Multivariate Quality Control

Using

Loss-Scaled Principal Components

A Thesis
Presented to
The Academic Faculty

by

Terrence E. Murphy

In Partial Fulfillment
of the Requirements for the Degree

Industrial and Systems Engineering
Georgia Institute of Technology
November 2004

Multivariate Quality Control

Using

Loss-Scaled Principal Components

Approved by:

Dr. Kwok-Leung Tsui, Co-Advisor
Industrial and Systems Engineering

Dr. David Goldsman
Industrial and Systems Engineering

Dr. Victoria C.P. Chen, Co-Advisor
University of Texas, Arlington

Dr. Roshan Vengazhiyil
Industrial and Systems Engineering

Dr. Janet K. Allen
Mechanical Engineering

Date Approved: 4 November 2004

DEDICATION

A Gris,

el bonbon de mi corazón,

y mis dos bonboncitos

Nolan James

y Mikaela

ACKNOWLEDGEMENTS

I thank my wife, Inginia Genao, for her unswerving support during these six long years. Thanks for your patience and for sticking to our deal. I want to thank my parents for giving me all the love and support they did while raising so many children. I want to express my special gratitude to my advisors, Dr. Kwok-Leung Tsui and Dr. Victoria Chen, for their support, patience and guidance these last 6+ years. Dr. Chen helped me through my first difficult years and I would not have persisted without her help and encouragement. You are a gifted teacher, mentor and researcher and your family and colleagues are fortunate to have you. I am grateful to Dr. Tsui for rescuing me after Dr. Chen's departure. I admire your intuitive grasp of research concepts and your gift for immediately getting to the heart of the matter. You consistently helped me see the larger picture when I got lost in the details. I thank Dr. Janet Allen for her detailed feedback and attention to detail when reviewing my writing and for help in so many ways during my collaboration with folks from the SRL. I thank the other members of my committee, Dr. Dave Goldsman and Dr. Roshan Vengazhiyil for your attention and feedback. Big ups for Yao Lin who taught me how to run the Kriging and MARS code and for showing me how to run iSIGHT and *IDEAS*.

I am grateful for the support and resources provided by OMED Educational Services during the last three years of my studies. OMED provided me with the funding, computers and physical space to perform my research in a pleasant and collegial atmosphere. I especially thank my boss Robert Hume for superb supervision and for maintaining a high standard of data integrity at OMED. I thank director S. Gordon Moore for his leadership in promotion of OMED's mission. I thank Jackie for her positive vibes, Patrice for her frequent help, Bruce for keeping my boxes running, Willie and Alvin for your thoughts on hip-hop, Letitia for Master Bryce, and Mr. Frank Stanley for keeping me in touch with the reality of American politics.

On a more personal note, many friends and colleagues have helped me survive the trip. Thanks Jerome for your friendship and for bringing OMED into my life. Thanks to Seoung Bum for help with many things. Thanks to Dan Lawrence and Tom Barker at RIT for encouraging me to pursue the doctorate despite the odds. A big shout out to the Kressmann-Kehoe family in Rochester and the Noyes-Buchanan clan in Holyoke. Finally, I want to thank the School of Industrial and Systems Engineering for financial support. I gratefully acknowledge support from the National Science Foundation, grant DMI-0100123.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

We consider a principal components based decomposition of the expected value of the multivariate quadratic loss function, i.e., MQL. The principal components are formed by scaling the original data by the contents of the loss constant matrix, which defines the economic penalty associated with specific variables being off their desired target values. We demonstrate the extent to which a subset of these "loss-scaled principal components", i.e., LSPC, accounts for the two components of expected MQL, namely the trace-covariance term and the off-target vector product. We employ the LSPC to solve a robust design problem of full and reduced dimensionality with deterministic models that approximate the true solution and demonstrate comparable results in less computational time. We also employ the LSPC to construct a test statistic called loss-scaled T^2 for multivariate statistical process control. We show for one case how the proposed test statistic has faster detection than Hotelling's T^2 of shifts in location for variables with high weighting in the MQL. In addition we introduce a principal component based decomposition of Hotelling's T^2 to diagnose the variables responsible for driving the location and/or dispersion of a subgroup of multivariate observations out of statistical control. We demonstrate the accuracy of this diagnostic technique on a data set from the literature and show its potential for diagnosing the loss-scaled T^2 statistic as well.

CHAPTER I

INTRODUCTION

In this chapter we describe the intent and structure of this thesis, which proposes a new principal component based method for applying the multivariate quadratic loss function (i.e., MQL) to multivariate quality control. In section 1 we explain the concept and motivation for the proposed method. In section 2 we review principal components and how they may be used to decompose the expected value of the MQL. In section 3 we describe the structure of the thesis.

1.1 Basic Concept of Thesis

This thesis proposes and examines a new way of employing principal components (PCs) in multivariate quality control. Quality control is commonly divided into off-line activities, synonymous with robust design (RD), and on-line procedures, also known as statistical process control. Both types of activity are vital for maintaining and improving quality in the increasingly competitive global marketplace. Most research in both areas of quality control has dealt with single variables. Since most complex systems are multivariate in nature, there is an increasing need for user friendly multivariate techniques.

The multivariate quadratic loss function (MQL) is arguably the most popular multivariate technique in static RD since it combines the losses of multiple responses into a single objective function which calculates the cost of poor quality. It also exploits the correlation structure of the responses and, as an extension of the mean squared error criterion used in linear regression, has intuitive appeal to a wide audience of engineers and scientists. Although Taguchi's techniques for univariate robust parameter design Taguchi and Wu (1980) are widely practiced, his multivariate approaches have not been well received. We propose decomposing expected multivariate quadratic loss into a set of dominant principal components which are then modelled as functions of the design variables themselves.

Although the loss function is not commonly associated with multivariate statistical process control (MSPC), a few pieces in the literature have discussed defining the acceptance region according to economic value. These include Montgomery and Klatt (1972) and Mohebbi and Hayre (1989). The latter presents a MSPC method based on multivariate quadratic loss. Tsui and Woodall (1993) propose a form of the multivariate exponentially weighted moving average (MEWMA) of Lowry et al (1992) which incorporates the loss constant matrix of multivariate quadratic loss. We propose an MQC statistic based on multivariate quadratic loss which can be decomposed into PCs for both dimensional reduction of ill-conditioned data as well as diagnosis of shifts in location and / or dispersion.

PCs are a purely multivariate technique in that they are specifically designed to work with multiply correlated random variables and do not exist in a univariate form. In addition to their notable mathematical properties, they sometimes identify non-directly observable multiple-response factors with meaningful physical interpretations. This idea of "latent factors" is borrowed from a related discipline known as "factor analysis", which is mathematically more sophisticated than regular principal component analysis.

We were motivated to harness the intuitive and mathematical strength of PCs to simplify the use of the MQL in both areas of quality engineering. To that end, this thesis can be seen as an extension of prior work employing the MQL in multivariate RD (i.e., Tsui (1999)) and MSPC (i.e., Tsui and Woodall (1993). In both areas we integrate the contents of the MQL into specially constructed PCs called loss-scaled principal components (LSPC). The resulting LSPC are uncorrelated and their full set completely replicates both components of expected MQL. A formal subset of the LSPC typically approximate the covariance and off-target vector components of expected MQL well. We also show in this thesis how both PCs and LSPCs can accurately diagnose shifts in location and or dispersion in MSPC.

1.2 Introduction to Principal Components

1.2.1 Motivation for Employing Principal Components (i.e., PCs)

Our work with PCs is motivated by the desire to simplify the search for an optimal design vector when the product possesses multiply correlated responses. PCs simplify this

optimization search in two significant ways.

The first simplification comes from the reduction of the dimensionality of the optimization problem from that of r correlated responses to that of a subset of $p < r$ uncorrelated factors, i.e., PCs. This reduction of dimensionality is potentially useful in the computational sense, since optimizing with a smaller number of models may result in significant savings of design time.

The second advantage is that in the physical sciences, PCs often identify an underlying, multi-variable factor that can't be identified by observing individual responses. Johnson and Wichern (1992) cite a study of weekly rates of return for five stocks over a two year period. Most of the variation in the five stocks is well explained by two principal components. The first PC represents a weighted average of the five stocks which can be interpreted as a general stock-market component. The second PC represents a contrast between a group of chemical stocks (e.g., Allied Chemical, du Pont and Union Carbide) and a group of oil stocks (e.g., Exxon and Texaco). This second component can be interpreted as an industry specific factor. The remaining three PCs, which correspond to the decreasingly ordered eigenvalues of the data's covariance matrix, do not lend themselves to a reasonable physical interpretation. This is a common finding for the PCs corresponding to smaller eigenvalues.

A second example comes from Jackson (1991), who describes how nine measurements of optical density in color film can be well represented by five PCs. Two of these PCs are of special importance since they represent color balance factors between pairings of colors. The first color-balance factor contrasts red and blue optical density measurements while the second does the same for green and magenta. While these color-balance factors are not directly monitor-able, they have a huge effect on human perception of photographic quality. For the Eastman Kodak Company, understanding and controlling these color balance factors directly translates into product quality. Hence the PC serves to identify and monitor an important quality characteristic not directly measurable at any single wavelength of light (i.e., color).

Flury (1988) discusses an interesting case regarding analysis of forged Swiss currency. By taking four physical measurements of the forged bills, two PCs explain nearly all the

variation in those response. The first PC is a linear combination of horizontal off-target measurements and the second PC of vertical off-target measurements. Here the PC analysis brought to light the fact that the rectangular nature of the bill allows the investigator to identify most discrepancies by examining the axes of the plane formed by the note. This example neatly demonstrates the geometric nature of PCs. The elegant mathematical properties of PCs are based upon their directionally specific principal axes, which have optimal properties with respect to the variability of multi-response data.

A second example by Johnson and Wichern (1992) examines weekly rates of return for five stocks listed on the New York Stock Exchange over a two year period. While the individual observations across the 100 weeks appear to be independent, the rates of return across the stocks are correlated due to their common dependence on general economic conditions. The majority of the variation across the five stocks is well represented by two PCs which represent market and industry factors. The three non-petroleum companies in the first PC represent general market conditions while the two petroleum stocks represent a specific oil-industry component.

These four examples highlight the inherently multivariate nature of PCs and their special ability to extract important multiple-response factors of interest. A common criticism of PCs is that they are simply a mathematical construct with no tangible physical meaning or interpretation. This criticism may be more valid in the social sciences, where more abstract notions like learning style or personality index are frequently encountered. Jackson (1991) claims that more often than not in the physical sciences, a practical interaction or dynamic can be identified as a PC. While we find this assertion intuitively attractive, we examine several mathematical characteristics of PCs that make them useful for working with multivariate data regardless of their physical interpretation.

1.2.2 How Principal Components Work

1.2.2.1 Definition of Principal Components

PCs can be thought of as a geometrically rigid rotation of the principal axes of the response data which transforms the r correlated responses into r newly uncorrelated variables.

The matrix concept used to rotate the axes is the fact that an r x r symmetric non-singular matrix such as $\mathbf{\Sigma_Y}$ can be converted into a diagonal matrix $\mathbf{L}$ simply by pre- and post-multiplying it by an orthonormal matrix $\mathbf{U}$ such that:

$$\mathbf{U^T \Sigma_Y U} = \mathbf{L}$$

where the diagonal elements of $\mathbf{L}$ and the columns of $\mathbf{U} = (\mathbf{U_1}, \ldots, \mathbf{U_r})$ are respectively the eigenvalues and orthonormal eigenvectors of $\Sigma_\mathbf{Y}$. The new, uncorrelated variables are simply the products of the columns of $\mathbf{U}$ by the vectors of response variables.

e.g.,

$$\begin{aligned} \mathbf{Z_1} &= \mathbf{U_1 Y} = u_{11}Y_1 + \cdots + u_{r1}Y_r \\ &\vdots \\ \mathbf{Z_r} &= \mathbf{U_r Y} = u_{1r}Y_1 + \cdots + u_{rr}Y_r \end{aligned}$$

Multiplying the response matrix $\mathbf{Y}$ by the orthonormal columns of the matrix $\mathbf{U}$ is geometrically equivalent to a principal axis rotation of the covariance matrix $\Sigma_\mathbf{Y}$ where the elements of the eigenvectors are the direction cosines of the new axes (i.e., the PCs) with respect to the original axes (i.e., the original response variables).

The covariance matrix of the PCs (i.e., $\Sigma_\mathbf{Z}$) is simply the diagonal matrix $\mathbf{L}$ since each PC is a linear combination of the elements of $\mathbf{Y}$:

$$\begin{aligned} \mathbf{Z} &= \mathbf{U^T Y} \\ &\Downarrow \\ \Sigma_\mathbf{Z} &= \mathbf{U^T \Sigma_Y U} \\ &= \mathbf{L} \end{aligned}$$

Using PCs is often a very practical data reduction method because the PCs have certain optimal properties for describing the variance of the response data set. Assuming that all the eigenvalues are of multiplicity one, the first PC is that linear combination of the individual responses which yields the maximum variance of all possible linear combinations.

The second PC is that linear combination yielding the next highest variance and so on in decreasing order. Any collection of the first p PCs accounts for a greater proportion of the total variation of the response data than any other set of p such linear combinations.

Furthermore because the covariance matrix of the PCs is diagonal, the decreasingly ordered eigenvalues are the variances of the respectively ordered PCs. This often means that a smaller number of the PCs account for all variation in the data-set reasonably attributed to the responses. Design decision makers most typically choose the number of PCs which account for some pre-determined proportion of the total sample variance, e.g., ninety percent.

1.2.2.2 How Many Principal Components to Use?

The question of how many PCs to retain is far from trivial and Jackson (1991) reviews no fewer than seven methods for answering this question. He discourages the decision rule based on explanation of a pre-ordained proportion of sample variance. We will defer answering this question for now since it is not yet clear to us which rule is the most appropriate for our robust design purposes

1.2.2.3 When are PCs from the Correlation Matrix Preferred?

While deriving PCs from covariance matrices simplifies both physical interpretation and computation of the statistics discussed, their use is not always appropriate. Jackson (1981) mentions two conditions for when it is appropriate to obtain PCs from the response correlation matrix rather than the covariance matrix.

- When response variables are measured in different units. Since the relative magnitudes of different units may be greatly different, the responses are standardized to remove the influence of incongruous response variances before calculating the PCs.

- Response variables are in the same units but of widely ranging variances. If the unevenness of the variances may give unreasonable weight to specific responses, then the PCs should be derived from the correlation matrix.

It is especially true in the social sciences, where units of measurement are often difficult to quantify precisely, that employing PCs derived from the correlation matrix is commonplace. However it is easy to imagine a product design situation where the controllable design variables are in different units such as temperature, force, volume, etc.

1.2.2.4 Contrasting with Factor Analysis

We take this opportunity to clarify how the commonly misconstrued terms Principal Component Analysis (PCA) and Factor Analysis (FA) are used in this thesis. Our terminology is taken from Jolliffe (2002).

PCA consists of forming PCs of the original responses defined as the product of the eigenvectors of the covariance/correlation matrix and the vector of multiple responses. These linear combinations of the responses are a mathematically elegant way of representing the diagonal terms of a covariance/correlation matrix and are examined for their potential interpretation as reasonable multi-response factors. While PCA is largely a descriptive and exploratory technique, it can be supplemented by inferential methods which add a statistical rigor to making decisions regarding its applicability.

The underlying idea of FA is that r observed random variables (i.e., $\mathbf{Y}$) can be expressed as linear combinations of $p < r$ hypothetical or common factors (i.e., f_k) per the following model:

$$\begin{aligned} y_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1p} f_p + \epsilon_1 \\ y_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2p} f_p + \epsilon_2 \\ &\vdots \\ y_r &= \lambda_{r1} f_1 + \lambda_{r2} f_2 + \cdots + \lambda_{rp} f_p + \epsilon_r \end{aligned}$$

where $\lambda_{jk} : \quad j = 1, 2, \ldots, r \ ; \ k = 1, 2, \ldots, p$ are constants called factor loadings and $e_j : \quad j = 1, 2, \ldots, p$ are error terms called specific factors because e_j is specific to y_j while the f_k are common to all the responses.

This equation can be rewritten in matrix form as:

$$\mathbf{Y} \;=\; \mathbf{\Lambda f} \;+\; \mathbf{e} \tag{1.1}$$

The following assumptions are associated with this model:

1. E[$\mathbf{e}$]= $\mathbf{0}$, E[$\mathbf{f}$]= $\mathbf{0}$, E[$\mathbf{Y}$]= $\mathbf{0}$,
2. E[$\mathbf{ee^T}$]= Ψ (diagonal), E[$\mathbf{fe^T}$]= $\mathbf{0}$, (a matrix of zeroes), E[$\mathbf{ff^T}$]= $\mathbf{I_p}$ (an identity matrix).

The second set of assumptions states respectively that the error terms are uncorrelated, the common factors and error terms are uncorrelated and that the common factors are orthogonal.

Estimation of the model terms in FA is done sequentially, starting with the parameters in Λ and Ψ. The covariance matrix for Equation (1.1) is:

$$\Sigma \;=\; \Lambda\Lambda^T + \Psi \tag{1.2}$$

where the sample covariance matrix is usually substituted for Σ. The two most popular ways of estimating the elements of Λ are PCA and maximum likelihood estimation (i.e., MLE).

PCA tends to do a better job of estimating the diagonal elements of the covariance or correlation matrix of the common factors whereas MLE generally creates better estimates of the off-diagonal components. Johnson and Wichern (1992) note that PCA tends to provide interpretable factors without need for orthogonal rotation. Since MLE does not usually provide physically interpretable factors, orthogonal rotation of the common factors is routinely employed. The goal of the rotation is to find the strongest contrasts between the elements of the common factors. The reason orthogonal rotation does not change the underlying covariance matrix is because of the following property:

$$\begin{aligned}
\Lambda^*\Lambda^{*T} &= (\Lambda\mathbf{T})(\Lambda\mathbf{T})^T \\
&= \Lambda\mathbf{T}\mathbf{T^T}\Lambda^T \\
&= \Lambda\Lambda^T
\end{aligned}$$

where $\mathbf{T}$ is any orthogonal transformation. The orthogonal transformation sought is that rigid rotation of Λ that optimizes a particular criterion and most statistical software packages allow the design decision maker to choose between Varimax, Quantmax, etc.

We now summarize the key differences between PCA and FA. PCA is a linear transformation of a covariance/correlation matrix, whose first few eigenvectors often have a reasonable multi-response interpretation. Although the response vector $\mathbf{Y}$ can be modelled using the PCs, that is not a necessary element of PCA. In the physical sciences the first few PCs commonly lend themselves to reasonable multi-response factors. If the PCs do not lend themselves to a rational interpretation, they are not employed. There is no rotation of the PCs to find a better physical meaning.

FA hypothesizes the existence of common factors which can not be observed and whose covariance matrix must comply with a number of modelling assumptions. Typically the common factors are not readily interpretable and need to be orthogonally rotated to find a tangible physical meaning. There are specific criteria used to choose the specific orthogonal transformation which are available in many statistical packages featuring FA. PCA and maximum likelihood estimation are the two most common techniques employed to model the covariance matrix of the common factors in FA.

PCA is largely a mathematical re-packaging of the covariance matrix whose eigenvectors sometimes simplify the explanation of complex systems. FA is a more formal covariance-modelling technique whose objective is the identification of non-observable multi-response common factors and uses orthogonal and sometimes oblique rotation to accomplish that goal.

1.2.2.5 What are Loss-Scaled Principal Components?

In the prior subsections of this chapter we've defined PCs and their relevant properties. While PCs are one of the oldest multivariate techniques in existence, their variance optimizing properties continue finding new application areas. In Chapter 3 we review multivariate robust design techniques and comment on their respective use of the correlation structure of the responses.

Of the multivariate techniques considered, minimizing the expected value of the MQL makes the best use of the correlation structure and can be decomposed into covariance and off-target vector product terms as shown below:

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \\ &= trace[\mathbf{A}\boldsymbol{\Sigma}_\mathbf{Y}] + [E\mathbf{Y}-\tau]^T\mathbf{A}[E\mathbf{Y}-\tau]. \end{aligned} \tag{1.3}$$

To minimize the two terms of the expected MQL, one must simultaneously consider the contributions of the loss matrix $\mathbf{A}$ as well as the response covariance/correlation matrix. The primary conceptual contribution of this thesis is the formation of PCs which simultaneously account for the loss constant matrix and the response covariance/correlation matrix in the MQL. Since the loss constant matrix $\mathbf{A}$ is symmetric and positive definite it can be decomposed by either Cholesky decomposition or diagonalized in the following way:

$$\mathbf{G^T A G} = \mathbf{\Lambda_A}$$

which allows $\mathbf{A}$ to be expressed as

$$\begin{aligned} \mathbf{A} &= \mathbf{G\Lambda_A G^T} \\ \mathbf{A} &= (\mathbf{G\Lambda_A^{\frac{1}{2}}})(\mathbf{G\Lambda_A^{\frac{1}{2}}})^\mathbf{T} \end{aligned}$$

where the columns of $\mathbf{G}$ are the eigenvectors of $\mathbf{A}$ and the diagonal elements of $\mathbf{\Lambda_A}$ are the eigenvalues of $\mathbf{A}$. If we then define $\mathbf{q} = (\mathbf{G\Lambda_A^{\frac{1}{2}}})$, we can express $\mathbf{A}$ in the following way:

$$\mathbf{A} = \mathbf{qq^T}$$

This allows Equation (1.3) to be written as follows:

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \\ &= E_(\mathbf{q^T Y}-\mathbf{q^T}\tau)^T(\mathbf{q^T Y}-\mathbf{q^T}\tau) \\ &= trace[\mathbf{q^T\Sigma_Y q}]+(E\mathbf{q^T Y}-\mathbf{q^T}\tau)^T(E\mathbf{q^T Y}-\mathbf{q^T}\tau) \\ &= trace[\mathbf{q^T\Sigma_Y q}]+[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]. \end{aligned} \tag{1.4}$$

Therefore multiplying the original response vectors by the matrix $\mathbf{q^T}$ incorporates the contents of matrix $\mathbf{A}$ and $\mathbf{\Sigma_Y}$ directly into the matrix $\mathbf{q^T\Sigma_Y q}$. We will see in the next section how this simplifies the expression of the covariance and off-target vector product terms on the right side of Equation (1.3).

Loss-scaled responses are simply $\mathbf{q^T Y}$ and their principal components are formed by multiplying $\mathbf{Y}$ by the eigenvectors of $\mathbf{q^T\Sigma_Y q}$.

This work is motivated by the desire to find out whether a subset of the loss-scaled principal components provide an acceptably accurate and faster way to solve for robust designs. This will be investigated in the application in chapter 5 and its ramifications for use in MSPC will be explored in chapter 7.

1.3 Structure of Thesis

The thesis is structured as follows. This first chapter describes the thesis at a high level, briefly reviews the underlying concepts behind principal components (PC) and defines loss-scaled principal components (LSPC). Chapter 2 represents an early, non-condensed version of a review paper on robust design which examined many articles in the applied statistical journals over the last twenty years. It is archived here to document the many short descriptions of these articles, some of which made it into the leaner version submitted for publication. Likewise the material in chapter 3 has been whittled down and submitted for a separate publication on multivariate robust design. The reader is forewarned that chapters 2 and 3 are archived here as documentation and in no way represent a sparse or economical packaging of this information. In Chapter 4 we examine the general case of how

well a subset of LSPC approximate expected MQL by looking at the possible combinations of loss constant and covariance matrices. In Chapter 5 we propose a procedure for static multivariate RD and demonstrate our proposed technique on a product design problem featuring 8 design variables and 6 correlated responses. We compare optimal design vector and computation time from our procedure to those yielded by traditional MQL. In Chapter 6 we first review multivariate SPC and then discuss how PCs can be used for diagnosing shifts in location and or dispersion. In Chapter 7 we propose a multivariate test static that combines Hotelling's T^2 with the loss constant matrix and compare its average run length (ARL) properties to Hotelling's T^2 statistic. We also demonstrate how the proposed test statistic can be decomposed to diagnose the individual variables driving location and or dispersion out of control. We conclude both Chapter 7 and this dissertation by suggesting future directions for this research in multivariate quality control.

CHAPTER II

A REVIEW OF ROBUST DESIGN FROM 1980 - 2000

Robust design (RD) became very popular after Genichi Taguchi's parameter design methods were introduced to US corporations in the 1980's. A broad definition of RD is the design of products and processes such that their performance variation is insensitive (i.e. robust) to changes in factors beyond reasonable control (i.e. noise factors). This paper tracks the evolution of RD from the designs and analysis of Taguchi to the variety of DOE, modelling and analysis approaches currently practiced in RD. The majority of articles stress the use of fractional factorial designs which incorporate control and noise variables into a combined array and fit a variety of fixed, random and mixed effects or GLM models directly to the process response. Several Bayesian techniques are mentioned as well as new techniques in non-linear optimization. Discussion of multivariate approaches to RD is deferred to Chapter 3.

2.1 Evolution of General Methodology in RD

2.1.1 Clarification of Taguchi's Parameter Design

This section features several references which helped to explain and clarify Taguchi's static parameter design as it became widely studied and practiced around the world.

Kacker (1985) defined Off-Line Quality Control, Parameter Design and Taguchi's approach to the latter. The central idea of parameter design is to choose design parameter settings which minimize the product or process performance sensitivity to uncontrollable sources of variation. He further describes Taguchi's designed experiments, performance measure, loss function and the two step procedure.

Taguchi's experimental design takes an orthogonal array for the controllable design parameters (i.e. an inner array of control factors) and crosses it with another orthogonal array for the factors beyond reasonable control (i.e. an outer array of noise factors). Hereafter

we refer to this design as the product array.

At each test combination of control factor levels, the entire noise array is run and a performance measure is calculated. Taguchi's recommended performance measure is called the signal-to-noise ratio (SNR) which is a quadratic function of the data. A model of the SNR is created as a function of the control parameters and then used to find the control factor levels that satisfy the desired optimization criterion. He offers a variety of SNRs from which to choose, depending on the final desired target value for the process mean. The three target choices for process mean are the smaller-the-better (STB), larger-the-better (LTB) and nominal-the-best (NTB). Each of these situations features a different SNR.

Taguchi's optimization goal is to minimize a loss function which is proportional to the squared difference between process mean and a target value. Taguchi's method of achieving this minimized loss is called "the two step process" because it relies on first finding the control factor levels that optimize the SNR, and secondly on using control factors independent of the SNR to shift the process mean to its desired target. Hereafter we call this method of analysis the loss model approach (LMA).

Hunter (1985) transmitted Taguchi's ideas to a wider audience by clarifying his choice of experimental design, the role of interactions and the effects of data transformations. Barker (1990) complements Hunter's interpretation with a universally readable (i.e. for statisticians and non-statisticians alike) introduction to the philosophy of Taguchi, clearly written in classical Taguchian terminology.

2.1.2 Description and Critique of Taguchi's RD Philosophy

Pignatiello (1988) stresses that one must examine both Taguchi's strategy and tactics to comprehend his contributions. He defines the Taguchi strategy as attempting to find a best design, which minimizes expected loss over an uncontrollable noise space. Taguchi tactics are the specific techniques recommended by Taguchi to implement this strategy and include the product array, the SNR and the loss model approach. While the literature is replete with constructive criticism of Taguchi's tactics, his strategy, that of an empirical application of decision making under uncertainty, is fundamentally sound.

In Nair (1992) twelve leading practitioners of RD provided their views regarding the philosophy and technical practices of Taguchi. In the following paragraphs we summarize comments from some participants in Nair's discussion regarding this philosophy. Nair uses the term robust design synonymously with Taguchi's parameter design.

Phadke states that robust design adds a new dimension to statistical experimental data by explicitly addressing how to economically reduce product variation while ensuring that laboratory findings hold up in customer environments.

Shin Taguchi states that the goal of parameter design is not to characterize the system but to achieve robust function. Box takes issue with Shin Taguchi's goal statement and feels engineers can best reduce variation by understanding the causal relations and mechanisms by which systems function. Box also credits Taguchi for his enormously valuable contribution in making robustness studies an integral part of the design of industrial products and processes.

Lorenzen applauds Taguchi for making the idea of robustness popular within the engineering community. While agreeing that emphasizing the importance of variance reduction to the larger community is of great value, Kacker reminds the reader that the specific techniques of parameter design are merely one option of many with which to pursue these objectives.

Lucas attributes the widespread adoption of parameter design to the simplicity of Taguchi's designs, linear graphs and loss functions. He views criticism of Taguchi's analysis as minor compared to his unparalleled success in getting engineers to run factorial experiments.

Tsui (1992) seconds the importance of getting user-friendly statistical tools into engineering hands and praises Taguchi for his contributions in this area. He also argues that more efficient designs, and modelling approaches which better reflect process function will do even more to bridge the gap between engineer and statistician. He specifically recommends replacing the product array with a single array integrating control and noise factors (i.e. the combined array) and modelling the process response rather than a summary measure such as the SNR. He furthermore recommends that the experimenter consider interaction

plots, Bayesian plots, and response surface methods to help pinpoint process settings which reduce performance variability.

Leon et al (1993) emphasize the crucial nature of choosing the appropriate response, control and noise variables and experimental design . They argue the best response is one that has a clear economic and mechanistic relationship to the process being optimized. The control and noise variables must encompass the relevant influences on process performance and the experimental design must allow for the modelling of these influences. Even though Taguchi's parameter design methods work in many cases, situations abound in which modelling a response other than the SNR is clearly superior.

Tsui (1996) compares Taguchi's loss model approach with that of the response model approach. Since the loss model approach presupposes a product array design, it gives the experimenter less flexibility to pick a more efficient design. More important is that while the model for the SNR may be a linear function of the control factors, the process variance may be a quadratic function of the same. In highly fractionated experiments this may create bias in the effects estimates which may lead to non-optimal choices of control factor settings. Only in circumstances where there are few or no significant control-noise interactions is this bias avoided.

In contrast the response model approach gives the experimenter more leeway in choosing experimental design and allows the modelling of the response as a function of both control and noise variables. This additional information on the influence of the noise factors enables exploitation of the control-noise interactions through simple graphs.

Compared to the loss model approach however, the response model suffers from a critical dependency on the accuracy of the fitted model. With the response model approach the experimenter bears more responsibility for correctly identifying significant control-noise interactions at the risk of increasing process variability. In summary the two approaches should yield similar results when the effects of the loss model approach are not seriously biased and the response model is a good fit.

Table 2.1: Factors in the Injection Molding Experiment

Control Factors	Noise Factors
A: cycle time	M: percentage re-grind
B: mold temperature	N: moisture content
C: cavity thickness	O: ambient temperature
D: holding pressure	
E: injection speed	
F: holding time	
G: gate size	

2.2 Design of Experiments in RD

In this section we present the views of a number of authors on how to decide whether a combined array or a product array is the best choice for conducting an investigation.

2.2.1 Product Arrays vs. Combined Arrays

We start by presenting an injecting molding example originally from Engel (1992). This case was chosen because it is a small, practical example which serves here to contrast the two types of arrays and has been analyzed in Steinberg and Bursztyn (1994) and Tsui (1996). This experiment consists of testing seven control factors (A,B,C,D,E,F,G) and three noise factors (M,N,O) which are described in Table 2.1.

In Table 2.2 we see the product array consisting of the inner (control) array with seven columns and the (noise) array represented by the three rows corresponding to the noise factors. The control array consists of a 2^{7-4} saturated fractional factorial design and the noise factor array consists of a 2^{3-1} saturated fractional factorial design. The entire experiment consists of a total of 2^{7-4} x $2^{3-1} = 32$ test runs.

The structure of the table reveals that for each of the four combinations of noise factor settings, the entire control array is repeated, yielding four response measurements for each test combination of the seven control factors. Using Taguchi methods, a summary statistic called the signal-to-noise ratio (SNR) would be calculated from the four responses measured at each of the control array test combinations. The data from a product array would then be analyzed with the LMA as described in Kacker (1985). Later in this paper we will show

Table 2.2: Injection Molding Experiment Design and Data

Run	Control Array							Noise Array			
							M	-1	-1	1	1
							N	-1	1	-1	1
							O	-1	1	1	-1
	A	B	C	D	E	F	G	Data			
1	-1	-1	-1	-1	-1	-1	-1	2.2	2.1	2.3	2.3
2	-1	-1	-1	1	1	1	1	0.3	2.5	2.7	0.3
3	-1	1	1	-1	-1	1	1	0.5	3.1	0.4	2.8
4	-1	1	1	1	1	-1	-1	2.0	1.9	1.8	2.0
5	1	-1	1	-1	1	-1	1	3.0	3.1	3.0	3.0
6	1	-1	1	1	-1	1	-1	2.1	4.2	1.0	3.1
7	1	1	-1	-1	1	1	-1	4.0	1.9	4.6	2.2
8	1	1	-1	1	-1	-1	1	2.0	1.9	1.9	1.8

Table 2.3: Variance-Effect Estimates and Their Aliases

Effect Estimates	Aliased Effects			
	m.e.	2-f.i.		
-.030	A	BC	DE	FG
.028	B	AC	DF	EG
.055	C	AB	DG	EF
-.027	D	AE	BF	CG
-.056	E	AD	BG	CF
.934	F	AG	BD	CE
.028	G	AF	BE	CD

how the data from the product array can also be analyzed with another technique called the response model approach, which is an alternative to Taguchi's recommended LMA.

The LMA assumes a model consisting only of control factors based on the presumption that the experimenter has chosen these factors so that the control-control factor interactions are negligible. Hence although each control factor is aliased with three control-control interactions (see Table 2.3), from this grouping of aliased effects only the individual control factor main effects are believed to influence the response.

Table 2.4 presents a combined array where the control and noise variables are columns within the same design matrix. This design is a 2^{10-5} fractional factorial design and consists of a total of 32 test runs. This particular design takes only one replicate of each combination

of the variables in order to maintain the same number of test runs as in the product array. When the experimenter gathers data with this type of design, he can not use the LMA for modelling since there is no repetition of all noise conditions at each combination of control factor settings as in the product array. With this experimental design the experimenter is restricted to the response model approach.

The data column is empty because the relevant experiment was only run as a product array. We remind the reader that this is only one possible combined array design and is placed here to illustrate its structural differences with the product array. There are many possible designs from which to choose, depending on the experimenter's experimental goals and knowledge of the process.

The response model approach assumes that the response can be modelled as a function of control and noise variables as well as the interactions between them. The experimenter chooses the specific combined array according to her best understanding of possible main effects and interactions that may influence the response.

The traditional practice in classical design of experiments is to pick a Resolution IV or higher design so that individual factors are aliased with three factor interactions, of which there are relatively few known physical examples. A Resolution V is even more desirable since it aliases the two factor interactions, of which many examples exist in nature, with three factor interactions. The general rule is that the experimenter is reasonably safe from missing important effects if the main effects and two factor interactions are only aliased with higher order interactions.

Table 2.5 shows the aliasing of the main effects from the combined array shown in Table 2.4. Since Table 2.4 is a Resolution IV design, the individual control and noise variables (i.e. main effects) are each aliased with three-factor interactions. Hence for estimating the effects of individual control and noise variables, the combined array of Table 2.4 is a better data gathering instrument than Table 2.3. This is because the main effects are aliased with three factor interactions rather than two factor interactions as in the product array.

However the estimation of main effects is not necessarily the best way to judge the value of a test design for RD. The control-noise interactions are generally regarded as having

Table 2.4: Injection Molding Experiment in Combined Array

Run	Combined Control and Noise Array										
	A	B	C	D	E	F	G	M	N	O	Data
1	-1	-1	-1	-1	-1	1	1	1	1	1	***
2	-1	-1	-1	-1	1	-1	-1	-1	-1	1	***
3	-1	-1	-1	1	-1	-1	-1	-1	1	-1	***
4	-1	-1	-1	1	1	1	1	1	-1	-1	***
5	-1	-1	1	-1	-1	-1	-1	1	-1	-1	***
6	-1	-1	1	-1	1	1	1	-1	1	-1	***
7	-1	-1	1	1	-1	1	1	-1	-1	1	***
8	-1	-1	1	1	1	-1	-1	1	1	1	***
9	-1	1	-1	-1	-1	-1	1	-1	-1	-1	***
10	-1	1	-1	-1	1	1	-1	1	1	-1	***
11	-1	1	-1	1	-1	1	-1	1	-1	1	***
12	-1	1	-1	1	1	-1	1	-1	1	1	***
13	-1	1	1	-1	-1	1	-1	-1	1	1	***
14	-1	1	1	-1	1	-1	1	1	-1	1	***
15	-1	1	1	1	-1	-1	1	1	1	-1	***
16	-1	1	1	1	1	1	-1	-1	-1	-1	***
17	1	-1	-1	-1	-1	1	-1	-1	-1	-1	***
18	1	-1	-1	-1	1	-1	1	1	1	-1	***
19	1	-1	-1	1	-1	-1	1	1	-1	1	***
20	1	-1	-1	1	1	1	-1	-1	1	1	***
21	1	-1	1	-1	-1	-1	1	-1	1	1	***
22	1	-1	1	-1	1	1	-1	1	-1	1	***
23	1	-1	1	1	-1	1	-1	1	1	-1	***
24	1	-1	1	1	1	-1	1	-1	-1	-1	***
25	1	1	-1	-1	-1	-1	-1	1	1	1	***
26	1	1	-1	-1	1	1	1	-1	-1	1	***
27	1	1	-1	1	-1	1	1	-1	1	-1	***
28	1	1	-1	1	1	-1	-1	1	-1	-1	***
29	1	1	1	-1	-1	1	1	1	-1	-1	***
30	1	1	1	-1	1	-1	-1	-1	1	-1	***
31	1	1	1	1	-1	-1	-1	-1	-1	1	***
32	1	1	1	1	1	1	1	1	1	1	***

Table 2.5: Combined Array Main Effects and Their Aliases

Aliased Effects				
Main Effects	3-Factor Interactions			
A	EFO	DFN	CFM	BFG
B	EGO	DGN	CGM	AFG
C	EMO	DMN	BGM	AFM
D	ENO	CMN	BGN	AFN
E	DNO	CMO	BGO	AFO
F	EAO	DAN	CAM	BAG
G	EBO	DBN	CBM	BAF
M	ECO	DCN	CBG	CAF
N	EDO	DCM	DBG	DAF
O	EDN	ECM	EBG	EAF

equal importance as the control effects for fine tuning the final control factor settings for minimal product variation. Hence evaluation of an experimental design for RD purposes must take into account the design's ability to estimate the control-noise interactions deemed most likely to affect product performance.

Having illustrated this example with both product and combined arrays, we proceed by reviewing the comments of some of the leading experts in RD on the pros and cons of using product and combined arrays.

Shoemaker et al (1991) point out that the product array dictates estimation of all two factor control interactions and higher order generalized control-noise interactions and is intended only for main effects models. The combined array allows the experimenter to choose interactions to be estimated and allows for models with main effects, two factor interactions and control-noise interactions which can be exploited towards reducing response variation. They state that significant runs savings are possible with the combined array due to the flexibility it affords experimenters in the estimation of effects.

In the following paragraphs we summarize comments from some participants in Nair (1992) regarding choice of experimental design.

Wu justifies Taguchi's choice of mixed level orthogonal arrays as a means of ensuring run-size economy. The simplest practical 3^{k-n} design has 27 runs while Taguchi's L18(3^7) has eighteen runs. He defends the use of linear graphs as a user-friendly way of capturing

solutions for non-statisticians, who may well be intimidated by aliasing tables. He stipulates that the Taguchi designs are appropriate only for simple, small applications and cautions that the Taguchi Resolution III designs represented by linear graphs guarantee neither maximum resolution nor minimum aberration.

Sacks and Welch assert that a single experimental array requires fewer runs and is therefore more practical for experimenters. Box echoes this by stating that the product array contains an excessive number of test runs.

Lorenzen makes a detailed comparison of the two types of arrays with regards to several criteria. He finds the product array more intuitive, a significant advantage for wider public employment. He finds the combined array less intuitive and dislikes its need to estimate missing noise combinations. He compares the designs in terms of detectability and robustness. He finds the two array types comparable in detectability, whereas using his own ad-hoc robustness measure to examine a number of cases, he finds the combined array superior for robustness. He balances this by saying that if the cost of each control factor combination is expensive, the combined array may be more expensive than the product array.

Kacker equates use of the combined array to classical regression decomposed into control and noise variables, albeit without satisfying the homogeneous variance requirement. In his experience product engineers can not be relied upon to prioritize important control-noise interactions, a working assumption of the combined array. He furthermore states that regression based on the combined array is more sensitive to missing data than the product array.

A final observation from Nair (1992) is that the product array is the more intuitive arrangement, while the combined array, even with replication often demands a smaller number of test runs. A few more references complement Nair's comprehensive discussion.

Ghosh and Derderian (1993) derive robustness measures for RD experimental layouts. They do this for both product and combined arrays, thereby allowing the experimenter to objectively decide whether the product or combined array provides a more robust option.

Lucas (1994) notes that Taguchi's product arrays can be considered response surface designs and compares them to standard and mixed resolution composite designs. Lucas

ties together Taguchi analysis and response surface analysis with an argument based on propagation of error. He remarks that a product array is size efficient when both inner and outer arrays are saturated. He goes on to show how a mixed resolution composite design has even smaller size by allowing the factorial portion to be Resolution III and using only star points for the control variables. He recommends the composite design using the smallest fraction of (2^p) with Resolution V or using a fractional factorial of mixed resolution.

When Taguchi and composite designs are of similar size, he recommends composite design for its ability to estimate interactions among control factors as as control-noise interactions. He concludes that the use of classical, statistically designed experiments can achieve same or better results than Taguchi's product arrays.

Finally, Wu and Hamada (2000) provide an intuitive approach to choosing between product and combined array based on an effect-ordering principle.

They judge an array by its estimation capacity, which is the number of strongly clear (i.e. not aliased with main effects or two or three factor interactions), clear (i.e. not aliased with main effects and two factor interactions) and eligible effects (i.e. not aliased with main effects) it can estimate.

They list the most important class of effects as that containing control-noise interactions, control main effects and noise main effects. The second highest class contains the control-control interactions and the control-control-noise interactions while the third and least important class contains the noise-noise interactions. Each type of interaction within a class grouping is considered of equal importance. The array which produces the highest number of clear effect estimates in the most important class is considered the best design.

Noting that the combined array is often touted as being more cost effective due to an implied smaller number of runs, Wu and Hamada (2000) place the cost comparison on a more objective basis by factoring in both cost per control setting and cost per noise replicate. They conclude the experimenter must prioritize the effects to be estimated and the realistic costs involved before deciding which type of array is optimal.

2.2.2 Designs for RD

The prior section focused on how to decide whether a product or combined array would best serve the experimenter's purposes. The common bond between product and combined arrays is that they both consist of orthogonal arrays. In this section we review two references which help the experimenter choose which specific orthogonal array to select as the building block for a combined or product array.

2.2.2.1 Standard Orthogonal Arrays and Related Tools

Tsui (1988) introduces the confounding table, i.e. a small chart showing which effects in an orthogonal array are confounded, as a general method for planning experiments. Because of their ability to represent more than two levels, the tables are applicable to a much wider selection of experiments than the linear graphs of Taguchi or the interaction graphs of Kacker and Tsui (1987). He restricts his examples to two and three level fractional factorials and defines the requirement set as a group of main effects and two-factor interactions that must remain unconfounded in the experiment. He gives detailed instruction for how to assign factors to specific columns in the orthogonal arrays, while honoring the requirement set. Two additional advantages of confounding tables is that they are not limited to any specific size of orthogonal array and can handle multiple and mixed level experiments as well.

Kacker and Tsui (1990) formally present the work first documented in Kacker and Tsui (1987), featuring a graphical technique for showing the confounding pattern of effects within a two level fractional factorial. The interaction graph is a graphical construct that identifies all interaction relationships among the columns of an orthogonal array (hereafter OA) for two level fractional factorials of sixteen runs or less. They provide step by step instructions for assigning factors to columns of the OA and list three major advantages of the interaction graphs when compared to the linear graphs of Taguchi.

The first advantage is providing complete information regarding confounding of interactions whereas linear graphs provide no information in this regard. Second is that linear graphs can always be derived from an interaction graph, while the converse is not true. Lastly is that interaction graphs can always be used to construct fractional factorial plans

of resolution IV. This simple graphical technique retains the user-friendly appeal of the linear graph while providing the experimenter with valuable confounding relationships.

Kacker et al (1991) define orthogonal arrays and describe how Taguchi's fixed element arrays are related to well known fractional factorial designs. The authors also review how Taguchi's mixed element orthogonal arrays are constructed.

2.2.2.2 Designs for Combined Arrays

Rosenbaum (1996) reinforces the efficiency claims of the combined array by giving a number of combined array designs which are smaller for a given orthogonal array strength or stronger for a given size. Hedayat and Stufken (1999) present a catalog of maximum possible strength two level combined array designs for a given number of factors and test runs. They use defining contrasts to obtain these arrays.

2.2.2.3 Designs for Product Arrays

For the experimenter inclined to use Taguchi's product array structure, there are many design choices available. These include the classic Taguchi arrays and several other variants which try to compensate for the oft cited complaint of excessive test runs. A trend in RD over the last ten years is for RSM practitioners to demonstrate how established RSM designs and analyses can be used to accomplish the goals of RD more precisely and efficiently than with the tools recommended by Taguchi.

Box and Jones (1992) point out that Taguchi's product array structure of inner (control) and outer (noise) arrays is a specific case of the split-plot design prominent in RSM. In split plot designs, a classification factor is one whose effect on the variance of the response is important to account for without need for precise estimation. The classification factor is confounded with the higher level block, i.e. the whole plot, while the other type of factor, the split plot factor, is confounded with partitions of the whole plot, i.e. the split plots. The experimenter usually has greater interest in reliably estimating the effect of the split plot factor on response variance.

Box and Jones assert that respectively placing the noise and control variables in whole and split plots maximizes the precision of the the control and control-noise effects, i.e. those

most important in correctly determining optimal control variable settings. This comes at the cost of less precise estimates of the noise effects. For RD purposes, they advocate an analysis approach which quantifies control by noise interactions to enable the experimenter to pick optimal control variable settings.

In his product array designs, Taguchi places the control variables in the whole plots and the noise variables in the split plots. This testing structure sacrifices information by spending its best estimation power on the effects of the noise variables rather than the control variables and the control-noise interactions.

Running the split plot per the recommendation of Box and Jones provides two major advantages for the RD experimenter. They are the valuable information gained on control-noise interactions and the potential of increased run efficiency. Using the split plot design in this fashion provides an attractive, alternative means of running the product array for the RD experimenter.

Borkowski and Lucas (1997) introduce a new class of product array designs called composite-mixed resolution designs (hereafter CMR). These are of minimal Resolution V in the control array and minimal Resolution III in the noise array. Size comparisons between CMR designs and Taguchi product array designs show CMR ranging from superior to competitive and D and G optimal efficiencies of the designs are included.

Bingham and Sitter (1999) introduce an algorithm that efficiently constructs minimum aberration two level fractional factorial split plot designs. They state that the algorithm is readily modified to produce fractional factorial and fractional factorial split plot designs in which the number of levels is the power of a prime. This enables the RD experimenter willing to consider the split plot structure great flexibility in choice of experimental design.

Bisgaard (2000) continues the work of Box and Jones (1992) by providing specific technical detail on the design and analysis of split plot designs based on two level fractional factorials. He stresses that choosing a split plot design must be driven by constraints on randomization or resources, not necessarily the need to simultaneously study control and noise variables. He seconds the recommendations of Box and Jones (1992) that control variables be placed in split plots and noise variables in whole plots to exploit the more

precise estimates of control-noise interactions.

He states that Taguchi style product array designs can be run as fully randomized experiments or with restrictions on randomization and shows how this style of execution enables labor saving strategies. He shows how to reduce the number of experimental runs by providing rules for finding the defining contrast from the generators of individual design matrices with and without split plot confounding. He concludes by giving guidelines on how to choose between competing designs based on cost.

2.2.2.4 D-Optimal Designs

This section provides an additional instance of where RSM practitioners show how RD can be efficiently and precisely executed with their methods. In this section several authors show how D-optimal designs can be exploited in RD experiments.

A D-optimal design minimizes the area of the confidence ellipsoids for parameters being estimated from an assumed model. A strength of D-optimal designs is their invariance to linear transformation of the terms in the model. The weakness of these computer generated designs is their dependence on the accuracy of the assumed model. Dumouchel and Jones (1994) seek to improve on the ad-hoc RSM methods for making a design less sensitive to model assumptions, such as center points and star-points .

They do this by supplying a theoretical foundation and algorithm for choosing an experimental design robust to model assumptions. They recommend models consisting of "p" primary or assumed factors and "q" potential or tentative terms. Their approach allows precise estimation of the primary terms and at least detectability for the potential terms. The use of prior information on the potential terms allows the experimenter to "hedge bets" on the assumed models and avoid the problem of singular estimation, since the run size "n" is less than the sum of p and q.

The coefficients of primary terms are assumed to have a diffuse prior distribution and potential terms are assumed to have a prior mean of zero and finite variance. Preliminary scaling and centering of the potential terms is necessary in order to choose an appropriate prior distribution. This Bayesian analog to the D-optimal design maximizes the determinant

of the design matrix and requires little modification of most D-optimal search algorithms. The major contribution here is the use of a proper prior distribution to attack the singular design problem and make the design less model dependent.

The authors provide a practical example where data from a known quadratic model is modelled with first order D-optimal, second order D-optimal and Bayesian D-optimal designs. The Bayesian D-optimal design shows greatly improved bias, mean squared error, prediction and lack of fit properties compared to the first order D-optimal design, and comparable results for the second order D-optimal design.

Atkinson and Cook (1995) extend the existing theory of D-optimal design to linear models with non-constant variance. They use the theory to create D-optimal designs for estimating mean, variance or both and show that it applies whether mean and variance are functions of the same or different variables. One common way of dealing with the model dependency of D-optimal designs is by examining a number of locally optimal designs over plausible values of the parameter to be estimated. In the case of significant differences between the locally optimal designs, Atkinson and Cook recommend choosing a compromise design with a Bayesian approach that approximates the pre-posterior loss.

Chang (1997) proposes an algorithm for generating near D-optimal designs for multiple response surface models. This algorithm differs from existing approaches in that it does not require prior knowledge or data based estimates of the covariance matrix to generate its designs. The author uses simulation results to show the designs are near D-optimal and non-dependent on the covariance structure.

For the case of heterogeneous variance, standard central composite designs (hereafter CCD) may not be optimal. Mays (1999) extends the quadratic model methodology of RSM to the case of heterogeneous variance. He uses D (i.e. maximal determinant) and I (i.e. minimal integrated prediction variance) optimality criteria to allocate a given number of test runs to different locations within a CCD design. Several variance structures are considered for two, three and four variable models.

2.2.2.5 Other Designs

The remaining references discuss types of designs used in RD which are not easily classified under the more common categories previously discussed.

Vining and Schaub (1996) attack the problem of finding separate linear models for process mean and variance. They use a two-pronged approach where variance is alternatively considered constant and with an assumed prior distribution. These approaches can be applied to the data from either combined or product array, as long as replication in the combined array or the replicated outer array of the product array provide sufficient data for the proposed variance model. They point out that separate models for mean and variance allow for better understanding of the process and for compromise choices between bias and process variance.

The authors use D-optimality to evaluate competing designs. Their comparison of the designs indicated that replicated fractional factorials of assumed constant variance best estimated the variance while the semi-Bayesian designs typically yield better estimates of the process response. The two approaches yielded designs of comparable performance for estimating the joint model except when the semi-Bayesian prior distribution was completely misspecified.

Pledger (1996) divides noise variables into observable and unobservable and argues that one's ability to observe selected noise variables in production should translate into better choices of optimal control settings. She describes a technique for introducing observable and unobservable noise factors into an experimental design to take advantage of this available information.

Rosenbaum (1999) focuses on the use of blocking in dispersion experiments, which he defines as an attempt to estimate the effects of control factors on the variance introduced by noise factors. He uses blocking to separate the control and noise variables in combined arrays, which he showed in Rosenbaum (1996) to be stronger for a given size than the corresponding product array designs.

Variance within blocks of noise variables that form an OA of strength 1 may be used to estimate dispersion effects. He builds a family of blocked designs of strength 3 by using

Hadamard matrices as within block arrays which are then completed and folded. Confounding and aliasing of these designs are presented. These designs provide for unbiased estimates for the main effects of control factors on variance introduced by noise factors under linear models where noise variables do not intersect.

Li and Nachtsheim (2000) present experimental designs which don't depend on the experimenter's prior determination of which interactions are most likely significant. They construct model robust factorial designs (hereafter MRFD) which are efficient for all models containing main effects and a pre-specified number of interactions. They compare these designs to the standard offerings from maximum resolution fractional factorials for sizes of 8, 12 and 16 and find them more robust to model mis-specification though not completely orthogonal.

They justify these designs with the sparsity of effects principle, assuming experimenters are concerned with a small number of unknown interactions. They recommend the MRFDs as alternatives to Resolution III and IV orthogonal designs in this situation. This design eliminates the need for the experimenter to choose a confounding scheme. They note that if the number of significant interactions exceeds the pre-specified number, then the designs are not robust.

2.3 Analysis of Experiments in RD

2.3.1 Optimization and Choice of Performance Measure

2.3.1.1 Signal-Noise-Ratio

When planning an experiment in RD, the experimenter must think through several things very carefully. We spoke earlier about different ways of deciding whether or not to use the product array advocated by Taguchi, or the combined array first formally discussed in Welch et al (1990) and extended in Shoemaker et al (1991). A related issue of equal importance is choosing which performance measure to model for achieving the desired optimization goal. Taguchi models the SNR to obtain control factor settings that satisfy the optimization goal of minimal quadratic loss.

