THE STUDY OF SIMULTANEOUS OPTIMIZATION FOR SEVERAL RESPONSES

by

ASTINI SALIHIMA

Bachelor of Science
Textile Technology Institute
Bandung, Indonesia
1972

Master of Science
Bogor Agriculture Institute
Bogor, Indonesia
1979

Master of Science
Iowa State University
Ames, Iowa
1989

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May 1993

THE STUDY OF SIMULTANEOUS OPTIMIZATION

FOR SEVERAL RESPONSES

Thesis Approved:

J. Leroy Folks

Thesis Adviser

P. L. Claypool

W. W. Warde

M. P. Terrell

Thomas C. Collins

Dean of the Graduate College

ACKNOWLEDGMENTS

Because of the mercy of GOD, ALLAH the Great, I was finally able to finish this dissertation. I would like to thank very much my committee members, Dr. J. Leroy Folks, Dr. W. D. Warde, Dr. P. L. Claypool and Dr. M. P. Terrell, for their time and efforts guiding me to finish my study. Especially, I appreciate Dr. J. Leroy Folks, my major adviser and the chairman of the committee, for his excellent guidance, patience, and never-ending support. He always encouraged me when I faced difficult work.

There are many instructors, faculty members, staffs of Statistics Department, friends, librarians, and others who helped me during my study at Oklahoma State University. I can only say thanks very much to them. My thanks are also addressed to Mrs. Frances Griffin who helped with proofreading and editing.

I would like to acknowledge the financial support I have received from the Indonesian Government, during my study in the U.S.A. This financial support is administered by BAPPENAS Overseas Training Office in Indonesia and by MUCIA/Indonesian Overseas Training Support Office at the Ohio State University. I appreciate The Institute of

Research and Development for Textile Industries in Bandung Indonesia, for permission to study abroad without losing my job.

I expressed my thanks to my late parents for their encouragement when I was young to reach the highest educational status. I appreciate my family for their support and sacrifices, especially my husband Abbas and my daughters Eka and Dwi, who live far away in my home country. Without their encouragement and support I would never have been able to begin this degree.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

My interest in optimization arises from involvement in quality improvement problems. Quality improvement can be achieved by determining optimum combinations of various control variables for producing a specific product. If the quality of a product is measured by only one property, we deal with optimization of one property (response). Actually, however, the quality of a product is measured by several properties. Therefore, the decision for choosing an appropriate combination of control variables become difficult, since many (sometime all) properties are affected simultaneously.

Optimization of one response (property) has been discussed and published widely by many researchers for both linear and nonlinear optimization (Wismer and Chattergy, 1979; Steuer, 1986; Mockus and Mockus, 1991). Kirkpatrick et al (1983) have proposed optimization by using simulated annealing for finding a global maximum point if the response function has at least one local maximum. Simulated annealing has been discussed in more detail by Bertsimas and Tsitsiklis (1992). Mockus and Mockus (1991), have proposed a Bayesian approach for global optimization for both unconstrained and constrained optimization.

Optimizing only one property does not seem good enough, since it often happens that optimizing one property adversely affects the other properties. It may happen that this optimization yields a product that has lower quality than before, therefore it is necessary to optimize all properties (responses).

Some researchers optimize the most important response and put constraints on the others . The solution for this optimization often lies on at least one of the boundaries of the constraints. Thus, this approach may not always give an optimum solution.

Taguchi (Ross 1988) was the first to introduce a two-step optimization for two-response cases (response mean and its variance). First, he minimizes the variance; then he sets the response mean close to a target value (Baker, 1986; Leon et al, 1987; and Ross, 1988). Taguchi's method has had great success in Japan. Many American manufacturers use his method while some scientists and statisticians have criticized or modified it (Leon et al 1987; Box 1988).

Most recently, scientists have paid attention to simultaneous optimization for several responses. Yet, it is still a perplexing problem and the procedure is complicated. A simultaneous optimization method usually cannot optimize all the responses. In general, there can be no single best optimum point for all individual responses. However, some points are definitely better than others. Consequently, we adopt a compromise, which leads to a consideration of the term "admissible points." The set of all admissible points is called the admissible set.

Every point x^* that belongs to an admissible set (in the sense of maximization) gives value $Y_i(x^*) > Y_i(x)$ for all i and for all x that lie outside the admissible set. Therefore, the characterization of sets of admissible solutions for optimizing multiple response functions is of particular interest. We may choose a point from many points in the admissible set. The choice is governed by decision makers, who will consider the advantages of trade-off among the responses.

One way to simplify the simultaneous optimization process is to apply a univariate approach. All original responses are combined into a single new response. Thus, using this approach reduces the multiple response problems to a single response situation, for which the methods of optimization are widely available.

Some combined response functions have been introduced for simultaneous optimization for several responses. Harrington (1965) introduced desirability functions and computed the geometric mean as a combined response function. Then Derringer and Suich (1980) extended the desirability functions to find better performances. Khuri and Conlon (1981) used a distance function as a measure of the deviation from the ideal optimum along with the variance-covariance matrix. Mockus and Mockus (1991) used a

Bayesian approach to maximize $\sum_{i=1}^{k} \alpha_i Y_i(\mathbf{x})$, the weighted sum of the original responses $Y_i(\mathbf{x})$, where $\alpha_i > 0$. Steuer (1986) discussed the weighted sum of original responses (first-order polynomial functions). Vos (1990) has introduced linear utility functions for choosing decisions in education cases. However, none of the researchers mentioned above has discussed the characterization of sets of admissible points for optimizing several second-order polynomial functions.

The main objectives of this thesis are to characterize sets of admissible points for several kinds of surfaces of the original responses and to determine conditions under which maximizing a combined response leads to an admissible solution. For this purpose, some lemmas, theorems, and definitions are developed in this thesis. The characterization is developed for both unconstrained and constrained optimization. The feasible region for constrained optimization is assume to be a closed convex set and that the original responses are limited to second-order polynomial functions of vector $\mathbf{x}$.

Each of the original responses has either a maximum, a minimum, or a saddle points. Since a combined response is derived from several original responses, we have several combinations of surfaces of the original responses. Those combinations are considered for characterizing the sets of admissible points. The characterization for two or three original responses as functions of x_1 and x_2, can be illustrated by graphs, otherwise by algebraic notations.

Chapter II contains some definitions related to optimization and the review of some earlier work in simultaneous optimization using a univariate approach. The characterization of sets of admissible points and the determination of conditions under which optimization of a combined response leads to an admissible point are developed in Chapter III. Chapter IV contains inferences related to the optimum point of maximization, using the convex combination method as a special case of the weighted sum method. This chapter also presents numerical examples about the confidence region of $\mathbf{x}$ for fixed α_i and the region of $\mathbf{x}$ if we impose constraints on α_i. Chapter V presents comparisons of the admissibility of 4 methods: the convex combination, Harrington's, Derringer-Suich's, and Khuri-Conlons's methods. Then the summary and conclusions are presented in Chapter VI.

CHAPTER II.

LITERATURE REVIEW

In this chapter some definitions for optimization and admissibility will be given. Before discussing optimization problems, it will be necessary to give some mathematical notation and definitions. The literature review is focused on multiple response optimization using a univariate approach.

Mathematical Preliminaries

Throughout this thesis a vector is denoted by a bold small letter and a matrix is denoted by a bold capital letter. Also the transpose of a vector $\mathbf{x}$ or a matrix $\mathbf{A}$ is denoted by $\mathbf{x}'$ or $\mathbf{A}'$.

Some definitions that relate to optimization and admissibility are presented in this section. For constrained optimization, we assume that the feasible region is a closed bounded convex set. The convex combination method is discussed later.

Definitions for a convex set and a convex combination are given below:

Definition II.1. A set S is convex if and only if for $x_i \in S$, a point $\hat{\mathbf{x}} = \sum_{i=1}^{k} \alpha_i x_i$ is also in set S, where

$\alpha_i > 0$ and $\sum_{i=1}^{k} \alpha_i = 1.$

Definition II.2. W is said to be a convex combination of $Y_i(x)$, $i = 1, 2, \ldots, k$, if

$$W = \sum_{i=1}^{k} \alpha_i Y_i(x), \ \alpha_i > 0 \text{ and } \sum_{i=1}^{k} \alpha_i = 1.$$

A second order polynomial response have three kinds of stationary points: maximum, minimum, and saddle. Their definitions are as follows:

Definition II.3. Let $Y(x)$ be a function over a closed set A in E^n . $Y(x)$ is said to have a global maximum point, at x^* if and only if

$$Y(x^*) \geq Y(x) \quad \text{for all } x \in A.$$

If $Y(x^*) > Y(x)$ for all $x \in A$, then the global maximum point is unique and is called the proper global maximum.

Definition II.4. Let $Y(x)$ be a function over a closed set A in E^n. Then $Y(x)$ is said to have a global minimum point at x_o if

$$Y(x_o) \leq Y(x), \quad \text{for every } x \in A.$$

If $Y(x_o) < Y(x)$ for all $x \in A$, then the global minimum is unique and is called the proper global minimum.

Definition II.5. Let $Y(x)$ be defined at all points in some δ neighborhood of $x_o \in E^n$. Then $Y(x)$ is said to have a local maximum point at x_o if there exists an ε, $0 < \varepsilon < \delta$, such that for all x, $0 < \| x - x_o \| < \varepsilon$,

$$Y(x_o) \geq Y(x).$$

δ neighborhood of x_o is the region that has radius δ from x_o. If $Y(x_o) > Y(x)$, then the local maximum is unique and is called the strong local maximum.

Definition II.6. Let $Z = f(x,y)$ be a differentiable function. Then Z is said to have a saddle point at (x_o, y_o) if there exists an $\varepsilon > 0$, such that for all x, $\|x - x_o\| < \varepsilon$ and for all y, $\|y - y_o\| < \varepsilon$,

$$f(x, y_o) \leq f(x_o, y_o) \leq f(x_o, y).$$

Admissibility become a particular focus in this thesis. We need to define the definition of admissible point and the set of admissible points.

Definition II.7. $x^o \in R$ is an admissible point for $y(x)$, $y \in E^k$ and $x \in E^p$, if and only if there does not exist $\hat{x} \in R$, such that $\hat{x}$ is better than x^o. In the sense of maximization, the above definition can be stated as $x^o \in R$ is admissible for $y(x) = \{Y_1(x), .., Y_k(x)\}$, $x \in E^p$, if and only if there does not exist $\hat{x} \in R$ such that

$$Y_i(\hat{x}) \geq Y_i(x^o), \ \forall \ i \text{ and}$$

$$Y_i(\hat{x}) > Y_i(x^o), \text{ for at least one } i.$$

Definition II.8. S is an admissible set for $y(x)$, $y \in E^k$ and $x \in E^p$, if and only if S is the set of all x^o, such that x^o is admissible for $y(x)$.

The gradient of Y(x) is important in determining the direction of the path of steepest ascent, the stationary point, the tangent path of $Y_1(x)$ and $Y_2(x)$. and the Hessian matrix, etc. Definitions for gradient and other related terms are given here.

Definition II.9. Let Y (x) be differentiable function of $x \in E^p$. The gradient of Y(x), denoted by $\nabla Y(x)$ is define as

$$\nabla Y(x) = \left(\frac{\partial Y}{\partial x_1} i_1, \ \frac{\partial Y}{\partial x_2} i_2, \ldots\ldots, \ \frac{\partial Y}{\partial x_p} i_p \right),$$

where i_1, i_2, ..., i_p are their coordinate axis.

Definition II.10. Let Y(x) be differentiable. A stationary point of Y(x) is a point that satisfies

$$\nabla Y(x) = 0.$$

Definition II.11. The Hessian matrix of Y(x), denoted by $H_{Y(x)}$, is the matrix of second derivative of Y(x) with respect to x. The element of $H_{Y(x)}$ is $H_{ij} = \dfrac{\partial^2 y}{\partial x_i \partial x_j}$.

As has been mentioned above, there are three kinds of stationary points: maximum, minimum, and saddle points. Let x_o be the stationary point. The nature of the stationary point for each response is determined by its Hessian matrix, as follows:

(1). Local maximum, if $H_{Y(x)}|_{x_o}$ is negative definite.

(2). Local minimum, if $H_{Y(x)|x_o}$ is positive definite.

(3). Saddle point, if $H_{Y(x)|x_o}$ is indefinite.

"Criterion cone" is generated by gradients of $Y_i(x)$, $i = 1, 2, \ldots, k$. In a univariate approach, k original responses are combined into one combined response. The gradient of the combined response should lie in the criterion cone of the original responses, so that the solution for maximizing the combined response leads to an admissible solution.

A criterion cone is important for optimizing several responses. Steuer (1986) defined the criterion cone as a convex cone generated by k response gradients (gradient of $Y_1, Y_2, \ldots, Y_k$ or $\nabla Y_i(x)$). The size of the criterion cone is defined by the number of linearly independent $\nabla Y_i(x)$, $i = 1, 2, \ldots, k$. If the number of linearly independent $\nabla Y_i(x)$ is $j \le k$, then the criterion cone is of dimension j. The null vector condition is in effect if there exists $\alpha_i > 0$, $\Sigma\alpha_i = 1$, such that $\Sigma\alpha_i\nabla Y_i(x) = 0$, $i = 1, \ldots, k$. The criterion cone and the null vector condition are shown in Figure 1 and Figure 2.

Some Methods for Multiobjective Optimization

Each of the original responses $Y_1, Y_2, \ldots, Y_k$ is considered as a function of p control variables, $x_1, x_2, \ldots, x_p$. This function may be unknown or sometimes it may be known from the engineering, physics, or chemistry.

