Integrated Robust Design Using Response Surface Methodology and Constrained Optimization

by

Lijun Chen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2008

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

System design, parameter design, and tolerance design are the three stages of product or process development advocated by Genichi Taguchi. Parameter design, or robust parameter design (RPD), is the method to determine nominal parameter values of controllable variables such that the quality characteristics can meet the specifications and the variability transmitted from uncontrollable or noise variables is minimized for the process or product. Tolerance design is used to determine the best limits for the parameters to meet the variation and economical requirements of the design.

In this thesis, response surface methodology (RSM) and nonlinear programming methods are adopted to integrate the parameter and tolerance design. The joint optimization method that conducts parameter design and tolerance design simultaneously is more effective than the traditional sequential process. While Taguchi proposed the crossed array design, the combined array design approach is more flexible and efficient since it combines controllable factors, internal noise factors, and external noise factors in a single array design. A combined array design and the dual response surface method can provide detailed information of the process through process mean and process variance obtained from the response model. Among a variety of cuboidal designs and spherical designs, standard or modified central composite designs (CCD) or face-centered cube (FCC) designs are ideal for fitting second-order response surface models, which are widely applied in manufacturing processes. Box-Behnken design (BBD), mixed resolution design (MRD), and small composite design (SCD) are also discussed as alternatives. After modeling the system, nonlinear programming can be used to solve the constrained optimization problem. Dual RSM, mean square error (MSE) loss criterion, generalized linear model, and desirability function approach can be selected to work with quality loss function and production cost function to formulate the object function for optimization. This research also extends robust design and RSM from single response to the study of multiple responses.

It was shown that the RSM is superior to Taguchi approach and is a natural fit for robust design problems. Based on our study, we can conclude that dual RSM can work very well with ordinary least squares method or generalized linear model (GLM) to solve robust parameter design problems. In addition, desirability function approach is a good selection for multiple-response parameter design problems. It was confirmed that considering the internal noise factors (standard deviations of the control factors) will improve the regression model and have a more appropriate optimal solution. In addition, simulating the internal noise factors as control variables in the combined array design is an

attractive alternative to the traditional method that models the internal noise factors as part of the noise variables.

The purpose of this research is to develop the framework for robust design and the strategies for RSM. The practical objective is to obtain the optimal parameters and tolerances of the design variables in a system with single or multiple quality characteristics, and thereby achieve the goal of improving the quality of products and processes in a cost effective manner. It was demonstrated that the proposed methodology is appropriate for solving complex design problems in industry applications.

Acknowledgements

I would like to thank my supervisor Professor Gordon Savage for his guidance, and my readers Professor Keith W. Hipel and Professor Stefan Steiner for their contributions. I would also like to acknowledge my parents and my wife for their support.

Dedication

This thesis is dedicated to my dearest wife, Wei.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Robust Design and Response Surface Methodology

Among the various applications of industry statistics, two of the most successful developments over the last 60 years are: response surface methodology (RSM) originating from the chemical industries in the 1950s and 1960s, and robust design, particularly robust parameter design (RPD), originating from manufacturing quality improvement initiatives in the 1980s (Steinberg 2008). RSM is a natural fit for the robust design problems in various industrial applications and academic researches.

In the 1980s, Genichi Taguchi (Taguchi and Wu (1985), and Taguchi (1986, 1987)) introduced robust design on quality engineering through the statistical design of experiments. The concepts of robust design and its realization methods are significant contributions to modern quality and process improvement. Taguchi defines three stages in product or process development: system design, parameter design, and tolerance design. In this thesis, we will work on robust parameter design, tolerance design, as well as their combination, called integrated robust design.

System design is the conceptual design, which proposes the principal system configuration based on existing knowledge and resources. Good system design can guarantee that the subsequent parameter and tolerance design stage are feasible and going in the right direction. Robust parameter design, or process robustness study in a manufacturing process, aims to achieve the requirements for the quality characteristics through determining nominal parameter values of controllable factors (or control variables) while minimizing the variability transmitted from uncontrollable factors (or noise variables). Control variables, which are denoted by $\mathbf{x}$, are easily controlled and manipulated, while noise variables, which are denoted by $\mathbf{z}$, are uncontrollable or difficult to control. It is desirable to make the responses, which are denoted by $\mathbf{y}$, robust or insensitive to the noise variables, while meeting the specification requirements. Tolerance design is employed to determine the best specification limits for the parameters with economic consideration. Due to the need for more resources to realize the tighter tolerance, the production cost usually increases as the tolerances of the control variables are reduced. In tolerance design, control variables with significant effects on the performance of the system should be identified, and their specification limits should be optimized in a yield-cost scenario. On the other hand, the insignificant control variables can be specified with wider variability to save the limited resources.

In this thesis, we focus on reducing variation of a process or product, which is referred to as a system, through proactive robust design methods. Statistical experiments based response surface methodology (RSM) is presented to solve the robust design problem and nonlinear programming methods are used for the constrained optimization. Figure 1-1 presents a framework using response surface methodology for the integrated robust design of a system with single or multiple quality characteristics. Though most of the processes are standard or generic, some new methods and comparisons will be discussed in this thesis.

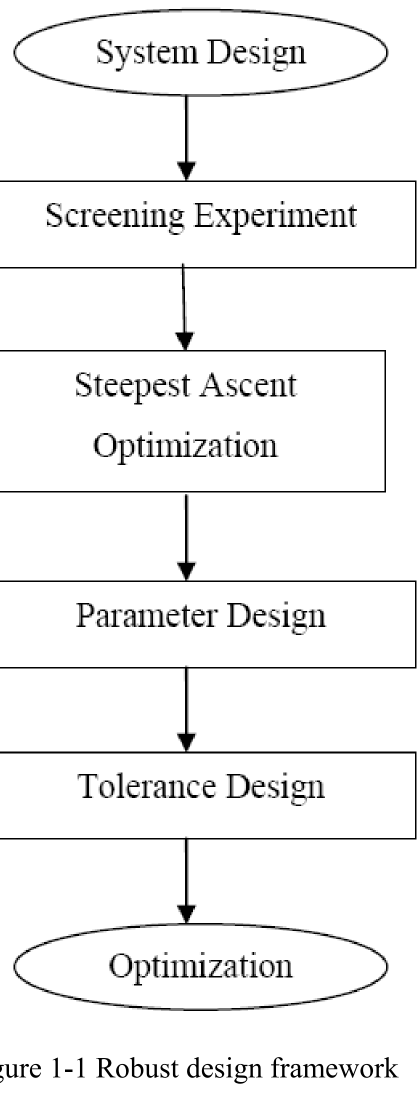

Figure 1-1 Robust design framework

1.2 Statement of Problem

Different kinds of variation exist everywhere and anytime in a production process. Reducing variation is one of the most important tasks for an engineer. We can improve the system performance through identifying the causes of variation and then taking actions to control them to meet our goal. Specifically, such kind of reactive method is one of the major problem solving techniques in manufacturing industry. However, robust design, as a proactive solution, is more effective and efficient than the traditional quality control methods. The philosophy of robust design is design for quality and reliability.

In general, we refer to Taguchi's robust design methodology as a two-stage sequential design approach. First, the designer invokes parameter design to determine the nominal values of the control variables and meet the target of the response. And then, tolerance design is performed to find the optimal tolerances around the parameters. Some engineers argue that the parameters and tolerances of design variables are affecting each other and competing in the total cost. A more competitive strategy is to optimize them simultaneously. We call the joint method integrated parameter and tolerance design, or robust design, and will study it intensely in this thesis.

Robust design is a very important concept for quality improvement and is widely adopted by scientists and engineers to improve the quality of a system and satisfy the customers. However, regarding the methodology of robust design, we have far more work to do to standardize the methods under different conditions and requirements, and to narrow the gap between the research and industry applications. Gremyr et al. (2003) conducted surveys to study the status of robust design methodology in the Swedish manufacturing industry. They revealed that only 17% out of the 105 sampled companies applied robust design methodology, though a majority of the companies were trying to minimize variation. In addition, their study shows that robust design is used mainly in large companies.

Taguchi laid a strong emphasis on variability reduction. His method includes two important parts: crossed array designs as the experimental strategy and signal-to-noise ratios (SNR) as the analysis method. One should be aware that despite criticisms put forward by Western statisticians concerning his approach, the methods advocated by Taguchi have been successfully applied to many real-life situations even it has unavoidable weaknesses (Nair et al. 1992, Myers and Montgomery 2002, Montgomery 2005). Response surface methodology (RSM) is one of the statistical approaches employed for robust design. RSM has become an important technique in the industrial world, particularly in process and product development. RSM is a collection of statistical and mathematical techniques useful for designing the statistical experiments, developing the regression models, and optimizing a process or product. The designer can apply RSM on new products or processes, as well as on existing ones.

RSM is an ideal approach to solve the robust design problem and optimize a product or process. It has many advantages over Taguchi's methods of experimental design and data analysis. However, many companies or engineers in the industry world do not fully understand the better approaches coming from the academic field and are still using Taguchi's methods, even though academia has discussed and reviewed robust design for over 20 years and statisticians have repeatedly compared Taguchi's methods with other new approaches (including RSM). The statisticians should focus on training people working in industry to clarify the misunderstanding and close the gap between academia and industry.

The dual response surface approach estimates two response models, one for the process mean, and another one for the process variance. The combined array design combines the design and noise variables into one single design, while Taguchi's crossed array includes two designs: one is inner array design consisting of the control variables, and the other is outer array design containing the noise variables. Dual RSM and combined array design are the most important RSM tools used for robust design. In this thesis, we apply dual RSM approach to construct the constrained optimization, and employ a variety of combined array designs, such as central composite design (CCD), Box-Behnken design (BBD), and face-centered cube (FCC), to fit the second-order regression models. Furthermore, we investigate functionality of the noise factors, consisting of external noise factors and internal noise factors. Whereas the external noise factors are the environmental variables that we cannot control, the internal noise factors represent the random variation of control variables due to the deviations of different components or manufacturing processes.

Parameter design is a major part of robust design, but the importance of tolerance design should not be ignored. Tolerance design allocates reasonable limits for the optimal parameters to balance the design and manufacturing applications. The goal is to optimize the parameters and tolerances of the control variables simultaneously. In general, the optimization is formulated as minimizing the total cost that includes two competing parts: the so called loss-of-quality cost and production cost. We can also use response mean, response variance, loss-of-quality cost, or production cost as objective function and others as constraints.

1.3 Objectives of Research

In this thesis, the following objectives will be fulfilled:

- The primary objective of this research is to determine appropriate statistical design of experiments for the combined array design to solve a robust design problem. The possible designs include factorial design (2^k) or fractional factorial design (2^{k-p}), central composite design (CCD), Box-Behnken design (BBD), face-centered cube (FCC), and mixed resolution design (MRD).
- The second objective is to develop regression models of the responses in terms of the design variables (control and noise variables). The models of process mean and variance are very important for data analysis.
- The third objective is the development of design optimization to obtain the optimal parameters and tolerances while improving conformance of the responses and reducing the total cost.
- After a general robust design framework is developed to apply the response surface methodology and constrained optimization method to solve practical robust design problems, we discuss some advanced topics on robust design.

1.4 Summary and Overview of Research

This chapter introduces the importance of robust design and response surface methodology to reduce variation and identifies the objectives of the research. Chapter 2 presents the literature review of the techniques and the application of robust design. The review covers the area of response surface methodology, statistical design of experiments, robust parameter design, robust tolerance design, and nonlinear programming. In Chapter 3, response surface methodology and statistical design of experiments are discussed in detail. Particularly, two important approaches of RSM, dual RSM and combined array design, are presented with practical examples. Generalized linear model (GLM) approach and design construction methods are discussed as well. In Chapters 4 some advanced topics are investigated, such as the functionality of the noise factors (external noise factors and internal noise factors), quality loss functions for single- and multiple-response, and nonlinear programming method for multiple responses. Chapter 5 presents two cases study to solve robust design problems. Summary of the research and plan for future research are provided in Chapter 6.

Chapter 2
Literature Review

Robust design, e.g. robust parameter design (RPD) and robust tolerance design, was pioneered by Genichi Taguchi in the 1980s (Taguchi and Wu (1985), and Taguchi (1986, 1987)). As an efficient and cost-effective engineering approach, it draws tremendous interest and attention among scientists and engineers. After three decades of development, robust design, particularly robust parameter design, has become a useful collection of tools to improve quality in diversified research areas and industrial applications. In this chapter, background information and literature review will be presented on Taguchi's philosophy and techniques, modern statistical design of experiments (SDE), response surface methodology (RSM), robust parameter design and tolerance design, and constrained optimization through nonlinear programming methods.

2.1 Taguchi's Philosophy and Techniques

After being introduced to the United States in the 1980s and having successful application in industry, Taguchi's techniques laid foundations and provided a philosophical basis for robust design. Many Western statisticians have reviewed and criticized Taguchi's approach to identify its weaknesses and limitations (Box 1988, Nair et al. 1992, Myers and Montgomery 2002, Montgomery 2005). Taguchi proposes three stages of robust design: system design, parameter design, and tolerance design. The goals of Taguchi's experimental design can be summarized as designing robust products or processes that are insensitive to environmental conditions (external noise factors), developing robust products that are insensitive to component variation (internal noise factors), and minimizing variation around a target value. The most important parts of Taguchi's philosophy are the reduction of variability and minimization of nonconformance cost. They are consistent with the modern continuous quality improvement philosophy.

Taguchi's design strategies include crossed or orthogonal arrays where the inner array that consists of the control variables **x**, is crossed with the outer array that contains the noise variables **z**. The inner or outer array is a factorial or fractional factorial design, and the outer array is crossed with every combination of the inner array. The orthogonal arrays are denoted by $L_r(m^k)$, where *r* represents the number of runs in the array, *k* is the number of variables, and *m* is the number of levels of each variable.

Consider a pull-off force experiment of a connector described by Byrne and Taguchi (1987) and Myers and Montgomery (2002). The experiment is a standard Taguchi parameter design that consists of four control factors (A, B, C, and D) and three noise factors (E, F, and G). The control factors are at three levels and the noise factors are at two levels. Since the inner array is a $L_9(3^4)$ design and the outer array is a $L_8(2^3)$ design, the total number of runs is $9 \times 8 = 72$. Table 2-1 presents the inner and outer array for the Taguchi parameter design. The inner array is a 3^{4-2} fractional factorial design and the outer array is a 2^3 factorial design.

Table 2-1 Taguchi Parameter Design

run	Inner Array $L_9(3^4)$				Outer Array $L_8(2^3)$									Responses	
					E	-1	-1	-1	-1	+1	+1	+1	+1		
					F	-1	-1	+1	+1	-1	-1	+1	+1		
	A	B	C	D	G	-1	+1	-1	+1	-1	+1	-1	+1	$\bar{y}$	SNR
1	-1	-1	-1	-1		y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	y_{18}	$\bar{y}_1$	SNR_1
2	-1	0	0	0		y_{21}	y_{22}	y_{23}					y_{28}	$\bar{y}_2$	SNR_2
3	-1	+1	+1	+1		y_{31}	y_{32}	y_{33}					y_{38}	$\bar{y}_3$	SNR_3
4	0	-1	0	+1		y_{41}	y_{42}	y_{43}					y_{48}	$\bar{y}_4$	SNR_4
5	0	0	+1	-1		y_{51}	y_{52}	y_{53}			…		y_{58}	$\bar{y}_5$	SNR_5
6	0	+1	-1	0		y_{61}	y_{62}	y_{63}					y_{68}	$\bar{y}_6$	SNR_6
7	+1	-1	+1	0		y_{71}	y_{72}	y_{73}					y_{78}	$\bar{y}_7$	SNR_7
8	+1	0	-1	+1		y_{81}	y_{82}	y_{83}					y_{88}	$\bar{y}_8$	SNR_8
9	+1	+1	0	-1		y_{91}	y_{92}	y_{93}					y_{98}	$\bar{y}_9$	SNR_9

The inner array design 3^{4-2} can accommodate linear and quadratic terms in each control variable, but there are no degrees of freedom left for estimating interactions between the control variables. Taguchi claims that it is possible to eliminate the control-by-control interactions either by correctly specifying the response and design factors or by using a sliding setting approach to choose factor levels. However, it is difficult to implement these two approaches unless we have a high level of process knowledge, which is unlikely (Montgomery 2005). The lack of consideration for the possible interactions between the control factors is a major drawback of Taguchi's method. Another concern of the Taguchi approach for parameter design is that the crossed array structure usually leads to a very large experiment as the number of design variables increases. In our example, 72 tests should be carried out by the cross array design. However, the combined array design, if we run all design variables at two levels, only needs 32 runs to conduct the experiment, and main effects and two-factor interactions between the control factors can be estimated without aliasing with any other main effects and two-factor interactions. This design meets the general requirement to consider the important control-by-control interactions in the regression model. The combined array design is 2_{IV}^{7-2}, and the design generators are $F = ABCD$ and $G = ABDE$. All the first-order main effects and two-factor interactions can be estimated, except that the control-by-noise interactions CE, CF, and CG are aliased with noise-by-noise interactions FG, EG, and EF, but in actual applications we often do not consider the interactions between the noise factors.

In addition, Taguchi suggests that we summarize the result of a response with the signal-to-noise ratio (SNR). In the analysis we treat the SNR as the response of the system. SNR is derived from a quadratic loss function and defined in three metrics: target-is-best, the-larger-the-better, or the-smaller-the-better. The goals of the three types of SNR are the same as their names: to achieve the nominal value, to maximize the response, and to minimize the response.

1. Target-is-best

$$SNR_T = 10\log(\frac{\bar{y}^2}{S^2})$$

2. The-larger-the-better

$$SNR_L = -10\log(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2})$$

3. The-smaller-the-better

$$SNR_S = -10\log(\frac{1}{n}\sum_{i=1}^{n} y_i^2)$$

Box (1988) and Nair et al. (1992) present excellent discussions on the limitations of the signal-to-noise ratios (SNR) and the analysis of these responses. Taguchi tries to use the SNRs as performance criteria while considering the process mean and variance. However, the SNRs usually confound location and dispersion, so the mean and variance contributions to the SNRs are confounded and we cannot analyze and evaluate them separately. We will illustrate in Chapter 3 that separate models for the process mean and variance, which are achieved through dual response surface approach, will provide a better understanding of the process.

Taguchi advocates the main-effects-only analysis through marginal means graphs, which plot SNR and $\bar{y}$ against the levels of the control factors. The decision can be made by "pick the winner" analysis. However, the approach of main effect means is appropriate only under the assumption that there are no interactions among the control variables. Obviously, in many cases, this assumption is not correct and the marginal means plots are misleading. Therefore, optimum solution cannot be guaranteed, even though a confirmation experiment, which is recommended by Taguchi, is run under inappropriate conditions.

Although Taguchi's methodology generated much discussion and debate, no one can deny the importance of his philosophy and his significant contributions to quality engineering. He advocates the robust design and extends the application of industrial statistics successfully to more industrial and manufacturing areas. However, his methods of experimental design and data analysis have critical weaknesses, and should be replaced with better approaches, such as the response surface methodology, in different research areas and industry applications.

2.2 Statistical Design of Experiments (SDE)

Statistical design of experiment (SDE), as a basis for response surface methodology, is an important engineering approach for robust design. It uses statistical techniques to plan the experimental strategy, decide the data analysis methods, and draw objective conclusions. According to Montgomery (2005), there have been four eras in its modern development. The first era, agricultural era, was led by Sir Ronald A. Fisher in the 1920s and early 1930s. He developed the three basic principles of experimental design: randomization, replication, and blocking. His major contributions include the factorial design concept and the analysis of variance (ANOVA), though they were particularly applied in the agricultural area.

The second era, namely industrial era, was pioneered and led by George Box through the development of response surface methodology (RSM). The origin of RSM is often attributed to the research done by Box and Wilson (1951). Over the next 30 years, RSM was widely used in research and development work in the chemical and process industries. However, the researchers did not realize that the statistical design of experiments could be applied in many other industries, such as in engineering and manufacturing processes.

In the 1980s, Genichi Taguchi advocated the robust design philosophy to make products or processes insensitive to the environmental noise factors and components deviation, determine the optimal values of the control factors, and minimize the variation of the responses. Robust design and statistical experimentation became essential tools for quality improvement in modern society, and were widely used in various industries, such as electronics, semiconductors, automotive, and aerospace manufacturing.

We are now in the fourth era of the statistical design of experiments. Many new approaches that are more effective and efficient than Taguchi's methods, have developed and applied successfully in the industrial world that includes almost all science and engineering areas, and even in financial and service fields.

The most widely used experimental designs for process optimization and improvement are two-level factorial and fractional factorial designs and other response surface designs, e.g. spherical designs or cuboidal designs. Montgomery (2005), Myers and Montgomery (2002), and Box and Draper (1987) present detailed information on the experimental designs.

2.3 Response Surface Methodology (RSM)

Montgomery (2005) defines RSM as "a collection of mathematical and statistical techniques useful for the modeling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize this response". The origin of RSM is often attributed to the research by Box and Wilson (1951), which changed dramatically the way that engineers, scientists, and statisticians approached industrial experimentation (Myers et al. 2004). A thorough discussion of response surface methodology (RSM) and its application is provided by Myers and Montgomery

(2002). There have been four extensive reviews of response surface methodology given chronologically as: Hill and Hunter (1966), Mead and Pike (1975), Myers, Khuri, and Carter (1989), and Myers et al. (2004).

The response surface approach is based on the assumption that

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon \tag{2-1}$$

where

y - response

f - unknown true response function

$x_1, x_2, \ldots, x_k$ - controllable input variables

ε - statistical error

For robust design problem, we can assume the response model involving control and noise factors as

$$y = f(\mathbf{x}) + h(\mathbf{x}, \mathbf{z}) + \varepsilon \tag{2-2}$$

where $f(\mathbf{x})$ consists of the control factors and $h(\mathbf{x}, \mathbf{z})$ involves noise factors and the interactions between noise and control factors.

RSM is a sequential experimental process that includes three design phases. The first phase is called a screening experiment that is designed to find the important factors and reduce the number of design variables. It is the preliminary step to make the subsequent experiments more applicable and efficient. The second phase uses the steepest ascent method mainly on first-order response surface models to optimize the process. The goal is to move the response toward the optimum by adjusting the important control variables. When the process is near the optimum, we can start the third phrase design to determine the optimum conditions for the process. Second-order response surface models will be used in the third phase to introduce curvature into the response function and obtain accurate approximations.

Myers and Montgomery (2002) summarized that RSM can solve three categories of industrial problems: fitting a response surface model over the region of interest, optimizing the quality characteristics and the process, and selecting operating conditions to meet specifications or customer requirements. RSM is applied in various areas, such as the semiconductor, electronic, automotive, chemical, pharmaceutical, financial, and service industries. Particularly, manufacturing has been achieving extensive quality improvement through the development of RSM and robust design. Two important RSM approaches for robust design are combined array designs and dual response surface approach, which will be explained next.

2.3.1 Combined Array Approach

The combined array approach is an efficient and widely used alternative to Taguchi's crossed array design. A combined array design introduces one array design that considers both control factors **x** and noise factors **z**. The combined array approach will vastly reduce the number of runs and can increase design accuracy by considering every interaction between the control factors. It deals with a single response model and is the basis for other response surface approaches, such as the dual RSM and generalized linear model. Welch et al. (1990) demonstrated the efficiency of the combined array approach for robust design. Myers, Khuri, and Vining (1992) illustrated that the single array design could be used for determining dual response surfaces.

The following example, which is taken from Myers and Montgomery (2002), illustrates the superiority of the combined array design over the crossed array design. Suppose there are four control factors x_1, x_2, x_3, and x_4, and two noise factors z_1 and z_2. One potential crossed array gives a total of 32 runs:

$$2_{IV}^{4-1} \times 2^2 \quad \text{(crossed array design)}$$

For a first-order regression model, the total 31 degrees of freedom include six main effects (x_1, x_2, x_3, x_4, z_1, and z_2), 12 two-factor interactions (control-by-control x_1x_2, x_1x_3, x_1x_4, control-by-noise x_1z_1, x_2z_1, x_3z_1, x_4z_1, x_1z_2, x_2z_2, x_3z_2, x_4z_2, and noise-by-noise z_1z_2), and 13 high-order control-by-noise interactions. Notice that other three control-by-control interactions x_2x_3, x_2x_4, and x_3x_4 cannot be considered in the design.

An alternative combined array design is 2_{VI}^{6-1} with 32 runs, and the defining relation is

$$I = x_1x_2x_3x_4z_1z_2 \quad \text{(combined array design)}$$

This combined array design estimates all six main effects and all 15 two-factor interactions for a first-order regression model. (The quadratic terms of control variables can also be easily included if we assume a second-order regression model.) Therefore, we have three important control-by-control interactions at the expense of three usually less important high-order control-by-noise interactions. The combined array design is more flexible and appropriate, particularly when the control-by-control interactions are important for the design, and this is the major drawback of Taguchi's crossed array design.

2.3.2 Experimental Designs for Combined Array

While two-level factorial and fractional factorial designs are the basis for the RSM and robust design, there are many other attractive experimental designs used in combined array design to fit both first-order and second-order models: for example, Plackett-Burman design, central composite design (CCD), Box-Behnken design (BBD), small composite design (SCD), face-centered cube (FCC), hybrid designs, mixed resolution designs (MRD), and other computer generated designs.

Box and Wilson introduced the concept of central composite design (CCD) in 1951. It is the most popular design of second-order models (Myers and Montgomery, 2002). In general, CCD consists of three parts: a two-level factorial (2^k) or fractional factorial design (resolution V), 2k axial or star points, and n_c center runs. Being the same as other RSM tools, CCD can be used as part of sequential design experiments. The factorial points contribute toward the estimation of linear and two-factor interaction terms, while the axial points are chosen based on the region of interest and region of operability and are mainly used to estimate quadratic terms (but not interaction terms). The center points provide information about the existence of curvature, and are used to estimate the pure error and quadratic terms of the model.

Mixed resolution design (MRD) is a relatively new kind of design. Borror (1998) proved that the mixed resolution design is the most efficient design with desirable variance properties and appears to be a very promising design strategy for problems that involve a large number of variables. Borkowski and Lucas (1997) discussed the optimality properties of mixed resolution design and provided a catalog of design construction. Borror and Montgomery (2000) compared the mixed resolution design with a typical inner/outer array design and concluded that the mixed resolution design is efficient and can lead to a useful prediction model involving both control and noise factors.

2.3.3 Dual Response Surface Approach

The dual response surface approach, which examines two response surfaces, one for the process mean and another one for the process variance, was first introduced by Myers and Carter (1973). Vining and Myers (1990) applied the dual response surface approach to robust design. They present robust design problems as constrained optimization problems that optimize the primary response under the constraint of the secondary response. The variance and mean can be formulated as primary and secondary response, respectively, or in another way, variance as secondary and mean as primary.

Myers, Khuri, and Vining (1992) obtained the process mean and variance model by applying unconditional expectation and variance operator to the response model. Myers and Montgomery (2002) introduced another method to estimate the process variance: the slope of the response surface in the direction of the noise variables. It is obvious that the larger the vector of slope, the larger the process variance. A third method to develop variance model, the delta method, is given by Rice (1995) and Myers and Montgomery (2002). The delta method is useful for complex models. It can be expressed by the so-called transmission-of-error formula:

$$Var_z[y(\mathbf{x},\mathbf{z})] = \sigma_z^2 \sum_{i=1}^{r} \left[\frac{\partial y(\mathbf{x},\mathbf{z})}{\partial z_i} \right]^2 + \sigma^2 \tag{2-3}$$

which is found by expanding y(**x, z**) as a first-order Taylor series about z = 0 and applying the variance operator. In the transmission-of-error formula, σ_z^2 is the variance of noise variables and σ^2 is the variance of residuals.

Lin and Tu (1995) proposed an alternative to the standard optimization of the dual response approach. Instead of optimizing the primary response while subject to a constraint of the secondary response, they use the mean square error (MSE) to combine the primary response and secondary response as a single objective function

$$MSE(y) = (\hat{\mu}_y - T)^2 + \hat{\sigma}_y^2 \tag{2-4}$$

where $\hat{\mu}_y$ and $\hat{\sigma}_y^2$ are the estimated mean and variance, and T is the target value of the response. As this method allows bias in the response, Copeland and Nelson (1996) improved it by adding a constraint on the mean value of the response to restrict the bias.

Giovagnoli and Romano (2008) introduced a modified dual response surface approach, which stochastically simulates the noise factors when their probabilistic behaviour is known. External noise variables are totally out of the designer's control, but internal noise variables, which represent random deviation in design parameters due to part-to-part variation induced by uncontrollable manufacturing errors, are partially controllable through specifying the tolerances of the control factor. Therefore, they use the standard deviations of internal noise factors as additional control factors in the experiment and simulated them accordingly. In this thesis, we will discuss the functionality of the noise factors in the following chapters.

Kim and Lin (1998) proposed a fuzzy model for dual response surface optimization. A membership function in fuzzy set theory is used to measure the experimenter's degree of satisfaction concerning the mean and standard deviation responses.

In many robust design problems, the noise variables are assumed to be continuous. Brenneman and Myers (2003) considered robust design problems where the noise factors are categorical, for example, different suppliers. They proposed the use of a multinomial distribution, and discussed the impact that the assumptions for continuous and categorical noise variables have on the robust settings and on the overall process variance estimate.

In summary, the combined array approach and the dual response approach are superior to Taguchi's methods for addressing the robust design problems in three areas: less experiment runs are needed, important control-by-control interactions are considered and other unimportant terms are ruled out, and the summary measures are more practical than signal-to-noise ratios.

2.3.4 Generalized Linear Models

Nelder and Lee (1991) and Myers, Khuri and Vining (1992) were the first to suggest applying generalized linear models (GLM) to analyzing Taguchi's types of experiments and modeling the variance in robust design. GLM is very useful in the dual RSM modeling for robust design. Myers and Montgomery (2002) defined GLMs as a unified collection of models that accommodate response distributions that follow the exponential family, such as the normal, Poisson, binomial, exponential, and gamma distribution. They point out that even if the basic response is normal, the modeling variance, either from replication or from non-replication experiments, is important to use with the GLM. In robust design, in which nonhomogeneous variance is common, GLM is used as an important tool for variance modeling. Lee and Nelder (2003) discussed the connections between robust parameter design and GLM, and illustrated the advantages of using GLM versus data transformation.

2.4 Robust Tolerance Design

Tolerance design specifies limits on the parameters of the control variables. When researchers conduct robust design on their systems, they mainly focus on parameter design, partially because the cost information for tolerance design is not straightforward. However, tolerance design is an important portion of robust design and its results will affect the target values of the parameters. Therefore, more studies are needed to generalize the design method.

Meng (2006) studied the drawback of the traditional robust design that carries out two-stage sequential parameter and tolerance design. It is assumed that the nominal values of the control factors obtained from the parameter design remain optimum for the subsequent tolerance design. However, if the magnitudes of the variation of the internal noise factors change in the tolerance design stage, the interactions between the control factors and noise factors most likely will change accordingly. As a result, the parameter setting will also most likely be different. Therefore, Meng proposed to conduct sequential parameter and tolerance design in the first round, and then another round, or even more rounds, of parameter design should be done based on the optimal parameter setting and variability setting from the first round design. He showed this method can achieve better optimum settings for the parameters. This is a modified Taguchi approach and extra resources are required.

Li and Wu (1999) proposed the single-stage integrated parameter and tolerance design and suggested this new approach is superior to Taguchi's two-stage sequential approach. They also point out a modified Taguchi approach, which performed the two-stage design iteratively until the tolerance value stops change, requires extra time and material. The new method is flexible enough to accommodate different cost structures for the component tolerances. Assuming that control factors **x** follow normal distribution $N(0,\sigma^2)$, it is common to define the tolerance *t* of control factors **x** to be

$$t = \pm 3\sigma \tag{2-5}$$

The relationship between (σ_y, μ_y) and (x_{i0}, t_i), where x_{i0} and t_i are the mean and tolerance for each control variable, is obtained by a first-order and second-order Taylor series expansion

$$y = f(x_1, x_2, \ldots, x_n)$$
$$\approx f(x_{10}, x_{20}, \ldots, x_{n0}) + \sum_{i=1}^{m} d_i (x_i - x_{i0}) + \sum_{i=1}^{m} d_{ii} (x_i - x_{i0})^2 \tag{2-6}$$

$$\sigma_y^2 \approx \sum_{i=1}^{m} d_i^2 \sigma_i^2 = \frac{1}{9} \sum_{i=1}^{m} d_i^2 t_i^2 \tag{2-7}$$

$$\mu_y \approx f(x_{10}, x_{20}, \ldots, x_{n0}) + \sum_{i=1}^{m} d_{ii} E((x_i - x_{i0})^2) \tag{2-8}$$

where $d_i = \frac{\partial f}{\partial x_i}\Big|_{x_i = x_{i0}}$ and $d_{ii} = \frac{1}{2} \frac{\partial^2 f}{\partial x_i^2}\Big|_{x_i = x_{i0}}$ are the first and second derivative of the function f evaluated at x_{i0}.

Romano et al. (2004) presented a general framework for the multiple response robust design problem when data are collected from a combined array design. Within their framework, both robust parameter design and tolerance design are handled in an integrated way. The optimization criterion is based on a single value in terms of a quadratic quality loss and producer loss, and it is selected in order to consider both statistical information and economic information related to the product or process.