The job of a performance measure is to tell how a process is functioning. The advantages

of the SNR advocated by Taguchi are twofold. First it has great intuitive appeal because it is measured at each combination of control factors and secondly is a concurrent statistic, i.e. it incorporates measures of location and dispersion into one value. Combined with the simplicity of the two step procedure, its intuitive appeal has been a powerful incentive for engineers and scientists to foray into the world of designed experiments while raising the quality of their products. Bridging the worlds of engineering and applied statistics is generally acknowledged as the greatest contribution of Taguchi.

Conversely the statistical community has been troubled by the unexplained connection between maximizing the SNR and achieving the stated goal of minimal quadratic loss. Recall that the two step procedure consisted of maximizing the SNR and then using control factors which have negligible effect on the SNR to bring the process mean to target. Taguchi did not rigorously spell out how this procedure guaranteed minimal quadratic loss.

The SNR is defined as:

$$SNR \text{ or } \eta = 10log_{10}\frac{E^2Y}{VarY} \tag{2.1}$$

A common choice of loss function is the quadratic loss function popularized by Taguchi,

$$L(Y,t) = A_0(Y-t)^2 \tag{2.2}$$

where A_0 is a constant. The average loss may be decomposed as:

$$\begin{aligned} R(\mathbf{C}) &= A_0E_{N,\epsilon}(Y-t)^2 \\ &= A_0[VAR_{N,\epsilon}(Y) + (E_{N,\epsilon}(Y)-t)^2] \end{aligned} \tag{2.3}$$

where $VAR_{N,\epsilon}(Y)$ and $E_{N,\epsilon}(Y)$ are the mean and variance of the response of the manufacturing process over random noise $\mathbf{N}$ and ϵ, which we refer to as the process mean and process variance in this paper. If the quadratic loss is believed to be a good approximation, the objective of robust design is to minimize the sum of the process variance and the square of the separation of process mean from the target.

The next few paragraphs demonstrate how modelling process response as a performance measure can lead to ways of modelling process mean and variance. From the decomposition of quadratic loss into variance and off-target bias squared, we see a direct relation to how minimizing variance and shifting the mean will optimize the average quadratic loss. We start by modelling the response as a function of the control and noise variables as follows:

$$Y = f(\mathbf{C}, \mathbf{N}) + \epsilon, \tag{2.4}$$

where $\mathbf{C} = (C_1, ..., C_p)^T$, $\mathbf{N} = (N_1, ..., N_q)^T$, and f is a transfer function. The control factors $\mathbf{C}$ are assumed to be fixed since they are controllable during production. Although the noise factors $\mathbf{N}$ are assumed to be random during production, they are often treated as fixed during the experiment to increase the efficiency of the experiment. The pure error ϵ represents the remaining variability of the manufacturing process that is not captured by the noise factors. We assume that the $\epsilon's$ are independently and identically distributed with mean zero and variance σ^2 and are independent of the noise factors.

Below we consider a simple approximation of the transfer function, an additive model (i.e. no control-control interactions) with control-by-noise interactions, to illustrate how we can fix the noise factors in the experiment to estimate the process mean and variance during production. Suppose the response Y for fixed noise factors in the experiment can be expressed by the following model:

$$Y|\mathbf{N} = \alpha_0 + \boldsymbol{\alpha}^T\mathbf{C} + \sum_{j=1}^{q}(\gamma_j + \boldsymbol{\beta_j}^T\mathbf{C})N_j + \epsilon \ , \tag{2.5}$$

where $Y|\mathbf{N}$ is the conditional random variable of Y given $\mathbf{N}$, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)^T$ and $\boldsymbol{\beta_j} = (\beta_{j1}, ..., \beta_{jp})^T$.

As mentioned earlier, $\mathbf{N}$ is treated as fixed in the experiment although it is random during on-line production. It follows that the conditional mean and variance for fixed noise are $E(Y|\mathbf{N}) = \alpha_0 + \boldsymbol{\alpha}^T\mathbf{C} + \sum_{j=1}^{q}(\gamma_j + \boldsymbol{\beta_j}^T\mathbf{C})N_j$, and $VAR(Y|\mathbf{N}) = \sigma^2$.

Suppose the $N_j's$ are independently distributed with mean zero and variance ${\sigma_{N_j}}^2$ during production. In order to use equation 2.5 to estimate the process mean and variance, we apply the following conditional expectation relationships:

$$E_{N,\epsilon}(Y) = E_N[E(Y|\mathbf{N})] = \alpha_0 + \boldsymbol{\alpha}^T\mathbf{C}, \tag{2.6}$$

$$\begin{aligned} VAR_{N,\epsilon}(Y) &= VAR_N[E(Y|\mathbf{N})] + E_N[VAR(Y|\mathbf{N})] \\ &= \sum_{j=1}^{q}(\gamma_j + \boldsymbol{\beta_j}^T\mathbf{C})^2{\sigma_{N_j}}^2 + \sigma^2. \end{aligned} \tag{2.7}$$

As mentioned above, equations 2.6 and 2.7 are good approximations of the true process mean and process variance only if equation 2.5 continues to be a good approximation of the true response model during on-line production where noise is random.

This discussion illustrates the important connection between the choice of performance measure and achievement of the optimization goal. Taguchi has given us a two step process to follow without justifying how it achieves minimal quadratic loss. Using the response as a performance measure allows modelling of process mean and variance in many cases, but these models are only as useful as the accuracy of the response model.

The choice of performance measure and achievement of the optimization goal are inextricably linked. The ideal performance measure will lend itself to a simple procedure, such as the two step procedure, which decomposes the optimization problem into simpler tasks. In the remainder of this section we present the thoughts of a number of authors on how an experimenter can choose the best performance measure for the optimization goal at hand.

Nair and Pregibon (1986) recommend choosing separate measures of location and dispersion based on system response. This typically implies the use of mean and variance but advise the experimenter to consider other measures such as median (location) and interquartile range (dispersion) which are more robust to outliers. They advocate the use of transformations to separate the location and dispersion effects and the supplementation of ANOVA with a number of graphical techniques allowing identification of the two types of effects.

Leon et al (1987) investigated how and when Taguchi's two step process actually achieves its stated optimization goal of minimal quadratic loss. Remember that Taguchi's first step consisted of maximizing the SNR and the second of moving the process mean to target. The second step of Taguchi's process assumes the existence of one or more control factors

that affect location without affecting dispersion. These control factors are referred to as adjustment factors.

Leon et al (1987) (hereafter LSK) show that when the system performance is of a form where the ratio of mean to sigma is fixed, Taguchi's two-step procedure does indeed minimize average quadratic loss. However for other underlying models the two step procedure may not achieve minimal average quadratic loss. LSK suggest an alternative to the SNR, the performance measure independent of adjustment (PerMIA). The PerMIA consists of control variables that affect dispersion and location and assumes the existence of other control factors (i.e. adjustment factors) which affect location and are independent of the PerMIA. A PerMIA can be found for models other than proportional mean and sigma and for loss functions other than the quadratic loss. The SNR then becomes a special case of the PerMIA.

When an adjustment parameter exists, a PerMIA can be derived directly from knowledge of the loss function and the general form of the performance measure model. The PerMIA generalizes the approach of the SNR so that the experimenter may take advantage of the simplifying two-step decomposition of the optimization problem for a wider range of models and loss functions.

Box (1988) agrees with Leon et al (1987) that the SNR is only appropriately used in conjunction with models where process sigma is proportional to process mean. In that case maximization of the SNR would achieve minimal quadratic loss by allowing adjustment of process mean to target.

Box promotes the use of transformations which, for common non-proportional relationships between process sigma and mean, would allow for the achievement of minimizing loss while moving mean to target. He recommends a lambda plot of the data to indicate the transformation likely needed to allow a functional separation of process mean and sigma. He argues that the SNRs for the STB and LTB cases are inadequate summaries of data and extremely inefficient measures of location.

He stresses the importance of distinguishing between two issues in RD. First is choosing an appropriate performance criterion and second is finding the best way to estimate it. He

feels that experimentation, use of simple graphical techniques and sequential testing are the best ways to solving the RD challenge instead of the use of rigid, preordained performance criteria such as the SNR.

In their comments on Box (1988), Shoemaker et al (1987) (hereafter STL) praise Box's use of the Lambda plot to find the transformation of variance such that adjustment factors independent of the transformed variance can be found. These adjustment factors allow the process mean to be moved to a target position for the minimization of quadratic loss. However STL assert that transformation may be unnecessary and prefer that adjustment of the mean take place on the untransformed data scale when possible.

Leon and Wu (1992) extend the PerMIA of LSK to a maximal PerMIA which can solve constrained minimization problems in a two step procedure similar to that of Taguchi. They defend the use of two-step procedures for their ability to model PerMIAs as a function purely of non-adjustment factors, which remain the same when the target is changed, as well as their transformation of constrained optimization problems into unconstrained ones. They also derive a theorem for finding maximal PerMIA for quadratic loss functions.

They cite the fallacy of assuming a quadratic loss function and mention asymmetric losses around a target as one instance of where the quadratic loss function is inappropriate. For non-quadratic loss functions they introduce general dispersion, location and off-target measures while developing a two step process. This general loss function two step procedure includes identification of adjustment functions, which they define as a function of design factors used to make adjustments of process location. Though Taguchi used process mean as an adjustment function, they point to the median as a possible alternative. They apply these new techniques in a number of examples featuring additive and multiplicative models with non-quadratic loss functions.

Ng and Tsui (1992) argue that quality loss (QL) should be expressed as a continuous function of a product characteristic. In order to model QL in this way, they recommend implementing a more accurate, complete and customer-oriented measure of yield. The current definition of yield equals the percent of process units passing final inspection. This implies that all passed units are equally acceptable to the next-in-line customer.

To that end they derive a measure called q-yield which accounts for amount of nonconforming units as well as variation from target among passed units. It does this by penalizing yield commensurate with the amount of variation measured within the passed units. The q-yield can be easily generalized to a general power function and can be extended to asymmetric tolerance situations. The minimization of this loss function would work to improve the quality seen by the next customer, rather than focus on getting more units to pass inspection.

Tsui and Li (1994) establish a two step (or multi-step) procedure for the STB and LTB problem based on the response model approach under certain conditions. They also show that the optimal solution from this two step (or multi-step) procedure is invariant to the choice of loss function from within the class of loss functions proposed by Box and Jones (1990).

Box and Jones (1990) considered the following average loss function as an overall measure of robust performance:

$$R_\lambda(\mathbf{C}) = A_0[\lambda\ (E(Y) - t)^2 + (1 - \lambda)\ VAR(Y)] \tag{2.8}$$

where $0 \leq \lambda \leq 1$ and $t = 0$ for the STB problem. Note that this average loss function is appropriate for the NTB and STB problems.

The procedure of Tsui and Li (1994) reduces the dimension of the optimization problem and allows future elimination of noise factors without re-optimization. They derive the procedure for the STB problem by first assuming that the response models for process mean and variance take the following forms:

$$E_{N,\epsilon}(Y) = E_N[E(Y|\mathbf{N})] = \alpha_0 + \boldsymbol{\alpha}^T\mathbf{C}, \tag{2.9}$$

$$\begin{aligned} VAR_{N,\epsilon}(Y) &= VAR_N[E(Y|\mathbf{N})] + E_N[VAR(Y|\mathbf{N})] \\ &= \sum_{j=1}^{q}(\gamma_j + \boldsymbol{\beta_j}^T\mathbf{C})^2{\sigma_{N_j}}^2 + \sigma^2. \end{aligned} \tag{2.10}$$

The two-step procedure consists of first minimizing process mean with respect to C1 and, while fixing the values of C1 derived in step one, minimizing process variance with respect to C2. They use arguments and proofs from Tsui (1993) and Leon and Wu (1992)

to show that the solutions obtained through this procedure minimize average quadratic loss. They extend the two step procedure into a multi-step procedure which divides process variance into contributions from bias, noise variables and pure error. They also demonstrate the procedural solution to be invariant to loss functions in the class introduced by Box and Jones (1990).

Anderson and Wu (1996) is one of the very few examples in the literature dealing with directional dispersion. They examine techniques for analyzing the influence of experimental factors on the dispersion of a directional response located on a unit circle. Several dispersion measures are considered and their relationships explored. Transformation of angular dispersion into a statistic measured on a linear scale is accomplished via circular variance (itself a measure of dispersion). After transformation analysis techniques are employed to determine individual factor influences on directional dispersion. This fits into RD by considering how noise and control variables are treated in this analysis.

Tsui (1996) generalizes Taguchi's two step procedure for the NTB problem into a multi-step procedure based on the response model approach. The procedure is derived, illustrated and shown to be invariant under the class of loss functions introduced by Box and Jones (1990).

The procedure is based upon decomposing average quadratic loss into process bias squared and variance as follows:

$$\begin{aligned} R(\mathbf{C}) &= A_0 E_{N,\epsilon}(Y-t)^2 \\ &= A_0[VAR_{N,\epsilon}(Y) + (E_{N,\epsilon}(Y)-t)^2] \end{aligned} \tag{2.11}$$

where Var(Y) is further decomposed into

$$\begin{aligned} VAR_{N,\epsilon}(Y) &= VAR_N[E(Y|\mathbf{N})] + E_N[VAR(Y|\mathbf{N})] \\ &= \sum_{j=1}^{q}(\gamma_j + \boldsymbol{\beta_j}^T\mathbf{C})^2{\sigma_{N_j}}^2 + \sigma^2. \end{aligned} \tag{2.12}$$

Here process variance is decomposed into q + 2 individual variance components contributed by q noise variables, bias squared and pure error variance. The contributions from the q noise variables serve as individual variance tuning factors, allowing tradeoff between bias and individual noise factors.

Moorhead and Wu (1998) state that although RD analysis techniques nearly always assume a quadratic loss function, in cases where loss is not quadratic the two step procedures of Taguchi and Tsui (1996) for NTB problems are not applicable. Here Moorhead and Wu develop modelling and analysis strategies for a general loss function where the quality characteristic follows a location-scale model. Their procedure adds a third step to the traditional two step process, an adjustment step which moves the mean to the side of the target with lower cost. Hence the optimal solution for the general loss function moves the final parameter settings in the direction of lower cost. Although limited by its need for a location-scale model, it applies to general loss functions and builds upon the familiar two step procedures used for quadratic functions.

Maghsoodloo (1990), derives and tabulates exact mathematical relationships between Taguchi's STB and LTB performance measures and his quality loss function. For the NTB case he provides an inequality between the SNR and the quality loss function. This paper provides more rigorous mathematical justification for some of Taguchi's methods.

Maghsoodloo and Lee (2000) deal with a situation common in many manufacturing processes, that of asymmetric tolerances for quality characteristics. Taguchi's societal quality loss is not minimized in this situation by the common design practices of choosing the smaller tolerance for both sides or setting process mean in the middle of the tolerances. The authors consider linear and quadratic loss functions for determining an optimal process mean which minimizes the expected value of the quality loss function.

2.3.1.2 Dual Response Approach

Taguchi's loss model approach (LMA) models a single, summary performance measure, the signal-to-noise ratio (SNR). A dual response approach is one that models two separate response functions, typically process mean and variance, and only works well when the two responses are functionally independent.

Vining and Myers (1990) discuss the use of a dual response approach within a RSM framework to provide a sequential platform for conducting RD. They borrow from Myers and Carter (1973) by designating the responses as primary and secondary and seek to

optimize the primary response subject to an acceptable value of the secondary response as a constraint. They define an appropriate region of interest through screening experiments and follow up with a second order response surface design to identify optimal control settings. While comparable to Kacker's PerMIA, this dual modelling applies more generally since it doesn't rely on a neat separation of control factors into those affecting mean and those affecting variance. This is a more general approach to the NTB situation.

Del Castillo and Montgomery (1993) extend the work of Vining and Myers (1990) by solving the dual response optimization using the generalized reduced gradient (GRG) algorithm, a standard nonlinear programming technique, to achieve the same goals. They claim this method is easier to use and more flexible than the dual response approach and can be applied to multiple response situations.

Lin and Tu (1995) seek to correct some of the deficiencies in Vining and Myers (1990) approach to solving the dual response optimization problem. They optimize via Lagrangian multipliers with a procedure based on mean square error and choose the best subset model rather than consistently adapting the full quadratic model for the primary and secondary response surface. Lin and Tu employ the example featured in Vining and Myers (1990) to compare the two procedures. They claim this method is not limited to polynomial models.

Copeland and Nelson (1996) solve the dual response optimization problem with the technique of direct function minimization. They use the Nelder-Mead simplex procedure and apply it to the LTB, STB and NTB cases.

DelCastillo et al (1997) present an algorithm in Fortran, the DRSALG, which finds a global optimal solution for dual response systems within a spherical region of interest. Its ability to guarantee a global optimum is an improvement over a general non-linear programming algorithmic tendency to converge on local optima.

Fan (2000) introduced the DR2 algorithm which guarantees a global optimal solution for non-degenerate problems and a near global solution for degenerate problems. Fan uses large sample simulation to show its effectiveness in locating near global optimal operating conditions for degenerate dual response problems.

2.3.1.3 Nonlinear Optimization

Whereas conventional non-linear programming techniques can be used to solve the problem of high-low tolerancing when the process transfer function is closed or well-simulated, Fathi and Palko (2000) introduce an approximation procedure for this problem for more general conditions. When the transfer function is not closed or the variables are discrete or categorical, their approximation procedure may be used. Several examples applying this procedure are presented.

2.3.2 Choice of Performance Measure Modelling

The third important decision the experimenter must grapple with is how to model the chosen performance measure. Since the final goal is optimization, the modelling methodology should ideally lend itself to a decomposition of the optimization problem like the two step procedure of Taguchi.

In sections 2.1 and 2.2 we respectively introduced the loss model approach (LMA) and the response model approach (RMA). We start this section by taking the same injection molding experiment and comparing the two modelling approaches by analyzing the response data with each. To demonstrate the LMA, the performance measure process variance was modelled as a function of the seven control variables presented in Table 2.3.

Tsui (1996) states that based on a main effects analysis, only factors C, E and F were found significant. Hence the first step of the two step procedure is readily completed by assigning these control variables the respective values of -1, 1 and -1 since this will minimize the average process variance. Step two of the LMA procedure is completed by finding values of the other variables that shift the process mean to the desired target.

The RMA models the same response data as a function of control and noise variables as well as their interactions. Tsui96JAS uses the method of Lenth (1989) to conclude that only factors A, D, G, C-N and E-N have significant effect on process variance. Including the main effects of each corresponding interaction yields the following model:

$$\begin{aligned} \hat{y} &= 2.25 + 0.425A + 0.063C - 0.282D + 0.144E \\ &\quad - 0.231G + 0.138N + 0.45CN - 0.419EN \end{aligned} \tag{2.13}$$

In order to study the biases of the effect estimates in Table 2.3, Tsui (1996) follows the approach in Box and Jones (1990) to estimate the variance of the response estimator represented by equation 2.13. He assumes for simplicity that the noise variables M, N and O in equation 2.13 are uncorrelated random variables with variances σ_M^2, σ_N^2, and σ_O^2. It follows that the variance of $\hat{y}$ over the noise can be estimated by

$$\begin{aligned} Var(\hat{y}) &= (0.138 + 0.45C - 0.419E)^2\sigma_N^2 \qquad (2.14) \\ &= (0.019 + 0.124C - 0.116E - 0.377CE + 0.203C^2 + 0.176E^2)\sigma_N^2 \end{aligned}$$

which is clearly a quadratic function of the control factors C and E. If equation 2.13 is believed to be adequate, equation 2.14 is a good estimate of the process variance. Thus we can find out the potential bias of the effect estimates in Table 2.3. According to equation 2.14, since the main effect of F is zero and interaction CE is nonzero, the estimate on line 6 (0.934) of Table 2.3 should be an estimate of the interaction CE rather than the main effect F. In other words, the main effect estimate of F is seriously biased with the estimate of CE.

Note that, as shown in Shoemaker and Tsui (1993), since noise factor N interacts with more than one control factor (C and E), the individual interaction plots should not be used to identify the "optimal" factor settings. Steinberg and Bursztyn (1993) have also studied Figure 2.1 and reached the same conclusion as Tsui (1996) regarding the optimal settings for control factors C and E. They have also provided a more complete data analysis including model diagnostics.

This example illustrates both the potential bias introduced into effects estimates by the LMA, and the dependency of the RMA on the goodness of model fit. The remainder of the section reviews selected references from the literature which provide detailed evaluations for choosing which modelling method to pursue.

In one of the first RD papers to advocate direct modelling of the process response instead of the SNR, Welch et al (1990) combined control and noise variables into a single orthogonal array. They modelled the response as a function of both control and noise variables and used the model to predict response as well as estimate the values of the quadratic loss function.

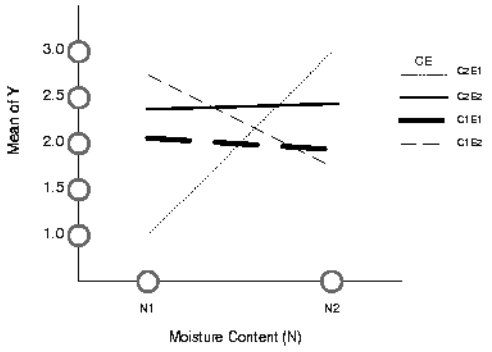

Figure 2.1: Interaction Plot for Cavity Thickness by Injection Speed

The authors employed the direct modelling of response to minimize quadratic loss in the design of VLSI circuits in computer experiments. They compare use of the response model approach with the loss model approach and conclude that in this application better results at lower cost are achieved with the response model approach.

Myers et al (1992) describe Taguchi's experimental designs and use of SNR and compare to the alternative approaches offered by response surface methodology (hereafter RSM). The combined array is compared to the product array and modelling of SNR compared to separate models for mean and variance. Mention is made of the growing number of options for modelling variance, including those presented in Box and Meyer (1986), Nair and Pregibon (1988) and Nelson and Lee (1991). They state that RSM satisfies most of the criticisms made of parameter design and allows use of a single performance measurement in the use of the loss functions of Box and Jones (1990). In addition RSM lends itself to the use of mixed models for random noise variables and fixed control variables.

In short they make the case that RSM provides a widely respected methodology for conducting RD, marrying the best of Taguchi's philosophy with a well established, sequential body of empirical experimentation.

Shoemaker and Tsui (1993) provide a formal basis for the graphical data-analytic approach presented in Shoemaker et al (1991) (hereafter called STW). In STW the response model approach (RMA) was compared to the loss model approach (LMA) of Taguchi and found to often be more efficient and helpful in pointing out control variable settings which would minimize variability caused by noise variables.

In STW they used an example with data from coating silicon on a wafer with several control and noise variables. They showed how using control-noise interaction plots could point out the most robust control factor settings. In this paper they state the conditions when using control-noise interaction plots is appropriate and also show how the c-n plots can actually increase process variability when used inappropriately.

The principal contribution of this paper is a detailed and justified procedure for when and how to use the c-n procedures and a decomposition of the process variance into specific c-n interactions called individual variability measures (IVMs).

When an experiment has high enough resolution, the sample individual variability measures (SIVMs) are unbiased estimators of the IVMs. If the design does not have high enough resolution, the SIVMs can be easily modified to provide an unbiased estimate. When the bias can not be removed, the bias term can be estimated from the estimated covariance matrix of the fitted effects.

They work though an analysis strategy and a constructed example modelling the noise variables as fixed effects and point out how and when to use the c-n plots and SIVMs to optimize robustness to the noise variables. They go on to examine the random effects cases with independence and with correlation between the noise and control variables and how this affects the IVMs.

This paper provides a necessary justification and broadening of the ideas presented in STW for using the RMA.

For the NTB case, Kirmani and Das Peddada (1993) use stochastic ordering to obtain alternative performance criteria to the SNR for finding the optimal control variables sought in RD. They show how stochastic ordering between two probability distributions can be used to compare two competing manufacturing processes. The comparison of the process is represented by the probability distributions of the respective loss functions. For a normally distributed process, the principle of stochastic ordering yields process variance and the non-centrality parameter as the two performance criteria by which to choose control variable settings. They also indicate how multivariate versions of these performance criteria can be found using stochastic ordering.

Tsui (1994) points out that modelling the response variance with the loss model approach (LMA) often creates unnecessary biases in the factorial effect estimates. This in turn may lead to non-optimal choices of the control factor settings. The quadratic terms of the log variance are negligible only when the magnitude of the control-noise interaction is much smaller than that of the noise main effect. Hence only in that case will linearizing the quadratic effects with a log transform effectively remove the bias from the model.

The variance of the response due to noise factors may be quadratic in the control factors even when the response model itself is linear for the same. A detailed example shows how, in

highly fractionated control arrays, the main effects are confounded with control by control interactions when using the LMA.

Tsui applies the response model approach (RMA) to the same problem and shows that for the best twenty combinations (of control factor settings) four of the control factors are not in the same settings as the optimal setting identified by the LMA. The RMA effects estimates are also unbiased. He mentions that since the RMA is more dependent on the adequacy of the model chosen,there is greater need to verify goodness of fit through normality plots, stepwise regression, Mallows' C-p statistics, prediction sum of squares, residual plots and when needed, data transformations.

Using two detailed industrial examples, Steinberg and Bursztyn (1994) show how choice of modelling with the LMA vs. the RMA causes different choices of optimal settings for control factors. The authors reinforce the assertion made in Tsui (1994), that use of the LMA in fractionated factorial control arrays may cause biased noise effects by their confounding with control-control interactions.

The first example is an injection molding experiment and the two methods indicate the same factors as having significant effect on the mean level of shrinkage but very different conclusions with respect to dispersion. They advise watching for separability, i.e. that noise variables interact with a single control factor. When not separable, the control-control interaction must be plotted against the levels of the noise factor that both control factors interact with, to find the best level for reducing variation.

The second experiment is injection molding of appliance handles where the LMA and RMA again produce different results. The significant finding in this example is that with the RMA it became evident early in the analysis that all the data from the first mold consisted of zero values. The RMA was able to separate location and dispersion effects whereas the LMA mixed them.

Khattre (1996) starts with a review of the approach of Myers et al (1992) (hereafter MKV) toward obtaining an appropriate response model that can be used to predict effects of the control and noise factors. He extends Myers et al (1992) by presenting a blocking of noise variables that allows an experiment to proceed when for physical reasons not all noise

variables can be simultaneously achieved.

Assuming noise variances are known or well estimated, the variance can be obtained as a response surface in the control variables. Using RSM and graphical techniques one obtains settings of control variables which are robust to noise variables.

Khattre points out that there exist many physical situations where the noise variables may not be simultaneously present, such as poor road conditions due to ice and high temperatures. He proposes blocking the noise variables in groups which can be simultaneously achieved. He runs two experiments which have the same control factor settings and different noise factors dependent on blocking.

During experimentation the underlying linear model is an assumed fixed effects model, whereas a mixed effects model is needed in to predict mean response and variance. The data from the two blocks is combined into a single layout with missing levels replaced by zeroes. The model coefficients are easily calculated with a regression package from which the variance surfaces are readily obtained. Orthogonality of design leads to significant simplification and can be extended to multiple blocks.

Myers et al (1997) develop and illustrate response surface methods to cover modelling of process mean and variance. The authors provide a methodology for calculating a confidence region on location of control factors for minimum process variance. They also show how mean and variance response surfaces can be combined to create prediction limits on future response and one sided tolerance intervals.

Assuming that noise variables are i.i.d. standard normally distributed, they show how knowledge of the control-noise interactions leads to a confidence region on process variance and prediction limits and one sided tolerance limits on future response values. They propose evaluating these prediction limits and tolerance intervals at competing areas of the design space to find the optimal settings.

2.3.3 Statistical Modelling and Analysis Tools

Earlier in this review paper we discussed how to choose experimental design, performance criteria and the modelling approach of the performance criteria. Typically the choice of

those procedures dictates the appropriate analysis and modelling techniques for determining optimal control variable settings. The references in the following section describe specific analysis and modelling techniques the RD experimenter may use depending on his earlier data gathering choices.

2.3.3.1 ANOVA and Linear Models

Box and Meyer (1986) show how unreplicated fractional factorials can be used to identify location and dispersion effects by using the principle of effect sparsity. They discuss aliasing of the location and dispersion effects and advocate the computation of variances from residuals obtained after least squares modelling of significant location effects. This removes the effects involved in the location models and allows for easier identification of dispersion effects.

After models for location and dispersion have been identified, maximum likelihood estimation can be used to refine the model for a more precise fit. Conditional on dispersion effects, the location effects are calculated by weighted least squares. Iterated cycles of calculated dispersion effects from residuals of location effect models continue until convergence. This procedure could be easily complemented by follow up experiments and hence provides a valuable and efficient screening mechanism for identifying important location and dispersion effects.

Nair and Pregibon (1988) cite Bartlett and Kendall (1946)'s introduction of the commonly used method, based on least squares analysis of the logarithm of within replication variance, to identify important dispersion effects. They secondly mention Box and Meyer (1986)'s pooling technique for unreplicated two-level experiments. Here Nair and Pregibon extend this pooling technique to replicated 2-level experiments and compare it to the fore-mentioned least squares analysis and maximum likelihood estimation (MLE).

While showing both techniques to be special cases of MLE under normal theory, they find the least squares analysis good for model identification but not so for estimation. They find the pooling technique to be generally biased and not good for model identification. They recommend using the least squares analysis for model identification and MLE for

parameter estimation.

Ullman (1989) introduces an extension of the analysis of means (ANOM) called the analysis of ranges (ANOR) which allows for a separate analysis of location and dispersion effects.

Maghsoodloo (1990) calculates the precise mathematical relationship between the SNRs for the STB and LTB cases and Taguchi's quadratic loss function. For the NTB case he provides an inequality.

2.3.3.2 Random and Mixed Effects

The simplest form of the linear model is the fixed effects model where the experimenter gathers information on all possible levels of a variable. The random effects model is used when the experimenter can only sample the levels of the variables and is commonly used to quantify the components of total process variance. Noise variables are random effects even though they are sometimes interpreted as fixed variables in RD experiments. A mixed effects model includes fixed and random effects and though computationally more complex than a fixed effects model, makes for a better prediction when noise variables are present. In this section we review two articles that discuss including random effects when modelling the process response in an RD experiment.

Sohn and Park (1998) consider two stage random effects regression models for process mean and variance. They use Empirical Bayes procedures to estimate treatment effects for mean and variance. The empirical Bayes procedure assumes parameters of the prior distribution are unknown and estimates them from the data using maximum likelihood estimation. Inferences are then made on the parameters using the likelihood ratio test to find significant effects.

Wolfinger and Tobias (1998) propose a general modelling paradigm for data from RD experiments which extends the traditional Gaussian linear mixed model for the case of heterogeneous variance. Mixed models are commonly required when blocking structures in data collection involve correlation between blocks which can be handled as random effects apart from residual error. In split plot design the whole plot effects and residual error are

usually modelled as zero mean, fixed variance random variables. The authors outline a general modelling framework involving location, dispersion and random effects and extend the mixed models to cover nonconstant variances for the whole plot and residual errors.

They define RD as the study of how to use control factors to make the process robust against noise and random factors. They describe random effects as those which directly model variance and covariance whereas dispersion effects determine the values of the parameters defining the distribution of the random effects. The framework forms an analysis model for the three types of effects, estimates the parameters and finally calculates an integrated model of the process mean and variance as a function of the control factors.

The advantages of this procedure include joint estimation of location, dispersion and random effects which are optimally weighted by likelihood estimates, accommodation of unbalanced data and flexible models for means, covariances and heterogeneous variances. Limitations of this approach include assumption of normality, dependence on model selection, and the loss of small location effects in the presence of large dispersion effects.

2.3.3.3 Analysis of Unreplicated Experiments

The most commonly cited advantage of the response model approach (RMA) is that it allows the use of more efficient experimental designs (i.e. the combined array). However that efficiency usually assumes there is no replication of the design. This section reviews references which provide methods of analyzing the data from un-replicated fractional factorial designs without replicate data points from which to estimate random error.

Box and Meyer (1986) use the sparsity principle to justify the use of unreplicated fractional factorial screening designs as a first step in a sequential methodology for RD. In this paper they present an analysis technique which complements normal probability plots for identifying significant effects from an unreplicated design. Their Bayesian approach assesses the size of contrasts by computing a posterior probability that each contrast is active. The prior distribution is characterized by two parameters representing the probability that a contrast is active and the standard deviation produced by an active contrast. Box and Meyer start with a prior probability of activity and assume normality of the significant

effects and deliver a non-zero posterior probability for each effect.

Lenth (1989) introduces a computationally simple and intuitively pleasing technique for measuring the size of contrasts in unreplicated fractional factorials. Box's method requires special software and uses a graphic known as the "Bayes Plot" to pick active effects. The Lenth method uses standard T statistics and contrast plots to indicate size and significance of the contrasts. The assumption of sparsity of effects justifies the specific calculation of the Margin of Error (ME) and simultaneous margins of error (SME) serve as decision limits for distinguishing active contrasts from inactive ones. Because of its elegant simplicity, the method of Lenth is commonly cited in RD case studies.

Berk and Picard (1991) examine the analysis of unreplicated saturated orthogonal arrays (i.e. designs where every degree of freedom is assigned to evaluate a contrast resulting in zero degrees of freedom for the error term). Lack of a proper error estimate negates the use of legitimate significance tests of factor effects. They propose a technique which still allows for a simple ANOVA based method which allows legitimate significance tests.

Sparsity of effects leads them to reserve approximately sixty percent of the total effects for construction of a baseline sum of squares which is used to compare the sums of squares of each observed effect. This results in legitimate significance tests, great ease of use and comparable performance to other statistically valid approaches in the examples analyzed by the authors.

Berg and Hynen (1997) develop another technique for testing significance of dispersion effects in unreplicated fractional factorials which differs from methods such as Box and Meyer (1986) in that it doesn't exploit the residuals remaining after removal of location effects. Data is assumed to be normally distributed with equal mean and different variances and dispersion effects are identified with alternative contrasts based on linear combinations of the design matrix columns. An F-statistic tests the null hypotheses that the variance of each column's high level factor data points is equal to that of its low-level factor data points. The F statistic is constructed as the ratio of the sum of contrasts of the high-level factor data to the sum of contrasts of the low-level factor data points for each column. Rejection of the null hypothesis indicates existence of a dispersion effect and the value of

the test statistic indicates the magnitude of that effect. The test is double sided and has demonstrated some sensitivity to non-normality in a simulation study by the authors.

Pan (1999) shows how failure to identify even small and moderate location effects can subsequently impair correct identification of dispersion effects when analyzing data from unreplicated fractional factorials. They explain formally how unidentified location effects influence the identification of dispersion effects and assert that these concerns do not exist in replicated experimental data.

Ye and Hamada (2000) propose a simple simulation method for estimating the critical values employed by Lenth in his method for testing significance of effects in unreplicated fractional factorial designs. They provide tables of these critical values and state this simulation method can be adapted for approximating critical values from other methods.

McGrath and Lin (2001) complement Pan (1999) by showing that if a model does not include all active location effects, the probability of falsely identifying significant dispersion factors is raised. They show analytically that without replication it is not possible to deconfound a dispersion effect from two location effects.

2.3.3.4 Bayesian Modelling

Just as RSM has been co-opted for use in RD, Bayesian methods of analysis are steadily finding wider employment in the statistical world as a useful alternative to frequentist methods. We've seen several articles where the Bayesian use of prior information has been used to respectively jump-start experimental design (Dumouchel and Jones (1994)), detection of location and dispersion effects (Box and Meyer (1986)) and modelling of random effects (Sohn and Park (1998)). In this section several references deal with Bayesian modelling of the data.

Chipman and Hamada (1996) use a Bayesian GLM to overcome the potential of infinite likelihood arising from general linear model (GLM) estimates of calculations from categorical data observed in fractional factorial designs. The Bayesian nature of their GLM easily models uncertainty in the parameters and the noise variables used in RD by assuming prior knowledge of the coefficients.

They use Gibbs sampling (Geman and Geman (1984)) to obtain marginal posteriors of the model parameters after conditioning on the data. The model fitting includes assessment of convergence of the Gibbs samples, variable selection techniques and examination of the posterior robustness to different priors. The authors demonstrate how to use the model to obtain the optimal control variable settings for RD.

Chipman (1998) applies Bayesian techniques to the data modelling and optimization stages of RD. The method he discusses is intended to serve as a supplement to the techniques of RMA and LMA for assessing the effect of parameter and model uncertainty. He uses the model selection methodology of Box and Meyer (1993) in conjunction with priors for variable selection with related predictors. For optimal choice of control factor settings he finds posterior distributions to assess the effect of model and parameter uncertainty. This methodology can be used to find the optimal control variable settings in RD or to assess the validity of a given set of such settings.

2.3.3.5 Generalized Linear Model

Up to this point, the linear modelling in this dissertation has assumed normality and constant variance. When the data does not demonstrate these properties, the most common approach is to transform the response data such that the transformed response complies with these assumptions. In many cases this is hard or impossible. The General Linear Model (GLM) was developed by Nelder and Wedderburn (1972) as a way of modelling data whose probability distribution is any member of the single parameter exponential family. The following brief introduction to GLMs is taken from Hamada and Nelder (1997).

GLMs generalize the classical linear model by allowing data distributions within the single parameter exponential family. For all these distributions the variance is a function of the mean and is represented as the product of a dispersion parameter and a variance function. The dispersion parameter represents that part of the variance not dependent on the mean and the variance function that part dependent on the mean.

Modelling with GLM comes down to correctly choosing the variance function, the link function and the terms in the linear predictor (i.e. the systematic effects). The link function

is a monotonic transformation of the assumed data distributional mean into the linear predictor and defines the scale on which the systematic effects are assumed to be additive. It is important to note that the distributional mean is transformed rather than the data itself. The choice of variance function to model the error component is entirely separate from the choice of link function to achieve linearity of the systematic effects.

The GLM is fitted by obtaining the maximum likelihood estimates for the coefficients to the terms in the linear predictor, which may contain continuous, categorical, interaction and polynomial terms. This section reviews articles which discuss the GLM and its relevance to solving RD problems.

Nelder and Lee (1991) argue that the GLM can extend the class of useful models for RD experiments to data-sets wherein a simple transformation can not necessarily satisfy the important criteria of normality, separation and parsimony. The GLM handles non-normality by embedding the normal distribution in the class of one-parameter exponential families and separation through a link function that defines a scale on which effects are assumed additive. This scale splits error specification from prediction by placing additivity on the predicted values rather than the data. For the dual response situation commonly encountered in RD, GLMs would be formulated for mean and variance.

After models are formulated and verified for consistency, a two step process of optimization is followed. First is choosing control variable settings for minimal variance and second is adjustment of the mean. Several examples illustrate how the link functions are chosen.

Engel and Huele (1996) integrate the GLM within the RSM approach to RD. Nonconstant variance is assumed and models for process mean and variance are obtained from a heteroscedastic linear model of the conditional process response. The pseudo-likelihood and logarithm methods for fitting the conditional response model via iterations of weighted least squares are presented as well as a simulation indicating that the optimal number of iterations depends on the magnitude of variance heterogeneity and degrees of freedom. The authors mention that nonlinear models and tolerances can also be studied with this approach.

Hamada and Nelson (1997) apply the techniques described in Nelson and Lee (1991) to

three quality improvement examples to emphasize the utility of the GLM in RD problems. They point out that the dispersion parameter and variance function are a natural expression of the separability desired for optimization of both process mean and variance. The linear predictor can accommodate continuous and categorical variables and the properties of linearity, additivity and aliasing are the same as for classical linear modelling. The generalization of residual sum of squares to the deviance function enables the GLM to function over its broader class of distributions.

Nelder and Lee (1998)restate the claim initially made in Nelder and Lee (1991) for the value of the GLM in joint modelling of mean and variance. They re-analyze the injection molding example of Enge (l992) and the welding strength example from Box and Meyer (1986) using the GLM to illustrate.

2.3.3.6 RSM and Regression Modelling

The tendency of RD experimenters to use RSM has been steadily growing since the advent of the RMA (see Welch et al (1990). This section features a non-parametric regression modelling approach to the dual response optimization problem and an overview of the current state of RSM.

Vining and Bohn (1998) look at using semi and fully non-parametric regression to model process variance in RD situations where there exist extremely noisy process variance functions. They use the product kernel estimator detailed in Mueller (1988) which minimizes the integrated mean square error, to allow weighted averaging to be applied one dimension at a time in a sequential process. The assumption of less structure in the non-parametric kernel estimator requires more replicates than a parametric model and the experimenter must balance number of replicates at each design point against available resources. In the semi-parametric approach they use nonparametric kernel estimators based on variances at points of replication to predict process variance at any point in the region of interest. They then use the inverses of the predicted design point variances as weights for weighted least squares estimation of a parametric linear response model. This approach exchanges a formal mode for process variance for better prediction while retaining the formal response model.

The fully nonparametric approach employs separate non-parametric kernel estimators for process response and variance. For any location in the region of interest the response is predicted by kernel regression on mean responses at the design points. The variance is predicted using a separate kernel regression performed on observed process variance at the points of replication. This latter approach exchanges formal models for response and variance for better prediction, assuming that standard linear models don't work well.

Myers (1999) summarizes the current state of RSM and its relation to RD over the last twenty years. He predicts that RSM will continue integrating use of the GLM and will expand in its treatment of multiple response situations and non-parametric and semi-parametric methods. The interaction between RSM and the areas of Bayesian experimental design, nonlinear optimization, mixed model analysis and quality engineering is expected to grow and continue refining RSM.

2.4 Dynamic Robust Design

Up to this point, this article has discussed only static RD, where the targeted response is a given, fixed level and is only affected by control and noise variables. In dynamic RD (hereafter DRD) there exists a third type of variable, the signal variable whose magnitude directly affects the mean value of the response. The experimental design recommended by Taguchi for DRD is the product array consisting of an inner control array crossed with an outer array consisting of a the sensitivity factors and a compound noise factor.

A common choice of dynamic loss function is the quadratic loss function popularized by Taguchi,

$$L(Y, t(M)) = A_0(Y - t(M))^2 \tag{2.15}$$

where A_0 is a constant. This loss function provides a good approximation to many realistic loss functions. It follows that the average loss becomes:

$$\begin{aligned} R(\mathbf{C}) &= A_0 E_M E_{N,\epsilon}(Y - t(M))^2 \\ &= A_0 E_M [VAR_{N,\epsilon}(Y) + (E_{N,\epsilon}(Y) - t(M))^2] \end{aligned} \tag{2.16}$$

Taguchi identifies dispersion and sensitivity effects by modelling SNR respectively as a function of control factors and sensitivity factors. His two step procedure for DRD recommends finding control factor settings to minimize SNR and, conditional on those settings, set other, non-SNR related control variables to adjust the process to the targeted sensitivity level. In this section we review articles discussing aspects of DRD.

Ghosh and Derderian (1995) introduce the concept of robustness of the experimental plan itself to the noise factors present when conducting DRD. For combined arrays they consider blocked and split-plot designs and for product arrays they consider univariate and multivariate models. The authors show how to obtain the optimal control variables settings for both types of experimental plans. In product arrays they do this by choosing setting which minimize the noise factor effects on process variability and for the combined array they attempt to minimize the interaction effects between control and noise factors.

Wassermann (1996) clarifies the use of the SNR for the dynamic case by explaining it in terms of linear modelling of process response. He expresses the dynamic response as a linear model consisting of a signal factor, the true sensitivity (β) at specific control variable settings and an error term. He obtains estimates of the sensitivity using ordinary least squares with variance represented by residual sum of squares (V_e). He references the generalized SNR for dynamic characteristics advocated by LNW95UP:

$$SNR \text{ or } \eta = 10log_{10}\frac{\beta^{\alpha}}{V_e} \tag{2.17}$$

and points out that the β term is the same he estimates using ordinary least squares. He illustrates Taguchi's approach modelling the sensitivity and dispersion effects under consideration.

Miller and Wu (1996) prefer the term signal-response system to dynamic robust design for its intuitive appeal and identify two distinct types of signal-response systems. They call them measurement systems and multiple target systems and this classification determines the performance measure used for finding the optimal control variable settings. A multiple target system is one where the value of the response is adjusted by adjusting a signal variable. A measurement system is a process used to estimate a characteristic of a particular sample.

The amount of the characteristic acts as an input signal which the system converts into a response value which can be a measurement. The purpose of a performance measure is to evaluate the validity of a signal-response relationship for a specific application.

Since a measurement system is intended to estimate a quantity of interest, it's reasonable to evaluate its performance with respect to precision of estimates. Miller and Wu discuss a number of performance measures based on assumptions regarding the variance of the response at different signal levels and note that Taguchi's dynamic SNR is an appropriate performance measure for some measurement systems. For multiple target systems they argue that maximizing Taguchi's dynamic SNR is not appropriate since it can result in greater response variation. They suggest that a reasonable performance measure for multiple target systems is a weighted average of the performance of the various system elements as measured in off-target distance. They discuss the use of performance measure modelling (PMM), the analog to the loss model approach in the static case, and response function modelling (RFM), the analog to the response model approach for the static case.

They illustrate the use of both PMM and RFM in an injection molding experiment. They recommend the RFM and an experimental design which crosses an inner combined array of control and noise factors with an outer array of different signal levels. Location and dispersion are measured at each line of the control-noise array and these parameters are modelled as a function of control and noise factors and standard procedures are used to identify and fit the model. The model is then used to find optimal control factor settings.

Lunani et al (1997) present two new graphical procedures for identifying suitable measures of location and dispersion in robust design situations with dynamic experimental designs. The data analysis is similar to that ofBox (1988) and Nair and Pregibon (1986) where the goal is to identify a suitable measure of dispersion to avoid confounding of dispersion and sensitivity effects. Taguchi identifies dispersion effects by fitting a linear model to the estimated SNR as a function of control factors using ANOVA or half normal plots to identify significant effects. The sensitivity effects are found by fitting a linear model to sensitivity measures as a function of control factors again using ANOVA or half-normal plots. Once dispersion and sensitivity effects are identified, one picks the control factor

settings to reduce variability and get as close as possible to the targeted sensitivity value. The authors conclude that inappropriate use of the SNR can lead to:

- spurious detection of dispersion effects due to their confounding with control-control interactions
- non-identification of important dispersion effects

They show that Taguchi's dynamic SNR is a special case of a more general variance relationship,

$$\sigma^2(d_i) = \beta^{\gamma}(d_i)\phi^2(d_i) \tag{2.18}$$

where $\gamma = 2$. They present two graphical methods for distinguishing dispersion effects and estimating the constant of the variance formula. These are the standard deviation plot (SS) and the gamma plot (GP). The standard deviation (SS) plot is an extension of the mean-variance plot used in static parameter design by Nair and Pregibon (1986). The gamma plot is similar to the lambda plot from Box (1988) where Box looked for the best transformation.

Given the variance model in equation 2.18, there is a log-linear relationship between the sensitivity measure β and the standard deviation σ_i:

$$\log \sigma(d_i) = \log \phi(d_i) + \frac{\gamma}{2} \log \beta(d_i) \tag{2.19}$$

The SS plot indicates different location factors when the different levels of the control factor produce a plot separated by magnitude of $log\beta(d_i)$. They indicate dispersion effect when there are two parallel plots from the different levels, indicating different intercepts and no dependence on level of $log\beta(d_i)$.

The gamma plot graphs different values of the T-statistics derived from plugging a range of gamma values into the variance equation above and indicates dispersion effects with negative slope and location effects with positive slope. The minimum of a control factor's plot indicates the ideal gamma setting for minimizing cross talk between location and dispersion effects.

McCaskey and Tsui (1997) show that Taguchi's two step procedure for dynamic systems is appropriate only for multiplicative models and develop a procedure for dynamic systems under an additive model. Given a goal of minimal quadratic loss, average loss can be decomposed into the sum of process variance and bias squared. For a dynamic system this equates to minimizing this sum while averaging over possible values of the signal.

They assume that the actual response function of the additive model for the dynamic system with signal M and response y has two components, the mean function and the noise function, which combine to form the following linear equation:

$$y = \beta(C_1, \mathbf{C}_2)M + e(\mathbf{N}; \mathbf{C}_2), \tag{2.20}$$

where $\mathbf{C} = (C_1, \mathbf{C}_2)$, with C_1 being the adjustment factor (a factor that has a significant effect on the variance but a negligible effect on the mean function), $\mathbf{C}_2$ being a vector of non-adjustment control factors, $E[e(\mathbf{N}; \mathbf{C}_2)] = 0$, and $VAR[e(\mathbf{N}; \mathbf{C}_2)] = \sigma^2(\mathbf{C}_2)$. The moving target function is assumed to be linear in M, i.e., $t(M) = \beta_0 M$.

Given the dynamic quadratic loss function and the additive model in equation 2.20, the average loss is:

$$\begin{aligned} R(C_1, \mathbf{C}_2) &\propto VAR[e(\mathbf{N}; \mathbf{C}_2)] + E_M[(\beta(C_1, \mathbf{C}_2)M - \beta_0 M)^2] \\ &= \sigma^2(\mathbf{C}_2) + (\beta(C_1, \mathbf{C}_2) - \beta_0)^2 E_M(M^2). \end{aligned} \tag{2.21}$$

Under the constraint that the slope function $\beta(C_1, \mathbf{C}_2)$ must be shifted to the ideal slope β_0, we define the following robust design problem:

$$Minimize_{C_1, \mathbf{C}_2}\ R(C_1, \mathbf{C}_2)\ subject\ to\ \beta(C_1, \mathbf{C}_2) = \beta_0. \tag{2.22}$$

It follows by the same argument in LSK (1987) and equation 2.20 that $R(C_1, \mathbf{C}_2)$ can be minimized by the following two-step procedure:

1. Find $\mathbf{C}_2^*$ that minimizes $\sigma^2(\mathbf{C}_2)$.

2. Shift the adjustment factor C_1 to C_1^* so that $\beta(C_1^*, \mathbf{C}_2^*) = \beta_0$.

It can be easily shown that $(C_1^*, \mathbf{C}_2^*)$ minimizes $R(C_1, \mathbf{C}_2)$ as the first term in equation 2.21 is minimized at $\mathbf{C}_2^*$ and the second term drops to zero at $(C_1^*, \mathbf{C}_2^*)$, where $\beta(C_1, \mathbf{C}_2)$ and $\sigma^2(\mathbf{C}_2)$ need to be estimated from experimental data.

Tsui (1999) compares the effect estimates obtained using the response model approach and the loss model approach for dynamic robust design problems. The loss model estimates the intercept, slope and variance parameters of a linear model of the response at each set of control factor settings. These parameter estimates are treated as separate responses from which significant control factor effects are identified and optimal control factor settings are derived. The variance is a quadratic function of the control factors even though response is a linear function in the control factors.

The response model approach estimates the parameters directly from the data, which is usually gathered from a combined array. The intercept, slope and variance functions are approximated from the parameter estimates and the optimal control factor settings identified. While the two approaches yield the same effect estimates for slope and intercept, they do not do so for variance.

In the static case Tsui and Li (1994) showed the potential for bias when estimating dispersion effects with the loss model approach. Here Tsui demonstrates the same potential bias exists when using the loss model approach in the dynamic case and argues that the loss model approach may hence result in non-optimal settings. This is especially true when the control array is Resolution III or IV. The response model approach provides information not yielded by the loss model approach on how specific control factor settings an reduce variance from individual noise factors as well as pure variance.

2.5 Other RD Applications

2.5.1 Tolerancing

This chapter has focused on RD which is synonymous with Taguchi's methods of parameter design. Taguchi has also made significant contributions in the area of tolerance design. This section reviews articles which examine developments in the techniques of tolerance design.

D'errico and Zaino (1988) describe Taguchi's approach to tolerance design and state that Taguchi's method provides good estimates of smooth transformation of the component variables. While Taguchi's method is simple, not requiring derivatives, Monte Carlo or second order Taylor series, it also does not easily handle non-normal component distributions

and returns only estimates of a distribution's moments rather than the distribution function itself. The authors propose a modification of Taguchi's approach based on a product Gaussian quadrature which provides better estimates of high order moments and outperforms the basic Taguchi method in most cases.

Bisgaard (1997) proposes using factorial experimentation as a more scientific alternative to trial and error for designing tolerance limits when mating components of assembled products. He employs combinations of pre- and post-fractionization of certain products of two level factorial designs and considers cost of experimentation when determining experimental runs.

Zhang and Wang (1998) formulate the robust tolerance problem as a mixed nonlinear optimization model and solve it using a simulated annealing algorithm. The optimal solution allocates assembly and machining tolerances so as to maximize the product's insensitivity to environmental factors.

Moskowitz et al (2001) develop parametric and non-parametric methods for finding economically optimal tolerance allocations for a multi-variable set of performance measures based on a common set of design parameters. The parametric method assumes distributional knowledge of design parameters and derives the tolerance allocation that jointly minimizes expected total cost of both supplier and manufacturer. The non-parametric method assumes partial information on the distribution of design parameters and derives mini-max tolerance allocation with respect to expected total cost. The authors consider the trade-off between supplier and manufacturer costs as an economic means of assessing tolerances and employ a quadratic loss function which includes both types of costs. They caution that tolerance design is a short term solution and no substitute for the goal of reducing undesired product variability, which is best addressed through continuous improvement activities.

2.5.2 Process Control

The techniques of RD are "off-line" in that they attempt to shift the responsibility for quality upstream into the product design. There are a number of ways that Taguchi's work has influenced production or "on-line" quality practices. This section reviews developments

in these process control practices.

