the flexibility to put different weights on the mean and variance parts for the simultaneous optimization.

3.1 Proposal of the Generalized Approach

3.1.1 Theoretical formulation on WMSE optimization

Assuming a more general response model with noise variables as shown in (3.1) below,

$$y(x,z) = u(x)\beta + z'\gamma + x'\Delta z + \epsilon \tag{3.1}$$

where x is the vector of the control variables. $u(x)$ is a vector of all the x terms in the mean model part. For quadratic models,

$$u(x) = \left(1, x_1, x_2, \cdots, x_k, x_1^2, x_2^2, \cdots, x_k^2, x_1x_2, \cdots, x_{k-1}x_k\right)'$$

β is the vector of all the model parameters in the mean model part. z is the vector of the noise variables. γ is the vector of the main effects of the noise variables. Δ is the interaction matrix as defined in equation (1.1). ϵ is the residual noise that is assumed to follow normal distribution with mean 0 and variance σ_ϵ^2. The introduction of $u(x)$ allows more complicated model terms within control variables like x_2^3.

To optimize both mean and variance, we could consider minimizing a weighted mean square error (WMSE) function as follows.

$$WMSE = \lambda_1(u(x)'\beta - T)^2 + \lambda_2(\gamma + \Delta'x)'(\gamma + \Delta'x)\sigma_z^2 \tag{3.2}$$

where T is the target mean for the response variable Y. λ_1 and λ_2 are different weights. $u(x)'\beta$ is the generalized representation of the model terms

related to the control variables. Using WMSE, we combined the optimization of mean and variance together into one function. The introduction of the weights λ_1 and λ_2 gives users more flexibility to adjust the objective function based on the needs.

If $\lambda_1 = \lambda_2$, then $WMSE = E[(Y - T)^2|x]$. If $\lambda_1 = 0$ and $\lambda_2 = 1$, then we have $WMSE = (\gamma + \Delta' x)'(\gamma + \Delta' x)\sigma_z^2$. For large, positive λ_1 and $\lambda_2 = 1$, then minimizing WMSE will produce approximately a minimum of $(\gamma + \Delta' x)'(\gamma + \Delta' x)\sigma_z^2$ subject to the constraint $\beta_0 + s(x)'\beta = T$. All of these optimization situations are useful for robust parameter design (RPD) problems. Hence, we can use WMSE as a generalized version of objective functions.

Since WMSE is quadratic in the model parameters, it is possible to write WMSE in the form

$$WMSE = q(x;\phi) = \lambda_1 T^2 + G(x)\phi + \phi' F(x)\phi \tag{3.3}$$

where $\phi' = (\beta', \gamma', \delta_{11}, ..., \delta_{hk})'$ is the vector of all regression model coefficients including both the mean part and the noise variable part, for some known matrices G and F. If the noise variables interact with the same set of control variables and the control variable part takes the quadratic form, then $G(x)$ and $F(x)$ can be expressed as follows. The expression for $F(x)$ will change if the model does not have full interaction terms (see fish patty

example).

$$G(x) = \begin{bmatrix} -2\lambda_1 T u(x)' & 0_{1\times npsi} \end{bmatrix} \tag{3.4}$$

$$F(x) = \begin{bmatrix} \lambda_1 u(x)u(x)' & 0 \\ 0 & \lambda_2 \sigma_z^2 a(x)a(x)' \otimes I_h \end{bmatrix} \tag{3.5}$$

where

$a(x) = (1, x')', x = (x_1, x_2, \cdots, x_k)'$ if $\psi = (\gamma', vec(\Delta')')'$

$i.e.$ $\psi = (\gamma_1, \gamma_2, \cdots, \gamma_h, \delta_{11}, \delta_{21}, \cdots \delta_{h1}, \cdots \delta_{1k}, \delta_{2k}, \cdots \delta_{hk})'$

$npsi$ is the length of ψ

δ_{ij} = the interaction coefficient for noise variable z_i and control variable x_j

k = the total number of control variables in the mean model.

h = the total number of noise variables.

β=$(\beta_0, \beta_1, \cdots, \beta_k, \beta_{11}, \cdots \beta_{kk}, \beta_{12}, \cdots, \beta_{k-1,k})'$

$\phi = (\beta', \psi')'$

Note the expression $F(x)$ will change if the order of the ψ elements is different.

Also note that the noise variance part $(\gamma + \Delta'\mathrm{x})'(\gamma + \Delta'\mathrm{x})\sigma_z^2$ here is writ-

ten as $\sigma_z^2 \psi' a(x)a(x)' \otimes I_h \psi$ because under full interaction model

$$\begin{aligned}
&(\gamma + \Delta' \mathrm{x})' (\gamma + \Delta' \mathrm{x}) = \psi' \big(a(x)' \otimes I_h\big)' \big(a(x)' \otimes I_h\big) \psi \\
&= \psi' (a(x) \otimes I_h) \big(a(x)' \otimes I_h\big) \psi \\
&= \psi' \big(a(x)a(x)'\big) \otimes I_h \psi
\end{aligned}$$

To find the confidence region for the x_0 point that minimizes WMSE, we can follow the same approach as indicated by Peterson et al. (2002). Let $\eta(\phi) = \min_{x \in R} q(x; \phi)$, where R is the experimental region. Consider the hypothesis test $H_0 : \eta(\phi) - q(x_0; \phi) = 0$. Then a confidence region for x_0 is $\{x_0 : H_0 : \eta(\phi) - q(x_0; \phi) = 0$ is not rejected at level $\alpha\}$. We can test $H_0 : \eta(\phi) - q(x_0; \phi) = 0$ by forming the confidence interval

$$\left[x_0 : \min_{\phi \in C_\phi} \left(\eta(\phi) - q(x_0; \phi)\right), \max_{\phi \in C_\phi} \left(\eta(\phi) - q(x_0; \phi)\right) \right] \tag{3.6}$$

where C_ϕ is a confidence region for ϕ. We call the confidence region in (3.6) the general confidence region. Spjøtvoll (1972) showed that such min-max confidence intervals for a family of functions have simultaneous coverage probability of at least $(1 - a)$ if C_ϕ is a $100(1 - \alpha)\%$ confidence region for ϕ. Note that Peterson et al. (2002) do not use the conservative Scheffes critical value, but instead utilize the appropriate critical value $kF(k, n - p)$ (see Peterson et al. (2002) for details.), where k is the number of the control variables, n is the total number of observations, and p is the total number of model parameter

estimates.

If (3.6) contains 0 we fail to reject H_0. If (3.6) does not contain 0 then we must reject H_0. Note that $\eta(\phi) - q(x_0;\phi) \leq 0$ by the definition of $\eta(\phi)$. In other words, the confidence interval (3.6) is either less than zero or equal to zero. The confidence interval contains zero when $q(x_0;\phi) = \eta(\phi)$. Hence, we only need to consider the upper bound of the confidence interval in (3.6). So if $\max_{\phi \in C_\phi}(\eta(\phi) - q(x_0;\phi)) < 0$ then we reject H_0; otherwise, $\max_{\phi \in C_\phi}(\eta(\phi) - q(x_0;\phi)) = 0$, and we fail to reject H_0. Now consider $\max_{\phi \in C_\phi}(\eta(\phi) - q(x_0;\phi))$. Note that

$$\begin{aligned}\max_{\phi \in C_\phi} \left(\eta(\phi) - q(x_0;\phi)\right) &= \max_{\phi \in C_\phi} \left(\min_{\mathrm{w} \in R} q(\mathrm{w};\phi) - q(x_0;\phi)\right) \\ &= \max_{\phi \in C_\phi} \left(\min_{\mathrm{w} \in R} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right)\end{aligned}$$

However, we could replace $\max\limits_{\phi \in C_\phi} \left(\min\limits_{\mathrm{w} \in R} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right)$ with $\min\limits_{\mathrm{w} \in R} \left(\max\limits_{\phi \in C_\phi} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right)$ to use $\min\limits_{\mathrm{w} \in R} \left(\max\limits_{\phi \in C_\phi} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right) < 0$ as a rejection criterion.

Note that $\min\limits_{\mathrm{w} \in R} \left(\max\limits_{\phi \in C_\phi} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right) \geq \max\limits_{\phi \in C_\phi} \left(\min\limits_{\mathrm{w} \in R} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right)$ by min-max theorem (See Bazaraa, Sherali, and Shetty, 2nd Edition, page 240, problem 6.44), which may make the confidence interval more conservative. Also, $\min\limits_{\mathrm{w} \in R} \left(\max\limits_{\phi \in C_\phi} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right)$ is much easier to deal with. In other words, if there exists a point w such that $\max\limits_{\phi \in C_\phi} \left[q(w;\phi) - q(x_0;\phi)\right] < 0$, then $\min\limits_{\mathrm{w} \in R} \left(\max\limits_{\phi \in C_\phi} \left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right) < 0$ is guaranteed.

In summary, the confidence interval

$$\{x : H_0 : \ \eta(\phi) - q(x;\phi) = 0 \ \text{ is not rejected at level } \alpha\}$$

can be determined by searching through the grid points in the experimental region or calibration region R. Only the point x_0 that satisfies the condition: $\min_{\mathrm{w}\in R}\left(\max_{\phi\in C_\phi}\left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right) \geq 0$ is in the confidence region. Otherwise, it is not. Note: It is possible for $\min_{\mathrm{w}\in R}\left(\max_{\phi\in C_\phi}\left[q(\mathrm{w};\phi) - q(x_0;\phi)\right]\right)$ to hold positive, negative, or zero values depending upon the values of w and x_0, but $\max_{\phi\in C_\phi}\left(\min_{w\in R}\left[q(w;\phi) - q(x_0;\phi)\right]\right) \leq 0$ is always held.

3.1.2 General Confidence Region Computation Algorithm

In this section we develop a computation algorithm for the general confidence region.

Notation 3.1. *Let x and w be the distinct points inside the constrained experimental region R.*

Definition 3.1. *A point w "dominates" point x if x is rejected out of the confidence region by w and is denoted as $w \mapsto x$.*

To make the computation more efficient, the following theorem and corollaries are developed.

Theorem 3.1. *If $f(x)$ and $g(x)$ are two functions of x and x is a vector in R, a real region, then* $\max_x (f(x) + g(x)) \leq \max_x (f(x)) + \max_x (g(x))$

Proof. Let $f(x_f) = \max_x(f(x))$ and $g(x_g) = \max_x(g(x))$.

Let $f(x_0) + g(x_0) = \max_x(f(x) + g(x))$, where $x_f, x_g, x_0, x \in R$.

If $x_0 = x_f = x_g$, then

$$\max_x(f(x) + g(x)) = f(x_0) + g(x_0) = \max_x(f(x)) + \max_x(g(x))$$

If $x_0 \neq x_f$ or $x_0 \neq x_g$, then

$$f(x_0) \leq f(x_f) = \max_x(f(x))$$

$$g(x_0) \leq g(x_g) = \max_x(g(x))$$

Hence,

$$f(x_0) + g(x_0) = \max_x(f(x) + g(x)) \leq \max_x(f(x)) + \max_x(g(x))$$

where the strict inequality holds when the maxima is unique.

□

Theorem 3.2. *If w dominates x and x dominates $\tilde{x}$ (where w, x, $\tilde{x}$ are distinct points in R), then w dominates $\tilde{x}$. In other words, the binary relation $\mapsto$ on the set R is transitive, that is, if $w \mapsto x$ and $x \mapsto \tilde{x}$, then $w \mapsto \tilde{x}$.*

Please note that this theorem may also apply to more general model forms as the proof below does not use the fact that the objective function $q(x; \phi)$ is a quadratic function of ϕ.

Proof. We define $w \mapsto x$ iff

$$\max_{\phi \in C_\phi} [q(w;\phi) - q(x;\phi)] < 0 \tag{3.7}$$

If $x \mapsto \tilde{x}$, then

$$\max_{\phi \in C_\phi} [q(x;\phi) - q(\tilde{x};\phi)] < 0 \tag{3.8}$$

Add (3.7) and (3.8) together, we get

$$\max_{\phi \in C_\phi} [q(w;\phi) - q(x;\phi)] + \max_{\phi \in C_\phi} [q(x;\phi) - q(\tilde{x};\phi)] < 0$$

By theorem 3.1,

$$\begin{aligned}\max_{\phi \in C_\phi} [q(w;\phi) - q(\tilde{x};\phi)] &= \max_{\phi \in C_\phi} [q(w;\phi) - q(x;\phi) + q(x;\phi) - q(\tilde{x};\phi)] \\ &\leq \max_{\phi \in C_\phi} [q(w;\phi) - q(x;\phi)] + \max_{\phi \in C_\phi} [q(x;\phi) - q(\tilde{x};\phi)]\end{aligned}$$

Therefore,

$$\max_{\phi \in C_\phi} [q(w;\phi) - q(\tilde{x};\phi)] < 0$$

i.e. $w \mapsto \tilde{x}$. □

Corollary 3.1. *Let ε be the set of points in R that are already eliminated (i.e. determined to be outside the C_R) at the current iteration process and let $R\backslash\varepsilon = R - \varepsilon$ (i.e., the set of points not eliminated so far). Then, for discarding the remaining $x \in R\backslash\varepsilon$, we just need to consider $w \in R\backslash\varepsilon$ as opposed to considering $w \in R$.*

Proof. The same proof by Cahya, del Castillo, Peterson (2004) applies and is also shown as the following: Let $x_1, \cdots, x_j$, where j is an integer, be the points in the set ε such that $x_1 \mapsto x_2 \mapsto \cdots \mapsto x_j$, and there are no other points $w \in \varepsilon$ such that $w \mapsto x_1$. Let $\tilde{x}$ be a point in the set $R \backslash \varepsilon$ such that $x_j \mapsto \tilde{x}$. Because $x_1 \in \varepsilon$, it implies that there exists a point $w \in R \backslash \varepsilon$ such that $w \mapsto x_1$. However, theorem 3.1 implies that $w \mapsto \tilde{x}$. Therefore, the points $x_1, x_2, \cdots, x_j$ are not needed for eliminating $\tilde{x}$. □

Corollary 3.2. *For all $x \notin C_R$, there exists $w \in C_R$ such that $w \mapsto x$*

The same proof in (Cahya, del Castillo, Peterson, 2004) holds.