2.5 Data Analysis and Optimization

Contour plots provide one of the most straightforward and effective ways to illustrate and analyze the response surface system. While response surface plot shows the relationship between the response and the design variables, contour plot is a two-dimensional or three-dimensional graph that shows the contours of constant response versus the design variables. The response surface plot and contour plot can be used together in the robust design analysis. We must bear in mind that contour plots are only used for estimation when two or three design variables exist. The more accurate nonlinear programming methods are available for performing formal optimization. Particularly, when there are more than three design variables, the contour plot method becomes awkward, as we could only select two variables in the two-dimensional plot and other variables must be held constant.

Giovannitti-Jensen and Myers (1989) and Myers et al. (1992) developed the variance dispersion graph (VDG), which plots the prediction variance information for the entire design region on a two-dimensional graph. From the VDG of a RSM design, a researcher has a “snapshot” regarding the stability of the prediction variance and has an alternative criterion to select the designs (Myers and Montgomery, 2002). The VDG illustrates a spherical design, such as a center composite design (CCD) and Box-Behnken design (BBD), with three graphic portions:

1. The spherical average prediction variance

$$V^r = \frac{N\Psi}{\sigma^2}\int_{U_r} Var[\hat{y}(x)]dx \tag{2-9}$$

where U_r implies integration over the surface of a sphere of a radius r, $\hat{y}(\mathbf{x})$ is the estimated response, and $\psi = (\int_{U_r} dx)^{-1}$.

2. Maximum prediction variance

$$\max_{x \in U_r}\left[\frac{NNar[\hat{y}(x)]}{\sigma^2}\right] \tag{2-10}$$

3. Minimum prediction variance

$$\min_{x \in U_r}\left[\frac{NNar[\hat{y}(x)]}{\sigma^2}\right] \tag{2-11}$$

VDG has been extended to other design scenarios, such as cuboidal designs. Myers et al. (1992), Borror (1998), and Myers and Montgomery (2002) illustrated how one can construct the plots and use VDGs to compare and evaluate response surface designs. Borror, Montgomery, and Myers (2002) developed a VDG methodology for designs for robust parameter designs that incorporate both control and noise variables. They produce VDGs for both the mean and the slope of the response surface model.

Vining and Myers (1990) first presented robust parameter design problems as constrained optimization problems. They used the dual response optimization technique, which optimizes the primary response under the constraint of the secondary response, to solve the problem. Del Castillo and Montgomery (1993) pointed out that the technique of Vining and Myers (1990) does not always produce local optima and therefore proposed the use of standard nonlinear programming techniques, specifically, the generalized reduced gradient (GRG) algorithm, to determine the optimal operating conditions. It is shown that this method can be more flexible and give better solution within the region of interest. There are many ways to use nonlinear programming techniques to formulate and solve single- or multiple-response optimization problems. Many software packages are available to accommodate a constrained optimization problem.

Fathi (1991) also used nonlinear programming techniques to solve the parameter design problem, but assumed that the functional relationship between the input parameters and the performance characteristic of interest is either known or could be well approximated. The specific implementation of the algorithmic strategy is based on conventional optimization techniques.

To simplify the robust design problem, many researchers usually deal with a single important response or quality characteristic. In many practical instances, however, multiple responses should be optimized simultaneously. Because the responses might be correlated, trade-offs among multiple responses should be studied to obtain the overall optimum setting of parameters where all responses are optimized or at least in desired ranges. Similar to single response optimization, contour plots can be used to estimate an optimum. Lind, Goldin, and Hickman (1960) illustrated this approach when examining the overlay of the contour plots for the responses. In general, different combinations of the design variables can result in different optimum conditions of the process, so the experimenter should choose the appropriate operating conditions from feasible possibilities using qualitative criterion.

Since the graphical approach is limited to two or perhaps three dimensions, more formal optimization methods should be developed for multiple quality characteristics. Derringer and Suich (1980) proposed the desirability function approach to simultaneously optimize multiple responses. This optimization technique uses direct search methods to find the optimum solutions and will obtain multiple optimum results. Another popular approach is to formulate the multi-response problem as a constrained optimization problem:

$$\begin{gathered} Min\ (or\ Max) \quad \hat{y}_1 \\ Subject\ to \\ l_i \le \hat{y}_i \le u_i,\ i = 2, 3, ..., m \\ x \in R \end{gathered} \tag{2-12}$$

where one response is the objective function and others are constraints. Two classes of nonlinear programming methods available in many software packages can be used to solve this kind of problem: direct search methods and mathematical optimization algorithms (such as the generalized reduced gradient method).

The optimization problem can be formulated in many other ways. Del Castillo (1996) presented a methodology for analyzing multi-response experiments that allows one to obtain optimal solutions that simultaneously satisfy confidence region constraints for all responses. The methodology consists of computing confidence regions for stationary points of quadratic responses and confidence cones for the direction of maximum improvement for linear responses. The stationary points are constrained to lie within the experimental region. The author showed that the confidence region is dependent upon the value of the Lagrange multiplier of the region's constraint. The value of the Lagrange multiplier is found by solving the Karush-Kuhn-Tucker (KKT) optimality conditions.

Another method to formulating the multiple-response optimization problem is the weighted sum method, which simultaneously optimizes all the responses

$$\begin{gathered} Min\ (or\ Max) \quad \sum_{i=1}^{m} w_i \hat{y}_i,\ i = 1, 2, ..., m \\ Subject\ to \\ x \in R \end{gathered} \tag{2-13}$$

where the w_i are weight coefficients. The drawback for this method is the difficulty in selecting the appropriate w_i.

All the methods we have discussed so far assume that the responses are independent or uncorrelated. When this assumption is inappropriate, other methods should be considered to solve the optimization problems. Vining (1998) extended the univariate quadratic loss function to a multivariate form based on the square error loss approaches. His loss function is

$$L = [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] \tag{2-14}$$

This method allows the experimenter to specify the directions of economic importance for the compromise optimum, when the variance-covariance structure of the responses is considered. Papers discussing the squared error loss approach include those by Khuri and Conlon (1981), Pignatiello (1993) and Ames et al. (1997).

When the objective function is discontinuous, stochastic, or highly nonlinear, standard optimization algorithms might not be appropriate to solve the problem, particularly for the multi-response problem. Genetic algorithms (GA) do not impose any continuity or smoothness demands on the objective function, nor are they deterred by discontinuities in the feasible space or the type of decision variables involved. Genetic algorithms put together dissimilar blocks of a solution until a combination satisfies the imposed requirements. The methods are based on natural selection, the process that drives biological evolution, to solve constrained and unconstrained optimization problems. The genetic algorithms use three main types of rules at each step to create the next generation from the current population: selection, crossover and mutation. Heredia-Langer (2004) discussed the use of genetic algorithms for the construction of model-robust experimental designs.

2.6 Reliability-Based Robust Design (RBRD)

The data driven method of robust design that we have discussed is a model-based approach that creates a mathematical model of the product or process such that the responses are functions of the design variables. While we focus on the response surface methodology and statistical design of experiment in this thesis, an alternative model-based approach is to use the reliability-based robust design (RBRD). RBRD is an integration of reliability-based design optimization (RBDO) and robust design to deal with uncertainties and optimize a process or product. The goal of RBRD is same as the statistical approaches we discussed before, and RBDO is a reliability-based (or conformance-based) design, which usually uses the reliability index β to express the probability of conformance. While we use RSM to optimize parameters and tolerances, reliability-based design optimization (RBDO) can also be used to as an effective alternative after the mechanistic or empirical model of the system is available. The RSM-based RBRD is a relatively new method to solve robust design problems through integration of RSM and RBRD.

Due to the inherent uncertainty and variability of the system (a product or process), the traditional deterministic optimal design cannot lead to effective and reliable results if we assume zero variability and do not consider uncertainty. Probabilistic uncertainty analysis has become an essential part of design and decision making process under uncertainty and risk. Uncertainties are unavoidable in a process or product, so variability and performance of the responses should be considered to ensure reliability and quality. Uncertainty can be classified into reducible and irreducible types. Irreducible uncertainty is a property of the system itself and describes the inherent randomness (variation) associated with a physical system or environment. It cannot be reduced even we have more knowledge or information about the system. Probabilistic uncertainty analysis is used to obtain the distribution of a response Y or the probability of $Y \le y$ given the distributions of random inputs $X = [X_1, X_2, ..., X_n]'$ and $Y = g(X)$. On the other hand, reducible uncertainty results from some level of ignorance or incomplete information about a system, and can be modeled by probability theory, or other theories such as evidence theory, possibility theory, and fuzzy sets. (Huang and Du, 2008).

Reliability-based design optimization can be integrated with RSM to perform optimization. There are three types of simulation methods to analyze probabilistic uncertainty: (1) sampling-based methods, such as Monte Carlo simulation that makes use of samples of random numbers; (2) second moment methods, like first-order second moment (FOSM), and (3) most probable point (MPP) reliability-based methods, such as the first-order reliability method (FORM). Huang and Du (2008) pointed out that the reliability-based design methods, such as the first-order reliability method (FORM), have the advantages of satisfactory accuracy and moderate computational cost compared with the other two kinds of simulation and approximate method. The total cost of the product or process can be used as the objective function of RBRD and it consists of two kinds of competing costs: the production cost and the so called lost-of-quality cost. RBDO includes two optimization processes: reliability (probabilistic uncertainty) analysis in the independent and standard normal random U-space, and design optimization in the original random V-space (Youn and Choi, 2004).

Chapter 3
Response Surface Methodology and Optimum Design

3.1 Introduction

Response surface methodology (RSM) is a superior alternative to Taguchi technique, and it works successfully to solve robust design problems. Two important approaches that contribute most to robust design are combined array designs for design construction and dual response surface approach for modeling and optimization formulation. RSM places both control and noise variables in a single experiment to avoid the disadvantages of the inner and outer array structure. We call this kind of design a combined array design, which results in smaller experiment comparing to crossed array design, and we can accommodate the important control-by-control and control-by-noise interactions conveniently in the response surface model. The dual response surface approach is a natural link between robust parameter design and constrained optimization. It provides an estimate of the mean and variance in two models. The mean model and variance model for the response can be chosen either as the primary or secondary response, respectively, and the goal is to optimize the primary response under some constraints on the secondary response. Thus, we have a variety of constrained optimization formulations in different scenarios, such as “the target is best”, “the larger the better”, and “the smaller the better” for the response.

3.2 System Design

Robust design is very useful for designing a product or process. The goal of a design is to develop high quality products or processes with low cost, so the product performance is insensitive to the variation of raw materials, manufacturing process, and external environment. Commonly we refer to products or processes as a system. System design is the conceptual design phrase, which proposes the principal system configuration based on existing knowledge, experience, and resources, to accomplish the desired functions and reliability of the system. A good system design is necessary for the success of subsequent design stages. In system design, the input (such as components or materials), process, control and noise factors, and output (one or more response variables) should be determined and planned. Their mechanistic relationship and realization methods should also be decided.

Consider an example to design a temperature controller described by Phadke (1989). McCaskey and Tsui (1997) and Savage (2008) discussed the functioning of the circuit in detail. Figure 3-1 shows the proposed temperature controller circuit. The circuit has four components: power supply, Wheatstone bridge, comparator, and relay. The relay switch turns the heater on and off and this action is operated by the comparator that sets its output voltage to either ground or E_0 according to the input voltage difference at points 1 and 2. This input voltage difference is established by the Wheatstone bridge comprising the thermistor R_T - that follows the room temperature - and the remaining three resistors (R_1, R_2, R_3) that set the desired temperature.

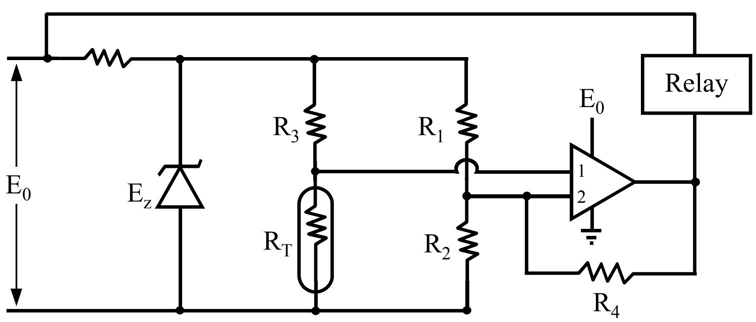

Figure 3-1 Temperature controller circuit

As the heater operates, the room temperature increases. When the temperature rises above a certain level, R_T drops below a threshold value R_{T-OFF}, so that the difference in the voltages between terminals 1 and 2 of the amplifier becomes negative. The relay is then actuated to turn the heater off. With no heat input, the temperature of the room begins to decrease if we assume it is colder outside, and the value of R_T increases. At a second threshold value R_{T-ON} the difference in the voltages between terminals 1 and 2 becomes positive, and the relay is actuated to turn the heater on. The actual values of R_{T-ON} and R_{T-OFF} can change due to variation in the values of various circuit components and the effects of noises.

The robust design problem is to make sure that the heater turns on or off when the environment temperature is as close as possible to set temperature around the target even though there is uncertainty in all of the components in the circuit. To formulate the robust system design problem, we classify resistor R_3 as the signal factor to achieve a desired value for R_T, and R_{T-ON} and R_{T-OFF} are the output responses. The potential control factors are the nominal values of R_1, R_2, R_4, diode voltage E_Z, and power supply voltage E_0. Ideally, the four resistors and voltage inputs (E_Z and E_0) are deterministic. However, the components do have variability and this affects the output voltage and its ability to maintain the proper room temperature. The noise factors are the manufacturing variations (tolerances) of the control factors. Recall that this kind of noise factor is the internal noise factor.

After the system design, the problem is identified and the system configuration is clear. In the subsequent robust parameter and tolerance design, the (ideally linear) relationship between the responses, control factors, and noise factors can be obtained and used to solve the problem.

3.3 Linear Regression Model

In the practical application areas, such as manufacturing systems, it is highly unlikely that enough knowledge of physical mechanism is available. It is more often that an empirical model is used as an approximation. An empirical model is built using the multiple regression method based on observations from the product or process. The linear regression model between response **y** and design variables **X** is describe as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3-1}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \text{ and } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}$$

The least squares estimation of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3-2}$$

Thus the fitted regression model and the fitted residuals can be denoted as

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \tag{3-3}$$

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} \tag{3-4}$$

The coefficient of multiple determination can be used to measure the performance of competing response surface models

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{3-5}$$

where SS_E is the residual (error) sum of squares, SS_R is the regression sum of squares, and SS_T is the total sum of squares. From the definition we can see that $0 \le R^2 \le 1$. A rule of thumb is that any R^2 value above 75% indicates an adequate RSM fit. However, it is possible for models that have large values of R^2 to yield poor predictions of new observations or estimates of the mean response; on the other hand, models with a low R^2 may fit the data reasonable well. It is unwise to place too much emphasis on R^2 alone, or to consider it without residual analysis. The residual from the least square fit, defined by $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$, is also very important for model adequacy checking. Residual plot can help decide whether the model assumptions are valid. In general, a normal probability plot of the residuals is used to check the normality assumption. If the residuals plot approximately along a straight line, there is no apparent problem with normality. If the plot indicates problem with the normality assumption, there may be mistakes of the data or model construction, or we should transform the response as a remedial measure.

In multiple linear regression, the tests of hypotheses about the model parameters are important to evaluate the model. The analysis of variance (ANOVA) is very helpful to test the significance of regression. It presents the results of a regression analysis in an ANOVA table, which contains various summary statistics of the data (such as the composition of the total variability for the response) to formulate the F-test.

3.4 Screening Experiment

Figure 3-2 shows the general process of the screening experiment. After system design, we usually have a long list of design variables that could be potentially important to affect the performance of the system. This leads to a screening experiment that studies the factors and aims to eliminate the insignificant ones. According to sparsity of effects principle, most systems are normally controlled by a few main effects and low-order interactions, and most high-order interactions are negligible without affecting the experiment results. Therefore, the experimenter can focus on the important factors with fewer tests in the subsequent experimental steps. A screening experiment is a necessary step to conduct robust design effectively and efficiently with limited resources.

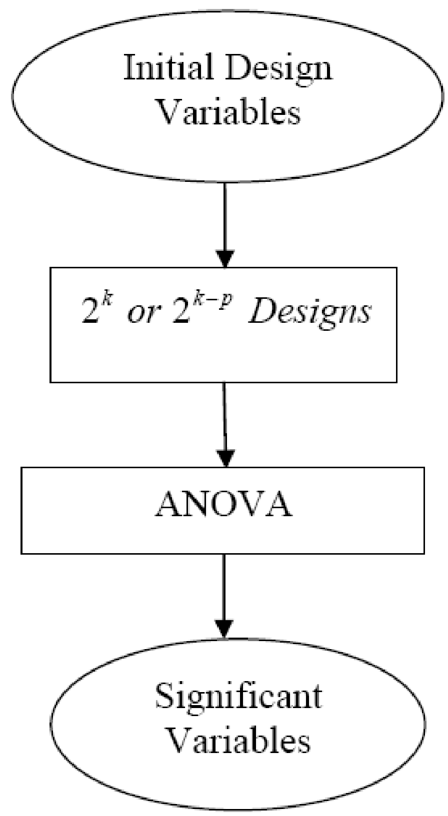

Figure 3-2 Process of screening experiment

Two-level factorial designs (2^k) or fractional factorial designs (2^{k-p}), particularly the latter, are very useful in screening experiments. As the number of design factors is large in the screening stage, the total number of treatments in a factorial design, even for only a single replicate, exceeds the available resources. For example, a full factorial 2^7 design requires 128 runs, which include 7 degrees of freedom to estimate the main effects and 21 degrees of freedom to estimate the two-factor interactions. The remaining 99 degrees of freedom are used for interactions of higher order, which are negligible based on sparsity of effects principle. Therefore, fractional factorial designs are more practical and widely used in industry.

A two-level fractional factorial design (2^{k-p}) contains k factors of interest and p independent design generators, which are also called words. The defining relation of the design includes the p generators and their $2^p - p - 1$ generalized interactions. The design generators should be selected to have the highest possible resolution for the design. A design is of resolution R if no p-factor effect is aliased with another effect containing less than $R - p$ factors. In general, the design resolution of 2^{k-p} is equal to the smallest number of letters in any design generator (word) in the defining relation. While degree of fractionation is satisfied, the higher resolution should be adopted. Design resolution III, IV, and V are very important, and Myers and Montgomery (2002) present a detailed selection on the two-level fractional factorial designs for $k \le 11$ factors and up to $n \le 128$ runs.

For example, consider a 2^{7-2} design. If we select I = ABCDF and I = ABDEG as the independent design generators, the generalized interaction is I = (ABCDF) (ABDEG) = CEFG, the complete defining relation is

$$\text{I} = \text{ABCDF} = \text{ABDEG} = \text{CEFG}$$

As the smallest number of letters in any design generator in the defining relation is four, we can conclude the design is of resolution IV (2_{IV}^{7-2}). The aliases of any effect are gained by multiplying the effect by each design generator (word) in the defining relation, and each effect has $2^p - 1$ aliases. For A and AB in our example, their aliases are

$$\text{A} = \text{BCDF} = \text{BDEG} = \text{ACEFG}$$
$$\text{AB} = \text{ACDF} = \text{ADEG} = \text{ABCEFG}$$

Example: Screening Experiment for Filtration Rate Experiment

Consider an example from Myers and Montgomery (2002) to study the filtration rate of a chemical product. The three control factors are pressure (x_1), concentration of formaldehyde (x_2), and stirring rate (x_3), and the noise factor is temperature (z_1). Each factor is present at two levels. The filtration rate is denoted as response y. We study the problem with a two-level fractional factorial design 2_{IV}^{4-1}.

Table 3-1 shows the 2_{IV}^{4-1} design with defining relation I = ABCD. The alias structure for the design is shown in Table 3-2.

Table 3-1 Filtration Rate Experiment with 2_{IV}^{4-1} Design (I = ABCD)

Run	A (z_1)	B (x_1)	C (x_2)	D (x_3) = ABC	Treatment	Filtration Rate
1	-	-	-	-	(1)	45
2	+	-	-	+	ad	100
3	-	+	-	+	bd	45
4	+	+	-	-	ab	65
5	-	-	+	+	cd	75
6	+	-	+	-	ac	60
7	-	+	+	-	bc	80
8	+	+	+	+	abcd	96

Table 3-2 Alias Structure for the 2_{IV}^{4-1} Design with I = ABCD

A = BCD B = ACD C = ABD D = ABC	AB = CD AC = BD BC = AD

The estimates of main effects and two-factor interactions are $l_A = 19$, $l_B = 1.5$, $l_c = 14$, $l_D = 16.5$, $l_{AB} = -1$, $l_{AC} = -18.5$, $l_{BC} = 19$. It is clear that A, C, D, AC, and BC are significant. As B is not significant, we may drop it as the main effect and replace the significant interaction BC with AD. The first-order regression model is

$$\hat{y}(\mathbf{x}, z_1) = 70.75 + 9.5z_1 + 7x_2 + 8.25x_3 - 9.25x_2z_1 + 9.5x_3z_1$$

Table 3-3 summarizes the analysis of variance (ANOVA) for this experiment and confirms that the effects of A, C, D, AC, and AD are significant. The model sum of squares is

$$SS_R = SS_A + SS_C + SS_D + SS_{AC} + SS_{AD} = 3065$$

and the coefficient of multiple determination is

$$R^2 = \frac{SS_R}{SS_T} = 99.8\%$$

Therefore, the regression model contributes most of the variability of the response. After the screening experiment, the model is simplified that only important control and noise variables and their interactions are kept for further optimization. Comparing with the first-order model obtained from the full replicate design in Section 3.6, the half fraction design has nearly identical estimate of the model.

Table 3-3 Analysis of Variance for Filtration Rate Experiment

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0
A	722	1	722	222.15
C	392	1	392	120.62
D	544.5	1	544.5	167.54
AC	684.5	1	684.5	210.62
AD	722	1	722	222.15
Error	6.5	2	3.25	
Total	3071.5	7		

3.5 Steepest Ascent Optimization

After the screening experiment, a first-order regression model is used to optimize the significant variables to satisfy the response requirement. The steepest ascent (or descent) method aims to find an optimal region of the response and leave the job for finding the optimal points to the further stages. It should be emphasized that quality improvement through analysis of designed experiments is usually an iterative experience. This is illustrated quite well in dealing with the strategy of steepest ascent.

Example: Steepest Descent Optimization of a Molded Die

We take an example from Myers and Montgomery (2002) to illustrate the process of steepest ascent method. The problem is to minimize shrinkage of a molded die (so it is steepest descent) by optimizing the setting of four factors: injection velocity (x_1), mold temperature (x_2), mold pressure (x_3), and back pressure (x_4). Table 3-4 shows the factor levels in natural and design units.

Table 3-4 Factor Levels in Natural and Design Units

Design Factors	Design Units	
	-1	+1
x_1	1.0	2.0
x_2	100	150
x_3	500	1000
x_4	75	120

Assume we have the first-order regression model

$$\hat{y} = 80 - 5.28x_1 - 6.22x_2 - 1.21x_3 - 1.07x_4$$

As all of the signs of the regression coefficients are minus, we will move the factors along the path of steepest descent in proportional to the magnitude of the regression coefficients. The regression coefficients of x_1 and x_2 are bigger than the ones of x_3 and x_4, so x_1 and x_2 will contribute more to the movement and one of them, say x_1, can be used to define the step size $\Delta x_1 = 1$, which is arbitrarily chosen to correspond to 0.5 in natural x_1 value. We define the step size in other *k-1* variables as

$$\Delta x_j = \frac{b_j}{b_i / \Delta x_i},\ j = 1,2,\ldots,k-1,\ i \neq j \tag{3-6}$$

Then we have other step sizes in design unit and natural unit

$$\Delta x_2 = \frac{b_2}{b_1 / \Delta x_1} = 1.178\ (\textit{design unit})$$
$$= 1.178 * 25 = 29.45\ (\textit{natural unit})$$

$$\Delta x_3 = \frac{b_3}{b_1 / \Delta x_1} = 0.23\ (\textit{design unit})$$
$$= 0.23 * 250 = 57.5\ (\textit{natural unit})$$

$$\Delta x_2 = \frac{b_4}{b_1 / \Delta x_1} = 0.203\ (design\ unit)$$
$$= 0.203 * 22.5 = 4.57\ (natural\ unit)$$

Table 3-5 presents the path of steepest descent in terms of design unit. After Base+6Δ, the experiment should be stopped as no further improvement can be achieved. After knowing the approximate optimal regions of the parameters, we can start the robust parameter design and tolerance design.

Table 3-5 Path of Steepest Descent

Step	x_1	x_2	x_3	x_4	$\hat{y}$
Base	0	0	0	0	80.00
Base+Δ	1	1.178	0.23	0.203	66.90
Base+2Δ	2	2.356	0.46	0.406	53.79
Base+3Δ	3	3.534	0.69	0.609	40.69
Base+4Δ	4	4.712	0.92	0.812	27.59
Base+5Δ	5	5.89	1.15	1.015	14.49
Base+6Δ	6	7.068	1.38	1.218	1.38
Base+7Δ	7	8.246	1.31	1.421	-11.72

3.6 Dual Response Surface Approach in Combined Array Designs

The response model containing control and noise variables can be expressed as

$$y(\mathbf{x},\mathbf{z}) = f(\mathbf{x}) + h(\mathbf{x},\mathbf{z}) + \varepsilon \tag{3-7}$$

where

f ($\mathbf{x}$) - the terms involving control variables only

h ($\mathbf{x}$, $\mathbf{z}$) - the terms involving noise variables and control-by-noise interactions

If we assume the noise variables have mean zero, variance σ_z^2, and covariance zero, the process mean is

$$E_z[y(\mathbf{x},\mathbf{z})] = f(\mathbf{x}) \tag{3-8}$$

and we can use Taylor series expansion or apply variance operator to obtain the transmission-of-error formula for the variance as:

$$Var_z[y(\mathbf{x},\mathbf{z})] = \sigma_z^2 \sum_{i=1}^{r} \left[\frac{\partial y(\mathbf{x},\mathbf{z})}{\partial z_i} \right]^2 + \sigma^2 \tag{3-9}$$

The mean and variance formulas indicate that the levels of the control variables will decide the process mean and variance, and we may solve the robust design problems by setting control variables only. The noise variables influence the response through the regression coefficients of the interactions between the control and noise variables.

In general, data from robust design experiment can be fitted by a quadratic or second-order model that would include all main linear effects of control and noise variables, control-by-control interactions, control-by-noise interactions, and pure quadratic effects of control variables. Thus the regression model can be

$$y(\mathbf{x},\mathbf{z}) = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} + \mathbf{z}'\boldsymbol{\gamma} + \mathbf{x}'\boldsymbol{\Delta}\mathbf{z} + \boldsymbol{\varepsilon} \tag{3-10}$$

where

y (**x, z**) – Response

x - $r_x \times 1$ Vector of control variables, $x' = \lfloor x_1, x_2, ..., x_{r_x} \rfloor$

z - $r_z \times 1$ Vector of noise variables, $z' = \lfloor z_1, z_2, ..., z_{r_z} \rfloor$

β_0 - Intercept

$\boldsymbol{\beta}$ - $r_x \times 1$ Vector of coefficients of first-order control factors

B - $r_x \times r_x$ Matrix of coefficients of second-order control variables and control-by-control interactions

$\boldsymbol{\gamma}$ - $r_z \times 1$ Vector of coefficients of first-order noise factors

$\boldsymbol{\Delta}$ - $r_x \times r_z$ Matrix of coefficients of control-by-noise interactions

$\boldsymbol{\varepsilon}$ - $N(0, \sigma^2)$ independent and identically distributed (i.i.d.) random errors

In many real-life applications, it is reasonable to assume the noise-by-noise interactions and second-order terms of noise variables are not significant, so they are not considered in the response model and the noise variables are assumed to be uncorrelated. Since noise factors are random variables, we can assume that the expected noise variables E (**z**) = 0 and the variance-covariance matrix of noise factors $Var(\mathbf{z}) = \sigma_z^2\mathbf{I}$. If we further assume the noise variables are centered at zero and scaled at ± 1 for $\pm\sigma_z$ in coded form, we can get $\sigma_z = 1$. From the second-order model we can have the response surface model of the process mean and variance as

$$E_z[y(\mathbf{x},\mathbf{z})] = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} \tag{3-11}$$

$$\begin{aligned} Var_z[y(\mathbf{x},\mathbf{z})] &= \sigma_z^2(\boldsymbol{\gamma}' + \mathbf{x}'\boldsymbol{\Delta})(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \sigma^2 \\ &= \sigma_z^2\mathbf{l}'(\mathbf{x})\mathbf{l}(\mathbf{x}) + \sigma^2 \end{aligned} \tag{3-12}$$

where $\mathbf{l}(\mathbf{x}) = \boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}$, σ_z^2 is variance of noise factors, and σ^2 is variance of residual errors. We notice that the noise factors are not contained in the model of process variance; however, coefficients of first-order noise factors and control-by-noise interactions play important roles in the variance model.

Notice $\mathbf{l}(\mathbf{x})$ is just the vector of partial derivatives of y(x, z), or the slope of the response surface in the direction of the noise variables

$$\mathbf{l}(\mathbf{x}) = \frac{\partial y(\mathbf{x},\mathbf{z})}{\partial \mathbf{z}}$$

And the larger the vector of the slope, the larger the process variance is. In addition, Δ, the matrix of coefficients of first-order control-by-noise interactions, should not be zero, so we have robust design problem.

We can use constrained optimization to find optimum operating condition in the formulation that the process mean is the constraint and the objective function is the process variance

$$\begin{gathered} Min \quad Var_z[y(\mathbf{x},\mathbf{z})] \\ subject\ to \\ E_z[y(\mathbf{x},\mathbf{z})] = m \\ \mathbf{x} \in R^k \end{gathered}$$

where m is the specific target value (or a specified range) of the mean response. This optimization of separate models (process mean and variance) can obtain better understanding of the process and is a superior alternative to Taguchi's data analysis approach, in which signal-to-noise ratios (SNR) are used as response variables and the experimenter seeks to maximize them. Three of SNRs are considered to be "standard" and widely applicable:

1. The target is best (SNR_T). The experimenter attempts to achieve a specific target value for the response and reduce the variability around the target value. SNR_T is to be maximized. For example, part dimension or component properties.
2. The larger the better (SNR_L). The experimenter attempts to maximize the response while maximize SNR_L . For example, durability measure.
3. The smaller the better (SNR_S). The experimenter attempts to minimize the response while maximize SNR_S . For example, cycle time and number of defects.

Table 3-6 presents a comparison for the different optimizing scenarios using Taguchi method and dual RSM to formulate the objective function and constraints. It is clear that dual RSM approach is more effective and efficient.

Another formulation of nonlinear programming is to minimize the estimated mean squared error (MSE) loss function as the objective function:

$$\begin{gathered} Min \quad MSE \\ subject\ to \\ \mathbf{x} \in R^k \end{gathered}$$

where MSE can be expressed as

$$E_z[y(\mathbf{x},\mathbf{z}) - T]^2 = \{E[y(\mathbf{x},\mathbf{z})] - T\}^2 + Var_z[y(\mathbf{x},\mathbf{z})]$$

This method simplifies the problem by admitting a little bias in the response mean and minimizing the response variability and MSE jointly. It is not an appropriate method when we want to keep the mean on target. We can improve this method by adding a constraint on the range of the response mean.

Table 3-6 Optimization Formulations with SNR and Dual RSM

<table>
<tr><th rowspan="2"></th><th rowspan="2">SNR</th><th colspan="2">Dual RSM</th></tr>
<tr><th>Method 1</th><th>Method 2</th></tr>
<tr><td>The target is best</td><td>$SNR_T = 10\log(\frac{\bar{y}^2}{S^2})$</td><td colspan="2">$Min \quad Var_z[y(\mathbf{x},\mathbf{z})]$
$subject\ to$
$E_z[y(\mathbf{x},\mathbf{z})] = m$
$\mathbf{x} \in R^k$
(Mean target m is known)</td></tr>
<tr><td>The larger the better</td><td>$SNR_L = -10\log(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2})$</td><td>$Max \quad E_z[y(\mathbf{x},\mathbf{z})]$
$subject\ to$
$Var_z[y(\mathbf{x},\mathbf{z})] = \sigma_0$
$\mathbf{x} \in R^k$
(Acceptable standard deviation σ_0 is known)</td><td>$Min \quad Var_z[y(\mathbf{x},\mathbf{z})]$
$subject\ to$
$E_z[y(\mathbf{x},\mathbf{z})] > m$
$\mathbf{x} \in R^k$
(Mean target m is a minimum value)</td></tr>
<tr><td>The smaller the better</td><td>$SNR_S = -10\log(\frac{1}{n}\sum_{i=1}^{n} y_i^2)$</td><td>$Min \quad E_z[y(\mathbf{x},\mathbf{z})]$
$subject\ to$
$Var_z[y(\mathbf{x},\mathbf{z})] = \sigma_0$
$\mathbf{x} \in R^k$
(Acceptable standard deviation σ_0 is known)</td><td>$Min \quad Var_z[y(\mathbf{x},\mathbf{z})]$
$subject\ to$
$E_z[y(\mathbf{x},\mathbf{z})] < m$
$\mathbf{x} \in R^k$
(Mean target m is a maximum value)</td></tr>
</table>

Example – Parameter Design Using Dual RSM and MSE

Retake the example to study the filtration rate of a chemical product from Myers and Montgomery (2002) to show the parameter design using dual response approach and mean square error loss function. We studied the problem with a two-level fractional factorial design 2_{IV}^{4-1} in Section 3.4. Recall that the three control factors are pressure (x_1), concentration of formaldehyde (x_2), and stirring rate (x_3), and the noise factor is temperature (z_1). Each factor is present at two levels. The filtration rate is denoted as response *y*. To compare the factorial design and fractional factorial design, we now use single replicate of a combined array 2^4 factorial design and the response data are provided in Table 3-7. The single response model is given by

$$y(x_1, x_2, x_3, z_1) = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i=1}^{3} \sum_{i<j} \beta_{ij} x_i x_j + \gamma_1 z_1 + \sum_{i=1}^{3} \delta_{1i} z_1 x_i + \varepsilon \qquad (3\text{-}13)$$

Table 3-7 Filtration Rate Experiment

Run Number	Factor				Filtration Rate (gal/hr)
	z_1 (A)	x_1 (B)	x_2 (C)	x_3 (D)	
1	-	-	-	-	45
2	+	-	-	-	71
3	-	+	-	-	48
4	+	+	-	-	65
5	-	-	+	-	68
6	+	-	+	-	60
7	-	+	+	-	80
8	+	+	+	-	65
9	-	-	-	+	43
10	+	-	-	+	100
11	-	+	-	+	45
12	+	+	-	+	104
13	-	-	+	+	75
14	+	-	+	+	86
15	-	+	+	+	70
16	+	+	+	+	96

Appendix A shows the contrast constants and the 15 factorial effect estimates for the 2^4 factorial designs. Based on these effects we can draw the normal probability plot as shown in Figure 3-3. From the normal probability plot we can find that the main effects of A (z_1), C (x_2), and D (x_3), and two-factor interactions AC (x_2z_1) and AD (x_3z_1) are significant. This can also be verified by ANOVA as Table 3-8. The coefficient of multiple determination R^2 is

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} = 96.6\%$$

That is, the regression model explains about 96.6% of the observed variability.