Adams and Woodall (1989) examine the premises of the optimal control policies derived from the economic model of Taguchi et al (1989) for the random walk case. Taguchi et al (1989) assume, based on Brownian motion, that the average number of items produced between process adjustments is proportional to the square of the control limit. Simple formulas for the sampling interval, the average number produced between process adjustments and the control limit are derived from this assumption and iteratively estimated. The authors find this procedure performs poorly under the random walk model for large sampling intervals and small control limits. They develop a generalized model for the random-walk case by modifying the model of Box and Jenkins (1963) to allow determination of the control limit and the sampling frequency. They also modify Taguchi's approach to allow a single estimation of the optimal control parameters based on a good estimate of process variance. Comparison of simulated results shows the modified Taguchi approach compares well to the generalized model approach. They recommend the generalized model approach to Taguchi's and note that this comparison was made assuming normality.

Nayebpour and Woodall (1993) compare the production line quality monitoring techniques of Taguchi et al (1989) with that of Gibra (1978). The authors propose an economic model assuming geometric process failure mechanisms and compare the optimal methods derived from this model with those advocated by Taguchi et al (1989).

2.5.3 Reliability

Reliability is the study of how to make products and processes function for longer periods of time with minimal interruption. It is a natural area for RD application and the Japanese auto industry has made huge strides in this area compared to its American counterpart. In this section several authors comment on the application of RD to reliability.

Hamada (1993) demonstrates the relevance of RD to reliability improvement. He recommends the response model approach for the additional information on control-noise interactions it provides and suggests alternative performance criteria to Taguchi's LTB signal-to-noise ratio for maximizing reliability.

Kuhn et al (2000) extend the methods of Myers et al (1992) for linear models and normally distributed data for achieving a robust process when the response is time to an event. They develop a failure-time model, usually non-linear, for time to event as a function of control and noise factors and construct unconditional performance measures which reflect the variation over the distribution of noise factors. These performance measures include the qth percentile of the failure-time distribution and unconditional mean and variance. They use graphical procedures and optimization routines to find control factor settings which optimize performance while reducing noise factor variance.

2.5.4 Case Studies in Manufacturing

We list some RD case studies in manufacturing documented in the literature.

Jiang et al (1991) discuss the application of RD to optimization of a robot's process capability (RPC). The RPC consists of four separate response variables which are modelled as functions of subsets of seven control factors. The authors use a product array design using time as a noise variable and taking twelve replicates of all test combinations of control and noise variables. They calculate SNR at each test combination and perform ANOVA on the SNR in some cases and directly on response data for others. They use a combination of loss model approach and response model approach techniques to calculate effect estimates and obtain the optimal control factor settings for the four response variables.

Mesenbrink et al (1994) applied the techniques of RD to optimize three performance measurements (i.e. mean, spatial uniformity and variance) of a high volume wave soldering process. They implemented a mixed level fractional factorial design to collect ordered categorical data regarding the soldering quality of component leads inserted through printed circuit boards. Representative soldered leads were sampled and grouped into several ordered classes of quality and were converted into continuous measurements via a scoring system. Main effect and interaction model terms were selected using regression and a nonlinear optimization routine used to identify the optimal settings for the discrete and continuous process variables. Significant quality improvement was obtained.

Chaajed and Lowe (2000) apply the techniques of RD to the problem of structured tool

management. Noting that the availability of specific tools affects product design options, machine loading, job batching, capacity scheduling and routing decisions, they stress the need to actively manage the design and selection of tools for minimizing cost. For the cases of tool selection and tool design they use Taguchi's quadratic loss function to find the most cost effective way for accomplishing the processing of a fixed number of punched holes in sheet metal products.

2.6 Discussion

RD is a collection of techniques whose aim is to design products and processes with minimal sensitivity to non-controllable factors. While the beauty of Taguchi's approaches to many non-statisticians has been its simplicity, the experimenter must be aware of the underlying assumptions to avoid biased results. This paper has described the pros and cons of Taguchi's parameter design and alternative practices from a variety of respected individuals in the field. In conclusion, there is no consistently simple way to conduct RD without some knowledge of experimental design, analysis and the process involved.

CHAPTER III

A REVIEW OF ROBUST DESIGN TECHNIQUES FOR MULTIPLE RESPONSES

Problems in engineering design often involve determining design variable settings to optimize individual product performance for multiple criteria, which are often in conflict. We review mathematically rigorous techniques from the statistical literature for finding a vector $\mathbf{x}$ of design variable settings, which produces an optimal compromise solution among a group of prioritized response variables. The best compromise solution is typically gained by optimizing an objective function, which incorporates the prioritized demands of multiple responses.

Chapter 2 provided a review of the prominent univariate approaches in RD. In this chapter we see that multi-response objective functions are usually constructed by combining the objective functions of univariate responses. A multivariate approach from the engineering literature called the compromise Decision Support Problem is also reviewed. Finally a table comparing the relative merits of the different multivariate approaches summarizes the article in a concise and user-friendly fashion.

3.1 Frame of Reference

One of the most important roles for an engineering designer is to make decisions about products that are being designed. The value of most products hinges upon their ability to satisfy multiple functional criteria. Typically a designer is asked to determine design variable settings to optimize product performance on these multiple criteria which are often in conflict. In this article, we call these criteria quality characteristics or responses and note that these are often of varying priority to the end user. This requires a design decision-maker (DDM) to prioritize and/or assign quantitative importance measures to responses in order to make the best compromise choices.

Our approach is founded on the idea of bounded rationality as proposed by Simon (1976). Unlike the ideal state commonly assumed for economic analyses, which assumes a designer has perfect and complete information, bounded rationality more closely reflects the actual state of the world in which information is uncertain, incomplete and complex, and where the number of potential courses of action is nearly infinite. If complete information were available, it would be possible to invoke single objective design making in all circumstances, as recommended by Hazelrigg (1996) and others. However, in recognition of the fact that the world is often less than ideal, we assert that in practice, rigorous methods for multi-criteria design decision making are essential.

In this chapter, we explore the formulation of engineering design decisions in the context of the more general, mathematically rigorous techniques documented in the statistical literature for finding a vector $\mathbf{x}$ of variable settings to yield an optimal compromise solution among a group of prioritized response variables. We will examine several attributes of each approach, including how the correlation structure of the multiple responses is utilized in the optimization process.

Some techniques assume that the multiple responses are independent of each other. This implies that variation in one response is not related to variation in any other response. While this assumption brings mathematical simplification to statistical analysis, it is not the reality of most design situations. For example, customer criteria for a car include fuel efficiency, cost, reliability, maneuverability, capacity, vehicle weight and driving comfort. These are clearly correlated since higher vehicle weight typically implies both higher capacity and lower fuel efficiency. Hence the assumption of response independence is an impracticality in many design problems.

Other techniques actively exploit the correlations of the responses as a source of information while searching for optimal design parameters. This is a statistical advantage since an additional source of response information, i.e., correlation structure, is being put to use. Lastly there are techniques which, while not actively harnessing the correlation information, are not hampered by an assumed independence of the responses. Regardless of how the response correlation structure is employed, all techniques examined assume that

ordering and weighting of responses are carried out by a single DDM and are transitive.

One of the commonly cited examples in the statistical literature is that of manufacturing a beef stew military field ration as detailed in Contreras (1995). In this case there are two important quality characteristics (i.e., responses), namely heating rate index and the lethality index. The heating rate index is the rate at which the product may be brought to sterilizing temperature and the lethality index is an indicator of microbiological safety. The five design variables include sauce viscosity, residual gas, solid to liquid ratio, net weight and speed of rotation of the food pouch during the heating process.

A DDM wants to choose the settings of these five variables so that the heating rate is as fast as possible, since this expedites manufacturing, and so that lethality index stays above a certain minimum to guarantee consumer safety. Furthermore, the DDM wants to minimize how far the lethality index rises above the required safety level since flavor deteriorates as this index rises. This last requirement is an example of a constraint within the multi-response optimization.

Even for the simple case of only two responses, the statistically based methods typically employ an objective function incorporating the relative importance of the two responses. The design goal is to identify the specific design variable settings that optimize the objective function. In most statistically based methods, the key to finding these optimal design parameters is choosing the appropriate objective function for the design situation at hand.

Important considerations when choosing the multi-response optimization approach include how many and which type of individual responses are handled, how they are weighted, the type of modelling used to represent individual responses or the objective function, the number of responses reasonably managed by the objective function and the specific optimization techniques which complement that function.

In Section 2 we compare the different multi-response objective functions formed by additive and multiplicative combination of the univariate objective functions. In Section 3 we review the compromise Decision Support Problem, a hybrid formulation incorporating concepts from both traditional mathematical programming and goal programming, which is used in engineering sectors for solving the multi-response optimization problem. In Section 4

we compare the different multi-response techniques with respect to a number of important metrics including their ability to manage constraints and how the optimal solutions are affected by shifts in target or specification.

3.2 Robust Design with Multiple Responses

In Chapter 2 we discussed loss and utility functions and showed how the relation between off-target and variance components underlies the loss function optimization strategies for single responses. Multi-response optimization typically combines the loss or utility functions of individual responses into a multi-variate function to evaluate the sets of responses created by a particular set of design variable settings.

This section is divided into two subsections which deal with the additive and multiplicative combination of loss and utility functions respectively.

3.2.1 Additive Combination of Univariate Loss Functions

In the multi-response situation it is assumed that the responses $(Y_1, ..., Y_r)$ are affected by the control and noise factors as follows:

$$Y_i = f_i(\mathbf{x}, \mathbf{N}, \theta_i) + \epsilon_i, \;\; i = 1, ..., r,$$

where $\mathbf{x} = (x_1, ..., x_p)^T$ is the vector of control factors, $\mathbf{N} = (N_1, ..., N_q)^T$ is the vector of noise factors, $\theta_\mathbf{i}$ is the vector of unknown response model parameters, and f_i is the transfer function for Y_i. The control factors are assumed to be fixed and represent the fixed design variables. The noise factors $\mathbf{N}$ are assumed to be random and represent the uncontrolled sources of variability in production. The pure error ϵ_i represents the remaining variability of the manufacturing process that is not captured by the noise factors, and is assumed to be normally distributed with zero mean and finite variance. A loss occurs if any of the responses (Y_i) deviates from its target t_i.

For univariate responses, expected squared-error loss is a convenient way to evaluate the loss caused by deviation from target because of its decomposition into squared off-target and variance terms. A natural extension of this loss function to multiple correlated responses is

the multivariate quadratic function of the deviation vector $(\mathbf{Y}-\tau)$ where $\mathbf{Y}=(Y_1,...,Y_r)^T$ and $\tau=(t_1,...,t_r)^T$, i.e.,

$$MQL(\mathbf{Y},\tau) \quad = \quad (\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau), \tag{3.1}$$

where $\mathbf{A}$ is a positive definite constant matrix. The values of the constants in $\mathbf{A}$ are related to the costs of non-optimal design, such as the costs related to repairing and/or scrapping non-compliant product. In general the diagonal elements of $\mathbf{A}$ represent the weights of the r characteristics and the off-diagonal elements represent the costs related to pairs of responses being simultaneously off-target.

It can be shown that, if $\mathbf{Y}$ follows a multivariate normal distribution with mean vector $E(\mathbf{Y})$ and covariance matrix $\Sigma_{\mathbf{Y}}$, the average (expected) loss can be written as:

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \\ &= trace[\mathbf{A}\Sigma_{\mathbf{Y}}] + [E(\mathbf{Y})-\tau]^T\mathbf{A}[E(\mathbf{Y})-\tau]. \end{aligned} \tag{3.2}$$

The simplest approach to solve the robust design problem is to apply algorithms to directly minimize the average loss function in Equation 3.2. Since the mean vector and covariance matrix are usually unknown, they can be estimated by the sample mean vector and sample covariance matrix or a fitted model based on a sample of observations of the multivariate responses. This strategy of optimizing the multivariate quadratic loss function directly employs the correlation structure of the responses in the trace component.

The off-target vector product $[E(\mathbf{Y})-\tau]^T\mathbf{A}[E(\mathbf{Y})-\tau]$ and $trace[\mathbf{A}\Sigma_{\mathbf{Y}}]$ are multi-variate analogs to the squared off-target component and variance of the univariate squared-error loss function.

One sees from this decomposition that moving all response means to target simplifies the expected multi-variate loss to the $trace[\mathbf{A}\Sigma_{\mathbf{Y}}]$ term. The trace-covariance term shows how the values of $\mathbf{A}$ and the covariance matrix $\Sigma_{\mathbf{Y}}$ directly affect the expected multi-variate loss. Higher variances and covariances of the individual responses raise expected loss while

the **A** values can be seen as weights of the individual response variances and two-response interactions. Hence weighting or prioritization of the individual responses within the multivariate loss function is determined by the values in these two matrices.

3.2.1.1 Additive Combination of Loss Functions

To demonstrate how this multivariate quadratic loss function additively combines the individual loss functions we look at the simplest multivariate case, that of two properly ordered responses.

$$\text{Let } \mathbf{Y} = [Y_1, Y_2],\ \tau = [t_1, t_2] \text{ and } \mathbf{A} = \begin{pmatrix} a1 & 0 \\ 0 & a2 \end{pmatrix}$$

Then

$$\begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix}^T \mathbf{A} \begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix} = a_1(Y_1 - t_1)^2 + a_2(Y_2 - t_2)^2.$$

For this simplest of cases, the multivariate quadratic loss function is equivalent to adding the individual squared-error loss functions of each response. This multivariate quadratic loss function becomes increasingly complex with larger numbers of responses and interactions indicated by non-zero terms in the off-diagonal elements of **A**. We repeat the same example with a non-zero off-diagonal element:

$$\text{Let } \mathbf{Y} = [Y_1, Y_2],\ \tau = [t_1, t_2] \text{ and } \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

Then

$$\begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix}^T \mathbf{A} \begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix}$$

$$= a_{11}(Y_1 - t_1)^2 + 2a_{12}(Y_1 - t_1)(Y_2 - t_2) + a_{22}(Y_2 - t_2)^2$$

3.2.1.2 The Mahalanobis Distance

Khuri and Conlon (1981) propose an algorithm for the optimization of a multi-response system which seeks the group of design settings that minimizes the Euclidean distance from a vector of idealized responses, i.e., the Mahalanobis distance (MD). The MD is a function of the estimated responses and their covariance structure.

Their procedure assumes that all response functions depend on the same set of design variables and can be represented by polynomial regression models of the same degree within the region of interest. They reduce the multiple responses to a linearly independent subset and calculate least squares estimates for these responses from the multi-response data set.

They express the r linearly independent response functions in the following multivariate form:

$$\mathbf{\Xi} = \mathbf{X\Theta} + \varepsilon \tag{3.3}$$

where $\mathbf{\Xi} = [\mathbf{Y_1}, \ldots, \mathbf{Y_r}]$ is the n x r matrix consisting of the r column vectors corresponding to the n observations of each of the responses . For each of the n input vectors there are r different response values which make up this n x r multi-response data matrix.

$\mathbf{X}$ is the n x p full column rank matrix consisting of of n rows (i.e., $\mathbf{x_1}, \ldots, \mathbf{x_n})^T$. Each row contains the p model terms consisting of the union of all polynomial model terms from the r different response models. These same p model terms however take on different values depending on the design vector corresponding to each row.

$\mathbf{\Theta} = [\ \theta_\mathbf{1}, \theta_\mathbf{2}, \ldots, \theta_\mathbf{r}\ ]$ is the p x r matrix consisting of r column vectors, each having p model parameters and $\varepsilon = [\epsilon_\mathbf{1}, \ldots, \epsilon_\mathbf{r}]$ is the n x r matrix consisting of the r column vectors corresponding to the error terms of the response models. The usual assumptions are that the rows of ε are statistically mutually independent, each having a zero mean vector and a common covariance matrix $\mathbf{\Sigma}$. An unbiased estimate of the covariance matrix (i.e., $\hat{\mathbf{\Sigma}}$) is typically used.

Each of the response column vectors $\mathbf{Y_1}, \ldots, \mathbf{Y_r}$ is represented by a polynomial response function of degree g of the following form:

$$\hat{\mathbf{Y}}_i(\mathbf{x_1}, \ldots, \mathbf{x_n}) \ = \mathbf{X}\hat{\theta}_\mathbf{i},\ \text{i} = 1, \ldots, \text{r} \tag{3.4}$$

where $\mathbf{X}$ is the matrix from Equation 3.3. It is of interest to note that since each row of the $\mathbf{X}$ matrix contains the union of model terms from all responses, the same $\mathbf{X}$ matrix is used for each response. Each individual response value, i.e., $Y_i(\mathbf{x_j})$ where Y_i indicates a specific response variable and $\mathbf{x_j}$ a specific design vector, can be modelled by the following

polynomial equation of degree g:

$$\hat{Y}_i(\mathbf{x_j}) = \mathbf{z_j^T}(\mathbf{x_j})\hat{\theta}_\mathbf{i} \tag{3.5}$$

where $\mathbf{z_j^T}(\mathbf{x_j})$ is the single row vector of dimension p from the $\mathbf{X}$ matrix of Equation 3.3 corresponding to $\mathbf{x_j}$.

Khuri and Conlon recommend the following distance measure, i.e., the MD:

$$MD[\hat{\mathbf{Y}}(\mathbf{x_j}), \tau] = \left[\frac{(\hat{\mathbf{Y}}(\mathbf{x_j}) - \tau)^T \mathbf{\Sigma^{-1}} (\hat{\mathbf{Y}}(\mathbf{x_j}) - \tau)}{\mathbf{z_j^T}(\mathbf{x_j})(\mathbf{X^T X})^{-1}\mathbf{z_j x_j}} \right]^{1/2} \tag{3.6}$$

where $\hat{\mathbf{Y}}^T(\mathbf{x_j}) = [\hat{Y}_1(\mathbf{x_j}), \ldots, \hat{Y}_r(\mathbf{x_j})]$ is the vector of estimated responses from a particular design vector $\mathbf{x_j}$, $\tau^T = [\tau_1, \tau_2, \ldots, \tau_r]$ is the vector of individual optimal responses and $(\mathbf{X^T X})^{-1}$ is the inverse of the squared design matrix.

Khuri and Conlon's optimal solution is the vector of design variables $\mathbf{x_j}$ which minimizes the distance measure $MD[\hat{\mathbf{Y}}(\mathbf{x_j}), \tau]$. For the case of potential fluctuation around the idealized response values, they propose a procedure for finding control variable settings which produce a mini-max solution for the distance metric involving a modified version of the same distance measure. Note that this distance function is a special case of the multivariate quadratic loss function of equation 3.2 where:

$$\mathbf{A} = \frac{\mathbf{\Sigma^{-1}}}{\mathbf{z_j^T}(\mathbf{x_j})(\mathbf{X^T X})^{-1}\mathbf{z_j}(\mathbf{x_j})} \tag{3.7}$$

Hence the MD procedure of Khuri and Conlon is a special case which implies a weighting of the individual responses based on the inverse covariance matrix $\mathbf{\Sigma^{-1}}$, the vector $\mathbf{z_j}^T(\mathbf{x_j})$ and the inverse of the squared design matrix $(\mathbf{X^T X})^{-1}$. Although this procedure does use the inverse covariance matrix of the responses, it only models the subset of linearly independent responses and so does not extract all the statistical information available.

3.2.1.3 Additive Formation of Multi-variate Loss Functions

In this section we briefly review the literature for examples of multivariate functions formed by the additive combination of univariate loss and utility functions. We list the cases in order of increasing complexity.

Kumar and coauthors (2000) suggest creating a multi-response utility function as the additive combination of utility functions from the individual responses. If Y_i is the value of response i, each response has utility function $P_i(Y_i)$ and the overall utility function is defined as:

$$P(Y_1, ..., Y_r) = \sum_{i=1}^{r} \omega_i P_i(Y_i) \tag{3.8}$$

where ω_i = weight of each response and $\sum_{i=1}^{r} \omega_i = 1$. Here the goal is to find the set of design variable settings that maximizes the overall utility function.

For cases where the target is the mid-specification point, Artiles (1996) proposes standardizing the squared-error loss function with the following proportionality constant:

$$A_0 = \left[\frac{2}{(USL_i - LSL_i)}\right]^2 \tag{3.9}$$

where LSL_i and USL_i are respectively the upper and lower specification limits for each Y_i.

With this constant the standardized squared-error loss function for a single response can be written as:

$$\begin{aligned} SLOSS(Y_i) &= \left[\frac{2}{USL_i - LSL_i}\right]^2 (Y_i - t_i)^2 \\ &= 4\left[\frac{Y_i - t_i}{USL_i - LSL_i}\right]^2 \end{aligned} \tag{3.10}$$

This standardized loss takes on the value zero at the target and the value one at the specification limits. A multivariate loss function is constructed simply as the sum of these dimensionless standardized loss functions. This total standard loss corresponds to the vector of responses $(Y_1, \ldots, Y_r)$ and is defined as:

$$TSLOSS(Y_1, ..., Y_r) \quad = \quad 4\sum_{i=1}^{r}\left[\frac{Y_i - t_i}{USL_i - LSL_i}\right]^2 \tag{3.11}$$

where t_i is the target value for each Y_i. They assume that all responses are uncorrelated. Since all individual standardized loss functions are simply added, they imply equal weighting of all responses.

Ames and coauthors (1997) proposed a multivariate loss function, the global quality loss function (GQL) as:

$$GQL(Y_1, ..., Y_r) = \sum_{i=1}^{r} \omega_i (Y_i - t_i)^2 \tag{3.12}$$

where the squared-error loss function of measured responses $(Y_1, \ldots, Y_r)$ and target values $(t_1, \ldots, t_r)$ are weighted by the constants $(\omega_1, \ldots, \omega_r)$. Clearly the GQL is a simple addition of the squared error losses of the individual responses with the scaling constants representing response priority.

For the subset of responses for which quadratic response surface models exist (i.e., $Y_1, \ldots, Y_m$), they define the process related quality loss function as:

$$GQLP(Y_1, ..., Y_m) = \sum_{i=1}^{m} \omega_i \left[Y_i(x_1, \ldots, x_p) - t_i + \epsilon_i \right]^2 \tag{3.13}$$

where the $Y_i(x_1, \ldots, x_p)$ are the responses approximated by quadratic response surface models and ϵ_i are the residuals of the models. When the ϵ_i are small compared to the off-target factors, the $GQLP$ can be defined as:

$$QLP(Y_1, \ldots, Y_m) = \sum_{i=1}^{m} \omega_i \left[Y_i(x_1, \ldots, x_p) - t_i \right]^2 \tag{3.14}$$

Minimizing the QLP also minimizes the $GQLP$ for the following two cases:

- ϵ_i are independent and of equal variance
- when off-target contributions to the loss function are significantly larger than the random error contributions

The authors point out that the first condition is a common assumption of response surface modelling and the second is a common characteristic during development of a product. In their photographic application, weighting is assigned by subjectively ranking the effect of a particular response being off-target. Their technique tacitly assumes uncorrelated responses.

3.2.1.4 Optimization of Multi-variate Loss Functions

For the expected multivariate quadratic loss of equation 3.2, Pignatiello (1993) introduced a two-step procedure for finding the design variable settings that minimize this composite cost of poor quality. His procedure assumes that the responses follow a multivariate normal distribution [i.e., MVN$(E(\mathbf{Y}), \mathbf{\Sigma_Y})$], are NTB and follow an additive model. His two-step

procedure involves minimizing the $trace[\mathbf{A\Sigma_Y}]$ and potentially adjusting the means of $E(\mathbf{Y})$ to target.

Tsui (1999) extended Pignatiello's two-step procedure to situations where responses may be NTB, STB or LTB. He divides the r responses into two subsets:

$\mathbf{Y_1} = (Y_1, \ldots, Y_{r_1})^T$, i.e., those responses whose means can be adjusted to target and,

$\mathbf{Y_2} = (Y_{r_1+1}, \ldots, Y_r)^T$, i.e., those responses whose means can't be adjusted to target.

He defines a corresponding division of the mean and target vectors as:

$E(\mathbf{Y})^T = (E(\mathbf{Y_1}), E(\mathbf{Y_2}))$ and,

$\tau^T = (\tau_1^T, \tau_2^T)$

and the corresponding partitioned components of the $\mathbf{A}$ matrix in equation 3.2 as:

$\mathbf{A_{11}}$, $\mathbf{A_{12}}$, $\mathbf{A_{21}}$, and $\mathbf{A_{22}}$.

Under the assumption that $\mathbf{A}$ is symmetric, the average loss can be written as:

$$\begin{aligned} E(MQL) &= trace[\mathbf{A\Sigma_Y}] + [E(\mathbf{Y_1}) - \tau_1]^T \mathbf{A_{11}} [E(\mathbf{Y_1}) - \tau_1] \\ &\quad + 2[E(\mathbf{Y_1}) - \tau_1]^T \mathbf{A_{12}} [E(\mathbf{Y_2}) - \tau_2] \\ &\quad + [E(\mathbf{Y_2}) - \tau_2]^T \mathbf{A_{22}} [E(\mathbf{Y_2}) - \tau_2] \\ &= trace[\mathbf{A\Sigma_Y}] + \mathbf{OT_1} + \mathbf{OT_{12}} + \mathbf{OT_2}. \end{aligned} \tag{3.15}$$

where $\mathbf{OT_i}$ refers to the respective off-target vector.

For the moment assume the covariance matrix of $\mathbf{Y}$ and the third off-target component are functions of the non-adjustment factors $\mathbf{x_2}$ only. This means $\mathbf{\Sigma_Y} = f(\mathbf{x_2})$ and $\mathbf{OT_2} = f(\mathbf{x_2})$, and assume furthermore that the adjustment factors $\mathbf{x_1}$ can be used to shift the mean vector $E(\mathbf{Y_1})$ to its target τ_1. Since the terms $\mathbf{OT_1}$ and $\mathbf{OT_{12}}$ drop to zero when $E(\mathbf{Y_1}) = \tau_1$, the average loss in equation 3.15 can be minimized by the following two-step procedure for this set of assumptions

1. find values of $\mathbf{x_2}$ that minimize $trace[\mathbf{A\Sigma_Y}] + \mathbf{OT_2}$, say $\mathbf{x_2^*}$,

2. at the values of $\mathbf{x_2^*}$ chosen in the first step, find values of $\mathbf{x_1}$ that shift the mean vector $E(\mathbf{Y_1})$ to its target τ_1.

Note that the assumptions above are parallel to the assumptions of the single characteristic problem under an additive model. Hence the two-step procedure above is only appropriate when the responses follow an additive model. In the same article Tsui additionally derives two-step procedures for constrained and unconstrained minimization of the multivariate quadratic loss function for non-additive models.

To this point we have examined squared-error loss functions whose expected value is decomposed into off-target and variance components. Ribeiro and ElSayed (1995) introduced a multivariate loss function which considers, in addition to off-target and variance components, a factor accounting for fluctuation in the supposedly fixed design variable settings. Use of this gradient loss function assumes models of each response $(Y_1, ..., Y_r)$ as a function of the process design variables $(x_1, ..., x_p)$ and estimates the variability induced on Y_i due to the variability of the process parameters using the following terms:

$$\hat{\sigma}^2_{Y_i} = \sum_{k=1}^{p} \hat{\sigma}^2_{x_k} \left(\frac{\partial Y_i}{\partial x_k} \right)^2 \tag{3.16}$$

for when the fluctuations in $(x_1, \ldots, x_p)$ are independent of each other, where $\hat{\sigma}^2_{Y_i}$ are the estimated variance of each Y_i, $\hat{\sigma}^2_{x_k}$ are the estimated fluctuations in the design parameters, and $\frac{\partial Y_i}{\partial x_k}$ are the model-predicted shifts in Y_i from the random variation of x_k. When the fluctuations in $(x_1, \ldots, x_p)$ are correlated, this variance term is defined as:

$$\hat{\sigma}^2_{Y_i} = \sum_{k=1}^{p} \hat{\sigma}^2_{x_k} \left(\frac{\partial Y_i}{\partial x_k} \right)^2 + \sum_{k \neq l} \hat{\rho}_{kl} \hat{\sigma}^2_{x_k} \hat{\sigma}^2_{X_l} \left(\frac{\partial Y_i}{\partial x_k} \right) \left(\frac{\partial Y_i}{\partial x_l} \right) \tag{3.17}$$

where $\hat{\rho}_{kl}$ is the estimated correlation between each pair of design variables x_k and x_l. The authors' multivariate gradient loss function is then the weighted sum of the individual gradient functions, which for the case of independent variation in the design settings can be expressed as:

$$MGL(\mathbf{x_j}) = \sum_{i=1}^{r} \omega_i \left[(Y_i - t_i)^2 + \hat{\sigma}^2_{Y_i} + \sum_{k=1}^{p} \hat{\sigma}^2_{x_k} \left(\frac{\partial Y_i}{\partial x_k} \right)^2 \right] \tag{3.18}$$

where $MGL(\mathbf{x_j})$ is the multivariate gradient loss function for a particular design vector $\mathbf{x_j}$, and ω_i and t_i are respectively the weights and targets of the individual responses. The authors allow for a very explicit, quantitative ranking of the responses through the ω_i term.

They find the optimal process parameters (i.e., design factor settings) through standard non-linear search techniques.

Ribeiro et al (2000) extend the gradient loss function of Ribeiro and ElSayed (1995) by adding a term for manufacturing costs. They first convert the dimensionless loss function values of Ribeiro and ElSayed (1995) into dollars by defining the proportionality constant κ as:

$$\kappa = \frac{\Delta\ Value}{\Delta MGL} = \frac{A_{mp} - B_{mp}}{MGL_A - MGL_B} \tag{3.19}$$

where A_{mp} and B_{mp} are the market prices of class A products, i.e., those with all responses close to target, and of class B products, i.e., those with at least one response out of specification, and where MGL_A and MGL_B are the values of the Ribeiro and ElSayed (1995) loss function corresponding to the class A and B products.

This proportionality constant is then multiplied by the Ribeiro and ElSayed (1995) loss function value to yield the equivalent lost dollar value resulting from a particular group of design settings:

$$C_Q(\mathbf{x}) = \kappa MGL(\mathbf{x}) \tag{3.20}$$

where $\mathbf{x}$ is a vector of design factor settings. They introduce manufacturing costs by starting with a multi-response experiment with r responses ($i = 1, \ldots, r$). They model manufacturing costs as:

$$C_M(\mathbf{x}) = \mathbf{x}^T\theta + \epsilon \tag{3.21}$$

where $\mathbf{x}$ is the design vector of p regressors, θ is a p dimensional vector of regression coefficients and ϵ the residual . Finally an extended multivariate loss function which includes costs of poor quality and manufacturing is defined as:

$$C(\mathbf{x}) = C_Q(\mathbf{x}) + C_M(\mathbf{x}) \tag{3.22}$$

The weighting of responses is accomplished directly through the weighting factor defined for the multivariate gradient loss function of Ribeiro and ElSayed (1995). The authors employ optimization techniques to find the vector $\mathbf{x}$ which minimizes this overall cost function.

The examples in this section demonstrate the additive combination of individual loss and utility functions to create a multivariate loss or utility function. Antony (2000) describes a novel approach to the multi-response optimization problem using principal components analysis (PCA) and some slightly confusing terminology. He defines a utility function (i.e., a function measuring positive value or worth) for each response and calls it the normalized quality loss function (NQL). While this technique does not simply add univariate utility functions to obtain a multivariate one, it does form linear combinations of individual response utility functions to explain the variability among them, enroute to finding the optimal vector of design variable settings.

Assuming models and loss functions for each of the r responses of interest, the (NQL) is calculated as:

$$NQL_{ij} = \frac{L_{i+} - L_{ij}}{L_{i+} - L_{i-}} \tag{3.23}$$

where

L_{ij} = quality loss for the i_{th} response at the j_{th} experimental run,
L_{i+} = maximum quality loss for the i_{th} response,
L_{i-} = minimum quality loss for the i_{th} response
which cause the NQL values to range from zero to one.

For each of the n experimental runs, the multi-response performance statistic Z_{1j} is the first principal component, i.e., the uncorrelated linear combination of highest eigenvalue, of the NQLs for all responses.

$$Z_{1j} \quad = \quad a_{1j}NQL_1 + a_{2j}NQL_2 + \cdots + a_{rj}NQL_r \tag{3.24}$$

with the restriction that $\sum_{i=1}^{r} a_{ij}^2 = 1$ for all $j = 1, \ldots, \ n$.

Z_1 is then modelled as a function of the design variables and optimization techniques are applied to this model to find the design variable settings that maximize it. Through normalization the differences in response scaling are removed but the technique does not describe any explicit way of weighting the individual responses. Assuming one uses the squared-error loss function for each response, proportionality constants can be chosen to rank the responses prior to their normalization.

3.2.2 Multivariate Utility Functions from Multiplicative Combination

In this section, a multivariate desirability function is constructed from the geometric average of the individual desirability functions of each response.

The geometric average of r components $(d_1, \ldots, d_r)$ is the r_{th} root of their products:

$$GA(d_1, \ldots, d_r) = \left[\prod_{i=1}^{r} d_i\right]^{\frac{1}{r}} \tag{3.25}$$

The GA is then a multiplicative combination of the individuals. When combining individual utility functions whose values are scaled between zero and one, the GA yields a value $\leq$ the lowest individual utility value. For rating the composite quality of a product, this prevents any single response from reaching an unacceptable value, since a very low value on any crucial characteristic (e.g., safety feature or cost) will render the entire product worthless to the end user.

To demonstrate the simplest case of a geometric average we show the case of combining two squared-error loss functions. For responses (Y_1, Y_2) with respective loss functions $L(Y_1)$ $= a_1(Y_1 - t_1)^2$ and $L(Y_2) = a_2(Y_2 - t_2)^2$ where (t_1, t_2) are the respective targets of the responses, the geometric average of the two loss functions is:

$$\begin{aligned} GA\left[(L_1), (L_2)\right] &= \sqrt{a_1(Y_1 - t_1)^2 a_2(Y_2 - t_2)^2} \\ &= \sqrt{a_1 a_2}\ (Y_1 - t_1)(Y_2 - t_2) \end{aligned} \tag{3.26}$$

which we recognize as the cross product of the roots of the squared-error loss functions of each response.

3.2.2.1 Modifications of the Desirability Function

In order to allow the DDM to place the ideal target value anywhere within the specifications, Derringer and Suich (1980) introduced a modified version of the desirability functions of Harrington (1965). Their desirability function for a one-sided specification is:

$$d_i = \begin{cases} 0 & Y_i \leq Y_{i*}, \\ \left[\frac{Y_i - Y_{i*}}{Y_i^* - Y_{i*}}\right]^{\phi_i} & Y_{i*} \leq Y_i \leq Y_i^*, \\ 1 & Y_i \geq Y_i^*. \end{cases} \tag{3.27}$$

where Y_{i*} and Y_i^* are the minimal and maximal acceptable levels of Y_i respectively, Y_i is the response predicted by a certain set of design variable settings, and ϕ_i is a positive constant whose increasing magnitude creates a correspondingly more convex desirability curve. That is, values of $\phi_i < 1$ create concave curves allowing higher desirability values with Y_i values relatively close to the minimal acceptable level, while values of $\phi_i > 1$ only allow high desirability values when Y_i is very close to the maximal acceptable level. When $\phi_i = 1$ the d_i value is a linear scale between the minimal and maximal acceptable values of the response.

Their desirability function for two-sided specifications is:

$$d_i = \begin{cases} \left[\frac{Y_i - Y_{i*}}{t_i - Y_{i*}}\right]^{\varphi_i} & Y_{i*} \leq Y_i \leq t_i, \\ \left[\frac{Y_i - Y_i^*}{t_i - Y_i^*}\right]^{\psi_i} & t_i \leq Y_i \leq Y_i^*, \\ 0 & Y_i < Y_{i*} \text{ or } Y_i > Y_i^*. \end{cases} \tag{3.28}$$

where t_i is the target value of Y_i, and φ_i and ψ_i are positive constants chosen by the DDM to indicate the importance of an individual response being close to its target. That is larger values of φ_i and ψ_i create a desirability curve with a sharper peak at the target value with more rapid drop-off as the response moves off-target. Lower values of φ_i and ψ_i create a flatter desirability curve that is much less sensitive to a response being off-target.

The desirability function of Harrington is a special case of the Derringer and Suich (1980)

desirability functions, which permit a target value t_i anywhere within the specification limits. Like Harrington, they do not provide for explicit weighting of the individual responses in the overall desirability function. Rather ranking is implied by the relative steepness of the gradients of the desirability curves which are in turn the results of specific choices of ϕ_i, φ_i and ψ_i.

Derringer (1994) added explicit weighting terms to the geometric average of the individual desirability functions as follows:

$$D = GA(d_1^{\omega_1}, \ldots, d_r^{\omega_r}) = \left(\prod_{i=1}^{r} d_i^{\omega_i} \right)^{\frac{1}{\sum \omega_i}} \tag{3.29}$$

where setting all the $\omega_i = 1$ yields the geometric mean of equation 3.25.

DelCastillo et al (1996) note that since the desirability functions of Harrington (1965) and Derringer and Suich (1980) are non-differentiable at the target points, only direct search optimization methods are applicable. Since the much more efficient gradient based methods require first order differentials at all points, they propose using a piecewise continuous desirability function in which the non-differentiable points are corrected using a local polynomial approximation.

Kim and Lin (2000) propose finding the vector of design variable settings $\mathbf{x}$ which maximizes the minimum level that the geometric average of the individual desirability functions may obtain. They state the multi-response optimization problem as:

$$\max_{\mathbf{x} \in \mathbf{\Omega}} (\iota) \tag{3.30}$$

where ι is the minimum degree of satisfaction subject to $d_i(\mathbf{x}) \geq \iota$ for $i = 1, 2, \ldots, r$ and for $\mathbf{x} \in \mathbf{\Omega}$, where $d_i(\mathbf{x})$ are the desirability functions of the individual estimated responses $Y_i(\mathbf{x})$. The goal is to identify the $\mathbf{x}$ which maximizes the minimum degree of satisfaction (ι) with respect to all the responses within the experimental region , i.e.,

$$\max_{\mathbf{x} \in \mathbf{\Omega}} (\min[d_1(\mathbf{x}), \ldots, d_r(\mathbf{x})]) \tag{3.31}$$

They assert that the advantage of this approach is that it does not assume any form or degree of the estimated response models and is insensitive to the potential dependence between responses. They contrast this approach with the method of Khuri and Conlon (1981)

which uses only the subset of independent responses and requires that all the independent responses have same order polynomial models of the same subset of design variables. Furthermore the ι term embodies the overall degree of satisfaction and allows for a quantitative way to compare the results induced by different $\mathbf{x}$ vectors.

They suggest a desirability function of the form:

$$d(z) = \begin{cases} \frac{\exp(\varsigma)-\exp(\varsigma z)}{\exp(\varsigma)-1}, & \text{if } \varsigma \neq 0, \\ 1-z, & \text{if } \varsigma = 0. \end{cases} \tag{3.32}$$

where ς is a constant ($-\infty \leq \varsigma \leq \infty$) called the exponential constant and z is a standardized parameter representing distance of the estimated response from its target in units of maximum allowable deviation.

z is calculated differently depending on whether the response is NTB, STB or LTB. For the NTB case with a symmetric desirability function, z is defined as:

$$\begin{aligned} z &= \frac{Y_i(\mathbf{x}) - t_i}{Y_i^{\max} - t_i} \\ &= \frac{Y_i(\mathbf{x}) - t_i}{t_i - Y_i^{\min}} \end{aligned} \tag{3.33}$$

for $Y_i^{\min} \leq Y_i(\mathbf{x}) \leq Y_i^{\max}$ and where t_i is the target of response i and $Y_i^{\min}$ and $Y_i^{\max}$ are respectively the minimum and maximum values of the individual response. This z function for the NTB case is easily modified for an asymmetric desirability function.

For the STB case:

$$z = \begin{cases} \frac{Y_i(\mathbf{x})-Y_i^{\min}}{Y_i^{\max}-Y_i^{\min}}, & \text{for } Y_i^{\min} \leq Y_i(\mathbf{x}) \leq Y_i^{\max} \end{cases} \tag{3.34}$$

For the LTB case:

$$z = \begin{cases} \frac{Y_i^{\max}-Y_i(\mathbf{x})}{Y_i^{\max}-Y_i^{\min}}, & \text{for } Y_i^{\min} \leq Y_i(\mathbf{x}) \leq Y_i^{\max} \end{cases} \tag{3.35}$$

It is easily verified that (-1 $\leq z \leq$ +1) for an NTB response and ($0 \leq z \leq 1$) for the STB and LTB responses. In all cases $d(z)$ is maximized when $z = 0$ which happens when $Y_i(\mathbf{x})$ is equal to the target value.

The choice of ς determines the relative concavity of the individual desirability curves. That is, increasingly negative values of ς produce decreasingly concave curves (i.e., more convex) and increasingly positive values of ς produce increasingly concave desirability curves. They define relative concavity as relative insensitivity to a response being off-target and relative convexity as sensitivity to the off-target distance.

For example, a response with $\varsigma = -5$ has a desirability curve whose values drop sharply with increasing off-target distance while a response with $\varsigma = +5$ has desirability values that change slowly as the response moves further off-target. The relative values of the ς variable approximate the weighting of the individual responses on the geometric average. Hence weighting is accomplished indirectly by choosing lower ς values for the responses of higher priority.

Kim and Lin (2000) furthermore incorporate a technique that accounts for the predictive ability of the individual response models. They do this by transforming the original ς values to ς' indicative of the predictive ability of the response models.

For example,

$$\varsigma' = \varsigma + (1 - R^2)(\varsigma^{\max} - \varsigma) \tag{3.36}$$

will decrease each ς value inversely with rising R^2, where R^2 is the standard coefficient of determination in linear regression and ς^{max} is a sufficiently large value of ς such that $d(z)$ with ς^{max} is extremely concave, hence having negligible effect on the optimization.

This makes the resulting desirability curve more convex, i.e., of higher priority in the geometric average, as the R^2 values increase. This effectively adjusts the relative weighting of each individual desirability function according to the predictive ability of the corresponding response model so that better predictive models get higher weighting than poorly predictive models. They demonstrate the attainment of a design vector yielding a higher overall desirability using the ς' transformation than that obtained with the original ς value. Although the authors use R^2 in their example, the DDM can use any preferred metric of predictive ability. In general the desirability function approach neither assumes response independence nor exploits the response correlation information.

3.2.3 The Non-Domination Search Technique

Loy et al (2000) extended the Dual Response Approach of Vining and Myers (1990) to the multi-response case by searching for groupings of design variable vectors which yield responses which are non-dominated with respect to each other for each of the r separate responses. This approach assumes a finite sample space with models for both mean and variance of each response so that they can be predicted for all responses from all possible design vectors.

This technique's relation to the Dual Response Approach is evident in the following formulations for the NTB, LTB and STB cases:

$$\begin{aligned} NTB: \quad & \min f_1(\mathbf{x}) = Y(\mathbf{x}) - \tau \\ & \min f_2(\mathbf{x}) = \hat{\sigma}_Y^2 \end{aligned} \tag{3.37}$$

$$\begin{aligned} LTB: \quad & \max f_1(\mathbf{x}) = Y(\mathbf{x}) \\ & \min f_2(\mathbf{x}) = \hat{\sigma}_Y^2 \end{aligned} \tag{3.38}$$

$$\begin{aligned} STB: \quad & \min f_1(\mathbf{x}) = Y(\mathbf{x}) \\ & \min f_2(\mathbf{x}) = \hat{\sigma}_Y^2 \end{aligned} \tag{3.39}$$

wherein each response $f_1(\mathbf{x})$ is optimized with the constraint of minimizing variance, i.e., $f_2(\mathbf{x}) = \sigma_Y^2$.

We now demonstrate the meaning of non-domination using the NTB case of equation 3.37 as an example. A vector of design variable settings (i.e., $\mathbf{x1}$) is said to dominate another design vector $\mathbf{x2}$ when no value of $[f_1(\mathbf{x2}) \ , \ f_2(\mathbf{x2})]$ is less than the corresponding element of $[f_1(\mathbf{x1}) \ , \ f_2(\mathbf{x1})]$, and at least one value of $f_1(\mathbf{x2})$ or $f_2(\mathbf{x2})$ is strictly greater than the corresponding element of $[f_1(\mathbf{x1}) \ , \ f_2(\mathbf{x1})]$.

When the first grouping of non-dominated design vectors (i.e., front) is formulated for each of the r responses, the DDM searches for the intersection of these r first fronts. This intersection set can range from the empty set to a large number of design vectors. For the empty set the DDM proceeds to examine the intersections of the successive fronts of the r responses. The authors apply this procedure to the military field ration study of Wurl and Albin (1999) and identify design vectors yielding comparable results to those identified using the expected loss approach of Pignatiello (1993) and the desirability approach of Derringer and Suich (1980).

This technique differs from the expected loss and desirability approaches in that it allows the DDM to proceed without subjectively prioritizing or combining the multiple response loss or utility functions into a single overall objective function. This means the DDM does not have to struggle with different engineering units or the conceptual challenge of combining different types of responses into a single objective function. Likewise the DDM does not have to consider the tradeoff between off-target and variance components of the same overall objective function.

This effectively shifts the required engineering judgement from the front end of the optimization process to the end of the process, when the DDM must consider the non-dominated design vectors against each other according to the same considerations usually used in the other approaches to weigh and combine the individual response functions. With respect to response correlation structure, this technique neither assumes response independence nor exploits this source of statistical information.

3.3 The Compromise Decision Support Problem

Up to this point, all the multi-response optimization techniques have been culled from the statistics literature. To contrast these approaches with an example from the engineering literature, we briefly review an important multi-response technique, which evolved from experimentation in the ship building and design industry. The compromise Decision Support Problem (cDSP) is a mathematical construct with which the conflicting goals in product design are modelled. The following high level description of the baseline, deterministic

Table 3.1: Baseline cDSP from Mistree et al (1993)

Given	p,	the number of system variables
	ζ,	the number of equality constraints
	η,	the number of inequality constraints
	ϑ,	the number of objectives
	$\Psi_i(\mathbf{x})$,	the goal achievement functions
	$\gamma_h(\mathbf{x})$,	the constraint functions
	G_i,	the target values for the goals
	$f_K(\delta_i)$,	function to be minimized at priority level K for preemptive form
	ω_i,	weight for the Archimedean form
Find	System Variables	$\mathbf{x} = (x_1, \ldots, x_p)$
	Deviation Variables	$\delta_i^-, \delta_i^+ \; for \; i = 1, \ldots, \vartheta$
Satisfy	Goals	$\Psi_i(\mathbf{x}) \; + \; \delta_i^- \; - \; \delta_i^+ \; = \; G_i$
	Constraints	$\gamma_h(\mathbf{x}) \; = \; 0 \; for \; h = 1, \ldots, \zeta$
		$\gamma_h(\mathbf{x}) \; \geq \; 0 \; for \; h = \zeta + 1, \ldots, \zeta + \eta$
	Bounds	$x_{l,min} \; \leq \; x_l \; \leq \; x_{l,max} \; for \; l = 1, \ldots, p$
	Other	$\delta_i^- \cdot \; \delta_i^+ = 0 \;\; and \;\; \delta_i^-, \delta_i^+ \; \geq \; 0$
Minimize	Deviation Function (DF)	Archimedean: $DF = \sum_{i=1}^{2\vartheta} (\omega_i^- \delta_i^- \; + \; \omega_i^+ \delta_i^+)$ where $\sum_{i=1}^{2\vartheta} \omega_i = 1$
		Preemptive: $DF = \left[f_1(\delta_i^-, \delta_i^+), \ldots, f_K(\delta_i^-, \delta_i^+) \right]$

cDSP is taken from Mistree et al (1993), Table 3.1.

The cDSP is a multi-objective decision model based on mathematical programming and goal programming. In the cDSP, values of design variables are determined to achieve a set of conflicting goals to the best extent possible while satisfying a set of constraints.

The cDSP consists of system variables, system constraints, system goals and deviation variables. Optimization is guided by a deviation function. The system variables $\mathbf{x}' = (x_1, \ldots, x_p)$ usually describe design variables of the system and each cDSP must have at least two system variables which may be continuous, discrete or Boolean. System variables are bounded to help the designer use experience-based judgement in formulating the problem.

System constraints and bounds define the feasible design space and are functions of the system variables only. The designer's aspiration for each response is represented by a system goal (G_i) which is then compared to the actual performance $\Psi_i(\mathbf{x})$ of that same response. The deviation variables (δ_i^-, δ_i^+) are respectively the level of under-achievement

or over-achievement of a goal. The goals are modelled as constraints in the following form:

$$\Psi_i(\mathbf{x}) \; + \; \delta_i^- \; - \; \delta_i^+ \; = \; G_i \tag{3.40}$$

The high level concept of the cDSP is to minimize the difference between the goal (G_i) of each objective and its actual performance $\Psi_i(\mathbf{x})$. This is accomplished by finding settings of the system variables that minimize the deviation function DF, the overall objective function of the cDSP, which is a function of the deviation variables of the system goals.

The deviation function DF is defined for two cases according to how the DDM prioritizes the design objectives:

Case 1: Archimedean Weighting

$$\min_{\mathbf{x}} DF = \begin{cases} \sum_{i=1}^{r} (\omega_i^- \delta_i^- \; + \; \omega_i^+ \delta_i^+) \\ \text{where } \sum_{i=1}^{r} \omega_i = 1 \\ \text{for symmetric weighting} \\ \text{and } \omega_i \geq 0; \; i = 1, \ldots, \; r \end{cases} \tag{3.41}$$

or

Case 2: Lexicographic Weighting

$$\min_{\mathbf{x}} DF = \Big\{ \left[f_1(\delta_i^-, \delta_i^+), \ldots, f_K(\delta_i^-, \delta_i^+) \right] \tag{3.42}$$

where ω_i^- and ω_i^+ are the potentially asymmetric weights placed respectively on the under-achievement or over-achievement of the respective design objectives (i.e., goals).

In this case, the DDM chooses the form of the deviation function based on his/her choice of Archimedean or Lexicographic weighting of the design objectives. In the Archimedean approach the DDM explicitly assigns the weights ω_i^- and ω_i^+ to reflect the importance of the individual goals. Lexicographic weighting (Ignizio (1982)) does not require the DDM to assign specific weights to the objectives. Rather goals are rank ordered in terms of their

priority and deviation variables in the highest priority goal are minimized first, followed by the deviation variables in the second priority goal, and so on.

While limited by space to discussing the deterministic baseline cDSP, there exist a number of extensions which reflect the cDSP's nature as a living construct, capable of being strengthened and/or specialized through augmentation. Bayesian and fuzzy logic versions of the cDSP (see respectively Vadde et al and Zhou et al (1992)) allow the DDM to use the cDSP while accommodating uncertainty regarding constraints, actual performance or weighting of goals in the problem formulation.

In summary the cDSP identifies the vectors of design variable settings that satisfy all constraints, i.e., all system constraints and system goals modelled as constraints per equation (4.1), and that minimize the chosen deviation function. Since the cDSP identifies a group of satisficing solutions, the optimal design solution can be viewed as a subset of the cDSP results. Like the desirability function and the Non-Domination search techniques, the cDSP makes no assumptions regarding response correlation and does not actively use response correlation structure in its optimization algorithm.

The apparent strength of the cDSP is its handling of highly constrained environments. It provides flexible decision support for achieving compromise among multiple goals while satisfying constraints and bounds. It is also domain independent and implementable with reasonable effort.

3.4 Comparing the Multi-response Optimization Techniques

In this section we compare the relative strengths and weaknesses of the four multi-response design optimization techniques which may be used in the solution of engineering design problems. These four techniques are expected loss, desirability, the non-domination search and the compromise Decision Support Problem.

A comparison of the four techniques with respect to the key metrics discussed in this paper is presented in the table on the next two pages which is a slight modification of the table in Loy et al (2000). We remind the reader that because all four techniques can handle the NTB, STB and LTB response types, this criterion is not included in the table.

Table 3.2: Comparing the Multi-response Optimization Techniques

Technique (columns) Metric (rows)	Expected Loss	Desirability	Non-Domination Search	cDSP
Max. number of input variables	Lim. only by experimental & computational resources	Lim. only by experimental & computational resources	Max. 6 recommended	Lim. only by experimental & computational resources
Use of Response Correlations	Directly Used	Not Used	Indirectly Used	Only if variance measure is a stated goal
Max. number of responses	Lim. only by experimental & computational resources	Lim. only by experimental & computational resources	Max. 3 recommended	Lim. only by experimental & computational resources
Experimental Replication Required?	Yes	No	No	Only if variance measure is a stated goal
Modeling Required	Overall exp. loss or mean and variance	Overall desirability or mean and variance	Mean and variance	Continuous or Boolean variable response models
Formulation of Mean	Included	Included	Included	Only if mean measure is a stated goal
Formulation of Variance	Explicit	Implicit	Included	Only if variance measure is a stated goal
Working Units of Objective Function	Monetary loss units	Dimensionless	Original engineering units	Dimensionless
Bounds on Response Values	Not required	Required	Not required	Required

Table 3.3: Addtional Criteria of Multi-response Optimization Techniques

Technique (columns) / Metric (rows)	Expected Loss	Desirability	Non-Domination Search	cDSP
Weighting of Responses	Specified thru loss function coefficients	Indirectly Specified thru Individual Desirability Curves	Implicit	Lexicographic and Archimedean Weights
Handling of Constraints	May not be possible	May not be possible	Relatively easy	Designed for many, differentiable constraints
Penalty for Deviation from Target	Quadratic Functions assumed, coefficients required	Explicit Functions Required	Implicit	Not Applicable
Trade-Off Between Off-Target and Variance	Equal	Not Applicable	Implicit	Not Applicable
Resulting Vector of Design Variable Settings	One $\boldsymbol{x}$ vector or none	One $\boldsymbol{x}$ vector or none	Alternative $\boldsymbol{x}$ vectors likely	One $\boldsymbol{x}$ vector or alternative $\boldsymbol{x}$ vectors depending on formulation
Applicable Optimization Settings	Nonlinear Search	Nonlinear Search	Exhaustive Non-Domination Search	Sequential and Adaptive Linear Programming
Consequence of New Target	Direct Search: Need to re-optimize, Two-step: no need to re-optimize	Need to re-optimize	Need to re-optimize	Need to re-optimize
Consequence of New Specification	No need to re-optimize	Need to re-optimize	Need to re-optimize	Need to re-optimize

When considering these approaches, major concerns include the software and computational resources necessary for carrying out the optimization algorithms. That topic is not addressed in this paper since we have focused on the statistical and optimizing properties of the approaches without detailed consideration of computational issues.