Corollary 3.2 implies that we can always find a point within the current confidence region to reject the points that do not belong to the final confidence region. Based on Corollary 3.1, once a point is rejected out of the confidence region, it does not need to be considered any more when we assess the next point because only the points in the confidence region could reject the next point of consideration. These Corollaries will speed up the computation significantly (see the time saved by Corollary 3.1 in Example 3.3.1 and Example 3.3.3). Based on the corollaries, the following three-step computation algorithm is used to find the general confidence region.

1. Grid the experimental or calibration region R;

2. Number the grid points from 1 to n; let the status of all the grid points

be in the confidence region;

3. Start with x_0. If the status of x_0 point is in, then pick up a point w that is different from x_0 point and the status of w point is also in. Compute $f = \max_{\phi \in C_\phi} q(w, x; \phi)$, where $q(w, x; \phi) = q(w; \phi) - q(x_0; \phi)$. see details below on how to compute f. If $f < 0$ then $\min_x \max_\phi q(w, x, \phi) < 0$ is guaranteed. Therefore x_0 point is rejected by w and the status of x_0 point is out (meaning out of the confidence region). No more searching for w is needed. Then move to check the next x point. If $f >= 0$, then use the next available w point to continue to check against the x_0 point until either x_0 is rejected or all the available w points have been checked.

Note: Fine grids are needed to ensure that enough w points are checked against x point. Or SAS IML nlpnms subroutine may be used to refine the search at the end. Just need to make sure that w stays within the experimental region R.

The computation of the quantity $f = \max_{\phi \in C_\phi} q(w, x; \phi)$ involves maximizing the quantity $q(w, x; \phi)$ over ϕ. It is critical that the true maximum of $q(w, x; \phi)$ is reached. Otherwise, false rejection could occur and lead to empty "holes" in the confidence region. Since we only care about the sign of the optimized quantity, not the actual value, one of the computation steps can be ended earlier. For any given x and w, once we find any ϕ such that $q(w, x; \phi) > 0$, then it is guaranteed that $\max_\phi q(w, x; \phi) > 0$. Therefore the

effort should focus on the case where $q(w, x; \phi) < 0$. In this case, the search for ϕ needs to continue until either $q(w, x; \phi) > 0$ is reached or all the possible $\phi \in C_\phi$ have been checked. Since the vector ϕ is a vector of multiple model parameters, it is impractical to check the maximum by gridding $\phi \in C_\phi$. Note that $q(w, x; \phi)$ can be written as a quadratic function of ϕ:

$$q(w, x; \phi) = (G(w) - G(x))\phi + \phi'(F(w) - F(x))\phi \tag{3.9}$$

Let $F(w, x) = F(w) - F(x)$. If $F(w, x)$ is negative definite, then we can solve for ϕ_{max}, where $q(w, x; \phi_{max}) = \max q(w, x; \phi)$, by letting the derivative of the $q(w, x; \phi)$ with respect to ϕ equal to zero. If $\phi_{max} \in C_\phi$, then $\max_{\phi \in C_\phi} q(w, x; \phi) = q(w, x; \phi_{max})$ by Theorem 3.3. Otherwise, the ϕ that maximizes $q(w, x; \phi)$ in C_ϕ is on the boundary of C_ϕ by Theorem 3.4.

If the ϕ that maximizes $q(w, x; \phi)$ in C_ϕ is on the boundary of C_ϕ, then Myers and Carter's algorithm (Myers and Carter, 1973) can be used to find $\max q(w, x; \phi)$. Myers and Carter's algorithm was developed to optimize a quadratic function subject to a quadratic constraint. In other words, both the objective function and the constraint function are quadratic functions of the parameter of interest. In this case, for any given x and w, the objective function $q(w, x; \phi)$ is a quadratic function of ϕ. The constraint function $(\phi - \hat{\phi})' V_{\hat{\phi}}^{-1} (\phi - \hat{\phi}) \leq c_\alpha^2$ is also a quadratic function of ϕ, where $c_\alpha^2 = kF(1 - \alpha, k, dfe)$, k is the number of the control variables, and dfe is the error degrees of freedom. If we let the critical value $c_\alpha^2 = pF(1 - \alpha, p, dfe)$, where p is the

dimension of ϕ, which is the total number of model parameters, we would get the exact confidence region where the confidence level is at least $1-\alpha$. However, as pointed out by Peterson (1993) and Peterson and Kuhn (2005), $pF(1-\alpha, p, dfe)$ is too conservative. Since this is a confidence region in control variables, $kF(1-\alpha, k, dfe)$ is chosen to be the critical value.

Steepest ascent(descent) algorithm developed by Peterson (1987) can be used to quickly find the maximum of $q(w, x; \phi)$ as well. This algorithm has been successfully used to compute the simultaneous confidence band for the ridge trace of a mean response surface in presence of noise variables (see Peterson and Kuhn (2005)). Peterson (1987) has also proved that the steepest ascent algorithm will converge to the solution of the Lagrangian system. However, it is not always guaranteed in practice to find the true maximum using steepest ascent algorithm. Myers and Carter's algorithm guarantees the optimum in practice although it generally takes longer to compute than steepest ascent algorithm. Therefore, Myers and Carter's algorithm was used to compute $\max q(w, x; \phi)$ if the optimal ϕ is on the boundary of C_ϕ. Note that $\max q(w, x; \phi) > 0$ is guaranteed if we find one $\phi^* \in C_\phi$ where $q(w, x; \phi^*) > 0$. Therefore, there is no need to find the true $\max q(w, x; \phi)$ and we can stop the computation early if $q(w, x; \hat{\phi}) > 0$.

Theorem 3.3. *Let ϕ_0 be the stationary point of the function $q(w, x, \phi)$, ϕ_{max} be the global maximum, and ϕ^* be the true maximum in C_ϕ. $\phi_0 = \phi_{max} = \phi^*$*

if the following two conditions are satisfied simultaneously: 1) $\phi_0 \in C_\phi$; 2) $F(w, x)$ is negative semi-definite (n.s.d).

See proof in Appendix A.5.

Theorem 3.4. *ϕ^* is on the boundary of C_ϕ if one or two of the two conditions is not satisfied. 1) $\phi_0 \in C_\phi$; 2) $F(w, x)$ is negative semi-definite (n.s.d).*

See proof in Appendix A.6.

3.2 Coverage rate

The nominal confidence level of the general confidence region 3.6 is $1 - \alpha$. However, the actual coverage rate could be higher than the nominal level since the upper bound of the confidence region is replaced by a potentially bigger quantity for computation purpose. To check the accuracy of the confidence region, the coverage rate is computed as follows. Assuming the empirical models based on the data sets are true, the optimal true solution point, x_0, is computed from the model. For each simulation, a data set was generated based on the model and the corresponding data design. Then the quantity, $\min_w \max_\phi q(w, x_0, \phi)$, is computed. If $\min_w \max_\phi q(w, x_0, \phi) < 0$, then $ID = 0$; otherwise, $ID = 1$. Do the simulation $10,000$ times. The coverage rate is defined as the ratio of the sum of ID to the total number of simulations. To compute $\min_w \max_\phi q(w, x_0, \phi)$, the quantity $\max_\phi q(w, x_0, \phi)$ needs to be

calculated for every w until a negative value is reached or the all the grid points in the experimental region are searched as w point. Therefore, the experimental region gridding is still needed for the coverage computation. Gridding can be done one time and stored in a matrix and then used repeatedly for each simulation. Because of this, the following steps are followed to compute the coverage rate.

1. Find the optional solution from the assumed true model, xmin.

2. Simulate the data set based on the same design and same model

3. Calculate the optional solution point based on the simulated data, w_0

4. Follow the previous procedure to only check xmin as the x point. i.e. the procedure includes checking using w_0 as well as any other grid points except xmin itself.

5. Repeat step 2 to 4 a large number of times. Record ID in each simulation. The coverage rate is the ratio of the sum of ID to the total number of simulation times.

3.3 Examples

3.3.1 TV example

This is an example originally from Myers, Montgomery, and Anderson-Cook (2009, p511). The objective of this study is to find the robust setting in control variables such that the television reception signal quality is high. Robust here means small variance around the optimized response value. The response signal quality is measured in decibels. The higher decibel number means better quality. A $3^2 \times 2^2$ crossed array design is used to model the effects of control and noise variables on the television signal quality. Two control factors are x_1, the number of tabs in a filter, and x_2, sampling frequency. Two noise factors are z_1, the number of bits in an image, and z_2, voltage. The noise variances are assumed to be 1 and covariance is zero. The model equation is:

$$\begin{aligned}\hat{y}(x,z) = 33.389 - 4.175x_1 + 3.748x_2 + 3.348x_1x_2 - 2.328x_1^2 - 1.867x_2^2 - 4.076z_1 \\ + 2.985z_2 - 2.324x_1z_1 + 1.932x_1z_2 + 3.268x_2z_1 - 2.073x_2z_2\end{aligned} \tag{3.10}$$

Note that this is a full interaction model. Therefore, the solution to the zero-gradient equation as well as WMSE is unique. To find the robust optimized region in control variables, Myers, etc. (2009) first calculated the zero-gradient confidence region. Although the highest response value (35.47) falls outside the zero-gradient confidence region, the mean response surface

is quite flat and only drops to 35 around the zero-gradient confidence region. Therefore, the zero-gradient confidence region can provide flexibility for the optimum condition. Kuhn constrained the variance to 0.81 (i.e. the minimum variance is 0.64), which resulted in an optimal mean of 35.3. Kuhn's confidence region for optimal mean turned out to be quite similar to Myers zero-gradient confidence region (i.e. MKG confidence region)(see Figure 3.1). Both of the confidence regions do not contain the highest predicted mean (35.47) point, which has a much higher process variance, 2.69. To compare with the Kuhns result, we set the mean target value to be 35.3 (which is the optimal mean value used by Kuhn) and weight ratio as 1 : 1 in our general confidence region approach. The resultant confidence region before and after Nelder Mead fine search is shown in Figure 3.2 and is again quite similar to Kuhns, but slightly narrower.

The efficiency of Corollary 3.1 was tested using this TV example with weight ratio 1 : 1. The cpu time using the corollary was 2 hours 31 minutes while it took 5 hours 45 minutes without the corollary.

The weight ratios of 1 : 10 and 10 : 1 have been computed as well. Weight ratio of 1 : 10 (Figure 3.3) (which puts more weight on the variance part) results in almost identical confidence region as weight ratio 1 : 1 (Figure 3.2). However, weight ratio 10 : 1 results in a quite different confidence region shape (see Figure 3.4). Since the grid levels are quite fine (the grid increment is 1%

Table 3.1: The coverage rate of the general confidence regions with different weight ratios

Weight Ratio	Coverage Rate
1 : 1	96.5%
1 : 10	96.0%
10 : 1	98.3%

of each factor range), the Nelder Mead refined search does not make practical differences as shown in Figure 3.2 to Figure 3.4.

All the confidence regions were computed at 95% confidence level. The coverage rates were computed for the general confidence regions and listed in Table 3.1. The Nelder Mead refined search was used in the coverage rate calculation. However, the addition of Nelder Mead procedure here did not make big difference due to the grids that are already fine enough. For instance, for weight ratio 1 : 10, the coverage rate before Nelder Mead procedure is 96.1%, only 0.1% higher than the one with Nelder Mead procedure.

3.3.2 Fish Patty Example

The data of this example came from Cornell, (1990). It is also shown in the appendix of the article by Steiner and Hamada (1997). The purpose of this study is to find the right combination of three fish species to get the best fish patty texture. The ideal fish patty texture value should lie between 2.0 and 3.5. The amount of each of the three fish species is represented by the three mixture factors: percent mullet (x_1), percent sheepshead (x_2), and percent

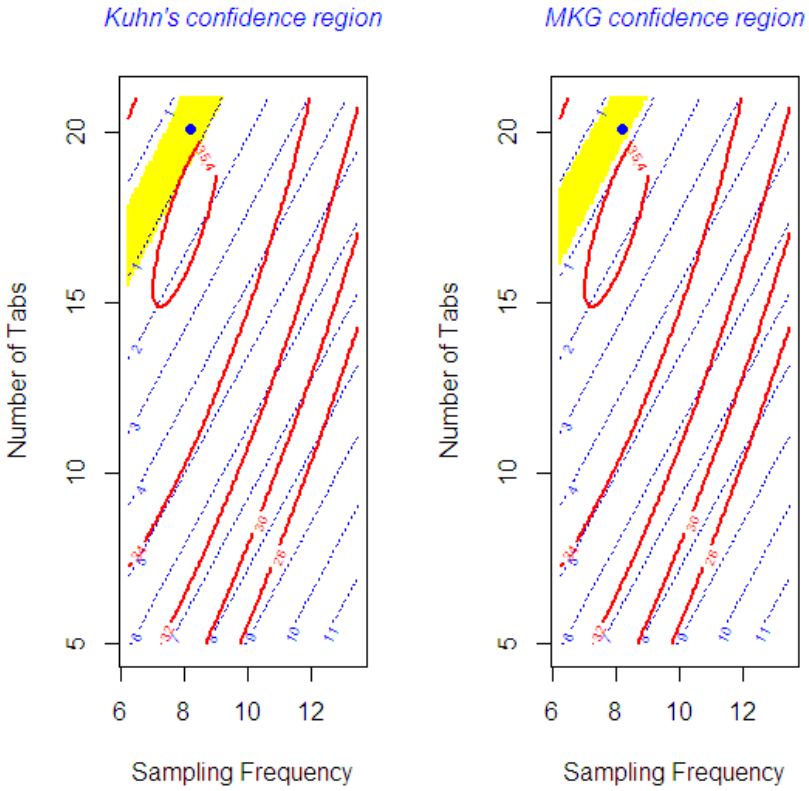

Figure 3.1: TV example: The Kuhn's and MKG confidence region. The optimal point (8.2, 20.12) based on Kuhn's method is shown as the blue dot for reference purposes.