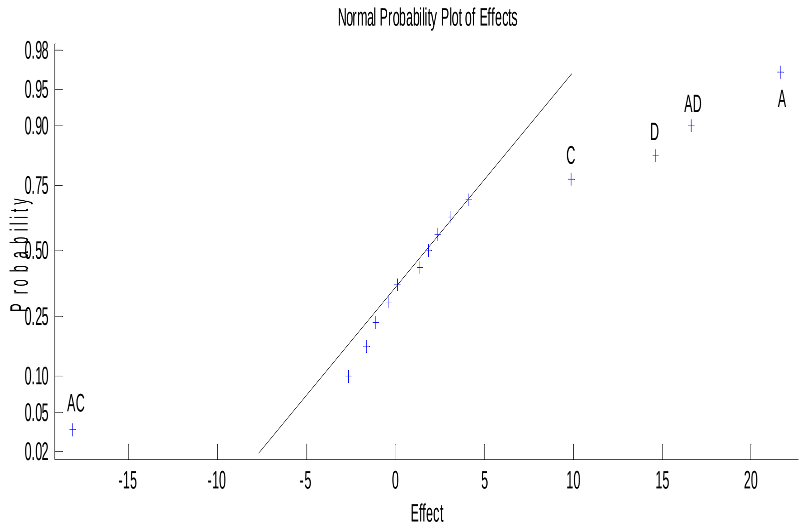

Figure 3-3 Normal probability plot of effects

Table 3-8 ANOVA of the Filtration Rate Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A	1870.563	1	1870.563	95.86
C	390.062	1	390.062	19.99
D	855.562	1	855.562	43.85
AC	1314.062	1	1314.062	67.34
AD	1105.563	1	1105.563	56.66
Error	195.125	10	19.513	
Total	5730.937	15		

As the average response $\bar{y}$ = 70.06, and the significant effects A = 21.625, C = 9.875, D = 14.625, AC = -18.125, and AD = 16.625, we can obtain the regression model of the response

$$\hat{y}(x, z_1) = 70.06 + (\frac{21.625}{2})z_1 + (\frac{9.875}{2})x_2 + (\frac{14.625}{2})x_3 - (\frac{18.125}{2})x_2 z_1 + (\frac{16.625}{2})x_3 z_1$$
$$= 70.06 + 10.81z_1 + 4.94x_2 + 7.31x_3 - 9.06x_2 z_1 + 8.31x_3 z_1$$

Figure 3-4 is the normal probability plot of the residuals. It shows all the points are close to the straight line and the analysis of the model is satisfactory.

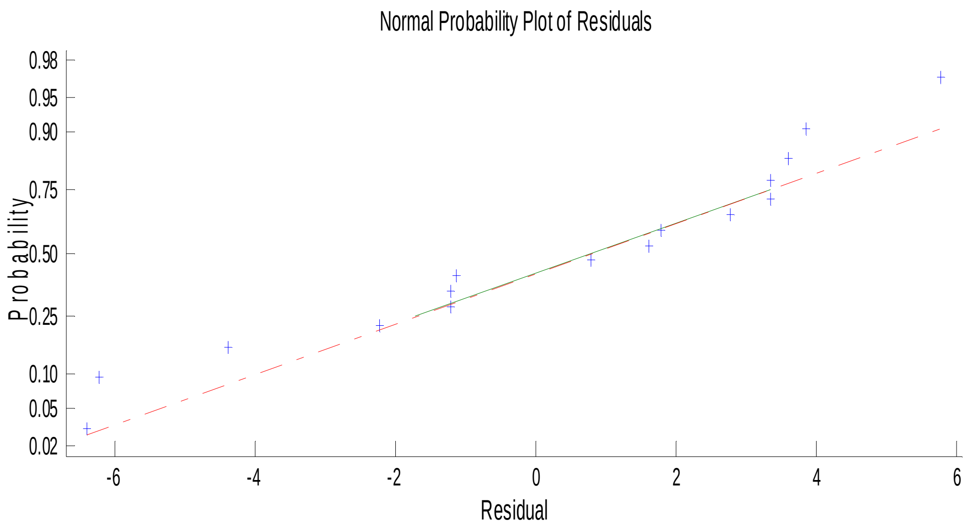

Figure 3-4 Normal probability plot of residuals

Suppose the noise factor z_1 is a random variable with mean zero and known variance $\sigma_z^2 = 1$. We also assume random residual term ε is NID $(0, \sigma^2)$. From above ANOVA, we know the residual variance $\hat{\sigma}^2$=19.51. If we assume that the fitted model above is adequate, and as $E(z_1) = 0$ and $E(\varepsilon) = 0$, the process mean model is

$$\hat{E}_z(y(x, z_1)) = 70.06 + 4.94x_2 + 7.31x_3$$

And the process variance is

$$\begin{aligned} Var_z[y(\mathbf{x}, \mathbf{z})] &= \sigma_z^2(\boldsymbol{\gamma}' + \mathbf{x}'\boldsymbol{\Delta})(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \sigma^2 \\ &= \sigma_z^2(10.81 - 9.06x_2 + 8.31x_3)^2 + \sigma^2 \\ &= 136.42 + 82.08x_2^2 + 69.06x_3^2 - 195.88x_2 \\ &\quad + 179.66x_3 - 150.58x_2x_3 \end{aligned}$$

We want to keep the mean at 75 and minimize the variance. The standard deviation (square root of the variance) of the response in the terms of the control variables is also called propagation of error (POE). Contour plots were used to analyze and estimate the robust design solution. Figure 3-5 to Figure 3-8 show the contour plots and response surface plots of the mean and standard deviation versus control variables x_2 and x_3. From the plots of response mean we find that the mean increases as either x_2 or x_3 increases. Furthermore we notice from the plots that the standard deviation increases as x_2 decreases or x_3 increases. Figure 3-9 illustrates the overlay plot of the mean and standard deviation of the response. We can estimate from the overlay plot that x_2 should be at high level and x_3 should be around the middle level (coded value 0).

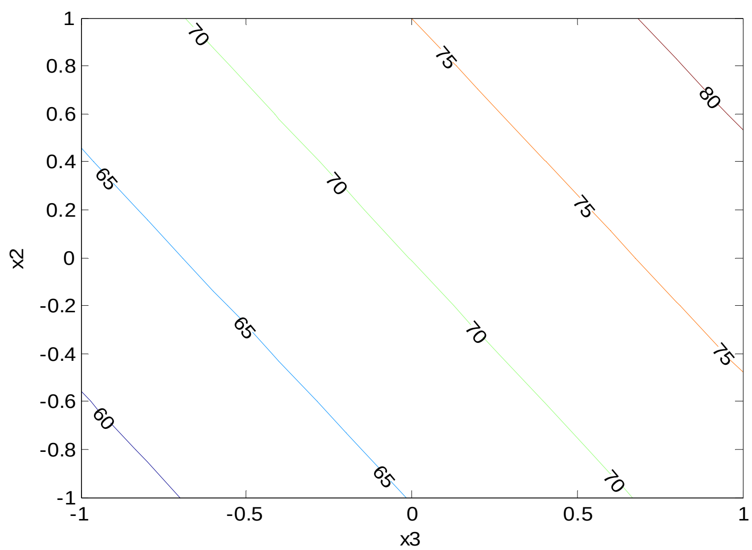

Figure 3-5 Contour plot of mean for filtration rate

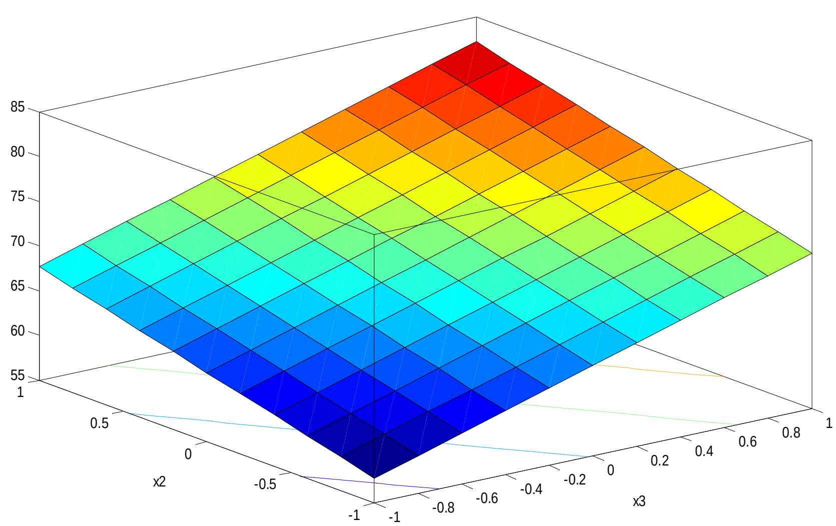

Figure 3-6 Response surface plot of mean for filtration rate

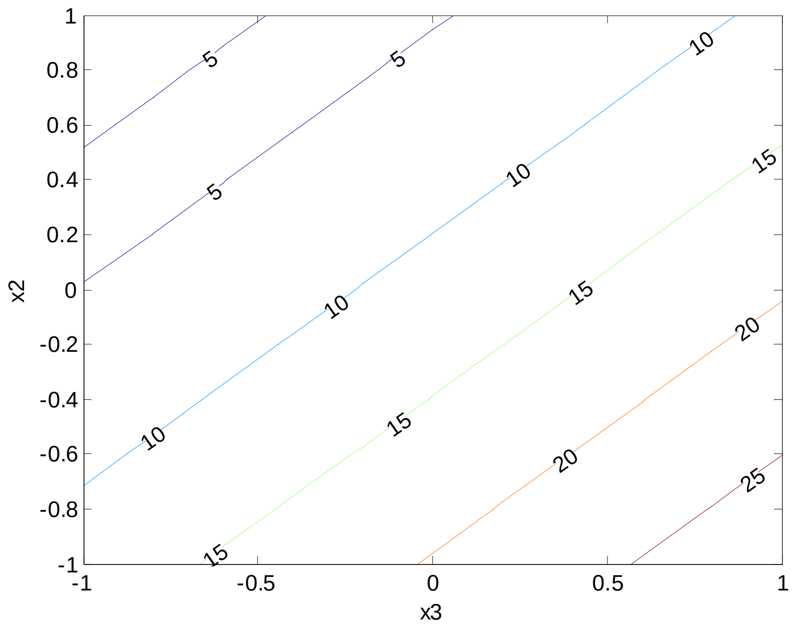

Figure 3-7 Contour plot of standard deviation for filtration rate

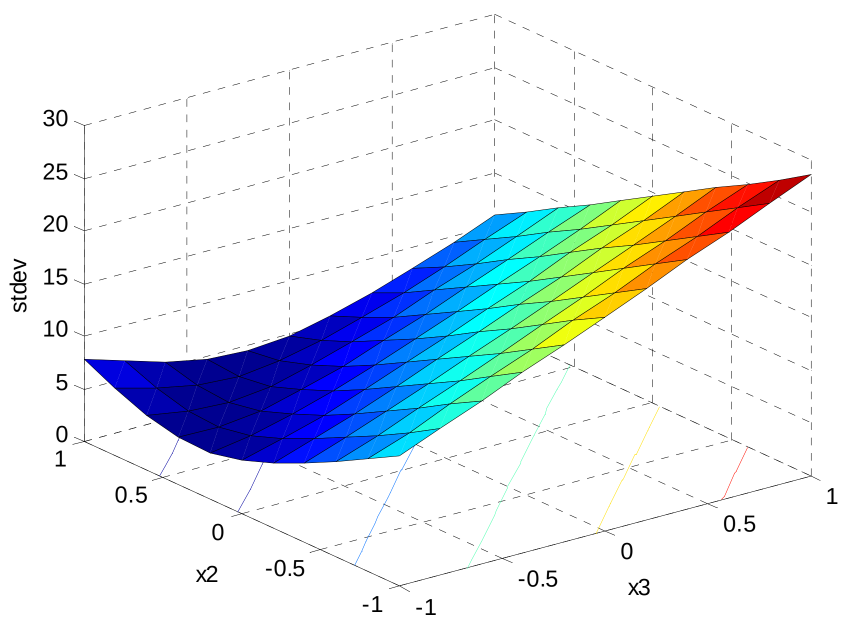

Figure 3-8 Response surface plot of standard deviation for filtration rate

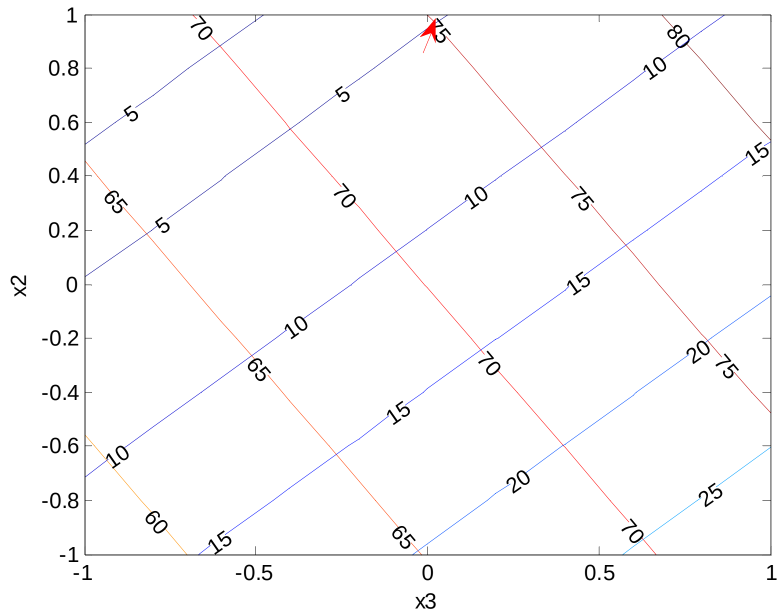

Figure 3-9 Overlay plot of mean and standard deviation contours

We must bear in mind that contour plots are used as informal estimation only. The more accurate nonlinear programming methods are available for performing formal optimization. Suppose we want to maintain the target filtration rate about m = 75 and minimize the variability around this value (target is best scenario), following two methods with different problem formulations were used to find optimal operating conditions.

First, we formulated process variance as objective function to optimize the target-is-best scenario. “fmincon” of Matlab was used to solve the following constrained optimization:

$$
\begin{gathered}
Min \quad Var_z[y(\mathbf{x},\mathbf{z})] \\
subject\ to \\
E_z[y(\mathbf{x},\mathbf{z})] = m \\
-1 \le x_2 \le 1 \;\; and \;\; -1 \le x_3 \le 1
\end{gathered}
$$

Second, we formulated estimated mean squared error (MSE) loss function as objective function in which the bias and variance are optimized together. Again, “fmincon” of Matlab was used to solve the following constrained optimization:

$$\min \quad E_z[y(\mathbf{x},\mathbf{z})-T]^2 = \{E[y(\mathbf{x},\mathbf{z})]-T\}^2 + Var_z[y(\mathbf{x},\mathbf{z})]$$
$$\textit{subject to}$$
$$-1 \le x_2 \le 1 \;\; and \;\; -1 \le x_3 \le 1$$

We obtained the optimal operating conditions as Table 3-9. The second method admits a little bias in the response mean, but reduces the response variability and MSE at the same time. Notice the nominal values of x_2 and x_3 from the two methods are very close.

Table 3-9 Optimal Operating Conditions of Filtration Rate Experiment

	x_2	x_3	$E_z[y(\mathbf{x},\mathbf{z})]$	$Var_z[y(\mathbf{x},\mathbf{z})]$	MSE
Min Var	1	0	75	22.62	22.62
Min MSE	1	-0.1187	74.1323	20.1413	20.8941

3.7 Generalized Linear Models (GLM)

The residuals from the least squares fit $\varepsilon_i = y_i - \hat{y}_i$ are very important to judge model adequacy. It is common to assume the residual variance to be constant σ^2 in regression model. However, it is not always true in practical problems. If the response variable is not normal, the residual variance is possibly nonconstant and ordinary least squares (OLS) approach is not appropriate to estimate the regression model. Recall the previous case study of the filtration rate of a chemical problem from Myers and Montgomery (2002) and plot the residuals from the regression model versus x_2, x_3, and $\hat{y}$ (Figure 3-10 to Figure 3-12). The funnel-shaped patterns imply possible inequality of residual variance.

Figure 3-10 Residuals versus x_2

Figure 3-11 Residuals versus x_3

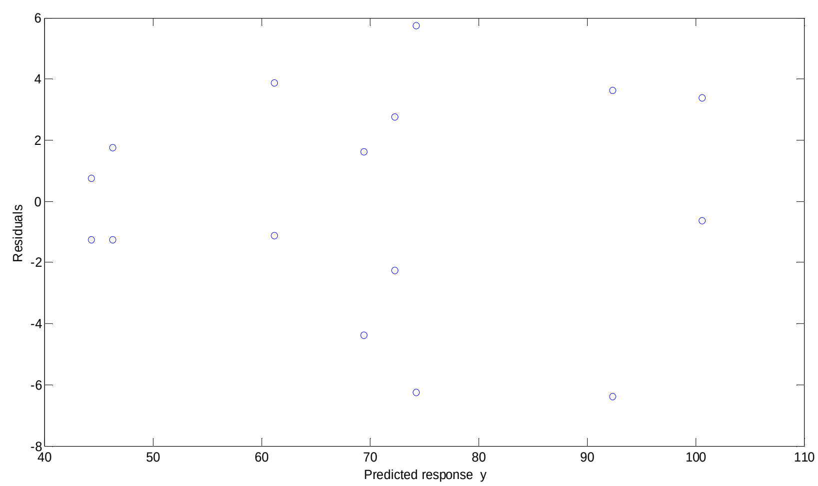

Figure 3-12 Residuals versus $\hat{y}$

While transformation of the response variables can be used to stabilize response variance and make the distribution of the response variable closer to the normal distribution, the generalized linear model (GLM) is a better tool in variance modeling and so in robust design. It helps solve problems involving exponential responses, such as responses with normal, gamma, exponential, Poisson, and binomial distributions. Even if the response is normal, the modeling variance, either from replication or from non-replication experiments, is important to use the GLM. The dual response models of process mean and variance can be achieved iteratively by GLM method.

The model of the residual variance can be expressed as

$$\sigma_i^2(\mathbf{x}) = E(\varepsilon_i^2) = \exp(\mathbf{x}'_{\mathbf{i}\boldsymbol{\sigma}}\boldsymbol{\beta}_{\boldsymbol{\sigma}}) \tag{3-14}$$

where *i* is the run number, and variance regression model $\mathbf{x}'_{\mathbf{i}\boldsymbol{\sigma}}\boldsymbol{\beta}_{\boldsymbol{\sigma}}$ is opposed to $\mathbf{x}'_{\mathbf{i}}\boldsymbol{\beta}$ that is used in ordinary regression model of the response to allow for possible difference between the model of variance and response. The iteratively reweighted least squares (IRLS) algorithm can be used for fitting the response model in the case of unreplicated experiments. The details of the IRLS algorithm are as follows:

1. Use ordinary least squares (OLS) with assumption of constant residual variation to obtain an initial fit to the regression model. The coefficient estimator for mean model $y_i = \mathbf{x}'_{\mathbf{i}}\boldsymbol{\beta} + \varepsilon_i$ is $\hat{\boldsymbol{\beta}}$.

2. Use the squared residuals from step 1, $\varepsilon_i^2 = (y_i - \hat{y}_i)^2 = (y_i - \mathbf{x}'_{\mathbf{i}}\hat{\boldsymbol{\beta}})^2$, to estimate the parameters, $\boldsymbol{\beta}_\sigma$, of the variance model using GLM methodology.

3. Use $v_i = \exp(\mathbf{x}'_{\mathbf{i}\boldsymbol{\sigma}}\hat{\boldsymbol{\beta}}_{\boldsymbol{\sigma}})$ as weights to compute $\mathbf{V} = diag(v_1, v_2, ..., v_d)$. The weighted least squares estimator is used as new coefficient estimator $\hat{\boldsymbol{\beta}}$ and

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

4. Go back to step 2 with the new coefficient estimator and residuals, and re-fit the variance model.

5. Continue to re-estimate the mean and variance until convergence to the maximum likelihood estimator for the coefficient vector, or a specified maximum number of iterations is reached.

Example: Filtration Rate Experiment with GLM Approach

We would now work on the previous filtration rate problem (Myers and Montgomery, 2002) with GLM method. To solve the residual variance, with the data from the initial fit, we used "glmfit" from Matlab while setting a gamma distribution with log link for the variance, and assumed a normal distribution with identity link for the mean. The link function links the variance or mean to the linear predictor. After five iterations of IRLS we can achieve convergence. The coefficient estimator $\hat{\boldsymbol{\beta}}_\sigma$ for the model of residual variance and the weighted least squares estimator $\hat{\boldsymbol{\beta}}$ for the response model are shown in Table 3-10 and Table 3-11.

Table 3-10 Coefficient Estimator of Residual Variance Model

Iteration	$\hat{\beta}_{\sigma 0}$	$\hat{\beta}_{\sigma}\ (x_2)$	$\hat{\beta}_{\sigma}\ (x_3)$
1	2.2636	0.6929	-0.2319
2	2.2	0.8334	-0.2975
3	2.1986	0.8474	-0.3081
4	2.1986	0.8487	-0.3097
5	2.1986	0.8488	-0.3099
6	2.1986	0.8488	-0.31

Table 3-11 Coefficient Estimator of Response Model

Iteration	$\hat{\beta}_0$	$\hat{\beta}\ (z_1)$	$\hat{\beta}\ (x_2)$	$\hat{\beta}\ (x_3)$	$\hat{\beta}\ (x_2 z_1)$	$\hat{\beta}\ (x_3 z_1)$
1	69.9856	10.7015	4.8094	7.6499	-9.2476	8.7999
2	69.9516	10.6523	4.7749	7.6963	-9.2973	8.8668
3	69.9466	10.6451	4.7695	7.7005	-9.3052	8.8729
4	69.9459	10.6441	4.7686	7.7008	-9.3064	8.8734
5	69.9458	10.644	4.7685	7.7009	-9.3066	8.8735
6	69.9458	10.644	4.7685	7.7009	-9.3066	8.8735

Therefore, the model of the residual variance is

$$\sigma_i^2(x_2, x_3) = \exp(\mathbf{x}'_{\mathbf{i}\boldsymbol{\sigma}}\boldsymbol{\beta}_{\boldsymbol{\sigma}}) = \exp(2.1986 + 0.8488x_2 - 0.31x_3)$$

Assuming the variance of noise variable z_1 is $\sigma_z^2 = 1$, the new regression model and process mean and variance model can be obtained as

$$\hat{y}(\mathbf{x}, \mathbf{z}) = 69.9458 + 10.644z_1 + 4.7685x_2 + 7.7009x_3 - 9.3066x_2 z_1 + 8.8735x_3 z_1$$

$$\hat{E}[y(\mathbf{x},\mathbf{z})] = 69.9458 + 4.7685x_2 + 7.7009x_3$$

$$\begin{aligned} Var_z[y(\mathbf{x},\mathbf{z})] &= \sigma_z^2(\boldsymbol{\gamma}' + \mathbf{x}'\boldsymbol{\Delta})(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \sigma^2 \\ &= \sigma_z^2(10.644 - 9.3066x_2 + 8.8735x_3)^2 + \exp(2.1986 + 0.8488x_2 - 0.31x_3) \\ &= (10.644 - 9.3066x_2 + 8.8735x_3)^2 + \exp(2.1986 + 0.8488x_2 - 0.31x_3) \end{aligned}$$

The problem has same goal as before: keep the mean at 75 and minimize the variance. Contour plots can be used for estimation before further formal optimization. Figure 3-13 and Figure 3-14 show response surface plot of the process mean and standard deviation. From the plot of process mean we find that the mean increases as either x_2 or x_3 increases. Next, we notice in the plot of the standard deviation that the square root of the variance increases as x_2 decreases or x_3 increases, and the process is more sensitive to change in x_3 than to change in x_2. Figure 3-15 illustrates the overlay plot of the mean and standard deviation of the response. We can estimate from the plot that x_2 should be at high level and x_3 should be around the middle level (coded value 0). The contour and response surface plots indicate the two methods, the ordinary linear regression model (OLS) method and the generalized linear model (GLM) method, have similar optimal conditions for x_2 and x_3 regarding the mean and standard deviation model. Nonlinear programming was performed next as formal constrained optimization.

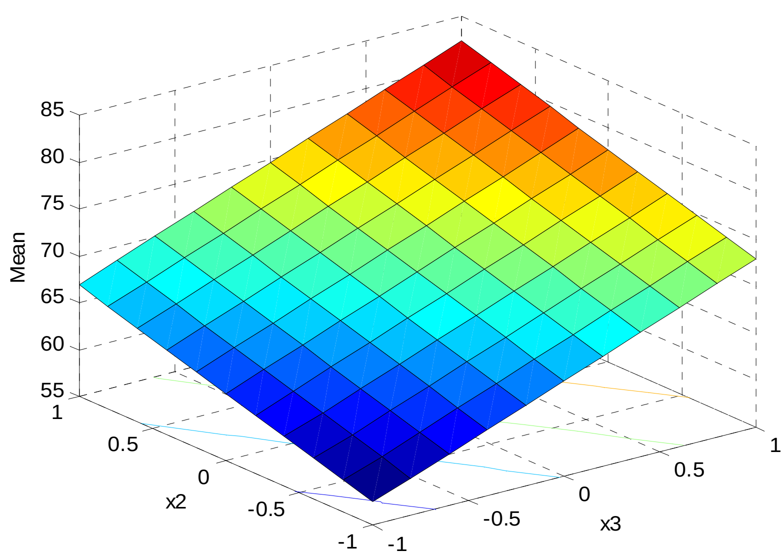

Figure 3-13 Response surface plot of mean filtration rate

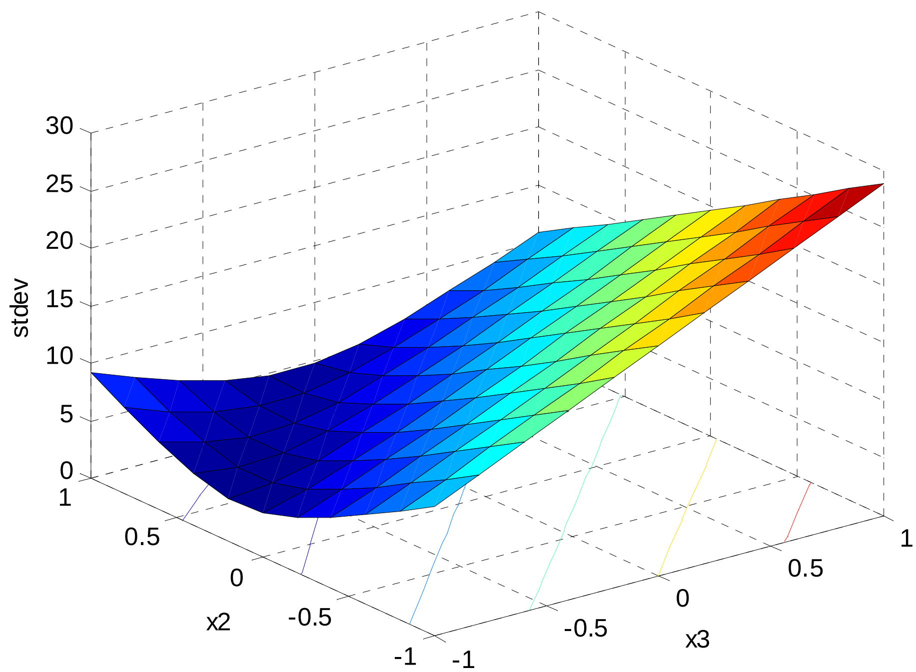

Figure 3-14 Response surface plot of standard deviation

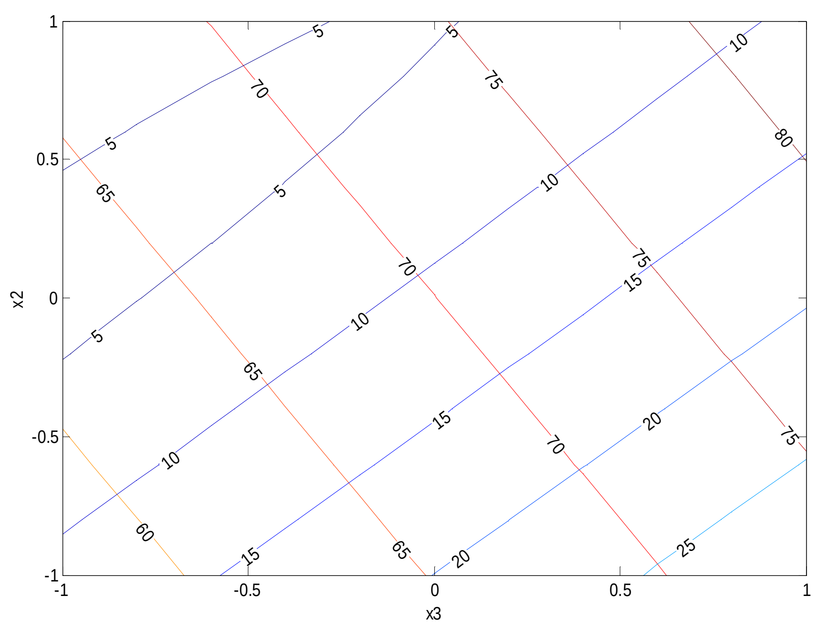

Figure 3-15 Overlay plot of mean and standard deviation contours

Same as before, suppose we want to maintain the target filtration rate about m = 75 and minimize the variability around this value (target is best scenario); following two methods with different problem formulations were used to find optimal operating conditions. First, we formulated process variance as objective function to optimize the target-is-best scenario. Matlab was used to solve the following constrained optimization:

$$\begin{gathered} Min \quad Var_z[y(\mathbf{x},\mathbf{z})] \\ subject\ to \\ E_z[y(\mathbf{x},\mathbf{z})] = 75 \\ -1 \le x_2 \le 1\ and\ -1 \le x_3 \le 1 \end{gathered}$$

Second, we formulated estimated mean squared error (MSE) loss function as objective function in which the bias and variance are optimized together. Again, Matlab was used to solve the following constrained optimization:

$$\begin{gathered} Min \quad E_z[y(\mathbf{x},\mathbf{z}) - T]^2 = \{E[y(\mathbf{x},\mathbf{z})] - T\}^2 + Var_z[y(\mathbf{x},\mathbf{z})] \\ subject\ to \\ -1 \le x_2 \le 1 \quad and \quad -1 \le x_3 \le 1 \end{gathered}$$

We obtained the optimal operating conditions as Table 3-12, where $\sigma^2_{residual}$ is the non-constant residual variance at the optimal conditions of the control variables. As before, the second method admits a little bias in the response mean, but at the same time reduces the response variability and MSE.

Table 3-12 Optimal Operating Conditions of Filtration Rate Experiment (GLM)

	x_2	x_3	$E_z[y(\mathbf{x},\mathbf{z})]$	$Var_z[y(\mathbf{x},\mathbf{z})]$	MSE	$\sigma^2_{residual}$
Min Var	1	0.0371	75	23.5971	23.5971	20.8197
Min MSE	1	-0.046	74.3597	22.226	22.636	21.3633

Table 3-13 presents the comparison of the optimal solutions from the two different variance modeling strategies (ordinary least squares method for constant residual variation, and generalized linear model method for non-constant residual variation) using two different optimizing formulations (dual RSM approach to minimizing process variance and MSE loss function).