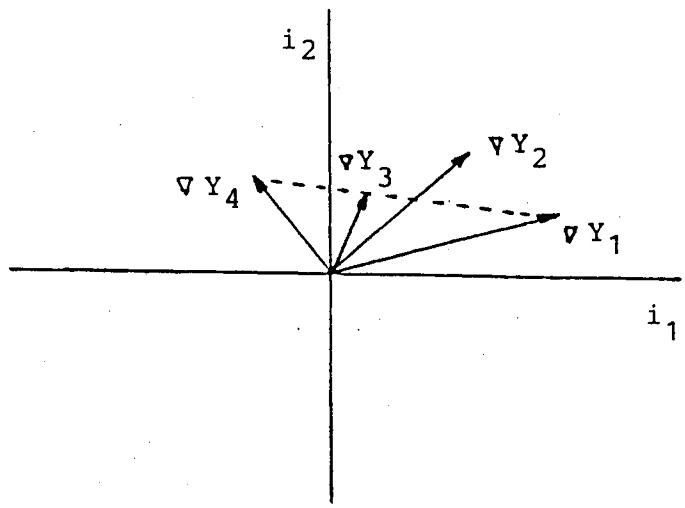


Figure 1. Criterion Cone of ∇Y_i

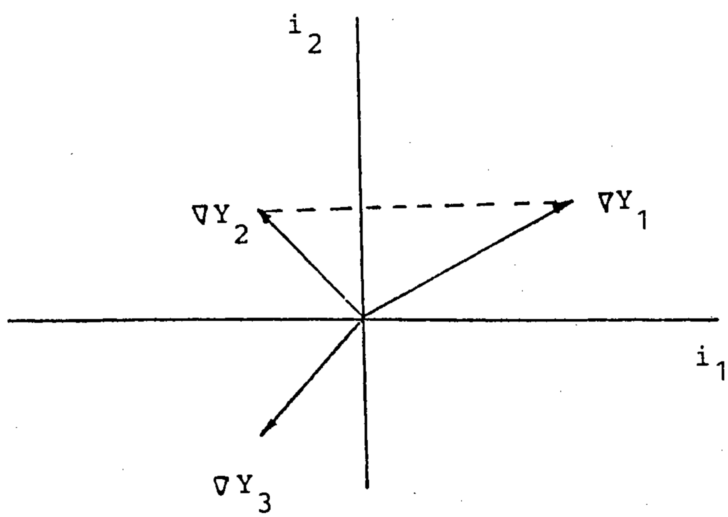


Figure 2. Null Vector Condition of a Criterion Cone

Maximizing an unknown function is attempted either (1) by fitting a function to experimental data and then maximizing the fitted function or (2) by using some empirical search procedure to try to find the maximum. Sometimes we may follow the search procedure and fit a function with a subsequent maximization of the fitted function.

Harrington (1965), Derringer and Suich (1980), and Khuri and Conlon (1981) used a single experimental design to estimate the function of $Y_i(x)$, i = 1, .., k and $x \in E^p$. Then they defined transformations based on $\hat{Y}_i(x)$, the estimator of $Y_i(x)$.

Harrington's Method

Let Y_i be the i^{th} response, Y_i^* be the upper specification limit and Y_{i*} be the lower specification limit of Y_i. $\hat{Y}_i(x)$ is the estimator of $Y_i(x)$ by using regression analysis (usually a second-order polynomial equation). Harrington's desirability functions are

$$d_i = \exp(-|Z_i|^n),\ n > 0$$

where

$$Z_i = \frac{\hat{Y}_i(x) - (Y_i^* + Y_{i*})/2}{Y_i^* - Y_{i*}} . \tag{2.1}$$

Then he maximizes the geometric mean of d_i defined by

$$D = \left(\prod_{i=1}^{k} d_i\right)^{1/k},\ i = 1, 2, \ldots., k \tag{2.2}$$

The choice of n is subjective and is governed by the importance the engineer places on each of the responses. We may choose a different value of n for every response. The greater values of n are assigned to the more important responses. More discussion concerning the admissibility of this method is presented in Chapter V.

Derringer-Suich's Method

Derringer and Suich (1980) modified Harrington's desirability functions to find better performances. Two cases will arise: one-sided and two-sided desirability functions. For one-sided cases, they considered the desirability functions given by

$$d_i = \begin{cases} 0 & \hat{Y}_i < Y_{i*} \\ \left[\dfrac{\hat{Y}_i - Y_{i*}}{Y_i^* - Y_{i*}} \right]^r & Y_{i*} \le \hat{Y}_i \le Y_i^*. \\ 1 & \hat{Y}_i > Y_i^* \end{cases} \qquad (2.3)$$

The engineer specifies the value of r and the minimum acceptable value of Y_{i*}. For one-sided cases, there is no highest acceptable value for $\hat{Y}$. However, from practical experience one may think that a value greater than Y_i^* lacks additional worth. on each of the responses. The more important the responses, the greater values of r are assigned to them. The values of r also define the speed of increase in d_i. If r = 1, the increase in d_i is constant. If r > 1, first d_i increases slowly when $\hat{Y}_i$ near Y_{i*}, then d_i increases rapidly when $\hat{Y}_i$ approach Y_i^*. If r < 1, the increase in d_i is opposite to that when $r > 1$ as shown in Figure 3a.

For two-sided cases, the desirability functions are

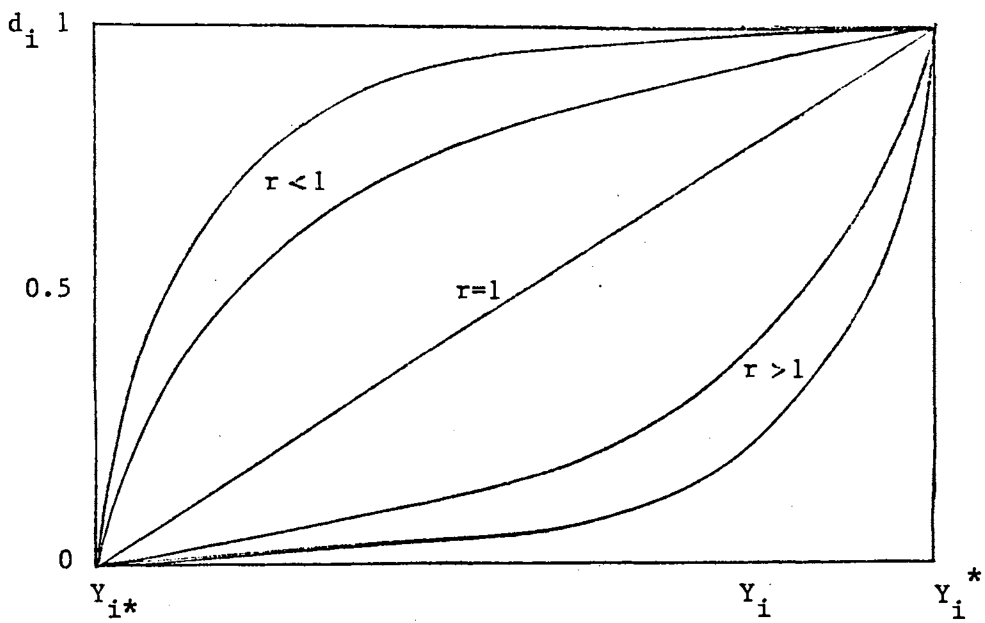
$$d_i = \begin{cases} \left[\dfrac{\hat{Y}_i - Y_{i*}}{c_i - Y_{i*}}\right]^s, & Y_{i*} \le \hat{Y}_i \le c_i \\ \left[\dfrac{\hat{Y}_i - Y_i^*}{c_i - Y_i^*}\right]^t, & c_i < \hat{Y}_i \le Y_i^* \\ 0 & , \quad \hat{Y}_i < Y_{i*} \text{ or } \hat{Y}_i \end{cases} \tag{2.4}$$

$$D_i = (\prod d_i)^{1/k}, \quad i = 1, 2, \ldots., k.$$

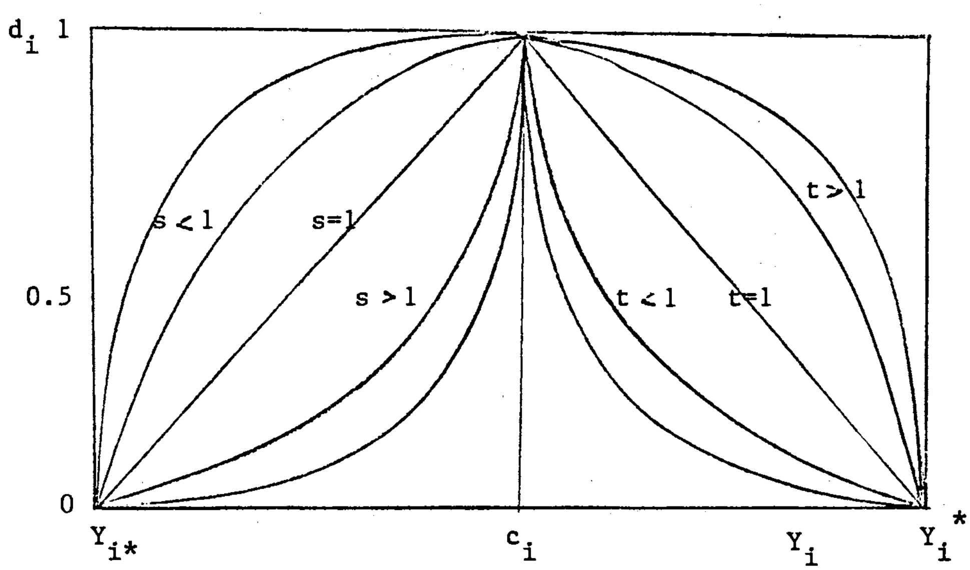
Again, the values of s and t are specified by the engineer. If all the responses are equally important or the increase in d_i are constant, then $s = t = 1$. The greater values of s or t are assigned to the more important responses. Figure 3b shows the relationship between d_i and s or t. If s or t is greater than 1, then the increase in d_i is slow, when $\hat{Y}_i$ is closer to Y_i^* (for s) or Y_{i*} (for t) than c_i is. In contrast, the increase in d_i is fast when $\hat{Y}_i$ is closer to c_i than Y_i^* or Y_{i*} is. If s or t less than 1, the increase in d_i is opposite to that if r or s is greater than 1. More discussion concerning the admissibility of this method is presented in Chapter V.

Khuri-Conlon's Method

Khuri and Conlon (1981) minimized the weighted distance of $\hat{Y}_i(x)$, the estimator of $Y_i(x)$, considered as a



a. One-sided Specification Limit



b. Two-sided Specification Limit

Figure 3. Effects of r, s, or t on d_i

point in r-dimensional space, from ϕ, the vector of individual optima. The distance function is denoted by

$$\rho[\hat{y}(x),\ \phi]=[\{\hat{y}(x)\ -\ \phi\}'\{\mathrm{var}(\hat{y}(x))\}^{-1}\{y(\hat{x})\ -\ \phi\}]^{1/2}. \qquad (2.5)$$

They used a polynomial regression function for $Y_i(x)$. From k responses they chose r linearly independent responses for simultaneous optimization. The polynomial equation for each $\hat{Y}_i(x)$ is found by using multiple regression analysis. If the elements of vector $y(x)$ are independent, then $\mathrm{var}\{\hat{y}(x)\}$ can be assumed to be a diagonal matrix.

Minimizing $\rho[\hat{y}(x),\phi]$ is equivalent to minimizing

$$\min_x \rho^2 = [\{\hat{y}(x)-\phi\}'\{\mathrm{diag}(\sigma_i^2 z_i'(x)(X'X)^{-1}z_i(x))\}^{-1}\{y(\hat{x})-\phi\}]$$

$$= \sum_{i=1}^{r}\{\hat{Y}_i(x)\ -\ \phi_i\}^2/(\sigma_i^2 c_i)$$

where $c_i = z_i'(x)(X'X)^{-1}z_i(x)\}$, and $z_i'(x)$ is a row vector of dimension m whoes first element is 1 and the remaining elements consist of power and cross-product of powers x_1, .., x_p as dictated by the polynomial model. X is the matrix of control variables x in constructing the regression analysis, whose first column is vector 1.

In minimizing $\rho[\hat{y}(x),\ \phi]$ in the above equation, ϕ was treated as a vector of constant. Khuri and Conlon (1981) also considered the randomness of ϕ. Here, if ϕ is the vector of individual optimum values of the random vector $\hat{y}(x)$, then ϕ is also a random vector. Let the true value of the individual optimum be a vector ζ; then the objective is to minimize $\rho[\hat{y}(x),\zeta]$. Since ζ is unknown, Khuri and

Conlon (1981) have decided to minimize the upper bound of the distance. They set up a confidence region about ζ with a certain degree of confidence. The region is denoted by D_ζ. For a fixed value of x in the experimental region R, then

$$\rho[\hat{y}(x),\zeta] \leq \max_{\eta \in D_\zeta} \rho[\hat{y}(x),\eta] \tag{2.6}$$

where η is a point in D_ζ. The right side of the equation overestimates the distance $\rho[\hat{y}(x),\zeta]$. However, the minimum of $\rho[\hat{y}(x),\zeta]$ over R cannot exceed the corresponding minimum of the upper bound, so

$$\min_{x \in R} \rho[\hat{y}(x),\zeta] \leq \min_{x \in R} \{\max_{\eta \in D_\zeta} \rho[\hat{y}(x),\eta]\} \tag{2.7}$$

$$\text{Let } d_o = \min_{x \in R}\{ \max_{\eta \in D_\zeta} \rho[\hat{y}(x),\eta]\}. \tag{2.8}$$

Then the minimum over R of the distance between $\hat{y}(x)$ and ζ is less then or equal to d_o. More discussion concerning the admissibility of this method is presented in Chapter V.

Linear Utility Function

Vos (1990) proposed a linear utility function for optimizing four different type of decision problems in education: selection, mastery, placement, and classification. The optimization is based on Bayesian decision theory, to search for a decision that maximizes the expected utility.

Let x_{ci} and y_{ci} denote the cut-off scores for sub-population i on observed test score variables X and Y, respectively. Let t_c be the cut-off score on the criterion score T, which is assumed to be equal for each population and which the decision maker set before the optimization process. The objective is to optimize x_{ci} and y_{ci} simultaneously, for a given t_c.