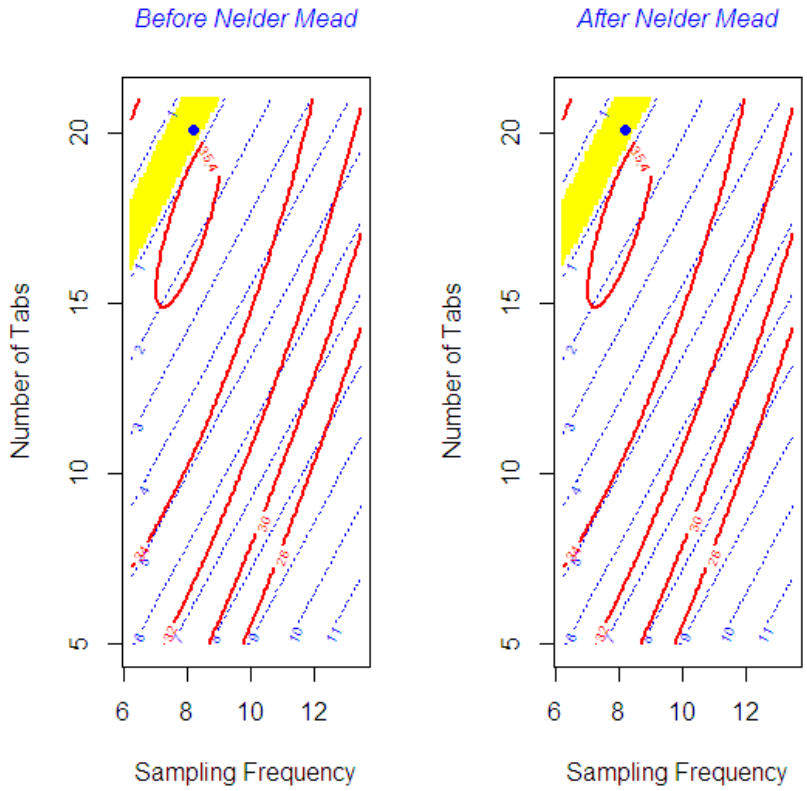

Figure 3.2: TV example: The general confidence region to optimize both the mean and the noise variance before and after Nelder Mead refined search. Weight ratio: 1 : 1. The optimal point (8.2, 20.12) based on Kuhn's method is shown as the blue dot for reference purposes.

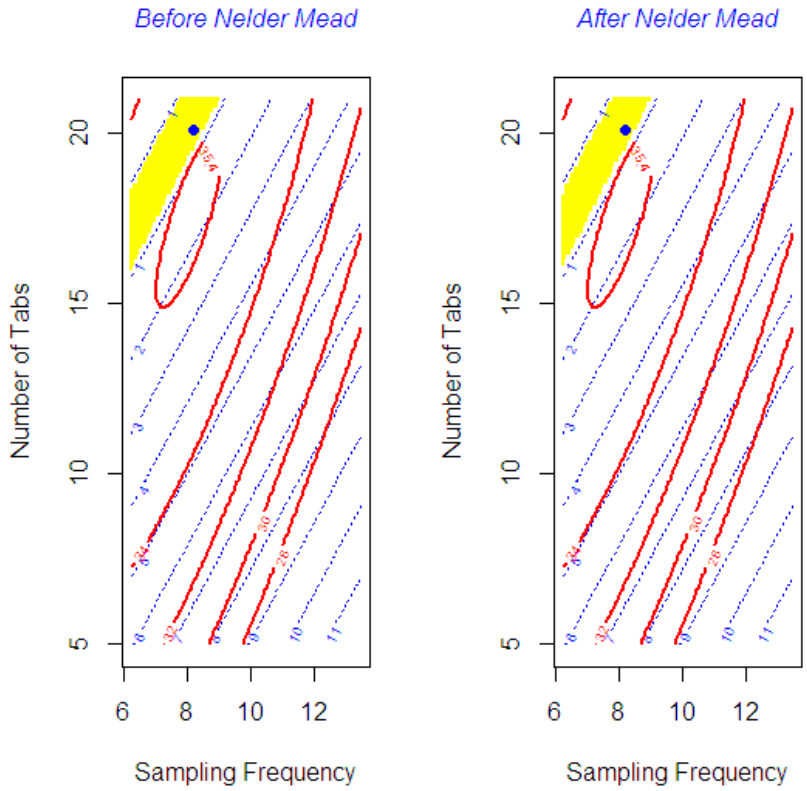

Figure 3.3: TV example: The general confidence region to optimize both the mean and the noise variance before and after Nelder Mead refined search. Weight ratio: $1:10$. The optimal point (8.2, 20.12) based on Kuhn's method is shown as the blue dot for reference purposes.

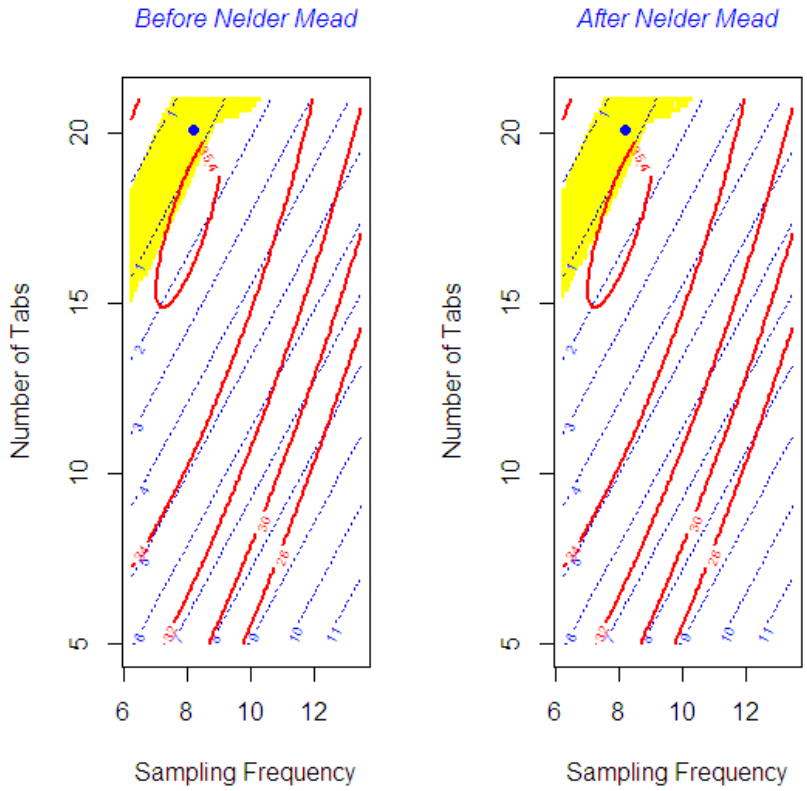

Figure 3.4: TV example: The general confidence region to optimize both the mean and the noise variance before and after Nelder Mead refined search. Weight ratio: $10:1$. The optimal point (8.2, 20.12) based on Kuhn's method is shown as the blue dot for reference purposes.

croaker (x_3). The study also involves three process factors: oven temperature (z_1), baking time (z_2), and deep frying time (z_3). Steiner and Hamada treated oven temperature (z_1), baking time (z_2) as noise variables because the oven temperature and baking time could be influenced by different customers. Since MSE is always low at $z_3 = -1$, Steiner and Hamada discussed the noise effect and mixture optimization by fixing $z_3 = -1$. Peterson and Kuhn (2005) applied ridge analysis to this example by assuming that the noise variables are independent, the corresponding variances are equal to $\frac{1}{9}$, and the target value equal to 2.5 at $z_3 = -1$. These are the assumptions used here as well. The model equation is also same as Peterson & Kuhn and Steiner & Hamadas and is shown below.

$$\begin{aligned}\hat{y} = {} & 2.86x_1 + 1.11x_2 + 2.03x_3 - 0.99x_1x_2 - 0.85x_1x_3 \\ & + z_1\,(0.44x_1 + 0.17x_2 + 0.19x_3 - 0.77x_1x_2) \\ & + z_2\,(0.64x_1 + 0.2x_2 + 0.4x_3) + z_1z_3\,(0.09x_3)\end{aligned} \tag{3.11}$$

Three weight ratios were computed for this example: 1 : 1; 100 : 1; and 1 : 100. The contour plots of mean and variance at $z_3 = -1$ were shown by Steiner and Hamada (1997, see Figure 2 and Figure 3) and are shown again in Figure 3.5 here. It can been seen that small variance favors high x_2 value while high mean values favor high x_1 area. Therefore the confidence region to

Table 3.2: The coverage rate of the general confidence regions with different weight ratios for the fish patty example

Weight Ratio	Coverage Rate
1 : 1	98.6%
100 : 1	98.5%
1 : 100	98.8%

minimize WMSE is toward high x_1 area for weight ratio 1 : 1 and 100 : 1, but it switches to high x_2 area when the weight ratio changes to 1 : 100 (Figure 3.6, 3.7, 3.8).

The coverage rates for the above general confidence regions are computed and summarized in Table 3.2. Since Nelder Mead refined search resulted in almost the same confidence regions as those without the refined search, the coverage rates were computed without Nelder Mead procedure. Since it is a mixture problem, polar transformation is used such that gridding can be done on the polar coordinates. The grid increment is 1 angle degree.

3.3.3 The General Confidence Region with Three Control Variables

This is the same example as shown in Section 2.3.2. Using target value equal to 50%, weight ratio equal to 1 : 1, grid increment equal to 1.25% of each factor range, the general confidence region is computed and shown in Figure 3.9. The coverage rate for this confidence region is 98.5%. Since the

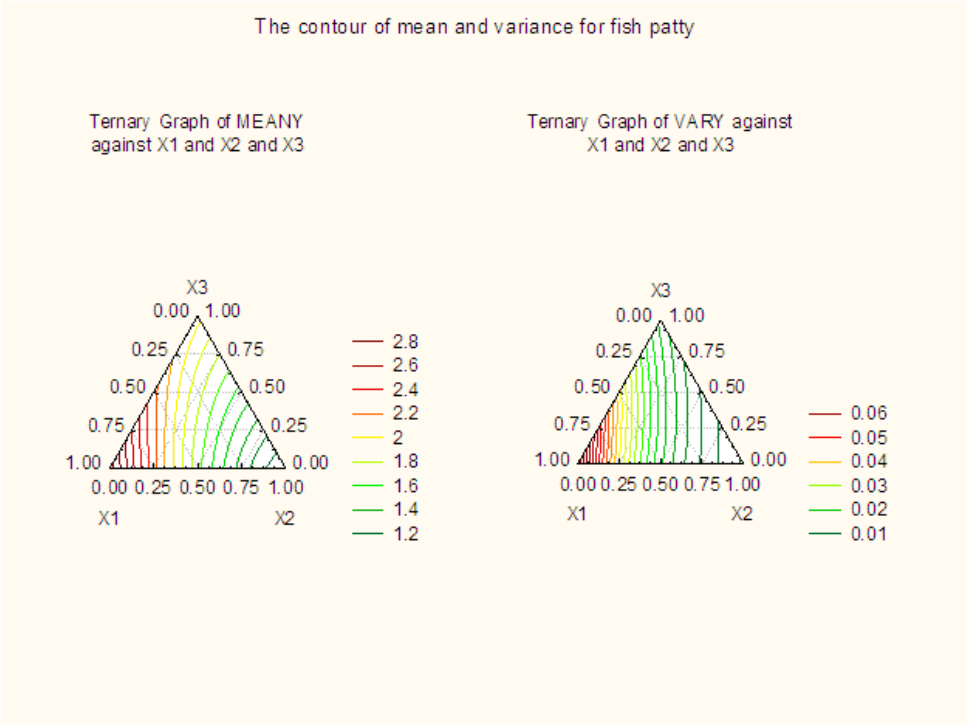

Figure 3.5: Fish Patty example: The contour plot of the mean and the variance of the fish patty model.

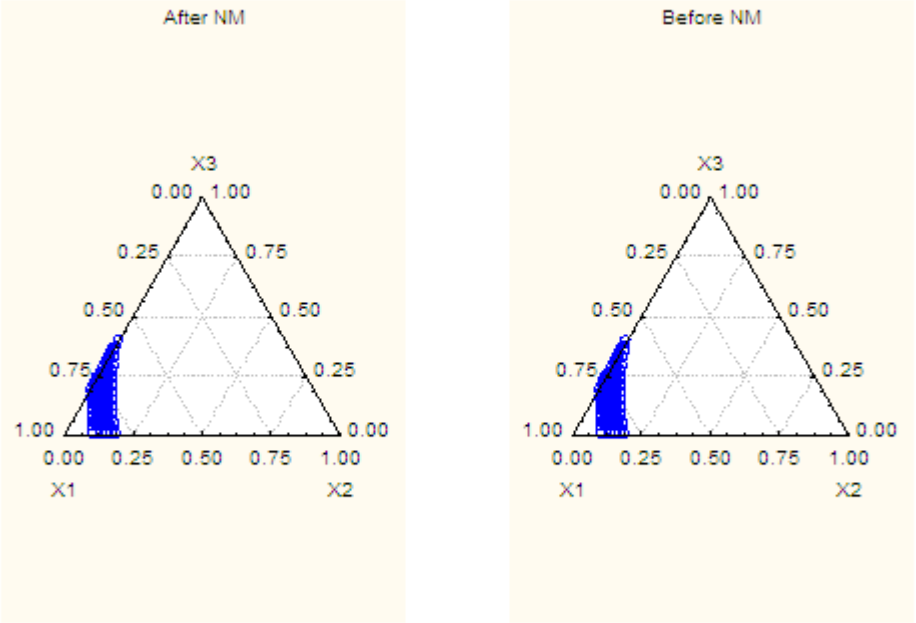

Figure 3.6: Fish Patty example: The general confidence region with weight ratio 1 : 1. Left is the confidence region after Nelder Mead refined search. The right one is the confidence region without Nelder Mead refined search.