Table 3-13 Optimal Operating Conditions by OLS and GLM

Optimizing formulation	Optimal Conditions	OLS	GLM
Min Variance	(x_2, x_3)	(1, 0)	(1, 0.0371)
	Mean	75	75
	Variance	22.62	23.5973
	MSE	22.62	23.5973
	$\sigma^2_{residual}$	19.513	20.8197
Min MSE loss function	(x_2, x_3)	(1, -0.1187)	(1, -0.046)
	Mean	74.1323	74.3597
	Variance	20.1413	22.226
	MSE	20.8941	22.636
	$\sigma^2_{residual}$	19.513	21.3633

Based on the comparison of the solutions, it is clear we obtained different regression models and optimal settings from GLM and OLS, and that means the assumption of constant residual variance does affect the accuracy of response modeling. While the residual variance is non-constant and changing along with the design variables (in this example x_2 and x_3), at the optimal operating conditions from GLM approach, the residual variation is different from (in this case higher than) the estimated constant residual variation obtained from the OLS approach. And this results in different process variance and MSE. In summary, when the residual variance is non-constant, we can use generalized linear model and dual RSM approach to obtain the regression model and formulate the constrained optimization for nonlinear programming.

3.8 Designs for Fitting Response Surfaces

As discussed before, Taguchi's cross array or product array designs, which use separate designs for control and noise factors, can be greatly improved with combined array designs, which put the control factors and noise factors in a single design. In this section, combined array designs are discussed for both spherical regions of interest and cuboidal regions of interest. Region of operability is the region in which the system (a product or process) can operate for the experiment. Region of interest or experimentation is a smaller region within the region of operability, as we may not explore the entire region of operability with a single or only a few experiments. Region of interest confines the ranges on the design variables, and may change from experiment to experiment.

When the region of interest and the region of operability are identical, cuboidal region of interest or a cuboidal design is appropriate. While the most popular cuboidal design is face-centered cube (FCC), other cuboidal designs, particularly computer-generated D-optimal or G-optimal designs are also useful in RSM. On the other hand, if we are interested in the variables with levels that are beyond the ranges given, which means the design variables are beyond the region of interest and within the region of operability, spherical region of interest or a spherical design is used. Central composite designs (CCD) and Box–Behnken designs (BBD) are standard choices for fitting a second-order response surface model in a spherical region of interest. Other spherical designs include small composite designs (SCD), mixed resolution designs (MRD), hybrid designs, and so on.

Myers and Montgomery (2002) listed ten desirable properties of response surface designs that the experimenters should take into account when they select the appropriate designs for robust design problem:

1. Result in a good fit of the model to the data.
2. Give sufficient information to allow a test for lack of fit.
3. Allow models of increasing order to be constructed sequentially.
4. Provide an estimate of pure experimental error.
5. Be insensitive (robust) to the presence of outliers in the data.
6. Be robust to error in control of design levels.
7. Be cost-effective.
8. Allow for experiments to be done in blocks.
9. Provide a check on the homogeneous variance assumption.
10. Provide a good distribution of $Var[\hat{y}(x)]/\sigma^2$.

Obviously, we cannot construct a single RSM design that satisfies all of these properties, and not all of the properties are required in every application. Therefore, design evaluation should be performed and trade-offs should be studied.

3.8.1 Design Construction and Analysis—Spherical Designs

A central composite designs (CCD) involves a two-level factorial (2^k) or fractional factorial (2^{k-p}) design, 2k axial points (denoted by $\pm\alpha$), and n_c center points. *k* represents the number of design variables in the model and *p* is the number of independent design generators. In general, a 2^{k-p} fractional factorial design can be used rather than a 2^k design to save limited resources. The estimation of linear terms and two-factor interactions is relied on the fractional factorial design, so a resolution V design, in which no main effect or two-factor interaction is aliased with any other main effects or two-factor interactions, is the best choice. Axial points lie on the axis. Note that in the axial portion of the design the factors are not varying simultaneously but rather in a one-factor-at-a-time array. As a result, no information regarding the interaction is provided by this portion of the design. However, the axial portion allows for efficient estimation of pure quadratic terms.

To make the design flexible, CCDs have two important parameters: the axial distance α and the number of center points n_c. The axial distance α is chosen based on the region of interest and region of operability, whereas the choice of the number of center point n_c influences prediction variance $Var[\hat{y}(x)]/\sigma^2$ in the region of interest. Axial distance α usually varies from 1 to $\sqrt{k}$. When α = 1, all axial points are on the face of the cube or hypercube (cuboidal region), and such kind of design is called face-centered cube (FCC) design, which will be discussed later. When $\alpha = \sqrt{k}$, all axial points will be placed on a common sphere (spherical region). While it is common to define axial distance $\alpha = \sqrt{k}$ in a spherical design, we can also choose $\alpha = \sqrt[4]{F}$, where F is the number of factorial points in the design. As the spherical design is rotatable, the value of the scaled prediction variance $Var[\hat{y}(x)]/\sigma^2$ for a second-order model is the same at all points that are equidistant from the design center.

Consider an example of a CCD with five variables, $\alpha = \sqrt{k} = 2.236$, $n_c = 3$, and the fractional factorial portion is 2_V^{5-1} . The standard CCD with 29 runs is shown in left portion of Table 3-14.

Table 3-14 Standard and Modified CCD with Five Variables

Run number	Standard CCD with 5 variables					Modified CCD with 2 control and 3 noise variables				
	x_1	x_2	x_3	x_4	x_5	x_1	x_2	z_1	z_2	z_3
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	-2.236	0	0	0	0	-2.236	0	0	0	0
18	2.236	0	0	0	0	2.236	0	0	0	0
19	0	-2.236	0	0	0	0	-2.236	0	0	0
20	0	2.236	0	0	0	0	2.236	0	0	0
21	0	0	-2.236	0	0	0	0	0	0	0
22	0	0	2.236	0	0	0	0	0	0	0
23	0	0	0	-2.236	0	0	0	0	0	0
24	0	0	0	2.236	0					
25	0	0	0	0	-2.236					
26	0	0	0	0	2.236					
27	0	0	0	0	0					
28	0	0	0	0	0					
29	0	0	0	0	0					

Since the axial points primarily contribute to the estimation of the quadratic terms, and generally in robust design problems we ignore the pure quadratic terms of the noise variables, we can modify the standard CCD by eliminating the axial points corresponding to noise variables. The right portion of this table illustrates the modified CCD with two control variables and three noise variables. The modified CCD is appropriate for estimation of the response surface and will also reduce the overall size of the design (from 29 to 23 runs in this example). The disadvantage of CCD or modified CCD is that as the number of variables increases, the experimental design becomes very large.

Box-Behnken designs (BBD) are developed to fit second-order models where it is essential to keep all factors at three levels and simultaneously preserve approximate rotatability. However, Borror (1998) pointed out that BBD is not an effective alternative to the CCD for robust design problems. This family of designs is based on balanced incomplete block designs where the variables are all at three levels. BBDs are strictly spherical and do not contain axial points. Therefore we cannot easily modify the BBD to accommodate the noise variables as is done in the modified CCD. However, BBD has the same design as long as the number of total variables is not changed. Table 3-15 presents the BBD with k = 5. The design consists of 40 factorial points and 3 centers point. BBD is same as CCD on the point that three to five center runs are recommended.

Table 3-15 BBD with Five Variables

Run number	BBD with 5 variables				
	x_1	x_2	x_3	x_4	x_5
1-4	±1	±1	0	0	0
5-8	±1	0	±1	0	0
9-12	±1	0	0	±1	0
13-16	±1	0	0	0	±1
17-20	0	±1	±1	0	0
21-24	0	±1	0	±1	0
25-28	0	±1	0	0	±1
29-32	0	0	±1	±1	0
33-36	0	0	±1	0	±1
37-40	0	0	0	±1	±1
41-43	0	0	0	0	0

Standard or modified CCD and BBD work very well to meet the desirable criteria of second-order responses surface designs. However, if cost prohibits the use of one of the standard designs with required number of experiment, saturated and near-saturated design can be used, such as small composite design (SCD). SCD is a smaller alternative to the CCD and is developed from the ideas of CCD. It will not become very large with an increase of the number of variables, because the factorial portion of SCD is neither a complete 2^k nor a resolution V fraction as in the CCD but is a resolution

III* fraction, which is a special 2_{III}^{k-p} design without a four-letter word among the defining relations (Myers and Montgomery, 2002).

The number of parameters in a second-order model, which contains first- and second-order design variables and their interactions, is

$$P = 1 + 2k + \frac{k(k-1)}{2} \tag{3-15}$$

We may use the SCD design that contains no less than P design points to fit this model. In the same example as in the CCD with five variables (k = 5), we have P = 21. This is a special case where we cannot use a resolution III* fraction: because the total of $2_{III}^{5-2} = 8$ points of fractional factorial portion, 10 axial points, and 3 center points, are 21 points, which are the same as the number of design parameters (P = 21) and will result in no degree of freedom for estimating lack of fit. As a result, we can use 2_V^{5-1} design with defining relation I = ABCDE. The design has 29 design points: 16 factorial points, 10 axial points, and 3 center points, and the degrees of freedom for lack of fit are 8. We notice that in the case of five design variables or two control variables and three noise variables, the SCD or modified SCD is exactly same as the CCD or modified CCD. For other numbers of design factors, particularly as the number of variables increases, the SCD will be significantly smaller than the CCD. The SCD is a very sound alternative to the CCD based on the number of runs. However, the standard or modified SCD has a potential problem that there is not always enough runs available to accommodate all the terms of interest in the response model of the robust design problem. SCD is not recommended if resources allow us to use either a CCD or BBD design.

Myers and Montgomery (2002) illustrated the differences between a CCD and SCD with a simple example. For k = 3, the standard CCD has 15 runs: 2^3 factorial points, 6 axial points, and 1 center point, whereas the SCD has 11 runs: 2^{3-1} factorial points (I = -ABC), 6 axial points, and 1 center point. Table 3-16 shows their design constructions. We may compare the efficiencies of model coefficients for CCD and SCD. From the matrix of $X'X_{SCD}$ and $X'X_{CCD}$ (Equation 3-16 and 3-17) of a second-order model we can see that in the CCD all linear main effects and two-factor interactions are mutually orthogonal, while in the SCD the main effects and two-factor interactions are not orthogonal (x_1 with x_2x_3, x_2 with x_1x_3, and x_3 with x_1x_2). These correlations will certainly affect the variances of the related regression coefficients.

Table 3-16 Standard CCD versus SCD with Three Variables

Run Number	Standard CCD			SCD		
	x_1	x_2	x_3	x_1	x_2	x_3
1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	1	1	-1
3	-1	1	-1	1	-1	1
4	1	1	-1	-1	1	1
5	-1	-1	1	-α	0	0
6	1	-1	1	α	0	0
7	-1	1	1	0	-α	0
8	1	1	1	0	α	0
9	-α	0	0	0	0	-α
10	α	0	0	0	0	α
11	0	-α	0	0	0	0
12	0	α	0			
13	0	0	-α			
14	0	0	α			
15	0	0	0			

Table 3-17 shows the scaled coefficient variances $N\,Var(b)/\sigma^2$ taken from appropriate diagonals of the matrix $N(X'X)^{-1}$ $(\alpha = \sqrt{3}\ and\ n_c = 3)$. The designs have close results for estimation of second-order coefficients (b_{ii}), but the results for estimation of linear and two-factor interactions are quite different (b_0, b_i, and b_{ij}). Therefore, we can conclude that the SCD is not a competitive design considering the efficiency for estimation of linear and two-factor interaction coefficients in the second-order models.

Table 3-17 Scaled Variances of Model Coefficients for CCD and SCD

	b_0	b_i	b_{ii}	b_{ij}
CCD (N = 17)	5.6666	1.2143	1.3942	2.1250
SCD (N = 13)	4.3333	2.1666	1.1074	5.4166

$$X'X_{SCD} = \begin{array}{c} \begin{array}{ccccccccc} x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1x_2 & x_1x_3 & x_2x_3 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 4+2\alpha^2 & 4+2\alpha^2 & 4+2\alpha^2 & 0 & 0 & 0 \\ 4+2\alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 \\ & 4+2\alpha^2 & 0 & 0 & 0 & 0 & 0 & -4 & 0 \\ & & 4+2\alpha^2 & 0 & 0 & 0 & -4 & 0 & 0 \\ & & & 4+2\alpha^2 & 0 & 0 & 0 & 0 & 0 \\ & & & & 4+2\alpha^2 & 0 & 0 & 0 & 0 \\ & & & & & 4+2\alpha^2 & 0 & 0 & 0 \\ & & & & & & 4 & 0 & 0 \\ & & & & & & & 4 & 0 \\ & & & & & & & & 4 \end{bmatrix} \end{array} \quad (3\text{-}16)$$

$$X'X_{CCD} = \begin{array}{c} \begin{array}{ccccccccc} x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1x_2 & x_1x_3 & x_2x_3 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 8+2\alpha^2 & 8+2\alpha^2 & 8+2\alpha^2 & 0 & 0 & 0 \\ 8+2\alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 8+2\alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 8+2\alpha^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 8+2\alpha^2 & 8 & 8 & 0 & 0 & 0 \\ & & & & 8+2\alpha^2 & 8 & 0 & 0 & 0 \\ & & & & & 8+2\alpha^2 & 0 & 0 & 0 \\ & & & & & & 8 & 0 & 0 \\ & & & & & & & 8 & 0 \\ & & & & & & & & 8 \end{bmatrix} \end{array} \quad (3\text{-}17)$$

3.8.2 Mixed Resolution Design (MRD)

A mixed resolution design (MRD) is a very competitive alternative to any of the standard or modified CCDs or SCDs. Same as the CCD, the mixed resolution design includes three portions: a two-level factorial (2^k) or fractional factorial (2^{k-p}) design, axial points, and center points. Borkowski and Lucas (1997) define the mixed-resolution fractional factorial design (2^{k-p}) as at least resolution V design in terms of the control factors, at least resolution III design among the noise factors, and none of the important control-by-noise interactions are aliased with any main effect or other two-factor interaction. The definition is reasonable since among the control factors, no main effects or two-factor interactions are aliased with any other main-effects or two-factor interactions. Meanwhile, among the noise factors, no noise main factors are aliased with other noise main factors, and the noise-by-noise interactions, which are not important to us, are aliased with each other and with main noise effects. Therefore, the terms of interest are not confounded with each other, and the mixed resolution designs select the defining relation for the fractional factorial portion of the design based on the factors and interactions that are important to estimate (Robinson et al. 2004). The axial and center points can be designed in the same way as the standard or modified CCD.

Consider a situation with three control variables and three noise variables. Myers and Montgomery (2002) used a 2^{6-2} mixed resolution design with defining relations

$$I = x_1x_2x_3z_1 = z_1z_2z_3 = x_1x_2x_3z_2z_3$$

The design is resolution III with regard to noise-by-noise interactions and resolution IV with regard to the control-by-control interactions. The design consists of 25 runs, which includes 16 runs in fractional factorial portion, six axial points (eliminating the axial points corresponding to noise variables as the modified CCD), and three center points. However, since some control-by-control interactions are aliased with control-by-noise interactions ($x_1x_2 = x_3z_1$, $x_1x_3 = x_2z_1$, and $x_2x_3 = x_1z_1$), this design construction is inappropriate for the robust design. If the practitioner is interested in estimating all the two-factor interactions, Borror (1998) presented a very efficient method to dealias the control-by-control interactions and control-by-noise interactions by augmenting the fractional factorial design with additional runs. A full foldover of the original design is not necessary. Adding a set of runs 17 to 22 can dealias the two-factor interactions. The set of runs are duplicates of runs 9 to 14 in the original factorial portion with the sign changed on only x_2, one of the four variables that involved in the two-factor interactions. Thus, all two-factor interactions can be estimated without alias with other two-factor interactions. This design requires 22 runs in fractional factorial portion, six axial points (eliminate the axial points of noise variables as the modified CCD), and three center points.

An alternative design is suggested by Borkowski and Lucas (1997) for three control variables and three noise variables. The design is 6A based on their design category. The design consists of 41 runs, which include 32 runs in fractional factorial portion, six axial points (eliminate the axial points of noise variables as the modified CCD), and three center points (assumed). The 2^{6-1} design is set up in five factors A, B, C, D, and E. The remaining one factor is F = ABCDE. Table 3-18 shows the mixed resolution designs from Borror (1998) and Borkowski and Lucas (1997).

Table 3-18 Comparison of Two MRDs (Three Control Variables and Three Noise Variables)

Run	Borror (1998) MRD						Borkowski and Lucas (1997) MRD					
	x_1	x_2	x_3	z_1	z_2	z_3	x_1	x_2	x_3	z_1	z_2	z_3
1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	1	-1	-1	1	-1	-1	-1	-1	1
3	-1	1	-1	1	-1	-1	-1	1	-1	-1	-1	1
4	1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1
5	-1	-1	1	1	-1	-1	-1	-1	1	-1	-1	1
6	1	-1	1	-1	-1	1	1	-1	1	-1	-1	-1
7	-1	1	1	-1	-1	1	-1	1	1	-1	-1	-1
8	1	1	1	1	-1	-1	1	1	1	-1	-1	1
9	-1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	1
10	1	-1	-1	1	1	1	1	-1	-1	1	-1	-1
11	-1	1	-1	1	1	1	-1	1	-1	1	-1	-1
12	1	1	-1	-1	1	-1	1	1	-1	1	-1	1
13	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1
14	1	-1	1	-1	1	-1	1	-1	1	1	-1	1
15	-1	1	1	-1	1	-1	-1	1	1	1	-1	1
16	1	1	1	1	1	1	1	1	1	1	-1	-1
17	-1	1	-1	-1	1	-1	-1	-1	-1	-1	1	1
18	1	1	-1	1	1	1	1	-1	-1	-1	1	-1
19	-1	-1	-1	1	1	1	-1	1	-1	-1	1	-1
20	1	-1	-1	-1	1	-1	1	1	-1	-1	1	1
21	-1	1	1	1	1	1	-1	-1	1	-1	1	-1
22	1	1	1	-1	1	-1	1	-1	1	-1	1	1
23	-2.38	0	0	0	0	0	-1	1	1	-1	1	1
24	2.38	0	0	0	0	0	1	1	1	-1	1	-1
25	0	-2.38	0	0	0	0	-1	-1	-1	1	1	-1
26	0	2.38	0	0	0	0	1	-1	-1	1	1	1
27	0	0	-2.38	0	0	0	-1	1	-1	1	1	1
28	0	0	2.38	0	0	0	1	1	-1	1	1	-1
29	0	0	0	0	0	0	-1	-1	1	1	1	1
30	0	0	0	0	0	0	1	-1	1	1	1	-1
31	0	0	0	0	0	0	-1	1	1	1	1	-1
32							1	1	1	1	1	1
33							-2.38	0	0	0	0	0
34							2.38	0	0	0	0	0
35							0	-2.38	0	0	0	0
36							0	2.38	0	0	0	0
37							0	0	-2.38	0	0	0
38							0	0	2.38	0	0	0
39							0	0	0	0	0	0
40							0	0	0	0	0	0
41							0	0	0	0	0	0

3.8.3 Design Construction and Analysis—Cuboidal Designs

Axial distance α usually varies from 1 to $\sqrt{k}$. When $\alpha = \sqrt{k}$, all axial points will be place on a common sphere (spherical region). On the other hand, when $\alpha = 1$, all axial points are on the face of the cube or hypercube (cuboidal region). The most popular cuboidal design is face-centered cube (FCC), which has the same advantages and disadvantages as the CCD. We also have a modified FCC to accommodate robust design problems in the same manner as the modified CCD. Therefore, as the number of variables increases, FCC or modified FCC becomes very large and costly. Another difference between cuboidal design and spherical design is the requirement for center runs. Myers and Montgomery (2002) studied the sensitivity of $N\,Var[\hat{y}(\mathbf{x})]/\sigma^2$ to the number of center runs and concluded that spherical designs need three to five center runs (n_c = 3 to 5) to produce stable results, but n_c=1 is sufficient to achieve stability in cuboidal designs, though n_c = 2 is slightly preferable.

3.8.4 Methods for Evaluating Response Surface Designs

While RSM can predict the response or estimate the mean response at a particular point, the variance of the prediction or prediction variance, which is a direct measure of the error resulted from the model at the point, is an important characteristic to study. The prediction variance is given as

$$Var[\hat{y}(\mathbf{x})] = \mathbf{x}^{(m)'}\mathbf{(X'X)^{-1}}\mathbf{x}^{(m)}\sigma^2 \tag{3-18}$$

where σ^2 is error variance, the vector $\mathbf{x}^{(m)}$ is a function of the location in the design variables at which one predicts the response, and the (*m*) reflects the nature of the model as **X** does and indicates that $\mathbf{x}^{(m)}$ is just **x** expanded to model space. For example, for k = 2 design variables and the model contains first-order x_1 and x_2 only, we have

$$\mathbf{x}^{(1)'} = [1, x_1, x_2]$$

For k = 2 design variables and a second-order model, we have

$$\mathbf{x}^{(2)'} = [1, x_1, x_2, x_1^2, x_2^2, x_1x_2]$$

The scaled prediction variance is

$$v(\mathbf{x}) = \frac{N\,Var[\hat{y}(\mathbf{x})]}{\sigma^2} = N\mathbf{x}^{(m)'}\mathbf{(X'X)}^{-1}\mathbf{x}^{(m)} \tag{3-19}$$

where N is the design size. It is clear that the scaled prediction variance depends on the experimental design ($(\mathbf{X'X})^{-1}$), the model, and the particular location ($\mathbf{x}^{(m)}$). For a rotatable design, the scaled prediction variance is the same for all points of $\mathbf{x}^{(m)}$ that have the same distance from the center point of the design. Therefore, the scaled prediction variance is constant at the axial points or on the spheres. Myers and Montgomery (2002) emphasized the importance of the idea of design rotatability to impose stability on $N\,Var[\hat{y}(\mathbf{x})]/\sigma^2$. They point out that rotatability or near-rotatability is often very easy to achieve without the sacrifice of other important design properties. However, it is not necessary to have exact rotatability in the second-order design. In other words, near-rotatable design is also acceptable. For example, the k = 3 CCD may have $\alpha = 1.682$ for the axial points. To stabilize the scaled prediction variance, the spherical or near-spherical designs need three to five center runs to avoid a severe imbalance through the design region. On the other hand, rotatability or near-rotatability is not important when the design region is cuboidal. The cuboidal design is suitable for the problems that have strict ranges for the design variables. Such kind of design has the same region of interest and region of operability, and the region is called a cube.

In addition to comparing the design size, design optimality criteria that are characterized by letters of the alphabet are often used in design evaluation and construction. Evaluation and comparison of RSM designs are based on the theoretical works of Kiefer (1961) and Kiefer and Wolfowitz (1959) that discuss design optimality in a measure theoretic approach. They provide the basis for much of the design optimality criteria that are used for design selection. Among the optimality criteria, D-optimality and G-optimality are the most commonly used.

Define the determinant as $|\mathbf{M}| = |\mathbf{X'X}|/N^p$, where N is the number of experimental runs and *p* is the number of parameters in the model. A D-optimality design will maximize $|\mathbf{M}|$ over all designs ξ. Therefore, the D-efficiency of a particular design ξ^* can be defined as

$$D_{eff} = \left(\frac{|M(\xi^*)|}{\underset{\xi}{Max}|M(\xi)|}\right)^{1/p} \tag{3-20}$$

G-optimality and the corresponding G-efficiency are associated with scaled prediction variance $v(\mathbf{x})$. The G-efficiency for a design is

$$G_{eff} = \frac{p}{\underset{x \in R}{Max}\, v(\mathbf{x})} = \frac{p}{\underset{x \in R}{Max}\, N\mathbf{x}^{(m)'}(\mathbf{X'X})^{-1}\mathbf{x}^{(m)}} \tag{3-21}$$

where p is the number of parameters in the model.

3.9 Conclusions

In this chapter, response surface methodology (RSM) and statistical design of experiment (SDE) are discussed to solve robust design problems. The RSM framework can lead to the solution of robust design naturally. The dual response surface approach and generalized linear models are growing extensively as modeling and analysis tools. To fit the response surface models, combined array designs are discussed for both spherical regions of interest and cuboidal regions of interest. The modified (or standard) central composite designs (CCD) and face-centered cubes (FCC) are appropriate for robust design and in general have attractive variance properties.

Chapter 4
Advanced Robust Design Topics

4.1 Internal and External Noise Factors

Robust design is an engineering methodology that determines the levels of control factors under consideration of noise factors. Taguchi defines and uses two kinds of noise factors in the experimental design: external noise factors and internal noise factors. The external noise factors are the uncontrollable environmental variables, such as temperature and humidity during manufacture, applied pressure or force, etc. Though they cannot be controlled in the process, we can set them at fix values in the experiments (particularly it is convenient to fix them in the computer experiments). The internal noise factors represent the random variation of control variables due to the deviations of different components or processes, for example, manufacturing error for each part or each batch of parts, deviations from nominal processing conditions (like ambient temperature), etc. They vary around the nominal values of the control variables and transmit the variability to the response.

It is understandable that the variation of the system can be greatly reduced if both external and internal noises are included in the parameter design. Thus, the optimum nominal settings from the parameter design will make the system robust (or less sensitive) to not only external noise factors, but also to internal noise factors. As a result, the response variance can be reduced. In each experimental design, the experimenter should evaluate the overall benefits and the restricted resources to determine the selection of the noise factors. If some sources of variability, such as the internal noise factors, are not included in the experiment, the design result may be inaccurate or even potentially lead to a wrong solution. Most researchers are focusing on the external noises factors and in many cases the internal noises are not included in the experimental design. Such kind of design will result in robust products or processes against external noises only. One of the major reasons for inadequate attention to internal noises is the assumption that external noises dominate the effects to the system; however, this assumption is not always true. For example, in the chemical or manufacturing industry, some control variables (environmental temperature, humidity, etc.) will certainly vary within the upper and lower limits, and the unavoidable deviations will produce variability in the response. In such kind of experimental design, the standard deviations of the control factors are the internal noise factors and contribute much to the variability of the response.

The practitioner could think it is expensive or hard to plan a design that includes variation of each control factor in the design. The concern should be clarified that commonly the number of significant control factors is not large in the final design stage (after the screening experiment) and not all internal noises should be considered. Li and Wu (1999) divided the components or variables into two groups: those with tolerance requirement and those without tolerance requirement. Both the nominal values and tolerances are of interest for the first type of variables, whereas the second type of variables, such as qualitative variables, bake time, and spin speed, do not have tolerance requirement since they do not change after the nominal values are set. Moreover, the internal noises from the same source, for example, supplier, manufacturing, and lot of materials, may share the same distribution or effect to the system, so the internal noise factors can be combined into one or few variables in the real design. Finally, computer simulation may make the experiments possible to perform without doing any physical experiments.

In the combined array design approach, control factors, internal noise factors (standard deviations or tolerances), and external noise factors are combined in one array. The internal noise factors are hard to control; however, they are partially controllable through controlling the components or subassemblies manufacturing. We can put a limit on the random variability of a control variable by specifying its tolerance. Therefore, in a combined array design, the internal noises can be combined with control factors and external noise factors as part of the noise factors, or even as part of control factors in the response surface model or dual response model, and the optimal parameters and variations (or tolerances) can be achieved through integrated optimization. If we use the internal noise factors (z_x) as part of noise factors, the new response model is

$$y(\mathbf{x}, \mathbf{z}_x, \mathbf{z}) = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{x}'\mathbf{B}\mathbf{x} + (\mathbf{z}_x, \mathbf{z})'\boldsymbol{\gamma} + \mathbf{x}'\boldsymbol{\Delta}(\mathbf{z}_x, \mathbf{z}) + \varepsilon \tag{4-1}$$

If we use the internal noise factors (z_x) as part of control factors, the response model changes to

$$y(\mathbf{x}, \mathbf{z}_x, \mathbf{z}) = \beta_0 + (\mathbf{x}, \mathbf{z}_x)'\boldsymbol{\beta} + (\mathbf{x}, \mathbf{z}_x)'\mathbf{B}(\mathbf{x}, \mathbf{z}_x) + \mathbf{z}'\boldsymbol{\gamma} + \mathbf{x}'\boldsymbol{\Delta}\mathbf{z} + \varepsilon \tag{4-2}$$

The models of process mean and variance will also change based on the different roles of the internal noise factors. In the second model we can get more information for the internal noise factors (second order terms and the interactions between internal and external noise factors). We will illustrate the roles of the internal noise factors with the following example.

4.1.1 Example: Robust Design for a Diesel Fuel Injector

Consider the following example taken from Meng (2006). The product under study is a diesel fuel injector, and the fuel injection system is a high pressure common rail system. It is instructive to examine a simple case, so only one performance (response) and six design variables are selected. The response is the injected fuel at a specific engine operating condition and the objective of the robust design is to adjust the control variables to minimize the injected fuel variation. We have two control variables: pilot valve seat diameter (x_1) and pilot valve minimum air gap (x_2), two internal noise factors: variation of pilot valve seat diameter (z_1) and pilot valve minimum air gap (z_2), and two external noise factors: temperature of fuel (z_3) and supply pressure (z_4). Table 4-1 lists their initial nominal values and tolerances. Recall that the internal noises result from variation in the control variables originating from the manufacturing process. In general, we assign the statistical tolerance of each design variable as three times of its standard deviation ($tol_i = \pm 3\sigma_i$).

Table 4-1 Initial Settings of the Control and Noise Factors

Variable	Nominal	Tolerance
x_1 (mm)	0.550	0.05
x_2 (mm)	0.080	±0.005
z_3 (degree)	25	±10
z_4 (bar)	1800	±5

The data of the experimental design is shown in Appendix B. The combined array design is a face-centered cube (FCC) design for a cuboidal design region. It consists of three distinct portions: fractional factorial design, axial points, and one center point. In this case the fractional factorial design is a resolution VI design 2_{VI}^{6-1} with defining relation I = $x_1x_2z_1z_2z_3z_4$ and the external noise factor supply pressure is $z_4 = x_1x_2z_1z_2z_3$. Since the axial points primarily contribute to the estimation of the quadratic terms and generally in robust design problems the pure quadratic terms of the noise variables are not significant, we used the modified FCC design by eliminating the axial points of the external noise factors (z_3 and z_4). However, we kept the axial points of internal noise factors (z_1 and z_2) as ±1 in the four experimental runs, to fit the model when we treat the internal noise factors as control factors. In FCC design, though two center runs are slightly preferable, one center run is quite sufficient for design stability.

We conducted the robust design in three situations with response surface approach:

Case a) No internal noises are considered. The control factors consist of x_1 and x_2, while the noise factors consist of z_3 and z_4.

Case b) Consider the internal noise factors as noise factors, as they are hard to control. This is the traditional way to construct the parameter design. The control factors include x_1 and x_2, while the noise factors consist of z_1, z_2, z_3, and z_4.

Case c) Consider the internal noise factors as control factors as an alternative for the parameter design. We define x_1, x_2, z_1, and z_2 as control factors , and z_3 and z_4 as noise factors.

Case a) No Internal Noises Considered in the Parameter Design

The design variables consist of x_1, x_2, z_3, and z_4. The second-order regression model is

$$\begin{aligned}\hat{y}(\mathbf{x},\mathbf{z}) = {} & 91.8008 + 1.3606x_1 - 0.5715x_2 + 0.2338x_1^2 + 0.0688x_2^2 + 0.0166x_1x_2 \\ & + 0.1566z_3 - 0.0772z_4 - 0.0091x_1z_3 - 0.0041x_1z_4 + 0.0016x_2z_3 \\ & + 0.0466x_2z_4\end{aligned}$$

The sum of squares of the residuals (SS_E) and the total sum of squares (SS_T) are

$$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 11.1369,\ SS_T = y'y - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 86.426$$

In this case, the coefficient of multiple determination is

$$R^2 = 1 - \frac{SS_E}{SS_T} = 0.8711$$

The model explains the variability reasonable well. The normal probability plot Figure 4-1 reveals no apparent problem with normality.

The process mean model is

$$\hat{E}_z[y(\mathbf{x},\mathbf{z})] = 91.8008 + 1.3606x_1 - 0.5715x_2 + 0.2338x_1^2 + 0.0688x_2^2 + 0.0166x_1x_2$$

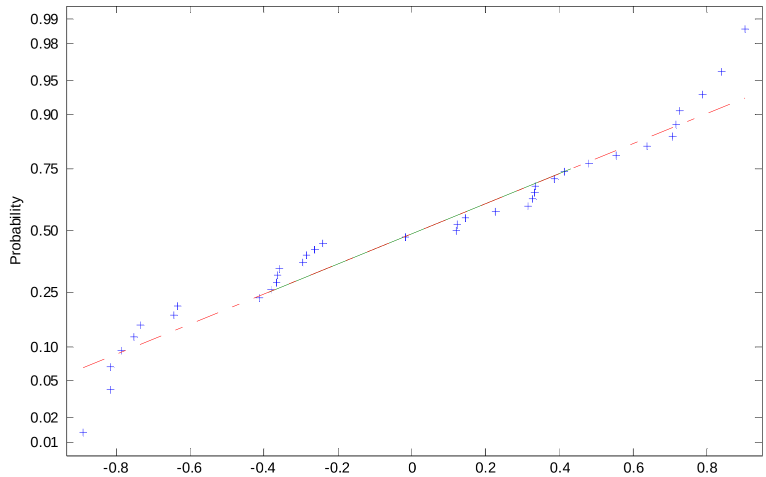

Figure 4-1 Normal probability of residuals for the response

The slopes in the direction of noise variables z_3 and z_4 are

$$l_{z_3}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_3} = 0.1566 - 0.0091x_1 + 0.0016x_2$$

$$l_{z_4}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_4} = -0.0772 - 0.0041x_1 + 0.0466x_2$$

The estimated response variance is $\sigma^2 = \frac{SS_E}{n-p} = 0.4455$ and we assumed $\sigma_{z_3} = \sigma_{z_4} = \sigma_z = 1$, then the process variance model is

$$\begin{aligned}\hat{Var}_z[y(\mathbf{x},\mathbf{z})] &= \sigma_z^2[l'_{z_3}(\mathbf{x})l_{z_3}(\mathbf{x}) + l'_{z_4}(\mathbf{x})l_{z_4}(\mathbf{x})] + \sigma^2 \\ &= 0.476 - 0.0022x_1 - 0.0067x_2 + 0.0001x_1^2 + 0.0022x_2^2 - 0.0004x_1x_2\end{aligned}$$

A relatively straightforward approach to optimizing the dual response models of the fuel injection system is to use the contour plot, response surface plot, and the overlay plot. Figure 4-2 and Figure 4-3 present the response surface plots of the mean model and standard deviation model (square root of the variance). Notice from these plots that the mean increases as x_1 increases and the mean does not change much as x_2 changes. It is clear that the standard deviation decreases as x_2 increases and the standard deviation does not change much as x_1 changes. In this problem it is required to keep the process mean between 91.7 and 92.1. Figure 4-4 shows an overlay plot of the contours of the mean and standard deviation. The tradeoffs can be seen from the overlay plot for better understanding of the process. To achieve the desired objective and satisfy the constraints, it is necessary to hold x_2 at the high level and x_1 around 0.5.