In his example, Vos defines the following decision

$$\delta(X,Y) = \begin{cases} a_0, & \text{for } X < x_{ci} \\ a_1, & \text{for } X \geq x_{ci},\ Y < y_{ci} \\ a_2, & \text{for } X \geq x_{ci},\ Y \geq y_{ci} \end{cases} \tag{2.9}$$

where a_0, a_1, a_2 stand for the actions to reject a student, to retain an accepted student, and to advance an accepted student, respectively. He states a combined decision problem as a linear functions in t for sub population i:

$$u_{ji}(t) = \begin{cases} b_{0i}(t_c - t) + d_{0i}, & \text{for } X < x_{ci} \\ b_{1i}(t - t_c) + d_{1i}, & \text{for } X \geq x_{ci},\ Y < y_{ci} \\ b_{2i}(t - t_c) + d_{2i}, & \text{for } X \geq x_{ci},\ Y \geq y_{ci} \end{cases} \tag{2.10}$$

where $d_{ji} < 0$ is the cost and b_{ji} is the slope of the linear regression, $j = 0, 1, 2$ and $i = 1, 2$. The parameters d_{ji} and b_{ji} have to be fixed before optimization.

Using the above utility function, he maximizes

$$E[u_i(T|x_{ci}\ y_{ci})]. \tag{2.11}$$

After several integrations, differentiations, and computations he found that the optimal cut-off scores (the boundaries of scores for making actions), x_{ci} and y_{ci} can be found by solving the following equations via numerical approximation methods.

$$\int_{x_{ci}}^{\infty} \left\{ (b_{2i} - b_{1i})[E_i(T|x,\ y_{ci})-t_c] + d_{2i} - d_{1i} \right\}$$

$$\left\{ z_i(x|y_{ci})\ dx \right\} = 0 \text{ and} \tag{2.12}$$

$$(b_{0i} + b_{1i})[E_i(T|x_{ci})-t_c] + d_{1i} - d_{0i} +$$

$$\int_{y_{ci}}^{\infty} \left\{ (b_{2i} - b_{1i})[E_i(T|x_{ci},\ y)-t_c] + d_{2i} - d_{1i} \right\}$$

$$\left\{ m_i(y|x_{ci})\ dy \right\} = 0, \tag{2.13}$$

where $z_i(x|y_{ci})$ and $m_i(y|x_{ci})$ are the posterior probability function of x given $Y = y_{ci}$ and the posterior probability function of $X = x_{ci}$, respectively.

He applies this procedure, for one or more of the following restrictions: (i) multiple populations, (ii) quota restrictions, (iii) multivariate test data, and (iv) multivariate criteria.

Taguchi's Method

This method is a two-step optimization of two responses (means and variance of a function of **x**). Taguchi classifies the control variables **x**, into two categories: dispersion factors and adjustment factors. Dispersion factors are control variables that affect the variance or both variance and means. Adjustment factors are control variables that affect only the means.

First, Taguchi (Baker, 1986; Byrne 1987; and Ross 1988) minimizes the variance with respect to dispersion factors, by maximizing a certain criterion called "signal-to-noise ratio" (later it is written as SN ratio). Then, he adjusts the adjustment factors, such that the mean value of the response is close to a given target value. Taguchi also claims that his method will lead to minimization of a quadratic loss function,

$$L(y, T) = C(y - T)^2,$$

where C is a constant, y is the value of a response Y, and T is a given target value.

There are three kinds of SN ratio:

(1). $SN_H = -10 \log \frac{1}{n} \sum_{i=1}^{k} 1/y_i^2$, where the higher value of the response of interest is better.

(2). $SN_L = -10 \log \frac{1}{n} \sum_{i=1}^{k} y_i^2$, where the lower value is better.

(3). $SN_T = 10 \log (E^2(y)/\mathrm{var}\ y)$, where the target value is the best.

Taguchi's method can only be applied if there are adjustment factors. Leon et al. (1987) have proven that Taguchi's method will lead to minimization of a quadratic loss function only if Y(x) can be written as

$$Y = \mu(\mathbf{d}, \mathbf{a})\varepsilon(\mathbf{N}, \mathbf{d}), \tag{2.14}$$

where $E(Y) = \mu(\mathbf{d}, \mathbf{a})$ is a strict monotone function of each component **a** for each **d**; **a** are adjustment factors, **d** are dispersion factors, and **N** are noise factors. The Noise factor affects the output of a response, but is difficult to control or its control causes high production costs.

Leon et al (1987) stated that model (2.14) holds, for example, if the noise affects the output, Y, uniformly over increments of time and distance. He also stated that model (2.14) gives the $\mathrm{var}(Y)/E^2Y$ which does not depend on **a**. In contrast, if the function of Y is replaced by

$$Y = \mu(\mathbf{d}, \mathbf{a}) + \varepsilon(\mathbf{N}, \mathbf{d}), \tag{2.15}$$

with $E(\varepsilon(\mathbf{N}, \mathbf{d}) = 0$ and $L(y, T) = (y - T)^2$, then under model (2.15), SN ratio depends on **a**. Therefore, Leon et al (1987) suggest using var(Y) instead of the SN ratio.

Box (1988) gives another alternative of the SN ratio for general cases. Suppose that $Y = g(y)$ is a "variance stabilizing transformation" such that $\sigma_Y \simeq \sigma_y\, g'(\mu)$ is independent of μ and suppose further that σ^2_Y is only affected by **d**. Then minimizing the quadratic loss

$$C(y - T)^2 = \min\ \{g'(\mu)\}^{-2}\ \sigma_y^{\,2} + (\mu(\mathbf{x})-T)^2$$

is possible in two steps. By using second-order Taylor expansion, then he found

$$\mu_{Yo} = g\left\{T + \frac{3}{2}\ [g'(T)]^{-3}\ g''(T)\sigma_{Yo}^{\,2}\right\}, \tag{2.16}$$

where g' and g'' denote the first and the second derivative of g, respectively and T is the target value. In a particular case, if $Y = y^{\gamma}$ $(\gamma \neq 0)$, then

$$\mu_{Y0} \simeq \left\{T\left[1 - \frac{3}{2}\,\frac{(1-\gamma)}{\gamma^2}\,\frac{\sigma_{Y0}^{2}}{T^{2\gamma}}\right]\right\}. \tag{2.17}$$

If $\gamma = 0$, the transformation is $Y = \ln y$, then

$$\mu_{Y0} \simeq \ln T - \frac{3}{2}\,\sigma_{Y0}^{2}. \tag{2.18}$$

In his article, Box (1988) explained how to find γ.

Some Methods for Single Response Optimization

After we combined the original responses into a combined response, the problem becomes one of single response optimization. The methods or procedures for single response optimization are reviewed briefly.

There are two branches of optimization: linear programming (for linear functions of vector $\mathbf{x}$) and nonlinear programming (for non linear functions of vector $\mathbf{x}$). The popular method for computing linear programming is the simplex method. Nonlinear programming may use gradient methods or quadratic programming if the response is a quadratic function.

A simple algorithm for optimization is the "dichotomous search". With this method we can choose x_1^n and x_2^n to be symmetric or not with respect to the mid point of interval $[a^n, b^n]$. In each iteration we can choose 2 or 3 points with equal intervals. Wismer and Chattergy (1979) discuss the algorithm and its expansion in detail.

If the response is a differentiable and a continuous function, then the gradient method can be used. In this method, the fastest search for maximization is in the direction of the path of steepest ascent. This direction is the positive gradient direction. Searching is continued until the gradient vector equals zero. To avoid reaching a local maximum or local minimum point, we shall repeat the search at various starting points. This method can be used for both unconstrained and constrained optimization (Wismer and Chattergy 1979).

Kirkpatrick et al (1983) proposed the simulated annealing algorithm for global optimization, to prevent optimization from reaching a local optimum. This method is based on a physical process whereby a solid is slowly cooled and spends a long time at temperatures near the freezing point. This process yields a stable configuration structure. In their example, the annealing schedule starts at a high temperature ($T_o = 10$), then cools exponentially, where

$$T_n = (T_1/T_o)^n \, T_o, \tag{2.19}$$

with the ratio $T_1/T_o = 0.9$.

More explanation about simulated annealing has been reported by Bertsimas and Tsitsiklis (1992). They show that the simulated annealing algorithm will converge in probability to the set of global optima, S^*, if and only if

$$\lim_{t \to \infty} T(t) = 0 \text{ and}$$

$$\sum_{t=1}^{\infty} \exp\,[-d^{*}/T(t)] = \infty, \tag{2.20}$$

where d^{*} is the smallest number such that every $i \in S$ communicates with S^{*} at height d^{*}. They state that "state i communicates with S^{*} at height h if there exists a path in S (with each element of the path being a neighbor of the preceding element) that starts at i and ends at some elements of S^{*}, and such that the largest value of J (cost function) along the path is $J(i) + h$." Bertsimas and Tsitsiklis (1992) are more interested in the probability that no state in S^{*} is visited during the execution of the algorithm than the value of $P(x(t^{*}) \notin S^{*})$. They found that it is at most $A/(t^{*})^{a}$, for a given cooling schedule $T(t) = d/\log t$, where $d > d^{*}$, and A and a are some positive constant. Therefore, it converges to zero if $t \longrightarrow \infty$.

Mockus and Mockus (1990) use a Bayesian approach for optimization. They use this method for both linear and nonlinear constraints. They show that the solution of this method converges to a global minimum for any continuous function Y(x) defined on a compact set R. For multiobjective optimization, they minimize the weighted sum of the objective function $Y_i(x)$, but do not explain the functions (whether linear, quadratic, or others). They also state that the result of minimization converges to the global minimum of $\sum_{i=1}^{k} \alpha_i Y_i$, where $\alpha_i > 0$ and $i = 1, \ldots, k$, for any continuous function $Y_i(x)$ and a compact feasible set R.

Confidence Region for Optimum Points

After the optimum point of a response is estimated in any optimization process, the next task is to determine its confidence region. From the regression analysis, we can estimate the vector β, the regression coefficients of the response. We can also estimate the variance-covariance matrix of $\hat{\beta}$. The procedure for constructing the confidence region for the stationary point of the response or any $g(\beta)$, has been suggested by Calter et al (1984, 1986). This construction is based on the $100(1 - \alpha)\%$ confidence region about $g(\beta)$, indicated by Rao (1973) as given below

$$P\left\{\min_{\beta\in U} g(\beta) \le g(\beta) \le \max_{\beta\in U} g(\beta)\right\} \ge 1 - \alpha,$$

where

$$U = \left\{\beta: (\hat{\beta}-\beta)'(\mathbf{X}'\mathbf{X})(\hat{\beta}-\beta)/qs^2 \le F_{\alpha,\ q,\ n-q}\right\}.$$

Let $Y(\mathbf{x})$ be a second-order polynomial function of $\mathbf{x}$. It can be written as a general linear model,

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times q}{\mathbf{X}}\ \underset{q\times 1}{\beta} + \underset{n\times 1}{\varepsilon},$$

$$\varepsilon \sim N(0, \sigma^2 I),$$

or in a quadratic form,

$$Y = \beta_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} + \varepsilon,$$

where $\mathbf{b} = (\beta_1, \beta_2, \ldots, \beta_p)$, $q = (p+1)(p+2)/2$, and

$$\mathbf{B} = \begin{bmatrix} \beta_{11i} & \beta_{12i}/2 & & \beta_{1pi}/2 \\ & \beta_{22i} & & \beta_{2pi}/2 \\ \text{symetric} & & & \\ & & & \beta_{ppi} \end{bmatrix}.$$

The stationary point $\mathbf{x}^*$ can be written as a function of β; it is defined as

$$\mathbf{x}^* = -\ 0.5\ \mathbf{B}^{-1}\mathbf{b}$$

For simplifying the computation, Carter et al (1984 and 1986) transform the confidence region about β to a multi-dimensional spherical region of radius

$$r = (qs^2F_{\alpha,\ q,\ n-q})^{1/2}.$$

The transformation is done by defining

$$\begin{aligned} (\hat{\beta}-\beta)'(\mathbf{X}'\mathbf{X})(\hat{\beta}-\beta) &= (\hat{\beta}-\beta)'\mathbf{PP}'(\mathbf{X}'\mathbf{X})\mathbf{PP}'(\hat{\beta}-\beta) \\ &= (\hat{\beta}-\beta)'\mathbf{P}\Lambda^{1/2}\Lambda^{1/2}\mathbf{P}'(\hat{\beta}-\beta) \\ &= \mathbf{z}'\mathbf{z}. \end{aligned} \tag{2.1}$$

Where $\mathbf{z}' = (\hat{\beta}-\beta)'\mathbf{P}\Lambda^{1/2}$

$\mathbf{P}$ = matrix of eigenvectors of $\mathbf{X}'\mathbf{X}$,

Λ = diag(eigenvalues of $\mathbf{X}'\mathbf{X}$),

$\mathbf{z} = (z_1,\ z_2,\ \ldots,\ z_q)$,

$z_1 = \rho \cos \theta_1,\ 0 \le \rho \le r$,

$z_2 = \rho \cos \theta_1 \sin \theta_2$,

$$\vdots$$

$$z_{q-1} = \rho \cos\theta_1 \cos\theta_2 \ \ldots\ldots \cos\theta_{q-2} \sin\theta_{q-1},$$

$$z_q = \rho \cos\theta_1 \cos\theta_2 \ \ldots\ldots\ldots \cos\theta_{q-2} \cos\theta_{q-1},$$

$$-0.5\,\pi \le \theta_i \le 0.5\,\pi,\ i = 1, 2, \ldots, q-2, \text{ and}$$

$$-\pi \le \theta_{q-1} \le \pi.$$

Then, every point of $\beta \in U$ can be defined. By evaluating any $g(\beta)$, we can construct the confidence region for $g(\beta)$.

The nature of the stationary point is determined by the eigenvalues of matrix **B**. If all the eigenvalues of **B** are positive, then the response has a minimum point. If all negative, then the response has a maximum point. If the eigenvalues have different signs, then the response has a saddle point. If one of the eigenvalues is very small or zero, then the response may not have a unique stationary point, but rather a stationary ridge. Therefore, information about the eigenvalues of **B** is useful for characterization of the stationary point.