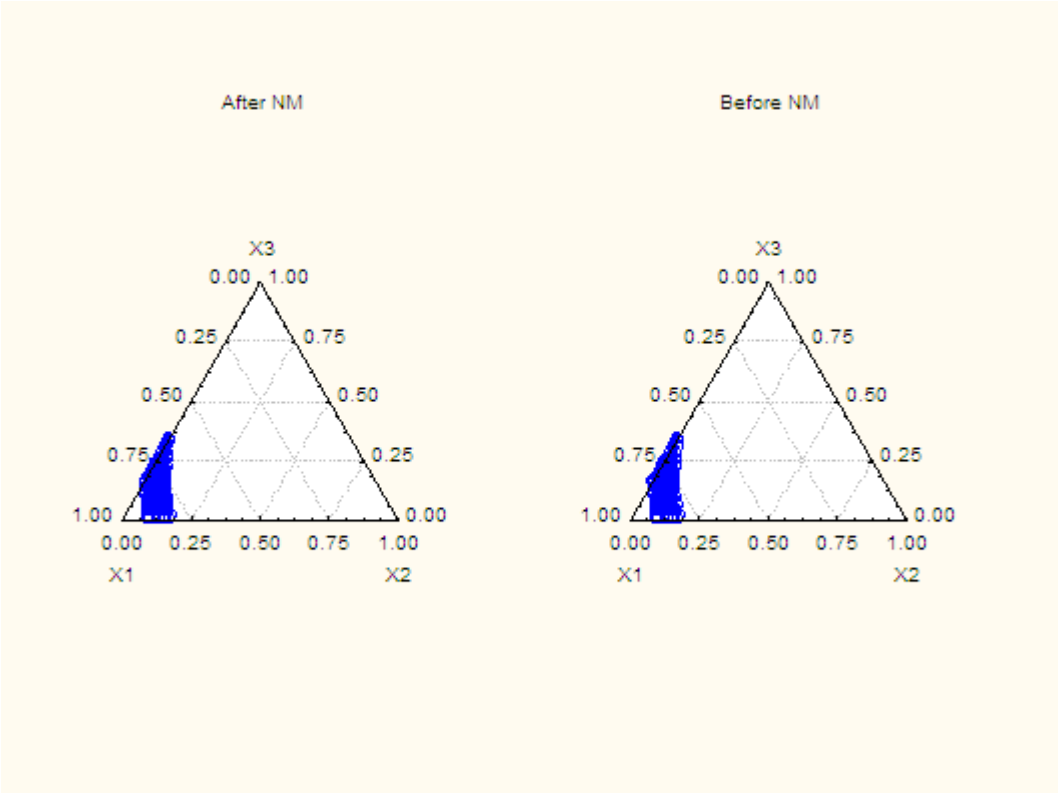

Figure 3.7: Fish Patty example: The general confidence region with weight ratio 100 : 1. Left is the confidence region after Nelder Mead refined search. The right one is the confidence region without Nelder Mead refined search.

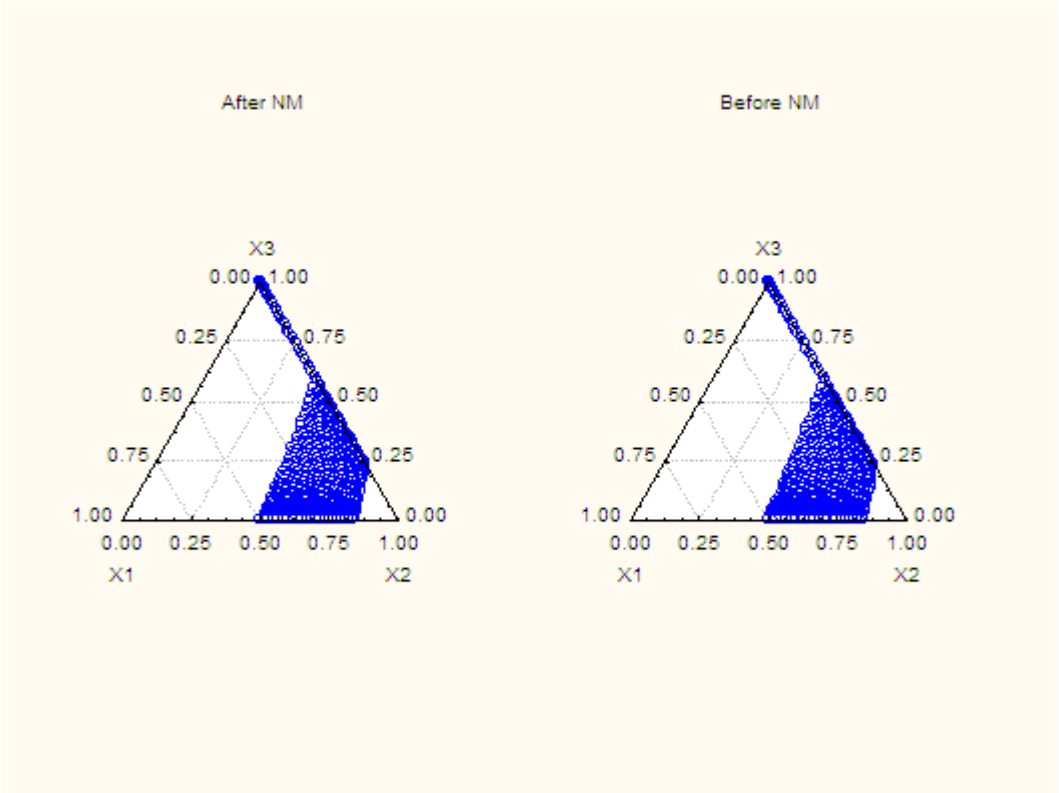

Figure 3.8: Fish Patty example: The general confidence region with weight ratio 1 : 100. Left is the confidence region after Nelder Mead refined search. The right one is the confidence region without Nelder Mead refined search.

grid levels are fine enough, the Nelder Mead refined search is not implemented. The efficiency of Corollary 3.1 was also tested with weight ratio 1 : 1. The cpu time using the corollary was 1 hours 41 minutes while it took 7 hours 18 minutes without the corollary.

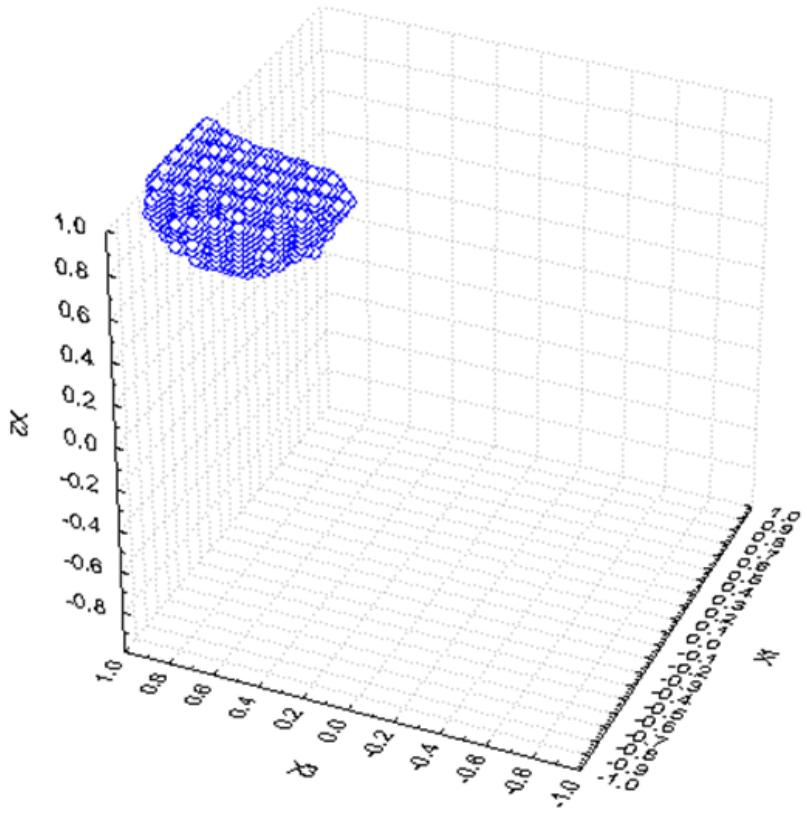

Figure 3.9: Three-control-variable example: The general confidence region with weight ratio 1 : 1.

3.4 Summary and Conclusion

The general confidence region in control variable settings to optimize WMSE is developed to address the practical need to optimize the response mean and

variance simultaneously. The response surface for the mean and variances may not always agree with each other. Unlike the current available techniques, this method provides flexibility to the practitioners in decision making at different situations by allowing the users to enter different weights for the mean and variance optimization. It can also be applied to more complicated studies like mixture experiments.

In addition, since the computation algorithm only requires the objective function, i.e. $WMSE$, to be the quadratic function of ϕ, the general confidence region can be applied directly to a more general model form as shown in the following:

$$y(x, z) = u(x)\beta + z'\gamma + t(x)'\Delta z + \epsilon \tag{3.12}$$

$t(x)$ is the vector of x terms that interact with the noise variables z. Note that the term $t(x)$ allows more complicated control and noise variable interaction forms like $x_1 x_2 z_1$.

CHAPTER 4

FUTURE RESEARCH: DUAL CONFIDENCE REGION

The zero-gradient confidence region and the general confidence region are useful in that they provide practitioners with the flexibility to choose the optimal operating conditions in control factors with desired confidence level. However, this does not mean that the resultant performance is the same as long as the control factors stay within their confidence region. Without further investigation, it is not clear how far the resultant response values are away from the target. Also, sometimes we get a wide confidence region in control factors, which could result from either a flat response surface or relatively high noise in the system. In the former case, a wide confidence region is good because it implies a robust region with almost equally good performance. In

the latter case, a wide confidence region implies poor quality of the system. We may want to improve the system to reduce the size of the confidence region. To differentiate between these two cases and get a better understanding of the response surface in the confidence region, a joint confidence region in both the control factors and the response is needed. We call this confidence region a dual confidence region. Section 4.1 shows that it is possible to apply the same computation used for general confidence region to solve for dual confidence region. A dual confidence region has never been developed before. It would have a significant impact if it can be calculated. Section 4.1 shows the formulation of the problem in the context of robust parameter design. If it works out, the same algorithm could be used for the dual confidence region without noise variables.

4.1 Formulation of the dual confidence region

Assuming that the model takes the following form:

$$y = u(x)'\beta + \gamma' z + t(x)'\Delta z + \varepsilon \tag{4.1}$$

Where x is the control variable vector, z is the noise variable vector. $u(x)'\beta$ is the mean model, and $t(x)'\Delta z$ is the noise model part. ε is the residual term. Here $u(x)$ and $t(x)$ are the vectors of the involved x terms for the mean model and the noise model part.

Now define $q(x, \phi)$ as the weighted sum of squares of errors (WMSE). i.e.

$$q(x, \phi) = \left(u(x)'\beta - T\right)^2 + \left(\gamma + \Delta' t(x)\right)' \left(\gamma + \Delta' t(x)\right) \sigma_z^2 \tag{4.2}$$

Where T is the target of the mean response and σ_z^2 is the variance of the noise variable. Let

$$\eta(\phi) = \min_w q(w, \phi) \tag{4.3}$$

Where ϕ is the vector of all the model parameters.

The $1-\alpha$ dual confidence region is defined as the collection of (x_0, y_0) such that the null hypothesis H_0 is not rejected at level α. H_0 is defined as the following:

$$H_0 : \begin{bmatrix} \eta\left(\phi\right) - q\left(x_0; \phi\right) \\ \\ u(x_0)'\beta - y_0 \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \tag{4.4}$$

This hypothesis is equivalent to the following:

$$H_0 : \quad \left(u(x_0)'\beta - y_0\right)^2 - \eta(\phi) + q(x_0, \phi) = 0 \tag{4.5}$$

Note that

$$\left(u(x_0)'\beta - y_0\right)^2 - \eta(\phi) + q(x_0, \phi) = \max_{w \in R} \left[\left(u(x)'\beta - y_0\right)^2 - \left(q(w, \phi) - q(x_0, \phi)\right)\right]$$

Therefore, (4.5) is also equivalent to:

$$H_0 : \quad \max_{w \in R} \left[\left(u(x_0)'\beta - y_0\right)^2 - q(w, \phi) + q(x_0, \phi)\right] = 0 \tag{4.6}$$

Let $Q(x_0, y_0, w, \phi) = (s(x_0)\beta - y_0)^2 - q(w, \phi) + q(x_0, \phi)$. It can be shown that $Q(x_0, y_0, w, \phi)$ is quadratic in ϕ. As shown in equation (3.3), $WMSE = q(x; \phi) = \lambda_1 T^2 + G_2(x)\phi + \phi' F_2(x)\phi$, where $G_2(x)$ and $F_2(x)$ are shown in Equation (3.4) and (3.5). Therefore, $q(x_0, \phi) - q(w, \phi)$ can be written as $G_2(w, x_0)\phi + \phi' F_2(w, x_0)\phi$, where $G_2(w, x_0) = G_2(x_0) - G_2(w)$ and $F_2(w, x_0) = F_2(x_0) - F_2(w)$. Now, let us express $(u(x_0)'\beta - y_0)^2$ in the same quadratic form:

$$(u(x_0)'\beta - y_0)^2 = y_0^2 + G_1(x_0, y_0)\phi + \phi' F_1(x_0)\phi \tag{4.7}$$