In Tables 3.2 and 3.3 the term expected loss encompasses all the variations on the additive combination of univariate loss functions to form a multi-response loss function. The optimization goal is to minimize expected loss and hence a number of direct search and two-step procedures have been discussed. Variations include a version which minimizes total cost, including both quality loss and manufacturing cost components. Another well known version is the Mahalanobis Distance which minimizes the Euclidean distance between the responses and their ideal target values and incorporates the covariance structure of the responses. Another version incorporates a term accommodating the variation from fluctuations in the design variables themselves hence providing another means of boosting the robustness of the final solution.

The biggest advantage of the expected loss approach is that it permits, under certain restrictions such as additive response models, normality and a partitioning of the design variables into adjustment and non-adjustment variables, a reduction in the dimensionality of the optimization procedure to a two-step procedure. The design solution obtained this way does not need complete re-optimization for changes in target or specifications since the assumption of adjustment factors allows for easy shifting of the response mean. Other advantages include explicit alternatives for handling mean, variance, cost and weighting of the individual responses and a wide variety of options of employing the response correlation structure. The disadvantages of expected loss are that it does not easily handle constraints and to date has not been used to consider a hierarchical structuring of loss functions corresponding to the assembly, subassembly and component levels of a product architecture.

In Tables 3.2 and 3.3 the term desirability encompasses all the variations on the multiplicative combination of univariate desirability functions to form a multi-response desirability function. The optimization goal is to maximize overall desirability and we've reviewed

versions of the desirability function compatible with reduced gradient search methods and a prioritizing of the individual responses through the selection of relative concavity of the individual desirability curves. Another version discussed weighting the contribution of individual response desirability functions according to the predictive value of their respective response models.

The advantages of the desirability approach are that it is an intuitive approach requiring no background in statistics and the combining of the individual desirabilities via the geometric average prevents any single response from reaching an unacceptable quality level. Its dimensionless units also remove the problem of different engineering units for different responses. The disadvantages of the desirability approach are that it does not explicitly measure cost of poor quality, does not easily incorporate constraints and to date has not been used to consider a hierarchical structuring of loss functions corresponding to the assembly, subassembly and component levels of a product architecture.

The advantages of the non-domination search technique include its formation of each response as a separate optimization problem which frees the DDM from needing to consider an overall multi-response objective function which combines different engineering units and response types. It also handles constraints in a relatively easy way and does not require the DDM to prioritize the individual responses before starting. Since it assumes a finite sample space it also generates responses for all possible design vectors and the optimal solution it derives is always one of the existing design vectors. It is also easily coded into an MS Excel spreadsheet which makes its application potentially universal.

Its major disadvantage is the limited number of design variables and responses easily managed. It is recommended for use with a maximum of six design variables and three responses which greatly limits its range of useful applications. Larger numbers of design variables and responses are likely to generate an impractically large number of non-dominated design vectors as potential solutions. For engineering design problems with tens, hundreds or even thousands of input variables, the non-domination search method is not practical.

In Tables 3.2 and 3.3 the term compromise Decision Support Problem (cDSP) includes the various versions of this technique which include Bayesian, fuzzy logic and utility function

versions which for space considerations have not been discussed in this paper. We have only discussed the deterministic version but in comparing this technique have mentioned the collective attributes of the multiple versions in the table.

The biggest advantage of the cDSP approach is that it readily accommodates highly constrained design environments. It was designed specifically for this purpose and allows the DDM great flexibility in setting the goals and constraints on the responses including response mean and variance. Additionally it lends itself to a hierarchical structuring of constraints for structurally complex product design and product platform design. Its primary disadvantage is its use of a specific sequential and adaptive linear programming algorithm (i.e., non-commercial software).

CHAPTER IV

LOSS-SCALED PRINCIPAL COMPONENTS

In this chapter we explore the mathematical properties of the loss-scaled principal components (i.e., LSPC). In section 1 we repeat the definition of LSPC presented in the first chapter. In section 2 we discuss the five variants of the general case and summarize their properties in a table. In section 3 we draw conclusions regarding the properties derived in Section 2.

4.1 What are Loss-Scaled Principal Components?

In the first chapter we defined PCs and their relevant properties. While PCs are one of the oldest multivariate techniques in existence, their variance optimizing properties continue finding new application areas. In Chapter 3 we reviewed multivariate robust design techniques and commented on their respective use of the correlation structure of the responses.

Of the multivariate techniques considered, minimizing the expected value of the MQL makes the best use of the correlation structure and can be decomposed into covariance and off-target vector product terms as shown below:

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \\ &= trace[\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{Y}}] + [E\mathbf{Y}-\tau]^T\mathbf{A}[E\mathbf{Y}-\tau]. \end{aligned} \tag{4.0}$$

To minimize the two terms of the expected MQL, one must simultaneously consider the contributions of the loss matrix $\mathbf{A}$ as well as the response covariance/correlation matrix. The primary conceptual contribution of this thesis is the formation of PCs which simultaneously account for the loss constant matrix and the response covariance/correlation matrix in the MQL. Since the loss constant matrix $\mathbf{A}$ is assumed to be symmetric and positive definite it

can be decomposed by either Cholesky decomposition or diagonalized in the following way:

$$\mathbf{G^T A G} \;=\; \mathbf{\Lambda_A},$$

where the columns of $\mathbf{G}$ are the eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda_A}$ is a diagonal matrix, whose diagonal elements are the eigenvalues of $\mathbf{A}$. This diagonalization allows $\mathbf{A}$ to be expressed as

$$\begin{aligned} \mathbf{A} &= \mathbf{G\Lambda_A G^T} \\ \mathbf{A} &= (\mathbf{G\Lambda_A^{\frac{1}{2}}})(\mathbf{G\Lambda_A^{\frac{1}{2}}})^{\mathbf{T}}. \end{aligned}$$

If we then define $\mathbf{q} = (\mathbf{G\Lambda_A^{\frac{1}{2}}})$, we can express $\mathbf{A}$ in the following way:

$$\mathbf{A} \;=\; \mathbf{qq^T}.$$

This allows Equation (4.0) to be written as follows:

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \\ &= E(\mathbf{q^TY}-\mathbf{q^T}\tau)^T(\mathbf{q^TY}-\mathbf{q^T}\tau) \\ &= trace[\mathbf{q^T\Sigma_Y q}] + (E\mathbf{q^TY}-\mathbf{q^T}\tau)^T(E\mathbf{q^TY}-\mathbf{q^T}\tau) \\ &= trace[\mathbf{q^T\Sigma_Y q}] + [\mathbf{q^T}(E\mathbf{Y}-\tau)]^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]. \end{aligned} \tag{4.-10}$$

Therefore multiplying the original response vectors by the matrix $\mathbf{q}^T$ incorporates the contents of matrix $\mathbf{A}$ and $\Sigma_\mathbf{Y}$ directly into the matrix $\mathbf{q}^T\Sigma_{\mathbf{Yq}}$. We will see in the next section how this simplifies the expression of the covariance and off-target vector product terms on the right side of equation 4.0.

Loss-scaled responses are simply $\mathbf{q^TY}$ and their principal components are formed by multiplying $\mathbf{Y}$ by the eigenvectors of $\mathbf{q}^T\Sigma_{\mathbf{Yq}}$. If we use $\mathbf{Z}$ to denote the full set of LSPC

formed from the covariance matrix of the loss-scaled responses, i.e., $\mathbf{q}^T\Sigma_{\mathbf{Yq}}$, then equation 4.0 can be expressed as:

$$\begin{aligned} E(MQL) &= E\left[(\mathbf{Z}-\tau_{\mathbf{Z}})^T(\mathbf{Z}-\tau_{\mathbf{Z}})\right] \\ &= trace[\Sigma_{\mathbf{Z}}] + (E\mathbf{Z}-\tau_{\mathbf{Z}})^T(E\mathbf{Z}-\tau_{\mathbf{Z}}). \end{aligned}$$

4.2 Decomposing Multivariate Quadratic Loss with Principal Components

In this section we examine how well a subset of PCs from the covariance/correlation matrix of a multiple response loss-scaled data-set represents the multivariate quadratic loss function (MQL) of the complete data set. We develop the general case by examining five specific combinations of response loss and covariance matrices. The first three cases construct PCs from the eigenvectors of the covariance matrix of the loss-scaled responses (i.e., $\mathbf{q^T\Sigma_Y q}$) and the last two from eigenvectors of the corresponding correlation matrix (i.e., $\mathbf{q^T R_Y q}$).

4.2.1 Case 1: PCs Derived from the Covariance Matrix of Loss-Scaled Responses, i.e. $\mathbf{q}^T\Sigma_{\mathbf{Yq}}$, When $\mathbf{A} = \mathbf{I}_r$ and $\Sigma_{\mathbf{Y}} \neq \mathbf{I}_r$

In this case the expected MQL is,

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T(\mathbf{Y}-\tau) \\ &= trace[\Sigma_{\mathbf{Y}}] + [E\mathbf{Y}-\tau]^T[E\mathbf{Y}-\tau]. \end{aligned}$$

We decompose the covariance matrix of the loss-scaled responses, in this case $\Sigma_{\mathbf{Y}}$, into its r characteristic vectors which form the columns of the matrix $\mathbf{U}$. These columns are by definition orthogonal and of unit length and are the coefficients of the r principal components of the responses $\mathbf{Y} = (Y_1, \ldots, Y_r)^T$, so

$$\begin{aligned} \mathbf{Z} &= \mathbf{U}^T(\mathbf{Y}) \\ \Rightarrow \Sigma_{\mathbf{Z}} &= \mathbf{U}^T\Sigma_{\mathbf{Y}}U. \end{aligned}$$

Inserting the entire vector of r PCs into the expected MQL yields

$$\begin{aligned} E(MQL) &= E(\mathbf{Z}-\tau_{\mathbf{Z}})^T(\mathbf{Z}-\tau_{\mathbf{Z}}) \\ &= trace[\Sigma_{\mathbf{Z}}] + (E\mathbf{Z}-\tau_{\mathbf{Z}})^T(E\mathbf{Z}-\tau_{\mathbf{Z}}) \end{aligned}$$

where $\tau_{\mathbf{Z}}$ is defined as the vector $\mathbf{U^T}\tau$.

The trace component can be re-expressed as

$$trace[\Sigma_{\mathbf{Z}}] \;=\; trace[\mathbf{U^T\Sigma_Y U}],$$

which because of the orthogonality of $\mathbf{U}$ is

$$\begin{aligned} trace[\Sigma_{\mathbf{Z}}] &= trace[\mathbf{U^T\Sigma_Y U}] \\ &= trace[\mathbf{\Sigma_Y U U^T}] \\ &= trace[\mathbf{\Sigma_Y}]. \end{aligned}$$

The off-target vector product can also be re-expressed as

$$\begin{aligned} (E\mathbf{Z}-\tau_{\mathbf{Z}})^T(E\mathbf{Z}-\tau_{\mathbf{Z}}) &= [E(\mathbf{U^T Y})-\mathbf{U^T}\tau]^T[E(\mathbf{U^T Y})-\mathbf{U^T}\tau] \\ &= [\mathbf{U^T}(E\mathbf{Y}-\tau)]^T[\mathbf{U^T}(E\mathbf{Y}-\tau)] \\ &= (E\mathbf{Y}-\tau)^T\mathbf{U U^T}(E\mathbf{Y}-\tau) \\ &= (E\mathbf{Y}-\tau)^T(E\mathbf{Y}-\tau). \end{aligned}$$

So in this case, the full set of r PCs yields exactly the same value of the expected MQL as the complete response data-set. This fact is true for all cases and is the fundamental reason why we can use PCs in this capacity. Since we wish to use the first p PCs instead of the full set, we need to know how using this lower dimensionality data-set affects the value of expected MQL.

Defining the matrix $\mathbf{Zp}$ as the matrix product of the first p columns of $\mathbf{U}$ i.e. $\mathbf{U_p}$, and $\mathbf{Y}$, which implies

$$trace[\Sigma_{\mathbf{Zp}}] \;= \Sigma_{i=1}^{p}\lambda_i$$

where the λ_i represent the eigenvalues of $\Sigma_{\mathbf{Z}}$ in order of descending magnitude. The sum of the first p eigenvalues accounts for a fixed proportion of the total sample variance equal to

$$\frac{\Sigma_{i=1}^{p}\lambda_i}{\Sigma_{i=1}^{r}\lambda_i}.$$

Hence the $trace[\mathbf{\Sigma_{Zp}}]$ component of the expected value of the MQL represents this same fixed proportion of the total response sample variance.

We re-express the off-target vector product ($OTVP$) as a function of this same subset of PCs

$$\begin{aligned}(E\mathbf{Zp}-\tau_{\mathbf{Zp}})^T(E\mathbf{Zp}-\tau_{\mathbf{Zp}}) &= [\mathbf{Up^T}(E\mathbf{Y}-\tau)]^T[\mathbf{Up^T}(E\mathbf{Y}-\tau)] \\ &= (E\mathbf{Y}-\tau)^T\mathbf{UpUp}^T(E\mathbf{Y}-\tau).\end{aligned}$$

Because $\mathbf{Up}$ is only a subset of the original r columns of $\mathbf{U}$, the matrix product $\mathbf{UpUp}^T$ does not simplify to the identity matrix as did the entire set of orthogonal columns in $\mathbf{UU}^T$.

However we can still estimate the fraction of the total OTVP that is accounted for by this subset of p PCs. We construct a fraction whose numerator is the $OTVP$ from Equation (4.-29) and whose denominator is the OTVP from the full-set of r PCs as follows:

$$\begin{aligned}\frac{(E\mathbf{Zp}-\tau_{\mathbf{Zp}})^T(E\mathbf{Zp}-\tau_{\mathbf{Zp}})}{(E\mathbf{Z}-\tau_{\mathbf{Z}})^T(E\mathbf{Z}-\tau_{\mathbf{Z}})} &= \frac{[(E\mathbf{Y}-\tau)]^T\mathbf{UpUp^T}[(E\mathbf{Y}-\tau)]}{[(E\mathbf{Y}-\tau)]^T\mathbf{UU^T}[(E\mathbf{Y}-\tau)]} \\ &= \frac{[(E\mathbf{Y}-\tau)]^T\mathbf{UpUp}^T[(E\mathbf{Y}-\tau)]}{[(E\mathbf{Y}-\tau)]^T[(E\mathbf{Y}-\tau)]} \\ &= \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2}. \qquad (4.\text{-}31)\end{aligned}$$

The scalar value of $OTVP$ is the sum total of the individual $OTVP$-contributions from the full set of LSPC, where $(c_1^2,\ldots,c_r^2)$ are the real, constant values contributed by the respective LSPC. The constants $(c_1^2,\ldots,c_p^2)$ are the scalar contributions from the subset of $p<r$ LSPC being utilized. The numerical value of equation (4.-31) depends on the specific off-target vector i.e., $(E\mathbf{Y}-\tau)$ being evaluated and is bounded as follows,

$$0 \leq \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2} \leq 1.$$

The proof of this property is presented in Appendix A. How well the first p terms in the numerator account for the entire quadratic form of the $OTVP$ depends on the specific structures of $\mathbf{A}$ and $\mathbf{\Sigma_Y}$.

In Appendix B we present the results of simulations from five different multivariate data-sets containing between three and six different responses. These data-sets are combined with between five and ten different loss constant matrices (i.e., LCM) before forming the LSPC. The first two LSPC from each combination of data-set/LCM are used to evaluate how well the $OTVP$ is approximated for twenty arbitrarily constructed off-target vectors. The simulations for Case 1, i.e., $\mathbf{A} = \mathbf{I}_r$, often result in a significant loss of information.

4.2.2 Case 2: PCs Derived from the Covariance Matrix of Loss-Scaled Responses, i.e. $\mathbf{q}^T\Sigma_{\mathbf{Yq}}$, When $\mathbf{A} \neq \mathbf{I}_r$ and $\Sigma_{\mathbf{Y}} \neq \mathbf{I}_r$

In this case the loss constant matrix $\mathbf{A}$ is a positive definite matrix of constants proportional to the costs associated with the responses being off-target. Since $\mathbf{A}$ is assumed to be symmetric and positive definite it can be decomposed in the following way:

$$\mathbf{A} \;=\; \mathbf{q}\mathbf{q}^T.$$

This allows the expected MQL to be written as

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \\ &= E_(\mathbf{q^T Y} - \mathbf{q^T}\tau)^T(\mathbf{q^T Y} - \mathbf{q^T}\tau) \\ &= trace[\mathbf{q^T \Sigma_Y q}] + (E\mathbf{q^T Y} - \mathbf{q^T}\tau)^T(E\mathbf{q^T Y} - \mathbf{q^T}\tau) \\ &= trace[\mathbf{q^T \Sigma_Y q}] + [\mathbf{q^T}(E\mathbf{Y}-\tau)]^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]. \end{aligned} \tag{4.-36}$$

From here on we'll refer to $\mathbf{q^T Y}$ as the loss-scaled response vector. The r principal components of the r loss-scaled responses are defined as

$$\mathbf{ZQ} \;=\; \mathbf{V^T}(\mathbf{q^T Y})$$

where the columns of $\mathbf{V} = (\mathbf{V_1}, \ldots, \mathbf{V_r})$ are orthogonal and of unit length and whose diagonal covariance matrix is

$$\begin{aligned} \Sigma_{\mathbf{ZQ}} &= \mathbf{V^T}\mathbf{\Sigma}_{(\mathbf{q^T Y})}\mathbf{V} \\ &= \mathbf{V^T}(\mathbf{q^T \Sigma_Y q})\mathbf{V}. \end{aligned}$$

It is simple to show that the trace and off-target vector product component of Equation (4.-36) obtained by substituting the entire set of PCs defined by **ZQ** are equal to their counterparts from the complete multi-response data-set by repeating the steps presented in Case 1.

We plan to use only the first p principal components of the loss-scaled responses to reduce the dimensionality of the optimization problem. We define the matrix **ZQp** as the matrix product of the first p columns of **V** i.e. **Vp**, with the loss-scaled response vector $\mathbf{q^T Y}$, which implies

$$trace[\Sigma_{\mathbf{ZQp}}] \quad = \Sigma_{i=1}^{p} \lambda_i$$

where the λ_i represent the characteristic values (i.e., eigenvalues) of $\mathbf{\Sigma_{ZQ}}$.

The sum of the first p eigenvalues accounts for a fixed proportion of the total sample variance equal to

$$\frac{\Sigma_{i=1}^{p} \lambda_i}{\Sigma_{i=1}^{r} \lambda_i} \ .$$

Hence the $trace[\Sigma_{\mathbf{ZQp}}]$ component of the expected value of the MQL represents this same fixed proportion of the total loss-scaled response sample variance. We can re-express the p dimensional approximation of the $OTVP$ as

$$\begin{aligned} (E\mathbf{ZQp} - \tau_{\mathbf{ZQp}})^T (E\mathbf{ZQp} - \tau_{\mathbf{ZQp}}) &= [\mathbf{Vp^T q^T}(E\mathbf{Y} - \tau)]^T [\mathbf{Vp^T q^T}(E\mathbf{Y} - \tau)] \\ &= [\mathbf{q^T}(E\mathbf{Y} - \tau)]^T \mathbf{VpVp^T} [\mathbf{q^T}(E\mathbf{Y} - \tau)]. \end{aligned} \tag{4.-42}$$

Because **Vp** is only a subset of the original r columns of **V** , the matrix product $\mathbf{VpVp}^T$ does not simplify to the identity matrix as the entire set of orthogonal columns does in $\mathbf{VV^T}$.

However we can still estimate the fraction of the total $OTVP$ that is accounted for by this subset of p PCs. We construct a fraction whose numerator is the $OTVP$ from

Equation (4.-42) and whose denominator is the $OTVP$ from the full-set of r PCs.

$$\begin{aligned}\frac{(E\mathbf{Zp}-\tau_{\mathbf{Zp}})^T(E\mathbf{Zp}-\tau_{\mathbf{Zp}})}{(E\mathbf{Z}-\tau_{\mathbf{Z}})^T(E\mathbf{Z}-\tau_{\mathbf{Z}})} &= \frac{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T\mathbf{VpVp^T}[\mathbf{q^T}(E\mathbf{Y}-\tau)]}{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T\mathbf{VV^T}[\mathbf{q^T}(E\mathbf{Y}-\tau)]} \\ &= \frac{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T\mathbf{VpVp}^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]}{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]}\end{aligned} \tag{4.-43}$$

The scalar value of $OTVP$ is the sum total of the individual $OTVP$-contributions from the full set of LSPC, where $(c_1^2,\ldots,c_r^2)$ are the real, constant values contributed by the respective LSPC. The constants $(c_1^2,\ldots,c_p^2)$ are the scalar contributions from the subset of $p<r$ LSPC being utilized. The numerical value of equation (4.-43) depends on the specific off-target vector i.e., $(E\mathbf{Y}-\tau)$ being evaluated and is bounded as follows,

$$0 \quad \leq \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2} \leq 1.$$

The proof of this property is the same as that of 4.-31 presented in Appendix A. How well the first p terms in the numerator account for the entire quadratic form of the $OTVP$ depends on the specific structures of $\mathbf{A}$ and $\mathbf{\Sigma_Y}$.

The simulations in Appendix B for specific Case 2 situations, such as when neither $\mathbf{A}$ nor $\hat{\Sigma}_{\mathbf{Y}}$ are diagonal, result in no significant loss of information.

It is interesting to note that the Mahalanobis Distance (MD) is a special example of the Case 2 MQL where $\mathbf{A}=\mathbf{\Sigma^{-1}}$.

$$MD(\mathbf{Y},\tau) \quad = \quad (\mathbf{Y}-\tau)^T\mathbf{\Sigma^{-1}}(\mathbf{Y}-\tau),$$

therefore

$$\begin{aligned}E(MQL) &= E(MD) \\ &= trace[\mathbf{\Sigma_Y^{-1}\Sigma_Y}]+(E\mathbf{Y}-\tau)^T\mathbf{\Sigma_Y^{-1}}(E\mathbf{Y}-\tau) \\ &= trace[\mathbf{I_r}]+(E\mathbf{Y}-\tau)^T\mathbf{\Sigma_Y^{-1}}(E\mathbf{Y}-\tau) \\ &= r+(E\mathbf{Y}-\tau)^T\mathbf{\Sigma_Y^{-1}}(E\mathbf{Y}-\tau)\end{aligned}$$

As the inverse of a positive definite matrix, $\mathbf{\Sigma_Y^{-1}}$ is positive definite and can be decomposed as

$$\mathbf{\Sigma_Y^{-1}} \quad = \quad \mathbf{hh^T}$$

where $\mathbf{h} = \Sigma_Y^{-\frac{1}{2}}$. This allows the off-target vector product to be re-expressed as

$$(E\mathbf{Y} - \tau)^T \mathbf{\Sigma_Y^{-1}} (E\mathbf{Y} - \tau) \quad = \quad [\mathbf{h^T}(E\mathbf{Y} - \tau)]^T [\mathbf{h^T}(E\mathbf{Y} - \tau)]$$

which in turn allows the expected MD to be expressed as

$$E(MD) \quad = \quad r + [\mathbf{h^T}(E\mathbf{Y} - \tau)]^T [\mathbf{h^T}(E\mathbf{Y} - \tau)].$$

Since r is a constant the choice of the vector of design variable values defined in Chapter 3, i.e., $\mathbf{x}$, which minimizes the expected MD, is that which minimizes the distance between the weighted responses and their targets. Weighting the responses in the MD is simply multiplying them by $\mathbf{h} = \Sigma_{\mathbf{Y}}^{-\frac{1}{2}}$. This expected MD is equivalent to Hotelling's T^2 for a single multivariate observation where the individual variables represent point estimates of the values of the $E\mathbf{Y}$ vector.

4.2.3 Case 3: PCs Derived from the Covariance Matrix of Loss Scaled Responses, i.e., $\mathbf{q}^T\Sigma_{\mathbf{Yq}}$, When $\mathbf{A} \neq \mathbf{I}_r$ and $\Sigma_{\mathbf{Y}} = \mathbf{I}_r$

In this case the expected MQL is

$$\begin{aligned} E(MQL) &= E(\mathbf{Y} - \tau)^T \mathbf{A} (\mathbf{Y} - \tau) \\ &= trace[\Sigma_{\mathbf{Y}} \mathbf{A}] + (E\mathbf{Y} - \tau)^T \mathbf{A} (E\mathbf{Y} - \tau) \\ &= trace[\mathbf{I}_r \mathbf{A}] + (E\mathbf{Y} - \tau)^T \mathbf{A} (E\mathbf{Y} - \tau) \\ &= trace[\mathbf{A}] + (E\mathbf{Y} - \tau)^T \mathbf{A} (E\mathbf{Y} - \tau) \end{aligned} \tag{4.-55}$$

It is simple to show that the trace and off-target vector product component from Equation (4.-55) are equal to their counterparts from the complete multi-response data-set by repeating the steps presented in Case 1. Here the covariance matrix of the r loss-scaled responses is numerically equal to the loss constant matrix $\mathbf{A}$ and the principal components of the loss-scaled responses are defined as

$$\mathbf{ZQ} \quad = \quad \mathbf{W}^T(\mathbf{q}^T\mathbf{Y})$$

where the columns of $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_r)$ are orthogonal and of unit length.

We plan to use only the first p principal components of the loss-scaled responses to reduce the dimensionality of the optimization problem. We define the matrix **ZQp** as the matrix product of the first p columns of **W** i.e. **Wp**, with the loss-scaled response vector $\mathbf{q}^T\mathbf{Y}$, which implies

$$trace[\Sigma_{\mathbf{ZQp}}] \quad = \Sigma_{i=1}^{p}\lambda_i,$$

where the λ_i represent the characteristic values (i.e., eigenvalues) of $\Sigma_{\mathbf{ZQ}}$. The sum of the first p eigenvalues accounts for a fixed proportion of the total sample variance equal to

$$\frac{\Sigma_{i=1}^{p}\lambda_i}{\Sigma_{i=1}^{r}\lambda_i} \quad .$$

Hence the $trace[\Sigma_{\mathbf{ZQp}}]$ component of the expected value of the MQL represents this same fixed proportion of the total sample variance of the loss-scaled responses. To examine how well the $OTVP$ is approximated by a subset of the PCs we first diagonalize $\mathbf{A}$,

$$\mathbf{A} \quad = \quad \mathbf{W}\mathbf{\Lambda_A}\mathbf{W^T}$$

where the columns of $\mathbf{W}$ are the ortho-normal eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda_A}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{A}$. For notational simplicity we let $\mathbf{O} = (E\mathbf{Y} - \tau)$. Next we write out the complete $OTVP$ as

$$\begin{aligned}
\mathbf{O^T A O} &= \mathbf{O^T W \Lambda_A W^T O} \\
&= \left[\mathbf{O^T W_1}, \ldots, \mathbf{O^T W_r}\right] \Lambda_{\mathbf{A}} \left[\mathbf{W_1^T O}, \ldots, \mathbf{W_r^T O}\right]^T \\
&= \Sigma_{i=1}^{r}\lambda_i[\mathbf{O^T W_i}]^2. \qquad (4.\text{-}60)
\end{aligned}$$

Therefore the $OTVP$ approximated by the first p columns of $\mathbf{W}$ is equal to $\Sigma_{i=1}^{p}\lambda_i[\mathbf{O^T W_i}]^2$. The scalar value of $OTVP$ is the sum total of the individual $OTVP$-contributions from the full set of LSPC, where $(c_1^2, \ldots, c_r^2)$ are the real, constant values contributed by the respective LSPC. The constants $(c_1^2, \ldots, c_p^2)$ are the scalar contributions from the subset of $p < r$ LSPC being utilized. The numerical value of equation 4.-60 depends on the specific

off-target vector i.e., $(E\mathbf{Y} - \tau)$ being evaluated and is bounded as follows,

$$0 \quad \leq \frac{c_1^2 + \cdots + c_p^2}{c_1^2 + \cdots + c_r^2} \leq 1.$$

The proof of this property is the same as that of 4.-31 presented in Appendix A. How well the first p terms in the numerator account for the entire quadratic form of the $OTVP$ depends on the specific structures of $\mathbf{A}$ and $\mathbf{\Sigma_Y}$.

In Appendix B we did not examine any data-sets for this case since the identity covariance matrix and its implied independence of responses would negate the relevance and utility of PCs.

4.2.4 Case 4: PCs Derived from the Correlation Matrix of Loss-Scaled Responses i.e., $\mathbf{q}^T\mathbf{R_{Yq}}$, When $\mathbf{A} = \mathbf{I}_r$ and $\mathbf{R_Y} \neq \mathbf{I}_r$

In section 1.2.2.3 of Chapter 1, we described the two special situations when it is preferable to derive PCs from the data's correlation matrix versus its covariance matrix. These are when the responses of interest are measured in different units or have widely ranging variances which would cause the first PC or two to account for nearly all variation in data as a linear combination of the higher-variance variable(s). During product design with multiple responses, these are very likely conditions .

Case 4 differs from Case 1 in that we employ the correlation matrix of $\mathbf{Y}$ rather than the covariance matrix. Just as in Case 1, the loss constant matrix $\mathbf{A}$ here is the identity matrix $\mathbf{I}_r$. First we express each individual response (y_i) observed in standard units, i.e., $y_i^* = \frac{y_i - Ey}{\sigma_{y_i}}$, and define the matrix $\mathbf{D}$ as a diagonal matrix of the square roots of the individual response variances.

We define the term MQL^* as the MQL where the response vector has been replaced by its standardized counterpart. This notation differentiates the MQL^* from the MQL , which is defined as a function of $\mathbf{Y}$ and is not invariant to transformations. Defining $\mathbf{R}$

as the original response correlation matrix, the expected value of MQL^* is defined as:

$$\begin{aligned} E(MQL^*) &= E[\mathbf{D}^{-1}(\mathbf{Y}-\mu_{\mathbf{Y}}-(\tau-\mu_{\mathbf{Y}}))]^T[\mathbf{D}^{-1}(\mathbf{Y}-\mu_{\mathbf{Y}}-(\tau-\mu_{\mathbf{Y}}))] \\ &= E[\mathbf{D}^{-1}(\mathbf{Y}-\tau)]^T[\mathbf{D}^{-1}(\mathbf{Y}-\tau)] \\ &= trace[\Sigma_{\mathbf{D}}^{-1}\mathbf{Y}]+[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]^T[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)] \\ &= trace[\mathbf{D}^{-1}\Sigma_{\mathbf{Y}}\mathbf{D}^{-1}]+[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]^T[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)] \\ &= trace[\mathbf{R}]+[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]^T[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)] \\ &= r+[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]^T[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)] \end{aligned} \tag{4.-66}$$

So if each response data point is standardized the expected value of the MQL^* results in the sum of r , the number of responses, and an $OTVP$ which is standardized by its multiplication with $\mathbf{D^{-1}}$. Since r is a constant the choice of $\mathbf{x}$ which minimizes the expected MQL^* is that which minimizes the distance between the standardized mean responses and their targets.

The correlation matrix $\mathbf{R}$ is simply the covariance matrix of the standardized response values. Hence the PCs derived from the correlation matrix are the product of the eigenvectors of the response correlation matrix $\mathbf{R}$ by the standardized response values. We define the r PCs derived from the correlation matrix as

$$\mathbf{Z_R} = \mathbf{K}^T\mathbf{Y}^*$$

where $\mathbf{Y}^* = \mathbf{D}^{-1}(\mathbf{Y}-E\mathbf{Y})$.

Plugging all r of the PCs derived from the correlation matrix into the expected value of the MQL^* and defining $\tau^* = \mathbf{D}^{-1}(\tau - E\mathbf{Y})$ yields

$$\begin{aligned} E(\mathbf{Z_R}-\tau_{\mathbf{ZR}})^T(\mathbf{Z_R}-\tau_{\mathbf{ZR}}) &= E(\mathbf{K}^T\mathbf{Y}^*-\mathbf{K}^T\tau^*)^T(\mathbf{K}^T\mathbf{Y}^*-\mathbf{K}^T\tau^*) \\ &= E[\mathbf{K^T}(Y^*-\tau^*)]^T[\mathbf{K^T}(Y^*-\tau^*)] \\ &= E[(\mathbf{Y}^*-\tau^*)]^T\mathbf{K}\mathbf{K^T}[(\mathbf{Y}^*-\tau^*)] \\ &= E[(\mathbf{Y}^*-\tau^*)]^T[(\mathbf{Y}^*-\tau^*)] \\ &= E[\mathbf{D}^{-1}(\mathbf{Y}-\tau)]^T[\mathbf{D}^{-1}(\mathbf{Y}-\tau)] \end{aligned} \tag{4.-70}$$

which we showed in Equation (4.-66) to equal the sum of a constant and the standardized $OTVP$. This shows that the set of all r PCs from the correlation matrix replicates exactly the expected value of the MQL^* for the standardized responses. It is interesting to note that this off-target vector product is equivalent to the Mahalanobis Distance when the responses are independent.

Once again our purpose is to reduce the dimensionality of optimization by examining the first p PCs. Defining the vector $\mathbf{Z_{Rp}}$ as the matrix product of the first p columns of $\mathbf{K}$, i.e. $\mathbf{Kp}$, by $\mathbf{Y}^*$ implies that

$$trace[\Sigma_{\mathbf{ZRp}}] \quad = \Sigma_{i=1}^{p}\lambda_i$$

where the λ_i represent the eigenvalues of $\Sigma_{\mathbf{ZR}}$ in order of descending magnitude. The sum of the first p eigenvalues accounts for a fixed proportion of the trace of the sample correlation matrix equal to

$$\frac{\Sigma_{i=1}^{p}\lambda_i}{\Sigma_{i=1}^{r}\lambda_i} \quad = \quad \frac{\Sigma_{i=1}^{p}\lambda_i}{r} \quad .$$

Since the trace of $\mathbf{R}$ is fixed, using p PCs derived from the correlation matrix to minimize expected MQL^* simplifies to minimizing the standardized $OTVP$. We re-express the standardized $OTVP$ as

$$\begin{aligned}(E\mathbf{Z_{Rp}} - \tau_{\mathbf{ZRp}})^T(E\mathbf{Z_{Rp}} - \tau_{\mathbf{ZRp}}) &= [\mathbf{Kp}^T\mathbf{D}^{-1}(E\mathbf{Y} - \tau)]^T[\mathbf{Kp}^T\mathbf{D}^{-1}(E\mathbf{Y} - \tau)] \\ &= [\mathbf{D}^{-1}(E\mathbf{Y} - \tau)]^T\mathbf{KpKp}^T[\mathbf{D}^{-1}(E\mathbf{Y} - \tau)].\end{aligned} \tag{4.-73}$$

Because $\mathbf{Kp}$ is only a subset of the original r columns of $\mathbf{K}$, the matrix product $\mathbf{KpKp}^T$ does not simplify to the identity matrix as the entire set of orthogonal columns does in $\mathbf{KK}^T$.

However we can still estimate the fraction of the total $OTVP$ that is accounted for by this subset of p PCs. We construct a fraction whose numerator is the $OTVP$ from

Equation (4.-73) and whose denominator is the $OTVP$ from the full-set of r PCs.

$$\begin{aligned}\frac{(E\mathbf{Z_{Rp}}-\tau_{\mathbf{ZRp}})^T(E\mathbf{Z_{Rp}}-\tau_{\mathbf{ZRp}})}{(E\mathbf{Z_R}-\tau_{\mathbf{ZR}})^T(E\mathbf{Z_R}-\tau_{\mathbf{ZR}})} &= \frac{[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]^T\mathbf{KpKp}^T[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]}{[\mathbf{D^{-1}}(E\mathbf{Y}-\tau)]^T\mathbf{KK}^T[\mathbf{D^{-1}}(E\mathbf{Y}-\tau)]} \\ &= \frac{[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]^T\mathbf{KpKp}^T[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]}{[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]^T[\mathbf{D}^{-1}(E\mathbf{Y}-\tau)]}\end{aligned} \tag{4.-74}$$

Just like the first three cases, the numerical value of equation 4.-74 depends on the specific off-target vector i.e., $(E\mathbf{Y}-\tau)$ being evaluated and is bounded as follows,

$$0 \leq \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2} \leq 1.$$

The proof of this property is the same as that of 4.-31 presented in Appendix A. How well the first p terms in the numerator account for the entire quadratic form of the $OTVP$ depends on the specific structures of $\mathbf{A}$ and $\Sigma_{\mathbf{Y}}$.

The simulations in Appendix B corresponding to Case 4, i.e., where $\mathbf{A}=\mathbf{I}_r$, suggest that using a subset of eigenvectors to approximate the $OTVP$ often results in significant loss of information.

4.2.5 Case 5: PCs Derived from the Correlation Matrix of Loss-Scaled Responses i.e. $\mathbf{q^T R_Y q}$, When $\mathbf{A}\neq\mathbf{I}_r$ and $\Sigma_{\mathbf{Y}}\neq\mathbf{I}_r$

This case differs from Case 2 only in that we employ the correlation matrix of the loss-scaled responses rather than their covariance matrix. The loss constant matrix $\mathbf{A}$ is positive definite and the expected value of the MQL^* is

$$\begin{aligned}E(MQL^*) &= E(\mathbf{Y}^*-\tau^*)^T\mathbf{A}(\mathbf{Y}^*-\tau^*) \\ &= trace[\mathbf{A}\Sigma_{\mathbf{Y}}^*]+(E\mathbf{Y}^*-\tau^*)^T\mathbf{A}(E\mathbf{Y}^*-\tau^*) \\ &= trace[\mathbf{AR}]+(E\mathbf{Y}^*-\tau^*)^T\mathbf{A}(E\mathbf{Y}^*-\tau^*).\end{aligned} \tag{4.-77}$$

In Case 2 we showed that $\mathbf{A}$ could be re-expressed by replacing the $\mathbf{Y}$ vector with the loss-scaled vector $\mathbf{q^TY}$ and in Case 4 we showed that $(\mathbf{Y}^*-\tau^*)$ can be re-written as $\mathbf{D}^{-1}(\mathbf{Y}-\tau)$.

Substituting these expressions in the expected MQL^* of Equation (4.-77) yields

$$\begin{aligned} E(MQL^*) &= trace[\mathbf{AR}] + (E\mathbf{Y}^* - \tau^*)^T\mathbf{A}(E\mathbf{Y}^* - \tau^*) \\ &= trace[\mathbf{q}^T\mathbf{Rq}] + [\mathbf{q}^T(E\mathbf{Y}^* - \tau^*)]^T[\mathbf{q}^T(E\mathbf{Y}^* - \tau^*)] \\ &= trace[\mathbf{q}^T\mathbf{Rq}] + [\mathbf{q}^T\mathbf{D}^{-1}(E\mathbf{Y} - \tau)]^T[\mathbf{q}^T\mathbf{D}^{-1}(E\mathbf{Y} - \tau)] \\ &= trace[\mathbf{q}^T\mathbf{D}^{-1}\Sigma_{\mathbf{Y}}\mathbf{D}^{-1}\mathbf{q}] + [\mathbf{q}^T\mathbf{D}^{-1}(E\mathbf{Y} - \tau)]^T[\mathbf{q}^T\mathbf{D}^{-1}(E\mathbf{Y} - \tau)]. \end{aligned} \tag{4.-80}$$

This new expression of expected MQL^* consists of the trace of the covariance matrix and the off-target vector product of $\mathbf{q}^T\mathbf{D}^{-1}\mathbf{Y}$, i.e., the loss-scaled standardized response vector. This motivates us to take the PCs of the loss-scaled standardized responses. We define the r PCs of the loss-scaled standardized responses as

$$\mathbf{ZQD} = \mathbf{L}^T(\mathbf{q}^T\mathbf{D}^{-1}\mathbf{Y})$$

where the columns of $\mathbf{L} = (\mathbf{L_1}, \ldots, \mathbf{L_r})$ are orthogonal and of unit length and whose diagonal covariance matrix is

$$\begin{aligned} \Sigma_{\mathbf{ZQD}} &= \mathbf{L}^T\Sigma_{(\mathbf{q}^T\mathbf{D}^{-1}\mathbf{Y})}\mathbf{L} \\ &= \mathbf{L}^T(\mathbf{q}^T\mathbf{D}^{-1}\Sigma_{\mathbf{Y}}\mathbf{D}^{-1}\mathbf{q})\mathbf{L}. \end{aligned}$$

Substituting $\mathbf{ZQD}$ in place of $\mathbf{q}^T\mathbf{D}^{-1}\mathbf{Y}$ in the expected MQL^* of Equation (4.-80) yields

$$E(\mathbf{ZQD} - \tau_{ZQD})^T(\mathbf{ZQD} - \tau_{ZQD}) = trace[\Sigma_{\mathbf{ZQD}}] + (E\mathbf{ZQD} - \tau_{\mathbf{ZQD}})^T(E\mathbf{ZQD} - \tau_{\mathbf{ZQD}}). \tag{4.-84}$$

It is simple to show that the trace and off-target vector product component from Equation (4.-84) are equal to their counterparts from the complete multi-response data-set by repeating the steps presented in Case 1.

We plan to use only the first p principal components of the loss-scaled standardized responses to reduce the dimensionality of the optimization problem. Next we examine how using this lower dimensionality data-set affects the expected value of the MQL^*. We define

$\mathbf{ZQDp}$ as the matrix product of the first p columns of $\mathbf{L}$ i.e. $\mathbf{Lp}$, by the loss-scaled standardized response matrix $\mathbf{q}^T\mathbf{D}^{-1}\mathbf{Y}$, which implies

$$trace[\Sigma_{\mathbf{ZQDp}}] \quad = \quad \Sigma_{i=1}^{p}\lambda_i$$

where the λ_i represent the eigenvalues of $\Sigma_{\mathbf{ZQD}}$.

The sum of the first p eigenvalues accounts for a fixed proportion of the total loss-scaled, standardized sample variance equal to

$$\frac{\Sigma_{i=1}^{p}\lambda_i}{\Sigma_{i=1}^{r}\lambda_i} \quad .$$

Since the trace of $\mathbf{AR_Y}$ is fixed, using p PCs derived from the correlation matrix to optimize expected MQL^* simplifies to minimizing the standardized off-target vector component. We now re-express the $OTVP$ as

$$\begin{aligned}(E\mathbf{Z_{Rp}} - \tau_{\mathbf{ZRp}})^T(E\mathbf{Z_{Rp}} - \tau_{\mathbf{ZRp}}) &= [\mathbf{Lp^T q^T D^{-1}}(E\mathbf{Y} - \tau)]^T[\mathbf{Lp^T q^T D^{-1}}(E\mathbf{Y} - \tau)] \\ &= [\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]^T\mathbf{LpLp^T}[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)].\end{aligned} \tag{4.-87}$$

Because $\mathbf{Lp}$ is only a subset of the original r columns of $\mathbf{L}$, the matrix product $\mathbf{LpLp}^T$ does not simplify to the identity matrix as the entire set of orthogonal columns does in $\mathbf{LL^T}$.

However we can still estimate the fraction of the total $OTVP$ that is accounted for by this subset of p PCs. We construct a fraction whose numerator is Equation (4.-87) and whose denominator is the $OTVP$ from the full-set of r PCs.

$$\begin{aligned}\frac{(E\mathbf{Z_{Rp}} - \tau_{\mathbf{ZRp}})^T(E\mathbf{Z_{Rp}} - \tau_{\mathbf{ZRp}})}{(E\mathbf{Z_R} - \tau_{\mathbf{ZR}})^T(E\mathbf{Z_R} - \tau_{\mathbf{ZR}})} &= \frac{[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]^T\mathbf{LpLp}^T[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]}{[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]^T\mathbf{LL}^T[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]} \\ &= \frac{[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]^T\mathbf{LpLp}^T[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]}{[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]^T[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]}\end{aligned} \tag{4.-88}$$

The numerical value of this fraction depends on the specific

Like the other cases, the numerical value of equation 4.-74 depends on the specific off-target vector i.e., $(E\mathbf{Y} - \tau)$ being evaluated and is equally bounded.

Table 4.1: LSPC From Covariance Matrices of Data

MATRIX CASES ⇓	Fraction of trace[$\Sigma\mathbf{A}$] Explained by p PCs	Expression for $OTVP$	Fraction of $OTVP$ Approx. by 2 PCs from Simulations in Appendix B
Case 1 $\mathbf{A} = \mathbf{I}_r$ $\mathbf{\Sigma_Y} \neq \mathbf{I}_r$	$\text{trace}[\Sigma_{\mathbf{Zp}}] = \frac{\Sigma_{i=1}^{p}\lambda_i}{\Sigma_{i=1}^{r}\lambda_i}$	$(E\mathbf{Y}-\tau)^T(E\mathbf{Y}-\tau)$	$0.51 \leq \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2} \leq 0.80$
Case 2 $\mathbf{A} = \mathbf{qq^T}$ $\mathbf{\Sigma_Y} \neq \mathbf{I}_r$	$\text{trace}[\Sigma_{\mathbf{ZQp}}] = \frac{\Sigma_{i=1}^{p}\lambda_i}{\Sigma_{i=1}^{r}\lambda_i}$	$[\mathbf{q}^T(E\mathbf{Y}-\tau)]^T[\mathbf{q}^T(E\mathbf{Y}-\tau)]$	$0.73 \leq \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2} \leq 0.97$
Case 3 $\mathbf{A} = \mathbf{qq^T}$ $\mathbf{\Sigma_Y} = \mathbf{I_r}$	$\text{trace}[\mathbf{Ap}] = \frac{\Sigma_{i=1}^{p}\lambda_i}{\Sigma_{i=1}^{r}\lambda_i}$	$[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]$	PCs Not Useful In This Case

The simulations in Appendix B suggest that using a subset of eigenvectors to approximate the $OTVP$ for situations related to Case 5, i.e., when neither $\mathbf{A}$ nor $\mathbf{R_Y}$ are diagonal, results in no significant loss of information.

There is no need to examine the case of using the correlation matrix when $\mathbf{A} \neq \mathbf{I}_r$ and $\Sigma_{\mathbf{Y}} = \mathbf{I}_r$ because when $\Sigma_{\mathbf{Y}=\mathbf{I_r}}$, then $\Sigma_{\mathbf{Y}} = \mathbf{I}_r = \mathbf{R_Y}$ which results in Case 3.

4.2.6 Summary of the Five Cases

Table 4.1 summarizes the cases for PCs derived from the covariance matrices and Table 4.2 summarizes the cases for PCs derived from correlation matrices.

Table 4.2: LSPC From Correlation Matrices of Data

MATRIX CASES ⇓	Fraction of trace[$\Sigma\mathbf{RA}$] Explained by p PCs	Expression for $OTVP$	Fraction of $OTVP$ Approx. by 2 PCs from Simulations in Appendix B
Case 4 $\mathbf{A} = \mathbf{I}_r$ $\mathbf{R_Y} \neq \mathbf{I}_r$	trace[$\Sigma_{\mathbf{Z_{Rp}}}$] $= \frac{p}{r}$	$[\mathbf{D^{-1}}(E\mathbf{Y} - \tau)]^T[\mathbf{D^{-1}}(E\mathbf{Y} - \tau)]$	$0.21 \leq \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2} \leq 0.73$
Case 5 $\mathbf{A} = \mathbf{qq^T}$ $\mathbf{R_Y} \neq \mathbf{I}_r$	trace[$\Sigma_{\mathbf{ZQDp}}$] $= \frac{\Sigma_{i=1}^{p}\lambda_i}{r}$	$[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]^T[\mathbf{q^T D^{-1}}(E\mathbf{Y} - \tau)]$	$0.71 \leq \frac{c_1^2+\cdots+c_p^2}{c_1^2+\cdots+c_r^2} \leq 0.96$

4.3 Conclusions

This chapter has defined and laid out some of the properties of the full set and subsets of LSPC. The following points are most important:

1. The full set of LSPC fully replicates expected MQL.

2. For subsets of LSPC, we can quantify exactly how much of the $trace[A\Sigma_{\mathbf{Y}}]$ is accounted for.

3. For subsets of LSPC, we can't quantify exactly how much of the $OTVP$ will be accounted for without knowing the specific off-target vector, since this quantity is derived completely from that vector.

4. In Appendix B we have quantified how much a subset of 2 LSPC accounts for of the $OTVP$ associated with twenty different off-target vectors selected for specific conditions over five data-sets ranging in dimension from 3 to 6.

5. For the cases with non-diagonal loss constant matrix and non-diagonal covariance/correlation

matrix, the subset of 2 LSPC accounts for between 71 and 97 percent of the $OTVP$.

CHAPTER V

LSPC IN STATIC MULTIVARIATE ROBUST DESIGN

In Chapter 4 we showed how well the LSPC approximate the expected MQL of the original multiple response data set at lower dimensionality. In this chapter we propose a technique for static multivariate RD using meta-models of the LSPC with deterministic, internal noise generated around the nominal values of each design vector chosen by the optimization algorithm.

In Section 1 we describe a product design problem with 8 predictor variables and 6 responses which inspired the concept of LSPC. We provide a high level description of how we solve this optimization problem using "meta-models" of the standardized responses. In Section 2 we describe a LSPC meta-model based procedure for static multivariate RD. In Section 3 we apply our LSPC-based procedure to the product design problem presented in Section 1. We compare the optimal design vector identified with models of the LSPC to that chosen by models of the standardized responses. We compare solutions from both sets of meta-models to that chosen by a highly accurate mechanical design simulator called *IDEAS*, which we take as the truly optimal response. We also compare the computational time needed by the competing meta-model methods and the sophisticated mechanical design simulator which we take as our standard solution. In Section 4 we make concluding remarks on the analytical and practical characteristics of the proposed approach.

5.1 Introduction to Meta-Modelling and Our Application

We start this chapter with a short history of how some of the ideas in this thesis evolved. In Fall '01 we started working with the Systems Realization Laboratory (i.e., SRL) of the G.W. Woodruff School of Mechanical Engineering at the Georgia Institute of Technology. The SRL creates technology and procedures for designing mechanical and thermal systems and collaborates on design projects in automation and aerospace engineering. Our goal

was to find engineering applications where RD concepts and ideas might be tested and implemented.

The SRL commonly simulates mechanical designs with sophisticated and highly accurate deterministic simulation models based on finite element analysis and geometric analysis. These simulation models tend to be computationally very demanding and include *Pro-Engineer*, *IDEAS* and *ABSYS* which are available to all GT students. They follow sequential design processes where the information gained influences the subsequent design stages.

In the early stages of a design, the designer is not expecting a highly accurate solution. She or he wants to reduce the broader design space to a smaller, localized design region where experimentation indicates the best responses may reside. During these early stages they commonly build "meta-models" using the accurate, deterministic simulation packages such as *IDEAS*. The huge savings in computational time afforded by the meta-models more than compensates for their loss of accuracy during the early stages of a design time-line. The SRL has experience using meta-models based on response surface methodology (RSM), Kriging and multivariate adaptive regression splines (MARS) to approximate the "true" responses, i.e., the output of the deterministic simulation package.

Because of the deterministic nature of the simulation output, meta-models stray from a key underlying statistical assumption for linear models, i.e., the response being a random variable with inherently constant variance. But since these meta-models are being used to narrow down a large design space, rather than for inferential purposes, it is a very practical and fruitful use of the model-fitting techniques. The variance in the responses is entirely a function of the variability in the assumed fixed design variable values rather than that of the response itself.

Though meta-models built around these deterministic processes are not valid for statistical inference, they are often of sufficient accuracy in early design stages. Later design stages rely on sophisticated modelling programs to verify mature designs, which removes the risks incurred by use of meta-models earlier in the process.

To bring this into perspective we discuss a specific design problem the SRL has solved

Figure 5.1: Electric Vehicle Chassis from Ortega (1998)

in the past. They have built finite-element and geometric simulation models using *IDEAS* for six characteristics of the chassis of an electric vehicle (EV). The following diagrams and description are borrowed from Lin (2001) who built upon the work of Ortega (1998). In Figure 5.1 we show a digital image of the electric vehicle chassis being designed.

The six characteristics to be designed are functions of the same eight design variables. Each of the design variables is the thickness of a structural element of the vehicle chassis. The design variables and their ranges are described in Table 5.1. Figure 5.2 shows the location of several of the design factors within the context of the vehicle chassis. The numbered items correspond to the design variables described in Table 5.1. The cutting and pasting of this figure prevented inclusion of all the design factors.

The six responses are highly correlated and are related to safety, strength and weight requirements of the final product. We will describe the responses below with some of their functional context in the following paragraphs.

1. BEND is the mid-span bending of the vehicle refers to the amount of deformation in the middle of the chassis in the event of a collision. This is a safety concern since less deformation translates into higher safety for vehicle passengers.