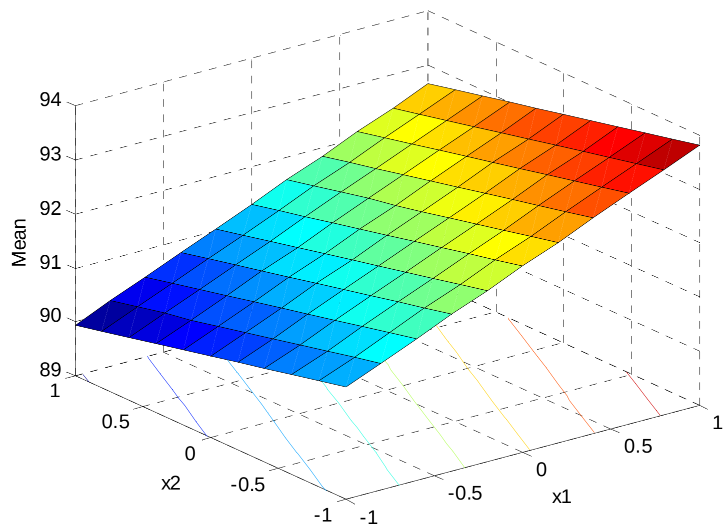

Figure 4-2 Response surface plot of the mean model

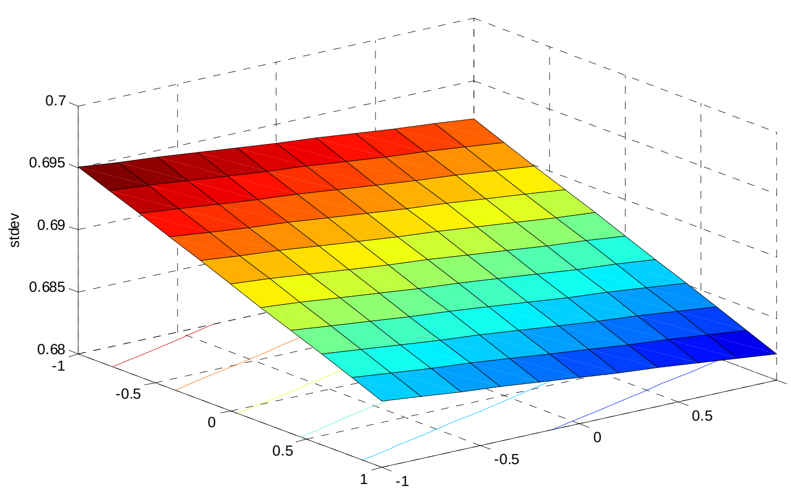

Figure 4-3 Response surface plot of the standard deviation model

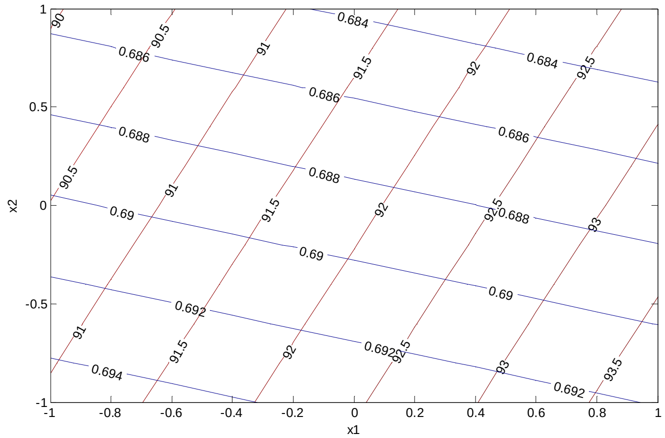

Figure 4-4 Overlay plot of mean and standard deviation contour

Constrained optimization was conducted to find the optimal operating conditions of the design variables x_1 and x_2. Recall that dual RSM and MSE loss function can be used to construct the optimization. In this case study, we used the dual RSM to formulate the objective function and constraint as following:

$$\begin{gathered} Min \quad \hat{Var}_z[y(\mathbf{x},\mathbf{z})] \\ subject\ to \\ 91.7 \le \hat{E}_z[y(\mathbf{x},\mathbf{z})] \le 92.1 \\ -1 \le x_1 \le 1 \quad and \quad -1 \le x_2 \le 1 \end{gathered}$$

We used Matlab to solve the optimization problem. The optimal settings are

$$x_1 = 0.5339,\ x_2 = 1,\ E_z[y(\mathbf{x},\mathbf{z})] = 92.1,\ Var_z[y(\mathbf{x},\mathbf{z})] = 0.4701$$

Case b) Consider Internal Noises as Noise Factors

In the second case, the control variables consist of x_1 and x_2, and the noise variables consists of z_1, z_2, z_3, and z_4. The second-order regression model is

$$\begin{aligned} \hat{y}(x,z) = {} & 91.9761 + 1.3606\mathrm{x}_1 - 0.5715\mathrm{x}_2 + 0.1435\mathrm{x}_1^2 - 0.0215\mathrm{x}_2^2 + 0.0166x_1x_2 + 0.1074z_1 - 0.42z_2 \\ & + 0.1566z_3 - 0.0772z_4 + 0.1253x_1z_1 - 0.1216x_1z_2 - 0.0091x_1z_3 - 0.0041x_1z_4 + 0.0709x_2z_1 \\ & - 0.0897\mathrm{x}_2z_2 + 0.0016x_2z_3 + 0.0466x_2z_4 \end{aligned}$$

The sum of squares of the residuals (SS_E) and the total sum of squares (SS_T) are

$$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 3.5751,\ SS_T = y'y - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 86.558$$

The coefficient of multiple determination is

$$R^2 = 1 - \frac{SS_E}{SS_T} = 0.9587$$

The model explains the variability better (95.87% of the variability) than the first case. The normal probability plot Figure 4-5 reveals no apparent problem with normality.

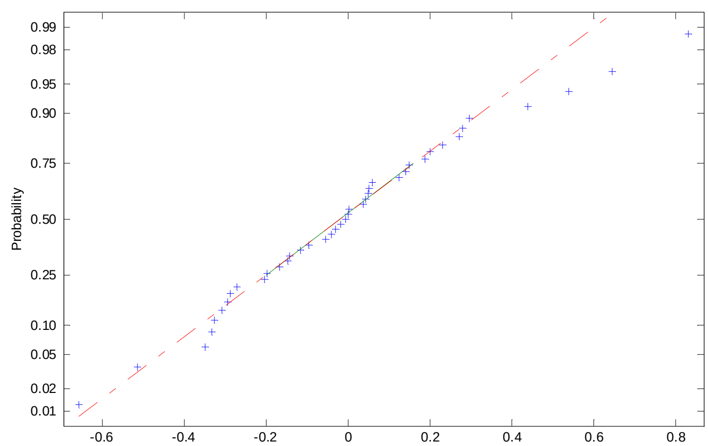

Figure 4-5 Normal probability of residuals for the response

The process mean model is

$$\hat{E}_z[y(\mathbf{x},\mathbf{z})] = 91.9761 + 1.3606x_1 - 0.5715x_2 + 0.1435x_1^2 + 0.0215x_2^2 + 0.0166x_1x_2$$

The slopes in the direction of noise variables z_1 , z_2 , z_3 , and z_4 are

$$l_{z_1}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_1} = 0.1074 + 0.1253x_1 + 0.0709x_2$$

$$l_{z_2}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_2} = -0.42 - 0.1216x_1 - 0.0897x_2$$

$$l_{z_3}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_3} = 0.1566 - 0.0091x_1 + 0.0016x_2$$

$$l_{z_4}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_4} = -0.0772 - 0.0041x_1 + 0.0466x_2$$

The estimated response variance is $\sigma^2 = \frac{SS_E}{n-p} = 0.1554$ and we assumed

$$\sigma_{z_1} = \sigma_{z_2} = \sigma_{z_3} = \sigma_{z_4} = \sigma_z = 1$$

then the process variance model is

$$\begin{aligned}\hat{Var}_z[y(\mathbf{x},\mathbf{z})] &= \sigma_z^2[l'_{z_1}(\mathbf{x})l_{z_1}(\mathbf{x}) + l'_{z_2}(\mathbf{x})l_{z_2}(\mathbf{x}) + l'_{z_3}(\mathbf{x})l_{z_3}(\mathbf{x}) + l'_{z_4}(\mathbf{x})l_{z_4}(\mathbf{x})] + \sigma^2 \\ &= 0.3738 + 0.1268x_1 + 0.0839x_2 + 0.0306x_1^2 + 0.0153x_2^2 + 0.0392x_1x_2\end{aligned}$$

Figure 4-6 and Figure 4-7 present the response surface plot of the mean model and the model of square root of the variance (standard deviation). Notice from these plots that the mean increases as x_1 increases and the mean does not change much as x_2 changes. It is clear that the standard deviation decreases as either x_1 or x_2 decreases. Again it is required to keep the process mean between 91.7 and 92.1. Figure 4-8 shows an overlay plot of the contours of the mean and standard deviation. The tradeoffs can be seen from the overlay plot for better understanding of the process. To achieve the desired objective and satisfy the constraints, it is necessary to hold x_2 at the low level and x_1 around -0.5.

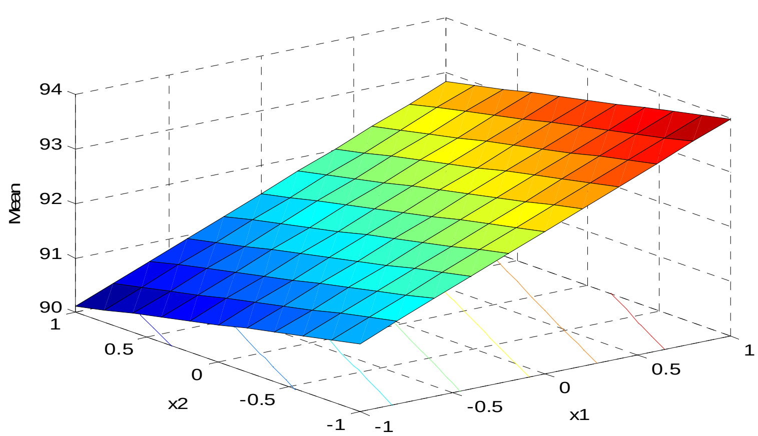

Figure 4-6 Response surface plot of the mean model

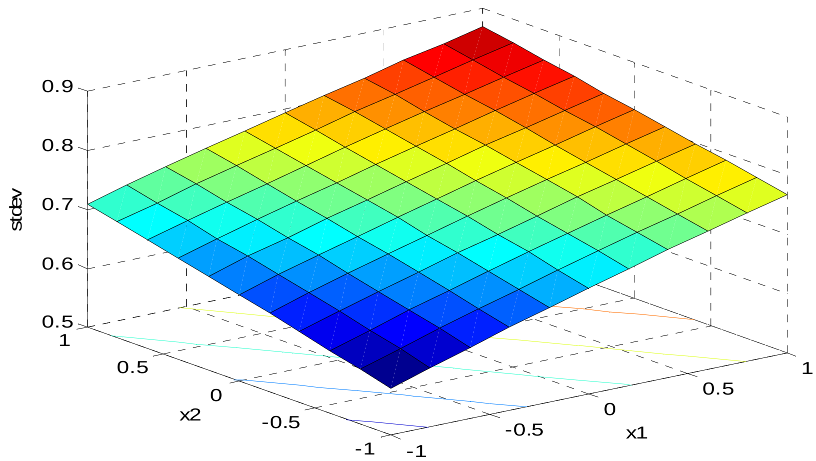

Figure 4-7 Response surface plot of the standard deviation model

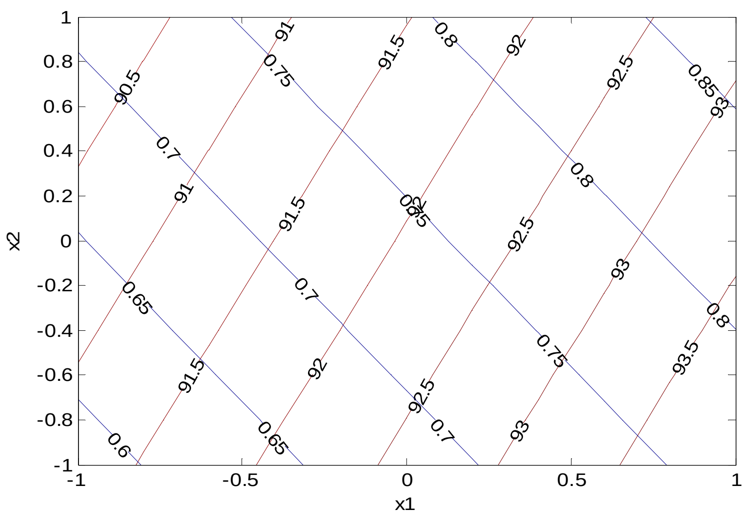

Figure 4-8 Overlay plot of mean and standard deviation contours

We conducted optimization to find the optimal setting of the design variables x_1 and x_2 through minimizing the process variance

$$\begin{gathered} Min \quad \hat{Var}_z[y(\mathbf{x},\mathbf{z})] \\ subject\ to \\ 91.7 \le \hat{E}_z[y(\mathbf{x},\mathbf{z})] \le 92.1 \\ -1 \le x_1 \le 1 \quad and \quad -1 \le x_2 \le 1 \end{gathered}$$

Matlab was used to solve the optimization problem. The optimal settings are

$$x_1 = -0.6614\,,\ x_2 = -1\,,\ E_z[y(\mathbf{x},\mathbf{z})] = 91.7\,,\ Var_z[y(\mathbf{x},\mathbf{z})] = 0.2606$$

Case c) Consider Internal Noises as Control Factors

In the third experiment, we try a new method by using the internal noise factors as control factors, so the model of mean and variance become the function of control factors and their tolerances. The control variables consists of x_1, x_2, z_1, and z_2, and the noise variables consists of z_3 and z_4. The second-order regression model is

$$\begin{aligned}\hat{y}(x,z) = {} & 91.9316 + 1.3606x_1 - 0.5715x_2 + 0.0414x_1^2 - 0.1236x_2^2 + 0.0166x_1x_2 + 0.1074z_1 \\ & - 0.42z_2 - 0.0336z_1^2 + 0.2914z_2^2 - 0.1234z_1z_2 + 0.1253x_1z_1 - 0.1216x_1z_2 \\ & + 0.0709x_2z_1 - 0.0897x_2z_2 + 0.1566z_3 - 0.0772z_4 - 0.0091x_1z_3 - 0.0041x_1z_4 \\ & + 0.0016x_2z_3 + 0.0466x_2z_4 + 0.0028z_1z_3 - 0.0278z_1z_4 + 0.0347z_2z_4 \\ & + 0.0022z_2z_4\end{aligned}$$

The sum of squares of the residuals (SS_E) and the total sum of squares (SS_T) are

$$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 2.7991,\ SS_T = y'y - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 86.558$$

The coefficient of multiple determination is

$$R^2 = 1 - \frac{SS_E}{SS_T} = 0.9677$$

This model contains more terms, and it explains the variability better than other two methods. The normal probability plot (Figure 4-9) reveals no apparent problem with normality.

The process mean model is

$$\begin{aligned}\hat{E}_z[(x,z)] = {} & 91.9316 + 1.3606x_1 - 0.5715x_2 + 0.0414x_1^2 - 0.1236x_2^2 + 0.0166x_1x_2 + 0.1074z_1 \\ & - 0.42z_2 - 0.0336z_1^2 + 0.2914z_2^2 - 0.1234z_1z_2 + 0.1253x_1z_1 - 0.1216x_1z_2 \\ & + 0.0709x_2z_1 - 0.0897x_2z_2\end{aligned}$$

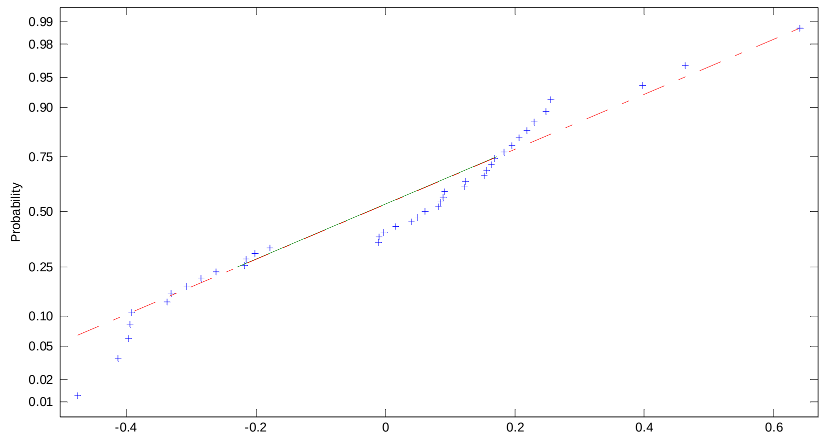

Figure 4-9 Normal probability of residuals for the response

The slopes in the direction of noise variables z_3 and z_4 are

$$l_{z_3}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_3} = 0.1566 - 0.0091x_1 + 0.0016x_2 + 0.0028z_1 + 0.0347z_2$$

$$l_{z_4}(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x},\mathbf{z})}{\partial z_4} = -0.0772 - 0.0041x_1 + 0.0466x_2 - 0.0278z_1 + 0.0022z_2$$

The estimated response variance is $\sigma^2 = \dfrac{SS_E}{n-p} = 0.1749$ and we assumed $\sigma_{z_3} = \sigma_{z_4} = \sigma_z = 1$,

then the process variance model is

$$\begin{aligned}\hat{Var}_z(y(x,z)) &= \sigma_z^2[l'_{z_3}(x)l_{z_3}(x) + l'_{z_4}(x)l_{z_4}(x)] + \sigma^2 \\ &= 0.2534 - 0.0022x_1 - 0.0067x_2 + 0.0052z_1 + 0.0105z_2 \\ &\quad + 0.0001x_1^2 + 0.0022x_2^2 + 0.0008z_1^2 + 0.0012z_2^2 \\ &\quad - 0.0004x_1x_2 + 0.0002x_1z_1 - 0.0006x_1z_2 \\ &\quad - 0.0026x_2z_1 + 0.0003x_2z_2 + 0.0001z_1z_2\end{aligned}$$

Figure 4-10 and Figure 4-11 present the response surface plot of the mean model and the model of square root of the variance (the standard deviation). To construct these plots, we held the two internal noise factors z_1 and z_2 at -1. Notice from these plots that the mean increases as x_1 increases and the mean does not change much as x_2 changes. It is clear that the standard deviation decreases as x_2

increases and the standard deviation does not change much as x_1 changes. Again in this problem it is required to keep the process mean between 91.7 and 92.1. Figure 4-12 shows an overlay plot of the contours of the mean and standard deviation. There are tradeoffs between the mean and variance. To achieve the desired objective and satisfy the constraints, it is necessary to hold x_2 at the high level and x_1 around 0.3.

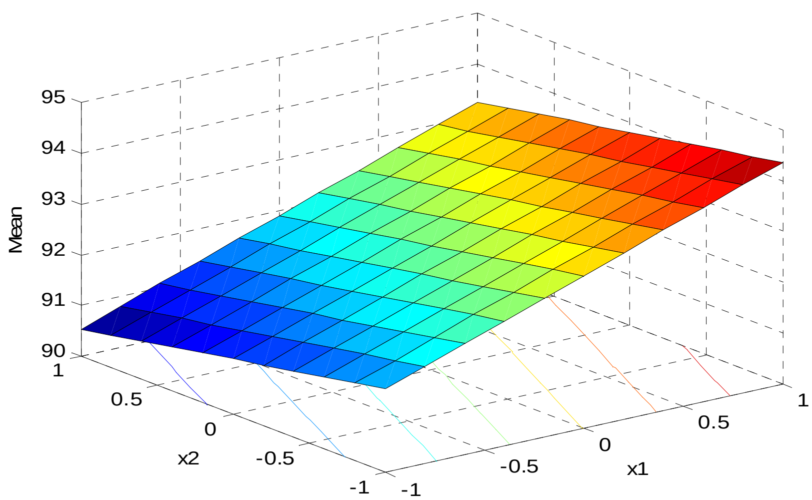

Figure 4-10 Response surface plot of the mean model

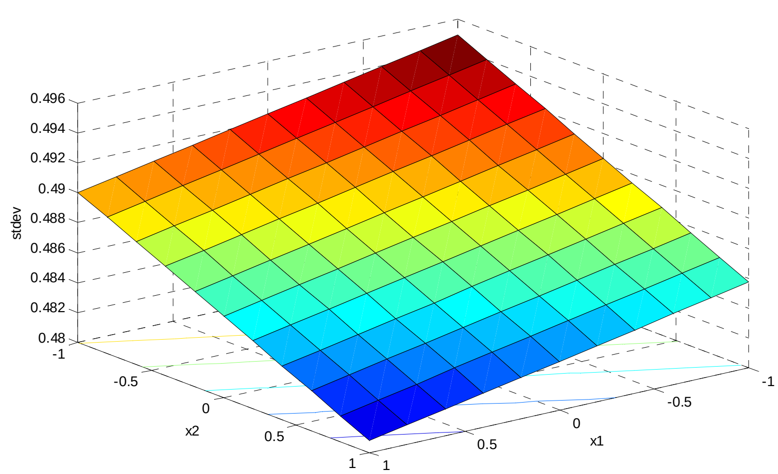

Figure 4-11 Response surface plot of the standard deviation model

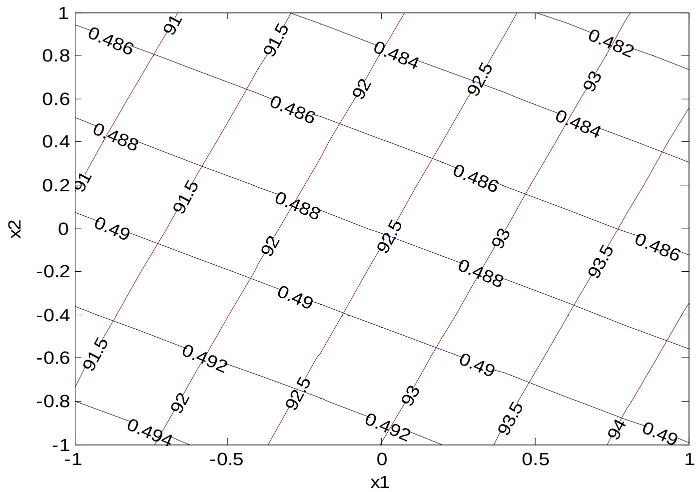

Figure 4-12 Overlay plot of mean and standard deviation contour

We constructed optimization as following to find the optimal settings of the design variables x_1 and x_2 through minimizing the process variance.

$$\begin{gathered} Min \quad \hat{Var}_z[y(\mathbf{x},\mathbf{z})] \\ subject\ to \\ 91.7 \le \hat{E}_z[y(\mathbf{x},\mathbf{z})] \le 92.1 \\ -1 \le x_1 \le 1 \ \ and \ \ -1 \le x_2 \le 1 \\ -1 \le z_1 \le 1 \ \ and \ \ -1 \le z_2 \le 1 \end{gathered}$$

Matlab was used again to solve the optimization problem. The optimal settings are

$$x_1 = 0.2871,\ x_2 = 1,\ E_z[y(\mathbf{x},\mathbf{z})] = 92.1,\ Var_z[y(\mathbf{x},\mathbf{z})] = 0.237$$

Table 4-2 shows the results from the above three cases. The comparison of the initial settings (control factors x_1 = 0 and x_2 = 0) and the different optimal operating conditions confirms that all of the three methods reduce the variability of the response through design optimization. When the internal noise factors are considered, as in the second and third case, the models are more adequate to express the response and the response variance can be reduced further. Consideration of the internal noise factors will increase the complex of the robust design problem and the simulation work can be time consuming. However, the accuracy is improved as illustrated in the case study. We can conclude that considering the internal noise factors as part of control variables in the combined array design is an attractive alternative to the traditional method that models the internal noise factors as part of the noise variables.

Table 4-2 Comparison of the Three Cases

	Case a) Ignore internal noises (no z_1 and z_2)	Case b) z_1 and z_2 as noise factors	Case c) z_1 and z_2 as control factors
x_1	0.5339	-0.6614	0.2871
x_2	1	-1	1
R^2	87.11%	95.87%	96.77%
$Var_z[y(\mathbf{x},\mathbf{z})]$ (initial)	0.476	0.3738	0.2534
$Var_z[y(\mathbf{x},\mathbf{z})]$ (optimal)	0.4701	0.2606	0.237
$E_z[y(\mathbf{x},\mathbf{z})]$	92.1	91.7	92.1

In the third case where the internal noise factors are used as control factors, though we may obtain the optimal setting of z_1 and z_2 (the variation of x_1 and x_2) from the optimization, we can not simply use them to compute the tolerance of x_1 and x_2. In this example, the optimal values of z_1 and z_2 are only optimization results without weighing the manufacturing cost and loss-of-quality cost. To decide the accurate tolerance range, economic information and cost study should be integrated with the parameter design.

4.2 Tolerance Design

Robust tolerance design is used to determine the best limits for the parameters to meet the variation requirement of the design. The ideal tolerances would be zero in a deterministic design if we do not consider the economical and physical constraints. In general we perform constrained optimization to minimize the objective function - total system cost (C_T), which is the sum of tolerance cost or production cost (C_P) and the so-called loss-of-quality cost (Q), with respect to the parameter constraints to find the optimal tolerances. We can also formulate the optimization in other ways, such as using response mean, response variance, or any one of the cost measures as the objective function under the constraints of others. If the variation (standard deviation) for a design variable is σ, it is common to assign the tolerance to be $\pm 3\sigma$.

The integrated robust parameter and robust design, which performs parameter design and tolerance design simultaneously, is a superior alternative to the traditional two-stage sequential parameter and tolerance design. However, it is still very popular to design parameters and tolerances separately, or in many cases, only parameter design is done. There might be three major reasons:

1. It is assumed that tolerance design does not affect the nominal values of control variables. This assumption will affect the accuracy of robust design, since the deviations of the control variables do influence the selection of the nominal values, even though the influence might be small.

2. It might be assumed that parameter design contributes mainly to the design of a system, so researchers focus on parameter design and have developed a systemic framework. There is much less research works have been done on tolerance design.

3. Tolerance design needs the information of cost that is difficult or impossible to obtain (such as the coefficients of the production cost and loss-of-quality cost). Because it is difficult to collect or quantify the data of cost related to different stages of the whole life cycle, only inaccurate or assumed values could be used. Therefore, tolerance design is more complex than parameter design, and the final decision should be made by experienced researchers and engineers based on practical conditions.

4.2.1 Total Cost

The total cost is the sum of loss-of-quality cost and production cost:

$$C_T = Q + C_P = \mathbf{K}[(\boldsymbol{\mu}_{\mathbf{y}} - \mathbf{T})^2 + \boldsymbol{\sigma}_{\mathbf{y}}^{\mathbf{2}}] + C_p \tag{4-3}$$

The production cost and the loss-of-quality cost are competing costs. Figure 4-13 is a simple example from Savage (2008) to show the relationship between cost and quality. We can see the trade-off between them clearly. In general, tight tolerance is preferred at the design stage but that leads to higher cost and more difficulties in the manufacturing stage. Therefore, the total cost provides a single objective function in terms of the means and tolerances of the control variables and optimization should be performed to find the minimum cost within acceptable quality level. This is a typical yield-cost scenario.

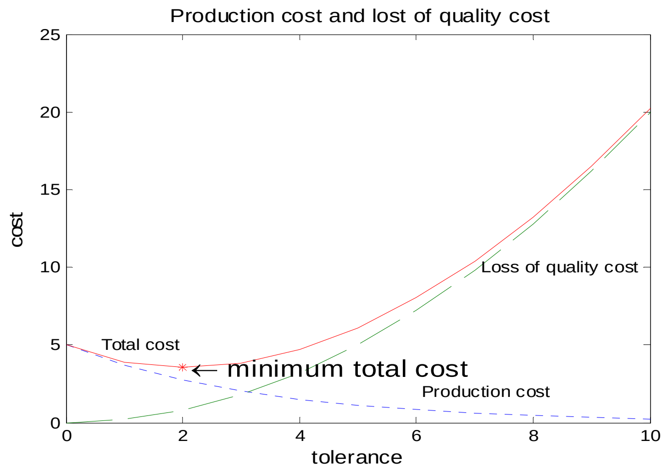

Figure 4-13 Production cost and loss-of-quality cost

4.2.2 Loss-of-Quality Cost

Taguchi (1986) contended that a loss-of-quality cost increases when product quality characteristics (responses) differ from the nominal. He creates a univariate quadratic loss function of the form

$$L(y) = k(y - T)^2 \tag{4-4}$$

where k is the cost coefficient, y is the response, and T is the nominal of the response. When the response is right on the target T, the loss will be zero. The smaller the deviation, the smaller the loss will be. In traditional quality thinking, there is no additional cost when y is within the upper and lower limits. Taguchi's loss function, however, suggests that even a small deviation of the response y will lead to a loss, so it is consistent with the continuous improvement philosophy of modern quality engineering. Figure 4-14 shows Taguchi's quadratic loss function.

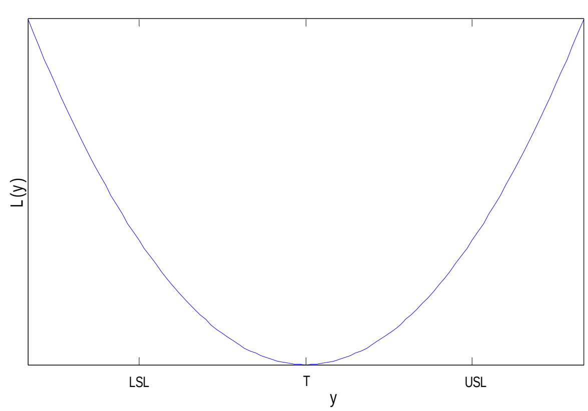

Figure 4-14 Taguchi's quadratic loss function

From the loss function, we can determine the expected loss function

$$Q = E[L(y)] = k[(\mu_y - T)^2 + \sigma_y^2] \tag{4-5}$$

where μ_y and σ_y^2 are the mean and variance of the response y. The cost coefficient *k* can be evaluated as

$$k = \frac{A_0}{\Delta_0^2} \tag{4-6}$$

where

A_0 - the cost of a response exceeding the critical levels. It includes the cost for rework, repair, scrap, market loss, and all other financial losses.

$\Delta_0 = \dfrac{USL - LSL}{2}$ - the response tolerance

In addition to the dual RSM approach, Taylor's series expansion can be used as an alternative to approximate the process mean and variance in terms of control variables and their standard deviations (or tolerances). Taylor approximation is useful to formulate the cost function for the constrained optimization, particularly when the internal noise factors (tolerances of the control variables) are not included in the experimental designs. Consider $y = g(x)$ and we have the mean and variance of random variable x, $E(x) = \mu$ and $Var(x) = \sigma^2$, the Taylor's approximation is

$$y \approx g(\mu) + (x - \mu)\left.\frac{dg(x)}{dx}\right|_{\mu} + \frac{1}{2}(x - \mu)^2 \left.\frac{d^2 g(x)}{dx^2}\right|_{\mu}$$

Thus, the process mean and variance can be obtained as

$$E(y) \approx g(\mu) + \frac{1}{2}\sigma^2 \left.\frac{d^2 g(x)}{dx^2}\right|_{\mu}$$

$$Var(y) \approx \sigma^2 [\left.\frac{dg(x)}{dx}\right|_{\mu}]^2$$

Next, extend the above approximations to the case where $y = g(\mathbf{x})$ and $\mathbf{x}$ is a vector of random variables, $\mathbf{x} = (x_1, x_2, ..., x_n)'$. The Taylor's approximation is

$$y \approx g[E(\mathbf{x})] + \sum_{i=1}^{n} \frac{\partial g(\mathbf{x})}{\partial x_i}\bigg|_{E(\mathbf{x})} [x_i - E(x_i)] + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 g(\mathbf{x})}{\partial x_x x_j}\bigg|_{E(\mathbf{x})} [x_i - E(x_i)][x_j - E(x_j)]$$

Therefore the process mean and variance can be obtained as

$$E(y) \approx g[E(\mathbf{x})] + \frac{1}{2}\sum_{i-1}^{n}\sum_{j=1}^{n} \frac{\partial^2 g(\mathbf{x})}{\partial x_i x_j}\bigg|_{E(\mathbf{x})} Cov(x_i x_j)$$

$$Var(y) \approx \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial g(\mathbf{x})}{\partial x_i}\bigg|_{E(\mathbf{x})} \frac{\partial g(\mathbf{x})}{\partial x_j}\bigg|_{E(\mathbf{x})} Cov(x_i x_j)$$

Vining (1998) extended the univariate quadratic loss function to multivariate form based on square error loss approach. His loss function is

$$L = [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] \tag{4-7}$$

where

$\hat{\mathbf{y}}(\mathbf{x})$ - vector of estimated responses

$\mathbf{C}$ - positive definite matrix of weight or costs

$\mathbf{T}$ - vector of target values of the responses.