Where $G_1(x_0, y_0) = \left(-2y_0 u(x_0)', 0_{1\times r_n}\right)$, r_n is the number of model parameters for the noise variable part. $F_1(x_0) = \begin{bmatrix} u(x_0)u(x_0)' & 0_{r_c\times r_n} \\ 0_{r_n\times r_c} & 0_{r_n\times r_n} \end{bmatrix}$, r_c is the number of model parameters on the mean part. Now let $Q(x_0, y_0, w, \phi) = (u(x_0)\beta - y_0)^2 - q(w, \phi) + q(x_0, \phi)$. Then $Q(x_0, y_0, w, \phi)$ can be written as:

$$y_0^2 + [G_1(x_0, y_0) + G_2(w, x_0)]\phi + \phi'[F_1(x_0) + F_2(w, x_0)]\phi \tag{4.8}$$

Notice that $G_1(x_0, y_0)$ is the only quantity that is affected by y_0. Therefore, $Q(x_0, y_0, w, \phi)$ is a quadratic function of ϕ. The same computation strategy used for the general confidence region in Chapter 3 can be applied here. The $1 - \alpha$ dual confidence region can be computed using the min max approach as before. Ie. for each pair of (x_0, y_0), we simply need to check if the following interval (4.9) contains zero. If it does, then (x_0, y_0) cannot be rejected;

otherwise, exclude (x_0, y_0) from the confidence region.

$$\left\{ \min_{\phi \in C_\phi} \max_{w \in R} Q\left(x_0, y_0, w, \phi\right), \max_{\phi \in C_\phi} \max_{w \in R} Q\left(x_0, y_0, w, \phi\right) \right\} \tag{4.9}$$

Where C_ϕ is the confidence region of the model parameter vector ϕ. Since this is a confidence region for both x_0 and y_0 and the dimension of x_0 is k, the critical value for C_ϕ is chosen to be $(k+1)F(1-\alpha, k+1, \nu)$. Here ν is the residual degrees of freedom. Note that the upper bound is always bigger than or equal to zero. We just need to check the lower bound. By min-max theorem (Bazaraa, Sherali, and Shetty, 2nd Edition, page 240, problem 6.44), $\min_{\phi \in C_\phi} \max_{w \in R} Q\left(x_0, y_0, w, \phi\right) \geq \max_{w \in R} \min_{\phi \in C_\phi} Q\left(x_0, y_0, w, \phi\right)$. The latter is easier to compute. By replacing the lower bound of (4.9) by $\max_{w \in R} \min_{\phi \in C_\phi} Q\left(x_0, y_0, w, \phi\right)$, we make the interval more conservative if not equal.

The computation only needs to focus on the lower bound of (4.9), which can be replaced by $\max_{w \in R} \min_{\phi \in C_\phi} Q\left(x_0, y_0, w, \phi\right)$. Therefore, the computation has two levels: one at ϕ level, which can be done via Myers and Carters algorithm because $Q(x_0, y_0, w, \phi)$ is a quadratic function of ϕ. The other is at the w level. Strategically, we could first consider using the estimated optimal solution that minimizes the estimated WMSE to reject the points in the experimental region. Let $w* =$ the point that minimizes the estimated WMSE. For any fixed pair of (x_0, y_0), first compute $\min_{\phi \in C_\phi} Q\left(x_0, y_0, w^*, \phi\right)$. If $\min_{\phi \in C_\phi} Q\left(x_0, y_0, w^*, \phi\right) > 0$, then stop and reject (x_0, y_0) out of the confidence region. Otherwise, continue the checking by letting w be any other point in the experimental region.

Note that if there exists a w such that $\min_{\phi \in C_\phi} Q(x_0, y_0, w, \phi) > 0$, then $\max_{w \in R} \min_{\phi \in C_\phi} Q(x_0, y_0, w, \phi) > 0$ is guaranteed. No more searching for w is needed. In other words, as long as the determination of the sign of $\min_{\phi \in C_\phi} Q(x_0, y_0, w, \phi)$ is right, then once x is rejected by some w, it is rejected for sure. No more refined search is needed. We only need to refine the search on the x points that are not rejected. Refined search strategy includes finer grids or/and Nelder Mead procedure. Note that if the real $\min_{\phi \in C_\phi} Q(x_0, y_0, w, \phi)$ is not reached, then the points could be rejected early. If the maximum at w level is not reached, then the points could be falsely accepted, which will lead to wider confidence region. Therefore, the optimization at both levels are important.

For dual confidence region, the maximization at w level can be done in the same way as for the general confidence region. However, the minimization at ϕ level will be more complicated than that for the general confidence region because the points in the dual confidence region are determined by both x and y. Also, how to grid the response value y in the experimental region needs more research as well.

REFERENCES

Bazaraa, M.S., Sherali, H.D., Shetty, C.M. (1993). *Nonlinear Programming Theory and Algorithms, second edition*, John Wiley & Sons, Inc.

Borror, C., Montgomery, D. (2002). Evaluation of Statistical Designs for Experiments Involving Noise Variables. *Journal of Quality Technology 34*, 54-70.

Box, G. E. P. and Jones, S. (1990). *Design Products That are Robust to Environment. Report No. 56*, Center for Quality and Productivity Improvement, University of Wisconsin, Madison, WI.

Cahya, S. del Castillo, E., Peterson, J.J. (2004). Computation of Confidence Regions for Optimal Factor Levels in Constrained Response Surface Problems. *Journal of Computational and Graphical Statistics, Volume 13, Number 2*, Pages 1-20

Cheng, A., Peterson, J.J., Chitturi, P., (2011). A Confidence Region for Zero-Gradient Solutions for Robust Parameter Design Experiments. *International Journal of Quality, Statistics, and Reliability, Volume 2011, Article ID 537543*, 11 pages

Chiao,C. Hamada, M. (2001). Analyzing Experiments with Correlated ultiple Responses. *Journal of Quality Technology, 33*, 451–465.

Cornell, J.A.(1990), *Experiments with mixtures, Designs, Models, and The Analysis of Mixture Data, 2nd edition*, John Wiley & Sons.

Fujikoshi, Y., Ulyanov, V., Shimizu, R. (2010). *Multivariate Statistics, High-Dimensional and Large-Sample Approximations*, John Wiley and Sons, Inc.

ICH Q8 (R2). (2009). *Pharmaceutical Development*, International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use

Kotz, S. and Johnson, N, L. (1982). Multivariate t-Distribution. *Encyclopedia of Statistical Sciences, 6*, 129–130, John Wiley and Sons, Inc., Hoboken, NJ.

Kuhn, Andrew M.(2003). Optimizing Response Surface Experiments with Noise Factors Using Confidence Regions. *Quality Engineering, 15*, 3, 419–426

Lucas, J. M. (1994). Achieving a Robust Process Using Response Surface Methodology. *Journal of Quality Technology, 26*, 248–260.

Miller, R. G. (1981). *Simultaneous Statistical Inference, 2nd edition*, Springer-Verlag New York Inc.

Miro-quesada, G., del Castillo, E., Peterson, J.J. (2004). A Bayesian Approach for Multiple Response Surface Optimization in the Presence of Noise Variables. *Journal of Applied Statistics Vol. 31, No. 3*, 251-270.

Montgomery, D. C. (1990-1991). Using Fractional Factorial Designs for Robust Process Development. *Quality Engineering, 3*, 193–205.

Montgomery, D. C. (2009). *Design and Analysis of Experiments, 7th ed.*, John Wiley and Sons, Inc.

Myers, R. H., Walter H. Carter Jr. (1973). Response Surface Tech-

niques for Dual Response Systems, *Technometrics, Vol. 15, No. 2*, pp. 301–317.

Myers, R. H., Khuri, A.I. , and Vining, G. G. (1992). Response Surface Alternatives to the Taguchi Robust Parameter Design Approach. *The American Statistician, 46*, 131–139.

Myers R. H., Kim Y, Griffiths K. L. (1997). Response surface methods and the use of noise variables. *Journal of Quality Technology, 29*, 429-440.

Myers, R. H., Montgomery, C. M., Anderson-Cook, C. M. (2009). *Response Surfaces Methodology, Process and Product Optimization Using Designed Experiments*, Wiley; New York

Nair, VN. (1992). Taguchi's parameter design: A panel discussion. *Technometrics, Vol 34, No 2*, 127–161.

Peterson, J.J. (1987). Computing Asymptotic Confidence Bands for Nonlinear Models. *Transactions of the Fourth Army Conference on Applied Mathematics and Computing, ARO Report 87-1, Ithaca, NY*, Cornell University, pp. 787–795

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

Peterson, J. J., Cahya, S., and del Castillo, E. (2002). A General Approach to Confidence Regions for Optimal Factor Levels of Response Surfaces. *Biometrics, 58*, 422–431

Peterson, J. J. , Kuhn, A. M. (2005). Ridge Analysis With Noise Variables. *Technometrics, August 2005, Vol 47, No 3*, page 277

Rao,C. R. , (1973). *Linear Statistical Inference and its application, 2nd edition*, p7, John Wiley & Sons, Inc.

Robinson, T., Borror, C., Myers, R.H., (2004). Robust Parameter Design: A Review. *Quality and Reliability Engineering International, 20*,81–101,

Shoemaker, A. C.; Tsui, K.-L.; and Wu, C. F. J. (1991). Economical Experimentation Methods for Robust Design. *Technometrics, 33*, 415–427.

Spjøtvoll, E. (1972). Multiple Comparison of Regression Functions. *The Annals of Statistics, 43*, 1076-1088.

Steiner, S. H., Hamada, M. (1997). Making mixtures robust to Noise and Mixing measurement Errors. *Journal of Quality Technology 1997, Vol. 29, No.4*

Taguchi G. (1986). *Introduction to Quality Engineering*, U-NIPUB/Kraus International: White Plains, NY.

Timm, N. H. (2002). *Applied Multivariate Analysis*, Springer-Verlag New York, N.Y.

Welch, W. J., Yu, T. K., Kang, S. M., and Sacks, J. (1990). Computer Experiments for Quality Control by Parameter Design. *Journal of Quality Technology, 22*, 15–22.

APPENDIX A

THE PROOFS FOR SOME OF THE THEOREMS

A.1 Proof for Theorem 2.1

The proof starts with a special case where we assume the intercept $\gamma = 0$ in the hypothesis: $\gamma + \Delta' x = 0$. Then we will generalize the result to the nonzero intercept case ($\gamma \neq 0$), which is what we want.

Part 1: zero-intercept case

When $h = 1$ and $k > h$, the equation $M(x)\psi = 0$ has more than one solution. The dimension of the solution to this equation is $d = k - h = k - 1$ (Rao,C. R. 1973). Since $\gamma = 0$ by assumption, M(x) can be written as $M(x) = w'L$, where $L_{d\times k}$ is a matrix with d basis vectors as its rows, w is a $d \times 1$ vector

that has one-to-one relationship to each solution x, and the k-th element of x can be expressed as a linear combination of the rest of the $k-1$ elements:

$$x_k = a_1 x_1 + a_2 x_2 + \cdots + a_{k-1} x_{k-1}$$

w and L can be expressed as follows:

$$w' = (x_1, x_2, \cdots, x_{k-1})$$

$$L_{d\times k} = \begin{pmatrix} 1 & 0 & \dots & 0 & a_1 \\ 0 & 1 & \dots & 0 & a_2 \\ & & \vdots & & \\ 0 & 0 & \dots & 1 & a_{k-1} \end{pmatrix}$$

Since $M(x)\psi = \Delta' x = \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k = 0$ and $x_k = -\frac{\delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{k-1} x_{k-1}}{\delta_k}$, then,

$$a_1 = -\delta_1/\delta_k$$

$$a_2 = -\delta_2/\delta_k$$

$$\vdots$$

$$a_{k-1} = -\delta_{k-1}/\delta_k$$

Hence, $L_{d\times(k)}$ depends on the true parameter vector ψ. Also, under the null

hypothesis: $M(x)\psi = 0$, the test statistic $Q_{\mathcal{L}}$ can be written as:

$$\begin{aligned} Q_{\mathcal{L}} &= \max_{x:M(x)\psi=0} (M(x)\hat{\psi})'(\widehat{Var(M(x)\hat{\psi})})^{-1}(M(x)\hat{\psi}) \\ &= \max_{x:x\psi=0} (x'\hat{\psi})'(\widehat{Var(x'\hat{\psi})})^{-1}(x'\hat{\psi}) \\ &= \max_{x:x\psi=0} (x'(\hat{\psi}-\psi))'(\widehat{Var(x'\hat{\psi})})^{-1}(x'(\hat{\psi}-\psi)) \\ &= \max_{x:x\psi=0} \frac{(x'(\hat{\psi}-\psi))^2}{\widehat{Var(x'\hat{\psi})}} \end{aligned} \tag{A.1}$$

where $x = w'L$, which is actually a d-dimensional vector in an k-dimensional space R. The last equal sign holds because both $(x'(\hat{\psi}-\psi))$ and $\widehat{Var(x'\hat{\psi})}$ are scalars. Based on Theorem 1 on page 65 of Miller's Simultaneous Statistical Inference (Miller, 1981, 2nd edition),

$$P\left\{\left|x'(\hat{\psi}-\psi)\right| < (dF^{\alpha}_{d,v})^{\frac{1}{2}}\widehat{Var(x'\hat{\psi})}^{\frac{1}{2}}, \forall x \in R\right\} = 1-\alpha$$

then,

$$P\left\{\frac{(x'(\hat{\psi}-\psi))^2}{\widehat{Var(x'\hat{\psi})}} < dF^{\alpha}_{d,v}, \forall x \in R\right\} = 1-\alpha$$

and

$$P\left\{\max_{x:x\psi=0}\frac{(x'(\hat{\psi}-\psi))^2}{\widehat{Var(x'\hat{\psi})}} < dF^{\alpha}_{d,v}\right\} = 1-\alpha$$

Hence, $Q_{\mathcal{L}}$ follows $dF(d,v)$.