Table 5.1: Design Variables and Ranges for Chassis Design from Ortega (1998)

Design Variable	**Part Name**	**Geometry**	**Minimum (mm)**	**Maximum (mm)**
$x1$	Lower Chassis Structure	Rect. Tube w70 x h100 x $x1$	1.0	8.0
$x2$	Rear End Long. Frame	Rect. Tube w70 x h80 x $x2$	1.0	8.0
$x3$	Upper Front End Structure	Rect. Tube w50 x h70 x $x3$	1.0	8.0
$x4$	Lower Front End Support Structure	Rect. Tube w50 x h50 x $x4$	1.0	8.0
$x5$	Floor Structure and Lower B-pillar	Hat w120 x h70 x b40 x $x5$	1.0	8.0
$x6$	Upper Front B-pillar	Hat w120 x h20 x b40 x $x6$	1.0	8.0
$x7$	Upper Rear B-pillar	Hat w60 x w20 x h40 x $x7$	1.0	8.0
$x8$	Floor Panels	Thin Shell $x8$	1.0	8.0

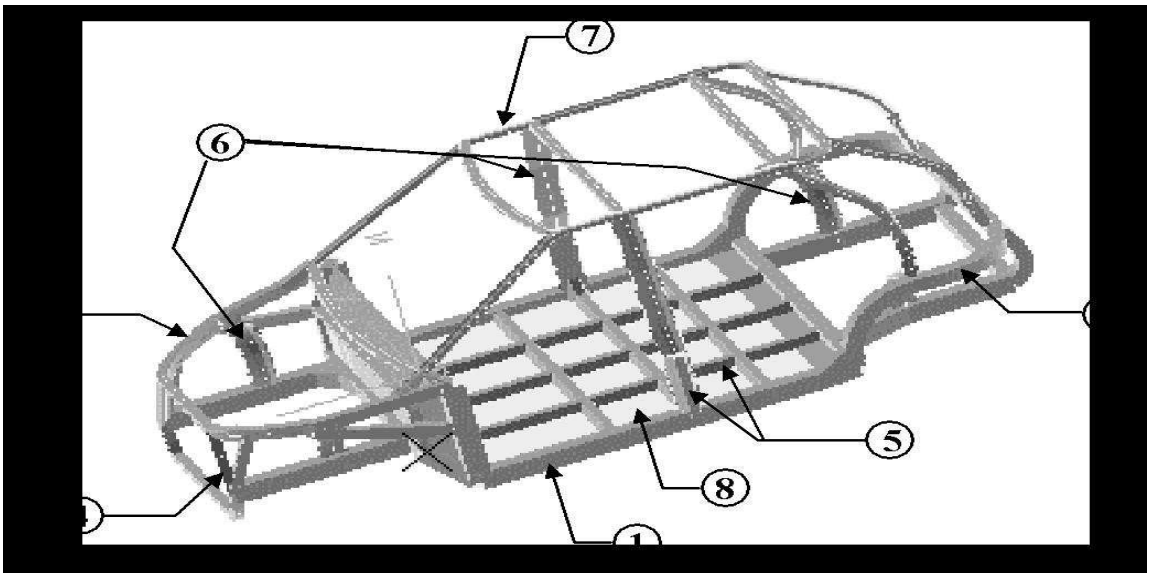

Figure 5.2: Design Factor Locations on Chassis from Ortega (1998)

2. CURB represents the magnitude of the force exerted on the vehicle's front wheels when the vehicle hits a bump or curb straight on with the brakes applied. This force has components directed vertically upward and horizontally in the rear direction. The designer wants to minimize both components of this force so as to provide the driver with as much vehicle control as possible during this sort of collision.

3. FRONT is the amount of deformation to the front of the chassis due to frontal impact during a head on collision. This deformation is to be minimized to provide the passenger with the largest "crush-free" zone possible.

4. MASS is the mass of the chassis and for purposes of fuel efficiency and vehicle longevity is to be minimized. While an important consideration to the consumer, it is typically of lower economic priority than responses directly related to passenger safety.

5. ROLL is related to the vehicle's deformation during vehicle roll-over and is to be minimized to afford the passenger greater protection.

6. TOR is the vehicle's torsional stiffness which is related to strength of the structural elements and is to be maximized.

Table 5.2: Electric Vehicle Chassis Response Characteristics

Responses ⇓	**Taguchi Type**	**Product Significance**	**Range**	**Target**
BEND	NTB	Safety	$0.00054 - 0.00125$	0.00058
CURB	NTB	Safety	$189M - 546M$	$192M$
FRONT	NTB	Safety	$125M - 406M$	$130M$
MASS	NTB	Gas Mileage	$235 - 656$	240
ROLL	NTB	Safety	$18M - 61M$	$20M$
TOR	NTB	Vehicle Strength	$0.00035 - 0.00129$	0.00125

The ranges and targets of the responses are characterized in Table 5.2. The first five responses are smaller-the-better (STB) and TOR is larger-the-better (LTB) per the Taguchi terminology defined in Chapter 2. Our original space-filling design had specific upper and lower bounds within which the meta-models for the responses were fit. To avoid extrapolation and the strange behavior often found at the margins of the data-space from which models are fit, we converted each response into a nominal-the-best (NTB) type response whose target is near the relevant extreme of the modelling space. This was intuitively more reasonable from both engineering and meta-modelling perspectives.

With *IDEAS*, calculating a vector of the six response values from a design vector of size 8 takes approximately 2 minutes. Solving for the design vector which optimizes some objective function of the six responses when starting from the center of the design space usually takes several hundred cycles. This equates to more than a day of computational

Table 5.3: Meta-Models for Approximating *IDEAS*

Model Type ⇓	**Nature of Model** ⇓	**Our Choices While Model-Fitting** ⇓	**Model Code or Software** ⇓
Response Surface Methodology	Linear Regression	Quadratic Model With No Interactions	Minitab and Matlab
Kriging	Multivariate Interpolation Model	Assumes Gaussian Process	Fortran Code by Simpson '96
MARS	Piecewise Polynomial	Limited to 2 Factor Interactions	C code by Chen '01

time. Adding an "external noise" matrix designed around the values of the design vector multiplies the time needed by *IDEAS* even more. A meta-model of each response computes each value in a few seconds allowing the same optimization problem to be solved in a few hours. In early design stages this rapid identification of a promising design region provides time and flexibility to other design stake-holders to react to the evolving design region.

Since the MQL is the most widely discussed objective function of RD in the statistics literature, we started thinking of ways to improve the use of meta-models with MQL. Our thinking was also influenced by the model methodology successfully used by the Systems Realization Laboratory. They have demonstrated successful results with RSM, Kriging and MARS. Table 5.3 contrasts how we employed these three model fitting approaches for this design problem.

RSM models use ordinary least squares which minimize the sum of squared error between the true responses and their predicted values. Kriging models assume a particular covariance structure of the multivariate data, typically normal, and form an interpolating model that

correctly predicts each of the points in the training data-set used to build them. MARS is fundamentally an extension of ordinary least squares which builds the piecewise polynomial that minimizes squared error or some other criterion.

Figure 5.3 depicts a high level view of how we implemented the solving of the robust design (i.e., optimization) problem with response meta-models.

Several details in Figure 5.3 need further documentation and explanation.

1. Why did we use the the optimization platform iSIGHT?

2. How does this deterministic optimization problem qualify as a RD problem?

We used the optimization platform iSIGHT because the SRL has considerable experience and trust in this product which has a full range of optimization algorithms as well as the ability to create experimental designs and build meta-models with RSM and Kriging. SRL also provided us with training and guidance. iSIGHT is made by Engineous Software Inc. (www.engineous.com) and interfaces well with custom made and commercially available simulation codes such as *IDEAS*. One of its most attractive features is its optimization advisor which surveys the parameters of the design problem and recommends the best optimization algorithms to select. Appendix C is a brief introduction to iSIGHT which is taken directly from the iSIGHT Basic Training Guide (2002).

RD concerns random variables possessing the $trace[A\Sigma_{\hat{\mathbf{Y}}(\mathbf{x})}]$ term of MQL calculated from the "noise replicates" associated with each particular design vector selected by the optimization algorithm. So far we've described only a purely deterministic optimization problem. This is a type II RD problem per the terminology of Chen et al (1996) where the response variance induced by each fixed design vector is artificially created by varying each design setting between 0.95 and 1.05 of the nominal values. These values are centered on the nominal design values and are meant to represent typical manufacturing tolerances for mechanical components.

We chose a 4-level Uniform Design constructed with 32 runs representing internal noise within each of the eight design variables. This uniform design was run for each design vector $\mathbf{x}$ chosen by the optimization algorithm in iSIGHT. The trace term was calculated

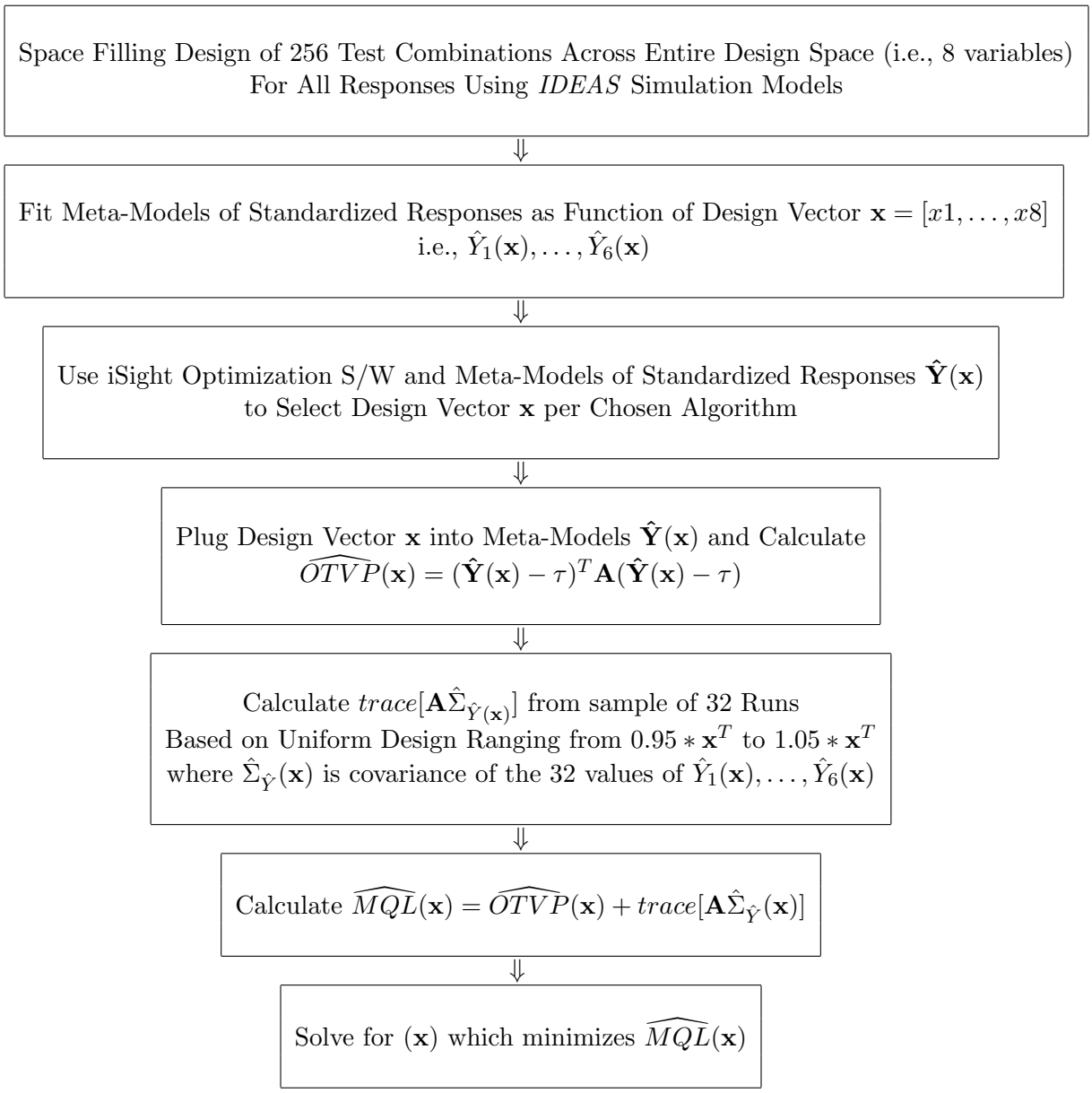

Figure 5.3: Using Response Meta-Models in $\widehat{MQL}$ to Solve for Design Vector $\mathbf{x}$

from the covariance matrix of the 32 resulting "internal noise" response vectors executed for each design vector being considered. The trace term was then added to the $\widehat{OTVP}$ to yield $\widehat{MQL}$ for each particular design vector.

We solved the design problem by minimizing the $\widehat{MQL}$ calculated with the meta-models of the standardized responses formed by the three different modelling approaches. Table 5.4 compares the optimal design vectors and corresponding $\widehat{MQL}$ values chosen by the three sets of response meta-models with those of the "true" optimal response calculated by *IDEAS*. The design values chosen by *IDEAS* and the modelling approach with lowest $\widehat{MQL}$ are in bold face print.

It is interesting to note that the more sophisticated modelling approaches, namely Kriging and MARS, selected the closest nominal design value to that of IDEAS in all cases. RSM tied MARS in selecting the best approximation for $x8$. This is not surprising given the RSM model was limited to a second order form with no interactions.

5.2 A Static MRD Procedure Using LSPC

After studying the way meta-models of the standardized responses are employed in Figure 5.3, we wondered whether there were multivariate techniques that might make this process more efficient. We speculated that instead of building meta-models of highly correlated individual responses, meta-models of PCs of the responses might allow a reduction of dimensionality of the problem leading to faster solution of the optimization problem. The models constructed for the PCs are assumed to provide a consistently good approximation throughout the design space. The same assumption applies to the meta-models representing the true response over the entire design space being considered.

The justification for direct modelling of PCs comes from Pearson (2001), who proved the following with regard to an r dimensional set of data. The hyper-plane defined by the first $p < r$ eigenvectors is that p dimensional surface, which out of all possible p dimensional surfaces, minimizes the sum of squared orthogonal distances between itself and the data-set.

Figure 5.4 shows how a 2 dimensional plane approximates a three dimensional scatterplot by minimizing the sum of squared orthogonal distances between individual points and

Table 5.4: Design Vectors and Computational Time from Response Meta-Models

Design Chosen⇓	*IDEAS* (off-target only)	**RSM**	**Kriging**	**MARS**
x1	**0.0055264**	0.0050753	**0.0056379**	0.0068702
x2	**0.0030129**	0.0046911	**0.0038799**	0.0030789
x3	**0.0015371**	0.001	**0.0027962**	0.0011942
x4	**0.0039918**	0.0046741	**0.0041407**	0.0038379
x5	**0.0042753**	0.0049817	**0.0036714**	0.0031200
x6	**0.0045712**	0.0051376	**0.0047241**	0.0051831
x7	**0.0010877**	0.0011156	**0.001001**	0.0011943
x8	**0.0044559**	0.004	**0.0057058**	0.004
MQL	**17.6368** (off-target only)	19.1172	**18.066811**	19.43
$\widehat{MQL}$	Not Used	18.2130	**18.0982**	17.5035
Runs to Solve = (Comp. Time)	525 (29 hours)	576 (81 min)	595 (342 min)	553 (314 min)

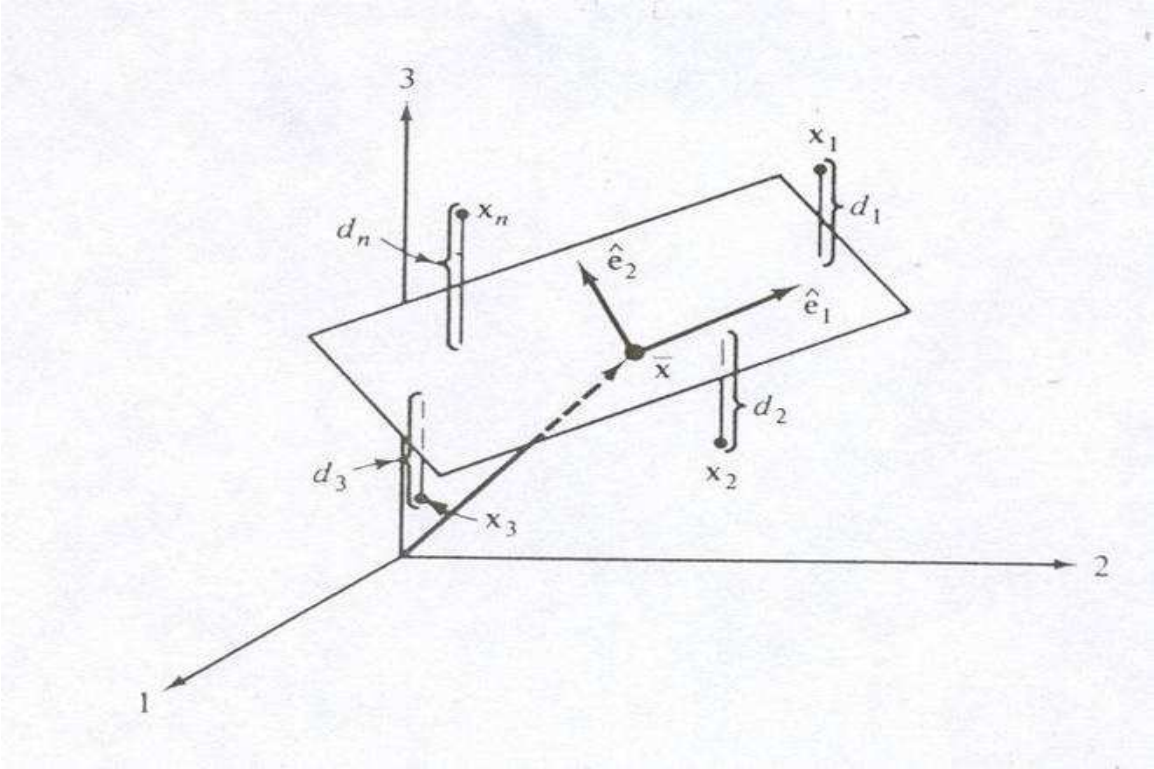

Figure 5.4: page 387 of Johnson and Wichern 1992

the plane formed by the first two eigenvectors of the covariance matrix of the data.

Pearson's proof makes no parametric assumption regarding the data other than the existence of finite variances and covariances of the variables concerned. We thought furthermore that models of the hyper-plane with minimal orthogonally squared distance might converge more quickly than a linear combination of six individual models created with ordinary least squares. This is purely intuitive but makes the idea more attractive to pursue.

We showed in Chapter 4 that the MQL can be decomposed into PCs called LSPC which incorporate the loss constant matrix (LCM) of multivariate quadratic loss. We also showed that for a given subset of LSPC we know what fraction of the trace covariance part of expected MQL is accounted for. We also showed that the same subset covers some fraction of the OTVP depending on the specific off-target vector. Therefore while the full set of LSPC completely replicates the value of MQL, there is some level of error associated with any formal subset of the LSPC.

We know that the expected MQL can be expressed and approximated as

$$\begin{aligned} E(MQL) &= E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \\ &= trace[\mathbf{A}\Sigma_{\mathbf{Y}}] + [E(\mathbf{Y})-\tau]^T\mathbf{A}[E(\mathbf{Y})-\tau] \\ &\approx trace[\Sigma_{\hat{\mathbf{Z}}\mathbf{p}}] + [\hat{\mathbf{Z}}\mathbf{p}-\tau_{\mathbf{Z}p}]^T[\hat{\mathbf{Z}}\mathbf{p}-\tau_{\mathbf{Z}p}]. \end{aligned}$$

where $trace[\Sigma_{\hat{\mathbf{Z}}\mathbf{p}}]$ corresponds to the trace-covariance term of expected MQL and $(\hat{\mathbf{Z}}\mathbf{p} - \tau_{\mathbf{Z}p})^T(\hat{\mathbf{Z}}\mathbf{p}-\tau_{\mathbf{Z}p})$ corresponds to the OTVP of expected MQL.

We propose a robust design optimization procedure which approximates the expected MQL with a subset of the LSPC as follows:

1. Select the first $p < r$ LSPC according to a pre-determined criterion. Using the matrix and vector notation of Chapter 4, we define this group of p LSPC as

$$\begin{aligned} \mathbf{Z}\mathbf{p} &= \mathbf{V}\mathbf{p}^{\mathbf{T}}(\mathbf{q}^{\mathbf{T}}\mathbf{Y}) \\ &= (Z_1,\ldots,Z_p)^T \ . \end{aligned}$$

2. Build a model for each LSPC as a function of the design vector $\mathbf{x}$ as follows,

$$\begin{aligned} \hat{Z}_1 &= f_{Z_1}(\mathbf{x}) \\ &\vdots \\ \hat{Z}_p &= f_{Z_p}(\mathbf{x}) \ . \end{aligned}$$

3. Approximate expected MQL with the models of the p LSPC,

$$E(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \approx trace[\Sigma_{\hat{\mathbf{Z}}\mathbf{p}}] + [\hat{\mathbf{Z}}\mathbf{p}-\tau_{\mathbf{Z}p}]^T[\hat{\mathbf{Z}}\mathbf{p}-\tau_{\mathbf{Z}p}].$$

4. Solve for the design vector $\mathbf{x}$ which

$$\min_{\mathbf{x}\epsilon\Omega}\left\{trace[\Sigma_{\hat{\mathbf{Z}}\mathbf{p}}] + [\hat{\mathbf{Z}}\mathbf{p}-\tau_{\mathbf{Z}p}]^T[\hat{\mathbf{Z}}\mathbf{p}-\tau_{\mathbf{Z}p}]\right\} \ .$$

Figure 5.5 depicts a high level view of how the optimization process works with meta-models of the LSPC. The models referenced in the second step of Figure 5.3 are plugged directly into the second component of MQL which estimates the OTVP. The $trace[\Sigma_{\hat{\mathbf{Z}}\mathbf{p}}]$

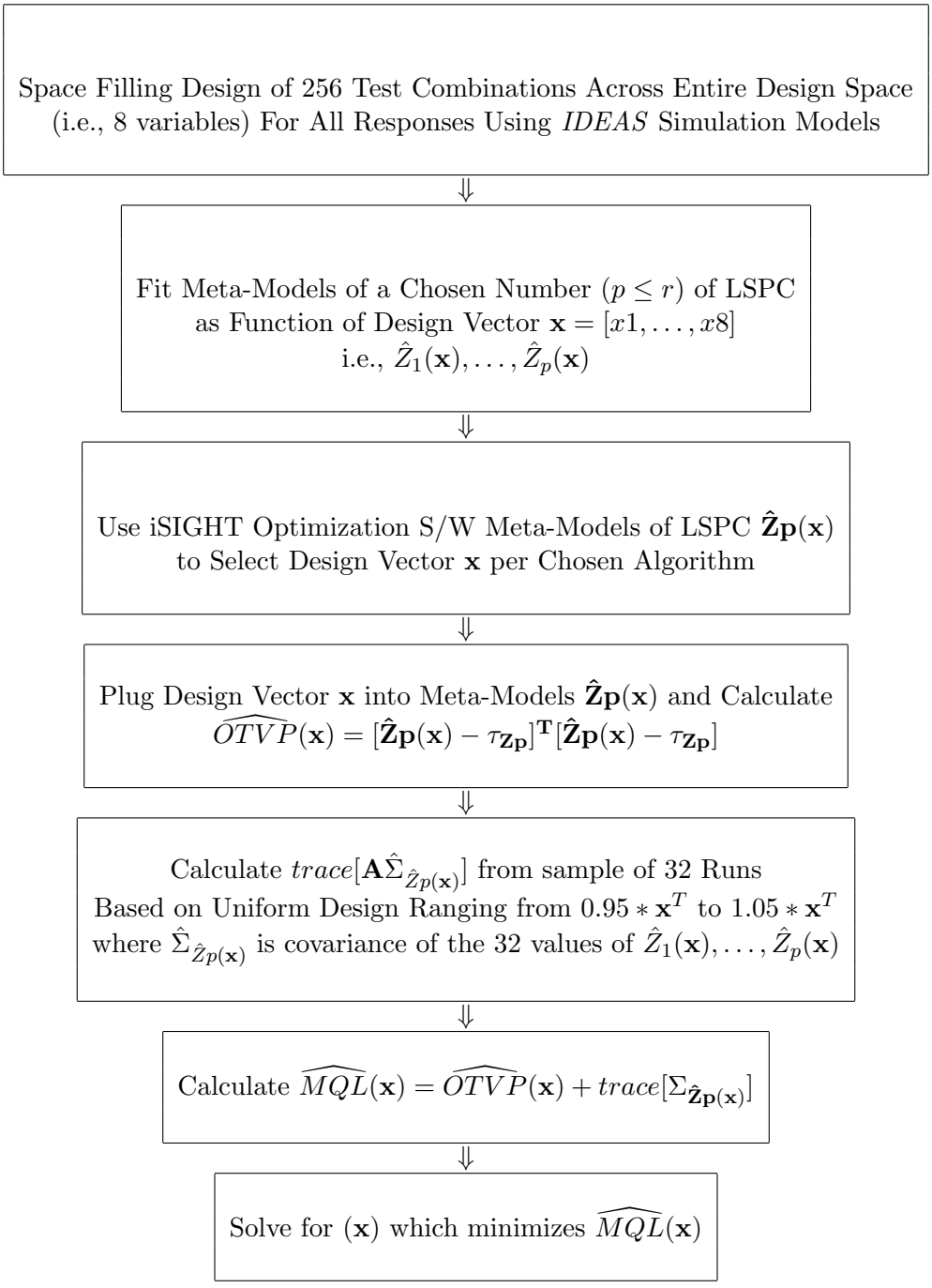

Figure 5.5: Using LSPC Meta-Models in $\widehat{MQL}$ to Solve for Design Vector $\mathbf{x}$

component is calculated by injecting a uniform design build around each design vector chosen by the optimization algorithm. This uniform design chooses from among four levels evenly spaced between 0.95 and 1.05 times the nominal values of each design vector. These runs are analogous to the outer design array of Taguchi designs and the covariance matrix of these "noise" replicates represent the sensitivity to unavoidable variation in the design settings. These are sources of variation in the manufacture of mechanical components so we choose a range of representative manufacturing variability such as +/- five percent. The trace term was then added to the $\widehat{OTVP}$ to yield $\widehat{MQL}$ for each particular design vector.

This procedure can be used iteratively for moving from global to localized design spaces.

5.2.1 Potential Advantages and Disadvantages of this Approach

There are two significant potential advantages to modelling the LSPC over modelling of the standardized responses to emulate a direct search for the design vector minimizing expected MQL:

- Assuming that $p < r$ and reasonable models for the LSPC per some criteria such as R^2, there may be reduced computational time.
- The LSPC meta-models may provide more accurate approximations of the true process than meta-models of the responses because of the optimal lower dimensional estimating property proved by Hotelling (1931).

The disadvantage of this approach is:

- The accuracy of the approximation of OTVP by a formal subset of LSPC varies with the specific off-target vector being evaluated.

5.3 Results of the Electric Vehicle Chassis Design with Meta-Models of Responses and LSPC

We found the optimal vector of eight design variable settings for the chassis of an electric vehicle which has six responses of interest. The responses are a mix of STB, NTB and LTB types and several have constraints. Because of the limitations on the responses from our

initial space fitting design and the desire to avoid extrapolation with the meta-models, we modified each response as a NTB type. The optimal design vector is that which minimizes $\widehat{MQL}$ as estimated by the respective meta-models employed.

We built meta-models of these responses using three different model types. The model types are RSM, Kriging and multi-adaptive regression splines (MARS). We also built models of the full set of six LSPC using the same three types of models. We plugged each set of response models from the three model-building approaches into the optimization algorithm and solved for the optimal design vector. We did the same with the three sets of LSPC models. We are interested in the final design vectors and the number of optimization cycles needed to find them. We repeated the process with subsets of 5 and 4 LSPC models to compare savings in computation and loss of accuracy.

A key concern when considering the use of PCs is whether to derive them from the covariance or correlation matrix of the data. Jackson (1985) stated that when the data (i.e., responses) are measured in different units or of greatly different magnitudes of variance, PCs should be derived from the correlation matrix. Otherwise the covariance matrix is preferred for retention of original units. It is obvious from the greatly varying response ranges in Table 5.1 that we need to derive our PCs from the response correlation matrix.

Choosing the correlation matrix implies the use of standardized response values which require that the values of the loss constant matrix (i.e., LCM) are defined accordingly. Whereas Pignatiello (1993) explained in detail how the values of the LCM reflect the cost of being one unit off target for the non-standardized responses, this data-set requires that our LCM values be defined as the cost of poor quality from being one standard deviation off-target in either direction. The off-diagonal values reflect the additional cost penalty incurred from that pair of responses being simultaneously one standard deviation off-target.

In order to show some consistency between the 3 model types we generated a 64 test combination latin hyper-cube (LHC) across the design space for purposes of estimating the accuracy of prediction. This LHC design was run in *IDEAS* and the response values were compared to the values predicted by each model type. These residual errors provide an estimate of prediction error and are presented in Table 5.5. In general, they do not show

Table 5.5: Prediction Error of LSPC Model Types

LSPC Modelled ⇓	**Pred. Error of RSM Models**	**Pred. Error of Kriging Models**	**Pred. Error of MARS Models**
LSPC1	2.2658	1.4988	2.2382
LSPC2	0.3607	0.2226	0.3587
LSPC3	6.0797	6.4973	7.7932
LSPC4	0.2587	0.1496	0.4099
LSPC5	1.6717	2.2921	2.4398
LSPC6	0.0102	0.0043	0.0072

significant differences in predictive ability across the 64-point verification design space.

5.3.1 Design Vectors Selected by MQL Approximated with Meta-Models of Responses and LSPC

A comparison of the resulting design vectors and their respective computation times yielded by meta-models of the standardized responses and the LSPC to estimate MQL will allow us to judge the utility of our approach in this problem. Three tables are organized to compare the design vectors selected by combinations of model type and meta-model as follows. Table 5.6 compares optimal design vector and computational time for the full sets of LSPC meta-models from three modelling approaches and *IDEAS*. Table 5.7 compares the same for meta-models of the first 5 LSPC and Table 5.8 does the same for meta-models of the first 4 LSPC.

Each of these tables compares the design vector as well as the number of optimization cycles needed to solve. Each table compares the "true" solution chosen by *IDEAS* with the solutions chosen by their analogues among the different model approaches. The value in each row closest to the values chosen by *IDEAS* is in bold face print, as is the value chosen by *IDEAS*, to visually show which model type yields the design variable settings closest to optimal values.

5.4 Summarizing Results and Conclusions

The use of LSPC is moderately successful in producing similar results in a shorter period of time. Due to time constraints, we only obtained the *IDEAS* solution corresponding to the off-target component of MQL. This solution was obtained in 525 iterations over a period of 29 hours. The robust solution includes the evaluation of a design vector's off-target component of loss followed by evaluation of 32 "internal noise" vectors centered around that same design vector. An individual cycle of evaluating the off-target off-target component of loss takes approximately 4 minutes. Adding the noise replicates multiplies that computational time by a factor of 16, meaning an individual cycle with noise relicates takes approximately 65 minutes. The estimated solution time for this RD problem in *IDEAS*, assuming the same number of iterations as for the deterministic off-target component of loss, is approximately 22 days. All past experiments have shown that the trace-covariance component of loss in this problem make up a small fraction of one percent of the MQL associated with an optimal design vector. Solving this problem with noise replicates is included in our list of future research tasks in Chapter 8.

The meta-models each take less than half a minute to calculate the responses and "noise replicates" for each design vector selected by the optimization algorithm. This resulted in total solution time of a few hours for each of the meta-model based techniques. While the meta-models of the responses provide significant savings in computational time, those savings can be extended by using either the full set or a subset of the LSPC of the data. For this example only, the following high level observations are noted :

1. The full set of response meta-models yields design vector values very close to those of

Table 5.6: Design Vectors and Comp. Time from Full Set of LSPC Meta-Models

Design Chosen ⇓	*IDEAS* (off-target only)	**RSM**	**Kriging**	**MARS**
x1	**0.0055264**	0.0049422	**0.0052196**	0.0063381
x2	**0.0030129**	0.0044993	**0.0042574**	0.0035278
x3	**0.0015371**	0.001	**0.001**	0.0013883
x4	**0.0039918**	0.0048697	**0.0053654**	0.0041146
x5	**0.0042753**	0.0048905	**0.0044986**	0.0039165
x6	**0.0045712**	0.0051205	**0.0052613**	0.0041814
x7	**0.0010877**	0.001	**0.001001**	0.0013006
x8	**0.0044559**	0.004	**0.0020382**	0.004
MQL	**17.6368** (off-target only)	18.6257	**18.4772**	19.0712
$\widehat{MQL}$	Not Used	17.7158	17.8017	20.3608
Runs to Solve = (Comp. Time)	525 (29 hours)	566 (114 min)	714 (460 min)	235 (126 min)

Table 5.7: Design Vectors and Comp. Time from Subset of 5 LSPC Models

Design Chosen ⇓	*IDEAS*	**RSM**	**Kriging**	**MARS**
x1	**0.0055264**	0.0049496	**0.0052916**	0.0063913
x2	**0.0030129**	0.0045719	**0.0043554**	0.0034834
x3	**0.0015371**	0.001	**0.001**	0.0013882
x4	**0.0039918**	0.0048394	**0.0051764**	0.0039860
x5	**0.0042753**	0.0048741	**0.0044571**	0.0039031
x6	**0.0045712**	0.0050785	**0.0052396**	0.0042977
x7	**0.0010877**	0.001	**0.0010092**	0.0013005
x8	**0.0044559**	0.004	**0.0021167**	0.004
MQL	**17.6368**	18.6531	**18.5196**	19.1005
$\widehat{MQL}$	Not Used	15.6985	15.5207	18.1618
Runs to Solve = (Comp. Time)	525 (29 hours)	513 (98 min)	229 (141 min)	280 (149 min)

Table 5.8: Design Vectors and Computational Time from Subset of 4 LSPC Models

Design Chosen ⇓	*IDEAS* (off-target only)	**RSM**	**Kriging**	**MARS**
x1	**0.0061328**	**0.0056615**	0.0065985	0.0064214
x2	**0.0041071**	**0.0043851**	0.0040401	0.0037083
x3	**0.0019598**	**0.0040582**	0.0053019	0.0029500
x4	**0.0034235**	**0.0041176**	0.0039482	0.0047322
x5	**0.0041462**	**0.0052525**	0.0045717	0.0046497
x6	**0.0051624**	**0.0041565**	0.0047048	0.0044643
x7	**0.0014794**	**0.0015581**	0.0018469	0.0018830
x8	**0.0030897**	**0.004**	0.0040139	0.004
MQL	**17.6368** (off-target only)	**22.6823**	25.5888	23.5363
$\widehat{MQL}$	Not Used	0.098	0.0094	0.1062
Runs to Solve = (Comp. Time)	525 29 hours	139 17 min	100 59 min	184 96 min

the *IDEAS* simulations,

2. The full set of response meta-models identifies the optimal design vector much faster than *IDEAS* but not as accurately,

3. Using the full set of LSPC models typically yields a design vector as close or closer to *IDEAS* than those chosen with the response models,

4. Using the full set of MARS models of LSPC yielded a design vector close to optimal in less than half the number of optimization runs and hence computational time of the analogous response models,

5. For RSM and Kriging models, approximately the same number of optimization runs were required by the full sets of responses and LSPC,

6. In all three modelling approaches, significant reductions in the number of optimization cycles was achieved by using either the full set or a subset of the LSPC models. This gain in computation time is typically accompanied by some loss of accuracy in the design vector.

7. In this example, the use of LSPC is feasible and useful in early design stages.

CHAPTER VI

DIAGNOSING MULTIVARIATE SPC WITH PCS

In the first section of this chapter we compare several different approaches to MSPC. In Section 2 we discuss some methods of diagnosing the variables responsible for driving individual multivariate observations out of control, with special attention to the technique of Kourti and MacGregor (1996). In Section 3 we apply the technique of Kourti and MacGregor (1996) to diagnose the individual variables driving a subgroup's location out of control and extend it to do the same for dispersion. In Section 4 we demonstrate our procedure and compare it to that of the MP charts of Fuchs and Benjamini (1994). In Section 5 we make concluding remarks about our proposed diagnostic technique.

6.1 An Overview of Techniques Used in MSPC

Products and processes with multiply correlated quality characteristics are commonly found in industry today. Jackson (1985) listed the three essential properties of a MSPC technique as:

1. Giving a clear indication of whether a process is in control,
2. Maintenance of a specific probability of Type I error,
3. Exploitation of the correlation structure.

The majority of multivariate techniques fall into the category of directionally invariant, where the average run length (ARL) behavior is related to the off-target vector of quality characteristics and their covariance matrix only through the non-centrality parameter (NCP) defined as

$$NCP \quad = \quad (E\mathbf{Y} - \tau)^T \Sigma_{\mathbf{Y}}^{-1} (E\mathbf{Y} - \tau),$$

where $E\mathbf{Y}$ is the vector of expected values of the quality characteristics and τ and $\Sigma_{\mathbf{Y}}$ are respectively the vector of target values and covariance matrix of the same quality characteristics.

Included under the umbrella of directionally invariant approaches are the χ^2 chart of Hotelling (1931), the method of principal components (PCs) as described in Jackson (1980) and Jackson (1981), the MCUSUM charts of Pignatiello and Runger (1990), the MEWMA chart of Lowry et al (1992), the M chart of Hayter and Tsui (1994) and the MTY decomposition, named after its authors Mason, Tracy and Young (1995).

In contrast to these directionally invariant techniques, there are only a few techniques relating ARL performance to specific directions of the off-target vector. These techniques either place a premium on detecting process shift in a pre-meditated direction (see Healy (1987) and Hawkins (1993)), or employ a loss function to assign directionally dependent financial costs to the distance separating individual and pairs of quality characteristics from their targets (see Mohebbi and Hayre (1989) and Tsui and Woodall (1993)).

All of the directionally invariant and variant techniques mentioned so far satisfy the three characteristics mentioned by Jackson (1985). They differ in their relative strengths regarding some important secondary characteristics which include sensitivity to the size of the process shift as well as identification of the specific quality characteristics responsible for the process shift. An additional consideration is their handling of the economic priorities of the quality characteristics.

It is generally agreed that there is no single "silver bullet" which uniformly achieves better results under all circumstances for the secondary characteristics just mentioned. For instance, while the MCUSUM and MEWMA are faster to detect small process shifts than the χ^2 chart of Hotelling, the latter is more reliable for detecting larger shifts. However neither MCUSUM, MEWMA nor the χ^2 chart identify the specific quality characteristics that are driving the process out of control. On the other hand, the MTY decomposition can clearly identify individual quality characteristics which are driving the process out of control, but becomes computationally cumbersome for a large number of quality characteristics.

Table 6.1: Comparison of MSPC Techniques

CRITERIA⇒ MSPC⇓ Technique	Detection of Small Shifts	Detection of Large Shifts	Directional Invariance	Economic Info. Used	Identify Causes	Assumes MVN($\mu, \mathbf{\Sigma}$)
Hotelling's T^2	Slow	Fast	Yes	No	No	Yes
MCUSUM	Fast	Usually Fast	Yes	No	No	Yes
MEWMA	Fast	Usually Fast	Yes	No	No	Yes
PCs	Slow	Fast	Yes	No	Yes	Yes
M	Slow	Fast	Yes	No	Yes	Yes
MYT	Slow	Fast	Yes	No	Yes	Yes

Tertiary considerations include whether a particular MSPC scheme can be made robust with respect to autocorrelation of the observed characteristics. While PCs remove the cross-correlation of the observations, they are not immune to their serial correlation. There is lots of research potential for making the diverse MSPC schemes robust to auto-correlation.

Tables 6.1 and 6.2 compare the merits of the well known MSPC approaches with respect to the properties mentioned over the following two pages. In Section 3 we propose an MSPC technique which extends the traditional use of PCs by incorporating the economic prioritization of the quality characteristics inherent to the MQL.

Table 6.2: Addtional Criteria of MSPC Techniques

CRITERIA⇒ MSPC⇓ Technique	Detection of Small Shifts	Detection of Large Shifts	Directional Invariance	Economic Info. Used	Identify Causes	Assumes MVN($\mu, \mathbf{\Sigma}$)
Cause Selecting	Slow	Fast	Yes	No	Yes	No
Loss Function	Fast	Fast	No	Yes	No	Yes
Regr. Adjustmt.	Fast	Fast	No	No	Yes	Yes
LSPC	Depends on Correlations and LCM	Depends on Correlations and LCM	No	Yes	Yes	Yes

References pertinent to the entries in Tables 6.1 and 6.2 are the following:

Hotelling's T^2	=	Hotelling (1951)
MCUSUM	=	Pignatiello and Runger (1990)
MEWMA	=	Lowry et al (1992)
PCs	=	Jackson (1985)
M	=	Hayter and Tsui (1994)
MYT	=	Mason, Tracy and Young (1995)
Cause Selecting	=	Wade and Woodall (1993)
Loss Function	=	Tsui and Woodall (1993)
Regression Adjustment	=	Healy (1987)

6.2 Diagnosing Causal Variables in MSPC

Multivariate statistical process control (MSPC) commonly uses Hotelling's T^2 statistic to indicate when a multivariate observation goes out-of-control. For independently and identically distributed multivariate normal observations, Mason, Tracy and Young (1995) and Kourti and MacGregor (1996) accurately diagnose which specific variables are driving the T^2 statistic out-of-control. For subgroups of multivariate observations, Jackson (1985) advocates decomposing the overall T^2 into independent T^2 statistics for separate monitoring of location and dispersion. We show how the PC-based procedure of Kourti and MacGregor (1996) diagnoses the specific variables shifting subgroup location and extend it to do the same for increased dispersion. The procedure is demonstrated on a data-set from Fuchs and Kenett (1998).

6.2.1 Background

A number of papers and books examine the state of multivariate statistical process control (MSPC). These include the review papers by Lowry and Montgomery (1995) and Mason et al (1997) and the books by Fuchs and Kenett (1998) and Mason and Young (2002). For multivariate processes monitored with Hotelling's T^2 statistic, these references state that principal components are sometimes useful in diagnosing which individual variables have driven a multivariate vector out-of-control. Several of these sources assert that the efficacy of principal component based diagnosis is contingent upon how well these components approximate a physically interpretable latent factor. Kourti and MacGregor (1996) show that under multivariate normality the normalized scores of the principal components can accurately identify the causal variables regardless of their physical meaning. In contrast, Mason and Young (2002) diagnose individually out-of-control vectors by orthogonal decomposition of the T^2 statistic.

When monitoring rational subgroups of multivariate observations, the techniques of Mason and Young (2002) and Kourti and MacGregor (1996) have limited diagnostic capability because these subgroups are prone to shifts of scale as well as location. Mason and Young (2002) state that their decomposition procedure does not discern between the

specific variables responsible for shifting location versus those responsible for driving dispersion out-of-control. We will show that the procedure of Kourti and MacGregor (1996) can be directly applied to diagnose the variables responsible for driving subgroup mean out-of-control and extended to independently select those variables driving subgroup dispersion out-of-control.

Of the works referenced here, only Fuchs and Kenett (1998) propose a technique for identifying the individual variables that drive a particular subgroup's variance out of control. We demonstrate the diagnostic potential of our procedure on a data-set from Fuchs and Kenett (1998) with out-of-control T^2 values attributable to out-of-control subgroup dispersion.

This section is organized as follows. Subsection 2 summarizes the principal component based procedure of Kourti and MacGregor (1996). Subsection 3 extends that procedure to detection of location and scale in the multivariate case. Subsection 4 demonstrates the new procedure on data from Fuchs and Kenett (1998) and concluding remarks are presented in Subsection 5.

6.2.2 Using PCs to Diagnose Shifts of Location in Single Observations

In this section, we briefly describe the method put forth by Kourti and MacGregor (1996) for identifying the specific process variables which cause an individual multivariate vector to go out-of-control. Assume that vector $\mathbf{Y}$ is independently multivariate normal of dimension r with mean vector $= \mu$ and covariance matrix $= \Sigma$, i.e.,

$$\mathbf{Y} \sim \quad N_r(\mu, \Sigma) \ .$$

The following $\chi^2_{(r)}$ statistic can be used to monitor the individual multivariate observations when the population mean and covariance matrix are known,

$$\chi^2_{(r)} \quad = \quad (\mathbf{Y} - \mu)^T \Sigma^{-1} (\mathbf{Y} - \mu).$$

More commonly the population parameters are not known, and Hotelling's T^2 for observation $\mathbf{Y}$ is defined as

$$T^2 \quad = \quad (\mathbf{Y} - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbf{Y} - \hat{\mu}) \ ,$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are respectively the estimates of the population mean vector and covariance matrix obtained from the historical data set in Phase I. Assumptions regarding the dependence between estimates gathered in Phase I and Phase II of the process determine whether T^2 has an F or β distribution.

Among others Jackson (1985) has shown that the T^2 value corresponding to multivariate data-point $\mathbf{Y}$ can be decomposed as:

$$\begin{aligned} T^2 &= \Sigma_{p=1}^{r} \frac{t_p^2}{\lambda_p} \\ &= \Sigma_{p=1}^{r} \frac{t_p^2}{\sigma_p^2} \\ &= \Sigma_{p=1}^{r} z_p^2 \ , \end{aligned}$$

where the t_p for $p = 1 \ldots r$ are the full set of r principal component scores obtained from $\hat{\Sigma}$, λ_p the corresponding eigenvalues, and σ^2 is a reminder that each eigenvalue is the variance of its related score. Dividing each principal component score by its standard deviation i.e., $\sigma_p = \sqrt{\lambda_p}$, yields the normalized principal component z_p. Therefore the terms being summed are the squares of the normalized scores.

Jackson (1980) introduced a diagnostic approach for identifying the specific variables driving the T^2 statistic out of control which employed univariate charts of the normalized scores defined in equation (6.-3). Jackson (1991) modified his 1980 procedure to examine the univariate normalized scores with Bonferroni bounds to monitor a type I error of at most α. The oft cited criticism aimed at Jackson's principal component-based diagnostic approach is its reliance on a reasonable physical interpretation of the normalized score being signalled.

Kourti and MacGregor (1996) addressed this limitation by plotting bar-charts of the variable contributions to the normalized score being signalled. They show that the individual normalized scores z_p can be written as

$$\begin{aligned} z_p &= \frac{\mathbf{u}_p^T(\mathbf{Y} - \mu)}{\lambda_p^{1/2}} \\ &= \Sigma_{j=1}^{r} \frac{u_{p,j}(Y_j - \mu_j)}{\lambda_p^{1/2}} \ , \end{aligned}$$

where the column vector $\mathbf{u}_p$ is eigenvector p and $\mathbf{Y} = [Y_1, \ldots, Y_r]^T$ the vector of quality characteristics. The $u_{p,j}$ are the components of $\mathbf{u}_p$ corresponding to each of the individual responses, Y_j for $j = 1 \ldots r$, which make up $\mathbf{Y}$. The μ_j are the means of the r individual variables making up the vector μ.

They further express the individual variable contribution of response Y_j to the normalized score z_p as

$$VarCont_{p,j} = \frac{u_{p,j}(Y_j - \mu_j)}{\lambda_p^{1/2}} .$$

Therefore the normalized scores are simply the scaled sum of the individual variable contributions and only those contributions that are of the same sign as the resultant sum can be significant. Variable contributions of the opposite sign reduce the scaled sum and therefore do not contribute to the score's significance. This allows us to write T^2 in the following way,

$$\begin{aligned} T^2 &= \Sigma_{p=1}^{r} z_p^2 \\ &= \Sigma_{p=1}^{r} \left[\Sigma_{j=1}^{r} \frac{u_{p,j}(Y_j - \mu_j)}{\lambda_p^{1/2}} \right]^2 , \end{aligned}$$

which shows its relation to the individual variable contributions to each normalized score.

Assuming that each multivariate observation $\mathbf{Y}$ has been centered around its historical estimate of mean $\hat{\mu}$, each t_p is approximately distributed as

$$t_p \sim N(0, \lambda_p).$$

This implies, by virtue of the standardizing divisor $\sqrt{\lambda_p}$, that the normalized scores are approximately standard normal i.e.,

$$z_p \sim N(0, 1).$$

Therefore any value of a normalized score approaching or exceeding three is statistically different from its historical estimate at an $\alpha = 0.0027$. A simple bar-chart of the normalized PCs quickly and accurately identifies which PCs are out of control. Bar-charts of the variable contributions defined in equation 6.-8 can then be used to diagnose the individual variables suspected of causing the event.

The procedure of Kourti and MacGregor (1996) consists of the following steps:

1. Out of limit value of T^2 detected for observation $\mathbf{Y}$.

2. Plot normalized scores of observation $\mathbf{Y}$. Bonferroni limits may be used as guides, α/r for each score chart where α is for the T^2 being monitored.

3. Plot a bar chart of the variable contributions to each significant normalized score.

4. Investigate those variables making a large contribution of the same sign as the significant normalized scores.

Mason et al (1997) state that due to its use of principal components, this approach is particularly useful for large and ill-conditioned data sets .

6.3 Using PCs to Diagnose Shifts in Subgroup Location and Dispersion

For the same reasons that subgroups of univariate data are collected, rational subgroups of multivariate observations yield more reliable process information than individual vectors. Research including that of Hawkins (1993), Kourti and MacGregor (1996) and Mason and Young (2002) can be used to diagnose the variables which drive the subgroup's overall T^2 out of control. However these do not typically differentiate between the variables shifting location versus those driving higher dispersion. In order to distinguish which variables cause which type of shift, we first decompose this statistic into two independent parts representing subgroup location and dispersion.

6.3.1 Decomposition of Hotelling's T^2

Consider a rational subgroup k consisting of the n individual multivariate data-points $\mathbf{Y}_{ki}$ where $i = 1 \ldots n$ and each $\mathbf{Y}_{ki}$ is of dimension r. The overall subgroup T^2 for subgroup k, i.e., T^2_{Ok}, is the sum of the T^2 of the n individual multivariate data-points that make up subgroup k. T^2_{Ok} is also known as the *Lawley-Hotelling* trace statistic (see Lawley (1938) and Hotelling (1951)) and has an χ^2 distribution with rn degrees of freedom (Jackson (1991)).

Jackson (1985) discussed the following decomposition of Hotelling's T^2_{Ok}

$$T^2_{Ok} = T^2_{Mk} + T^2_{Dk} ,$$

where T^2_{Mk} tests whether the sample mean of this subgroup has shifted away from the estimated population mean and is defined as

$$T^2_{Mk} = n(\bar{\mathbf{Y}}_k - \hat{\mu})^T \hat{\Sigma}^{-1} (\bar{\mathbf{Y}}_k - \hat{\mu}),$$

where $\bar{\mathbf{Y}}_k$ is the vector whose components are the subgroup averages of the n observations in subgroup k. The distribution of T^2_{Mk} is asymptotically $\chi^2_{(r)}$, the same as that of the T^2 of an individual observation (Jackson (1985)).

T^2_{Dk} tests whether the variance within this subgroup is significantly greater than historical variance and is defined as

$$T^2_{Dk} = \Sigma_{i=1}^{n} (\mathbf{Y}_{ki} - \bar{\mathbf{Y}}_k)^T \hat{\Sigma}^{-1} (\mathbf{Y}_{ki} - \bar{\mathbf{Y}}_k).$$

The distribution of T^2_{Dk} is asymptotically $\chi^2_{r(n-1)}$ where r is the number of quality characteristics in each multivariate observation and n the number of observations in the subgroup (Jackson (1985)).

Whereas T^2_{Ok} contains all the information in the subgroup, Jackson (1985) argues that it is of little diagnostic value since when significant, the analyst must immediately ascertain whether it is T^2_{Mk}, the multivariate analog of the $\bar{\mathbf{x}}$ chart, or T^2_{Dk}, the multivariate analog of the r chart, that is responsible.

6.3.2 Applying Kourti and MacGregor (1996)'s Procedure to Subgroup Location and Dispersion

In this section we show why the PC-based procedure of Kourti and MacGregor (1996) diagnoses the specific quality characteristics shifting subgroup location and extend it to do the same for increased dispersion. As in their procedure, we assume that the multivariate observations are independently and identically distributed as multivariate normal. Since PC's are derived from decomposition of the estimated covariance matrix $\hat{\Sigma}$, we first re-express $\hat{\Sigma}^{-1}$ in terms of the eigenvectors and eigenvalues of $\hat{\Sigma}$.

Since $\hat{\Sigma}$ is positive definite, it can be expressed as follows,

$$\hat{\Sigma} \;=\; \mathbf{u}\Lambda\mathbf{u}^{\mathbf{T}} \;,$$

where $\mathbf{u}$ is an r by r orthogonal matrix whose columns are the eigenvectors of $\hat{\Sigma}$ and Λ an r by r diagonal matrix containing the pertinent eigenvalues. The inverse of $\hat{\Sigma}$ can be expressed in similar manner as

$$\begin{aligned}\hat{\Sigma}^{-1} &= \mathbf{u}\Lambda^{-1}\mathbf{u}^{\mathbf{T}} \\ &= [\mathbf{u}\Lambda^{-1/2}][\mathbf{u}\Lambda^{-1/2}]^T \;.\end{aligned}$$

The covariance matrix of subgroup mean $\bar{\mathbf{Y}}_{\mathbf{k}}$ is $\frac{\hat{\Sigma}}{n}$. Its inverse can also be expressed as

$$\begin{aligned}\left(\frac{\hat{\Sigma}}{n}\right)^{-1} &= \mathbf{v}\Lambda_{\bar{\mathbf{Y}}}^{-1}\mathbf{v}^{\mathbf{T}} \\ &= [\mathbf{v}\Lambda_{\bar{\mathbf{Y}}}^{-1/2}][\mathbf{v}\Lambda_{\bar{\mathbf{Y}}}^{-1/2}]^T \;,\end{aligned}$$

where v are the eigenvectors of $\frac{\hat{\Sigma}}{n}$ and the diagonal elements of $\Lambda_{\bar{\mathbf{Y}}}$ the corresponding eigenvalues.