Since λ, the eigenvalues of **B**, are computed from a random matrix **B**, then λ are also random variables. Carter et al. (1990) proposed a procedure to construct the confidence region for λ, the eigenvalues of **B**. They give an approximate $100(1-\alpha)\%$ confidence region for λ as

$$\hat{C}_\lambda = \left\{\lambda \in R^p : (\hat{\lambda} - \lambda)'(\hat{\mathbf{H}}'\mathbf{V}\hat{\mathbf{H}})^{-1}(\hat{\lambda} - \lambda) \le \chi^2_{p,\,1-\alpha}\right\}, \qquad (2.22)$$

where λ = the vector of eigenvalues of **B**,

$\mathbf{V}$ = p x p matrix of variance-covariance of vec($\hat{\mathbf{B}}$),

$$\mathbf{H}' = \begin{bmatrix} \text{vec}'(2\mathbf{d}_1\mathbf{d}_1' - \text{diag}(\mathbf{d}_1\mathbf{d}_1')) \\ \vdots \\ \text{vec}'(2\mathbf{d}_p\mathbf{d}_p - \text{diag}(\mathbf{d}_p\mathbf{d}_p')) \end{bmatrix}, \text{ and}$$

$\mathbf{d}_i$(i = 1, 2, .., p) = the i^{th} eigenvector of $\mathbf{B}$.

The confidence interval for the i^{th} ordered eigenvalue of $\mathbf{B}$ is given by

$$\hat{\lambda}_{(i)} - Z_{1-\alpha/2} [\ \mathbf{e}_i'\ \hat{\mathbf{H}}'\mathbf{V}\hat{\mathbf{H}}\ \mathbf{e}_i]^{0.5} \leq \lambda_{(i)}$$

$$\leq \hat{\lambda}_{(i)} + Z_{1-\alpha/2} [\ \mathbf{e}_i'\ \hat{\mathbf{H}}'\mathbf{V}\hat{\mathbf{H}}\ \mathbf{e}_i]^{0.5}, \qquad (2.23)$$

where $\lambda_{(1)} < \lambda_{(2)} < \ldots < \lambda_{(p)}$, and $\mathbf{e}_i$ is a p x 1 vector of zeros with a 1 in the i^{th} row. For a small sample size, they recommend using $F_{1-\alpha,p,n-q}$, instead of $\chi^2_{p,1-\alpha}$, where q = (p+1)(p+2)/2.

CHAPTER III

ADMISSIBLE SETS

Characterization of Sets of Admissible Points

When considering problems as a multivariate one, it is obvious that, in general, there can be no single best solution for all individual responses. Still, some solutions are definitely better than others. We are drawn naturally to consider the sets of points for which there are no "better" points. "Admissible points" and "admissible set" are defined in Definition II.7 and Definition II.8 in Chapter II, respectively.

"Admissible set" is a terminology standard in much of statistical decision theory. In mathematical economics the term "pareto optimal set" is used, for example in models of welfare economics where no consumer can be made "better off" without making another consumer "worse off." Mockus and Mockus (1991) also used the term "pareto optimal set" in multiobjective optimization using the Bayesian approach. Another term for admissible points is "efficient points," used by Karlin (1959). Folks and Antle (1965) also used the term "efficient points" in arriving at an optimum allocation of units to strata when there were multiple responses.

We shall now describe the sets of admissible points in a case where we have k response variables, Y_i, i = 1, 2, ..., k as a function of p control variables, X_1, ..., X_p. Characterizations are developed for both unconstrained and constrained optimization, also for several cases of surfaces of $Y_i(x)$. Geometric visualizations are shown for k = 2, 3 and p = 2.

Unconstrained Optimization

<u>p = 2 and k = 2</u>. Let $Y_i(x)$, i = 1, 2 and $x \in E^2$ be second-order polynomial functions of x. We want to maximize both Y_1 and Y_2. The tangent path of Y_1 and Y_2 seems like a natural candidate as the admissible set. The part of the tangent path of Y_1 and Y_2 that is admissible depends on the nature of the surfaces of Y_1 and Y_2, described in Theorem III.1.

The equation of the tangent path can be derived from the function of Y_1 and Y_2. Let $Y_1 = a_0 + a'x + x'\mathbf{A}x$, and $Y_2 = b_0 + b'x + x'\mathbf{B}x$. Let x^* belong to the tangent path of $Y_1(x)$ and $Y_2(x)$. Then $\exists$ a number λ, such that

$$\nabla Y_1(x^*) + \lambda \nabla Y_2(x^*) = 0 \;, \quad \Rightarrow \tag{3.1}$$

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + 2\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} +$$

$$\lambda\left[\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + 2\begin{bmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{bmatrix}\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}\right] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Then the equation of the tangent path of Y_1 and Y_2 is given by

$$\frac{a_1+2a_{11}x_1+2a_{12}x_2}{a_2+2a_{12}x_1+2a_{22}x_2} = -\frac{b_1+2b_{11}x_1+2b_{12}x_2}{b_2+2b_{12}x_1+2b_{22}x_2}. \quad (3.2)$$

Since Y(x) is a quadratic form, if Y(x) has a maximum or a minimum point, then its contours are close. The set bounded by the closed contour is a closed convex set.

Theorem III.1. Let $Y_1(x)$ and $Y_2(x)$ be second-order polynomial functions. Let $x \in E^2$. The admissible region for $Y_1(x)$ and $Y_2(x)$ is a part of their tangent path, described as follows:

(1). If both $Y_i(x)$ have maximum points, then the admissible region is the part of the tangent path between the stationary points of $Y_1(x)$ and $Y_2(x)$.

(2). If $Y_1(x)$ has a maximum point and $Y_2(x)$ has a minimum point, then the admissible region is the part of the tangent path of $Y_1(x)$ and $Y_2(x)$ from the maximum point of $Y_1(x)$ to infinity, when $Y_1(x)$ lies on the right sides of $Y_2(x)$.

(3). If $Y_1(x)$ has a maximum point and $Y_2(x)$ has a saddle point then the admissible region is the part of the tangent path from the maximum point of $Y_1(x)$ to infinity.

(4). If both $Y_1(x)$ and $Y_2(x)$ have saddle points, then the admissible region may or may not exist. If it exists, it is the tangent path that does not pass through their stationary points.

(5). If $Y_1(x)$ has a minimum point and $Y_2(x)$ has a saddle point, then the admissible region is an empty set.

(6). If both $Y_1(x)$ and $Y_2(x)$ have minimum points, then the admissible region is an empty set.

Proof. Suppose x^* is not admissible for $Y_i(x)$; then there exists x^o such that

$$Y_i(x^o) \geq Y_i(x^*), \ \forall \ i \text{ and}$$

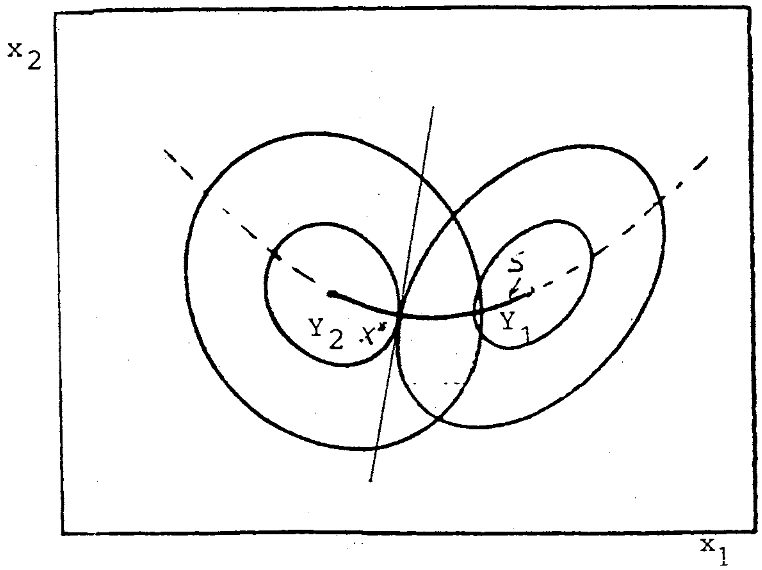
$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i. \tag{3.3}$$

First, we consider case (1). From Figure 4a, let a part of the tangent path of Y_1 and Y_2 between their maximum points be the set S. Take any $x^* \in S$. At x^*, $\nabla Y_1(x^*)$ and $\nabla Y_2(x^*)$ are perpendicular to the tangent line of their contours. Every other point x^o that lies on the left side of or on the tangent line gives value $Y_1(x^o) < Y_1(x^*)$. Every other point that lies on the right side of or on the tangent line gives value $Y_2(x^o) < Y_2(x^*)$. So, if we move from x^* to any x^o, then

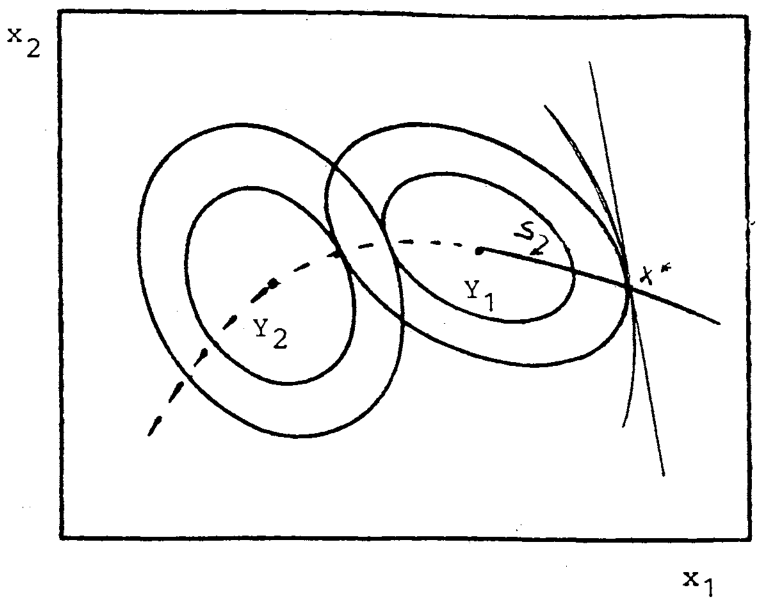
$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i,$$

which is a contradiction.

For case (2), suppose Y_1 has a maximum point and Y_2 has a minimum point. Let a part of the tangent path of Y_1 and Y_2 from the maximum point of Y_1 to infinity be the set S, as shown in Figure 4b. Every other point x^o that lies outside the contour of Y_1 drawn from x^* gives value $Y_1(x^o) < Y_1(x^*)$. Every other point x^o that lies inside the contour of Y_2