Part 2: Intercept-case

The intercept case can be proved in the following way as implied by Miller (Miller, 1981, page 113).

When intercept term is significant, $\psi' = (\delta_0, \delta_1, \delta_2, \cdots, \delta_p)$, where δ_0 is the intercept. $w' = (1, x_1, x_2, \cdots, x_{k-1})$. We can convert the intercept case to no-intercept case by multiplying w by w_0^* assuming that $w_0^* \neq 0$. In other words, if we let $w^{*'} = (w_0^*, w_1^*, \cdots, w_{k-1}^*)$, where $w_i^* = x_i w_0^*$, then there is a $1-1$ correspondence between w^* and w. The dimension for w^* is $d+1 = k$, where $d = k - h = k - 1$ Hence,

$$\begin{aligned} \max_{w \in R^{d-1}, w_0 = 1} \frac{(w'L(\hat{\psi} - \psi))^2}{s^2(w'L\Omega L'w)} &= \max_{w \in R^{d-1}, w_0 = 1, w_0^* \neq 0} \frac{(w_0^* w'L(\hat{\psi} - \psi))^2}{s^2(w_0^* w'L\Omega L'w w_0^*)} \\ &= \sup_{w^* \in R^d, w_0^* \neq 0} \frac{(w^{*'}L(\hat{\psi} - \psi))^2}{s^2(w^{*'}L\Omega L'w^*)} \end{aligned} \tag{A.2}$$

Equation (A.2) becomes into a no-intercept case. Therefore, the critical value is $(d+1)F(1-\alpha, (d+1), v)$ for $w_0^* \neq 0$.

A.2 Proof for Theorem 2.2 and Corollary 2.1

Part 1: For no-intercept case (i.e. $\gamma = 0$):

Let s^2 be the sample estimate of σ^2, then the test statistic $Q_{\mathcal{L}}$ can be expressed as:

$$
\begin{aligned}
Q_{\mathcal{L}} &= \max_{x:M(x)\psi=0} (M(x)(\hat{\psi}-\psi))'(\widehat{Var(M(x)\hat{\psi})})^{-1}(M(x)(\hat{\psi}-\psi)) \\
&= \max_{x:M(x)\psi=0} (M(x)(\hat{\psi}-\psi))'(s^2 M(x)\Omega M(x)')^{-1}(M(x)(\hat{\psi}-\psi)) \\
&= \max_{x:M(x)\psi=0} (M(x)\frac{\hat{\psi}-\psi}{\sigma})'((I_h \otimes (w'L))\Omega(I_h \otimes (w'L))')^{-1}(M(x)\frac{\hat{\psi}-\psi}{\sigma})\frac{\sigma^2}{s^2} \\
&= \max_{M(x)\psi=0} (M(x)z)'((I_h \otimes (w'L))\Omega(I_h \otimes (w'L))')^{-1}(M(x)z)\frac{\sigma^2}{s^2} \\
&= \max_{M(x)\psi=0} (Zx)'((I_h \otimes (w'L))\Omega(I_h \otimes (w'L))')^{-1}(Zx)\frac{\sigma^2}{s^2} \\
&= \max_{w:w\in R} w'LZ'((I_h \otimes (w'L))\Omega(I_h \otimes (w'L))')^{-1}(ZL'w)\frac{\sigma^2}{s^2}
\end{aligned}
$$

where R is the d dimensional real space. $z_{p\times 1} = \frac{\hat{\psi}-\psi}{\sigma} = (z_{11}z_{12}\cdots z_{1k}\cdots z_{h1}\cdots z_{hk})'$, $z_{ij} = \frac{\hat{\delta_{ij}}-\delta_{ij}}{\sigma}$, $z \sim N(0,\Omega)$. Z is a $h\times k$ matrix. The relationship between vector z and matrix Z is that $vec(Z') = z$. i.e.

$$
Z' = \begin{pmatrix}
z_{11} & z_{21} & \cdots & z_{h1} \\
z_{12} & z_{22} & \cdots & z_{h2} \\
 & \vdots & & \\
z_{1k} & z_{2k} & \cdots & z_{hk}
\end{pmatrix} \tag{A.3}
$$

By assumption $\Omega_{p\times p} = V_1 \otimes V_2$, where V_1 is a $h \times h$ covariance matrix, and V_2 is a $k \times k$ covariance matrix when the intercept is zero. Then $Z' \sim MN(0_{k\times h}, V_1, V_2)$, where $MN(0_{k\times h}, V_1, V_2)$ stands for the matrix normal distribution (See the definition and the property of matrix normal distribu-

tion in Timm, 2002, page 90-91) with mean $0_{k\times h}$ and two covariance matrices: V_1 and V_2.

Now the test statistic $Q_{\mathcal{L}}$ above can be written as:

$$\begin{aligned} Q_{\mathcal{L}} &= \frac{v}{U}\max_{w:w\in R} w'LZ'((I_h\otimes(w'L))(V_1\otimes V_2)(I_h\otimes(w'L))')^{-1}(ZL'w) \\ &= \frac{v}{U}\max_{w:w\in R} w'LZ'(V_1\otimes(w'LV_2L'w))^{-1}(ZL'w) \\ &= \frac{v}{U}\max_{w:w\in R}\frac{w'LZ'V_1^{-1}ZL'w}{w'LV_2L'w} \\ &= \frac{v}{U}\lambda_{max}(D) \end{aligned}$$

where $D=(LV_2L')^{-1}LZ'V_1^{-1}ZL'$, v is the degrees of freedom for error, $U=\frac{vs^2}{\sigma^2}\sim\chi^2(v)$. (For a proof of last equality see the result in Problem 22.1 in Rao 1973, p74.)

Note that $\lambda_{max}(D)=\lambda_{max}(A)$, where $A=(LV_2L')^{-\frac{1}{2}}LZ'V_1^{-1}ZL'(LV_2L')^{-\frac{1}{2}}$ This is true because of the following:

By the definition of eigenvalue,

$$|D-\lambda_{max}I_d|=0 \tag{A.4}$$

Note LV_2L' is positive definite and symmetric. Hence,$(LV_2L')^{\frac{1}{2}}$ is positive definite and symmetric, and $(LV_2L')^{-1}$ is positive definite and symmetric as well. Therefore, $(LV_2L')^{-1}=(LV_2L')^{-\frac{1}{2}}(LV_2L')^{-\frac{1}{2}}$. Now multiply both sides of equation A.4 by $\left|(LV_2L')^{+\frac{1}{2}}\right|$ from left and then multiply by $\left|(LV_2L')^{-\frac{1}{2}}\right|$ from right, we get

$$\left|(LV_2L')^{-\frac{1}{2}}(LZ'V_1^{-1}ZL')(LV_2L')^{-\frac{1}{2}}-\lambda_{max}I_d\right|=0$$

Note that $A = (LV_2L')^{\frac{-1}{2}}(LZ'V_1^{-1}ZL')(LV_2L')^{\frac{-1}{2}}$. Hence, $\lambda_{max}(A) = \lambda_{max}(D)$. $A \sim Wishart(I_d, d, h)$ as proved below. Therefore, the distribution of $Q_{\mathcal{L}}$ can be obtained by simple Monte Carlo simulation from Wishart and χ^2 distribution.

To prove $A \sim Wishart(I_d, d, h)$, first prove $Z'V_1^{-1}Z \sim Wishart(V_2, k, h)$.

Since V_1^{-1} is a covariance matrix, V_1^{-1} is symmetric, nonsingular, and positive definite and $V_1^{-1} = V_1^{-\frac{1}{2}}V_1^{-\frac{1}{2}}$, where $V_1^{-\frac{1}{2}}$ is also symmetric. Then $(Z'V_1^{-1}Z) = Z'V_1^{-\frac{1}{2}}V_1^{-\frac{1}{2}}Z = Y'Y$, where $Y_{h\times k} = V_1^{-\frac{1}{2}}Z$.

Next, prove each row of Y, $Y_{(i)} \sim N(0, V_2)$ and $Y_{(i)}$ and $Y_{(j)}$ are independent, $\forall i \neq j$. This can be proved by proving $vec(Y') \sim N(0, I_h \otimes V_2)$. Let

$$V_1^{-\frac{1}{2}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1h} \\ a_{21} & a_{22} & \dots & a_{2h} \\ & \vdots & & \\ a_{h1} & a_{h2} & \dots & a_{hh} \end{pmatrix}$$

Then

$$Y = V_1^{-\frac{1}{2}}Z' = \begin{pmatrix} \sum_1^h a_{1i}z_{i1} & \sum_1^h a_{1i}z_{i2} & \dots & \sum_1^h a_{1i}z_{ik} \\ \sum_1^h a_{2i}z_{i1} & \sum_1^h a_{2i}z_{i2} & \dots & \sum_1^h a_{2i}z_{ik} \\ & \vdots & & \\ \sum_1^h a_{hi}z_{i1} & \sum_1^h a_{hi}z_{i2} & \dots & \sum_1^h a_{hi}z_{ik} \end{pmatrix}$$

Therefore,

$$vec(Y') = \begin{pmatrix} \sum_1^h a_{1i} z_{i1} \\ \sum_1^h a_{1i} z_{i2} \\ \vdots \\ \sum_1^h a_{1i} z_{ik} \\ \vdots \\ \vdots \\ \sum_1^h a_{hi} z_{i1} \\ \sum_1^h a_{hi} z_{i2} \\ \vdots \\ \sum_1^h a_{hi} z_{ik} \end{pmatrix}$$

that is, $vec(Y') = V_1^{-\frac{1}{2}} \otimes I_k z$. Since $z \sim N(0, V_1 \otimes V_2)$, $vec(Y') \sim N(0, (V_1^{-\frac{1}{2}} \otimes I_k)(V_1 \otimes V_2)(V_1^{-\frac{1}{2}} \otimes I_k)')$. Hence, $vec(Y') \sim N(0, I_h \otimes V_2)$. In other words, the variance of $vec(Y')$ can be written as:

$$Var(vec(Y')) = \begin{pmatrix} V_2 & 0 & \dots & 0 \\ 0 & V_2 & \dots & 0 \\ & \vdots & & \\ 0 & 0 & \dots & V_2 \end{pmatrix}$$

Hence, $Y_{(i)} \sim N(0, V_2)$ and $Y_{(i)}$ and $Y_{(j)}$ are independent, $\forall i \neq j$ because the covariance of $Y_{(i)}$ and $Y_{(j)}$ is zero $\forall i \neq j$.

By the definition of Wishart distribution (Fujikoshi, Y., et al, 2010), $Y'Y \sim$

$Wishart(V_2, k, h)$. Then $LZ'V_1^{-1}ZL' = LY'YL' \sim Wishart(LV_2L', k, h)$. Hence, $A \sim Wishart((LV_2L')^{-\frac{1}{2}}LV_2L'(LV_2L')^{-\frac{1}{2}}, d, h)$. i.e. $A \sim Wishart(I_d, d, h)$.

Part 2: Intercept case (i.e. none of the main effects of noise variables is zero)

Follows the same logic as in the proof for Theorem 2.1.

Next prove Corollary 2.1

If the design is orthogonal, then $\Omega = cI$, where c is a constant, I is the $p \times p$ identity matrix. And cI can be written as $cI = \sqrt{c}I_h \otimes \sqrt{c}I_{k+1}$. Hence, the assumtion $\Omega = V_1 \otimes V_2$ is satisfied and Theorem 2.2 holds for this situation.

A.3 Proof of (2.6)

Note that,

$$\max_{\psi \in C_\psi(b_\alpha)} \max_{x:M(x)\psi=0} Q(x;\hat{\Psi}) = \max_{\{(x;\psi):M(x)\psi=0,\psi\in C_\psi(b_\alpha)\}} Q(x;\hat{\Psi}) \tag{A.5}$$

Adapting the proof of Theorem 2.1 in Peterson, Cahya, and del Castillo (2002), it follows that for any critical value, b_α,

$$C_x(b_\alpha) = \{x : M(x)\psi = 0, \psi \in C_\psi(b_\alpha)\} = \{x : Q(x;\hat{\psi}) \leq b_\alpha\}$$

Since $Q(x;\hat{\psi})$ is not a function of ψ , it follows directly that

$$\max_{\{(x;\psi):M(x)\psi=0,\psi\in C_\psi(b_\alpha)\}} Q(x;\hat{\Psi}) = \max_{x\in C_x(b_\alpha)} Q(x;\hat{\Psi}) \tag{A.6}$$

So (2.6) follows directly from (A.5) and (A.6).

A.4 Proof for $C_x(c_\alpha) = C_{\mathcal{L}}(c_\alpha)$

Recall that $C_x(c_\alpha) = \{x : Q(x;\hat{\psi}) \leq c_\alpha\}$ and $C_{\mathcal{L}}(c_\alpha) = \{\mathcal{L} : Q(x;\hat{\psi}) \leq c_\alpha, \text{for all } x \in \mathcal{L}\}$. By definition, $x \in C_{\mathcal{L}}(c_\alpha)$ implies that $x \in C_x(c_\alpha)$. Next, we will show that if $x \in C_x(c_\alpha)$, then $x \in C_{\mathcal{L}}(c_\alpha)$. This can be proved by contradiction. Suppose that there exist some x's such that $x \in C_x(c_\alpha)$, but $x \notin C_{\mathcal{L}}(c_\alpha)$. Then there exists at least one x point, say x^*, in $C_x(c_\alpha)$ such that there is no linear subspace that satisfies two conditions: 1) it contains x^*; 2) for every point x in this subspace $Q(x;\hat{\psi}) \leq c_\alpha$. Define $C_\psi(c_\alpha) = \{\psi : (\psi - \hat{\psi})'\hat{V}_{\hat{\psi}}(\psi - \hat{\psi}) \leq c_\alpha\}$ and consider the set, $C_x^* = \{x : M(x)\psi = 0, \psi \in C_\psi(c_\alpha)\}$. It follows using a proof analogous to that in Theorem 2.1 in Peterson, Cahya, and del Castillo (2002) that $C_x(c_\alpha) = C_x^*$. Therefore, $x^* \in C_x^*$. This then implies that there exists a $\psi^* \in C_\psi(c_\alpha)$ such that $M(x^*)\psi^* = 0$. If $h < k$, then there exists a line or hyperplane $\mathcal{L}^*$ such that all the x's that satisfy $M(x)\psi^* = 0$ are in $\mathcal{L}^*$. Hence, $x^* \in \mathcal{L}^*$. Again using the proof of Theorem 2.1 in Peterson et al. (2002), it follows that for any given point x, $Q(x;\hat{\psi}) \leq c_\alpha$ if and only if $M(x)\psi = 0, \text{for some } \psi \in C_\psi(c_\alpha)$. Therefore $Q(x;\hat{\psi}) \leq c_\alpha, \text{for all } x \in \mathcal{L}^*$. In other words, $\mathcal{L}^*$ satisfies the above two conditions, which is a contradiction.