6.3.2.1 Applying Kourti and MacGregor (1996) to Subgroup Location

Plugging the diagonalized form of $\frac{\hat{\Sigma}}{n}$ into the decomposition of subgroup T_{Mk} yields

$$\begin{aligned}T_{Mk} &= [\mathbf{v}^T\Lambda_{\bar{\mathbf{Y}}}^{-1/2}(\bar{\mathbf{Y}}_{\mathbf{k}} - \hat{\mu})]^T[\mathbf{v}^T\Lambda_{\bar{\mathbf{Y}}}^{-1/2}(\bar{\mathbf{Y}}_{\mathbf{k}} - \hat{\mu})] \\ &= \Sigma_{p=1}^{r}\left[\frac{\mathbf{v}_p^T(\bar{\mathbf{Y}}_{\mathbf{k}} - \hat{\mu})}{\pi_p}\right]^T \;,\end{aligned}$$

which we recognize as the sum of squares of normalized scores based on the vector of differences between the subgroup mean vector $\bar{\mathbf{Y}}_{\mathbf{k}}$ and the estimated mean vector $\hat{\mu}$. The eigenvectors $\mathbf{v_p}$ and eigenvalues π_p in equation 6.-21 are derived from the covariance matrix of $\bar{Y}_k$, i.e., $\frac{\hat{\Sigma}}{n}$.

We define these normalized score terms based on subgroup location as normalized scores of location (NSL) as follows,

$$NSL_{p,k} \;=\; \left[\frac{\mathbf{v}_p^T(\bar{\mathbf{Y}}_{\mathbf{k}} - \hat{\mu})}{\pi_p}\right] \;.$$

The NSL are approximately standard normal so any value approaching or exceeding three is statistically different from its population mean at an $\alpha = 0.0027$. A simple bar-chart of the NSL quickly and accurately identifies which are out of control. For a single multivariate observation, i.e., $n = 1$, this reduces to equation (6.-6), which Kourti and MacGregor (1996) use to diagnose variables responsible for driving a single observation out-of-control. Just as variable contributions to normalized scores are used to diagnose individual observations, variable contributions to the NSL are defined as follows, i.e., $VCNSL$,

$$VCNSL_{p,k,j} = \frac{v_{p,j}(\bar{Y}_{k,j} - \hat{\mu}_j)}{\pi_p^{1/2}} ,$$

where $v_{p,j}$ is the element of eigenvector p corresponding to the individual variable j and $\bar{Y}_{k,j}$ is the average of the Y_j from each observation in subgroup k. We can now write out T^2_{Mk} as

$$T^2_{Mk} = \Sigma_{p=1}^{r} \left[\Sigma_{j=1}^{r} \frac{v_{p,k,j}(\bar{Y}_{k,j} - \hat{\mu}_j)}{\pi_p^{1/2}} \right]^2 ,$$

which shows its relation to the individual variable contributions to the squared $NSL_{p,k}$.

For $n > 1$, the procedure of Kourti and MacGregor (1996) diagnoses the variables responsible for driving subgroup mean out-of-control by following these steps:

1. A subgroup with an out-of-control value of T^2_{Mk} is detected at some α.
2. Plot a bar-chart of the $NSL_{p,k}$. Bonferroni limits may be used as guides, α/r for each score where α is for the T^2 being monitored.
3. For each significant $NSL_{p,k}$, plot a bar-chart of the $VCNSL_{p,k,j}$.
4. Investigate those variables making a large contribution of the same sign as the statistically significant $NSL_{p,k}$.

6.3.2.2 Extending Kourti and MacGregor (1996) to Subgroup Dispersion

Plugging the diagonalized form of $\hat{\Sigma}^{-1}$ into the decomposition of subgroup T_{Dk} yields

$$\begin{aligned} T^2_{Dk} &= \Sigma_{i=1}^{n} [\mathbf{u}_p^T \Lambda^{-1/2} (\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)]^T [\mathbf{u}_p^T \Lambda^{-1/2} (\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)] \\ &= \Sigma_{i=1}^{n} \Sigma_{p=1}^{r} \left[\frac{\mathbf{u}_p^T (\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)}{\lambda_p^{1/2}} \right]^2 , \end{aligned}$$

which we recognize as the sum of squares of normalized scores calculated from the vector of differences between each individual observation $\mathbf{Y}_{k,i}$ and the subgroup mean vector $\bar{\mathbf{Y}}_k$. We define scores of dispersion (ScD) as

$$ScD_{p,k,i} \quad = \quad \mathbf{u}_p^T(\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k) \ ,$$

where p denotes the principal component, k the subgroup and i the individual observation within the subgroup.

Since the average value of $(\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)$ within any subgroup equals zero, we assume that the $ScD_{p,k,i}$ are approximately normal as follows,

$$ScD_{p,k,i} \quad \sim \quad N(0, \lambda_p) \ .$$

The $ScD_{p,k,i}$ are normalized by dividing by the square root of the associated eigenvalue. We define normalized scores of dispersion ($NSD_{p,k,i}$) as follows,

$$NSD_{p,k,i} \quad = \quad \frac{\mathbf{u}_p^T(\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)}{\lambda_p^{1/2}} \ ,$$

which are approximately standard normal, i.e.,

$$NSD_{p,k,i} \quad \sim \quad N(0, 1) \ .$$

We can now re-write the decomposition of T^2_{Dk} as the sum of the squares of the $NSD_{p,k,i}$ from the n observations in subgroup k as follows,

$$\begin{aligned} T^2_{Dk} \quad &= \quad \Sigma_{i=1}^n \Sigma_{p=1}^r \left[\frac{\mathbf{u}_p^T(\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)}{\lambda_p^{1/2}} \right]^2 \\ &= \quad \Sigma_{i=1}^n \Sigma_{p=1}^r \left[NSD_{p,k,i} \right]^2 \\ &= \quad \Sigma_{p=1}^r \Sigma_{i=1}^n \left[NSD_{p,k,i} \right]^2 \ . \end{aligned}$$

Based on the approximate standard normality of the $NSD_{p,k,i}$, we can show that the sum of their squares within subgroup k are approximately distributed as $\chi^2_{(n-1)}$, i.e.,

$$\Sigma_{i=1}^n \left[NSD_{p,k,i} \right]^2 \quad \sim \quad \chi^2_{(n-1)} \ . \tag{6.-35}$$

An outline of the proof of equation 6.-35 is given in Appendix A.

Therefore plotting the sums of squared $NSD_{p,k,i}$ allows us to quickly identify which of the $NSD_{p,k,i}$ have significant sample variance. We next decompose the $NSD_{p,k,i}$ with significant sample variance into contributions from individual variables. For each observation $\mathbf{Y}_{k,i}$ in subgroup k, individual variable contributions to the normalized scores of subgroup dispersion ($VCNSD_{p,k,i,j}$) are defined as follows:

$$VCNSD_{p,k,i,j} = \frac{u_{p,j}(Y_{k,i,j} - \bar{Y}_{k,j})}{\lambda_p^{1/2}} ,$$

where the subscript p indicates the specific score, k the subgroup, i the specific observation and j the individual variable. $Y_{k,i,j}$ is the value of variable Y_j in observation $\mathbf{Y}_{k,i}$ and $\bar{Y}_{k,j}$ the average of the Y_j values within subgroup k. In our procedure, we will plot the standard deviations of the variable contributions to each $NSD_{p,k}$ whose sum of squared $NSD_{p,k,i}$ is statistically significant.

For subgroups of $n > 1$, we propose the following extension of Kourti and MacGregor (1996) to diagnose the variables responsible for driving subgroup dispersion out-of-control:

1. Detect some subgroup k with an out-of-control value of T^2_{Dk},
2. Plot $\Sigma_{i=1}^{n} [NSD_{p,k,i}]^2$ for all $NSD_{p,k}$ on a bar-chart.
3. Look for values of $\Sigma_{i=1}^{n} [NSD_{p,k,i}]^2$ that exceed the critical value of $\chi^2_{(n-1)}$ chosen for a specific α. Bonferroni limits may be used as guides, α/r for each score chart where α is for the T^2_{Dk} being monitored.
4. For each $NSD_{p,k}$ with a statistically significant value of $\Sigma_{i=1}^{n} [NSD_{p,k,i}]^2$, calculate the standard deviations of the $VCNSD_{p,k,i,j}$ from the n observations of $NSD_{p,k}$ in subgroup k.
5. Plot a bar chart of the standard deviations computed in step 3.
6. Investigate those variables whose standard deviations of $VCNSD_{p,k,i,j}$ are the largest.

6.4 Demonstration and Comparison of the Procedure

To demonstrate and contrast techniques in MSPC, Fuchs and Kenett (1998) employ a data-set they call Case 1 consisting of seventy observations of six dimensions each. This data-set

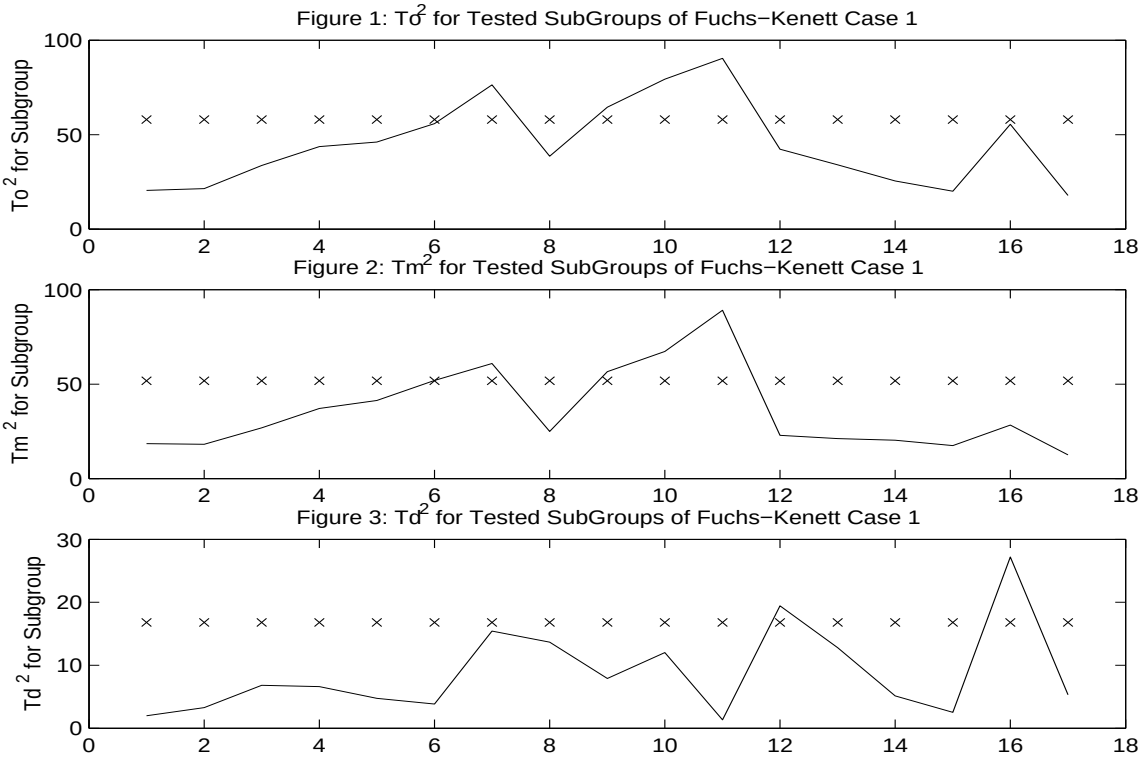

Figure 6.1: T^2_{Ok}, T^2_{Mk} and T^2_{Dk} for first 17 Phase II Subgroups

is divided into thirty five subgroups of two apiece with the first fifteen used to generate historical estimates of the mean vector and the covariance matrix. The last twenty are used as the Phase II data-set for comparison against the historical values.

Figure 6.1 plots T^2_{Ok}, T^2_{Mk} and T^2_{Dk} values for the first 17 of the Phase II subgroups. Critical values at $alpha = 0.01$ for the three statistics are indicated by the horizontal lines of the letter x. These three figures emphasize the point made by Jackson (1991) that a significant value of T^2_{Ok} typically requires further inquiry as to whether location, dispersion or both are driving the subgroup out-of-control.

We will focus on subgroups 11 and 16 whose T^2_{Ok} are significant at $alpha = 0.01$. Subgroup 11 has a significant T^2_{Mk} and non-significant T^2_{Dk} while the converse is true for subgroup 16.

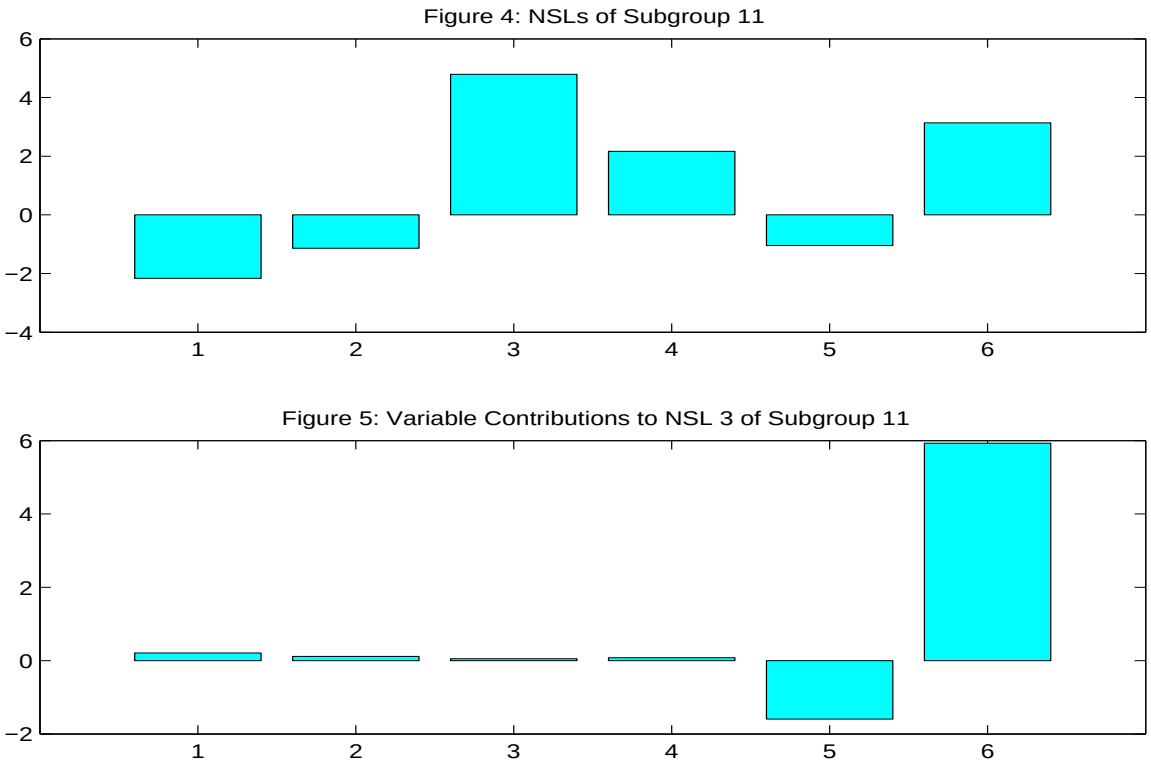

Figure 6.2: NSLs and Var. Contr. to NSL3 in Phase II Subgroup 11

6.4.1 Demonstration of the PC-based Diagnostic Procedures

Figure 6.2 plots the results of applying the procedure of Kourti and MacGregor (1996) for diagnosing location to subgroup 11.

In Figure 6.2 the third NSL from subgroup 11 is most significant followed by NSL 6. Remember that the NSL are approximately standard normal so that any value approaching or exceeding 3 is very significant. Variable 6 makes the largest same sign contribution to the third NSL. By adjusting the value of variable 6 in one of the subgroup 11 data points closer to its historical mean, the value of T^2_{M11} decreased from 89.09 to 40.87, well below the critical value of 51.75.

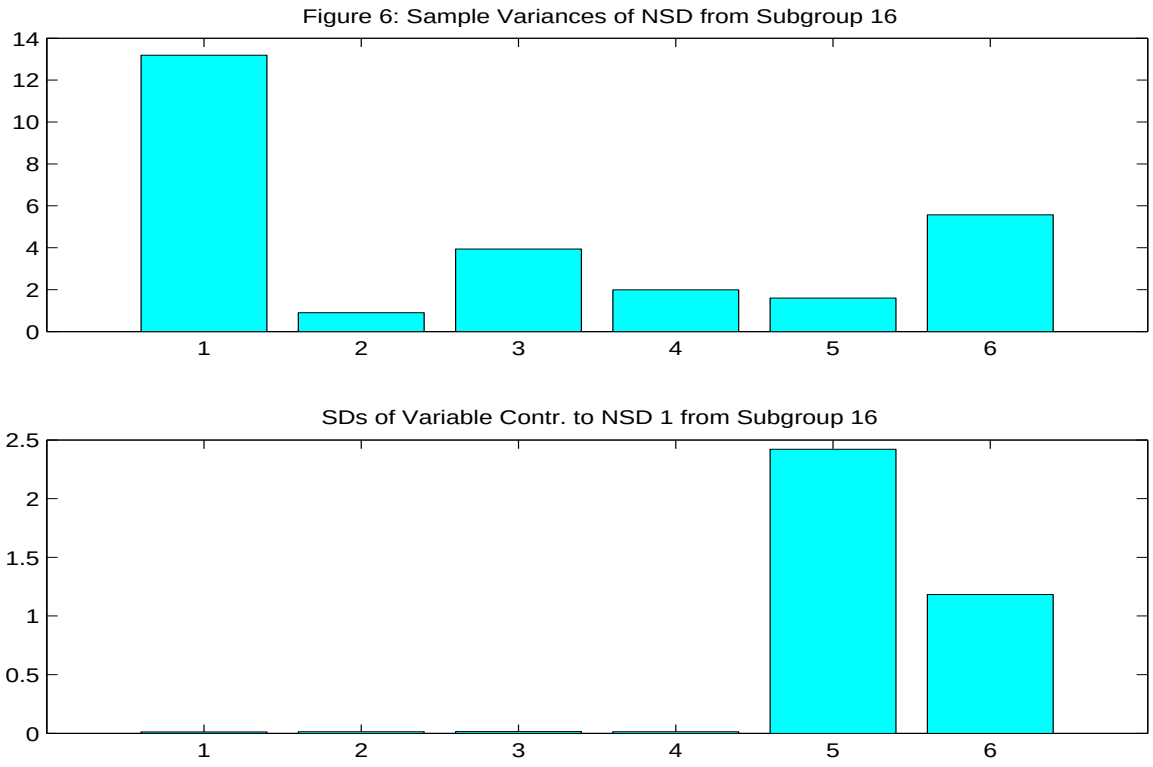

Figure 6.3: Sum of Squared NSD and SD(VarContr) to NSD1 in Subgroup 16

Figure 6.3 shows the results of applying our procedure for diagnosing dispersion to subgroup 16. Note that since $n = 2$ in this example, the product of $(n-1) * S^2(NSD_{p,k,i})$ is simply the sample variance of the $NSD_{p,k,i}$.

In Figure 6.3 the sample variance of the first NSD from subgroup 16 is of highest significance and the contributions of variables 5 and 6 have the largest variance in decreasing order. By adjusting the values of variables 5 and 6 in one of subgroup 16's observations closer to their subgroup averages, the value of T^2_{Dk} decreased from 27.19 to 11.92, well below the critical value of 16.81.

6.4.2 Comparing Techniques for Diagnosis of Subgroup Dispersion

Of the works referenced in this paper, only Fuchs and Kenett (1998) demonstrate a technique for identifying the individual variables that drive a particular subgroup's variance out of control. They employ a graphical diagnostic tool called the multivariate profile (MP) chart which can be constructed for separate diagnosis of shifts in location and variation.

MP charts were introduced by Fuchs and Benjamini (1994) and position the baseline of a miniature bar-chart of scaled deviations at the vertical magnitude of the respective T^2_{Mk} or T^2_{Dk} statistic. The miniature bar-chart for location, which is vertically positioned at the magnitude of the subgroup's T^2_{Mk} value, displays the scaled deviation of each individual variable's subgroup mean with respect to standard values of location. The miniature bar-chart for dispersion, which is vertically positioned at the magnitude of the subgroup's T^2_{Dk} value, displays the scaled deviation of each individual variable's subgroup dispersion with respect to standard values of dispersion.

The miniature bar-charts are essentially individual variable charts of measures of location and dispersion which do not consider correlation and are not directly related to the decomposition of the relevant T^2 statistic. The MP chart's simultaneous presentation of the multivariate T^2 statistic and bar-charts of univariate scaled deviation provide an informative snapshot of subgroup behavior which facilitates quick identification of the suspect variables.

Figure 6.4 below is a reproduction of page 130 of Fuchs and Kenett (1998) which plots the MP charts for dispersion for the last twenty subgroups of the Case 1 data-set. Notice that a curve formed by connecting the baselines of each subgroup's MP dispersion chart mimics the shape of the T^2_{Dk} chart in Figure 6.1.

Notice that subgroups 16 and 18 are out of control at an α of 0.0027 and that subgroups 7, 8, 10, 12 and 13 are out of control at an α of 0.05. The MP dispersion chart suggests that the individual variables with higher scaled deviation be investigated first. The degree of darkness of the individual bars indicates the level of significance with darker bars having greater significance.

Notice that the miniature bar-charts of subgroups 10, 16 and 18 each have a few individual variables whose deviations are very dark and tower above the others. This means that those individual variables, namely variables 5 and 6 in subgroups 10 and 16 and variables 1 through 4 in subgroup 18, are clearly indicated as responsible for driving the subgroup dispersion out of control. For these three subgroups, where the subgroup dispersion of individual variables differs extremely from standard values, our procedure diagnoses the same

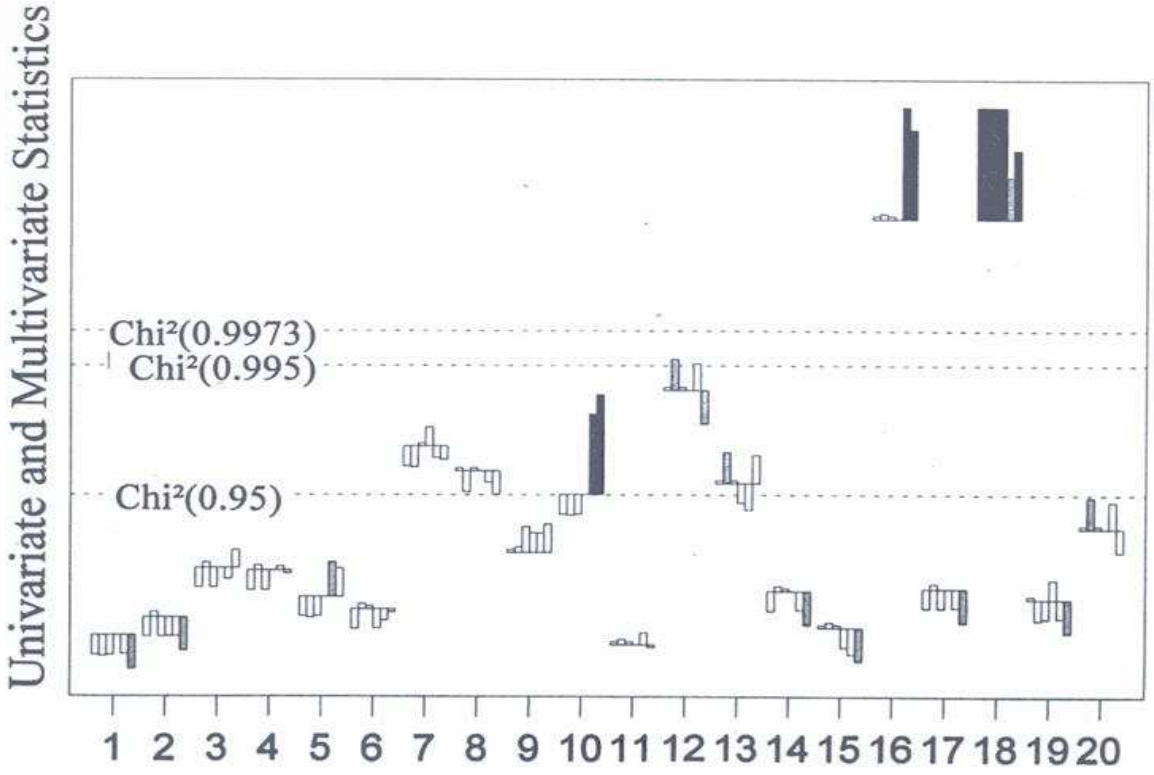

Figure 6.4: MP Dispersion Charts for the 20 Phase II Subgroups

variables as the MP chart for dispersion.

We verified that for subgroups 10, 16, and 18, the same causal variables identified by both procedures were responsible by adjusting one or more of their values closer to the subgroup mean before re-calculating T^2_{Dk}. This resulted in T^2_{Dk} values well below the critical values for all reasonable α levels.

For subgroups 7, 8, 12 and 13 there are no individual variables in the MP chart for dispersion whose scaled deviations differ extremely from standard values. Nonetheless the miniature bar-charts suggest first investigating those variables of higher scaled deviation. To contrast our PC-based diagnostic procedure for subgroup dispersion with the MP dispersion charts, we restrict our discussion to subgroup 8. The T^2_{D8} value of subgroup 8 is significant at $\alpha = 0.05$ with a critical value of 12.59.

The MP dispersion chart of subgroup 8 suggests causality from variables 2 and 6 due to their larger magnitudes of scaled deviation from standard values of dispersion. The diagnosis from our procedure presented in Figures 9 and 10 produces different results.

Figure 6.5 demonstrates how the sample variance of the sixth NSD is the most significant. Remember that since $n = 2$ in this example, the product of $(n-1) * S^2(NSD_{p,k,i})$ is simply the sample variance of the $NSD_{p,k,i}$. It also shows that the contributions of variables 1, 3 and 4 have the largest variation in decreasing order. To compare the accuracy of the two diagnoses, we modify the original subgroup 8 data to reduce the differences between the identified variables and their subgroup means. Since variables 1 and 3 are diagnosed by our procedure, we halve the distance between them and their subgroup means in one of subgroup 8's observations. This reduces the T^2_{D8} value from 13.67 to 4.69, well below the critical value of 12.59.

The MP chart for dispersion points us toward variables 2 and 6. From Figure 10 we see that the standard deviation of variable 2 in subgroup 8 is zero, which explains why in the MP chart for dispersion, the scaled deviation for this variable is a large negative value. The large negative value means that in this subgroup, variable 6's deviation is much lower than its standard value. Although this makes variable 6 a non-factor in driving T^2_{D8} out of control, it may indicate another problem since zero sample variance is somewhat unusual.

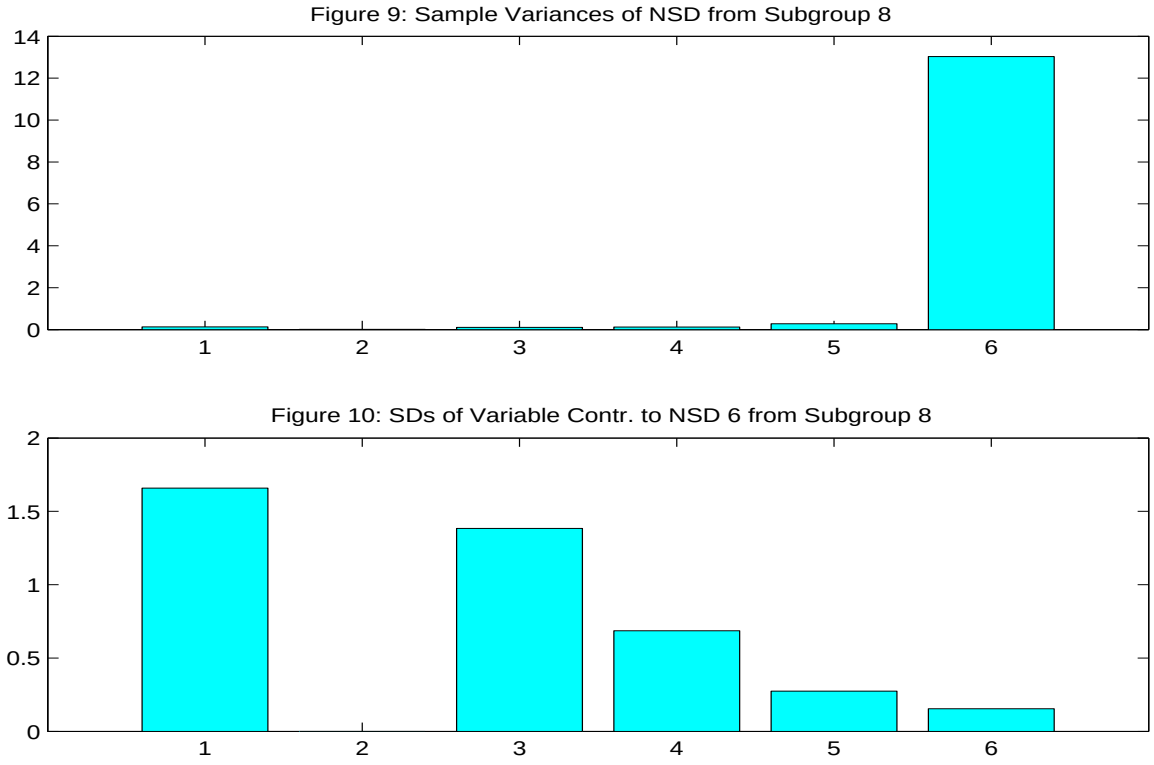

Figure 6.5: Sum of Squared NSD and SD(VarContr) to NSD6 in Subgroup 8

Halving the distance between the values of variable 6 and their subgroup mean reduces the T^2_{D8} value from 13.67 to 13.25, still above the critical value of 12.59. This indicates that for this specific subgroup, the PC-based procedure for diagnosing subgroup dispersion is more accurate.

6.5 Concluding Remarks

There are several reliable techniques for identifying the individual variables responsible for driving the T^2 value of a multivariate process out-of-control. Less research has been published regarding the diagnosis of subgroups of multivariate observations, which are prone to shifts of scale as well as location.

The MP charts of Fuchs and Benjamini (1994) are a helpful graphical instrument for investigating the potentially causal variables of rational subgroups whose location and or dispersion have shifted away from standard values. In their discussion, Fuchs and Kenett (1998) indicate that the scaled deviations of the MP dispersion chart must be interpreted with the correlational structure of the variables for accurate diagnosis. In contrast, our proposed procedure is directly related to the decomposition of T^2_{Dk}, which integrates the correlation structure of the individual variables.

We have extended the PC-based technique described in Kourti and MacGregor (1996) to diagnose the causal variables of a particular subgroup's shift in dispersion. This technique is simply implemented and in some cases more efficient than the corresponding MP chart owing to its incorporation of the correlation structure of the quality characteristics being monitored.

Kourti and MacGregor (1996) also note that once the T^2 statistic has gone out of control, a deviation at the chosen level of significance has already been detected. This obviates the need for precise control limits on the univariate charts of the normalized scores which are now merely guides for pinpointing the causal variables. For the same reason the diagnostic technique proposed here remains a viable method of identifying causal variables in MSPC.

CHAPTER VII

LSPC IN MULTIVARIATE SPC

In this chapter we propose a multivariate control chart based on LSPC that combines the diagnostic and dimension-reducing strengths of PCs with the directionally dependent economic weighting of the MQL. In the first section we describe the test statistic, called the loss-scaled T^2 statistic, i.e., LST^2, and how it can be decomposed to independently monitor subgroup location and dispersion. In section 2 we show that the diagnosis of individual variables driving subgroup location and/or dispersion out of control can be accomplished by decomposing the LST^2 into normalized LSPC. In Section 3 we compare its ARL properties to those of Hotelling's T^2. In Section 4 we summarize the chapter.

7.1 An MSPC Chart based on LSPC

Mohebbi and Hayre (1989) proposed defining in-control and out-of-control areas before selecting the multivariate statistic to monitor a process. They demonstrated that these areas could be defined according to the economic costs of the different quality characteristics being off-target. They employed a loss function such that the ARL properties were related to a loss-based non-centrality parameter defined as

$$LNCP \;=\; (E\mathbf{Y}-\tau)^T\mathbf{A}(E\mathbf{Y}-\tau)\;,$$

where $\mathbf{A}$ is the loss constant matrix of the MQL. It is interesting to note that this $LNCP$ is equal to the off-target vector product of the expected MQL examined in Chapter 4.

Tsui and Woodall (1993) employed a version of the MEWMA chart (Lowry et al (1992)) which substituted $\mathbf{A}$ for the inverse covariance matrix to create the test statistic. This combination of the loss function with the MEWMA allows faster detection of shifts in the directions of higher economic priority than the test statistic of Mohebbi and Hayre (1989). The ARL performance of this test statistic is related to the same $LNCP$ as that of Mohebbi and Hayre (1989).

We propose a multivariate SPC based on LSPCs with the following properties:

- a test statistic combining the loss constant matrix (LCM) and Hotelling's T^2 which we call Loss-Scaled T^2 (LST^2),
- the LST^2 decomposes into independent LSPC,
- applying the extension of Kourti and MacGregor (1996) presented in Chapter 6 allows the LSPC to identify the specific quality characteristics driving the process location and/or dispersion out-of-control regardless of physical interpretation,
- for certain combinations of correlation structure and LCM the LST^2 detects shifts in location faster than Hotelling's T^2.

7.1.1 How the LST^2-based Chart Works

We propose the following test statistic for a single multivariate observation $\mathbf{Y}$, which we call the Loss-Scaled T^2 statistic (LST^2):

$$LST^2 \quad = \quad [\mathbf{q^T}(\mathbf{Y}-\tau)]^T\hat{\Sigma}^{-1}[\mathbf{q^T}(\mathbf{Y}-\tau)]$$

where

$$\mathbf{A} \quad = \quad \mathbf{qq^T} \ ,$$

where τ is the vector of target values, $\hat{\Sigma}^{-1}$ is the inverse covariance matrix of the multivariate observations and $\mathbf{A}$ is the loss constant matrix (LCM) defined as part of the expected MQL in Equation 4.0. Not knowing the distribution of this statistic, we use the percentile values of large sample simulations as control limits.

In Chapter 6 we discussed a decomposition of Hotelling's T^2 into two independent parts representing location and dispersion. Consider a rational subgroup k consisting of the n individual multivariate data-points $\mathbf{Y}_i$ where $i = 1 \ldots n$. The overall subgroup LST^2 for subgroup k, i.e., LST^2_{Ok}, is the sum of the LST^2 of the n individual multivariate data-points that make up subgroup k. For the LST^2_{Ok} we also use the percentile values of large sample simulations to determine control limits.

We mimic Jackson (1981)'s decomposition of Hotelling's T^2_{Ok} for LST_{Ok}

$$LST^2_{Ok} \quad = \quad LST^2_{Mk} + LST^2_{Dk} \ ,$$

where LST^2_{Mk} tests whether the loss-scaled sample mean of this subgroup has shifted away from its standard value and is defined as

$$LST^2_{Mk} \quad = \quad n[\mathbf{q^T}(\bar{\mathbf{Y}}_k - \hat{\mu})]^T \hat{\Sigma}^{-1} [\mathbf{q^T}(\bar{\mathbf{Y}}_k - \hat{\mu})],$$

where $\bar{\mathbf{Y}}_k$ is the vector whose components are the subgroup averages of each quality characteristic. The percentile values of large sample simulations are used as control limits for LST^2_{Mk}.

LST^2_{Dk} tests whether the loss-scaled variance within this subgroup is significantly greater than its standard value and is defined as

$$LST^2_{Dk} \quad = \quad \Sigma^n_{i=1} [\mathbf{q^T}(\mathbf{Y}_{ki} - \bar{\mathbf{Y}}_k)]^T \hat{\Sigma}^{-1} [\mathbf{q^T}(\mathbf{Y}_{ki} - \bar{\mathbf{Y}}_k)].$$

where the $\mathbf{Y}_{ki}$ are the $i = 1 \ldots n$ observations in subgroup k. The percentile values of large sample simulations are used as control limits for LST^2_{Dk}.

Jackson (1985) argues that T^2_O is of little diagnostic value since when significant, the analyst must immediately ascertain whether it is T^2_M, the multivariate analog of the $\bar{\mathbf{x}}$ chart, or T^2_D, the multivariate analog of the r chart, that is responsible. We suspect that a similar relationship exists between LST^2_O, LST^2_M and LST^2_D. We therefore recommend simultaneous monitoring of LST^2_M and LST^2_D in order to expedite diagnosis, which we discuss in the next section.

7.2 Diagnosing with Normalized LSPC

In section 2 of Chapter 6 we extended the technique of Kourti and MacGregor (1996) to diagnose the individual variables driving either T^2_M or T^2_D out of control. The same procedure can be directly applied to the LST^2_M and LST^2_D with the drawback that unlike normalized PCs, normalized LSPC do not have a standard normal distribution.

Since we can not take critical values from a given statistical distribution, we use the percentiles from large sample simulations. However we can still apply the procedure proposed in Chapter 6 by taking the LSPCs making the largest contributions to the LST_M^2 or LST_D^2 statistics and decomposing them into individual variable contributions. We will demonstrate that this allows us to correctly identify the individual variables responsible for driving either LST^2 statistic out of control in the case 1 data-set of Fuchs and Kenett (1998). The reader will remember we used this data-set in Chapter 6 to contrast our PC-based diagnostic procedure with the MP charts of Fuchs and Benjamini (1994).

Figure 6.1 showed the T_M^2 and T_D^2 values of the last twenty subgroups of case 1 of Fuchs and Kenett (1998) based on the covariance matrix of the first fifteen subgroups. The critical value of T_M^2 for $\alpha = 0.01$ equals 51.74 and is calculated as follows,

$$UCL = \frac{p(k+1)(n-1)}{k(n-1)-p+1} F_{p,kn-k-p+1,0.01}$$

where $p = 6$ is the dimensionality of the data, $k = 15$ is the number of subgroups used to estimate population mean and covariance, $n = 2$ is the number of multivariate observations per subgroup and $\alpha = 0.01$. The critical value T_D^2 for $\alpha = 0.01$ equals 16.81, taken from the $\chi^2_{6(2-1)}$ distribution.

Figure 7.1 plots the T_M^2 and T_D^2 for the first 17 of the Phase II subgroups from case 1 of Fuchs and Kenett (1998). The lines of triangles indicate the critical values at $\alpha = 0.01$.

The first step in considering the use of the LST_M^2 and LST_D^2 is to choose a specific loss constant matrix (LCM). We used the following LCM because it offers diverse weighting in both diagonal and non-diagonal positions: A3 =

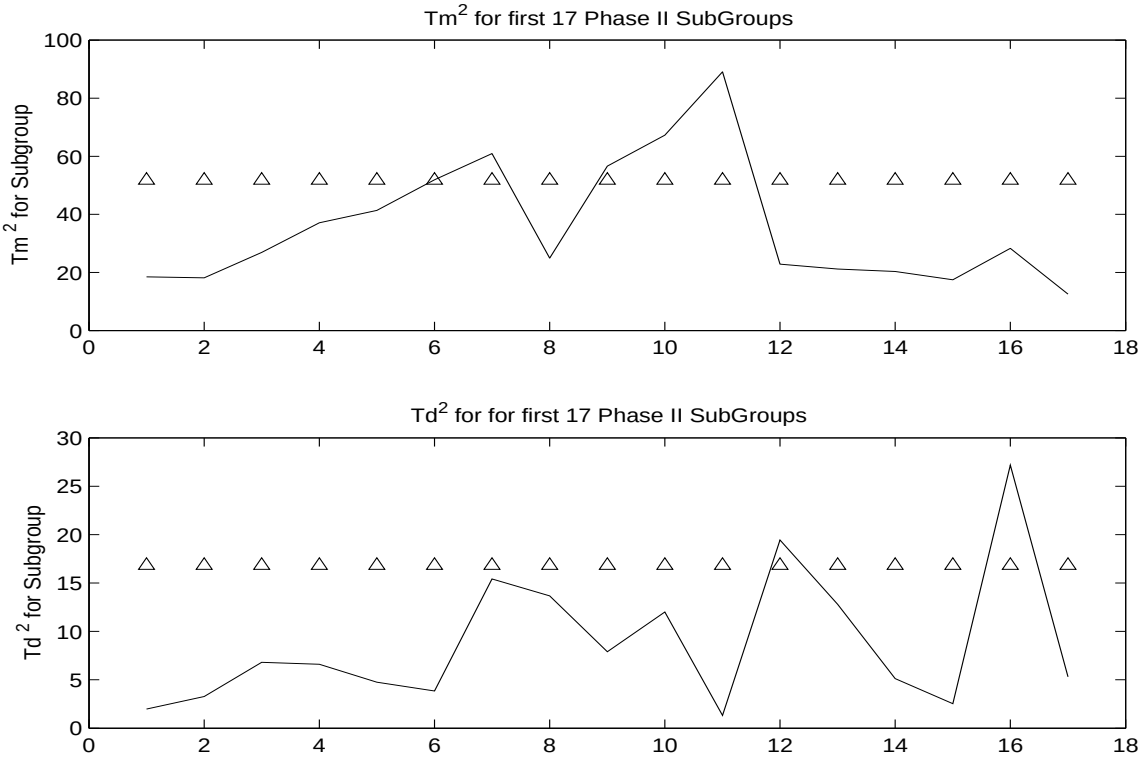

Figure 7.1: T^2_{Mk} and T^2_{Dk} for first 17 Phase II Subgroups

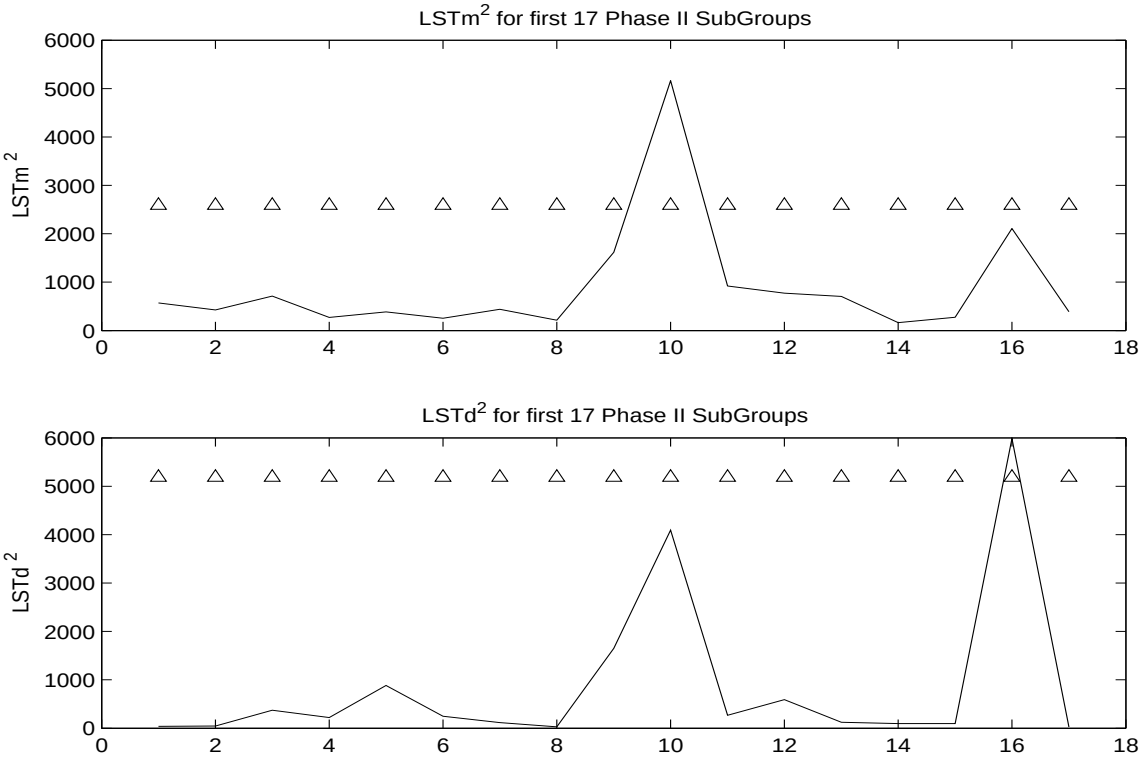

Figure 7.2: LST_M^2 and LST_D^2 for first 17 Phase II Subgroups

$$\begin{pmatrix} 2 & 2 & 0 & 1 & 1 & 1 \\ 2 & 9 & 1 & 0 & 1 & 1 \\ 0 & 1 & 4 & 1 & 1 & 1 \\ 1 & 0 & 1 & 4 & 1 & 1 \\ 1 & 1 & 1 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

In Figure 7.2 we plot the LST_M^2 and LST_D^2 statistics for the same subgroups of data whose T_M^2 and T_D^2 values are displayed in Figure 7.1. These LST_M^2 and LST_D^2 values are calculated using A3 as the LCM.

We will focus on subgroup 10 for location since this is the only subgroup with both T_M^2 and LST_M^2 statistics out of control. From Figures 7.1 and 7.2 we note that the T_M^2 for subgroup 10 is out of control and the LST_M^2 of subgroup 10 is the most out of control of all subgroups. Subgroup 10 therefore allows a direct comparison between

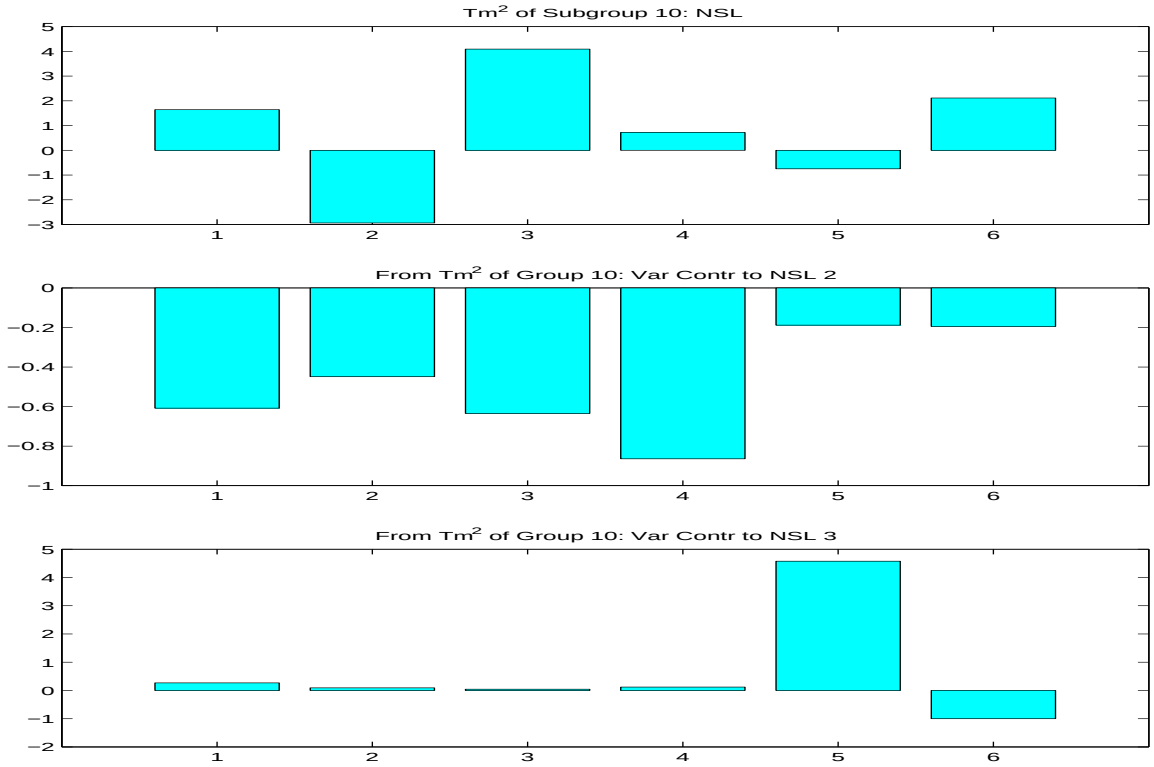

Figure 7.3: T^2_M of Subgroup 10: NSL and and Var. Contr. to NSLs 2 and 3

the diagnostic results of applying the procedure of Chapter 6 to the T^2_M and LST^2_M statistics. It is interesting to note that the T^2_M for subgroup 11 is the most out of control of all subgroups, whereas the LST^2_M of subgroup 11 is not out of control. This large difference in the relative LST^2_M signals produced by subgroups 10 and 11 is due to the weighting of the variables in the LCM.

Figure 7.3 shows three bar-charts related to the six normalized scores of location that make up the T^2_M of subgroup 10. The first bar-chart shows that normalized scores 3 and 2 are the most significant since they exceed a magnitude of 3. The second and third bar-charts shows that variable 5 makes like-signed contributions to the significant normalized scores.

This suggests that variable 5 is driving the location of subgroup 10 out of control. We verified this by changing the values of variable 5 to halve the distance between their subgroup mean and their calibration mean which decreased the T^2_M value from 67.3

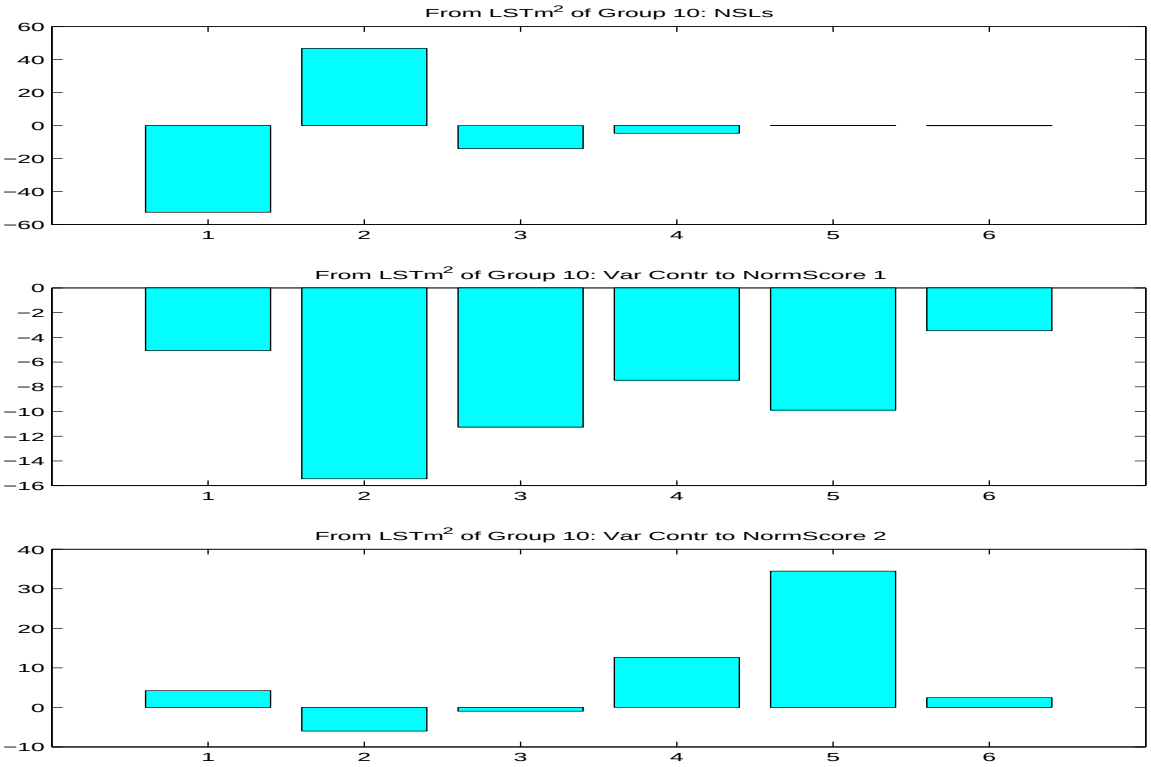

Figure 7.4: LST_M^2 of Subgroup 10: NSL and and Var. Contr. to NSLs 1 and 2

to 12.25, well below the critical value of 51.75.

Now we apply the diagnostic procedure of Chapter 6 to the LST_M^2 value of subgroup 10. Figure 7.4 shows three bar-charts related to the six NSL that make up the LST_M^2 of subgroup 10. The first bar-chart shows the relative magnitudes of the normalized LSPC and the second and third bar-charts show the individual variable contributions to the two largest of the normalized LSPC.

In Figure 7.4 we see that all six variables make same sign contributions to $NSL1$, with variable 2 being the largest and variables 3 and 5 close behind. We also see that variable 5 makes same sign contributions to both NSLs 1 and 2. By adjusting the values of variable 5 to create a subgroup average half the distance from its calibration mean we lower subgroup 10's LST_M^2 value from 5164 to 3053, still well above the critical value of 2586.54. By performing comparable individual adjustments to variables 2 and 3 we respectively reduce subgroup 10's LST_M^2 values to 4584 and 4496. Therefore

variable 5 is still the single most important variable for bringing subgroup 10's LST_M^2 value in control.

By adjusting variable 5 so that subgroup mean is very close to the standard value, subgroup 10's LST_M^2 value is reduced to 1866, well below the critical value of 2586.54. Making similar adjustments to variables 2 and 3 bring subgroup 10's LST_M^2 value respectively to 4351 and 4112.