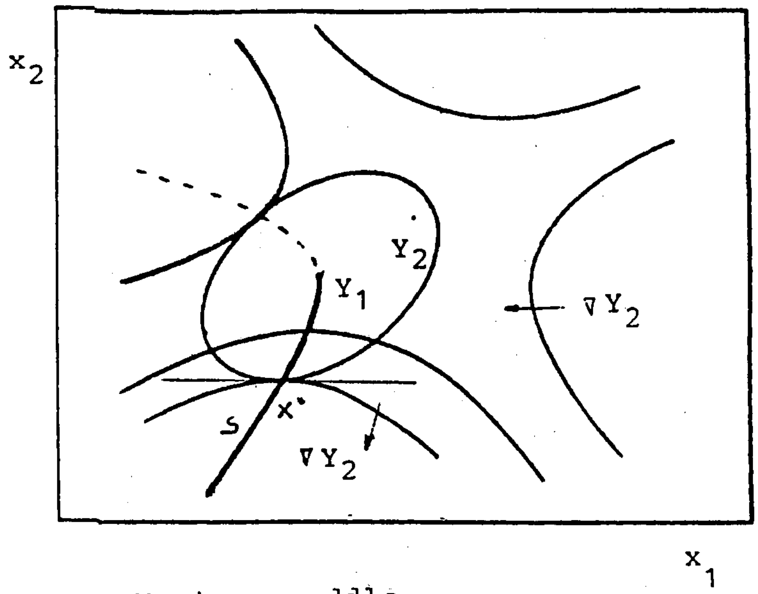


a. Maximum-maximum

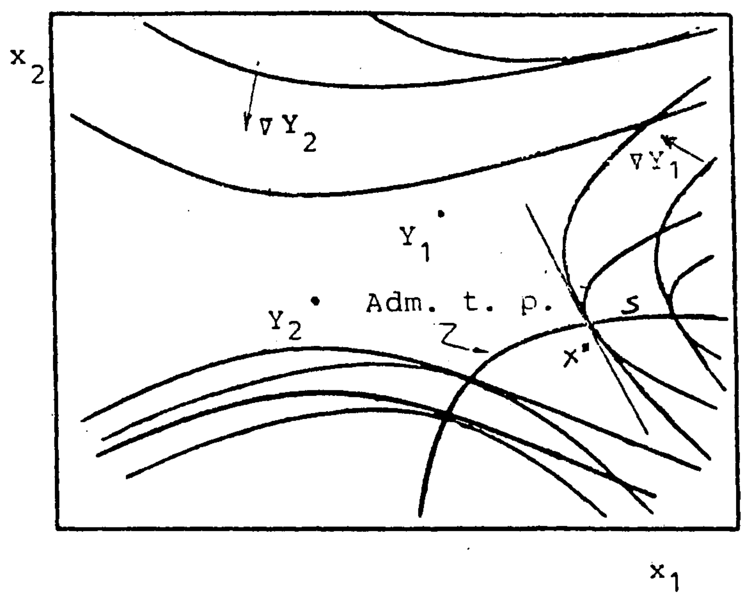


b. Maximum-minimum

Figure 4. Admissible Tangent Path

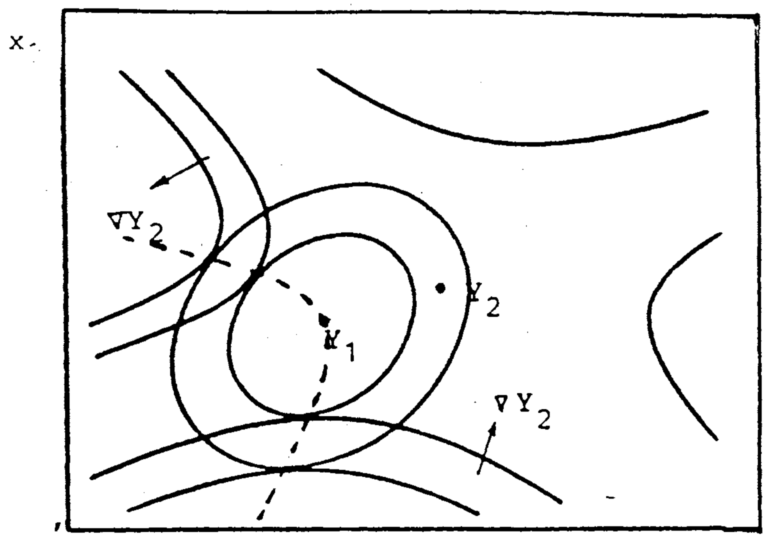


c. Maximum-saddle

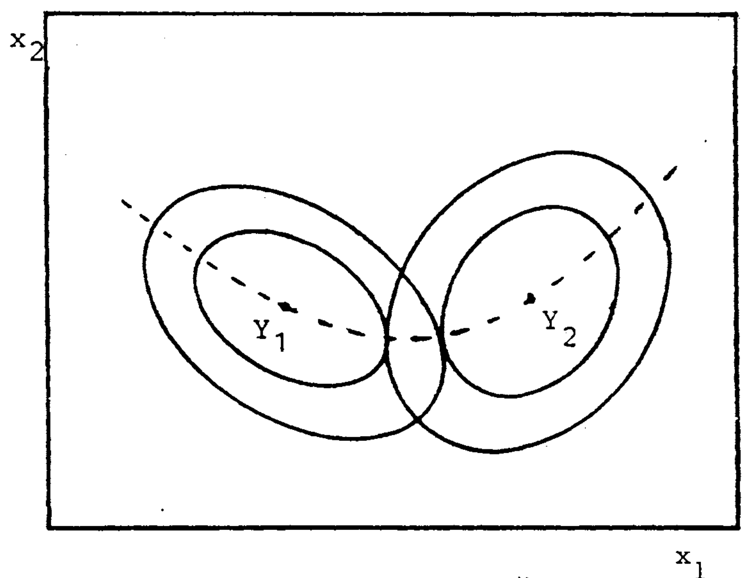


d. Saddle-saddle

Figure 4. (Continued)



e. Minimum-saddle (S = ∅)



f. Minimum-minimum (S = ∅)

Figure 4. (continued)

drawn passing through x^* gives value $Y_2(x^o) < Y_2(x^*)$. So, if we move from x^* to any x^o, then

$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i,$$

which contradicts Equation (3.3). By the same way as case (1) or as case (2), we can show for the other cases that if we move from x^* to any x^o, then

$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i,$$

which contradicts Equation (3.3).

$\underline{p \geq 2 \text{ and } k = 2}$. Let $Y_i(x)$, $i = 1, 2$ and $x \in E^p$, be second-order polynomial functions. For $p = n$, the contours of $Y_1(x)$ and $Y_2(x)$ lie in n-dimensional space; therefore, their tangent path also lies in n-dimensional space. So, Theorem III.1 can be extended for $p = 3, 4, \ldots\ldots., n$. Admissible regions when $p = 3$ and $k = 2$, for two kinds of surface combinations, maximum-maximum and maximum-saddle, are shown in Figure 5.

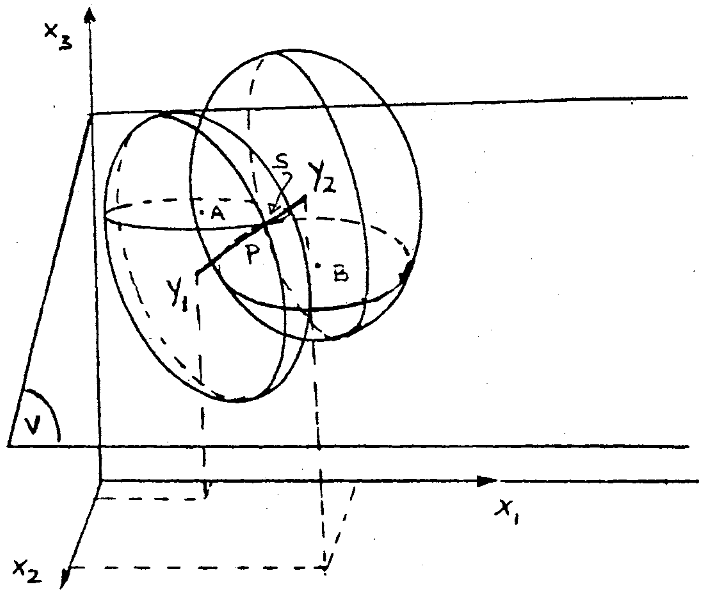
For determining the equation of the tangent path of $Y_1(x)$ and $Y_2(x)$, $x \in E^p$, we recall Equation (3.1). From that equation, then

$$(\mathbf{a} + 2\mathbf{A}\mathbf{x}^*) + \lambda(\mathbf{b} + 2\mathbf{B}\mathbf{x}^*) = 0 \text{ or}$$

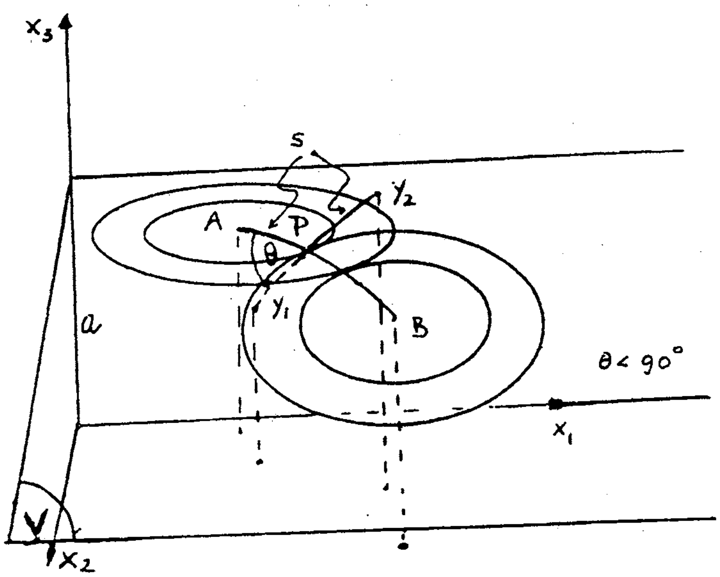
$$2(\mathbf{A} + \lambda\mathbf{B})\mathbf{x}^* = -(\mathbf{a} + \lambda\mathbf{b}).$$

Then, the general equation of the tangent path of Y_1 and Y_2 is given by

$$\mathbf{x} = -0.5(\mathbf{A} + \lambda\mathbf{B})^{-1}(\mathbf{a} + \lambda\mathbf{b}), \qquad (3.4)$$

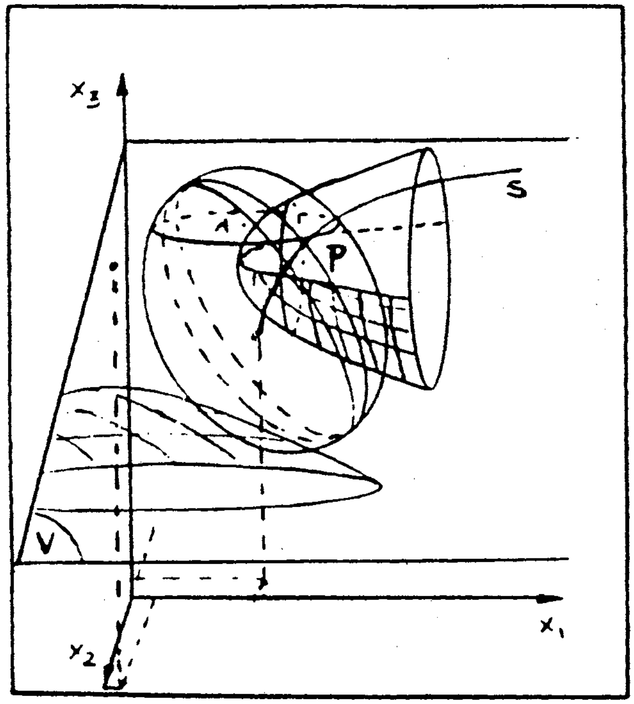


a. Maximum-maximum in 3-dimensional Space

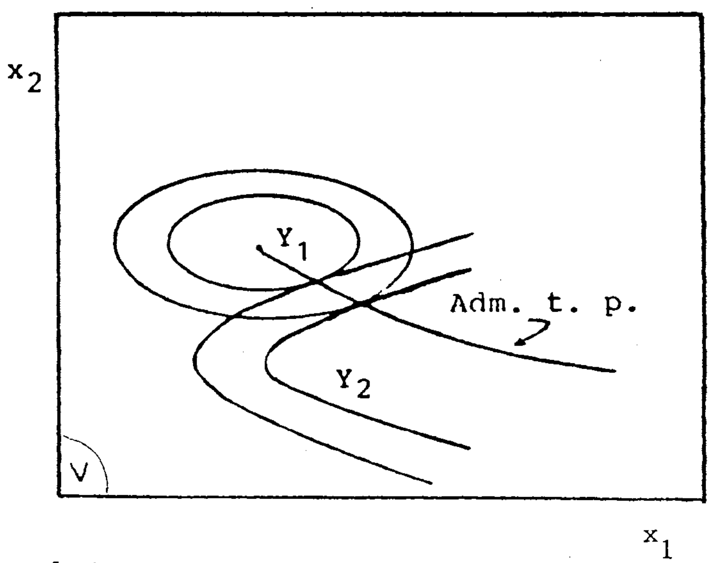


b. Maximum-maximum in 3-dimensional Space with Constraint $x_3 > a$

Figure 5. Admissible Region for $Y_1(\mathbf{x})$ and $Y_2(\mathbf{x})$ in 3-dimensional Space



c. Maximum-saddle in 3-Dimensional Space



d. Maximum-saddle with Constraint $x_3 = a$

Figure 5. (Continued)

for some λ and $x \in E^p$.

<u>$p = 2$ and $k = 3$</u>. Here we have three responses and three pairs of tangent paths. We are only interested in the admissible part of these tangent paths for characterizing the admissible region for $Y_i(x)$. The following theorem characterizes the admissible region.

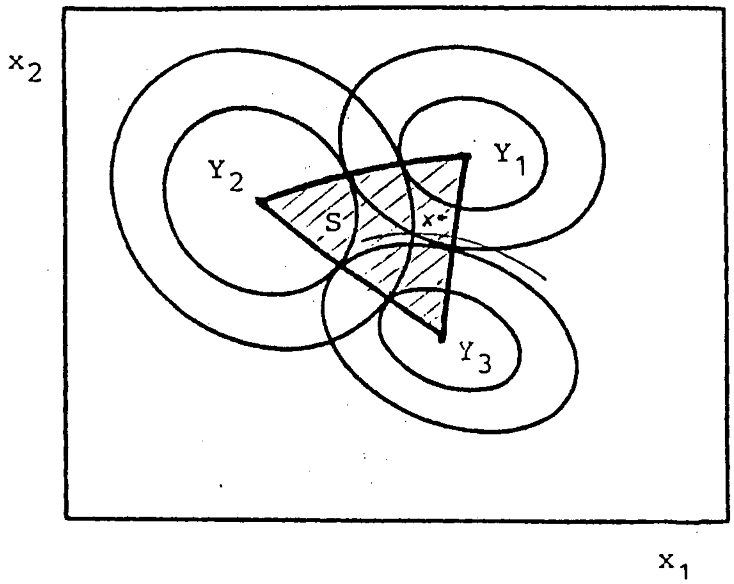
Theorem III.2. Let $Y_i(x)$, $i = 1, 2, 3$ and $x \in E^2$, be second-order polynomial functions of x. If at least two pairs of $Y_i(x)$ and $Y_{i'}(x)$ have admissible tangent paths, then the admissible region for $Y_i(x)$ is the closed region bounded by the admissible tangent path of each pair of $Y_i(x)$ and $Y_{i'}(x)$, $i < i'$.