A.5 Proof for Theorem 3.3

There are two ways to prove this theorem. The first one used the convex theorem from Bazarra and Shetty's book. The second used canonical analysis.

1. If $F(w,x)$ is negative semi-definite, then $-F(w,x)$ is positive semi-definite. By Bazarra and Shetty (1993, p91, Theorem 3.3.7), $-q(w,x;\phi) = \phi^{'}(-F(w,x))\phi - G(w,x)\phi$ is convex. Since ϕ_0 is the stationary point for $q(w,x;\phi)$ by assumption, it is also the stationary point for $-q(w,x;\phi)$. By Bazarra and Shetty (1993, Theorem 3.4.2, p101), ϕ_0 is also the global minimum for the convex function $-q(w,x;\phi)$. i.e. $-q(w,x;\phi_0) \leq -q(w,x;\phi), \phi \in C_\phi$. Therefore, $q(w,x;\phi_0) \geq q(w,x;\phi), \phi \in C_\phi$. i.e. $\phi_0 = \phi_{max} = \phi^*$.

2. This result can also be proved via canonical analysis as shown by Myers, Montgomery, and Anderson-Cook(2009). Note $q(w,x;\phi_u) = q(w,x;\phi_0) + \sum_1^n \lambda_i u_i^2$, where λ_i is the eigenvalue of $F(w,x)$, n is the total number of eigenvalues of $F(w,x)$. And u is a vector of canonical variables. $u = P^{'}(\phi_u - \phi_0)$, where P is the matrix whose columns are the normalized eigenvectors associated with eigenvalues of $F(w,x)$. (see page 224 of Myers, Montgomery, and Anderson-Cook, Response Surface Methodology, Process and Product Optimization Using Designed Experiments, third edition, 2009). Note that vector u is simply a transformation of ϕ_u.

u and ϕ_u actually represent the same point in C_ϕ. The stationary point is the maximum since the eigenvalues are either negatives or zeros due to the assumption that $F(w,x)$ is negative semi-definite (n.s.d). Since $\phi_0 \in C_\phi$, then $\phi_0 = \phi_{max} = \phi^*$. Note that if $F(w,x)$ is negative semi-definite (n.s.d), then the stationary point is not unique. The stationary points would then form a stationary ridge, which included both interior points and boundary points.

A.6 Proof for Theorem 3.4

The proof is done via contradiction through the following cases using the fact that $q(w,x;\phi_u) = q(w,x;\phi_0) + \sum_1^n \lambda_i u_i^2$, where u is a vector of canonical variables. $u = P^{'}(\phi_u - \phi_0)$.

Case 1 Assuming the stationary point $\phi_0 \notin C_\phi$ but $F(w,x)$ is negative semi-definite (n.s.d). Since $q(w,x;\phi_u) = q(w,x;\phi_0) + \sum_1^n \lambda_i u_i^2$ and $F(w,x)$ is negative semi-definite, the eigenvalues are either negatives or zeros. Therefore, $q(w,x;\phi_{max}) = q(w,x;\phi_0)$ where the corresponding $u_{max} = 0$. $u_{max} = P^{'}(\phi_{max} - \phi_0)$. Suppose that the true maximum in C_ϕ is ϕ^*, and ϕ^* is an interior point. Then $q(w,x,\phi^*) = q(w,x,\phi_0) + \sum_{i=1}^n \lambda_i u_i^{*2} \geq q(w,x;\phi), \phi \in C_\phi$ since ϕ^* is the maximum in C_ϕ. However, since point u^* is inside C_ϕ and $q(w,x;\phi)$ is continuous, there always exists another

point y, where $y = P^{'}(\phi_y - \phi_0)$ and $\phi_y \in C_\phi$, such that one of its elements, say y_1, has smaller absolute value than the absolute value of u_1^* while all other elements are the same. Since all the λ_is are negative, this would lead to a $q(w, x; \phi_y) = q(w, x; \phi_0) + \sum_{i=1}^{n} \lambda_i y_i^2$ value bigger than $q(w, x; \phi^*)$, which is a contradiction. Therefore, ϕ^* has to be on the boundary of C_ϕ.

Case 2 Assuming $\phi_0 \notin C_\phi$ but $F(w, x)$ is positive semi-definite (p.s.d). When $F(w, x)$ is positive semi-definite (p.s.d), there is no global maximum. Instead, ϕ_0 is a global minimum. Suppose that the true maximum in C_ϕ is ϕ^*, and ϕ^* is an interior point. $q(w, x; \phi^*) = q(w, x; \phi_0) + \sum_{i=1}^{n} \lambda_i u_i^{*2} \geq q(w, x; \phi), \phi \in C_\phi$. Since point u^* is inside C_ϕ and $q(w, x; \phi)$ is continuous, we can always find a point, say, ϕ_y in C_ϕ such that $q(w, x; \phi_y) > q(w, x; \phi^*)$. For instance, w.l.g. let $\lambda_1 > 0$. There always exists another point y, where $y = P^{'}(\phi_y - \phi_0)$, such that one of its element, say y_1, has bigger absolute value than the absolute value of u_1^* while all other elements are the same. This would lead to a $q(w, x; \phi_y)$ value bigger than $q(w, x; \phi^*)$, which is a contradiction. Therefore, ϕ^* has to be on the boundary of C_ϕ.

Case 3 Assuming $\phi_0 \in C_\phi$ but $F(w, x)$ is positive semi-definite (p.s.d). The proof for this case is same as Case 2.

Case 4 Assuming $F(w, x)$ is indefinite. If $F(w, x)$ is indefinite, then the stationary point ϕ_0 is a saddle point (Myers, Montgomery, and Anderson-Cook, 2009). Therefore the maximum is NOT ϕ_0. Let ϕ^* b be the true maximum in C_ϕ , and ϕ^* is an interior point. Since $q(w, x; \phi^*) = q(w, x; \phi_0) + \sum_1^n \lambda_i u_i^{*2} \geq q(w, x; \phi), \phi \in C_\phi$. Without loss of generality, let $\lambda_1 > 0$, then we can always find a point y in C_ϕ, where $y = P^{'}(\phi_y - \phi_0)$. such that y_1 has a bigger absolute value than the absolute value of u_1^* while all other elements are the same since u is inside C_ϕ. This would lead to a $q(w, x; \phi_y) = q(w, x; \phi_0) + \sum_1^n \lambda_i y_i^2$ value bigger than $q(w, x; \phi^*)$, which is a contradiction Therefore, ϕ^* has to be on the boundary.

APPENDIX B

THE LONG TABLES

B.1 The Critical Values Based on Corollary 2.1

Table B.1: Monte Carlo simulation of the critical values based on Corollary 2.1

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
2	1	1	99.04	18.52	8.58
3	1	1	35.27	10.31	5.58
4	1	1	20.86	7.68	4.53

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
5	1	1	16.14	6.63	4.08
6	1	1	13.88	6.07	3.83
7	1	1	12.27	5.60	3.59
8	1	1	11.09	5.29	3.42
9	1	1	10.62	5.15	3.38
10	1	1	10.30	4.85	3.19
20	1	1	7.75	4.35	2.97
30	1	1	7.32	4.10	2.86
40	1	1	7.08	4.05	2.84
50	1	1	7.15	4.09	2.82
60	1	1	7.11	3.99	2.81
70	1	1	6.97	3.92	2.78
80	1	1	6.94	3.97	2.79
90	1	1	7.06	3.96	2.76
100	1	1	7.10	4.02	2.78
2	1	2	194.0	37.21	17.98
3	1	2	64.02	19.38	11.14

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
4	1	2	36.49	13.89	8.66
5	1	2	26.60	11.60	7.56
6	1	2	21.97	10.16	6.90
7	1	2	19.01	9.49	6.52
8	1	2	17.24	8.94	6.22
9	1	2	15.76	8.47	6.01
10	1	2	15.78	8.44	5.92
20	1	2	11.70	7.04	5.23
30	1	2	10.98	6.66	4.96
40	1	2	10.40	6.36	4.84
50	1	2	9.75	6.35	4.82
60	1	2	9.82	6.21	4.80
70	1	2	9.66	6.35	4.90
80	1	2	9.71	6.23	4.81
90	1	2	9.80	6.20	4.74
100	1	2	9.50	6.12	4.70
2	1	3	298.5	56.76	27.27

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
3	1	3	88.21	27.85	16.10
4	1	3	51.63	19.88	12.63
5	1	3	37.07	16.26	10.92
6	1	3	29.13	14.30	9.96
7	1	3	25.54	13.11	9.33
8	1	3	22.59	12.13	8.74
9	1	3	20.64	11.54	8.45
10	1	3	18.95	10.98	8.07
20	1	3	14.71	9.44	7.07
30	1	3	13.13	8.67	6.72
40	1	3	12.97	8.40	6.69
50	1	3	12.75	8.43	6.66
60	1	3	12.56	8.36	6.63
70	1	3	12.03	8.33	6.52
80	1	3	12.02	8.29	6.52
90	1	3	12.36	8.17	6.54
100	1	3	12.31	8.12	6.49

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
2	1	4	400.8	78.26	36.70
3	1	4	116.4	37.12	21.69
4	1	4	62.34	25.34	16.41
5	1	4	45.07	20.82	14.05
6	1	4	36.10	18.06	12.74
7	1	4	32.82	16.62	11.89
8	1	4	27.76	15.33	11.22
9	1	4	26.12	14.59	10.79
10	1	4	24.39	14.14	10.42
20	1	4	17.67	11.58	9.10
30	1	4	16.01	10.76	8.55
40	1	4	15.31	10.49	8.43
50	1	4	14.57	10.10	8.16
60	1	4	14.85	10.06	8.11
70	1	4	14.77	10.26	8.18
80	1	4	14.44	10.00	8.10
90	1	4	13.65	9.88	8.01

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
100	1	4	13.85	9.75	7.83
2	1	5	511.3	96.68	46.34
3	1	5	141.0	45.49	26.72
4	1	5	78.23	31.21	20.17
5	1	5	56.65	25.22	17.27
6	1	5	43.79	21.98	15.55
7	1	5	37.06	19.77	14.47
8	1	5	32.81	18.31	13.54
9	1	5	30.27	17.41	13.01
10	1	5	28.74	16.66	12.55
20	1	5	20.73	13.27	10.64
30	1	5	18.09	12.52	10.23
40	1	5	17.90	12.31	10.10
50	1	5	16.95	11.95	9.77
60	1	5	16.64	11.64	9.63
70	1	5	16.32	11.56	9.67
80	1	5	16.06	11.68	9.61

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
90	1	5	16.31	11.66	9.71
100	1	5	16.40	11.36	9.46
2	2	1	211.7	39.13	18.17
3	2	1	61.07	19.12	10.95
4	2	1	36.45	13.94	8.67
5	2	1	26.44	11.60	7.57
6	2	1	21.82	10.21	6.92
7	2	1	19.49	9.56	6.55
8	2	1	17.22	8.92	6.24
9	2	1	16.23	8.55	6.07
10	2	1	15.65	8.38	5.98
20	2	1	11.92	7.06	5.12
30	2	1	11.07	6.48	4.85
40	2	1	10.46	6.49	4.94
50	2	1	9.80	6.26	4.70
60	2	1	10.53	6.30	4.87
70	2	1	9.54	6.20	4.64

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
80	2	1	9.36	6.27	4.74
90	2	1	10.17	6.27	4.75
100	2	1	9.28	5.98	4.64
2	2	2	361.2	69.40	33.57
3	2	2	101.7	32.31	18.99
4	2	2	58.37	23.10	14.80
5	2	2	40.90	18.71	12.71
6	2	2	33.96	16.22	11.47
7	2	2	27.87	14.77	10.58
8	2	2	25.05	13.72	10.01
9	2	2	23.13	13.04	9.65
10	2	2	21.18	12.45	9.37
20	2	2	17.14	10.53	8.18
30	2	2	15.04	10.06	7.75
40	2	2	14.16	9.26	7.28
50	2	2	13.78	9.18	7.53
60	2	2	13.33	9.09	7.28

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
70	2	2	13.42	9.09	7.29
80	2	2	12.84	8.95	7.22
90	2	2	12.90	8.98	7.18
100	2	2	12.76	9.00	7.29
2	2	3	473.9	95.48	46.42
3	2	3	142.5	45.26	26.65
4	2	3	74.17	30.73	20.11
5	2	3	53.50	24.96	17.17
6	2	3	43.52	21.75	15.41
7	2	3	35.82	19.17	14.10
8	2	3	31.84	18.08	13.40
9	2	3	29.91	17.17	12.88
10	2	3	25.82	15.81	12.02
20	2	3	19.22	13.10	10.54
30	2	3	18.27	12.30	10.06
40	2	3	16.58	11.87	9.75
50	2	3	16.23	11.62	9.54