The diagnosis of the T_M^2 value of subgroup 10 showed that a moderate adjustment of only variable 5 (i.e., half distance) brought the subgroup back into control with respect to location. To bring subgroup 10's LST_M^2 value into control, variable 5 has to be modified so that its subgroup mean is nearly equal to the historical mean. However, bringing the subgroup means of variables 2 and 5 half way closer to their historical means reduces the LST_M^2 value to 2414, below the critical value of 2586.54. The difference here is that the loss constant matrix A3 places highest weight on variable 2 which has the value 9 in the second diagonal position and the second highest weight of 4 on variable 5 in the loss constant matrix. Bringing the subgroup means of variables 3 and 5 half way closer to their historical means reduces the LST_M^2 value to 2487, also below the critical value of 2586.54.

We will focus on subgroup 16 for dispersion, which allows a direct comparison between the diagnostic results of T_D^2 and LST_D^2, since both statistics are out of control with respect to their critical values. Figure 7.5 contains two bar-charts, the first displays the sum of squares of the six normalized scores of dispersion that make up the T_D^2 of subgroup 16. Since the sum of squared NSD follow a $\chi^2_{(n-1)}$ distribution and $n = 2$, the critical value for $\alpha = 0.01$ is 6.63. Hence only $NSD1$ is significant. The second bar-chart shows the standard deviations of individual variable contributions to $NSD1$. Note that the standard deviations of contributions to $NSD1$ from variables 5 and 6 are largest in decreasing order. By adjusting the values of variables 5 and 6 in one of subgroup 16's observations closer to their subgroup averages, the value of T_{Dk}^2 decreased from 27.19 to 11.92, well below the critical value of 16.81.

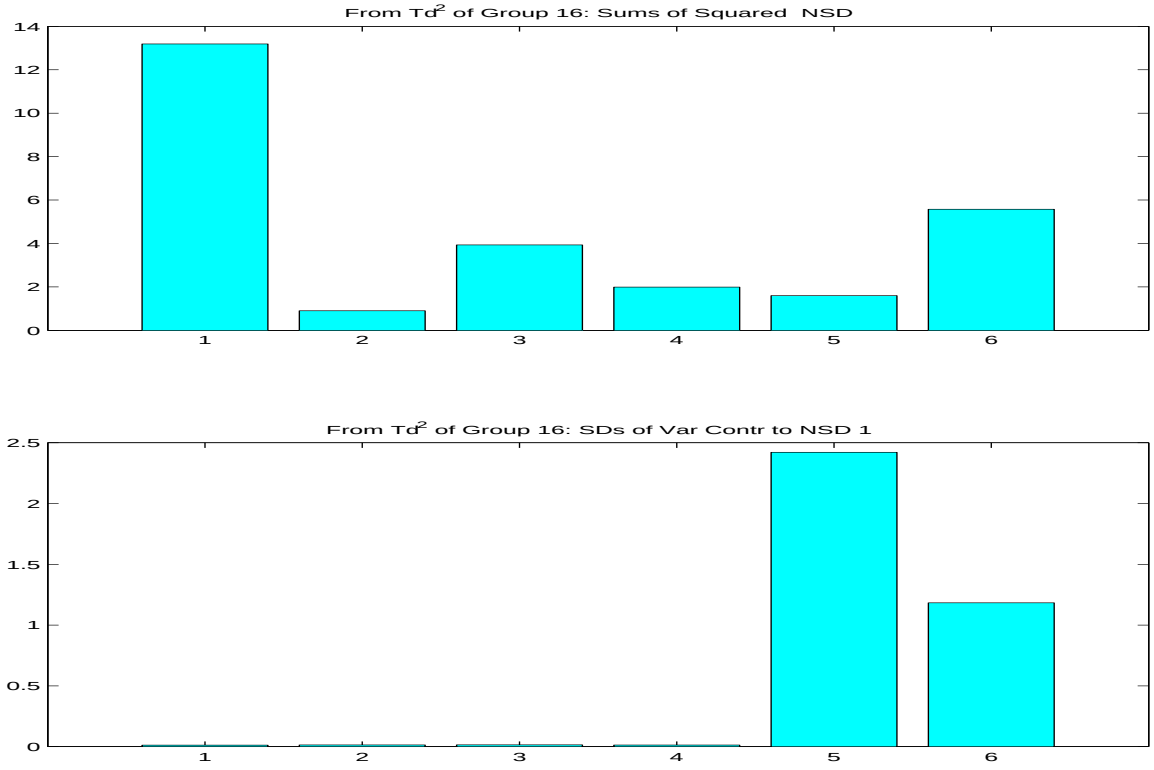

Figure 7.5: T_D^2 of Subgroup 16: Sum of Squared NSD and SD(VarContr) to NSD1

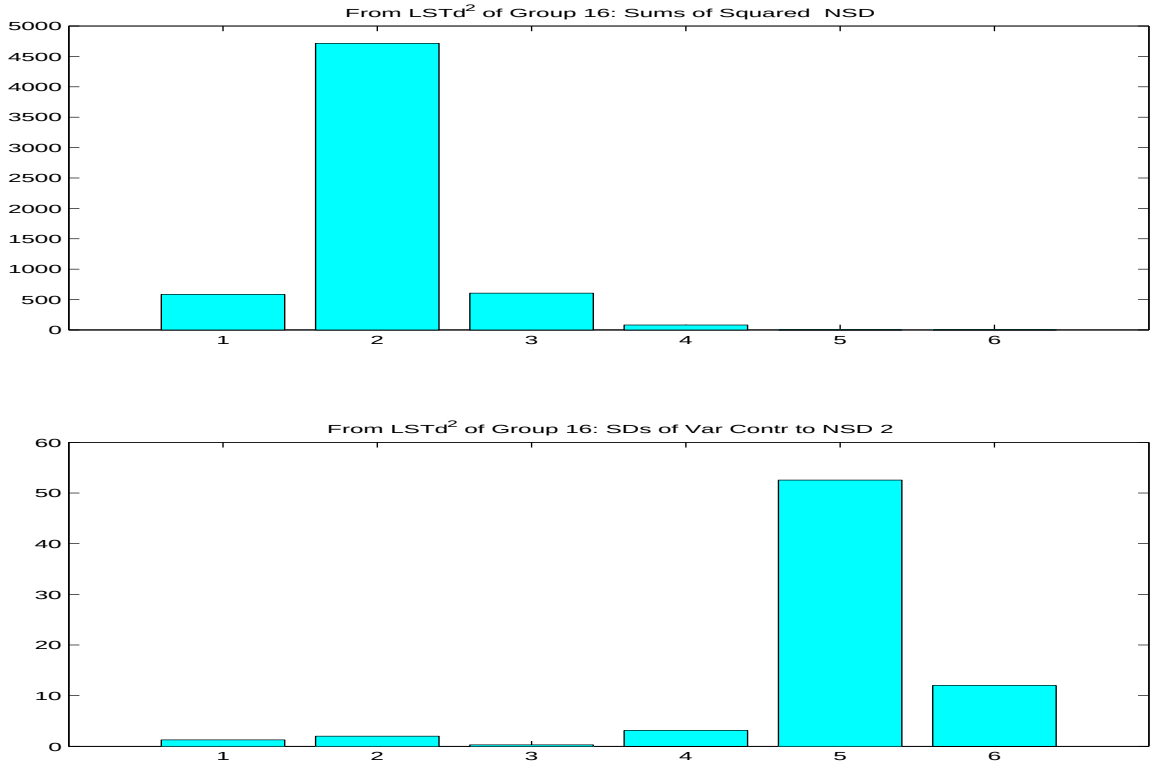

Figure 7.6: LST_D^2 of Subgroup 16: Diagnosis with NSD

We now diagnose the LST_D^2 value of subgroup 16. Figure 7.6 has two bar-charts related to the LST_D^2 of subgroup 16. The first bar-chart graphs the sum of squares of the NSD and the second depicts the standard deviations of the individual variable contributions to $NSD2$, which has the largest sum of squared values in subgroup 16. Figure 7.6 suggests that variable five is responsible for driving the LST_D^2 of subgroup 16 out of control. We verified this by halving the distance between the values of variable 5 and their subgroup mean, which reduced the LST_D^2 of subgroup 16 from 5984 to 1277, well under the critical value of 5184. This is the same variable identified as driving the T_D^2 of subgroup 16 out of control.

In summary, we've shown that for this particular case the PC-based diagnostic procedure introduced in Chapter 6 accurately diagnoses the variables driving the respective LST_M^2 and LST_D^2 statistics out of control. We demonstrated this by comparing the diagnostic results with those yielded by analyzing the T_M^2 and T_D^2 statistics from the

same data.

7.3 Comparing LST^2 and Hotelling's T^2

We compare the average-run-length (ARL) properties of the LST_M^2 and LST_D^2 statistics by comparing them to those of Hotelling's T^2. We do this with simiulated $MVN(\mathbf{0}, \mathbf{I}_6)$ data-sets and induce shifts in location and dispersion.

After many trials we detected that the ARL performance of the LST_M^2 and LST_D^2 statistics depends on the data's correlation structure, the relative magnitudes of individual variances of the data and the weighting of the values in the LCM.

In each case we chose upper control limits corresponding approximately to an $\alpha = 0.001$ since this level consistently provided a reasonable range of ARL values for the various statistics.

We illustrate the potential of our approach with the results in Tables 7.1 and 7.2. Table 7.1 compares the ARL performance of our approach with that of Hotelling's T^2 for detecting shifts in location, using simulated $MVN(\mathbf{0}, \mathbf{I}_6)$ data, which possess the following favorable conditions for faster detection with LST_M^2:

1. no cross correlation between the individual variables,
2. individual variables with equal variance,
3. dramatically different diagonal weighting in the Loss-Constant Matrix (LCM) and
4. off-diagonal terms equal to zero in the LCM.

The following LCM was used in combination with the simulated $MVN(\mathbf{0}, \mathbf{I}_6)$ data-sets,

A5 =

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 9 \end{pmatrix}.$$

Notice that this LCM lightly weights variables 1, 2 and 3 at the same value of 1. Variables 4, 5 and 6 are heavily weighted with the same value of 9, a dramatic increase over the weighting of the first three variables. This specific hierarchy of weighting will be important when we discuss the results in Tables 7.1 and 7.2 which follow immediately.

Table 7.1 compares the average run length (ARL) performance of our approach with that of Hotelling's T^2 for detecting shifts in location of $+1/2$, $+1$ and $+2$ standard deviations for the six individual variables from simulated $MVN(\mathbf{0}, \mathbf{I_6})$ data. The simulations consisted of 400 cycles of 10k subgroups of $n = 5$ each at an $\alpha = 0.001$ for both tables of results. The number in parentheses after ARL is the standard error of the estimate of ARL.

Note that LST_M^2 detects shifts in location faster for $v4 - v6$ while T_M^2 detects shifts in location faster for $v1 - v3$, in accordance with their weighting in the LCM. This simulated example was picked because of the data's correlation structure and its easily interpreted weighting values within the LCM. We make the following high-level observations regarding the results in Table 7.1:

1. Compared to Hotelling's T_M^2, LST_M^2 demonstrates a modest increase in speed of detection of the location shifts for the heavily weighted variables, i.e., $v4, v5$ and $v6$.
2. Compared to LST_M^2, Hotelling's T_M^2 displays much faster speed of detection of the location shifts for the lightly weighted variables, i.e., $v1, v2$ and $v3$.

3. Neither T_D^2 nor LST_D^2 show a large change in speed of detection for the shifts of location examined.

Table 7.2 compares the average run length (ARL) performance of the LST_D^2 with that of Hotelling's T_D^2 for detecting the doubling of variance of individual variables from simulated $MVN(\mathbf{0}, \mathbf{I_6})$ data. The simulations consisted of 400 cycles of 10k subgroups of $n = 5$ each at an $\alpha = 0.001$ for all results. The numbers in parentheses are the standard errors of each estimated ARL.

We make the following high level observations regarding the results in Table 7.2:

1. Compared to Hotelling's T_D^2, LST_D^2 demonstrates a modest increase in speed of detection of the doubled variance for the heavily weighted variables, i.e., $v4, v5$ and $v6$.
2. T_M^2's speed of detection for the heavily weighted variables is the same as of LST_D^2.
3. Compared to LST_D^2, Hotelling's T_D^2 displays much faster speed of detection of the doubled variance for the lightly weighted variables, i.e., $v1, v2$ and $v3$.
4. For the heavily weighted variables, LST_M^2 shows a very small increase in speed of detection of the doubled variance compared to T_M^2.
5. For the lightly weighted variables T_M^2 has the fastest speed of detection.
6. The statistics designed to detect location, i.e., T_M^2 and LST_M^2, consistently detect the doubled variance of individual variables faster than their dispersion related counterparts, i.e., T_D^2 and LST_D^2.
7. As expected for this $MVN(\mathbf{0}, \mathbf{I_6})$ data, the ARL of the T_M^2 and T_D^2 statistics are the same for each individual variable across all shifts of location and variance.

7.4 Comments Regarding LST^2 in MSPC

In this chapter we've introduced a new multivariate test statistic which incorporates the economic weighting of the MQL into Hotelling's T^2 statistic. In section 2 we introduced a LSPC-based diagnostic procedure which identifies the individual variables

Table 7.1: Average Run Length (s.e.) for Shifts in Location

Ind. Var.	**Loc. Shift**	T^2_M	LST^2_M	T^2_D	LST^2_D
$v1$	+1/2s.d.	285.9(2.6)	1063.4(22)	1096(31.2)	1181(28)
	+1s.d.	33.8(0.1)	855.5(15)	1044(21)	1144.8(22.6)
	+2s.d.	2.1(0.002)	386(4.2)	893.9(15)	1106.9(24)
$v2$	+1/2s.d.	285(3.0)	1060(25)	1095.9(24)	1171.3(25)
	+1s.d.	33.8(0.1)	860.2(14)	1074.3(23)	1136(24)
	+2s.d.	2.1(0.001)	381.3(4)	894.7(17)	1076.3(19)
$v3$	+1/2s.d.	282.5(2.5)	1044.4(18)	1070.6(21)	1154.1(22)
	+1s.d.	33.8(0.1)	853.3(15)	1104.4(25.3)	1170.1(23.5)
	+2s.d.	2.1(0.001)	377.8(3.9)	887.6(14)	1095(21)
$v4$	+1/2s.d.	281.3(2.6)	166.4(1.2)	1118.9(33.4)	1124.9(27)
	+1s.d.	33.9(0.1)	17.7(0.0)	1049(23)	1064.4(25)
	+2s.d.	2.1(0.002)	1.5(0.001)	894(15)	874.6(17.6)
$v5$	+1/2s.d.	277.6(2.3)	165.7(1)	1116.7(24)	1097.6(22)
	+1s.d.	33.7(0.1)	17.6(0.0)	1057.7(20)	1051.1(19)
	+2s.d.	2.1(0.001)	1.6(0.004)	891(15)	866.6(14)
$v6$	+1/2s.d.	275.8(2.3)	196.3(1.3)	1017.3(18)	1060.7(21)
	+1s.d.	27.2(0.1)	17.4(0.04)	817.3(14)	910.2(15)
	+2s.d.	1.6(0.001)	1.4(0.0005)	357.2(4)	518.7(6.3)

Table 7.2: Average Run Length (s.e.) for 2x Shifts in Variance

Individual Variables $\Downarrow$	T_D^2	LST_D^2	T_M^2	LST_M^2
v1	103.7(0.6)	1086(28)	49.8(0.2)	195.3(1.4)
v2	104.2(0.6)	1121 (25)	49.8(0.2)	193.2(1.3)
v3	102.9(0.6)	1033(18)	50(0.2)	193.7(1.4)
v4	102.9 (0.6)	49.4(0.2)	49.8(0.2)	37.6(0.1)
v5	104.5(0.5)	49.8(0.2)	49.8(0.2)	37.5(0.1)
v6	104.5 (0.6)	49.4 (0.2)	49.9(0.2)	37.3(0.1)

responsible for driving a subgroup's location and/or dispersion out of control with respect to the LST^2 statistic. Since this technique is a direct decomposition of the corresponding LST_M^2 or LST_D^2 statistic, it exploits the correlation structure of the data as well as the loss constant matrix. It therefore provides economic prioritization of individual variables while identifying those responsible for shifts in subgroup location and or dispersion.

We believe the coupling of faster detection of specific variables of higher economic priority with the ability to identify which variables drive the monitoring statistics out of control make the LST_M^2 and LST_D^2 statistics attractive alternatives for certain combinations of data-sets and loss constant matrices.

7.5 Future Research Directions

In this section we outline a number of potential areas to extend and expand the work started in this thesis. We break these down into ideas for extending work in multivariate robust design and multivariate SPC as well as areas of personal interest not related to this thesis.

7.5.1 Future Research in Multivariate Robust Design

The following is a list of some of the avenues of investigation that have surfaced during the work documented in this thesis.

1. We will run the electric vehicle chassis design problem with *IDEAS* including the noise replicates, which will take approximately 20 of continuous running for a final comparison with the results obtained using meta-models of responses and LSPC.
2. We want to examine a combination of using the meta-models of the LSPC examined in Chapter 5 with some experimental design techniques that seek to identify the most promising area in the design space for building meta-models. This may allow a reduction in the time needed for design and modelling in early stages of

the design cycle. This will entail combining some elements of this thesis with that of Yao Lin of the SRL.

3. We wish to find a problem with many more design variables than the six in our electric vehicle chassis design, to show whether the meta-models of LSPC can greatly reduce the computational time of high dimensional problems in robust design.
4. Examine whether LSPC can be used to simplify dynamic MRD.
5. Examine how different modelling techniques (e.g., MARS, Kriging, Wavelets, etc.) and objective functions (e.g., expected MQL) interact when choosing the optimal design vector.
6. Investigate other new areas of multivariate robust design such as multivariate steepest descent and multivariate signal-to-noise techniques.

7.5.2 Future Research in Multivariate SPC

The following are a few questions that have piqued our curiosity while working on this thesis.

1. Are there other specific combinations of data type and loss constant matrix for which the LST_M^2 consistently detects shifts in subgroup location and dispersion faster than T_M^2 without a huge loss in detection for the lightly weighted variables? The advantages of detecting heavily weighted variables faster in Chapter 7 is accompanied by a large loss of sensitivity in the lightly weighted variables that may seriously undermine its usefulness.
2. Can the PC-based diagnostic procedure introduced in Chapter 6 be integrated into automatic process control so as to automatically diagnose and correct the individual variables responsible for driving multivariate location and/or dispersion out of control?
3. Consider how to account for auto-correlation in MSPC when using PCs with or without a loss function. Some ideas in the literature involve calculating a set of

eigenvectors, i.e., the coefficients of principal components, that optimally describe the variance across a range of autocorrelation matrices. This work employs the work of Flury (1988) to find the eigenvectors that optimize the variance across multiple covariance matrices.

7.5.3 Future Research Areas of Interest

The following areas are things have interested me during my study of statistics during the past ten years. Any new feature or application for these well established disciplines hold research potential for me since there is so much to learn.

1. Modelling including use of the generalized linear model (GLM), non-linear modelling techniques, wavelets and others.
2. Multivariate statistics in general seem more practical since most real life situations involve many factors that are inter-related.
3. Biostatistics interests me because I have undergraduate training in biomedical engineering, chemistry and biology.
4. Applications of experimental design. I have a good grounding in fractional factorials and some experience with space filling designs, but enjoyed studying nested, split-plot and other types of experimental design which I've not yet had occasion to use.
5. Data mining is interesting to me for the endless ways in which it can be used to enhance decision making based on large sets of data.

APPENDIX A

PROOFS OF EQUATIONS 4.-31 AND 6.-35

A.1 Proof of Equation 4.-31

Equation (4.-31) represents the following fraction,

$$\begin{aligned}\frac{(E\mathbf{Zp}-\tau_{\mathbf{Zp}})^T(E\mathbf{Zp}-\tau_{\mathbf{Zp}})}{(E\mathbf{Zp}-\tau_{\mathbf{Zp}})^T(E\mathbf{Zp}-\tau_{\mathbf{Zp}})} &= \frac{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T\mathbf{VpVp^T}[\mathbf{q^T}(E\mathbf{Y}-\tau)]}{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T\mathbf{VV^T}[\mathbf{q^T}(E\mathbf{Y}-\tau)]} \\ &= \frac{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T\mathbf{VpVp}^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]}{[\mathbf{q^T}(E\mathbf{Y}-\tau)]^T[\mathbf{q^T}(E\mathbf{Y}-\tau)]}\end{aligned}$$

The numerical value of this fraction depends on the specific off-target vector i.e., $(E\mathbf{Y}-\tau)$ being evaluated. However the numerator and denominators of this fraction can be decomposed into the sums of coefficients of the underlying basis vectors. We first re-express the matrix product $\mathbf{VpVp^T}$ as a function of its underlying bases. The matrix $\mathbf{Vp}$ consists of the column vectors $\mathbf{V_1},\ldots,\mathbf{V_p}$:

$$\mathbf{Vp} = \begin{pmatrix} \mathbf{V_1} & \ldots & \mathbf{V_p} \end{pmatrix}$$

and because each column vector of $\mathbf{Vp}$ is orthogonal,

$$\mathbf{Vp^TVp} = \begin{pmatrix} \mathbf{V_1^T} \\ \vdots \\ \mathbf{V_p^T} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{V_1} & \ldots & \mathbf{V_p} \end{pmatrix} = \mathbf{I} \text{ of dimension } p$$

However the matrix product of $\mathbf{VpVp^T}$ does not equal the identity matrix when $p < r$:

$$\mathbf{VpVp^T} = \begin{pmatrix} \mathbf{V_1} & \ldots & \mathbf{V_p} \end{pmatrix} \begin{pmatrix} \mathbf{V_1^T} \\ \vdots \\ \mathbf{V_p^T} \end{pmatrix}$$

and when the matrix product $\mathbf{VpVp^T}$ is multiplied by a response vector $\mathbf{Y}$:

$$\begin{aligned}
\mathbf{VpVp^TY} &= \begin{pmatrix} \mathbf{V_1} & \dots & \mathbf{V_p} \end{pmatrix} \begin{pmatrix} \mathbf{V_1^T} \\ \vdots \\ \mathbf{V_p^T} \end{pmatrix} \mathbf{Y} \\
&= \begin{pmatrix} \mathbf{V_1} & \dots & \mathbf{V_p} \end{pmatrix} \begin{pmatrix} \mathbf{V_1^TY} \\ \vdots \\ \mathbf{V_p^TY} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{V_1}(\mathbf{V_1^TY}) \\ \vdots \\ \mathbf{V_p}(\mathbf{V_p^TY}) \end{pmatrix}
\end{aligned}$$

we recognize the orthogonal projection of $\mathbf{Y}$, an r dimensional vector, onto the span $\{\mathbf{V_1}, \dots, \mathbf{V_p}\}$. This directly implies that the matrix product $\mathbf{VpVp^T}$ is diagonalized in the following form:

$$\mathbf{VpVp^T} = \mathbf{V\Lambda V^T}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of dimension r whose first p eigenvalues are equal to one and whose remaining $(r-p)$ eigenvalues are equal to zero. The vectors $\mathbf{V_1}$ through $\mathbf{V_p}$ are eigenvectors with eigenvalue one and the vectors $\mathbf{V_{p+1}}$ through $\mathbf{V_r}$ are eigenvectors with eigenvalue zero.

This specific diagonal form allows us to write the numerator and denominator of Equation (4.-43) as the sums of specific terms.

The loss-scaled off-target vector $\mathbf{q^T}(E\mathbf{Y} - \tau)$ is defined over an r dimensional space. It can then be expressed as a linear combination of the r basis vectors that define the space.

$$\mathbf{q^T}(E\mathbf{Y} - \tau) = c_1\mathbf{w_1} + \dots + c_r\mathbf{w_r}$$

where the c_i are the coefficients and the $\mathbf{w_i}$ are orthonormal basis vectors. We have shown that the matrix product $\mathbf{VpVp^T}$ has r eigenvalues and eigenvectors such

that:

$$\mathbf{VpVp^T w_i} \;=\; \lambda_i \mathbf{w_i}$$

where all λ_i equal zero or one and

$$\mathbf{w_i^T w_k} = \begin{cases} 1 & \text{if } i = k \ , \\ 0 & \text{if } i \neq k \ . \end{cases}$$

The denominator of Equation (4.-43) is simply the Euclidean norm of $\mathbf{q^T}(E\mathbf{Y} - \tau)$

$$\begin{aligned} [\mathbf{q^T}(E\mathbf{Y} - \tau)]^T[\mathbf{q^T}(E\mathbf{Y} - \tau)] &= \| \mathbf{q^T}(E\mathbf{Y} - \tau) \|^2 \\ &= (c_1\mathbf{w_1} + \cdots + c_r\mathbf{w_r})^T(c_1\mathbf{w_1} + \cdots + c_r\mathbf{w_r}) \\ &= c_1^2 + \cdots + c_r^2 \end{aligned}$$

The numerator of Equation (4.-43) can be written as the product of $[\mathbf{q^T}(E\mathbf{Y} - \tau)]^T$ and $\mathbf{VpVp^Tq^T}(E\mathbf{Y} - \tau)$. We first write out $\mathbf{VpVp^Tq^T}(E\mathbf{Y} - \tau)$ as:

$$\begin{aligned} \mathbf{VpVp^Tq^T}(E\mathbf{Y} - \tau) &= c_1\mathbf{VpVp^Tw_1} + \cdots + c_r\mathbf{VpVp^Tw_r} \\ &= c_1\lambda_1\mathbf{w_1} + \cdots + c_r\lambda_r\mathbf{w_r} \end{aligned}$$

and we then write out $[\mathbf{q^T}(E\mathbf{Y} - \tau)]^T\mathbf{VpVp^Tq^T}(E\mathbf{Y} - \tau)$ as:

$$\begin{aligned} [\mathbf{q^T}(E\mathbf{Y} - \tau)]^T\mathbf{VpVp^Tq^T}(E\mathbf{Y} - \tau) &= (c_1\mathbf{w_1} + \cdots + c_r\mathbf{w_r})^T(c_1\lambda_1\mathbf{w_1} + \cdots + c_r\lambda_r\mathbf{w_r}) \\ &= c_1^2\lambda_1 + \cdots + c_r^2\lambda_r \\ &= c_1^2 + \cdots + c_p^2 \end{aligned}$$

Equation (4.-43) can now be written as:

$$\frac{[\mathbf{q^T}(E\mathbf{Y} - \tau)]^T\mathbf{VpVp}^T[\mathbf{q^T}(E\mathbf{Y} - \tau)]}{[\mathbf{q^T}(E\mathbf{Y} - \tau)]^T[\mathbf{q^T}(E\mathbf{Y} - \tau)]} \;=\; \frac{c_1^2 + \cdots + c_p^2}{c_1^2 + \cdots + c_r^2}$$

A.2 Outline of Proof of Equation 6.-35

It is well known that for any normally distributed random variable V, i.e.,

$$V \;\sim\; N(\mu, \sigma^2) \ ,$$

that

$$\frac{(n-1)S^2(V)}{\sigma^2} \quad \sim \quad \chi^2_{(n-1)} \ .$$

where $S^2(V)$ is the sample variance of V calculated from a random sample of size n. In equation 6.-31 we stated that the $NSD_{p,k,i}$ are approximately standard normal. Substituting $NSD_{p,k,i}$ for V yields,

$$\frac{(n-1)S^2(NSD_{p,k,i})}{\sigma^2_{NSDp,k,i}} \quad \sim \quad \chi^2_{(n-1)} \ , \tag{A.-24}$$

where $S^2(NSD_{p,k,i})$ can be simply expressed as

$$S^2(NSD_{p,k,i}) \quad = \quad \frac{\Sigma_{i=1}^{n} \left[\frac{\mathbf{u}_p^T(\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)}{\lambda_p^{1/2}} \right]^2}{(n-1)} \ , \tag{A.-23}$$

since the sample average of the $\frac{\mathbf{u}_p^T(\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)}{\lambda_p^{1/2}}$ terms is zero. The term $(n-1)S^2(NSD_{p,k,i})$ is equal to the sum of the squared $NSD_{p,k,i}$, i.e., $\Sigma_{i=1}^{n} \left[\frac{\mathbf{u}_p^T(\mathbf{Y}_{k,i} - \bar{\mathbf{Y}}_k)}{\lambda_p^{1/2}} \right]^2$.

APPENDIX B

OTVP FROM 2 LSPC IN SAMPLE DATA-SETS

We have shown in Appendix A that the proportion of the off-target vector product (OTVP) accounted for by a subset of LSPC falls between 0 and 1. However the exact proportion depends entirely on the specific off-target vector i.e., $(E\mathbf{Y} - \tau)$ being evaluated. The tables in this Appendix represent the verification of how well the a subset of LSPC represent the OTVP over a range of cases.

In order to provide a sense of how well the OTVP is approximated by a subset of LSPC, the following tables show a group of 20 off-target vectors and what proportion of their OTVP is accounted for by the first two LSPC of a variety of data-sets and loss constant matrices.

The data sets ranged in dimension from 3 to 6 and are taken from JW92PH. In the following tables, the abbreviation JWp283 means the data comes from page 283 of this reference, etc. The six dimensional loss constant matrices (LCM) employed are presented on the following pages. The smaller dimensional data sets used the corresponding sub-matrices of the LCMs listed.

Loss Constant Matrices for Dimension 6

$$AI = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A1 = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}$$

$$A2 = \begin{pmatrix} 1 & 1 & 2 & 0 & 1 & 1 \\ 1 & 5 & 1 & 1 & 1 & 1 \\ 2 & 1 & 9 & 0 & 1 & 1 \\ 0 & 1 & 0 & 4 & 0 & 1 \\ 1 & 1 & 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 1 & 1 & 4 \end{pmatrix}$$

$$A3 = \begin{pmatrix} 2 & 2 & 0 & 1 & 1 & 1 \\ 2 & 9 & 1 & 0 & 1 & 1 \\ 0 & 1 & 4 & 1 & 1 & 1 \\ 1 & 0 & 1 & 4 & 1 & 1 \\ 1 & 1 & 1 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Additional Loss Constant Matrices for Dimension 6

$$A4 = \begin{pmatrix} 2 & 2 & 0 & 1 & 2 & 1 \\ 2 & 9 & 1 & 0 & 2 & 1 \\ 0 & 1 & 9 & 2 & 2 & 1 \\ 1 & 0 & 2 & 9 & 2 & 1 \\ 2 & 2 & 2 & 2 & 9 & 2 \\ 1 & 1 & 1 & 1 & 2 & 1 \end{pmatrix}$$

$$A5 = \begin{pmatrix} 1 & 2 & 0 & 1 & 1 & 1 \\ 2 & 9 & 1 & 0 & 2 & 1 \\ 0 & 1 & 9 & 2 & 2 & 2 \\ 1 & 0 & 2 & 9 & 2 & 1 \\ 1 & 2 & 2 & 2 & 9 & 2 \\ 1 & 1 & 2 & 1 & 2 & 1 \end{pmatrix}$$

Table B.1: Proportion of OTVP from 2 LSPC: JWp183 and LCM A1

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5]$	0.9026
$[0, 0.5, 0]$	0.8396
$[0.5, 0, 0]$	0.7065
$[0.5, 0, 0.5]$	0.9026
$[0, 0, 1]$	0.7065
$[0, 1, 0]$	0.7569
$[1, 0, 0]$	0.9929
$[1, 0, 1]$	0.9026
$[0, 0.5, 1]$	0.9658
$[0, 1, 0.5]$	0.7217
$[1, 0, 0.5]$	0.8995
$[1, 0.5, 1]$	0.9026
$[0, 0, 2]$	0.8396
$[0, 2, 0]$	0.7065
$[2, 0, 0]$	0.8995
$[2, 1, 2]$	0.9929
$[0, 1, 2]$	0.9658
$[0, 2, 1]$	0.7217
$[2, 0, 1]$	0.9026
$[2, 2, 2]$	0.9657
Avg. OTVP Explained =	0.8493

Table B.2: Proportion of OTVP from 2 LSPC: JWp183 and LCM A2

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5]$	0.9855
$[0, 0.5, 0]$	0.9924
$[0.5, 0, 0]$	0.8673
$[0.5, 0, 0.5]$	0.9047
$[0, 0, 1]$	0.9855
$[0, 1, 0]$	0.9924
$[1, 0, 0]$	0.8674
$[1, 0, 1]$	0.9047
$[0, 0.5, 1]$	0.9984
$[0, 1, 0.5]$	0.9978
$[1, 0, 0.5]$	0.8653
$[1, 0.5, 1]$	0.9694
$[0, 0, 2]$	0.9855
$[0, 2, 0]$	0.9924
$[2, 0, 0]$	0.8673
$[2, 1, 2]$	0.9694
$[0, 1, 2]$	0.9984
$[0, 2, 1]$	0.9978
$[2, 0, 1]$	0.8653
$[2, 2, 2]$	0.9904
Avg. OTVP Explained =	0.9498

Table B.3: Proportion of OTVP from 2 LSPC: JWp183 and LCM A3

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5]$	0.9086
$[0, 0.5, 0]$	0.9751
$[0.5, 0, 0]$	0.4306
$[0.5, 0, 0.5]$	0.5342
$[0, 0, 1]$	0.9086
$[0, 1, 0]$	0.9751
$[1, 0, 0]$	0.4306
$[1, 0, 1]$	0.5342
$[0, 0.5, 1]$	0.9784
$[0, 1, 0.5]$	0.9971
$[1, 0, 0.5]$	0.3749
$[1, 0.5, 1]$	0.7992
$[0, 0, 2]$	0.9086
$[0, 2, 0]$	0.9751
$[2, 0, 0]$	0.4306
$[2, 1, 2]$	0.7992
$[0, 1, 2]$	0.9784
$[0, 2, 1]$	0.9971
$[2, 0, 1]$	0.3749
$[2, 2, 2]$	0.9245
Avg. OTVP Explained =	0.7617

Table B.4: Proportion of OTVP from 2 LSPC: JWp183 and LCM A4

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5]$	0.9697
$[0, 0.5, 0]$	0.9494
$[0.5, 0, 0]$	0.7308
$[0.5, 0, 0.5]$	0.8049
$[0, 0, 1]$	0.9697
$[0, 1, 0]$	0.9494
$[1, 0, 0]$	0.7308
$[1, 0, 1]$	0.8049
$[0, 0.5, 1]$	0.997
$[0, 1, 0.5]$	0.9848
$[1, 0, 0.5]$	0.7253
$[1, 0.5, 1]$	0.9184
$[0, 0, 2]$	0.9697
$[0, 2, 0]$	0.9494
$[2, 0, 0]$	0.7308
$[2, 1, 2]$	0.9184
$[0, 1, 2]$	0.9970
$[0, 2, 1]$	0.9848
$[2, 0, 1]$	0.7253
$[2, 2, 2]$	0.9715
Avg. OTVP Explained =	0.8891

Table B.5: Proportion of OTVP from 2 LSPC: JWp183 and LCM A5

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5]$	0.9710
$[0, 0.5, 0]$	0.9913
$[0.5, 0, 0]$	0.5944
$[0.5, 0, 0.5]$	0.8090
$[0, 0, 1]$	0.9710
$[0, 1, 0]$	0.9913
$[1, 0, 0]$	0.5944
$[1, 0, 1]$	0.8090
$[0, 0.5, 1]$	0.9945
$[0, 1, 0.5]$	0.9989
$[1, 0, 0.5]$	0.6744
$[1, 0.5, 1]$	0.9316
$[0, 0, 2]$	0.9710
$[0, 2, 0]$	0.9913
$[2, 0, 0]$	0.5944
$[2, 1, 2]$	0.9316
$[0, 1, 2]$	0.9945
$[0, 2, 1]$	0.9989
$[2, 0, 1]$	0.6744
$[2, 2, 2]$	0.9757
Avg. OTVP Explained =	0.8731

Table B.6: Proportion of OTVP from 2 LSPC: JWp279 and LCM A1

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.9116

Table B.7: Proportion of OTVP from 2 LSPC: JWp279 and LCM A2

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.8848

Table B.8: Proportion of OTVP from 2 LSPC: JWp279 and LCM A3

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.8760

Table B.9: Proportion of OTVP from 2 LSPC: JWp279 and LCM A4

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.8915

Table B.10: Proportion of OTVP from 2 LSPC: JWp279 and LCM A5

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.8938

Table B.11: Proportion of OTVP from 2 LSPC: JWp279 and LCM A6

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.7130

Table B.12: Proportion of OTVP from 2 LSPC: JWp279 and LCM A7

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.7434

Table B.13: Proportion of OTVP from 2 LSPC: JWp279 and LCM A8

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.7485

Table B.14: Proportion of OTVP from 2 LSPC: JWp279 and LCM A9

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.7454

Table B.15: Proportion of OTVP from 2 LSPC: JWp279 and $\mathbf{I}_6$

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7476
$[0, 0.5, 0, 0.5]$	0.8245
$[0.5, 0, 0, 0.5]$	0.8366
$[0.5, 0, 0.5, 0.5]$	0.9664
$[0, 0, 1, 1]$	0.9986
$[0, 1, 0, 1]$	0.8245
$[1, 0, 0, 1]$	0.8366
$[1, 0, 1, 0.5]$	0.9353
$[0, 0.5, 1, 1]$	0.9779
$[0, 1, 0.5, 1]$	0.9175
$[1, 1, 0.5, 1]$	0.9906
$[1, 1, 1, 1]$	0.9998
$[0, 0, 2, 1]$	0.9623
$[0, 2, 0, 1]$	0.8182
$[2, 0, 0, 2]$	0.8366
$[2, 1, 2, 0]$	0.9310
$[0, 1, 2, 2]$	0.9779
$[0, 2, 1, 2]$	0.9175
$[2, 0, 1, 2]$	0.9335
$[2, 2, 2, 2]$	0.9998
Avg. OTVP Explained =	0.7305

Table B.16: Proportion of OTVP from 2 LSPC: JWp283 and LCM A1

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1, 0.5]$	0.8912
$[0, 0.5, 0, 0.5, 1, 0.5]$	0.9271
$[0.5, 0, 0, 0.5, 1, 0.5]$	0.9591
$[0.5, 0, 0.5, 0.5, 1, 0.5]$	0.9766
$[1, 1, 1, 1, 2, 0]$	0.9667
$[0, 1, 0, 1, 2, 1]$	0.9271
$[1, 0, 0, 0, 2, 1]$	0.9115
$[1, 0, 1, 0.5, 2, 1]$	0.9665
$[0, 0.5, 1, 1, 2, 1]$	0.9418
$[0, 1, 0.5, 1, 1, 2]$	0.9563
$[1, 1, 0.5, 1, 1, 1]$	0.9998
$[1, 0, 1, 1, 1, 0]$	0.9895
$[0, 0, 2, 1, 2, 2]$	0.9566
$[0, 2, 0, 1, 2, 2]$	0.9610
$[2, 2, 1, 2, 0, 0]$	0.9636
$[2, 1, 2, 0, 2, 2]$	0.9919
$[0, 1, 2, 2, 2, 2]$	0.9667
$[0, 2, 0, 2, 1, 1]$	0.9714
$[2, 0, 1, 2, 1, 1]$	0.9970
$[2, 2, 2, 0, 1, 0]$	0.9527
Avg. OTVP Explained =	0.9587

Table B.17: Proportion of OTVP from 2 LSPC: JWp283 and LCM A2

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1, 0.5]$	0.8912
$[0, 0.5, 0, 0.5, 1, 0.5]$	0.9271
$[0.5, 0, 0, 0.5, 1, 0.5]$	0.9591
$[0.5, 0, 0.5, 0.5, 1, 0.5]$	0.9766
$[1, 1, 1, 1, 2, 0]$	0.9667
$[0, 1, 0, 1, 2, 1]$	0.9271
$[1, 0, 0, 0, 2, 1]$	0.9115
$[1, 0, 1, 0.5, 2, 1]$	0.9665
$[0, 0.5, 1, 1, 2, 1]$	0.9418
$[0, 1, 0.5, 1, 1, 2]$	0.9563
$[1, 1, 0.5, 1, 1, 1]$	0.9998
$[1, 0, 1, 1, 1, 0]$	0.9895
$[0, 0, 2, 1, 2, 2]$	0.9566
$[0, 2, 0, 1, 2, 2]$	0.9610
$[2, 2, 1, 2, 0, 0]$	0.9636
$[2, 1, 2, 0, 2, 2]$	0.9919
$[0, 1, 2, 2, 2, 2]$	0.9667
$[0, 2, 0, 2, 1, 1]$	0.9714
$[2, 0, 1, 2, 1, 1]$	0.9970
$[2, 2, 2, 0, 1, 0]$	0.9527
Avg. OTVP Explained =	0.9282

Table B.18: Proportion of OTVP from 2 LSPC: JWp283 and LCM A3

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1, 0.5]$	0.8912
$[0, 0.5, 0, 0.5, 1, 0.5]$	0.9271
$[0.5, 0, 0, 0.5, 1, 0.5]$	0.9591
$[0.5, 0, 0.5, 0.5, 1, 0.5]$	0.9766
$[1, 1, 1, 1, 2, 0]$	0.9667
$[0, 1, 0, 1, 2, 1]$	0.9271
$[1, 0, 0, 0, 2, 1]$	0.9115
$[1, 0, 1, 0.5, 2, 1]$	0.9665
$[0, 0.5, 1, 1, 2, 1]$	0.9418
$[0, 1, 0.5, 1, 1, 2]$	0.9563
$[1, 1, 0.5, 1, 1, 1]$	0.9998
$[1, 0, 1, 1, 1, 0]$	0.9895
$[0, 0, 2, 1, 2, 2]$	0.9566
$[0, 2, 0, 1, 2, 2]$	0.9610
$[2, 2, 1, 2, 0, 0]$	0.9636
$[2, 1, 2, 0, 2, 2]$	0.9919
$[0, 1, 2, 2, 2, 2]$	0.9667
$[0, 2, 0, 2, 1, 1]$	0.9714
$[2, 0, 1, 2, 1, 1]$	0.9970
$[2, 2, 2, 0, 1, 0]$	0.9527
Avg. OTVP Explained =	0.9102

Table B.19: Proportion of OTVP from 2 LSPC: JWp283 and LCM A4

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1, 0.5]$	0.8912
$[0, 0.5, 0, 0.5, 1, 0.5]$	0.9271
$[0.5, 0, 0, 0.5, 1, 0.5]$	0.9591
$[0.5, 0, 0.5, 0.5, 1, 0.5]$	0.9766
$[1, 1, 1, 1, 2, 0]$	0.9667
$[0, 1, 0, 1, 2, 1]$	0.9271
$[1, 0, 0, 0, 2, 1]$	0.9115
$[1, 0, 1, 0.5, 2, 1]$	0.9665
$[0, 0.5, 1, 1, 2, 1]$	0.9418
$[0, 1, 0.5, 1, 1, 2]$	0.9563
$[1, 1, 0.5, 1, 1, 1]$	0.9998
$[1, 0, 1, 1, 1, 0]$	0.9895
$[0, 0, 2, 1, 2, 2]$	0.9566
$[0, 2, 0, 1, 2, 2]$	0.9610
$[2, 2, 1, 2, 0, 0]$	0.9636
$[2, 1, 2, 0, 2, 2]$	0.9919
$[0, 1, 2, 2, 2, 2]$	0.9667
$[0, 2, 0, 2, 1, 1]$	0.9714
$[2, 0, 1, 2, 1, 1]$	0.9970
$[2, 2, 2, 0, 1, 0]$	0.9527
Avg. OTVP Explained =	0.8925

Table B.20: Proportion of OTVP from 2 LSPC: JWp283 and LCM A5

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1, 0.5]$	0.8912
$[0, 0.5, 0, 0.5, 1, 0.5]$	0.9271
$[0.5, 0, 0, 0.5, 1, 0.5]$	0.9591
$[0.5, 0, 0.5, 0.5, 1, 0.5]$	0.9766
$[1, 1, 1, 1, 2, 0]$	0.9667
$[0, 1, 0, 1, 2, 1]$	0.9271
$[1, 0, 0, 0, 2, 1]$	0.9115
$[1, 0, 1, 0.5, 2, 1]$	0.9665
$[0, 0.5, 1, 1, 2, 1]$	0.9418
$[0, 1, 0.5, 1, 1, 2]$	0.9563
$[1, 1, 0.5, 1, 1, 1]$	0.9998
$[1, 0, 1, 1, 1, 0]$	0.9895
$[0, 0, 2, 1, 2, 2]$	0.9566
$[0, 2, 0, 1, 2, 2]$	0.9610
$[2, 2, 1, 2, 0, 0]$	0.9636
$[2, 1, 2, 0, 2, 2]$	0.9919
$[0, 1, 2, 2, 2, 2]$	0.9667
$[0, 2, 0, 2, 1, 1]$	0.9714
$[2, 0, 1, 2, 1, 1]$	0.9970
$[2, 2, 2, 0, 1, 0]$	0.9527
Avg. OTVP Explained =	0.8864

Table B.21: Proportion of OTVP from 2 LSPC: JWp283 and LCM I

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1, 0.5]$	0.8912
$[0, 0.5, 0, 0.5, 1, 0.5]$	0.9271
$[0.5, 0, 0, 0.5, 1, 0.5]$	0.9591
$[0.5, 0, 0.5, 0.5, 1, 0.5]$	0.9766
$[1, 1, 1, 1, 2, 0]$	0.9667
$[0, 1, 0, 1, 2, 1]$	0.9271
$[1, 0, 0, 0, 2, 1]$	0.9115
$[1, 0, 1, 0.5, 2, 1]$	0.9665
$[0, 0.5, 1, 1, 2, 1]$	0.9418
$[0, 1, 0.5, 1, 1, 2]$	0.9563
$[1, 1, 0.5, 1, 1, 1]$	0.9998
$[1, 0, 1, 1, 1, 0]$	0.9895
$[0, 0, 2, 1, 2, 2]$	0.9566
$[0, 2, 0, 1, 2, 2]$	0.9610
$[2, 2, 1, 2, 0, 0]$	0.9636
$[2, 1, 2, 0, 2, 2]$	0.9919
$[0, 1, 2, 2, 2, 2]$	0.9667
$[0, 2, 0, 2, 1, 1]$	0.9714
$[2, 0, 1, 2, 1, 1]$	0.9970
$[2, 2, 2, 0, 1, 0]$	0.9527
Avg. OTVP Explained =	0.8086

Table B.22: Proportion of OTVP from 2 LSPC: JWp354 and LCM A1

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7300
$[0, 0.5, 0, 0.5]$	0.8878
$[0.5, 0, 0, 0.5]$	0.8976
$[0.5, 0, 0.5, 0.5]$	0.9004
$[0, 0, 1, 1]$	0.8758
$[0, 1, 0, 1]$	0.8878
$[1, 0, 0, 1]$	0.8976
$[1, 0, 1, 0.5]$	0.8773
$[0, 0.5, 1, 1]$	0.9724
$[0, 1, 0.5, 1]$	0.9747
$[1, 1, 0.5, 1]$	0.9984
$[1, 1, 1, 1]$	0.9940
$[0, 0, 2, 1]$	0.8686
$[0, 2, 0, 1]$	0.8698
$[2, 0, 0, 2]$	0.8976
$[2, 1, 2, 0]$	0.8943
$[0, 1, 2, 2]$	0.9724
$[0, 2, 1, 2]$	0.9747
$[2, 0, 1, 2]$	0.9094
$[2, 2, 2, 2]$	0.9940
Avg. OTVP Explained =	0.9137

Table B.23: Proportion of OTVP from 2 LSPC: JWp354 and LCM A2

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.9765
$[0, 0.5, 0, 0.5]$	0.9967
$[0.5, 0, 0, 0.5]$	0.6244
$[0.5, 0, 0.5, 0.5]$	0.8194
$[0, 0, 1, 1]$	0.7481
$[0, 1, 0, 1]$	0.9967
$[1, 0, 0, 1]$	0.6244
$[1, 0, 1, 0.5]$	0.9019
$[0, 0.5, 1, 1]$	0.9071
$[0, 1, 0.5, 1]$	0.9876
$[1, 1, 0.5, 1]$	0.9692
$[1, 1, 1, 1]$	0.9724
$[0, 0, 2, 1]$	0.8713
$[0, 2, 0, 1]$	0.9657
$[2, 0, 0, 2]$	0.6244
$[2, 1, 2, 0]$	0.9643
$[0, 1, 2, 2]$	0.9071
$[0, 2, 1, 2]$	0.9876
$[2, 0, 1, 2]$	0.7285
$[2, 2, 2, 2]$	0.9724
Avg. OTVP Explained =	0.8773

Table B.24: Proportion of OTVP from 2 LSPC: JWp354 and LCM A3

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7148
$[0, 0.5, 0, 0.5]$	0.7911
$[0.5, 0, 0, 0.5]$	0.3131
$[0.5, 0, 0.5, 0.5]$	0.7644
$[0, 0, 1, 1]$	0.9676
$[0, 1, 0, 1]$	0.7911
$[1, 0, 0, 1]$	0.3131
$[1, 0, 1, 0.5]$	0.7560
$[0, 0.5, 1, 1]$	0.9720
$[0, 1, 0.5, 1]$	0.9180
$[1, 1, 0.5, 1]$	0.8059
$[1, 1, 1, 1]$	0.8799
$[0, 0, 2, 1]$	0.9545
$[0, 2, 0, 1]$	0.9041
$[2, 0, 0, 2]$	0.3131
$[2, 1, 2, 0]$	0.8492
$[0, 1, 2, 2]$	0.9720
$[0, 2, 1, 2]$	0.9180
$[2, 0, 1, 2]$	0.5888
$[2, 2, 2, 2]$	0.8799
Avg. OTVP Explained =	0.7683

Table B.25: Proportion of OTVP from 2 LSPC: JWp354 and LCM A4

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7666
$[0, 0.5, 0, 0.5]$	0.6739
$[0.5, 0, 0, 0.5]$	0.3506
$[0.5, 0, 0.5, 0.5]$	0.8346
$[0, 0, 1, 1]$	0.9739
$[0, 1, 0, 1]$	0.6739
$[1, 0, 0, 1]$	0.3506
$[1, 0, 1, 0.5]$	0.8857
$[0, 0.5, 1, 1]$	0.9733
$[0, 1, 0.5, 1]$	0.8825
$[1, 1, 0.5, 1]$	0.7843
$[1, 1, 1, 1]$	0.8915
$[0, 0, 2, 1]$	0.9776
$[0, 2, 0, 1]$	0.8192
$[2, 0, 0, 2]$	0.3506
$[2, 1, 2, 0]$	0.9536
$[0, 1, 2, 2]$	0.9733
$[0, 2, 1, 2]$	0.8825
$[2, 0, 1, 2]$	0.6478
$[2, 2, 2, 2]$	0.8915
Avg. OTVP Explained =	0.7769

Table B.26: Proportion of OTVP from 2 LSPC: JWp354 and LCM A5

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.2300
$[0, 0.5, 0, 0.5]$	0.9336
$[0.5, 0, 0, 0.5]$	0.9026
$[0.5, 0, 0.5, 0.5]$	0.9221
$[0, 0, 1, 1]$	0.8872
$[0, 1, 0, 1]$	0.9336
$[1, 0, 0, 1]$	0.9026
$[1, 0, 1, 0.5]$	0.8015
$[0, 0.5, 1, 1]$	0.9471
$[0, 1, 0.5, 1]$	0.9977
$[1, 1, 0.5, 1]$	0.9715
$[1, 1, 1, 1]$	0.9828
$[0, 0, 2, 1]$	0.6782
$[0, 2, 0, 1]$	0.9336
$[2, 0, 0, 2]$	0.9026
$[2, 1, 2, 0]$	0.8468
$[0, 1, 2, 2]$	0.9471
$[0, 2, 1, 2]$	0.9977
$[2, 0, 1, 2]$	0.9632
$[2, 2, 2, 2]$	0.9828
Avg. OTVP Explained =	0.8832

Table B.27: Proportion of OTVP from 2 LSPC: JWp354 and LCM A6

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.1278
$[0, 0.5, 0, 0.5]$	0.6167
$[0.5, 0, 0, 0.5]$	0.4339
$[0.5, 0, 0.5, 0.5]$	0.1760
$[0, 0, 1, 1]$	0.1295
$[0, 1, 0, 1]$	0.6167
$[1, 0, 0, 1]$	0.4339
$[1, 0, 1, 0.5]$	0.2659
$[0, 0.5, 1, 1]$	0.3608
$[0, 1, 0.5, 1]$	0.6619
$[1, 1, 0.5, 1]$	0.6074
$[1, 1, 1, 1]$	0.5237
$[0, 0, 2, 1]$	0.1482
$[0, 2, 0, 1]$	0.8390
$[2, 0, 0, 2]$	0.4339
$[2, 1, 2, 0]$	0.5247
$[0, 1, 2, 2]$	0.3608
$[0, 2, 1, 2]$	0.6619
$[2, 0, 1, 2]$	0.2988
$[2, 2, 2, 2]$	0.5237
Avg. OTVP Explained =	0.4373

Table B.28: Proportion of OTVP from 2 LSPC: JWp354 and LCM A7

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.0155
$[0, 0.5, 0, 0.5]$	0.8933
$[0.5, 0, 0, 0.5]$	0.8567
$[0.5, 0, 0.5, 0.5]$	0.7222
$[0, 0, 1, 1]$	0.0143
$[0, 1, 0, 1]$	0.8933
$[1, 0, 0, 1]$	0.8567
$[1, 0, 1, 0.5]$	0.7892
$[0, 0.5, 1, 1]$	0.5717
$[0, 1, 0.5, 1]$	0.8897
$[1, 1, 0.5, 1]$	0.9076
$[1, 1, 1, 1]$	0.8671
$[0, 0, 2, 1]$	0.0171
$[0, 2, 0, 1]$	0.9675
$[2, 0, 0, 2]$	0.8567
$[2, 1, 2, 0]$	0.8744
$[0, 1, 2, 2]$	0.5717
$[0, 2, 1, 2]$	0.8897
$[2, 0, 1, 2]$	0.8051
$[2, 2, 2, 2]$	0.8671
Avg. OTVP Explained =	0.7063