Proof. When $p = 2$ and $k = 2$, a pair of $Y_i(x)$ and $Y_{i'}(x)$ will have an admissible tangent path if one of $Y_i(x)$ has a maximum point and the other has either a maximum, a minimum, or a saddle point. Let the closed region bounded by the admissible tangent paths of $Y_i(x)$ and $Y_{i'}(x)$, $i < i'$, be the set S. Take any $x^* \in S$. Suppose $x^* \in S$ is not admissible; then $\exists\, x^o$ such that

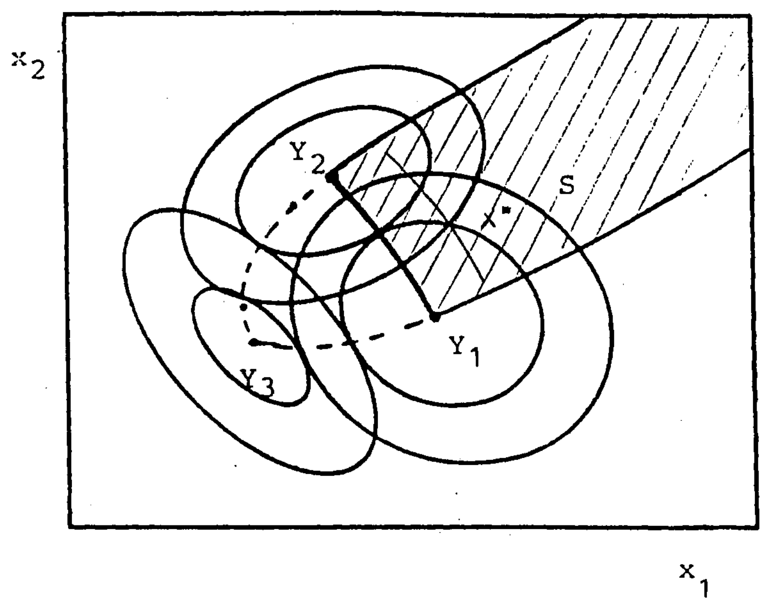
$$Y_i(x^o) \geq Y_i(x^*),\ \forall\ i \text{ and}$$

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i.$$

First we consider that all $Y_i(x)$ have maximum points, as shown in Figure 6a. The contours of $Y_i(x)$ are drawn passing through x^*. Every other point x that lies outside the

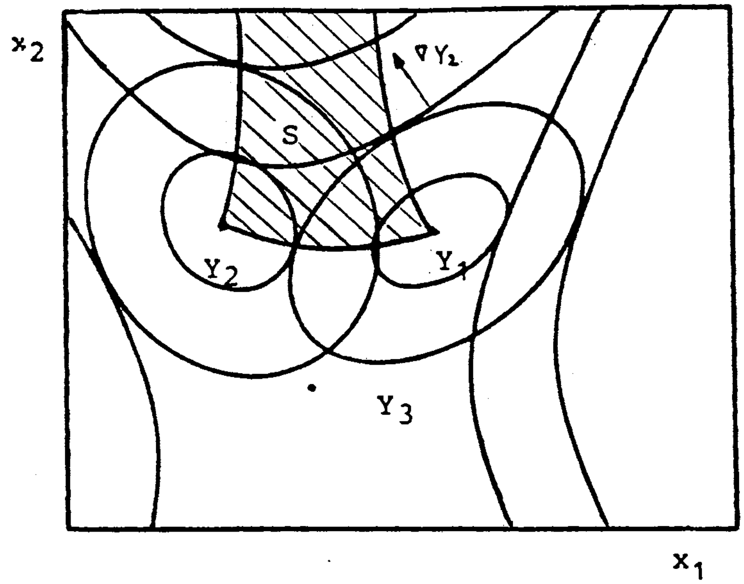


a. Maximum-maximum-maximum

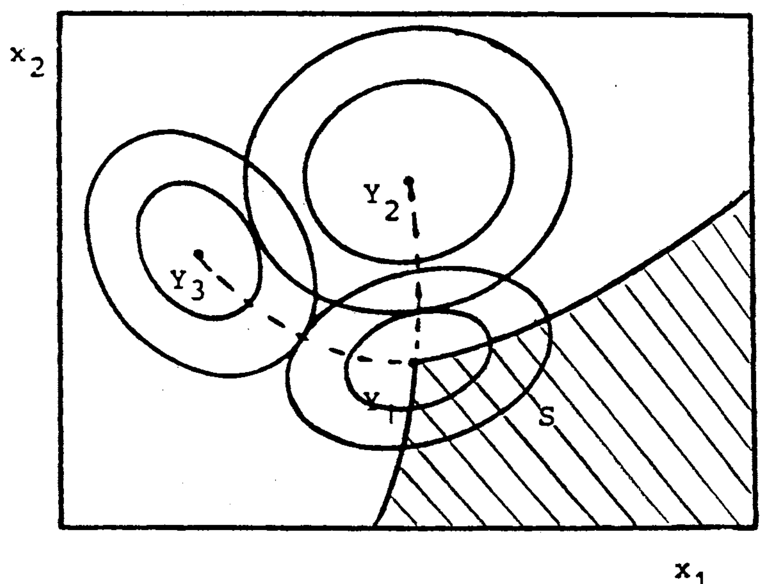


b. Maximum-maximum-minimum

Figure 6. Admissible Region for Three Responses

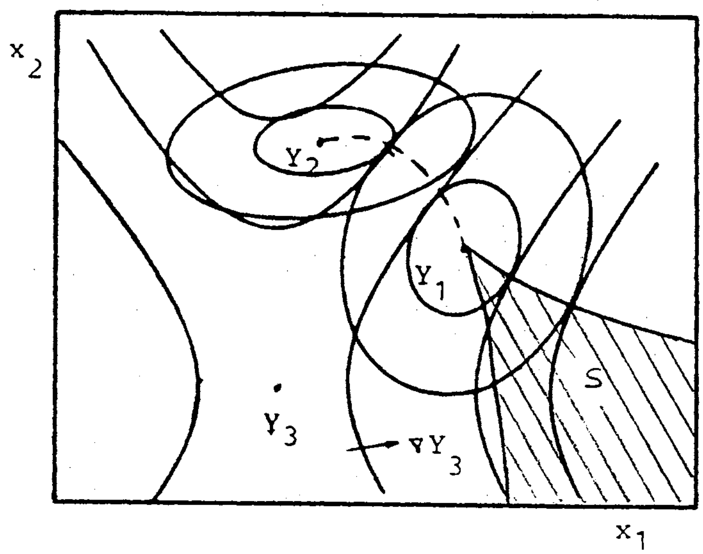


c. Maximum-maximum-saddle

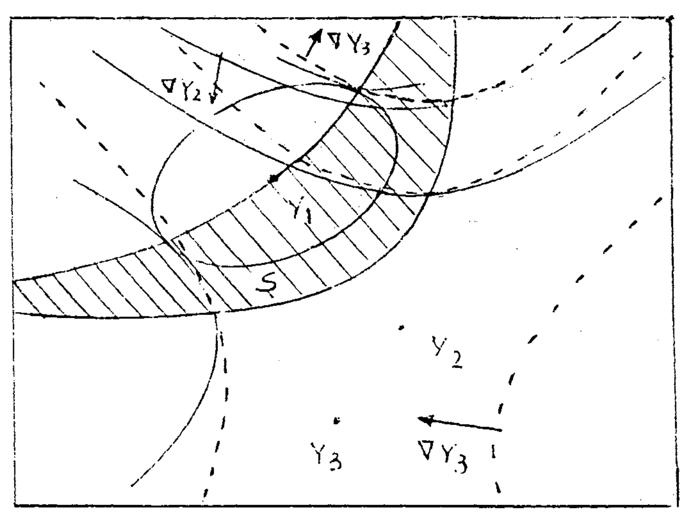


d. Maximum-minimum-minimum

Figure 6. (Continued)



e. Maximum-minimum-saddle



f. Maximum-saddle-saddle

Figure 6. (Continued)

contours of $Y_i(x)$ gives value $Y_i(x) < Y_i(x^*)$. Let T_i be the set of all x that lie outside the contour of Y_i. Then we can see that $x^o \neq x^*$ belongs to $\bigcup_i T_i$. It implies that $x^o \in T_i$ for at least one i. So, if we move to any other point x^o, then

$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i,$$

which is a contradiction. By the same way, we can show for the other combinations of surfaces that, if we move from x^* to other point x^o, then

$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i,$$

which is a contradiction.

<u>$p = 2$ and $k \geq 3$</u>. For more than 3 responses, we can look at a group of 3 distinct responses at a time. Then the admissible region can be figured out easily. It is described in the following theorem.

Theorem III.3. Suppose there are k responses $Y_i(x)$. Let S_j be the admissible region for every group of three responses. Then the admissible region for all $Y_i(x)$, i = 1, 2, .., k is defined by

$$S = \bigcup_{j=1}^{m} S_j, \text{ where } m = \frac{k!}{(k-3)!3!}. \tag{3.5}$$

Proof. Take any point $x^* \in S$. Suppose x^* is not admissible for $Y_i(x)$, i = 1, ..., k; then there exists x^o, such that

$$Y_i(x^o) \geq Y_i(x^*), \text{ for } \forall i \text{ and}$$

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i.$$

Since $x^* \in S$, then x^* belongs to at least one $S_j \neq \emptyset$. It implies that

$$Y_i(x^*) > Y_i(x^o), \text{ for at least one } i,$$

which is a contradiction.

$\underline{p \geq 2}$ and $\underline{k \geq 2}$. In general cases when x belongs to p-dimensional space, $p > 3$, it is impossible to characterize the admissible set geometrically. One possibility to characterize the admissible set is by algebraic notation or equations, stated in the following theorem.

Theorem III.4. Let $Y_i(x)$, $i = 1, 2, .., k$ and $x \in E^p$ be second-order polynomial functions of x. Take any $\alpha_i > 0$. If $S = \left\{x^*: x^* \text{ satisfies } \nabla\left[\sum_{i=1}^{k}\alpha_i Y_i(x^*)\right] = 0 \text{ and } \sum_{i=1}^{k}\alpha_i A_i \text{ is a negative definite matrix}\right\}$, then S is admissible region for all $Y_i(x)$. A_i is the Hessian matrix of $Y_i(x)$, $i = 1, .., k$.

Proof: Since x^* satisfies $\nabla\left[\sum_{i=1}^{k}\alpha_i Y_i(x^*)\right] = 0$ and $\sum_{i=1}^{k}\alpha_i A_i$ is a negative definite matrix, then x^* is a maximum point of $\sum_{i=1}^{k}\alpha_i Y_i(x)$. Suppose $x^* \in S$ is not admissible; then $\exists\, x^o$ such that

$$Y_i(x^o) \geq Y_i(x^*), \ \forall i \text{ and}$$

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i.$$

Since $\alpha_i > 0$, then

$$\sum_{i=1}^{k} \alpha_i Y_i(x^*) < \sum_{i=1}^{k} \alpha_i Y_i(x^o) \tag{3.6}$$

which contradicts the fact that x^* is a maximum point of $\sum_{i=1}^{k} \alpha_i Y_i(x)$. Then, by Definition II.8, the set S is an admissible region for all $Y_i(x)$.

Corollary III.1. If $\sum_{i=1}^{k} \alpha_i Y_i(x)$ attains its global maximum value at x^*, then $x^* \in S$.

Constrained Optimization

Let $Y_i(x)$, i = 1, 2, .., k and $x \in E^p$ be second-order polynomial functions of x. Let the feasible region R be a nonempty, closed, convex set bounded by $g_n(x) = 0$, n = 1, 2,..., m. Let S be an admissible region for $Y_i(x)$ for unconstrained optimization. Then two cases arise; $R \cap S$ is either an empty set or a nonempty set.

p = 2 and k = 2

$R \cap S$ is a Nonempty Set. Before we develop theorems concerning admissible sets, we need to present two definitions and four lemmas. As mentioned in Chapter II, if at least of $Y_i(x)$ has a saddle point, then it may happen that the subset of the admissible region lies in the interior of R, but outside $R \cap S$. Definition III.1 and III.2 will help the description of such a subset.

Definition III.1. Let one of $Y_i(\mathbf{x})$ have a saddle point. Let none of $Y_i(\mathbf{x})$ have a minimum point. A pseudo-admissible tangent path is a part of the inadmissible tangent path, for which $\nabla Y_i(\mathbf{x})$ and $\nabla Y_{i'}(\mathbf{x})$, $i \neq i'$, have the opposite direction.

Definition III.2. Let $Y_i(\mathbf{x})$ have a saddle point. A pair of opposite contours is a pair of contours of $Y_i(\mathbf{x})$ that have the same value but lie on different sides of the stationary point of $Y_i(\mathbf{x})$.

Since $Y(\mathbf{x})$ has a saddle point, its contours are open. A contour of $Y(\mathbf{x})$ is drawn passing through $\mathbf{x}^*$. If the direction of $\nabla Y(\mathbf{x}^*)$ does not go toward the stationary point of $Y(\mathbf{x})$, then any $\mathbf{x}^o$ that lies in between a contour and its opposite contour gives value

$$Y(\mathbf{x}^o) < Y(\mathbf{x}^*) \tag{3.7}$$

Part of the boundary of R is also a candidate for a subset of the admissible region. The following three lemmas will describe it. All $Y_i(\mathbf{x})$ in the following lemmas and theorems are second-order polynomial functions of $\mathbf{x}$.

Lemma III.1. Let $Y_i(\mathbf{x})$ have the highest value at $\tilde{\mathbf{x}}_i \in$ boundary of R, $g_n(\mathbf{x}) = 0$, $i = 1, \ldots, k$ and $n = 1, 2, .., m$. Suppose $Y_i(\mathbf{x})$ does not have a tangent point on $\mathbf{x}_{oi} \in g_n(\mathbf{x}) = 0$, where $\mathbf{x}_{oi} \neq \tilde{\mathbf{x}}_i$. Let the boundary of R that contains $\tilde{\mathbf{x}}_i$ intersect S at point T_i. Let the part of the boundary of R between $\tilde{\mathbf{x}}_i$ and T_i be the set A_i. Then A_i is a subset of the feasible admissible region for $Y_i(x)$.