The expected loss function (Vining 1998) is given by

$$E(L) = \{E[\hat{\mathbf{y}}(\mathbf{x})] - \mathbf{T}\}'\mathbf{C}\{E[\hat{\mathbf{y}}(\mathbf{x})] - \mathbf{T}\} + Trace[\mathbf{C}\Sigma_{\hat{\mathbf{y}}(\mathbf{x})}] \tag{4-8}$$

And the estimated expected loss function is

$$\hat{E}(L) = [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] + Trace[\mathbf{C}\Sigma_{\hat{\mathbf{y}}(\mathbf{x})}] \tag{4-9}$$

where

$[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]$ -- loss for any predict value away from target

$Trace[\mathbf{C}\Sigma_{\hat{\mathbf{y}}(\mathbf{x})}]$ -- loss generated from the quality of the prediction

$\sum_{\hat{\mathbf{y}}(\mathbf{x})}$ -- variance-covariance matrix for the estimated responses

Let variance-covariance matrix for the responses be $\sum = \left[\hat{\sigma}_{ij}^2\right]$, which is known in advance or can be obtained through maximum likelihood estimation based on the experimental data. Recall from Section 3.8.4 the prediction variance is given as

$$Var[\hat{\mathbf{y}}(\mathbf{x})] = \mathbf{x}^{(m)'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{(m)}\boldsymbol{\sigma}^2$$

where $\mathbf{X}$ is the model matrix for the response, the vector $\mathbf{x}^{(m)}$ is a function of the location in the design variables at which one predicts the response, and the (m) reflects the model as $\mathbf{X}$ does. In the case of multiple responses, the prediction variance-covariance matrix for the estimated responses is

$$\sum_{\hat{\mathbf{y}}(\mathbf{x})} = \mathbf{x}^{(m)'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{(m)} \sum \tag{4-10}$$

The object of the optimization is to find the nominal values which minimize expected loss function E(L). The loss-of-quality cost is determined by the cost coefficient $\mathbf{C} = \mathbf{K}\,\hat{\sum}^{-1}$, where $\mathbf{K}$ is a matrix with the diagonal elements reflecting the economic importance of each response and the off-diagonal elements measuring the correlation of the responses. The estimated expected loss function can be given as

$$\begin{aligned}\hat{E}(L) &= [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] + Trace[\mathbf{C}\sum_{\hat{\mathbf{y}}(\mathbf{x})}] \\ &= [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] + Trace\{\mathbf{C}\mathbf{x}^{(m)'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{(m)}\left[\hat{\sigma}_{ij}^2\right]\}\end{aligned} \tag{4-11}$$

or

$$\begin{aligned}\hat{E}(L) &= [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] + Trace[\mathbf{K}\,\hat{\sum}^{-1}\,\mathbf{x}^{(m)'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{(m)}\sum] \\ &= [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] + Trace[\mathbf{K}\mathbf{x}^{(m)'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{(m)}]\end{aligned} \tag{4-12}$$

The optimization problem for the expected loss will consider quality of the responses and model prediction. It can be solved through nonlinear programming method in following form:

$$\begin{aligned}&Min \quad \hat{E}(L) \\ &subject\ to \\ &\mathbf{x} \in R\end{aligned} \tag{4-13}$$

Example: Multivariate Loss Approach for Multi-response Optimization

We take an example from Myers and Montgomery (2002) to show the squared error loss approach for multi-response optimization. The goal of the polymer experiment is to find the parameters of three variables in a chemical process to meet the requirements for two responses. The design variables are time (x_1), reaction temperature (x_2), and amount of catalyst (x_3). The responses are conversion of a polymer (y_1) and thermal activity (y_2).

Table 4-3 shows the experimental results. The acceptable range for y_1 is 80 to 100 and that for y_2 is 55 to 60. As the experimenters want to maximize y_1 and achieve a target value of y_2 (57.5), the target value for y_1 is at upper limit ($T_1 = 100$), and nominal for y_2 is at the midpoint ($T_2 = 57.5$).

Table 4-3 Experimental Results of Polymer Experiment

Run number	x_1 Time	x_2 Temperature	x_3 Catalyst	y_1 Conversion	y_2 Activity
1	-1	-1	-1	74	53.20
2	1	-1	-1	51	62.90
3	-1	1	-1	88	53.40
4	1	1	-1	70	62.60
5	-1	-1	1	71	57.30
6	1	-1	1	90	67.90
7	-1	1	1	66	59.80
8	1	1	1	97	67.80
9	-1.682	0	0	76	59.10
10	1.682	0	0	79	65.90
11	0	-1.682	0	85	60.00
12	0	1.682	0	97	60.70
13	0	0	-1.682	55	57.40
14	0	0	1.682	81	63.20
15	0	0	0	81	59.20
16	0	0	0	75	60.40
17	0	0	0	76	59.10
18	0	0	0	83	60.60
19	0	0	0	80	60.80
20	0	0	0	91	58.90

This experiment uses standard CCD design, which has $2^k = 2^3 = 8$ factorial points, 6 axial points ($\alpha = \sqrt[4]{F} = \sqrt[4]{8} = 1.682$), and 6 center points, to fit the second-order models. We used regression model to obtain the estimated response $\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{X}\boldsymbol{\beta}$, where the least squares estimator $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. The prediction equations for conversion (y_1) and thermal activity (y_2) are

$$\hat{y}_1 = 81.09 + 1.03x_1 + 4.04x_2 + 6.20x_3 - 1.83x_1^2 + 2.94x_2^2 - 5.19x_3^2 + 2.13x_1x_2 + 11.38x_1x_3 - 3.88x_2x_3$$

$$\hat{y}_2 = 60.51 + 3.58x_1 + 2.23x_3$$

The graphic methods were used to estimate the optimal operating conditions. Figure 4-15 and Figure 4-16 present the response surface plots of the two responses, and Figure 4-17 shows the overlay contour plot of the two responses. Since the model of $\hat{y}_1$ includes three design variables and $\hat{y}_2$ includes two design variables, we fixed x_2 (at 1.68) to plot the response surfaces of $\hat{y}_1$ or $\hat{y}_2$ versus x_1 and x_3. From the plot of $\hat{y}_1$ we find that x_1 and x_3 are competing: as x_1 increases $\hat{y}_1$ increases, while as x_2 increases $\hat{y}_1$ decreases. The overlay plot shows $\hat{y}_1$ and $\hat{y}_2$ are also competing. We can reach an approximate solution from these plots. The formal nonlinear programming can decide the trade-offs.

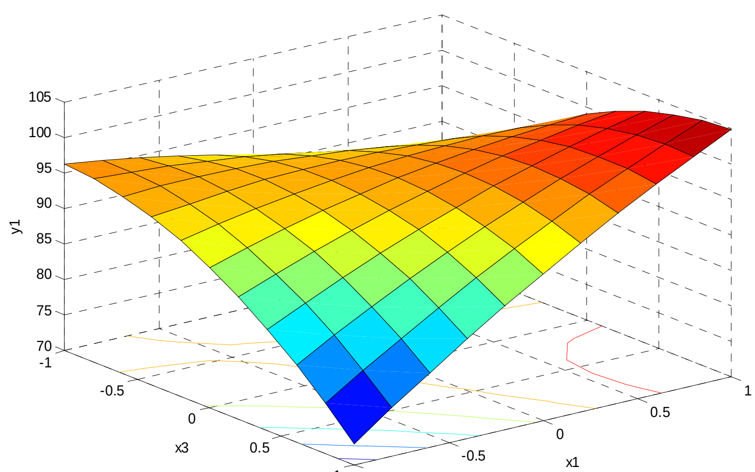

Figure 4-15 Response surface plot of conversion y_1

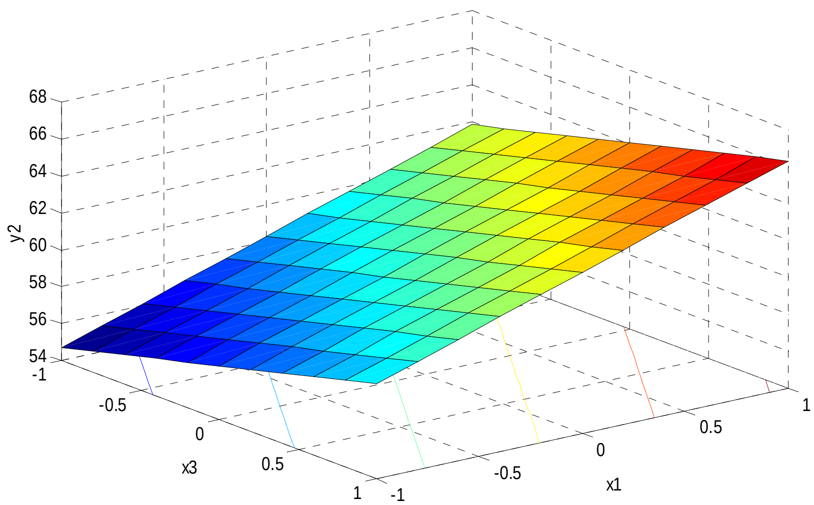

Figure 4-16 Response surface plot of activity y_2

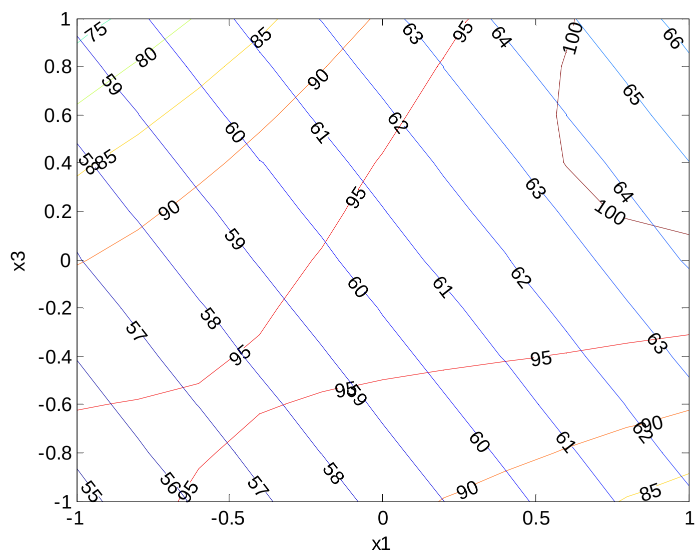

Figure 4-17 Overlay contour plot of conversion y_1 and activity y_2

As $\hat{\sigma}_{ij} = \hat{\varepsilon}_i' \hat{\varepsilon}_j' / N$, where $\hat{\varepsilon}$ is the residual vector from the ordinary least squares (OLS) estimation of each response, we can estimate the variance-covariance matrix of the responses as

$$\hat{\Sigma} = \begin{bmatrix} 11.1240 & -0.5450 \\ -0.5450 & 2.2012 \end{bmatrix}$$

In this example, we assume **K** = **I**, so cost efficient **C** is

$$C = \hat{\Sigma}^{-1} = \begin{bmatrix} 0.0910 & 0.0225 \\ 0.0225 & 0.4599 \end{bmatrix}$$

Next, we can perform the constrained optimization within experimental region to find the optimal values of the design variables:

$$Min \;\; \hat{E}(L) = [\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{T}] + Trace[\mathbf{x}^{(m)'}(\mathbf{X'X})^{-1}\mathbf{x}^{(m)}\mathbf{K}]$$
$$subject\ to$$
$$-1.68 \le x_1, x_2, x_3 \le 1.68$$

The optimal conditions of the design variables and responses are

x_1	x_2	x_3	y_1	y_2
-0.4041	1.682	-0.4863	95.2077	57.9789

And the loss for any value that is out of specification is 2.0921, loss generated from the quality of the prediction is 0.7593, and the total loss is 2.8514.

The settings of the two responses show that while both of y_1 and y_2 are close to their target values, y_1 obtains relatively better result than y_2. This means our choice of **C** gives more weight to y_1 than to y_2. If we assume that it is necessary to make the thermal activity (y_2) as close to the target as possible, more weight should be put on y_2 than on y_1. For example, suppose we assume the covariance of the responses is zero and select matrix of **K** and **C** as

$$\mathbf{K} = \begin{bmatrix} 11.124 & -0.545 \\ -54.5 & 220.12 \end{bmatrix} \text{ and } \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix}$$

The optimal conditions of the design variables and responses are obtained as

x_1	x_2	x_3	y_1	y_2
-0.5218	1.4535	-0.4552	91.8952	57.627

The loss for any value that is out of specification is 67.3014, loss generated from the quality of the prediction is 122.07, and the total loss is 189.3713.

Change **K** and **C** to put much more weight on y_2, for example, let

$$K = \begin{bmatrix} 11.124 & -0.545 \\ -109 & 440.24 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & 0 \\ 0 & 200 \end{bmatrix}$$

The optimal conditions of the design variables and responses are obtained as

x_1	x_2	x_3	y_1	y_2
-0.5484	1.2225	-0.4227	88.9769	57.5929

The loss for any value that is out of specification is 123.2338, loss generated from the quality of the prediction is 161.5993, and the total loss is 284.8331.

Notice that after increasing the weight on y_2, y_2 becomes closer to the target but y_1 is decreased and away from the target. Additionally, from the cost data we can see the non-conformance penalty and the loss generated from the quality of the prediction will be increased after we increase the weight on y_2. Therefore, the experimenter should determine the trade-offs carefully based on overall consideration of the relative weights, variability, and costs.

4.2.3 Production Cost (Tolerance Cost)

Tolerance design suggests that the nominal values of the control variables are deterministic and their limits or variances should be allocated. This is a necessary step for robust design, because the parameter design, as we have presented so far, would obtain the optimum parameters without giving acceptable deviations. Even with consideration of loss-of-quality cost from previous subsection, the tolerances (or variances) would go to zero to achieve minimum quality loss if we do not consider the tolerance cost. This is not possible in real applications. Therefore, the tolerance cost should be added in the objective function for optimization to balance the robust design. In general, tolerance cost is just the cost involving the manufacturing process.

From experience we can conclude that a smaller tolerance value (tighter tolerance) leads to higher cost, due to the need for more manpower and energy to manufacture the product with low variability. Table 4-4 shows various functions proposed to describe the cost-tolerance relationship (Chase et al.

1990). The two constants (a and b) are cost coefficients determined by the particular manufacturing process.

Table 4-4 Proposed Cost-Tolerance Models

Model Name	Cost Model
Linear	$a - b \times Tol$
Reciprocal	$a + b / Tol$
Reciprocal Squared	$a + b / Tol^2$
Reciprocal Power	$a + b / Tol^k$
Exponential	$b \times e^{-m \times Tol}$
Exponential / Reciprocal Power	$b \times e^{-m \times Tol} / Tol^k$
Piecewise Linear	$a_i - b_i \times Tol_i$

4.3 Multiple Response Optimization – Desirability Function Approach

The two-stage sequential robust design optimizes the response over the parameters of control variables with fixed tolerances and then optimizes the response over the tolerances of control variables with optimal parameters found in the previous stage, whereas the one-stage integrated robust design conducts optimization of the response over the parameters and tolerances of control variables simultaneously. The superiority of the joint method is obvious: the optimization will lead to the solution of integrated robust design with higher accuracy and less time and energy.

To simplify the robust design problem, the researcher usually works on robust design problem with single important response or quality characteristic. In many practical instances, however, multiple competing responses should be optimized simultaneously. For example, in some industrial settings, such as semiconductor manufacturing, more than ten response variables are not unusual in a process. Because the responses might be correlated, trade-offs among multiple responses should be studied to obtain overall optimum setting of the parameters that all responses are optimized or at least in desired ranges.

While graphic methods are very popular to determine an initial estimate, the nonlinear programming method is the formal approach to solve the constrained optimization problem. In general, desirability function approach and loss function approach are two of the most important methods to formulate the optimization criteria. We have introduced the loss function approach for both single and multiple response optimization problems, and in the next subsection, the desirability function approach will be discussed.

Derringer and Suich (1980) proposed the desirability function approach to simultaneously optimize multiple responses. This optimization technique weights the responses in accordance with their deviations to the target, upper, and lower specifications. Each response is converted into an individual desirability function d_i that varies between 0 and 1, where $d_i = 0$ if the response is in an unacceptable range, $d_i = 1$ if the response is on target, and $0 < d_i < 1$ otherwise, such that d_i increases as the individual response is moving close to the target. The optimization objective is to maximize the overall desirability

$$D = (d_1 d_2 ... d_m)^{1/m} \tag{4-14}$$

where m is the number of responses. The constraint is the experimental region of the control variables.

In the scenario of "the target is best",

$$d = \begin{cases} 0 & y < L \\ \left(\dfrac{y-L}{T-L}\right)^{r_1} & L \le y \le T \\ \left(\dfrac{U-y}{U-T}\right)^{r_2} & T \le y \le U \\ 0 & y > U \end{cases} \tag{4-15}$$

In the scenario of "the smaller the better",

$$d = \begin{cases} 1 & y \le T \\ \left(\dfrac{U-y}{U-T}\right)^{r} & T < y \le U \\ 0 & y > U \end{cases} \tag{4-16}$$

In the scenario of "the larger the better",

$$d = \begin{cases} 0 & y < L \\ \left(\dfrac{y-L}{T-L}\right)^{r} & L \le y < T \\ 1 & y \ge T \end{cases} \tag{4-17}$$

In the above desirability functions, T is the target, U and L are upper and lower specification, and r, r_1, and r_2 are weights. In general the weights are assumed as 1 to get linear desirability functions. Figure 4-18 shows the two-sided desirability functions with different weights r. We can see the weighting shapes are in the form of tents that peak at the target with value 1 and flattens to zero beyond the upper and lower specifications.

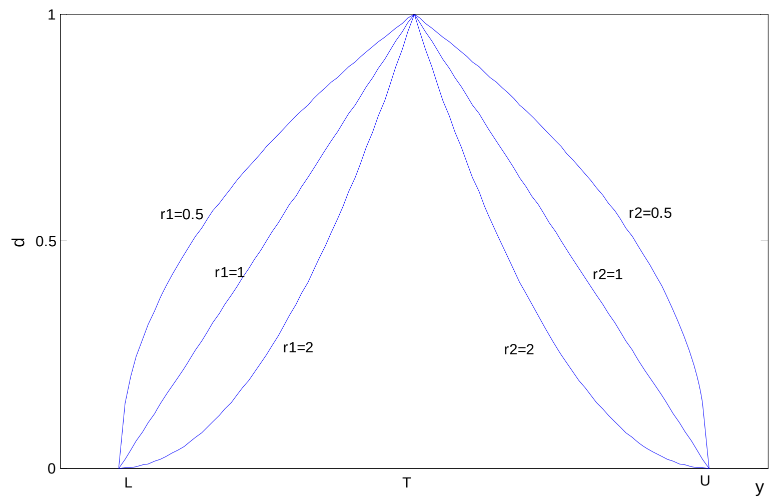

Figure 4-18 Desirability functions (the target is best)

Multiple response optimization will always involve comprise between individual responses. In general, direct search methods are used to maximize the overall desirability. Therefore, we may find multiple optimal values with different starting points. Final selection can be made based on practical and overall considerations.

Example: Desirability Function Approach for Multi-response Optimization

We retake the example from Myers and Montgomery (2002) to optimize the chemical process with two responses: polymer (y_1) and thermal activity (y_2). Recall from Section 4.2.2 the regression models are

$$\hat{y}_1 = 81.09 + 1.03x_1 + 4.04x_2 + 6.20x_3 - 1.83x_1^2 + 2.94x_2^2 - 5.19x_3^2$$
$$+ 2.13x_1x_2 + 11.38x_1x_3 - 3.88x_2x_3$$

$$\hat{y}_2 = 60.51 + 3.58x_1 + 2.23x_3$$

We assumed the weights $r = r_1 = r_2 = 1$. The desirability function of response y_1 (the larger the better) is

$$d_1 = \begin{cases} 0 & y_1 < 80 \\ \dfrac{y_1 - 80}{100 - 80} & 80 \le y_1 < 100 \\ 1 & y_1 \ge 100 \end{cases}$$

The desirability function of response y_2 (the target is best) is

$$d_2 = \begin{cases} 0 & y_2 < 55 \\ \dfrac{y_2 - 55}{57.5 - 55} & 55 \le y_2 \le 57.5 \\ \dfrac{60 - y_2}{60 - 57.5} & 57.5 \le y_2 \le 60 \\ 0 & y_2 > 60 \end{cases}$$

The overall desirability function is

$$D = (d_1 \times d_2)^{1/2}$$

Optimization was formulated to find the optimal conditions of the control variables x_1, x_2, and x_3 as

$$\begin{gathered} Max \quad D = (d_1 \times d_2)^{1/2} \\ subject\ to \\ -1.682 \le x_i \le 1.682,\ i = 1, 2, 3 \end{gathered}$$

We used Matlab to solve the optimization problem. Using different starting points we can obtain different local optima. It should be noted that the multiple responses optimization cannot guarantee to find the global optimum. In general we simply this problem by changing the starting points of the search to find the local optima and making decision through overall consideration.

Considering the overall requirements for this problem, we selected the optimal setting as:

x_1	x_2	x_2	y_1	y_2	d_1	d_2	D
-0.4887	1.682	-0.5653	95.1813	57.5	0.7591	1	0.8712

With this solution, y_1 is close to the upper limit and y_2 is on target, whereas the overall desirability is the highest. One should bear in mind that the trade-offs are always there, thus, we can not improve any criterion without deteriorating a value of at least one other.

Chapter 5
Case Study

5.1 Case Study 1 – Grating Spectroscope

We take an example from Savage and Seshadri (2003) that outlines a design of a mechanism for controlling a grating diffraction spectroscope. In the multi-link mechanism, a wavelength λ is related to the displacement S that is set by a stepping motor subsystem. To calibrate the mechanism, we need two wavelengths λ_1 and λ_2 as the responses in the robust design and they are corresponding to the two step positions S_1 and S_2. In addition to step position (S_1 or S_2), we have other four design variables: off-set distance K_1, grating angle K_2, and machining dimensions L and C. The overall uncertainty of the six design variables deems that the problem be posed as a robust design problem to find parameters and tolerances of K_1 and K_2 to provide sufficient conformance of the responses at best cost.

Table 5-1 shows the mean and tolerance of the design factors. There are five design variables for each response. In the table, the means and tolerances for K_1 and K_2 are initial estimates only. In addition, the minimum tolerance of K_1 and K_2 are $Tol_{1\min} = 0.005mm$ and $Tol_{2\min} = 0.01°$.

Table 5-1 Control and Noise Factors for the Spectroscope

Design variables	Mean	Tolerance
Off-set setting K_1 (mm)	230.6	±0.05
Grating angle K_2 (°)	76	±0.05
Dimension C (mm)	51	±0.05
Dimension L (mm)	211.25	±0.05
Step position S_1 (mm)	34.5325	±0.001
Step position S_2 (mm)	43.4150	±0.001

Screening experiment was done first to eliminate the unimportant variables. Table 5-2 shows the screening experimental design. We used a 2_V^{5-1} fractional factorial design in which no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction. The defining relation is I = ABCDE. Every main effect is aliased with a four-factor interaction

$$l_{K_1} \to K_1 + K_2LCS \,,\; l_{K_2} \to K_2 + K_1LCS \,,\; l_L \to L + K_1K_2CS$$

$$l_C \to C + K_1K_2LS \,,\; l_S \to S + K_1K_2LC$$

where S is either S_1 or S_2 corresponding to response λ_1 or λ_2, respectively.

Table 5-2 2_V^{5-1} Screening Experimental Design

Run	K_1	K_2	L	C	S_1	S_2	λ_1	λ_2
1	-1	-1	-1	-1	1	1	530.0405	1062.7259
2	1	-1	-1	-1	-1	-1	523.9381	1056.5859
3	-1	1	-1	-1	-1	-1	535.1537	1067.5846
4	1	1	-1	-1	1	1	529.2925	1061.6892
5	-1	-1	1	-1	-1	-1	535.733	1068.4727
6	1	-1	1	-1	1	1	529.8697	1062.5724
7	-1	1	1	-1	1	1	541.0837	1073.5683
8	1	1	1	-1	-1	-1	534.9829	1067.4312
9	-1	-1	-1	1	-1	-1	528.8503	1060.4051
10	1	-1	-1	1	1	1	522.9993	1054.5188
11	-1	1	-1	1	1	1	534.2029	1065.5057
12	1	1	-1	1	-1	-1	528.115	1059.3831
13	-1	-1	1	1	1	1	534.7692	1066.3788
14	1	-1	1	1	-1	-1	528.6791	1060.2512
15	-1	1	1	1	-1	-1	539.8812	1071.2353
16	1	1	1	1	1	1	534.0318	1065.3518

Every two-factor interaction is aliased with a three-factor interaction:

$$l_{K_1K_2} \to K_1K_2 + LCS \,,\; l_{K_2C} \to K_2C + K_1LS$$

$$l_{K_1L} \to K_1L + K_2CS\,,\; l_{K_2S} \to K_2S + K_1LC$$

$$l_{K_1C} \to K_1C + K_2LS\,,\; l_{LC} \to LC + K_1K_2S$$

$$l_{K_1S} \to K_1S + K_2LC\,,\; l_{LS} \to LS + K_1K_2C$$

$$l_{K_2L} \to K_2L + K_1CS\,,\; l_{CS} \to CS + K_1K_2L$$

The estimates of the main effects for the response λ_1 are

$$l_{K_1} \to K_1 + K_2LCS = -5.9758$$

$$l_{K_2} \to K_2 + K_1LCS = 5.2331$$

$$l_L \to L + K_1K_2CS = 5.8048$$

$$l_C \to C + K_1K_2LS = -1.0707$$

$$l_S \to S + K_1K_2LC = 0.1195$$

The estimates of the main effects for the response λ_2 are

$$l_{K_1} \to K_1 + K_2LCS = -6.0116$$

$$l_{K_2} \to K_2 + K_1LCS = 4.9798$$

$$l_L \to L + K_1K_2CS = 5.8579$$

$$l_C \to C + K_1K_2LS = -2.2$$

$$l_S \to S + K_1K_2LC = 0.1202$$

From the estimates of the main effects for the two responses, it is reasonable to conclude that K_1, K_2, L, and C are significant. Therefore S_1 and S_2 were deleted from the list of significant design variables and assigned with nominal values 34.5325 mm and 43.4150 mm, respectively. The sums of squares for effects of response λ_1 are

$$SS_{K_1} = \frac{(-47.8061)^\wedge 2}{16} = 142.8391$$

$$SS_{K_2} = \frac{41.8645^\wedge 2}{16} = 109.5395$$

$$SS_L = \frac{46.4381^\wedge 2}{16} = 134.7812$$

$$SS_C = \frac{(-8.5652)^\wedge 2}{16} = 4.5852$$

$$SS_T = \lambda_1'\lambda_1 - \frac{(\sum_{i=1}^{16} \lambda_1)^\wedge 2}{16} = 391.8025$$

$$SS_E = SS_T - SS_{K_1} - SS_{K_2} - SS_L - SS_C = 0.0575$$

The sums of squares for effects of response λ_2 are

$$SS_{K_1} = \frac{(-48.0928)^\wedge 2}{16} = 144.5575$$

$$SS_{K_2} = \frac{39.8385^\wedge 2}{16} = 99.1941$$

$$SS_L = \frac{46.8633^\wedge 2}{16} = 137.2606$$

$$SS_C = \frac{(-17.6004)^\wedge 2}{16} = 19.3608$$

$$SS_T = \lambda_2'\lambda_2 - \frac{(\sum_{i=1}^{16} \lambda_2)^\wedge 2}{16} = 400.4312$$

$$SS_E = SS_T - SS_{K_1} - SS_{K_2} - SS_L - SS_C = 0.0582$$

The analysis of variance (ANOVA) of the screening experiment is summarized in Table 5-3. It is confirmed that K_1, K_2, L, and C are significant.

Table 5-3 Analysis of Variance for Screening Experiment

	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F_0
λ_1	K_1	142.8391	1	142.8391	27469.06
	K_2	109.5395	1	109.5395	21065.29
	L	134.7812	1	134.7812	25919.46
	C	4.5852	1	4.5852	881.7692
	Error	0.0575	11	0.0052	
	Total	391.8025	15		
λ_2	K_1	144.5575	1	144.5575	27275
	K_2	99.1941	1	99.1941	18715.87
	L	137.2606	1	137.2606	25898.23
	C	19.3608	1	19.3608	3652.981
	Error	0.0582	11	0.0053	
	Total	400.4312	15		

In the following robust design, K_1 and K_2, which are adjusted to visualize the wavelengths, are used as control factors. C and L are uncontrollable and uncorrelated because of the different machining processes and are used as noise factors. Table 5-4 shows the data for this experiment. The experiment is a modified face-centered cube (FCC) design, which is a combined array design for a cuboidal region of interest. The FCC design includes three parts:

1. As there are only 4 design factors, we have enough resource to do a single replicate 2^4 two-level factorial design.
2. 2*2=4 axial points on the face of the cube (cuboidal region). Recall that in the modified FCC design the axial points for the noise variables will be removed, because axial points primarily contribute to the estimation of the quadratic terms, and generally in robust design problems we rule out the pure quadratic terms of the noise variables.

3. One center points. Though two center points are preferable, on center point is sufficient enough for stability.

Table 5-4 Experimental Design Data for the Multi-link Mechanism

Run	K_1	K_2	L	C	λ_1	λ_2
1	-1	-1	-1	-1	529.9807	1062.6657
2	1	-1	-1	-1	523.998	1056.6461
3	-1	1	-1	-1	535.2135	1067.6448
4	1	1	-1	-1	529.2327	1061.629
5	-1	-1	1	-1	535.7928	1068.5329
6	1	-1	1	-1	529.8098	1062.5122
7	-1	1	1	-1	541.0239	1073.5081
8	1	1	1	-1	535.0427	1067.4914
9	-1	-1	-1	1	528.91	1060.4652
10	1	-1	-1	1	522.9396	1054.4588
11	-1	1	-1	1	534.1432	1065.4457
12	1	1	-1	1	528.1747	1059.4431
13	-1	-1	1	1	534.7095	1066.3187
14	1	-1	1	1	528.7388	1060.3112
15	-1	1	1	1	539.9409	1071.2953
16	1	1	1	1	533.9721	1065.2918
17	-1	0	0	0	534.9635	1066.9832
18	1	0	0	0	528.9878	1060.9716
19	0	-1	0	0	529.3589	1061.487
20	0	1	0	0	534.592	1066.4668
21	0	0	0	0	531.9756	1063.9773

The second-order RSM prediction models for the two responses are as following:

$$\hat{\lambda}_1 = 531.9752 - 2.9879K_1 + 2.6165K_2 + 0.0007K_1^2 + 0.0005K_2^2 + 0.0005K_1K_2 + 2.9024L - 0.5353C + 0.0031K_1C - 0.0004K_2L$$

$$\hat{\lambda}_2 = 1063.9765 - 3.0058K_1 + 2.4899K_2 + 0.0013K_1^2 + 0.0008K_2^2 + 0.001K_1K_2$$
$$+ 2.929L - 1.1C - 0.0003K_1L + 0.0033K_1C - 0.001K_2L + 0.0004K_2C$$

The sum of squares due to residual and total sum of squares for λ_1 and λ_2 are

$$SS_{E1} = \sum_{i=1}^{n}(y_{1i} - \hat{y}_{1i})^2 = 0.00016,\ SS_{E2} = \sum_{i=1}^{n}(y_{2i} - \hat{y}_{2i})^2 = 0.00019$$

$$SS_{T1} = y_1'y_1 - \frac{(\sum_{i=1}^{n} y_{1i})^2}{n} = 423.29,\ SS_{T2} = y_2'y_2 - \frac{(\sum_{i=1}^{n} y_{2i})^2}{n} = 430.84$$

The coefficients of multiple determination of the models for the two responses are

$$R_1^2 = 1 - \frac{SS_{E1}}{SS_{T1}} = 0.9999996,\ R_2^2 = 1 - \frac{SS_{E2}}{SS_{T2}} = 0.99999956$$

So the models explain the variability of the response very well.