Table B.29: Proportion of OTVP from 2 LSPC: JWp354 and LCM A8

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.7158
$[0, 0.5, 0, 0.5]$	0.5113
$[0.5, 0, 0, 0.5]$	0.4520
$[0.5, 0, 0.5, 0.5]$	0.4700
$[0, 0, 1, 1]$	0.4923
$[0, 1, 0, 1]$	0.5113
$[1, 0, 0, 1]$	0.4520
$[1, 0, 1, 0.5]$	0.6069
$[0, 0.5, 1, 1]$	0.7327
$[0, 1, 0.5, 1]$	0.7485
$[1, 1, 0.5, 1]$	0.7395
$[1, 1, 1, 1]$	0.7927
$[0, 0, 2, 1]$	0.6422
$[0, 2, 0, 1]$	0.6931
$[2, 0, 0, 2]$	0.4520
$[2, 1, 2, 0]$	0.8860
$[0, 1, 2, 2]$	0.7327
$[0, 2, 1, 2]$	0.7485
$[2, 0, 1, 2]$	0.4272
$[2, 2, 2, 2]$	0.7927
Avg. OTVP Explained =	0.6300

Table B.30: Proportion of OTVP from 2 LSPC: JWp354 and LCM A9

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.0175
$[0, 0.5, 0, 0.5]$	0.8936
$[0.5, 0, 0, 0.5]$	0.8567
$[0.5, 0, 0.5, 0.5]$	0.7221
$[0, 0, 1, 1]$	0.0157
$[0, 1, 0, 1]$	0.8936
$[1, 0, 0, 1]$	0.8567
$[1, 0, 1, 0.5]$	0.7891
$[0, 0.5, 1, 1]$	0.5736
$[0, 1, 0.5, 1]$	0.8905
$[1, 1, 0.5, 1]$	0.9167
$[1, 1, 1, 1]$	0.8798
$[0, 0, 2, 1]$	0.0190
$[0, 2, 0, 1]$	0.9676
$[2, 0, 0, 2]$	0.8567
$[2, 1, 2, 0]$	0.8841
$[0, 1, 2, 2]$	0.5736
$[0, 2, 1, 2]$	0.8905
$[2, 0, 1, 2]$	0.8049
$[2, 2, 2, 2]$	0.8798
Avg. OTVP Explained =	0.7091

Table B.31: Proportion of OTVP from 2 LSPC: JWp354 and LCM I

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0]$	0.9919
$[0, 0.5, 0, 0.5]$	0.7128
$[0.5, 0, 0, 0.5]$	0.5026
$[0.5, 0, 0.5, 0.5]$	0.7134
$[0, 0, 1, 1]$	0.4223
$[0, 1, 0, 1]$	0.7128
$[1, 0, 0, 1]$	0.5026
$[1, 0, 1, 0.5]$	0.8723
$[0, 0.5, 1, 1]$	0.6871
$[0, 1, 0.5, 1]$	0.7772
$[1, 1, 0.5, 1]$	0.7982
$[1, 1, 1, 1]$	0.8426
$[0, 0, 2, 1]$	0.7287
$[0, 2, 0, 1]$	0.9156
$[2, 0, 0, 2]$	0.5026
$[2, 1, 2, 0]$	0.9864
$[0, 1, 2, 2]$	0.6871
$[0, 2, 1, 2]$	0.7772
$[2, 0, 1, 2]$	0.6226
$[2, 2, 2, 2]$	0.8426
Avg. OTVP Explained =	0.7299

Table B.32: Proportion of OTVP from 2 LSPC: JWp623 and LCM A1

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1]$	0.7276
$[1, 0.5, 0, 0.5, 1]$	0.9073
$[0.5, 0, 2, 0.5, 1]$	0.9045
$[0.5, 0, 0.5, 0.5, 1]$	0.9146
$[0, 2, 2, 1, 2]$	0.9512
$[0, 1, 0.5, 1, 2]$	0.9522
$[1, 2, 0, 0, 2]$	0.8666
$[1, 0, 1, 0.5, 2]$	0.8802
$[0, 0.5, 2, 1, 2]$	0.9083
$[0, 1, 0.5, 1, 2]$	0.9522
$[1, 1, 0.5, 0, 1]$	0.9003
$[1, 0, 2, 1, 2]$	0.9273
$[0, 0, 2, 1, 2]$	0.8708
$[0, 2, 0, 1, 2]$	0.9700
$[2, 2, 1, 2, 0]$	0.9650
$[2, 1, 2, 0, 2]$	0.9158
$[0, 1, 2, 2, 2]$	0.9624
$[0, 2, 0, 2, 1]$	0.9888
$[2, 0, 1, 2, 1]$	0.9681
$[2, 2, 2, 0, 1]$	0.9230
Avg. OTVP Explained =	0.9178

Table B.33: Proportion of OTVP from 2 LSPC: JWp623 and LCM A2

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1]$	0.9417
$[1, 0.5, 0, 0.5, 1]$	0.8557
$[0.5, 0, 2, 0.5, 1]$	0.9988
$[0.5, 0, 0.5, 0.5, 1]$	0.9591
$[0, 2, 2, 1, 2]$	0.9547
$[0, 1, 0.5, 1, 2]$	0.9701
$[1, 2, 0, 0, 2]$	0.6781
$[1, 0, 1, 0.5, 2]$	0.9463
$[0, 0.5, 2, 1, 2]$	0.9982
$[0, 1, 0.5, 1, 2]$	0.9701
$[1, 1, 0.5, 0, 1]$	0.7824
$[1, 0, 2, 1, 2]$	0.9921
$[0, 0, 2, 1, 2]$	0.9978
$[0, 2, 0, 1, 2]$	0.9174
$[2, 2, 1, 2, 0]$	0.9958
$[2, 1, 2, 0, 2]$	0.8989
$[0, 1, 2, 2, 2]$	0.9974
$[0, 2, 0, 2, 1]$	0.9996
$[2, 0, 1, 2, 1]$	0.9179
$[2, 2, 2, 0, 1]$	0.8619
Avg. OTVP Explained =	0.9317

Table B.34: Proportion of OTVP from 2 LSPC: JWp623 and LCM A3

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1]$	0.7992
$[1, 0.5, 0, 0.5, 1]$	0.8335
$[0.5, 0, 2, 0.5, 1]$	0.9750
$[0.5, 0, 0.5, 0.5, 1]$	0.8871
$[0, 2, 2, 1, 2]$	0.9571
$[0, 1, 0.5, 1, 2]$	0.8926
$[1, 2, 0, 0, 2]$	0.9374
$[1, 0, 1, 0.5, 2]$	0.8768
$[0, 0.5, 2, 1, 2]$	0.9524
$[0, 1, 0.5, 1, 2]$	0.8926
$[1, 1, 0.5, 0, 1]$	0.9822
$[1, 0, 2, 1, 2]$	0.9777
$[0, 0, 2, 1, 2]$	0.9504
$[0, 2, 0, 1, 2]$	0.9309
$[2, 2, 1, 2, 0]$	0.9421
$[2, 1, 2, 0, 2]$	0.9909
$[0, 1, 2, 2, 2]$	0.9804
$[0, 2, 0, 2, 1]$	0.9938
$[2, 0, 1, 2, 1]$	0.8637
$[2, 2, 2, 0, 1]$	0.9883
Avg. OTVP Explained =	0.9302

Table B.35: Proportion of OTVP from 2 LSPC: JWp623 and LCM A4

Off-Target Vector $\Downarrow$	**Proportion of OTVP Explained by 2 LSPC**
$[0, 0, 0.5, 0, 1]$	0.6787
$[1, 0.5, 0, 0.5, 1]$	0.8673
$[0.5, 0, 2, 0.5, 1]$	0.9821
$[0.5, 0, 0.5, 0.5, 1]$	0.8372
$[0, 2, 2, 1, 2]$	0.9398
$[0, 1, 0.5, 1, 2]$	0.8866
$[1, 2, 0, 0, 2]$	0.8512
$[1, 0, 1, 0.5, 2]$	0.7877
$[0, 0.5, 2, 1, 2]$	0.9420
$[0, 1, 0.5, 1, 2]$	0.8866
$[1, 1, 0.5, 0, 1]$	0.9098
$[1, 0, 2, 1, 2]$	0.9249
$[0, 0, 2, 1, 2]$	0.9275
$[0, 2, 0, 1, 2]$	0.9175
$[2, 2, 1, 2, 0]$	0.9292
$[2, 1, 2, 0, 2]$	0.9143
$[0, 1, 2, 2, 2]$	0.9870
$[0, 2, 0, 2, 1]$	0.9979
$[2, 0, 1, 2, 1]$	0.8864
$[2, 2, 2, 0, 1]$	0.9451
Avg. OTVP Explained =	0.8999

APPENDIX C

HOW TO USE ISIGHT V6.0

iSIGHT is a generic software shell that improves the efficiency of the design process. Within this software platform, design problems can be defined and simulation codes from a variety of sources can be coupled within a description file (i.e., *.desc). After building the description file, iSIGHT can be used to monitor and analyze the design runs performed by a specific description file.

iSIGHT provides the user with capability in the following areas:

- Experimental Design including fractional factorials, central composite designs and space filling designs.
- Model fitting approaches including response surface models and Kriging.
- Optimization algorithms including gradient based methods, linear and non-linear programming, mixed integer, genetic algorithms and many more.
- A structure for linking simulation codes in many languages and from many commercial packages. These include the standard mechanical design packages such as *IDEAS*, *Pro-Engineer*, *ABSYS* and others. Languages easily integrated include Basic, Fortran, C, C-plus, Matlab, etc.
- Additional computations can be performed on the data during transfer of data between the simulation codes and iSIGHT.

  The iSIGHT graphical user interface is made of four modules that accommodate the varying parts of defining, building, measuring and analyzing design processes. These include:

  - Task Manager,
  - Process Integration,

- Problem Definition and
- Solution Monitor.

Appendix H of the official iSIGHT training manual iSIGHTbtg spells out the following steps as quick way to start using iSIGHT.

1. Process Integration
   - Set up data flow within the Process Integration module.
   - Add a simcode block.
     * Define path to simcode.
     * Define path to input and output files.
   - Write data into input file from iSIGHT.
   - Read data from output file to iSIGHT.
   - Verify correct data flow between input, simcode and output.
   - Save description file and exit Process Integration.
2. Problem Definition
   - From the **Parameters** dialog box:
     * Choose to maximize or minimize the objective functions,
     * Enter constraints and boundaries for inputs and outputs,
     * Enter initial values for all inputs.
   - From the **Task Plan** dialog box:
     * Define a plan,
     * Consult iSIGHT'S built in advisor if helpful,
     * Select one or more design exploration techniques,
     * Adjust settings.
   - From the **Database** dialog box:
     * Name the current database,
     * Select criteria for data to be stored in database,

        * Make selections regarding database lookup options,
        * Define lookup order if using an additional database,
        * Set tolerances.
    - From the **Rules** dialogue box, configure rules if needed (optional).
    - From the **Approximations** dialog box (optional):
        * Give model a name,
        * specific simcode for exact (vs. approximate) analysis,
        * Select model type,
        * Define an approximation model,
        * Select inputs,
        * Initialize model,
        * Determine and specify evaluation mode.
    - Reconfigure the Task Plan appropriately
    - Save description file.
3. Set output options with Solution Monitor
    - Add one or more pages with the appropriate number of cells. On each page:
        * Add one or more graphs,
        * For each graph select type of graph,
        * parameters to be displayed,
        * colors and symbols to be used,
        * location of graph on page.
    - Add one of more tables.
    - For each table select type.
    - For each table select parameters.
    - for each table select location on page.
    - Save graphics configuration template as a .cfg file.
    - Ensure entire page is selected in the left pane of the dialog box.

- To reuse configuration, select **Insert/Template** ... not **File/Open**...
- Split screen between Graphics and Data Browser, if desired
- Note that Data Browser displays only one thing during runtime
- Leave Solution Monitor dialog box open

4. Executing Task from within Task Manager
   - Click the **Execute** button,
   - Note that the gear icon changes to a running man during simcode execution,
   - Leave Task Manager open and return to Solution Monitor to observe data flow,
   - Examine log file for troubleshooting,
   - When design exploration session is complete, examine data in Solution Monitor,
   - Note tables and graphs for successful completion,
   - Load database from file in Solution Monitor's Data Browser,
   - Database can be filtered by rows or columns,
   - Data can be re-ordered,
   - Additional graphs can be created from database,
   - If required, return to "Step 2" and reformulate problem.

REFERENCES

[1] Adams, B. and Woodall, W., "An Analysis of Taguchi's On-Line Process-Control Procedure Under a Random-Walk Model," *Technometrics*, vol. 31, pp. 401–409, Nov 1989.

[2] Allen, J., Krishnamachari, R., Masetta, J., Pearce, D., Rigby, D., and Mistree, F., "Fuzzy compromise: An effective way to solve hierarchical design problems," *Structural Optimization*, vol. 4, pp. 21–43, 1992.

[3] Ames, A., Mattucci, N., and MacDonald, S., e. a., "Quality Loss Functions For Optimization Across Multiple Response Surfaces," *Journal of Quality Technology*, vol. 29, pp. 339–346, Jul 1997.

[4] AndereRendon, J., Montgomery, D., and Rollier, D., "Design of Mixture Experiments Using Bayesian D-Optimality," *Journal of Quality Technology*, vol. 29, pp. 451–463, Oct 1997.

[5] Anderson, C. and Wu, C., "Dispersion Measures and Analysis for Factorial Directional Data with Replicates," *Applied Statistician*, vol. 45, no. 1, pp. 47–61, 1996.

[6] Anderson, T., "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, pp. 122–148, 1963.

[7] Antony, J., "Multi-Response Optimization in Industrial Experiments Using Taguchi's Quality Loss Function and Principal Component Analysis," *Quality and Reliability Engineering International*, vol. 16, pp. 3–8, 2000.

[8] Antreich, K., Graeb, H., and Wieser, C., "Circuit Analysis and Optimization Driven by Worst Case Distances," *IEEE Trans. Computer-Aided Des. Integrated Circuits Syst.*, vol. 13, no. 1, pp. 57–71, 1994.

[9] Artiles-Leon, N., "A Pragmatic Approach to Multiple-Response Problems Using Loss Functions," *Quality Engineering*, vol. 9, no. 2, pp. 213–220, 1996.

[10] Atkinson, A. and Cook, R., "D-Optimum Designs for Heteroscedastic Linear Models," *Journal of the American Statistical Association*, vol. 90, no. 429, pp. 204–212, 1995.

[11] Barker, T., "Quality Engineering By Design: Taguchi's Philosophy," *Quality Progress*, pp. 32–42, Dec 1986.

[12] Barker, T. B., *Engineering Qualitly by Design Interpreting the Taguchi Approach*, vol. 113 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc. and ASQC Quality Press, 1990.

[13] Barlow, E., Hunter, L., and Proschan, F., "Optimum Checking Procedures," *Journal of the Society of Industrial and Applied Mathematics*, vol. 11, pp. 1078–1095, 1963.

[14] Bartlett, M., "Tests of significance in factor analysis," *B. J. Psych. Stat. sec.*, vol. 3, pp. 77–85, 1950.

[15] Bartlett, M., "The effect of standardization on a χ^2 approximation in factor analysis," *Biometrika*, vol. 38, pp. 337–344, 1951.

[16] Bartlett, M., "A note on the multiplying factors for various χ^2 approximations," *J. Roy. Stat. Soc. B*, vol. 16, pp. 296–298, 1954.

[17] Bartlett, M. and Kendall, D., "The Statistical Ananlysis of Variance-Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society, Series B*, vol. 8, pp. 128–138, 1946.

[18] Bendell, A., *Taguchi Methods Proceedings of the 1988 European Conference*. Elsevier Science Publishers LTD, 1989.

[19] Bendell, A., Disney, J., and Pridmore, W., *Taguchi Methods: Applications in World Industry*. IFS Publications and Springer-Verlag, 1989.

[20] Berger, J. O., *Statistical Decision Theory*. Wiley, 1995.

[21] Bergman, B. and Hynen, A., "Dispersion Effects from Unreplicated Designs in the 2(k-p) Series," *Technometrics*, vol. 39, pp. 191–198, May 1997.

[22] Berk, K. N. and Picard, R. R., "Significance Tests for Saturated Orthogonal Arrays," *Journal of Quality Technology*, vol. 23, pp. 79–89, Apr 1991.

[23] Bingham, D. and Sitter, R., "Minimum-Aberration Two-Level Fractional Factorial Split-Plot Designs," *Technometrics*, vol. 41, pp. 62–70, Feb 1999.

[24] Bisgaard, S., "Designing Experiments for Tolerancing Assembled Products," *Technometrics*, vol. 39, pp. 142–152, May 1997.

[25] Bisgaard, S., "The Design and Analysis of 2(k-p) x 2(q-r) Split Plot Experiments," *Journal of Quality Technology*, vol. 32, pp. 39–56, Jan 2000.

[26] Borkowski, J. and Lucas, J., "Designs of Mixed Resolution for Process Robustness Studies," *Technometrics*, vol. 39, pp. 63–70, Feb 1997.

[27] Box, G. and Jones, S., "Designs for Minimizing the Effects of Environmental Variables," tech. rep., University of Wisconsin, 1990.

[28] Box, G., "Signal-to-Noise Ratios, Performance Criteria, and Transformations," *Technometrics*, vol. 30, pp. 1–17, Feb 1988.

[29] Box, G. and Jones, S., "Split-Plot Designs for Robust Product Experimentation," *Journal of Applied Statistics*, vol. 19, no. 1, pp. 3–26, 1992.

[30] Box, G. and Meyer, R., "Dispersion Effects from Fractional Designs," *Technometrics*, vol. 28, pp. 19–27, Feb 1986.

[31] Box, G., "Discussion of "Off-Line Quality Control , Parameter Design and the Taguchi Method" by R.N. Kackar," *Journal of Quality Technology*, vol. 17, pp. 189–190, 1985.

[32] Box, G. and Jenkins, G., "Further contributions to Adaptive Quality Control: Simultaneous Estimation of Dynamics: Non-Zero Costs," *Bulletin of the International Statistical Institute*, vol. 34, pp. 943–974, 1963.

[33] Box, G. and Meyer, R. D., "Dispersion Effects From Fractional Designs," *Technometrics*, vol. 28, pp. 19–28, Feb 1986.

[34] Box, G. and Meyer, R., "Finding the Active Factors in Fractionated Screening Experiments," *Journal of Quality Technology*, vol. 25, pp. 94–105, 1993.

[35] Brinkley, P., Meyer, K., and Lu, J., "Combined Generalized Linear Modelling Non-Lineaer Programming Approach to Robust Process Design - A Case Study in Circuit Board Quality Improvement," *Applied Statistician*, vol. 45, no. 1, pp. 99–110, 1996.

[36] Bullington, K. E., Hool, J. N., and Maghsoodloo, S., "A Simple Method for Obtaining Resolution IV Designs for Use with Taguchi's Orthogonal Arrays," *Journal of Quality Technology*, vol. 22, pp. 260–264, Oct 1990.

[37] Chang, S., "An Algorithm to Generate near D-Optimal Designs for Multiple Response Surface Models," *IIE Transactions*, vol. 29, pp. 1073–1081, Dec 1997.

[38] Chang, S. I. and Shivpuri, R., "A Multiple-Objective Decision-Making Approach for Assessing Simultaneous Improvement in Die Life and Casting Quality in a Die Casting Process," *Quality Engineering*, vol. 7, no. 2, p. 371, 1994.

[39] Chapman, R. E., "Photochemistry multiple response co-optimization," *Quality Engineering*, vol. 8, no. 1, p. 31, 1995.

[40] Chen, W., Allen, J., Tsui, K., and Mistree, F., "A Procedure for Robust Design: Minimizing Variation Caused by Noise Factors and Control Factors," *Transactions of the ASME*, vol. 118, pp. 478–485, Dec 1996.

[41] Chhajed, D. and Lowe, T., "Tooling Choices and Product Performance," *IIE Transactions*, vol. 32, pp. 49–57, Jan 2000.

[42] Chipman, H., "Handling Uncertainty in Analysis of Robust Design Experiments," *Journal of Quality Technology*, vol. 30, pp. 11–17, Jan 1998.

[43] Chipman, H. and Hamada, M., "Bayesian Analysis of Ordered Categorical Data from Industrial Experiments," *Technometrics*, vol. 38, pp. 1–10, Feb 1996.

[44] Contreras, L., Bruins, H., and Trout, J., "Heating Characteristics of Beef Stew in MRE Pouches as Influenced by Critical Process and Product Parameters," working paper 100, Center for Advanced Food Technology, Rutgers University, 1995.

[45] Contreras, L., Bruins, H., and Trout, J., "Heating Characteristics of Beef Stew in MRE Pouches as Influenced by Critical Process and Product Parameters," Tech. Rep. 100, Rutgers University, 1995.

[46] Copeland, K. and Nelson, P., "Dual Response Optimization via Direct Function Minimization," *Journal of Quality Technology*, vol. 28, pp. 331–336, Jul 1996.

[47] Das, P., "Concurrent Optimization of Multiresponse Product Performance," *Quality Engineering*, vol. 11, no. 3, pp. 365–368, 1999.

[48] Del Castillo, E. and Montgomery, D. C., "A Nonlinear Programming Solution to the Dual Response Problem," *Journal of Quality Technology*, vol. 25, pp. 199–204, Jul 1993.

[49] DelCastillo, E., Fan, S., and Semple, J., "The Computationof Global Optima in Dual Response Systems," *Journal of Quality Technology*, vol. 29, pp. 347–353, Jul 1997.

[50] DelCastillo, E., Montgomery, D., and McCarville, D., "Modified Desirability Functions for Multiple Response Optimization," *Journal of Quality Technology*, vol. 28, pp. 337–345, Jul 1996.

[51] D'ERRICO, J. R. and ZAINO, N. A., "Statistical Tolerancing Using a Modification of Taguchi's Method," *Technometrics*, vol. 30, pp. 397–405, Nov 1988.

[52] DERRINGER, G. and SUICH, R., "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, vol. 12, pp. 214–219, Oct 1980.

[53] DERRINGER, G. C., "A Balancing Act:Optimizing a Product's Properties ," *Quality Progress*, pp. 51–58, June 1994.

[54] DUFFY, J., LIU, S., and MOSKOWITZ, H., E. A., "Assessing Multivariate Process/Product Yield Via Discrete Point Approximation," *IIE Transactions*, vol. 30, pp. 535–543, Jun 1998.

[55] DUMOUCHEL, W. and JONES, B., "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model," *Technometrics*, vol. 36, pp. 37–47, Feb 1994.

[56] ELLEKJAER, M. R., FULLER, H. T., and LADSTEIN, K., "Analysis of Unreplicated Split-Plot Experiments with Multiple Responses," *Quality Engineering*, vol. 10, no. 1, p. 25, 1997.

[57] ENGEL, J., "Modeling Variation in Industrial Experiments," *Applied Statistics*, vol. 41, pp. 579–593, 1992.

[58] ENGEL, J. and HUELE, A., "A Generalized Linear Modeling Approach to Robust Design," *Technometrics*, vol. 38, pp. 365–373, Nov 1996.

[59] Engineous Software, Inc., *iSIGHT Basic Training Guide Version 6.0*, 2002.

[60] FAN, S., "A Generalized Global Optimization Algorithm for Dual Response Systems," *Journal of Quality Technology*, vol. 32, pp. 444–456, Oct 2000.

[61] FATHI, Y. and PALKO, D., "A Mathematical Model and a Heuristic Procedure for the Robust Design Problem with High-Low Tolerances," *IIE Transactions*, vol. 33, pp. 1121–1127, Dec 2000.

[62] FLURY, B., *Common Principal Components and Related Multivariate Models*. John Wiley and Sons, 1988.

[63] FOGLIATTO, F. S. and ALBIN, S. L., "Variance of Predicted Response as an Optimization Criterion in Multiresponse Experiments," *Quality Engineering*, vol. 12, no. 4, pp. 523–533, 2000.

[64] FUCHS, C. and BENJAMINI, Y., "Multivariate Profile Charts for Statistical Process Control," *Technometrics*, vol. 36, pp. 182–195, 1994.

[65] FUCHS, C. and KENETT, R. S., *Multivariate Quality Control Theory* . NY: Marcel Dekker, 1998.

[66] GEMAN, S. and GEMAN, D., "Stochastic Relaxation, Gibbs Distributions, and the Bayesian REstoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[67] GHOSH, S. and DERDERIAN, E., "Robust Experimental Plan and Its Role in Determining Robust Design Against Noise Factors," *The Statistician*, vol. 42, pp. 19–28, 1993.

[68] GHOSH, S. and DERDERIAN, E., "Determination of Robust Design Against Noise Factors and in Presence of Signal Factors," *Commun. Statist.-Simula.*, vol. 24, no. 2, pp. 309–326, 1995.

[69] Gibra, I., "Economically Optimal Determination of the Parameters of np-Control Charts," *Journal of Quality Technology*, vol. 10, pp. 12–19, 1978.

[70] Goik, P., Liddy, J. W., and Taam, W., "Use of Desirability Functions to Determine Operating Windows for New Product Designs," *Quality Engineering*, vol. 7, no. 2, p. 267, 1994.

[71] Hamada, M., "Reliability Improvement Via Taguchi's Robust Design," *Quality and Reliability Engineering International*, vol. 9, pp. 7–13, 1993.

[72] Hamada, M. and Nelder, J., "Generalized Linear Models for Quality-Improvement Experiments," *Journal of Quality Technology*, vol. 29, pp. 292–304, Jul 1997.

[73] Hamada, M. and Wu, C., *xperiments: Planning, Optimizing, etc.* Wiley, 2000.

[74] Harrington, E. C. J., "The Desirability Function," *Industrial Quality Control*, vol. 21, pp. 494–498, Apr 1965.

[75] Hawkins, D. M., "Regression adjustment for variables in multivariate quality control," *Journal of Quality Technology*, vol. 25, pp. 170–182, July 1993.

[76] Hayter, A. J. and Tsui, K.-L., "Identification and quantification in multivariate quality control problems," *Journal of Quality Technology*, vol. 26, pp. 197–208, July 1994.

[77] Hazelrigg, G., *Systems engineering an approach to information-based design.* Upper Saddle River, NJ: Prentice Hall, 1996.

[78] Healy, J., "A note on multivariate CUSUM procedures," *Technometrics*, vol. 29, pp. 409–412, 1987.

[79] Hedayat, A. and Stufken, J., "Compound Orthogonal Arrays," *Technometrics*, vol. 41, pp. 57–61, Feb 1999.

[80] Hernandez, G., Allen, J., and Mistree, F., "The compromise decision support problem: Modelling the deviation function as in physical programming," *Engineering Optimization*, vol. 33, no. 4, pp. 445–471, 2001.

[81] Hotelling, H., "The generalization of Student's ratio," *Ann. Math. Stat.*, vol. 2, pp. 360–378, 1931.

[82] Hotelling, H., "A generalized T test and measure of multivariate dispersion," *Proceedings of the 2nd Berkeley Symp. of Math. Stat. and Prob.*, vol. 1, pp. 23–41, 1951.

[83] Hunter, J. S., "Statistical Design Applied to Product Design," *Journal of Quality Technology*, vol. 17, pp. 210–250, Oct 1985.

[84] Ignizio, J., *Linear Programming in Single and Multi-Objective Systems.* Prentice Hall, 1982.

[85] Jackson, J. E., "Principal Components and Factor Analysis: Part I-Principal Components," *Journal of Quality Technology*, vol. 12, pp. 201–213, October 1980.

[86] Jackson, J. E., "Principal Components and Factor Analysis: Part II-Additional Topics Related to Principal Components," *Journal of Quality Technology*, vol. 13, pp. 46–58, January 1981.

[87] Jackson, J. E., "Multivariate Quality Control," *Commun. Statist. Meth.*, vol. 14, no. 11, pp. 2657–2688, 1985.

[88] Jackson, J. E., *A User's Guide to Principal Components*. NY: John Wiley and Sons, Inc., 1991.

[89] Jackson, J. and Mudholkar, G., "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, pp. 341–349, 1979. Addendum, Vol. 22, p. 136 in 1980.

[90] Jiang, B., Chen, D., Black, J., and Hool, J., "Taguchi-Based Methodology for Determining Optimizing Robot Process Capability," *IIE Transactions*, vol. 23, pp. 169–184, Jun 1991.

[91] Johnson, R. A. and Wichern, D. W., *Applied Multivariate Statistical Analysis*. Prentice-Hall, 3 ed., 1992.

[92] Jolliffe, I., *Principal Component Analysis*. NY: Springer, 2002.

[93] Joseph, V. R., "Quality Loss Functions for Nonnegative Variables and Its Applications," Tech. Rep. 380, University of Michigan, December 2001.

[94] Kackar, R. N., "Off-Line Quality Control, Parameter Design, and the Taguchi Method," *Journal of Quality Technology*, vol. 17, pp. 176–200, Oct 1985.

[95] Kacker, R. and Tsui, K.-L., "Interaction Graphs: Graphical Aids for Planning Experiments." 1987.

[96] Kacker, R. N., Lagergren, E. S., and Filliben, J. J., "Taguchi's Fixed-Element Arrays are Fractional Factorials," *Journal of Quality Technology*, vol. 23, pp. 107–116, Apr 1991.

[97] Kacker, R. N. and Tsui, K.-L., "Interaction Graphs: Graphical Aids for Planning Experiments," *Journal of Quality Technology*, vol. 22, pp. 1–14, Jan 1990.

[98] Karandikar, H. and Mistree, F., "The Integration of Information from Design and Manufacture Through Decision Support Problems," *Applied Mechanics Reviews*, vol. 44, no. 11, pp. 150–159, 1991.

[99] Khattre, R., "Robust Parameter Design: A Response Surface Approach," *Journal of Quality Technology*, vol. 28, pp. 187–198, Apr 1996.

[100] Khuri, A. and Conlon, M., "Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions," *Technometrics*, vol. 23, pp. 363–375, Nov 1981.

[101] Kim, K.-J. and Lin, D. K., "Simultaneous Optimization of Mechanical Properties of Steel by Maximizing Exponential Desirability Functions," *Applied Statistics*, vol. 49, no. 3, pp. 311–325, 2000.

[102] Kim, Y. and Myers, R., "A Response Surface Approach to Data Analysis in Robust Parameter Design," Tech. Rep. 92-12, Virginia Polytechnic Institute & State University, 1992.

[103] Kirmani, S. and Das Peddada, S., "Stochastic Ordering Approach to Off-Line Quality Control," *Applied Statistics*, vol. 42, no. 2, pp. 271–281, 1993.

[104] Kourti, T. and MacGregor, J. F., "Multivariate SPC methods for process and product monitoring," *Journal of Quality Technology*, vol. 28, pp. 409–428, October 1996.

[105] Kuhn, A., Carter, W., and Myers, R., "Incorporating Noise Factors into Experiments with Censored Data," *Technometrics*, vol. 42, pp. 376–383, Nov 2000.

[106] Kumar, P., Barua, P., and Gaindhar, J., "Quality Optimization (Multi-Characteristics) Through Taguchi's Technique and Utility Concept," *Quality and Reliability Engineering International*, vol. 16, pp. 475–485, 2000.

[107] Lawley, D., "A generalization of Fisher's z-test," *Biometrika*, vol. 30, pp. 180–187, 1938.

[108] Lawley, D., "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, vol. 43, pp. 128–136, 1956.

[109] Ledolter, J. and Swersey, A., "Dorian Shainin's Variables Search Procedure: A Critical Assessment," *Journal of Quality Technology*, vol. 29, pp. 237–247, Jul 1997.

[110] Lenth, R. V., "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics*, vol. 31, pp. 469–473, Nov 1989.

[111] Leon, R. V., Shoemaker, A. C., and Kackar, R. N., "Performance Measures Independent of Adjustment: An Explanation and Extension of Taguchi's Signal-to-Noise Ratios," *Technometrics*, vol. 29, pp. 253–265, Aug 1987.

[112] Leon, R., Shoemaker, A., and Tsui, K., "A Systematic Approach to Planning for a Designed Industrial Experiment - Discussion," *Technometrics*, vol. 35, pp. 21–24, Feb 1993.

[113] Leon, R. and Wu, C., "Theory of Performance Measures in Parameter Design," *Statistica Sinica*, vol. 2, pp. 335–358, Jul 1992.

[114] Li, W. and Nachtsheim, C., "Model-Robust Factorial Designs," *Technometrics*, vol. 42, pp. 345–352, Nov 2000.

[115] Lin, D. and Tu, W., "Dual Response Surface Optimization," *Journal of Quality Technology*, vol. 28, pp. 496–496, Oct 1995.

[116] Lin, Y., "Robust design goal formulations and metamodeling techniques," Master's thesis, Georgia Institute of Technology, 2000.

[117] Lowry, C. A. and Montgomery, D. C., "A review of multivariate control charts," *IIE Transactions*, vol. 27, pp. 800–810, 1995.

[118] Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E., "A multivariate exponentially weighted moving average control chart," *Technometrics*, vol. 34, pp. 46–53, February 1992.

[119] Lucas, J. M., "How to Achieve a Robust Process Using Response Surface Methodology," *Journal of Quality Technology*, vol. 26, pp. 248–260, Oct 1994.

[120] Lunani, M., Nair, V., and Wasserman, G. S., "Robust Design of Dynamic Systems: An Applied Approach to Identifying Suitable Measures of Dispersion." 1995.

[121] Lunani, M., Nair, V., and Wasserman, G., "Graphical Methods for Robust Design with Dynamic Characteristics," *Journal of Quality Technology*, vol. 29, pp. 327–338, Jul 1997.

[122] Maghsoodloo, S. and Li, M., "Optimal Asymmetric Tolerance Design," *IIE Transactions*, vol. 32, pp. 1127–1137, Dec 2000.

[123] Maghsoodloo, S., "The Exact Relation of Taguchi's Signal-to-Noise Ratio to His Quality Loss Function," *Journal of Quality Technology*, vol. 22, pp. 57–67, Jan 1990.

[124] Mason, R. L., Champ, C. W., Tracy, N. D., Wierda, S. J., and Young, J. C., "Assessment of multivariate process control techniques," *Journal of Quality Technology*, vol. 29, pp. 140–162, April 1997.

[125] Mason, R. L., Tracy, N. D., and Young, J. C., "Decomposition of T^2 for multivariate control chart interpretation," *Journal of Quality Technology*, vol. 27, pp. 99–108, April 1995.

[126] Mason, R. L. and Young, J. C., *Multivariate Statistical Process Control with Industrial Applications*. Alexandria, VA and Philadelphia, PA: ASA-SIAM, 2002.

[127] Mays, D., "Optimal Central Composite Deisgns in the Presence of Dispersion Effects," *Journal of Quality Technology*, vol. 31, pp. 398–407, Oct 1999.

[128] McCaskey, S. and Tsui, K.-L., "Analysis of Dynamic Robust Design Experiments," *International Journal of Production Research*, vol. 35, no. 6, pp. 1561–1574, 1997.

[129] McCullough, P. and Nelder, J., *Generalized Linear Models*. Chapman and Hall, NY, 2 ed., 1989.

[130] McGrath, R. and Lin, D., "Confounding of Location and Dispersion Effects in Unreplicated Fractional Factorials," *Journal of Quality Technology*, vol. 33, pp. 129–139, Apr 2001.

[131] Mesenbrink, P., Lu, J., McKenzie, R., and Taheri, J., "Characterization and Optimization of a Wave-Soldering Process," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1209–1217, 1994.

[132] Miller, A. and Wu, C., "Parameter Design for Signal-Response Systems: A Different Look at Taguchi's Dynamic Parameter Design," *Statistical Science*, vol. 11, no. 2, pp. 122–136, 1996.

[133] Mistree, F., Hughes, O., and Bras, B., "The Compromise Decision Support Problem and the Adaptive Linear Programming Algorithm," *Structural Optimization: Status and Promise*, pp. 247–286, 1993. AIAA.

[134] Mohebbi, C. and Hayre, L., "Multivariate control charts: a loss function approach," *Sequential Analysis*, vol. 8, no. 3, pp. 253–268, 1989.

[135] Montgomery, D. C. and Klatt, P., "Economic design of T^2 control charts to maintain current control of a process," *Man. Sci.*, vol. 19, pp. 76–89, 1972.

[136] Montgomery, D. C. and Klatt, P., "Minimum cost multivariate quality control tests," *AIEE Transaction*, vol. 4, pp. 103–110, 1972.

[137] Moorhead, P. and Wu, C., "Cost-Driven Parameter Design," *Technometrics*, vol. 40, pp. 111–119, May 1998.

[138] Moskowitz, H., Plante, R., and Duffy, J., "Multivariate Tolerance Design Using Quality Loss," *IIE Transactions*, vol. 33, pp. 437–448, Jun 2001.

[139] Mueller, H., *Nonparametric Regression Analysis of Longitudinal Data*. Springer-Verlag, Berlin, Germany, 1988.

[140] Murphy, T. E., Tsui, K.-L., and Allen, J. K., "Compromise design settings for multiple responses," 2002.

[141] Myers, R. H., Khuri, A. I., and Carter, W. H., "Response Surface Methodology: 1966-1988," *Technometrics*, vol. 31, pp. 137–157, May 1989.

[142] Myers, R., "Response Surface Methodology -Current Status and Future Directions," *Journal of Quality Technology*, vol. 31, pp. 30–44, Jan 1999.

[143] Myers, R. and Carter, W.H., J., "Response Surface Techniques for Dual Response Systems," *Technometrics*, vol. 15, pp. 301–317, 1973.

[144] Myers, R., Khuri, A., and Vining, G., "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," *The American Statistician*, vol. 46, pp. 131–139, May 1992.

[145] Myers, R., Kim, Y., and Griffiths, K., "Response Surface Methods and the Use of Noise Variables," *Journal of Quality Technology*, vol. 29, pp. 429–440, Oct 1997.

[146] Myers, R. and Montgomery, D., "A Tutorial on Generalized Linear Models," *Journal of Quality Technology*, vol. 30, pp. 423–424, Oct 1998.

[147] Nair, Vijayan N., e. a., "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, vol. 34, pp. 127–161, May 1992.

[148] Nair, V. N. and Pregibon, D., "Analyzing Dispersion Effects From Replicated Factorial Experiments," *Technometrics*, vol. 30, pp. 247–257, Aug 1988.

[149] Nair, V. and Pregibon, D., "A Data Analysis Strategy for Quality Engineering Experiments," *ATT Technical Journal*, vol. 65, pp. 73–84, 1986.

[150] Nayebpour, M. and Woodall, W., "An Analysis of Taguchi Online Quality-Monitoring Procedures for Attributes," *Technometrics*, vol. 35, pp. 53–60, Feb 1993.

[151] Nelder, J. and Lee, Y., "Generalized Linear Models for the Analysis of Taguchi Type Experiments," *Applied Stochastic Models and Data Analysis*, vol. 7, pp. 107–120, 1991.

[152] Nelder, J. and Lee, Y., "Joint Modeling of Mean and Dispersion," *Technometrics*, vol. 40, pp. 168–171, May 1998.

[153] Nelder, J. and Wedderburn, R., "Generalized Linear Models," *Journal of the Royal Statistical Society*, no. A132, pp. 370–384, 1972.

[154] Ng, K. and Tsui, K.-L., "Expressing Variability and Yield with a Focus on the Customer," *Quality Engineering*, vol. 5, no. 2, pp. 255–267, 1992.

[155] Ortega, R. A., "Robust design goal formulations and metamodeling techniques," Master's thesis, Georgia Institute of Technology, 1998.

[156] Pan, G., "The Impact of Unidentified Location Effects on Disperison-Effects Identification from Unreplicated Factorial Designs," *Technometrics*, vol. 41, pp. 313–326, Nov 1999.

[157] Pearson, K., "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, pp. 559–572, 1901.

[158] Pignatiello, J., "Strategies for Robust Multiresponse Quality Engineering," *IIE Transactions*, vol. 25, pp. 5–15, May 1993.

[159] Pignatiello, J. J., "An Overview of the Strategy and Tactics of Taguchi," *IIE Transactions*, vol. 20, pp. 247–254, Sep 1988.

[160] Pignatiello, J. J. and Runger, G. C., "Comparisons of multivariate CUSUM charts," *Journal of Quality Technology*, vol. 22, pp. 173–186, July 1990.

[161] Plante, R., "Process Capability: A Criterion for Optimizing Multiple Response Product and Process Design," *IIE Transactions*, vol. 33, pp. 497–509, Jun 2001.

[162] Pledger, M., "Observable Uncontrollable Factors in Parameter Design," *Journal of Quality Technology*, vol. 28, pp. 153–162, Apr 1996.

[163] Reddy, P., Nishina, K., and Babu, A. S., "Unification of Robust Design and Goal Programming for Multi-Response Optimization - A Case Study," *Quality and Reliability Engineering International*, vol. 13, pp. 371–383, 1997.

[164] Ribeiro, J. and Albin, S. L., "An Objective Function of Multiple Quality Characteristics for Parameter Optimization Experiments," IE Working Paper 93-111, Dept. of Industrial Engineering, Rutgers University, 1993.

[165] Ribeiro, J. and Elsayed, E., "A Case Study on Process Optimization Using the Gradient Loss Function," *International Journal of Production Research*, vol. 33, no. 12, pp. 3233–3248, 1995.

[166] Ribeiro, J. L., Fogliatto, F. S., and ten Caten, C. S., "Minimizing Manufacturing and Quality Costs in Multiresponse Optimization," *Quality Engineering*, vol. 13, no. 2, pp. 191–201, 2000.

[167] Rosenbaum, P., "Some Useful Compound Dispersion Experiments in Quality Design," *Technometrics*, vol. 38, pp. 354–364, Nov 1996.

[168] Rosenbaum, P., "Blocking in Compound Dispersion Experiments," *Technometrics*, vol. 41, pp. 125–134, May 1999.

[169] Ross, P. J., *Taguchi Techniques for Quality Engineering*. McGraw-Hill, 1988.

[170] Sahni, N. S., Eide, O., and Naes, T., "An Application of Multivariate Analysis in Product Development in the Food Industry," *Quality Engineering*, vol. 11, no. 4, pp. 579–586, 1999.

[171] Savage, G. and Swan, D., "Probabilistic Robust Design with Multiple Quality Characteristics," *Quality Engineering*, vol. 13, no. 4, pp. 629–640, 2001.

[172] Schott, J., "Testing the equality of the smallest latent roots of a correlation matrix," *Biometrika*, vol. 75, pp. 794–796, 1988.

[173] Seepersad, C. S., "The utility-based compromise decision support problem with applications in product platform design," Master's thesis, Georgia Institute of Technology, 2001.

[174] Seepersad, C. S., Mistree, F., and Allen, J., "A Quantitative Approach for Designing Multiple Platforms for and Evolving Portfolio of Products," Design Automation Conference, 2002.

[175] Shoemaker, A. and Tsui, K.-L., "Response Model Analysis for Robust Design Experiments," *Commun. Statist.-Simula.*, vol. 22, no. 4, pp. 1037–1064, 1993.

[176] Shoemaker, A. C., Tsui, K.-L., and Leon, R. V., "Discussion of Box's Signal-to-Noise Ratios, etc.,February 1988 in Technometrics," *Technometrics*, vol. 30, pp. 19–21, Feb 1988.

[177] Shoemaker, A. C., Tsui, K.-L., and Wu, C. J., "Economical Experimentation Methods for Robust Design," *Technometrics*, vol. 33, pp. 415–427, Nov 1991.

[178] Simon, H., *Administrative behavior*. New York, NY: The Free Press, 1976.

[179] Singpurwalla, N., Barlow, R., Berkeley, U., Lindley, D., Phadke, M., and Pignatiello, J., "A Bayesian Perspective on Taguchi Approach to Quality Engineering and Tolerance Design," *IIE Transactions*, vol. 24, pp. 18–32, Nov 1992.

[180] Sohn, S. and Park, C., "Random Effects Linear Models for Process Mean and Variance," *Journal of Quality Technology*, vol. 30, pp. 33–39, Jan 1998.

[181] Srinivasan, R., Karandikar, H., and Mistree, F., "Understanding Design-Manufacture Interaction using Compromise Decisions Support Problems Parts I,II, and III: A Formulation for Composite Pressure Vessels," *Computers and Structures*, vol. 40, no. 3, pp. 679–717, 1991.

[182] Steinberg, D. M. and Bursztyn, D., "Dispersion Effects in Robust-Design Experiments with Noise Factors," *Journal of Quality Technology*, vol. 26, pp. 12–20, Jan 1994.

[183] Steinberg, D. and Bursztyn, D., "Noise Factors, Dispersion Effects, and Robust Design," tech. rep., Department of Statistics, Tel-Aviv University, Israel, 1993.

[184] Steiner, S. and Hamada, M., "Making Mixtures Robust to Noise and Mixing Measurement Errors," *Journal of Quality Technology*, vol. 29, pp. 441–450, Oct 1997.

[185] Suhr, R. and Batson, R. G., "Constrained Multivariate Loss Function Minimization," *Quality Engineering*, vol. 13, no. 3, pp. 475–483, 2001.

[186] Taguchi, G., Elsayed, E., and Hsaing, T., *Quality Engineering in Production Systems*. McGraw-Hill, NY, 1989.

[187] Taguchi, G. and Wu, Y., *Introduction to Off-Line Quality Control*. Nagoya, Central Japan Quality Control ASsociation, Japan, 1980.

[188] Taguchi, G., *Introduction to Quality Engineering: Designing Quality into Products and Processes*. Kraus International Publications, White Plains, NY, 1986.

[189] Tong, L.-I., Houng, J.-Y., and Chen, J.-Y., "Optimizing Dynamic Multi-Response Problems Using the Dual Response Surface Method," *Quality Engineering*, vol. 14, no. 1, p. 115, 2001.

[190] Tong, L.-I. and Hsieh, K.-L., "A Novel Means of Applying Neural Networks to Optimize the Multiresponse Problem," *Quality Engineering*, vol. 13, no. 1, pp. 11–18, 2000.

[191] Tong, L.-I. and Su, C.-T., "Optimizing Multi-response Problems in the Taguchi Method by Fuzzy Multiple Attribute Decision Making," *Quality and Reliability Engineering International*, vol. 13, pp. 25–34, 1997.

[192] Tsui, K., "An Overview of Taguchi Method and Newly Developed Statistical Methods for Robust Design," *IIE Transactions*, vol. 24, pp. 44–57, Nov 1992.

[193] Tsui, K.-L., "Strategies for Planning Experiments Using Orthogonal Arrays and Confounding Tables," *Quality and Reliability Engineering International*, vol. 4, pp. 113–122, Jan 1988.

[194] Tsui, K.-L., "A Multi-Step Analysis Procedure for Robust Design, Industrial and Systems Engineering Research Report J-93-03," tech. rep., Georgia Institute of Technology, 1993.

[195] Tsui, K.-L., "Avoiding Unnecessary Bias in Robust Design Analysis," *Computational Statistics and Data-Analysis*, vol. 18, pp. 535–546, 1994.

[196] Tsui, K.-L., "A Critical Look at Taguchi's Modelling Approach for Robust Design," *Journal of Applied Statistics*, vol. 23, no. 1, pp. 81–95, 1996.

[197] Tsui, K.-L., "A Multi-Step Analysis Procedure for Robust Design," *Statistica Sinica*, vol. 6, pp. 631–648, 1996.

[198] Tsui, K.-L., "Interpretation of Process Capability Indices and Some Alternatives," *Quality Engineering*, vol. 9, no. 4, pp. 587–596, 1997.

[199] Tsui, K.-L., "Modeling and Analysis of Dynamic Robust Design Experiments," *IIE Transactions*, vol. 31, pp. 1113–1122, Sep 1999.

[200] Tsui, K.-L., "Robust Design Optimization for Multiple Characteristic Problems," *International Journal of Production Research*, vol. 37, no. 2, pp. 433–445, 1999.

[201] Tsui, K.-L., Goh, T., Xie, M., and Loy, C., "A Non-Domination Search Approach to Multi-Response Design of Experiments." 2001.

[202] Tsui, K.-L. and Li, A., "Analysis of Smaller and Larger the Better Robust Design Experiments," *International Journal of Industrial Engineering*, vol. 1, no. 3, pp. 193–202, 1994.

[203] Tsui, K.-L. and Woodall, W. H., "Multivariate Control Charts Based on Loss Functions," *Sequential Analysis*, vol. 12, no. 1, pp. 79–92, 1993.

[204] Ullman, N. R., "The Analysis of Means for Signal and Noise," *Journal of Quality Technology*, vol. 21, pp. 111–127, Apr 1989.

[205] Vadde, S., Allen, J., and Mistree, F., "Compromise decision support problems for hierarchical design under uncertainty," *Computers and Structures*, vol. 52, no. 4, pp. 645–658, 1994.

[206] Vadde, S., Krishnamachari, R., Allen, J., and Mistree, F., "The bayesian compromise decision support problem for hierarchical design involving uncertainty," *ASME Journal of Mechanical Design*, vol. 116, pp. 388–395, 1994.

[207] Vining, G. G. and Myers, R. H., "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach," *Journal of Quality Technology*, vol. 22, pp. 38–45, Jan 1990.

[208] Vining, G., "A Compromise Approach to Multi-Response Optimization," *Journal of Quality Technology*, vol. 30, pp. 309–313, Oct 1998.

[209] Vining, G. and Bohn, L., "Response Surfaces for the Mean and Variance Using a Nonparametric Approach," *Journal of Quality Technology*, vol. 30, pp. 282–291, Jul 1998.

[210] Vining, G. and Schaub, D., "Experimental Designs for Estimating both Mean and Variance Functions," *Journal of Quality Technology*, vol. 28, pp. 135–147, Apr 1996.

[211] Wade, M. R. and Woodall, W. H., "A review and analysis of cause-selecting control charts," *Journal of Quality Technology*, vol. 25, pp. 161–169, July 1993.

[212] Wang, Y., Lin, D., and Fang, K., "Designing Outer Array Points," *Journal of Quality Technology*, vol. 27, pp. 226–241, Jul 1995.

[213] Wasserman, G. S., "Parameter Dsign with Dynamic Characteristics: A Regression Perspective," *Quality and Reliability Engineering International*, vol. 12, pp. 113–117, 1996.

[214] Welch, W. J., Yu, T.-K., Kang, S. M., and Sacks, J., "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, vol. 22, pp. 15–22, Jan 1990.

[215] Wolfinger, R. and Tobias, R., "Joint Estimation of Location, Disperison, and Random Effects in Robust Design," *Technometrics*, vol. 40, pp. 62–71, Feb 1998.

[216] Woodall, W. and Montgomery, D., "Research Issues and Ideas in Statistical Process Control," *Journal of Quality Technology*, vol. 31, no. 4, pp. 376–386, 1999.

[217] Woodall, W. H., Koudelik, R., Tsui, K.-L., Kim, S. B., Stoumbos, Z. G., and Carvounis, C. P., "A Review and Analysis of the Mahalanobis-Taguchi System." 2001.

[218] Wu, C. J. and Hamada, M., *Experiments Planning, Analysis, and Parameter Design Optimization*, ch. 10, pp. 461–466. Wiley-Interscience, 2000.

[219] Wurl, R. and Albin, S. L., "A Comparison of Multiresponse Optimization: Sensitivity to Parameter Selection," *Quality Engineering*, vol. 11, no. 3, pp. 405–415, 1999.

[220] Ye, K. and Hamada, M., "Critical Values of the Lenth Method for Unreplicated Factorial Designs," *Journal of Quality Technology*, vol. 32, pp. 57–66, Jan 2000.

[221] Zhang, C. and Wang, H., "Robust Design of Assembly and Machining Tolerance Allocations," *IIE Transactions*, vol. 30, pp. 17–29, Jan 1998.

[222] Zhou, Q., Allen, J., and Mistree, F., "Decisions under uncertainty: The fuzzy compromise decision support problem," *Engineering Optimization*, vol. 20, pp. 21–43, 1992.

VITA

Terrence E. Murphy was born on April 15, 1961 in Indianapolis, the sixth of ten children. He received his bachelor degree in Biomedical and Electrical Engineering Technology in May 1983 from Temple University in Philadelphia, PA.

From 1983 through 1989 , he worked as a quality control engineer and training consultant for the Eastman Kodak Company. In 1989 and 1990 he explored teaching at the junior high level at a private middle school in Ajijic, Mexico. In 1992 he returned to Eastman Kodak Company as a manufacturing engineer for a line of clinical blood analyzers and through 1998 worked in various capacities in design and six-sigma projects. In that same period he completed a M.S. in Applied and Mathematical Statistics from the Rochester Institute of Technology, which was awarded in May of 1998. From 1998 through 2004 he completed a Ph.D. at the School of Industrial and Systems Engineering at the Georgia Institute of Technology. His area of concentration is engineering statistics and his thesis focused on new applications of principal components to areas in robust design and multivariate statistical process control.

During his studies he received funding as a student teacher and teaching assistant for the first three years. From $2002 - 2004$ he worked as a graduate research assistant at OMED Education Services at Georgia Tech. He is pursuing a position in academia and wishes to teach and conduct research in statistics at an institution that highly values both teaching and research.