Proof. From Figure 7.a, Suppose $x^* \in A_i$ is not admissible for $Y_i(x)$;then there exists $x^o \in R$ such that

$$Y_i(x^o) \geq Y_i(x^*),\ \forall\ i \text{ and}$$

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i.$$

First, we consider that both $Y_i(x)$ have maximum points. Since $\tilde{x}_i$ is the maximum point of $Y_i(x)$, then $Y_i(x^o) \leq Y_i(\tilde{x}_i)$. If $x^* = \tilde{x}_i$, it is obvious that $Y_i(x^o) \leq Y_i(x^*)$. If $x^* \neq \tilde{x}$, look at the contours of $Y_1(x)$ and $Y_2(x)$ drawn passing through x^*. Let the set of all points that lie outside the contour of $Y_1(x)$ be T_1 and the set of all points that lie outside the contour of $Y_2(x)$ be T_2. For every $x \in (T_1 \cup T_2)$ then

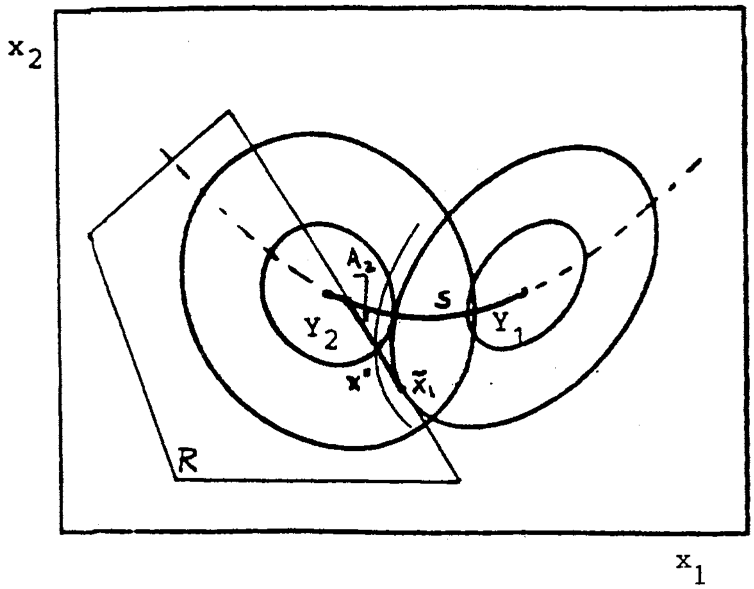
$$Y_1(x) < Y_1(x^*) \text{ or } Y_2(x) < Y_2(x^*) \text{ or both.}$$

We can see that $(R-x^*) \subset (T_1 \cup T_2)$; if we move from x^* to any $x^o \in (R-x^*)$ then $x^o \in (T_1 \cup T_2)$; it implies that

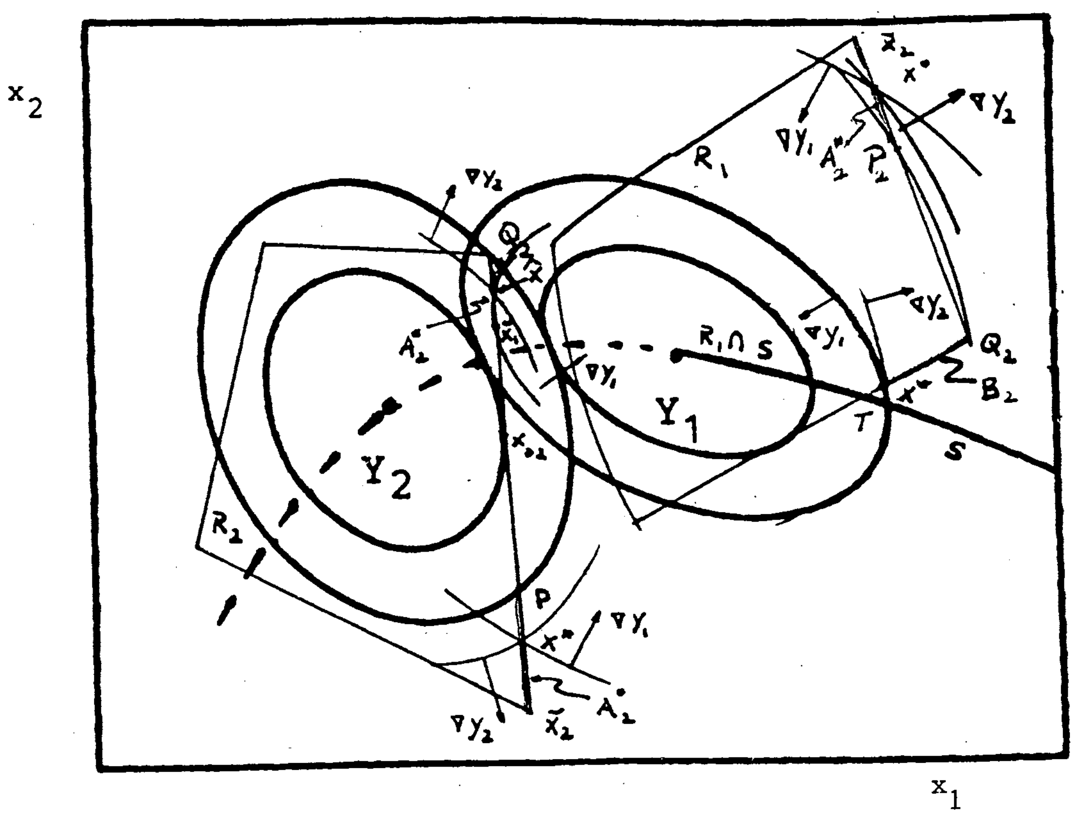
$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i,$$

which is a contradiction. By the same way we can prove this theorem for the other combinations of surfaces.

Lemma III.2. Let $Y_i(x)$, $i = 1, 2$, have the highest value at $\tilde{x}_i \in$ boundary of R, $g_n(x) = 0$ and $\tilde{x}_i \notin S$. Let this $Y_i(x)$ have a tangent point $x_{oi} \in g_n(x) = 0$, where $x_{oi} \neq \tilde{x}_i$. Let $g_n(x) = 0$ which contains this point be the smooth curve

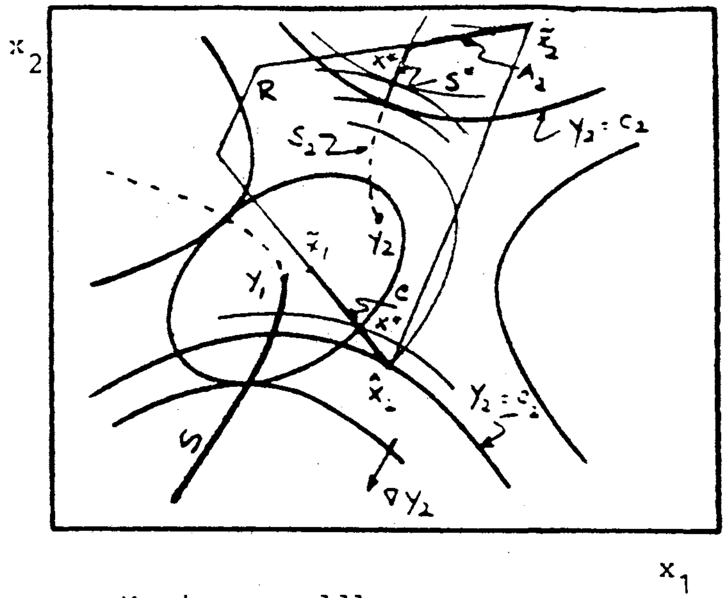


a. Maximum-maximum

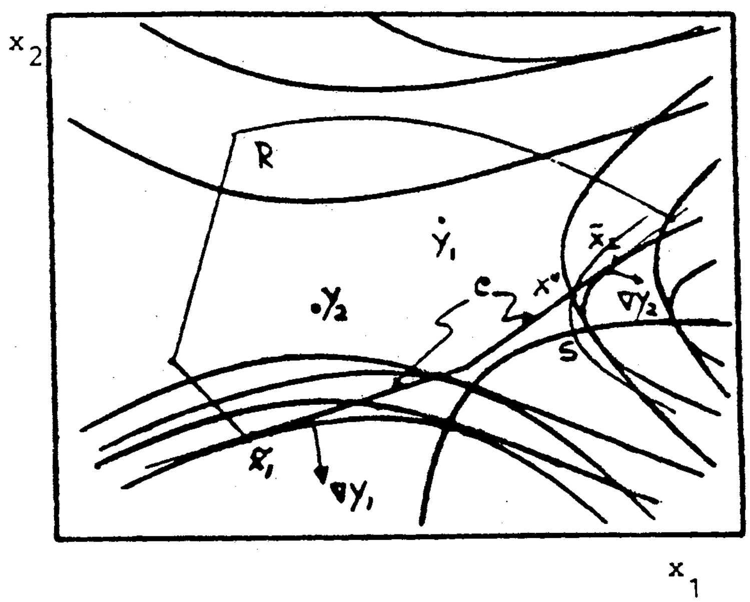


b. Maximum-minimum

Figure 7. Admissible Region for two Responses with Constraints

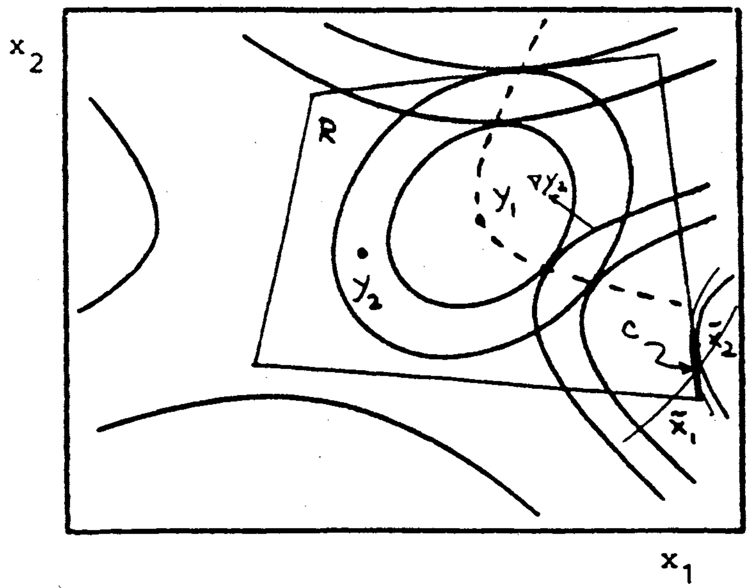


c. Maximum-saddle

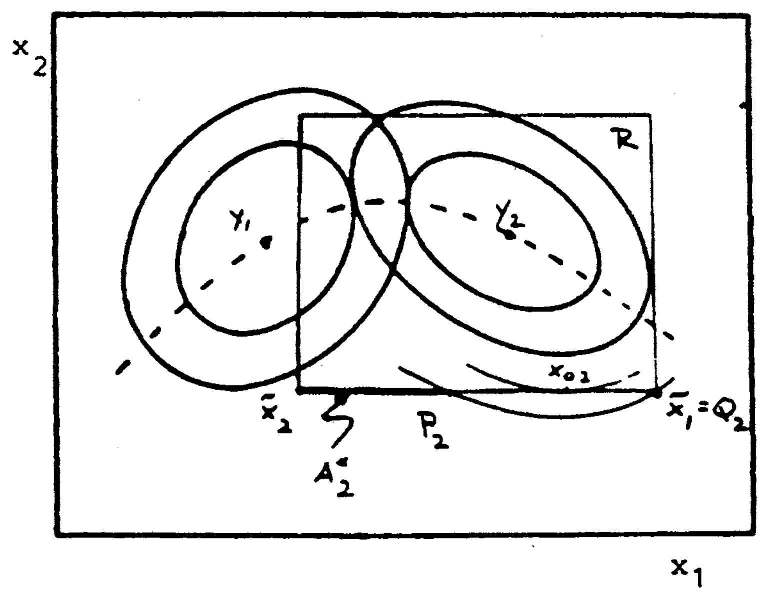


d. Saddle-saddle

Figure 7. (Continued)



e. Minimum-saddle



f. Minimum-minimum

Figure 7. (Continued)

$Q_i - \tilde{x}_i$. Let the contour of $Y_i(x)$ that passes through Q_i intersect this curve at a point P_i. Let set A_j^* be defined as:

(1). The curve P_i-$\tilde{x}_i$ (excluding the point P_i), if the curve Q_i-$\tilde{x}_i$ does not contain another $\tilde{x}_{i'}$, $i \neq i'$, and Q_i is closer to S than $\tilde{x}_i$ is.

(2). The curve P_i-$\tilde{x}_i$ unions the curve Q_i-$\tilde{x}_{i'}$, if the curve Q_i-$\tilde{x}_i$ contains another $\tilde{x}_{i'}$ and x_{oi} lies between $\tilde{x}_i$ and $\tilde{x}_{i'}$, $i \neq i'$.

(3). The curve $\tilde{x}_i$-$\tilde{x}_{i'}$, if the curve Q_i-$\tilde{x}_i$ contains another $\tilde{x}_{i'}$ and x_{oi} lies between Q_i and $\tilde{x}_{i'}$, $i \neq i'$.

Then the set A_j^* is a subset of the feasible admissible region for all $Y_i(x)$.

Proof. Suppose $x^* \in R_1$ or $x^* \in R_2$ is not admissible for $Y_i(x)$; then there exists x^o such that

$$Y_i(x^o) \geq Y_i(x^*), \ \forall \ i \text{ and}$$

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i.$$

From Figure 7b, take any point $x^* \in$ curve P_i-$\tilde{x}_i$ (excluding point P_i), then $Y_2(x^*) > Y_2(Q_i)$. If $x^o = Q_i$ or $x^* = \tilde{x}_2$, it is obvious that $Y_2(x^o) < Y_2(x^*)$. If $x^o \neq Q_i$ and $x^* \neq \tilde{x}_i$, then look at the contours of $Y_1(x)$ and $Y_2(x)$ drawn passing through x^*. Let the set of all points that lie outside the contour of $Y_1(x)$ be T_1 and the set of all points that lie inside the contour of $Y_2(x)$ be T_2 (since $Y_2(x)$ has a minimum point). For every $x \in (T_1 \cup T_2)$ then

$$Y_1(x) < Y_1(x^*) \text{ or } Y_2(x) < Y_2(x^*) \text{ or both.}$$

We can see that $(R-x^*) \subset (T_1 \cup T_2)$; so that if we move from x^* to any $x^o \in (R-x^*)$, then

$$Y_1(x^o) < Y_1(x^*) \text{ or } Y_2(x^o) < Y_2(x^*)$$

which is a contradiction. Look at R_2 in Figure 7b; it is shown that x_{o2} lies between $\tilde{x}_1$ and $\tilde{x}_2$. Take any $\hat{x} \in$ curve $Q_i-\tilde{x}_{i'}$. The contour of $Y_1(x)$ and $Y_2(x)$ are drawn passing through $\hat{x}$. By the same way as above, for any $x^o \in R$ and $x^o \neq \hat{x}$, then

$$Y_1(x^o) < Y_1(\hat{x}) \text{ or } Y_2(x^o) < Y_2(\hat{x}) \text{ or both}$$

which is a contradiction.

Lemma III.3. Let any boundary of R intersect the admissible tangent path of $Y_i(x)$ and $Y_{i'}(x)$, $i \neq i'$, at point T. Let the part of this boundary that does not lie between $S \cap R$ and both contours of $Y_i(x)$ and $Y_{i'}(x)$, $i \neq i'$, which are drawn through T, be the set B_i. Then B_i is a subset of the feasible admissible region for $Y_i(x)$.