The RSM mean models for the two responses are:

$$\hat{E}(\lambda_1) = 531.9752 - 2.9879K_1 + 2.6165K_2 + 0.0007K_1^2 + 0.0005K_2^2 + 0.0005K_1K_2$$

$$\hat{E}(\lambda_2) = 1063.9765 - 3.0058K_1 + 2.4899K_2 + 0.0013K_1^2 + 0.0008K_2^2 + 0.001K_1K_2$$

We simplified the regression models of the two responses to first-order and included only the control variables:

$$\hat{\lambda}_1 = 531.9762 - 2.9879K_1 + 2.6165K_2$$

$$\hat{\lambda}_2 = 1063.9784 - 3.0058K_1 + 2.4899K_2$$

To find the relationship between the variances of responses and the tolerances of control factors, we used Taylor series expansion

$$\sigma_\lambda^2 = \begin{bmatrix} \frac{\partial\lambda}{\partial K_1} & \frac{\partial\lambda}{\partial K_2} \end{bmatrix} \begin{bmatrix} \sigma_{K_1}^2 & Cov_{K_1K_2} \\ Cov_{K_1K_2} & \sigma_{K_2}^2 \end{bmatrix} \begin{bmatrix} \frac{\partial\lambda}{\partial K_1} \\ \frac{\partial\lambda}{\partial K_2} \end{bmatrix}$$

Assuming $Cov_{K_1K_2} = 0$, $Tol_{K_1} = 3\sigma_{K_1}$, and $Tol_{K_2} = 3\sigma_{K_2}$, then

$$\sigma_\lambda^2 = \frac{1}{9}\begin{bmatrix} \frac{\partial\lambda}{\partial K_1} & \frac{\partial\lambda}{\partial K_2} \end{bmatrix}\begin{bmatrix} Tol_{K_1}^2 & 0 \\ 0 & Tol_{K_2}^2 \end{bmatrix}\begin{bmatrix} \frac{\partial\lambda}{\partial K_1} \\ \frac{\partial\lambda}{\partial K_2} \end{bmatrix}$$

Though the two responses are correlated, to make the problem simple, we assumed the off-diagonal coefficients of the cost matrix for the two responses are zero. Thus the total loss is

$$Q = E[L(y)] = k_1[(\mu_{\lambda_1} - T_{\lambda_1})^2 + \sigma_{\lambda_1}^2] + k_2[(\mu_{\lambda_2} - T_{\lambda_2})^2 + \sigma_{\lambda_2}^2] \tag{5-1}$$

The loss-of-quality cost Q is the cost to recalibrate any nonconforming mechanism, either on-line in the factory or on-site at the customer. We used RSM mean models as μ_{λ_1} and μ_{λ_2}. The variance $\sigma_{\lambda_1}^2$ and $\sigma_{\lambda_2}^2$ can be obtained from the Taylor series expansion. The target values of the responses are $\lambda_1 = 532\ nm$ and $\lambda_2 = 1064\ nm$, and both responses have tolerances of $\pm 1\ nm$. In this case the recalibration is done in the factory and the unit cost is $A_0 = \$500$. Cost coefficient *k* can be evaluated as

$$k_1 = k_2 = \frac{A_0}{\Delta_0^2} = 500$$

The production cost is evaluated as

$$C_P = a_1 + \frac{b_1}{Tol_{K_1}} + a_2 + \frac{b_2}{Tol_{K_2}} + C_{Step} + C_{CL} + C_I$$

where

a_1 and b_1 - production cost coefficients of K_1

a_2 and b_2 - production cost coefficients of K_2

C_{Step} - production cost of the stepping motor subsystem (depends on the designated tolerance)

C_{CL} - cost for machining the two noise factors: dimensions C and L

C_I - cost for inspection of the assembled mechanism

In the analysis, we set $a_1 = \$10$, $a_2 = \$10$, $b_1 = 0.5$ \$-mm, $b_2 = 1.0$ \$-degree, C_{Step} = \$110, C_{CL} = \$340, and C_I = \$50 (we assumed every unit is inspected after assembly by checking wavelengths against step positions using test samples).

The single objective function used in the optimization is the total cost

$$C_T = Q + C_P$$

Thus, the optimization problem can be formulated as the minimization of the total cost subject to constraints of the mean of the responses and the experimental regions of the control factors:

$$\begin{gathered} Min \quad C_T \\ subject\ to \\ \lambda_{1\min} \le \hat{E}(\lambda_1) \le \lambda_{1\max}, \lambda_{2\min} \le \hat{E}(\lambda_2) \le \lambda_{2\max} \\ K_1, K_2 \in R \\ Tol_{K_1} \ge Tol_{K_1\min}, Tol_{K_2} \ge Tol_{K_2\min} \end{gathered}$$

We obtained the optimal conditions as following:

$$K_1 = 0.00064,\ K_2 = 0.01021,\ Tol_{K_1} = 0.008556,\ Tol_{K_2} = 0.01199$$

$$Mean_{\lambda_1} = 531.999999,\ Mean_{\lambda_2} = 1063.999995$$

$$Q = 70.9125,\ C_P = 649.8263,\ C_T = 720.7388$$

Notice the tolerance of K_1 and K_2 are tighter (smaller) than the initial settings (0.05 and 0.05). It is reasonable because comparing with the machining process of C and L, adjusting K_1 and K_2 is less expensive. The nominal values of K_1 and K_2 are close to the initial conditions (middle level).

If we do not inspect the assembly in the factory and recalibration is done on site, we have $C_I = 0$ and the unit cost is $A_0 = 2000$ ($K_1 = K_2 = 2000$). After optimization, the optimal conditions are

$$K_1 = 0.0005985,\ K_2 = 0.01016,\ Tol_{K_1} = 0.00539,\ Tol_{K_2} = 0.01,$$

$$Mean_{\lambda_1} = 531.999998,\ Mean_{\lambda_2} = 1064.0000015$$

$$Q = 162.345,\ C_P = 650.767\ ,\ C_T = 813.112$$

Because the cost coefficients are increased, the tolerance of K_1 and K_2 are decreased to the minimum level to balance the total cost. We notice the loss-of-quality cost (Q) is increased a lot in the second case. It is understandable that the cost for action taken outside of the factory is more expensive than the cost of internal inspection. Our results for the nominal values and tolerances are close to the solutions of Savage and Seshadri (2003) that solved this problem with a probabilistic robust design method.

5.2 Case Study 2 – Elastic Element of a Force Transducer

The following example is from a case study of Romano et al. (2004). The robust design is applied for the design of the elastic element of a force transducer (Figure 5-1). The design of the element is intended to minimize the transducer's inaccuracy, which originates from two major sources: non-linearity and hysteresis, when a compressive load is applied to the elastic element. There is a trade-off between the effects of non-linearity and hysteresis, which can be studied to improve the design of the transducer. The robust design experiment was simulated using a finite element program (Romano et al. 2004).

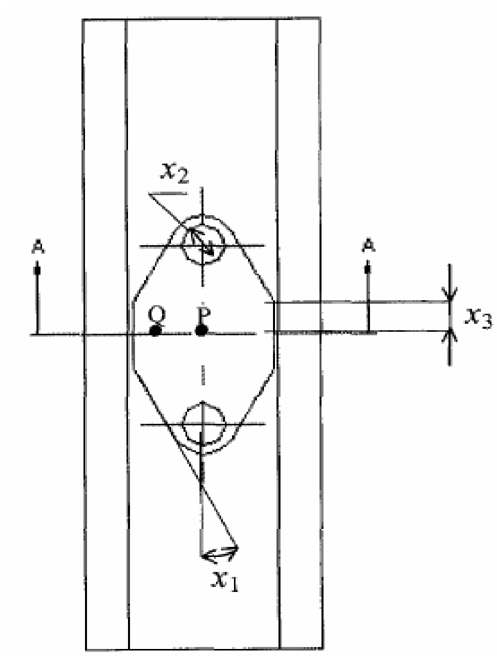

Figure 5-1 Elastic element of the force transducer

The two indicators that quantify the effect of non-linearity and the hysteresis are defined as the responses y_1 and y_2, respectively. The nonlinearity effect is the ratio between longitudinal strain (retrieved at the center P of the measuring area in the y direction) and transversal strain (at a point Q which is 10mm off center in the x direction). The hysteresis indicator is the ratio between maximum strain on the measuring area and longitudinal strain at point P. We seek to achieve a target value of y_1 and minimize y_2. While the acceptable range for y_1 is 0.9 to 1.1, and the target value is the midpoint 1, the acceptable range for y_2 is 1 to 3, and the target value is the lower limit 1. These limits meet standard specifications for the force transducer.

Table 5-5 shows the natural and coded levels of the control and noise factors. In this problem, control factors are the three parameters defining the element configuration with three levels: lozenge angle (x_1), bore diameter (x_2), and half-length of the vertical segment (x_3). Though we can assign more levels for x_1 and x_2 to represent their variability, the corresponding standard deviations of the control factors are used as separated noise factors for the robust design. The noise factors in this design include two internal noise factors only: the standard deviation of the lozenge angle from its nominal value (z_1) and the standard deviation of the bore diameter from its nominal value (z_2). These internal noise factors are determined by the corresponding machining processes and so independent. They are assumed to be normally distributed random variables with zero mean and variance σ_1^2 and σ_2^2.

Table 5-5 Levels of Control and Noise Factors

Factors	Levels		
	-1	0	1
$x_1(\circ)$	15	30	45
$x_2(mm)$	8	11	14
$x_3(mm)$	7	9	11
$z_1(\circ)$	-1.5	0	1.5
$z_2(mm)$	-0.25	0	0.25

Table 5-6 shows the data from this experiment. (The last two columns present the estimated responses $\hat{y}_1$ and $\hat{y}_2$ obtained from the regression model). The experiment is a modified face-centered cube (FCC) design, which is a combined array design for a cuboidal region of interest. The FCC design includes three parts:

Table 5-6 Multiple-response Experimental Design for the Force Transducer

Run	x_1	x_2	x_3	z_1	$z_2 = x_1 x_2 x_3 z_1$	y_1	y_2	$\hat{y}_1$	$\hat{y}_2$
1	-1	-1	-1	-1	1	1.81	1.1	1.81	1.23
2	-1	-1	-1	1	-1	1.69	1.11	1.70	1.07
3	-1	-1	1	-1	-1	1.9	1.07	1.92	1.16
4	-1	-1	1	1	1	1.78	1.07	1.76	1
5	-1	1	-1	-1	-1	1.8	1.47	1.79	1.33
6	-1	1	-1	1	1	1.63	1.18	1.63	1.17
7	-1	1	1	-1	1	1.92	1.41	1.91	1.6
8	-1	1	1	1	-1	1.78	1.58	1.79	1.44
9	1	-1	-1	-1	-1	1.36	1.57	1.34	1.77
10	1	-1	-1	1	1	1.22	2.03	1.22	1.93
11	1	-1	1	-1	1	1.48	1.38	1.49	1.46
12	1	-1	1	1	-1	1.44	1.68	1.42	1.62
13	1	1	-1	-1	1	0.693	3.37	0.69	3.42
14	1	1	-1	1	-1	0.616	3.75	0.62	3.57
15	1	1	1	-1	-1	0.95	2.81	0.94	2.77
16	1	1	1	1	1	0.817	2.83	0.82	2.93
17	-1	0	0	0	0	1.79	1.24	1.79	1.25
18	1	0	0	0	0	1.03	2.46	1.07	2.43
19	0	-1	0	0	0	1.53	1.23	1.54	1.2
20	0	1	0	0	0	1.22	1.73	1.23	2.08
21	0	0	-1	0	0	1.3	1.63	1.3	1.74
22	0	0	1	0	0	1.44	1.67	1.46	1.55
23	0	0	0	0	0	1.38	1.73	1.38	1.64
24	0	0	0	0	0	1.39	1.74	1.38	1.64
25	0	0	0	0	0	1.4	1.74	1.38	1.64

1. A 2_V^{5-1} two-level fractional factorial design with defining relation $I = x_1 x_2 x_3 z_1 z_2$ and setting of the fifth factor is $z_2 = x_1 x_2 x_3 z_1$. In the resolution V design there is no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction.
2. $2 \times (5-2) = 6$ axial points on the face of the cube (cuboidal region). Recall that in the modified FCC design the axial points for the noise variables are eliminated, because axial points primarily contribute to the estimation of the quadratic terms, and generally in robust design problems we do not consider the pure quadratic terms of the noise variables.
3. Three center points. Since the finite element simulation model considers random noise effects, repeated runs for the center points produce different results.

To compare the results from different methods, first we used the same models as from Romano et al. (2004), which include first-order and second-order effects of control factors, two-factor interactions and three-factor interactions among the control factors, linear effects of the noise factors, and control-by-noise interactions. The RSM prediction models for the two responses are as following:

$$\begin{aligned}\hat{y}_1 &= 1.38 - 0.361x_1 - 0.155x_2 + 0.0771x_3 + 0.0481x_1^2 - 0.148x_1x_2 + 0.0218x_1x_3 \\ &\quad + 0.0130x_2x_3 - 0.0588z_1 - 0.0116z_2 + 0.01x_1z_1\end{aligned}$$

$$\begin{aligned}\hat{y}_2 &= 1.64 + 0.592x_1 + 0.438x_2 - 0.095x_3 + 0.0201x_1^2 + 0.301x_1x_2 - 0.143x_1x_3 \\ &\quad - 0.0844x_1x_2x_3 + 0.0794x_1z_1\end{aligned}$$

The sum of squares due to residual and the total sum of squares are

$$SS_{E1} = \sum_{i=1}^{n}(y_{1i} - \hat{y}_{1i})^2 = 0.0046, \ SS_{E2} = \sum_{i=1}^{n}(y_{2i} - \hat{y}_{2i})^2 = 0.3866$$

$$SS_{T1} = y_1'y_1 - \frac{(\sum_{i=1}^{n} y_{1i})^2}{n} = 3.3182, \ SS_{T2} = y_2'y_2 - \frac{(\sum_{i=1}^{n} y_{2i})^2}{n} = 12.5005$$

The coefficients of multiple determination are

$$R_1^2 = 1 - \frac{SS_{E1}}{SS_{T1}} = 0.9986, \ R_2^2 = 1 - \frac{SS_{E2}}{SS_{T2}} = 0.9691$$

that show the models explain the variability of the response well. A check of the normality assumption is made by constructing normal probability plots of the residuals. Figure 5-2 and Figure 5-3 reveal no apparent problem with normality.

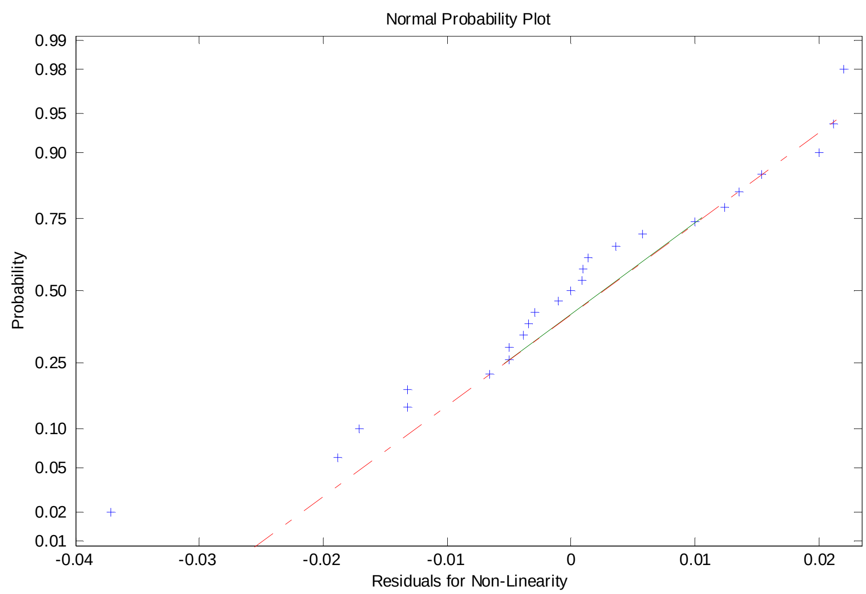

Figure 5-2 Normal probability plot of residuals for non-linearity (y_1)

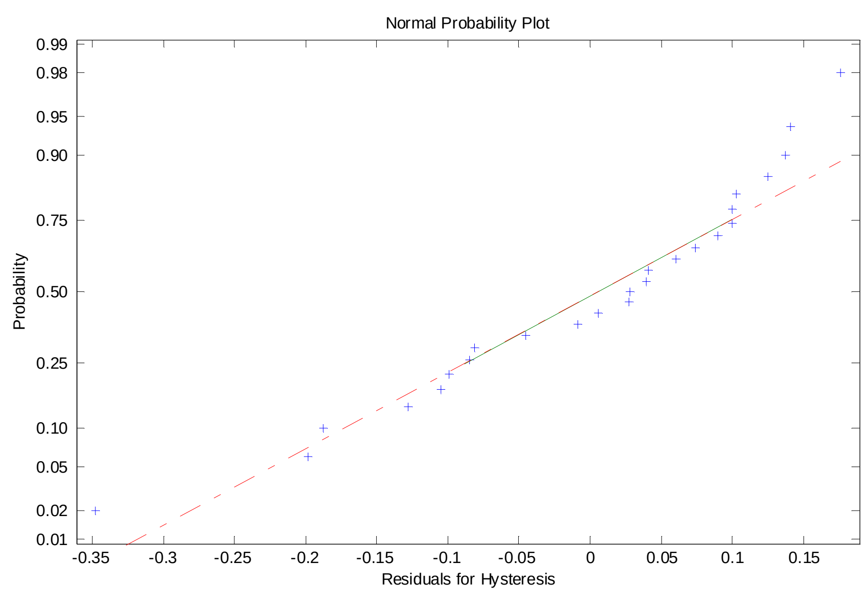

Figure 5-3 Normal probability plot of residuals for hysteresis (y_2)

The estimated mean models are:

$$\hat{E}(y_1) = 1.38 - 0.361x_1 - 0.155x_2 + 0.0771x_3 + 0.0481x_1^2 - 0.148x_1x_2 + 0.0218x_1x_3 + 0.0130x_2x_3$$

$$\hat{E}(y_2) = 1.64 + 0.592x_1 + 0.438x_2 - 0.095x_3 + 0.0201x_1^2 + 0.301x_1x_2 - 0.143x_1x_3 - 0.0844x_1x_2x_3$$

And the estimated variance models are:

$$\hat{Var}(y_1) = (-0.0588 + 0.01x_1)^2\sigma_{z_1} + (-0.0116)^2\sigma_{z_2} + \sigma_{y_1}^2$$

$$\hat{Var}(y_2) = (0.0794x_1)^2\sigma_{z_1} + \sigma_{y_2}^2$$

As $\hat{\sigma}_{y_1}^2 = \dfrac{SS_{E1}}{n-p} = \dfrac{0.0046}{25-11} = 0.00033$, $\hat{\sigma}_{y_2}^2 = \dfrac{SS_{E2}}{n-p} = \dfrac{0.3866}{25-9} = 0.0242$, and we assumed $\sigma_{z_1} = \sigma_{z_2} = 1$, the variance models can be given as

$$\hat{Var}(y_1) = (-0.0588 + 0.01x_1)^2 + (-0.0116)^2 + 0.00033$$

$$\hat{Var}(y_2) = (0.0794x_1)^2 + 0.0242$$

In the following subsections, we present six methods to optimize this robust design problem: integrated parameter and tolerance design, integrated parameter and tolerance design with GLM approach, integrated parameter and tolerance design with Taylor's series expansion, parameter design through desirability function approach, parameter design through dual response method, and parameter design through mean square error (MSE) loss function.

5.2.1 Integrated Robust Design with Cost Consideration

The loss-of-quality cost is obtained through multivariate quality loss function:

$$Q = [\hat{\mathbf{y}}(\mathbf{x},\mathbf{z}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x},\mathbf{z}) - \mathbf{T}] + Trace[\mathbf{C}\textstyle\sum_{\hat{\mathbf{y}}(\mathbf{x},\mathbf{z})}]$$

where targets of the two responses $T_1 = 1$ and $T_2 = 1$. Romano et al. (2004) give the cost matrix **C** as

$$\mathbf{C} = \begin{bmatrix} 30000 & 1500 \\ 1500 & 75 \end{bmatrix}$$

They claim the cost coefficients on the diagonal are obtained by assigning a penalty of 200% of the unit price ($150) to the quality loss and the off-diagonal elements are selected arbitrarily to quantify the compensation or increase of the loss due to deviations from target associate with each pair of the estimated responses. The quality loss function includes two parts: the first part is cost due to non-conformance, and the second part is the penalty imposed by the quality of prediction.

The variance-covariance matrix for the responses is obtained through maximum likelihood estimation based on the experimental data as

$$\Sigma = \left[\hat{\sigma}_{ij}^2\right] = \begin{bmatrix} 0.00018217 & 0.000026678 \\ 0.000026678 & 0.0155 \end{bmatrix}$$

In the case of multiple responses, variance-covariance for the predicted responses is

$$\Sigma_{\hat{\mathbf{y}}(\mathbf{x},\mathbf{z})} = (\mathbf{x},\mathbf{z})^{(m)'}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x},\mathbf{z})^{(m)}\,\Sigma$$

So the estimated expected loss function can be given as

$$\hat{E}(L) = [\hat{\mathbf{y}}(\mathbf{x},\mathbf{z}) - \mathbf{T}]'\mathbf{C}[\hat{\mathbf{y}}(\mathbf{x},\mathbf{z}) - \mathbf{T}] + Trace[(\mathbf{x},\mathbf{z})^{(m)'}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x},\mathbf{z})^{(m)}\mathbf{K}]$$

where

$$\mathbf{K} = \mathbf{C}\Sigma = \begin{bmatrix} 5.0497 & 23.9836 \\ 0.2755 & 1.3575 \end{bmatrix}.$$

Notice we used the unbiased estimated responses in the loss function, so this cost can be expressed in terms of the control factors and internal noise factors (tolerances), which will also be used in the equation of production cost.

The production cost is a function of the two internal noises σ_1 and σ_2 (standard deviations of the two control factors: lozenge angle and bore diameter):

$$C_P = \frac{0.15}{\sigma_1 - 0.003} + \frac{0.1}{\sigma_2 - 0.007}$$

The production cost determines the costs of the two main machining processes (milling and boring). From practical considerations for the machining ability, there are two minimum tolerance constraints: $\sigma_1 > 0.003$ and $\sigma_2 > 0.007$.

The single objective function used herein is the total cost $C_T = Q + C_P$, the sum of the production cost and loss-of-quality cost. The constrained optimization can be performed by minimizing the total cost function subject to constraints on the process means and variances, which are obtained previously from dual response models, and the specifications of the control and noise variables:

$$\begin{gathered} Min \quad C_T \\ subject\ to \\ 0 \le \hat{Var}(y_i) \le V_{j,\max} \\ y_{i,\min} \le \hat{E}(y_i) \le y_{j,\max},\ j = 1, 2 \\ \sigma_{i,\min} < \sigma_i \le 1,\ i = 1, 2 \\ x \in D \end{gathered}$$

We can also formulate the optimization by combining the mean and variance constraints into a single constraint:

$$\begin{gathered} Min \quad C_T \\ subject\ to \\ \min\{(y_{j,\max} - \hat{E}[y_j(x,z)]), (\hat{E}[y_j(x,z)] - y_{j,\min})\} - 3\sqrt{\hat{Var}[y_i(x,z)]} \ge 0, \qquad j = 1, 2 \\ \sigma_{i,\min} < \sigma_i \le 1,\ i = 1, 2 \\ x \in D \end{gathered}$$

Romano et al. (2004) presented the solution (in coded levels) of the integrated robust design as

$$x_1 = 0.473,\ x_2 = 0.801,\ x_3 = -1,\ \sigma_1 = 0.0074,\ \sigma_2 = 0.057,$$

$$Var_{y_1} = 0.0034,\ Var_{y_2} = 0.0256,\ Mean_{y_1} = 0.942,\ Mean_{y_1} = 2.5778,$$

$$MSE_{y_1} = 0.0068,\ MSE_{y_2} = 2.515.$$

The total cost is \$22.40 that includes two portions: the production cost $C_P = \$4.1$, and loss of quality cost Q = \$18.3, which is made up of the nonconformance cost (\$16.6) and the penalty for the quality of prediction (\$1.7).

Cost matrix is very important to the solution of the problem. If we arbitrarily change **C** to $\begin{bmatrix} 27500 & 1500 \\ 1500 & 85 \end{bmatrix}$, different solutions can be obtained as following:

$$x_1 = 0.5673,\ x_2 = 0.6701,\ x_3 = -0.6468, \sigma_1 = 0.9276,\ \sigma_2 = 0.4922,$$

$$Var_{y_1} = 0.0033,\ Var_{y_2} = 0.0262,$$

$$Mean_{y_1} = 0.9670,\ Mean_{y_2} = 2.5832.$$

The production cost is $C_P = 0.3683$, the nonconformance cost is 8.4144 and the penalty for lack-of-fit is 2.0426. We find this solution relaxes the tight tolerances as the previous one, and the means and variances are improved as while. At the same time, both production cost and loss-of-quality cost are reduced.

As discussed before, the economic information, such as the cost matrix, is very important for the tolerance design but difficult to obtain. Practically, we should bear in mind that if we consider both diagonals and off-diagonals of the cost matrix in the multiple responses design, the general rules for the single response tolerance design may be not applicable. For example, in single response tolerance design, tighter tolerances of the design variables result in lower loss-of-quality cost and higher production cost. In multiple responses design, however, the conclusion should be drawn carefully. When the tolerances change, the parameters and the competing responses will change also. As a result, while production cost still follows the same rule, loss-of-quality quality might change differently for each individual situation. In this example, we can have another optimal solution if we restrict the maximum value of σ_1 and σ_2 as 0.5:

$$x_1 = 0.6033,\ x_2 = 0.6992,\ x_3 = -0.6811,\ \sigma_1 = 0.5,\ \sigma_2 = 0.4783,$$

$$Var_{y_1} = 0.0032,\ Var_{y_2} = 0.0265,$$

$$Mean_{y_1} = 0.9412,\ Mean_{y_2} = 2.6513.$$

The production cost is $C_P = 0.514$, the nonconformance cost is 8.9422 and the penalty for the quality of prediction is 1.5633. As the standard deviations (tolerances) are decreased, the production cost is increased, but loss-of-quality cost is increased also, though not very much.

5.2.2 Integrated Robust Design with GLM Approach

Figure 5-4 and Figure 5-5 present the plots of residuals versus the predicted response $\hat{y}_1$ and $\hat{y}_2$. Both of them exhibit funnel-shaped patterns that imply possible inequality of variance. So we used RSM with generalized linear model (GLM) approach to model the responses.

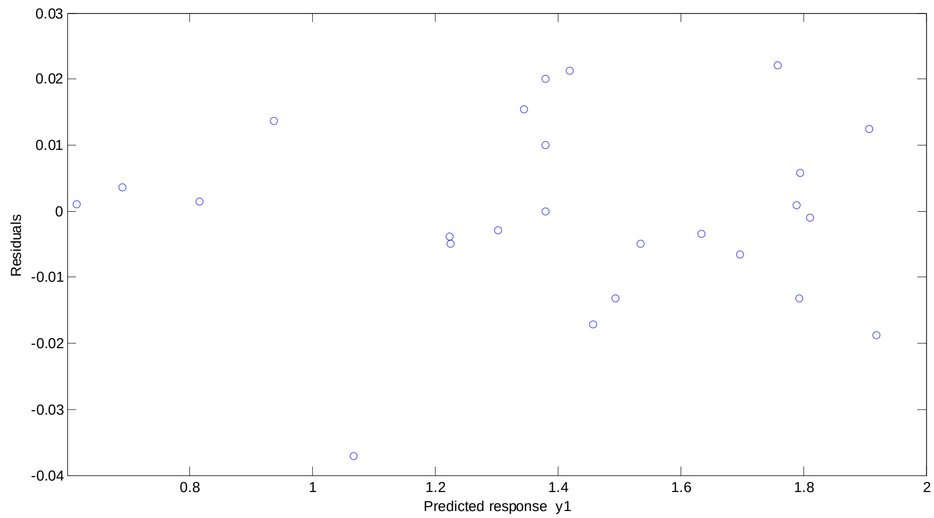

Figure 5-4 Plot of residuals versus predicted response $\hat{y}_1$

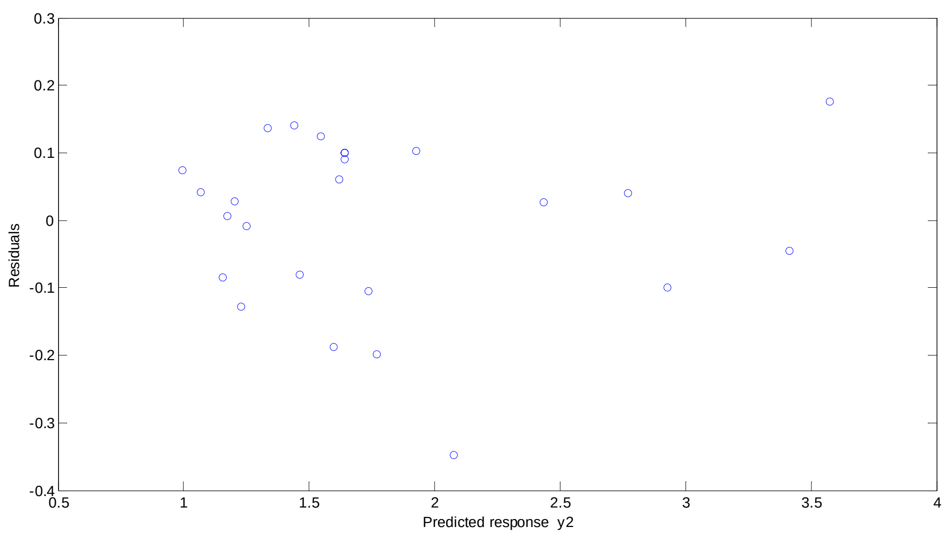

Figure 5-5 Plot of residuals versus predicted response $\hat{y}_2$

The RSM prediction models for the two responses are as following:

$$\hat{y}_1 = 1.3819 - 0.3584x_1 - 0.1564x_2 + 0.0791x_3 + 0.0501x_1^2 - 0.1502x_1x_2 + 0.0229x_1x_3 + 0.0113x_2x_3 - 0.0587z_1 - 0.0116z_2 + 0.01x_1z_1 - 0.0079x_1z_2 - 0.0062x_2z_1$$

$$\hat{y}_2 = 1.7139 + 0.5743x_1 + 0.4497x_2 - 0.1065x_3 + 0.1699x_1^2 - 0.0934x_2^2 + 0.0277x_3^2 + 0.2633x_1x_2 - 0.0991x_1x_3 - 0.053x_2x_3 + 0.0656z_1 - 0.0419z_2 + 0.0794x_1z_1 + 0.0169x_1z_2 - 0.0306x_2z_1 - 0.0606x_2z_2 - 0.0144x_3z_2$$

The models for the two non-constant residual variances are

$$\hat{\sigma}_{y_1}^2 = \exp(-10.4299 + 3.4309x_1 - 0.3587x_2 + 1.4379x_3)$$

$$\hat{\sigma}_{y_2}^2 = \exp(-4.8033 - 0.6144x_1 + 0.755x_2 - 0.7225x_3)$$

The estimated mean models are:

$$\hat{E}_{y_1} = 1.3819 - 0.3584x_1 - 0.1564x_2 + 0.0791x_3 + 0.0501x_1^2 - 0.1502x_1x_2 + 0.0229x_1x_3 + 0.0113x_2x_3$$

$$\hat{E}_{y_2} = 1.7139 + 0.5743x_1 + 0.4497x_2 - 0.1065x_3 + 0.1699x_1^2 - 0.0934x_2^2 + 0.0277x_3^2 + 0.2633x_1x_2 - 0.0991x_1x_3 - 0.053x_2x_3$$

And the estimated variance models are:

$$\hat{Var}_{y_1} = (-0.0587 + 0.01x_1 - 0.0062x_2)\hat{}2 + (-0.0116 - 0.0079x_1)\hat{}2 + \exp(-10.4299 + 3.4309x_1 - 0.3587x_2 + 1.4379x_3)$$

$$\hat{Var}_{y_2} = (0.0656 + 0.0794x_1 - 0.0306x_2)\hat{}2 + (-0.0419 + 0.0169x_1 - 0.0606x_2 - 0.0144x_3)\hat{}2 + \exp(-4.8033 - 0.6144x_1 + 0.755x_2 - 0.7225x_3)$$

We assumed $C = \begin{bmatrix} 27500 & 1500 \\ 1500 & 85 \end{bmatrix}$ and used the same optimization method as before to obtain the optimal operating conditions as following:

$$x_1 = 0.3793,\ x_2 = 1,\ x_3 = -0.4486,\ \sigma_1 = 1,\ \sigma_2 = 1,$$

$$Var_{y_1} = 0.004,\ Var_{y_2} = 0.0314,$$

$$Mean_{y_1} = 0.9953,\ Mean_{y_2} = 2.5064.$$

The production cost is $C_p = 0.2512$, the nonconformance cost is 6.9926 and the penalty for quality of prediction is 0.1451. We notice this solution has loosened the tolerances that result in lower production cost (easier for the machining operations), and the mean for $\hat{y}_1$ is very close to the goal. At the same time, because we improve the accuracy of the estimate for the residual variances and response surface models through GLM approach, the penalty for quality of prediction is reduced vastly.