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
60	2	3	16.01	11.54	9.48
70	2	3	15.44	11.44	9.49
80	2	3	16.05	11.50	9.43
90	2	3	15.24	11.26	9.28
100	2	3	15.61	11.15	9.39
2	2	4	599.0	123.5	58.96
3	2	4	169.8	55.68	33.21
4	2	4	94.98	38.69	25.35
5	2	4	66.47	31.07	21.47
6	2	4	51.87	26.86	19.30
7	2	4	44.43	23.86	17.57
8	2	4	38.45	22.03	16.62
9	2	4	35.38	20.85	15.78
10	2	4	32.91	19.78	15.19
20	2	4	23.31	16.28	13.06
30	2	4	20.48	14.80	12.22
40	2	4	19.30	14.13	11.81

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
50	2	4	18.32	13.77	11.42
60	2	4	19.04	13.73	11.49
70	2	4	18.74	13.61	11.36
80	2	4	17.95	13.36	11.13
90	2	4	17.87	13.32	11.31
100	2	4	17.59	13.18	11.19
2	2	5	737.8	147.7	71.96
3	2	5	210.2	68.52	40.71
4	2	5	107.7	45.68	29.98
5	2	5	78.76	36.43	25.50
6	2	5	61.55	31.45	22.54
7	2	5	51.38	28.11	20.84
8	2	5	45.35	26.03	19.75
9	2	5	41.02	24.27	18.66
10	2	5	38.10	23.55	18.28
20	2	5	26.29	18.29	15.10
30	2	5	24.22	17.18	14.26

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
40	2	5	22.70	16.32	13.86
50	2	5	21.24	15.75	13.33
60	2	5	20.65	15.62	13.33
70	2	5	20.58	15.33	13.12
80	2	5	19.76	15.15	13.00
90	2	5	19.74	15.23	12.98
100	2	5	19.84	15.17	13.01
2	3	1	288.4	56.03	27.50
3	3	1	88.01	27.63	16.07
4	3	1	50.08	19.97	12.56
5	3	1	37.05	16.36	10.85
6	3	1	29.86	14.27	9.87
7	3	1	25.01	12.97	9.15
8	3	1	23.02	12.20	8.76
9	3	1	21.33	11.64	8.40
10	3	1	19.12	11.02	8.09
20	3	1	14.51	8.97	6.98

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
30	3	1	13.45	8.78	6.91
40	3	1	12.68	8.45	6.67
50	3	1	12.41	8.50	6.65
60	3	1	12.46	8.42	6.56
70	3	1	11.83	8.27	6.49
80	3	1	12.09	8.05	6.45
90	3	1	11.75	8.13	6.47
100	3	1	12.02	8.05	6.43
2	3	2	513.8	98.84	46.93
3	3	2	139.7	44.78	26.45
4	3	2	76.92	31.03	20.09
5	3	2	53.05	24.64	16.92
6	3	2	42.91	21.71	15.38
7	3	2	36.41	19.55	14.20
8	3	2	32.46	18.06	13.42
9	3	2	29.49	16.94	12.84
10	3	2	26.53	16.05	12.28

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
20	3	2	20.14	13.28	10.70
30	3	2	17.63	12.28	9.96
40	3	2	16.88	12.06	9.94
50	3	2	16.85	11.87	9.56
60	3	2	15.92	11.39	9.35
70	3	2	16.00	11.66	9.50
80	3	2	16.16	11.33	9.55
90	3	2	15.70	11.24	9.37
100	3	2	15.07	11.09	9.17
2	3	3	666.8	129.6	63.08
3	3	3	185.1	59.94	35.64
4	3	3	98.28	40.84	26.65
5	3	3	71.11	32.76	22.53
6	3	3	54.44	27.95	20.02
7	3	3	45.53	25.19	18.58
8	3	3	40.20	23.07	17.31
9	3	3	36.97	21.58	16.59

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
10	3	3	34.19	20.83	16.03
20	3	3	23.55	16.66	13.61
30	3	3	21.79	15.26	12.54
40	3	3	20.20	14.59	12.33
50	3	3	20.09	14.59	12.14
60	3	3	19.63	14.26	12.03
70	3	3	19.15	14.00	11.79
80	3	3	18.50	13.95	11.89
90	3	3	18.51	13.92	11.82
100	3	3	18.87	13.92	11.74
2	3	4	851.9	162.8	78.71
3	3	4	218.8	72.59	43.41
4	3	4	120.8	49.32	32.45
5	3	4	81.84	39.76	27.40
6	3	4	64.58	33.38	24.16
7	3	4	54.84	30.33	22.43
8	3	4	48.21	27.59	20.90

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
9	3	4	43.99	26.05	20.02
10	3	4	39.98	24.40	19.33
20	3	4	27.46	19.49	15.96
30	3	4	24.31	17.98	15.13
40	3	4	23.41	17.12	14.72
50	3	4	22.26	16.59	14.27
60	3	4	22.25	16.48	14.24
70	3	4	21.71	16.34	14.05
80	3	4	21.22	16.40	14.17
90	3	4	21.29	16.32	13.92
100	3	4	20.97	16.13	13.81
2	3	5	1001	190.6	94.28
3	3	5	257.4	84.84	50.79
4	3	5	139.7	58.38	38.18
5	3	5	97.91	46.43	32.43
6	3	5	78.08	39.97	28.88
7	3	5	64.39	35.35	26.37

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
8	3	5	56.59	32.44	24.61
9	3	5	50.33	30.38	23.40
10	3	5	46.61	28.72	22.57
20	3	5	32.56	22.36	18.49
30	3	5	28.58	20.83	17.53
40	3	5	26.83	19.86	16.96
50	3	5	25.20	19.10	16.36
60	3	5	24.61	18.65	16.03
70	3	5	24.23	18.85	16.27
80	3	5	23.97	18.63	16.04
90	3	5	24.46	18.52	15.99
100	3	5	23.55	18.17	15.70
2	4	1	425.8	77.56	36.74
3	4	1	116.1	37.05	21.45
4	4	1	64.15	25.43	16.42
5	4	1	45.56	20.78	14.11
6	4	1	36.66	18.30	12.84

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
7	4	1	31.46	16.45	11.90
8	4	1	28.33	15.40	11.25
9	4	1	26.06	14.53	10.80
10	4	1	23.69	13.59	10.26
20	4	1	16.98	11.31	8.94
30	4	1	15.71	10.56	8.54
40	4	1	14.51	10.07	8.19
50	4	1	14.75	10.29	8.42
60	4	1	14.54	10.23	8.17
70	4	1	14.23	10.13	8.08
80	4	1	14.70	9.98	8.07
90	4	1	13.80	9.84	8.00
100	4	1	13.55	9.73	7.86
2	4	2	608.0	123.1	59.51
3	4	2	175.4	56.42	33.90
4	4	2	94.14	38.43	25.13
5	4	2	66.63	30.89	21.38

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
6	4	2	53.10	26.65	19.03
7	4	2	44.21	23.94	17.61
8	4	2	38.81	21.84	16.51
9	4	2	35.17	20.72	15.81
10	4	2	33.79	19.96	15.25
20	4	2	24.17	16.11	13.16
30	4	2	20.30	14.59	12.07
40	4	2	19.55	14.13	11.79
50	4	2	18.61	13.68	11.60
60	4	2	18.44	13.82	11.63
70	4	2	18.35	13.46	11.25
80	4	2	18.35	13.45	11.40
90	4	2	18.28	13.24	11.17
100	4	2	17.35	13.06	11.20
2	4	3	832.4	162.7	78.75
3	4	3	222.4	72.81	43.73
4	4	3	117.3	48.28	32.16

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
5	4	3	83.69	39.74	27.64
6	4	3	65.06	33.84	24.45
7	4	3	53.63	30.10	22.47
8	4	3	48.43	27.84	21.05
9	4	3	44.83	26.30	20.14
10	4	3	40.45	24.49	19.07
20	4	3	27.97	19.51	16.11
30	4	3	24.76	17.99	15.01
40	4	3	23.56	17.26	14.56
50	4	3	22.89	17.15	14.48
60	4	3	21.68	16.70	14.29
70	4	3	21.57	16.24	13.91
80	4	3	21.39	16.35	13.98
90	4	3	21.25	16.33	13.97
100	4	3	21.27	16.12	13.86
2	4	4	996.7	195.3	93.36
3	4	4	270.9	89.25	53.06

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
4	4	4	146.6	59.59	39.44
5	4	4	97.66	47.20	32.78
6	4	4	79.75	40.28	29.09
7	4	4	65.04	36.23	27.02
8	4	4	57.64	33.30	25.46
9	4	4	50.67	30.90	24.05
10	4	4	45.62	28.42	22.52
20	4	4	32.41	22.50	18.84
30	4	4	28.17	20.77	17.82
40	4	4	27.36	20.22	17.19
50	4	4	25.98	19.43	16.72
60	4	4	25.54	19.36	16.66
70	4	4	24.98	19.15	16.51
80	4	4	24.06	18.99	16.37
90	4	4	24.39	18.69	16.34
100	4	4	23.61	18.36	16.02
2	4	5	1196	229.2	111.2

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
3	4	5	314.3	104.1	62.50
4	4	5	168.4	69.30	45.61
5	4	5	117.6	54.75	38.47
6	4	5	87.90	46.56	33.98
7	4	5	74.87	41.25	30.80
8	4	5	66.05	37.65	28.82
9	4	5	59.20	35.60	27.62
10	4	5	54.53	33.36	26.87
20	4	5	35.93	25.76	21.72
30	4	5	32.13	23.54	20.10
40	4	5	28.58	22.36	19.59
50	4	5	29.01	22.14	19.06
60	4	5	28.13	21.60	18.71
70	4	5	28.27	21.53	18.69
80	4	5	26.41	20.90	18.26
90	4	5	26.34	20.85	18.42
100	4	5	26.05	20.63	18.19

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
2	5	1	514.0	96.74	46.19
3	5	1	142.2	44.31	26.18
4	5	1	75.51	31.40	20.15
5	5	1	56.50	25.22	17.25
6	5	1	43.74	21.95	15.45
7	5	1	37.70	19.96	14.46
8	5	1	33.26	18.43	13.61
9	5	1	29.71	17.35	13.01
10	5	1	27.67	16.19	12.37
20	5	1	20.57	13.58	10.83
30	5	1	18.61	12.54	10.10
40	5	1	17.64	12.04	9.87
50	5	1	16.97	12.00	9.76
60	5	1	17.14	11.91	9.65
70	5	1	16.81	11.81	9.67
80	5	1	16.24	11.68	9.67
90	5	1	15.86	11.35	9.34

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
100	5	1	15.41	11.33	9.46
2	5	2	754.8	151.1	72.21
3	5	2	212.6	66.98	39.57
4	5	2	112.6	45.76	29.82
5	5	2	77.33	36.55	25.33
6	5	2	59.90	31.20	22.68
7	5	2	51.31	28.18	20.95
8	5	2	45.06	26.09	19.70
9	5	2	41.18	24.34	18.60
10	5	2	36.94	22.97	17.89
20	5	2	26.78	18.65	15.15
30	5	2	23.59	16.99	14.09
40	5	2	21.71	16.00	13.68
50	5	2	21.18	15.78	13.50
60	5	2	21.18	15.87	13.45
70	5	2	20.62	15.57	13.21
80	5	2	20.87	15.50	13.26

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
90	5	2	20.38	15.22	13.15
100	5	2	19.99	15.11	13.01
2	5	3	949.8	187.0	90.44
3	5	3	274.0	86.98	51.66
4	5	3	144.2	58.44	38.42
5	5	3	99.83	46.44	32.44
6	5	3	76.14	39.62	28.74
7	5	3	62.42	34.96	26.16
8	5	3	54.79	32.17	24.58
9	5	3	50.82	30.23	23.42
10	5	3	47.55	29.11	22.64
20	5	3	33.49	22.94	18.89
30	5	3	27.81	20.36	17.26
40	5	3	26.21	19.63	16.77
50	5	3	25.15	19.19	16.53
60	5	3	24.73	18.88	16.24
70	5	3	24.34	18.66	16.20

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
80	5	3	24.02	18.55	16.04
90	5	3	23.31	18.22	15.95
100	5	3	23.57	18.24	15.91
2	5	4	1173	230.8	113.1
3	5	4	315.0	101.4	60.91
4	5	4	164.3	68.92	45.41
5	5	4	115.8	54.86	38.32
6	5	4	90.16	45.86	33.61
7	5	4	76.50	41.86	31.26
8	5	4	66.05	37.88	29.04
9	5	4	58.78	35.63	27.64
10	5	4	54.74	33.08	26.27
20	5	4	37.43	26.03	21.92
30	5	4	32.32	23.83	20.29
40	5	4	30.54	22.94	19.53
50	5	4	29.19	22.09	19.12
60	5	4	27.72	21.61	18.71

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
70	5	4	27.20	21.33	18.68
80	5	4	27.12	21.20	18.49
90	5	4	26.95	20.81	18.28
100	5	4	26.31	20.58	18.20
2	5	5	1384.20	265.43	128.95
3	5	5	358.08	117.58	70.23
4	5	5	189.52	79.45	52.71
5	5	5	129.19	62.62	43.83
6	5	5	102.75	53.05	38.98
7	5	5	85.56	47.66	35.75
8	5	5	73.73	43.21	33.22
9	5	5	65.51	40.27	31.40
10	5	5	60.75	37.83	29.67
20	5	5	41.92	29.85	24.89
30	5	5	34.87	26.52	22.72
40	5	5	33.34	25.48	22.23
50	5	5	31.66	24.72	21.38

Continued on Next Page...

Table B.1 – Continued

v	**h**	**d**	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
60	5	5	31.31	24.16	21.04
70	5	5	31.85	24.15	21.17
80	5	5	30.13	23.54	20.73
90	5	5	29.88	23.37	20.76
100	5	5	29.68	23.45	20.67