5.2.3 Integrated Robust Design with Taylor's Series Expansion

Taylor's series expansion can be used as an alternative for the dual RSM approach by approximating the process mean and variance in terms of control variables and their standard deviations (or tolerances). The two RSM prediction models that include only the three control variables are

$$\begin{aligned}\hat{y}_1 = {} & 1.3773 - 0.3608x_1 - 0.1547x_2 + 0.0771x_3 + 0.0423x_1^2 + 0.0073x_2^2 + 0.0023x_3^2 \\ & - 0.1484x_1x_2 + 0.0217x_1x_3 + 0.013x_2x_3\end{aligned}$$

$$\begin{aligned}\hat{y}_2 = {} & 1.6604 + 0.5917x_1 + 0.4383x_2 - 0.095x_3 + 0.2469x_1^2 - 0.1231x_2^2 + 0.0469x_3^2 \\ & + 0.3006x_1x_2 - 0.1431x_1x_3 - 0.0331x_2x_3\end{aligned}$$

Using Taylor approximation, we can obtain process mean as

$$\begin{aligned}\mu_{y_1} = E(\hat{y}_1) & \approx g[E(\mathbf{x})] + \frac{1}{2}\sum_{i-1}^{3}\sum_{j=1}^{3}\left.\frac{\partial^2 g(\mathbf{x})}{\partial x_i x_j}\right|_{E(\mathbf{x})} Cov(x_i x_j) \\ & = g[E(\mathbf{x})] + \frac{1}{2}(0.0423 * 2 * \sigma_1^2 + 0.0073 * 2 * \sigma_2^2 + 0.0023 * 2 * \sigma_3^2) \\ & = g[E(\mathbf{x})] + 0.0423\sigma_1^2 + 0.0073\sigma_2^2 + 0.0023\sigma_3^2\end{aligned}$$

$$\begin{aligned}\mu_{y_2} = E(\hat{y}_2) & \approx g[E(\mathbf{x})] + \frac{1}{2}\sum_{i-1}^{3}\sum_{j=1}^{3}\left.\frac{\partial^2 g(\mathbf{x})}{\partial x_i x_j}\right|_{E(\mathbf{x})} Cov(x_i x_j) \\ & = g[E(\mathbf{x})] + \frac{1}{2}(0.2469 * 2 * \sigma_1^2 - 0.1231 * 2 * \sigma_2^2 + 0.0469 * 2 * \sigma_3^2) \\ & = g[E(\mathbf{x})] + 0.2469\sigma_1^2 - 0.1231\sigma_2^2 + 0.0469\sigma_3^2\end{aligned}$$

and process variance as

$$\sigma_{y_1}^2 = Var(\hat{y}_1) \approx \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\partial g(\mathbf{x})}{\partial x_i}\bigg|_{E(\mathbf{x})} \frac{\partial g(\mathbf{x})}{\partial x_j}\bigg|_{E(\mathbf{x})} Cov(x_i x_j)$$

$$= (-0.3608 + 0.0423 * 2x_1 - 0.1484x_2 + 0.0217x_3)\sigma_1^2 + (-0.1547 + 0.0073 * 2x_2 - 0.1484x_1 + 0.013x_3)\sigma_2^2 + (0.0771 + 0.0023 * 2x_3 + 0.0217x_1 + 0.013x_2)\sigma_3^2$$

$$\sigma_{y_2}^2 = Var(\hat{y}_2) \approx \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\partial g(\mathbf{x})}{\partial x_i}\bigg|_{E(\mathbf{x})} \frac{\partial g(\mathbf{x})}{\partial x_j}\bigg|_{E(\mathbf{x})} Cov(x_i x_j)$$

$$= (0.5917 + 0.2469 * 2x_1 + 0.3006x_2 - 0.1431x_3)\sigma_1^2 + (0.4383 - 0.1231 * 2x_2 + 0.3006x_1 - 0.0331x_3)\sigma_2^2 + (-0.095 + 0.0469 * 2x_3 - 0.1431x_1 - 0.0331x_2)\sigma_3^2$$

We assume the off-diagonal coefficients of the cost matrix for the two responses are zero, so the loss-of-quality cost is

$$Q = k_1[(\mu_{y_1} - T_1)^2 + \sigma_{y_1}^2] + k_2[(\mu_{y_2} - T_2)^2 + \sigma_{y_2}^2]$$

and we give a new production cost function that adds the standard deviation of x_3 as

$$C_P = \frac{0.15}{\sigma_1 - 0.003} + \frac{0.1}{\sigma_2 - 0.007} + \frac{0.1}{\sigma_3 - 0.01}$$

If we assume the cost coefficients $k_1 = 30000$ and $k_2 = 75$, through constrained optimization we can find the optimal operating conditions as

$$x_1 = 0.3602,\ x_2 = 1,\ x_3 = -0.4511, \sigma_1 = 0.0238,\ \sigma_2 = 0.0397,\ \sigma_3 = 0.0625$$

$$Var_{y_1} = 0.000233,\ Var_{y_2} = 0.0011,$$

$$\hat{y}_1 = 1.0083,\ \hat{y}_2 = 2.4196$$

The production cost is $C_P = 12.1895$, and the loss-of-cost is 160.3298. Using Taylor's series expansion, even without simulating the internal noises in the experimental designs, we can obtain the model of process mean and variance, and then achieve the parameters and tolerances of the control variables through the integrated robust design. Therefore, Taylor approximation is an effective alternative for the dual RSM approach in robust design.

5.2.4 Parameter Design through Desirability Function Approach

Parameter designs do not consider the economic constraints, so they are usually performed separately or as a first step for robust design. Robust parameter design (RPD) is the most important part of robust design. In the following subsections, we present three different parameter design methods to solve the same problem of a force transducer. Matlab is used to solve the multiobjective optimization problem. Because the multiple responses are competing, these parameter designs may not give a global optimal solution and tradeoff is always involved, which means an improvement in one response requires a degradation of another. The set of solutions can be called noninferior or Pareto optimal solutions. The final decision should be made by the experienced process owner based on practical conditions.

In this subsection, desirability functions are used to optimize the two responses simultaneously. We selected the weights $r = r_1 = r_2 = 1$. The desirability function of response y_1 (the target is best) is

$$d_1 = \begin{cases} 0 & y_1 < 0.9 \\ \left(\dfrac{y_1 - 0.9}{1 - 0.9}\right)^1 & 0.9 \le y_1 \le 1 \\ \left(\dfrac{1.1 - y_1}{1.1 - 1}\right)^1 & 1 \le y_1 \le 1.1 \\ 0 & y_1 > 1.1 \end{cases}$$

The desirability function of response y_2 (the smaller the better) is

$$d_2 = \begin{cases} 1 & y_2 \le 1 \\ \left(\dfrac{3 - y_2}{3 - 1}\right)^1 & 1 < y_2 \le 3 \\ 0 & y_2 > 3 \end{cases}$$

The overall desirability function is

$$D = (d_1 \times d_2)^{1/2}$$

Formulated optimization to find the optimal setting of the control variables x_1, x_2, and x_3

$$\begin{gathered} Max \quad D = (d_1 \times d_2)^{1/2} \\ subject\ to \\ -1 \le x_i \le 1, i = 1, 2, 3 \end{gathered}$$

We used Matlab to solve the optimization problem. Using different starting points we can obtain different local optima. It should be noted that the multiple responses optimization cannot guarantee to find the global optimum. In general we simply solve this problem by changing the starting points of the search to find the local optima and making decision later through overall consideration.

Considering the overall requirements for this problem, we selected the optimal setting as:

$$x_1 = 0.4150\,,\ x_2 = 0.664\,,\ x_3 = -1\,,$$

$$Mean_{y_1} = 1.0151\,,\ Var_{y_1} = 0.0034\,,\ MSE_{y_1} = 0.0037\,,$$

$$Mean_{y_2} = 2.4716\,,\ Var_{y_2} = 0.0253\,,\ MSE_{y_2} = 2.191$$

$$d_1 = 0.8494\,,\ d_2 = 0.2642\,,\ D = 0.4737$$

With this solution, y_1 is very close to target and y_2 has the smallest value, while the overall desirability is big among others. One should bear in mind that the trade-offs are always there, thus, we can not improve any criterion without deteriorating a value of at least one other.

5.2.5 Parameter Design through Dual RSM Approach

With dual response surface approach, the variances of the two responses should be optimized simultaneously, so a suitable multiobjective optimization method is very important. The relative importance of the multiple objectives is generally not known until the system's best capabilities are determined and tradeoffs between the objectives are fully understood. The design for multiple-response optimization should express the problem correctly and solve the problem effectively. In order to simplify the nonlinear multiobjective problem, we selected the weighted sum method that converts multiple objective functions into a single one by summing the weighted objectives.

As in the first method, the mean and variance models are

$$\begin{aligned}\hat{E}(y_1) = {} & 1.38 - 0.361x_1 - 0.155x_2 + 0.0771x_3 + 0.0481x_1^2 - 0.148x_1x_2 + 0.0218x_1x_3 \\ & + 0.0130x_2x_3\end{aligned}$$

$$\begin{aligned}\hat{E}(y_2) = {} & 1.64 + 0.592x_1 + 0.438x_2 - 0.095x_3 + 0.0201x_1^2 + 0.301x_1x_2 - 0.143x_1x_3 \\ & - 0.0844x_1x_2x_3\end{aligned}$$

$$\hat{Var}(y_1) = (-0.0588 + 0.01x_1)^2 + (-0.0116)^2 + 0.00033$$

$$\hat{Var}(y_2) = (0.0794x_1)^2 + 0.0242$$

The MSE loss functions are

$$MSE_1 = E_z(y_1(x,z)-1)^2 = (\hat{E}(y_1)-1)^2 + \hat{Var}(y_1)$$

$$MSE_2 = E_z(y_2(x,z)-1)^2 = (\hat{E}(y_2)-1)^2 + \hat{Var}(y_2)$$

$\hat{Var}(y_1)$ and $\hat{Var}(y_2)$ are the functions of x_1 only. Figure 5-6 shows the contour plot of the variance models $\hat{Var}(y_1)$ and $\hat{Var}(y_2)$, and the combined $w_1\hat{Var}(y_1)+w_2\hat{Var}(y_2)$ (in this case because y_1 has a larger weight than y_2, we assumed $w_1 = 50$ and $w_2 = 1$). From the contour plot it is clear that $x_1 = 1$ is the solution to minimize $w_1\hat{Var}(y_1)+w_2\hat{Var}(y_2)$.

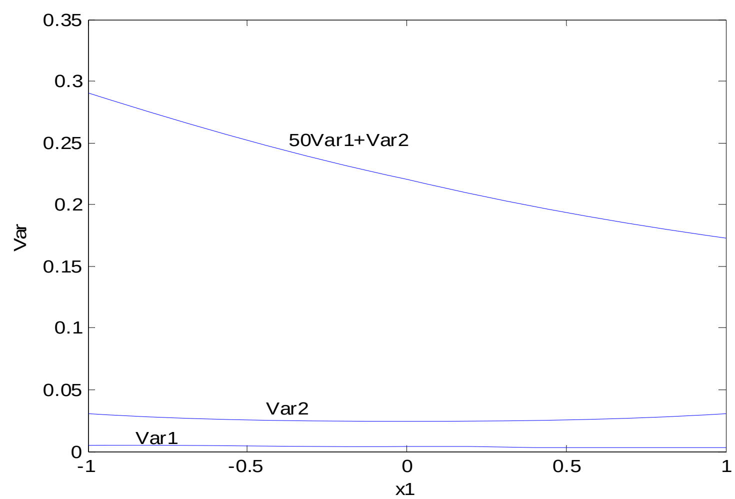

Figure 5-6 Contour plot of the variance models

The formal constrained optimization is

$$\begin{gathered} Min \quad w_1\hat{Var}(y_1)+w_2\hat{Var}(y_2) \\ subject\ to \\ 0.9 \le \hat{E}(y_1) \le 1.1, 1 \le \hat{E}(y_2) \le 3 \\ -1 \le x_i \le 1, i = 1, 2, 3 \end{gathered}$$

where w_1 and w_2 are weights of the two objectives (variances) to measure the tradeoffs between them. In our case, non-linearity obviously has larger weight than hysteresis, so we arbitrarily chose w_1 = 50 and $w_2 = 1$. As $w_1 \hat{Var}(y_1) + w_2 \hat{Var}(y_2)$ is the function of x_1 only, we fix the optimal value of x_1 as $x_1 = 1$ and present the response surface plots in Figure 5-7 for $\hat{E}(y_1)$ and $\hat{E}(y_2)$ versus x_2 and x_3.

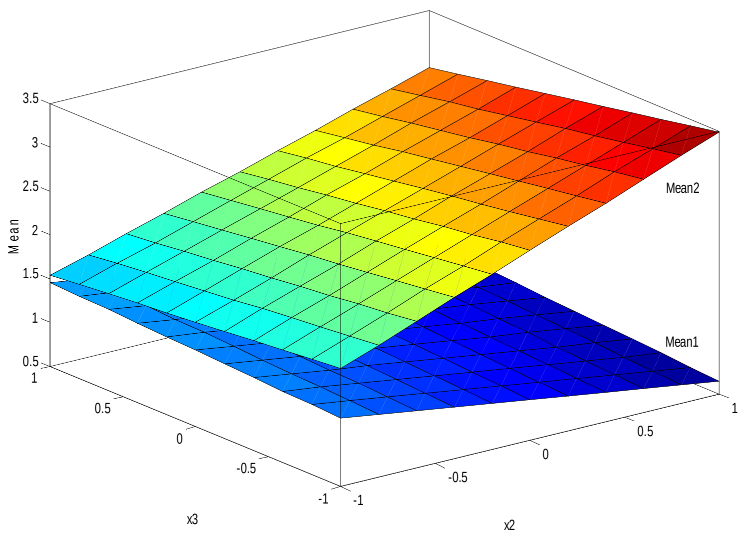

Figure 5-7 Response surface plot for $\hat{E}(y_1)$ and $\hat{E}(y_2)$ versus x_2 and x_3 ($x_1 = 1$)

Matlab was used to solve the optimization problem. The optimal operating conditions are

$$x_1 = 1,\ x_2 = -0.2815,\ x_3 = -0.9963,$$

$$Var_{y_1} = 0.0028,\ Var_{y_2} = 0.0305,\ Mean_{y_1} = 1.0575,\ Mean_{y_2} = 2.4384,$$

$$MSE_{y_1} = 0.0061,\ MSE_{y_2} = 2.0996.$$

We chose this optimal solution because $\hat{E}(y_1)$ is very close to the target, and $\hat{E}(y_2)$ is small among the solutions. The values of variances and MSEs are close to that of the desirability function approach.

5.2.6 Parameter Design through Mean Square Error (MSE) Loss Function

Recall the nonlinear programming method to minimize the estimated mean squared error (MSE) loss function. The individual MSE loss function is

$$MSE_i = E_z[y_i(\mathbf{x},\mathbf{z}) - T_i]^2 = \{E[y_i(\mathbf{x},\mathbf{z})] - T_i\}^2 + Var_z[y_i(\mathbf{x},\mathbf{z})] \tag{5-2}$$

This method admits a little bias in the response mean, but reduces the response variability and MSE at the same time. To simultaneously optimize all m responses, we can combine the responses as a single function for optimization

$$\hat{MSE} = \sum_{i=1}^{m} w_i MSE_i, \tag{5-3}$$

where w_i is weighting coefficient. Since we knew in advance y_1 is associated with a larger weight than y_2 in this case study, we assumed the weights as $w_1 = 50$ and $w_2 = 1$.

The mean and variance models are

$$\begin{aligned}\hat{E}(y_1) &= 1.38 - 0.361x_1 - 0.155x_2 + 0.0771x_3 + 0.0481x_1^2 - 0.148x_1x_2 + 0.0218x_1x_3 \\ &\quad + 0.0130x_2x_3\end{aligned}$$

$$\begin{aligned}\hat{E}(y_2) &= 1.64 + 0.592x_1 + 0.438x_2 - 0.095x_3 + 0.0201x_1^2 + 0.301x_1x_2 - 0.143x_1x_3 \\ &\quad - 0.0844x_1x_2x_3\end{aligned}$$

$$\hat{Var}(y_1) = (-0.0588 + 0.01x_1)^2 + (-0.0116)^2 + 0.00033$$

$$\hat{Var}(y_2) = (0.0794x_1)^2 + 0.0242$$

The MSE loss functions are

$$MSE_1 = E_z(y_1(x,z) - 1)^2 = (\hat{E}(y_1) - 1)^2 + \hat{Var}(y_1)$$

$$MSE_2 = E_z(y_2(x,z) - 1)^2 = (\hat{E}(y_2) - 1)^2 + \hat{Var}(y_2)$$

The multiobjective optimization is formulated as

$$\begin{gathered} Min \quad 50MSE_1 + MSE_2 \\ subject\ to \\ 0.9 \le \hat{E}(y_1) \le 1.1, 1 \le \hat{E}(y_2) \le 3 \\ -1 \le x_i \le 1, i = 1, 2, 3 \end{gathered}$$

Since we have three control factors, one factor should be fixed to draw the response surface plots. Figure 5-8 presents the response surface plots for MSE_1, MSE_2, and the combined objective function $50MSE_1 + MSE_2$ versus x_1 and x_2, while the third factor x_3 is fixed at -1.

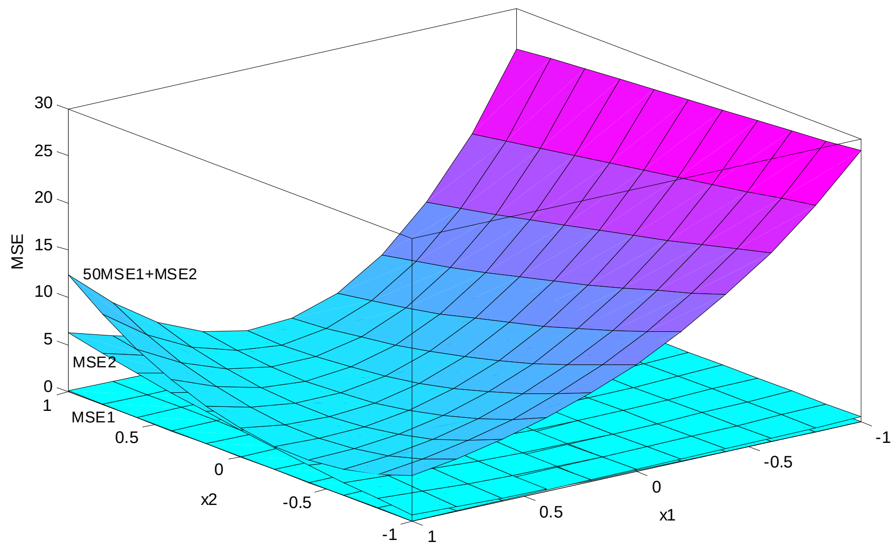

Figure 5-8 Response surface plots for MSE_1 and MSE_2 versus x_1 and x_2 ($x_3 = -1$)

Matlab was used to solve the optimization problem. Same as other methods, there is no unique solution to this problem and tradeoff is always involved, which means an improvement in one response requires a degradation of another. All the solutions from this method are within a small feasible region of the design space. We selected the following solution:

$$x_1 = 0.4317,\ x_2 = 0.3408,\ x_3 = -1,$$

$$\hat{Var}(y_1) = 0.0034,\ \hat{Var}(y_2) = 0.0254,\ \hat{E}(y_1) = 1.0675,\ \hat{E}(y_2) = 2.2959,$$

$$MSE_{y_1} = 0.0080,\ MSE_{y_2} = 1.7045$$

Comparing the three parameter design methods, we notice all of them need to determine the weights arbitrarily while it is not easy to justify the weights in the experimental design. However, the weights for the desirability functions are much simpler to decide and make less difference for the final solutions. Therefore, desirability function approach is recommended for the multiple-response parameter design.

Chapter 6
Summary and Future Research

6.1 Summary

In this thesis, integrated robust design was studied as a cost-effective method to improve quality in product design and manufacturing. Robust parameter design works on the control factors and noise factors to optimize the parameters of control factors and minimize the variability transmitted from the internal and external noise factors. The goal of parameter design is to fulfill the requirements of the quality characteristics or the responses. Robust tolerance design aims to balance the quality level and overall cost to achieve high quality with low cost. Response surface methodology is used in system design and manufacturing to construct, model, and analyze the products or processes. The final optimum solutions are obtained through nonlinear programming methods.

It was shown that the RSM is superior to Taguchi approach and is a natural fit for robust design problems. Factorial or fractional factorial designs are the basis for the cuboidal designs and spherical designs. A variety of RSM designs were introduced and compared. Standard or modified central composite design (CCD) and face-centered cube (FCC) are two of the best choices for second-order robust design problems.

Dual RSM, mean square error (MSE) loss criterion, and desirability functions can be selected to combine with quality loss function and production cost function to formulate the optimization problems. Based on our study, we can conclude that ordinary least squares (OLS) method is suitable for modeling constant residual variation, while generalized linear model (GLM) method can be used to fit non-constant residual variation. Dual RSM and ordinary least squares (OLS) or generalized linear model (GLM) work together very well in robust design to obtain the regression model and formulate the constrained optimization. In addition, desirability function approach is a good choice to solve multiple-response parameter design problems.

Internal noises factors, or the standard deviations of the control factors, were included in the experimental design to compare with the design that does not include the internal noises. It was confirmed that considering the internal noise will improve the accuracy of the solution. In addition, considering the internal noise factors as part of control variables in the combined array design is an attractive alternative to the traditional method that models the internal noise factors as part of the noise variables.

Tolerance design is an important part of robust design and its results will affect the nominal values of the parameters. Therefore, the integrated robust design, which carries out parameter design and tolerance design simultaneously, is more competitive than the traditional sequential designs. This research identified the shortage and existing problem for tolerance design. The univariate and multivariate quadratic loss functions are effective to formulate the objective function of cost.

A framework of integrated robust design was proposed and discussed in detail. Two real world case studies were considered. It was demonstrated that the proposed methodology is appropriate for solving complex design problems in industry applications.

6.2 Recommendations for Future Research

In this research, we assumed the noise variables are independent, so the interactions among the noise variables are not included in the response surface model. In real applications, the lack of attention adequately dealing with these potential interactions may lead to critical mistakes. Therefore, dependence among the noise variables should be investigated further.

While relatively large amount of work has been done on robust parameter design, not enough has been focused on the research of robust tolerance design, even though the tolerance design will affect the parameter design and the overall robust design. The loss-of-quality cost formulated through the loss function and the production cost, particularly their cost coefficients, are questionable in the practice. The loss-of-quality cost is difficult to obtain because the economic information covers the areas of design, production, sales, and customer service. Furthermore, since some kinds of cost may associate with the whole life cycle of the product or process, it is impossible to have the accurate cost data. As a result, more studies should be performed on tolerance design to find a generalized and effective methodology.

Multiple response optimization problems usually result in a set of optimal operating conditions or a desirable range. More works can be done to choose the appropriate approach and standardize the method to make final selection when detailed process knowledge is or is not available.

Computer simulation models based on finite element analysis or network simulation have been widely used in RSM and robust design as an economic alternative to physical experiments. The RSM optimization can be carried out on the metamodel to determine the optimum conditions for the real system. For complex systems, Latin hypercube design and other space-filling designs could be researched further to improve the design modeling and analysis.

This research used response surface methodology to solve robust design problems on the product design and manufacturing only. Further research can extend to other areas, such as financial applications, environmental studies, supply chain, service industry, and so on.

Appendix A

Contrast Constants and Effect Estimates for the Filtration Rate Experiment

Observation	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
(1)=45	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+
a=71	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-
b=48	-	+	-	-	-	+	+	-	-	+	+	+	-	+	-
ab=65	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
c=68	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-
ac=60	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
bc=80	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
abc=65	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-
d=43	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-
ad=100	+	-	-	+	-	-	+	+	-	-	+	-	-	+	+
bd=45	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
abd=104	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-
cd=75	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
acd=86	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-
bcd=70	-	+	+	+	-	-	-	+	+	+	-	-	-	+	-
abcd=96	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Effect Estimates	21.625	3.125	9.875	14.625	0.125	-18.125	16.625	2.375	-0.375	-1.125	1.875	4.125	-1.625	-2.625	1.375

Appendix B

Experimental Design Data of the Injected Fuel

Run number	x_1	x_2	z_1	z_2	z_3	z_4	y
1	-1	-1	-1	-1	-1	-1	91.27
2	1	1	-1	-1	-1	-1	93.11
3	1	-1	1	-1	-1	-1	94.84
4	-1	1	1	-1	-1	-1	90.65
5	1	-1	-1	1	-1	-1	93.11
6	-1	1	-1	1	-1	-1	89.28
7	-1	-1	1	1	-1	-1	90.47
8	1	1	1	1	-1	-1	92.51
9	1	-1	-1	-1	1	-1	94.62
10	-1	1	-1	-1	1	-1	90.68
11	-1	-1	1	-1	1	-1	92
12	1	1	1	-1	1	-1	93.88
13	-1	-1	-1	1	1	-1	92.32
14	1	1	-1	1	1	-1	92.34
15	1	-1	1	1	1	-1	93.88
16	-1	1	1	1	1	-1	89.99
17	1	-1	-1	-1	-1	1	94.08
18	-1	1	-1	-1	-1	1	90.44
19	-1	-1	1	-1	-1	1	91.6
20	1	1	1	-1	-1	1	93.45

21	-1	-1	-1	1	-1	1	91.46
22	1	1	-1	1	-1	1	91.91
23	1	-1	1	1	-1	1	93.45
24	-1	1	1	1	-1	1	89.58
25	-1	-1	-1	-1	1	1	91.01
26	1	1	-1	-1	1	1	93.25
27	1	-1	1	-1	1	1	94.94
28	-1	1	1	-1	1	1	91.01
29	1	-1	-1	1	1	1	93.25
30	-1	1	-1	1	1	1	89.65
31	-1	-1	1	1	1	1	90.74
32	1	1	1	1	1	1	92.66
33	-1	0	0	0	0	0	90.41
34	1	0	0	0	0	0	93.54
35	0	-1	0	0	0	0	92.2
36	0	1	0	0	0	0	91.42
37	0	0	-1	0	0	0	92.01
38	0	0	1	0	0	0	91.79
39	0	0	0	-1	0	0	92.25
40	0	0	0	1	0	0	92.2
41	0	0	0	0	0	0	91.92

References

Ames, A. E., Mattucci, N., MacDonald, S., Szonyi, G., and Hawkins, D. M. (1997), "Quality Loss Functions for Optimizations across Multiple Response Surfaces", Journal of Quality Technology 29, 339-346.

Andersson, P. (1997), "On Robust Design in the Conceptual Design Phase: A Qualitative Approach", Journal of Engineering Design, 8:1, 75-89.

Borkowski, J. J., and Lucas, J. M. (1997), "Designs of mixed resolution for process robustness studies", Technometrics, 39, 63-70.

Borror, C. M. (1998), "Response Surface Methods for Experiments Involving Noise Variables", Ph.D. Dissertation, Department of Industrial Engineering, Arizona State University, Tmpe, AZ.

Borror, C. M., and Montgomery, D. C. (2000), "Mixed Resolution Designs as Alternatives to Taguchi Inner/Outer Array Designs for Robust Design Problems", Quality and Reliability Engineering International, 16, 117-127.

Borror, C. M., Montgomery, D. C., and Myers, R. H. (2002), "Evaluation of Statistical Designs for Experiments Involving Noise Variables", Journal of Quality Technology, 34, 54-70.

Box, G. E. P. (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations" (with discussion), Technometrics, 30, 1-40.

Box, G. E. P. and Draper N. R. (1987), Empirical Model Building and Response Surfaces, Jhon Wiley & Sons, New York.

Box, G. E. P. and Wilson, K. B. (1951), "On the Experimental Attainment of Optimum Conditions", Journal of the Royal statistical Society, Series B, 13, 1-45.

Brenneman, W. A. and Myers, W. R. (2003), "Robust Parameter Design with Categorical Noise Variables", Journal of Quality Technology, 35(4), 335-341.

Chase, K. W., Greenwood, W. H., Loosli, B. G., and Hauglund L. F. (1990), "Least cost tolerance allocation for mechanical assemblies with automated process selection", ASME Manufacturing Review, 3(1), 45-59.

Copeland, K. A. F., and Nelson, P. R. (1996), “Dual response optimization via direct function minimization”, Journal of Quality Technology, 28, 331-336.

Del Castillo, E. (1996), “Multiresponse Process Optimization via Constrained Confidence Regions”, Journal of Quality Technology 28(1), 61-70.

Del Castillo, E. and Montgomery D. C. (1993), “A nonlinear programming solution to the dual response problem”, Journal of Quality Technology, 25, 199-204.

Derringer, G. C. and Suich, R. (1980), “Simultaneous Optimization of Several Response Variables”, Journal of Quality Technology, 12, 214-219.

Fathi, Y. (1991), “A nonlinear programming approach to the parameter design problem”, European Journal of Operational Research, 53, 371–381.

Giovagnoli, A., and Romano, D. (2008), “Robust design via simulation experiments: a modified dual response surface approach”, Quality and Reliability Engineering International, 24, 401-416.

Giovannitti-Jensen, A. and Myers, R. H. (1989), “Graphical Assessment of the Prediction Capability of Response Surface Designs”, Technometrics, 31, 159-171.

Gremyr, I., Arvidsson, M., and Johansson, P. (2003), “Robust DesignMethodology: Current Status in Swedish Manufacturing Industry”, Quality and Reliability Engineering International, 19, 285-293.

Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M., and Borror, C. M. (2004), “Model-Robust Optimal Designs: A Genetic Algorithm Approach”, Journal of Quality Technology, 36, No. 3.

Hill, W. J., and Hunter, W. G. (1966), “A Review of Response Surface Methodology: A Literature Review”, Technometrics, 8, 571-590.

Huang, B. and Du, X. (2008), “Probabilistic uncertainty analysis by mean-value first order Saddlepoint Approximation”, Reliability Engineering and System Safety, 93, 325-336.

Khuri, A. I. and Conlon, M. (1981), “Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions”. Technometrics, 23, 363-375.

Kiefer. J. (1961), “Optimum Designs in Regression Problems. II”, Annals of Mathematical Statistics, 29, 298-325.

Kiefer, J., and Wolfowitz, J. (1959). “Optimum Designs in Regression Problems,” Annals of Mathematical Statistics, 30, 271-294.

Kim, K. and Lin, D. K. J. (1998), “Dual response surface optimization: A fuzzy modeling approach”, Journal of Quality Technology, 30, 1-10.

Lee, Y., and Nelder, J. A. (2002), “Robust design via generalized linear models”, Journal of Quality Technology, 35, 2-12.

Li, W., and Wu, C. F. J. (1999), “An integrated method of parameter design and tolerance design”, Quality Engineering, 11, 417-425.

Lin, D. K. J., and Tu, W. (1995), “Dual response surface optimization”, Journal of Quality Technology, 27, 34-39.

Lind, E. E., Goldin, J., and Hickman, J. B. (1960), “Fitting Yield and Cost Response Surfaces", Chemical Engineering Progress, 56, 62.

Mccaskey, S. D. and Tsui, K. L. (1997), “Analysis of Dynamic Robust Design Experiments”, International Journal of Production Research, 35, 1561-1574.

Mead, R., and Pike, D. J. (1975), “A Review of Response Surface Methodology from a Biometrics Viewpoint”, Biometrics, 31, 803-851.

Meng, J. (2006), “Integrated Robust Design Using Computer Experiments and Optimization of a Diesel HPCR Injector”, Ph.D. Dissertation, Department of Industrial and Manufacturing Engineering, The Florida State University.

Montgomery, D. C. (2005), Design and Analysis of Experiments (6th edition). John Wiley & Sons, New York.

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

Myers, R. H., Khuri, A. I., and Carter, W. H. (1989), “Response Surface Methodology: 1966-1988”. Technometrics, 31, 137-157.

Myers, R. H., Khuri, A. I., and Vining, G. G. (1992), “Response surface alternatives to the Taguchi robust parameter design approach”, The American Statistician, 46, 131–139.

Myers, R. H. and Montgomery, D. C. (2002), Response Surface Methodology (2nd edition), John Wiley & Sons, New York.

Myers, R. H., Montgomery, D. C., and Vining, G. G. (2002), Generalized Linear Models with Applications in Engineering and the Sciences, John Wiley & Sons, New York.

Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M., and Kowalski, S. M. (2004), "Response Surface Methodology: A Retrospective and Literature Survey", Journal of Quality Technology, Vol. 36, No. 1.

Myers, R. H., Vining, G. G., Giovannitti-Jensen, A., and Myers, S. L. (1992), "Variance Dispersion Properties of Second-Order Response Surface Designs", Journal of Quality Technology, 24, 1-11.

Nair, V. N., editor (1992), "Taguchi's Parameter Design: A panel Discussion", Technometrics, 34, 2, 127-161.

Nelder, J. A. and Lee, Y. (1991), "Generalized Linear Models for the Analysis of Taguchi Type of Experiments", Applied Stochastic Models and Data Analysis, 7, 107-120.

Phadke, M. S. (1989), Quality Engineering Using Robust Design, Englewood Cliffs, NJ: Prentice-Hall.

Pignatiello, J. J., Jr. (1993), "Strategies for Robust Multiresponse Quality Engineering", IIE Transactions, 25, 5-15.

Rice, J. (1995), Mathematical Statistics and Data Analysis (2nd edition), Wadsworth & Brooks/Cole, Pacific Grove, CA.

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

Romano, D., Varetto, M., and Vicario, G (2004), "Multiresponse Robust Design: A General Framework Based on Combined Array", Journal of Quality Technology, Jan., Vol. 36, Iss. 1, 27.

Savage, G. J. (2008), Model-Based Robust Design, Department of Systems Design Engineering, University of Waterloo.

Savage, G. J., and Seshadri, R. (2003), "Minimizing Cost of Multiple Response Systems by Probabilistic Robust Design", Quality Engineering, 16:1, 67-74.

Steinberg, D. M., editor (2008), "The Future of Industrial Statistics: A Panel Discussion", Technometrics, 50, 2, 103-127.

Taguchi, G., (1986), Introduction to Quality Engineering, UNIPUB/Kraus International, White Plains, NY.

Taguchi, G., (1987), System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost, UNIPUB/Kraus International, White Plains, NY.

Taguchi, G. and Wu, Y. (1985), Introduction to Off-Line Quality Control, Nagoya, Japan, Central Japan Quality Control Association.

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

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

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

Youn, B. D. and Choi, K .K. (2004), "Selecting Probabilistic Approaches for Reliability-Based Design Optimization", AIAA Journal, Vol. 42, No. 1.