Proof. Take any point $x^* \in B_i$. Suppose x^* is not admissible for $Y_i(x)$; then there exists $x^{oo} \in R$ such that

$$Y_i(x^{oo}) \geq Y_i(x^*), \ \forall\ i = 1, 2 \text{ and}$$

$$Y_i(x^{oo}) > Y_i(x^*), \text{ for at least one } i.$$

Since B_i does not lie between $S \cap R$ and both contours of $Y_1(x)$ and $Y_2(x)$, then B_i and $(S \cap R)$ lie on different sides of the tangent line drawn from T. Take any $\mathbf{x}^*$ ∈ B_i. Suppose B_i lies on the right side, as shown in Figure 7b. The contours of $Y_1(x)$ and $Y_2(x)$ are drawn passing through x^*. Let the set of all points that lie outside the contour of $Y_1(x)$ be T_1 and the set of all points that lie inside the contour of $Y_2(x)$ be T_2 (since $Y_2(x)$ has a minimum point). For every $x \in (T_1 \cup T_2)$ then

$$Y_1(\mathbf{x}) < Y_1(\mathbf{x}^*) \text{ or } Y_2(\mathbf{x}) < Y_2(\mathbf{x}^*) \text{ or both.}$$

We can see that $(R-\mathbf{x}^*) \subset (T_1 \cup T_2)$; so that if we move from x^* to any $x^{oo} \in (R-x^*)$, then

$$Y_1(x^{oo}) < Y_1(x^*) \text{ or } Y_2(x^{oo}) < Y_2(x^*) \text{ or both,}$$

which is a contradiction.

The following lemma describes the subset of the admissible region in the interior of R, but outside $R \cap S$. This kind of subset may exist if at least one $Y_i(\mathbf{x})$ has a saddle point.

Lemma III.4. Let at least one of $Y_i(\mathbf{x})$ have a saddle point. Let none of Y_i have a minimum point. Let the region bounded by pseudo-admissible tangent paths be the set S_2. Let $Y_i(\mathbf{x})$ have the highest value on R at $\tilde{\mathbf{x}}_i$. Every contour of $Y_i(\mathbf{x})$ that passes through S_2 has its opposite contour on the other side of the stationary point. Let the highest value of these opposite contours that intersect R at $\hat{\mathbf{x}}_i$ be c_i. Let the set of the contours that pass through S_2 of the boundary of R between $\tilde{\mathbf{x}}_{i'}$ and $\hat{\mathbf{x}}_{i'}$, $i \neq i'$, be a set C.

If the contour of $Y_{i'}(x)$ drawn passing through $\hat{x}_i$ does not intersect S_3, then set C is also a subset of the feasible admissible region.

Proof. First we want to prove that S^* is admissible. Take any point $x^* \in S^*$, as shown in Figure 7c. Suppose x^* is not admissible for $Y_i(x)$; then $\exists\ x^o$ such that

$$Y_i(x^o) \geq Y_i(x^*),\ \forall\ i \text{ and}$$

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i.$$

By the same way as the proof in Lemma III.1, the contours of $Y_1(x)$ and $Y_2(x)$ are drawn passing through x^*. Every point x^o that lies outside the contour of $Y_1(x)$ or in between the contour of $Y_2(x)$ and its opposite contour (because $Y_2(x)$ has a saddle point then x^o satisfies Inequality (3.7)), gives value

$$Y_1(x^o) < Y_1(x^*) \text{ or } Y_2(x^o) < Y_2(x^*) \text{ or both}$$

which is a contradiction.

Secondly, we want to prove that C is admissible. Take any $x^* \in C$. Since $\hat{x}_2$ is the maximum point of $Y_2(x)$ on $(R \cap S_3^c)$, it is obvious that $Y_2(x)^o) < Y_2(\hat{x})$, for all $x^o \in (R \cap S_3^c)$. If $x^* \neq \hat{x}_2$, then look at the contour of $Y_1(x)$ and $Y_2(x)$ drawn passing through x^*. Let the set of all points that lie outside the contour of $Y_1(x)$ be T_1 and the set of all points that lie in between the contour of $Y_2(x)$ and its opposite contour be T_2 (since $Y_2(x)$ has a saddle point). By the same way as the above proof, every $x \in (T_1 \cup T_2)$ satisfies

$$Y_1(x) < Y_1(x^*) \text{ or } Y_2(x) < Y_2(x^*) \text{ or both.}$$

We can see that $(R-x^*) \subset (T_1 \cup T_2)$; so that if we move from x^* to any $x^o \in R$, then

$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i,$$

which is a contradiction.

Note: Instead of S, $S_2 \cap S_3$, can be applied to Lemma III.1, III.2, III.3 and III.5 for defining whether a part of the boundary of R is a subset of an admissible region.

Now we can develop a theorem, based on those lemmas for characterizing the admissible region for $Y_i(x)$.

Theorem III.5. Let S be an admissible region for unconstrained optimization. Let the feasible region R be a closed convex set. Let A_i, A_j^*, B_i, and S^* be defined as those in Lemma III.1, III.2, III.3, and III.4. Then the feasible admissible region for $Y_i(x)$ is the union of $(S \cap R)$, all A_i, all A_j^*, all B_i, and S^*. Illustrations are shown in Figure 7a and 7b.

<u>$S \cap R$ is an Empty Set</u>. The following definition defines 4 kinds of boundaries for feasible region R, which will be used in the next lemma and theorem.

Definition III.3. The upper, lower, right, and left boundaries of R are described sequentially as follows:

(1). The upper part of the boundary of R, from A, B, to C, shown in Figure 8a.

(2). The lower part of the boundary of R, from A, B, C, to D, shown in Figure 8b.

(3). The right part of the boundary of R from A, B, C, to D, shown in Figure 8c.

(4). The left part of the boundary of R from A, B, C, to D, shown in Figure 8d.

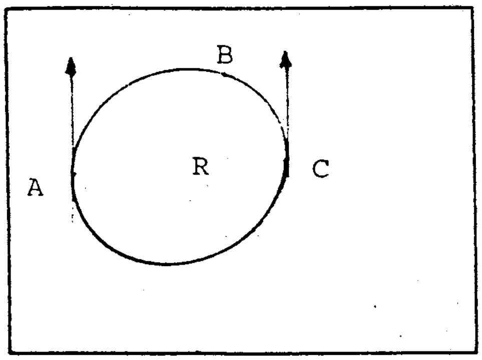
Lemma III.5. Let $Y_i(x)$ have the highest value at $\tilde{x}_i \in$ boundary of R, $g_n(x)=0$. Let S^*, as defined in Lemma III.4, be an empty set. Let none of $Y_i(x)$ have a tangent point at $x^*_i \in g_n(x)=0$, where $x^*_i \neq \tilde{x}_i$. If R lies below, above, on the left side, or on the right side of S, then the subset of the feasible admissible region for $Y_i(x)$ is a part of the upper, lower, right, or left boundaries of R from $\tilde{x}_{1'}$, $\tilde{x}_{2'}$, ..., $\tilde{x}_{k'}$, respectively, where 1', 2', ... , k' are permutations of 1, 2, ..., k. Let this set be denoted as set C.

Proof. From Figure 7d, take any point $x^* \in$ C. Suppose x^* is not admissible for $Y_i(x)$, then $\exists$ x^o such that

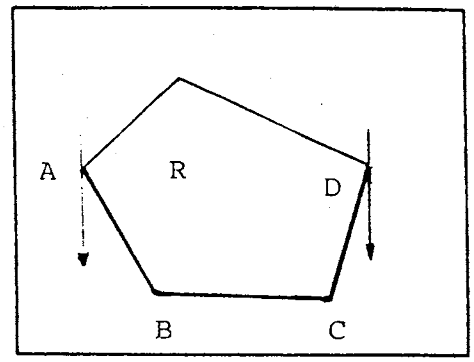
$$Y_i(x^o) \geq Y_i(x^*),\ \forall\ i \text{ and}$$

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i.$$

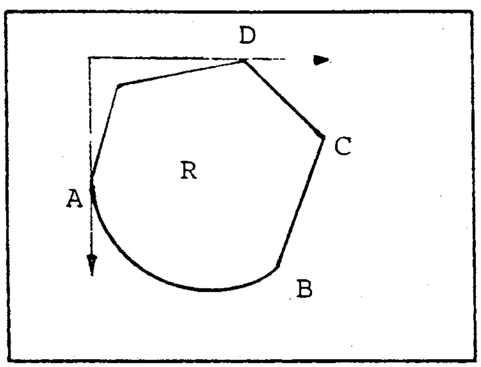
By the same way as the proofs in previous lemmas, if we move from x^* to any x^o, then



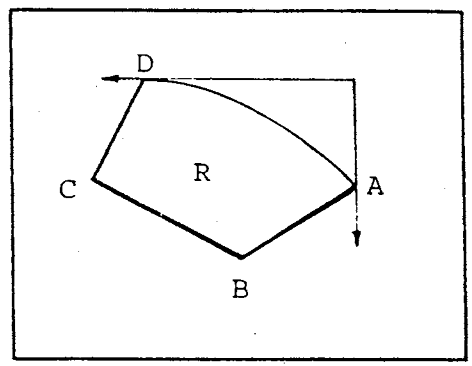
a. Upper Boundary



b. Lower Boundary



c. Right Boundary



d. Left Boundary

Figure 8. Four Kinds of the Boundaries of R

$$Y_i(x^o) < Y_i(x^*), \text{ for at least one } i$$

which is a contradiction.

The following theorem uses the previous lemmas for characterizing the admissible region for constrained optimization. The proof is obvious, so that it is not included here.

Theorem III.6. Let S be an admissible region for unconstrained optimization. Let a feasible region R be a closed convex set. Let $R \cap S$ be an empty set. Let C, A_i, A_j^*, and S^* be defined as those in Lemma III.5, III.1, III.2, and III.4. Then, the feasible admissible region for $Y_i(x)$ is the union of C, all A_i, all A_j^*, and S^*. Illustrations are given in Figure 7 (c through f).

<u>If $p = 2$ and $k \geq 3$</u>

The following lemma shows the property of the union of admissible sets. The purpose of this lemma is to explain the characterization of the admissible set for multiple responses (more than 2). The characterization is developed by expanding the previous theorems that are available for two or three responses.

Lemma III.6. Let set A be the admissible region for Y_1 and Y_2. Let set B be the admissible region for Y_2 and Y_3. Then $A \cup B$ is a subset of the admissible region for Y_1, Y_2, and Y_3.

Proof. Take any point $x^o \in A \cup B$. Suppose x^o is not admissible for $Y_i(x)$, i = 1, 2, 3; then there exists x^*, such that

$$Y_i(x^*) \geq Y_i(x^o), \text{ for } i = 1, 2, 3 \text{ and}$$

$$Y_i(x^*) > Y_i(x^o), \text{ for at least one } i.$$

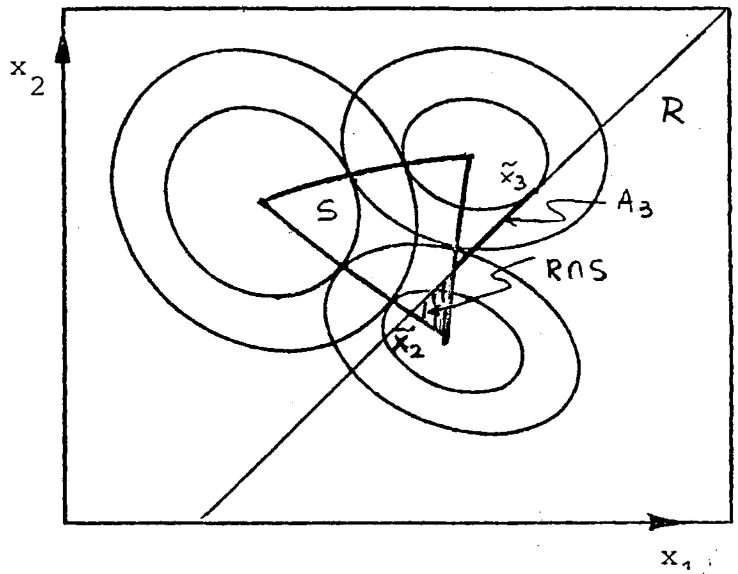
Since $x^o \in A \cup B$, then x^o belongs to either A or B, or both. It implies that

$$Y_i(x^o) > Y_i(x^*), \text{ for at least one } i,$$

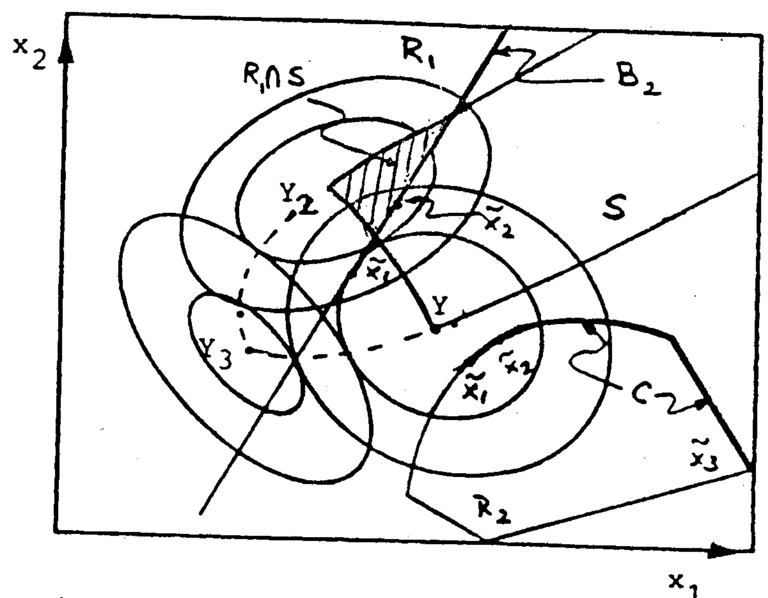
which is a contradiction.

<u>$S \cap R$ is a Nonempty Set</u>. By using the previous lemmas, the following theorem can be developed. Since the proof is obvious by using Lemma III.6, we shall not include it here. This theorem characterizes the admissible region for constrained optimization if $R \cap S$ in a nonempty set.

Theorem III.7. Let S be an admissible region for unconstrained optimization. Let the feasible region R be a closed convex set. Let sets A_i, A_j^*, B_i and S^* be defined as those in Lemma III.1, III.2, III.3, and III.4. Then the feasible admissible region for $Y_i(x)$ is the union of $S \cap R$, all A_i, all A_j^*, all B_i, and S^*. Illustrations are given in Figure 9.

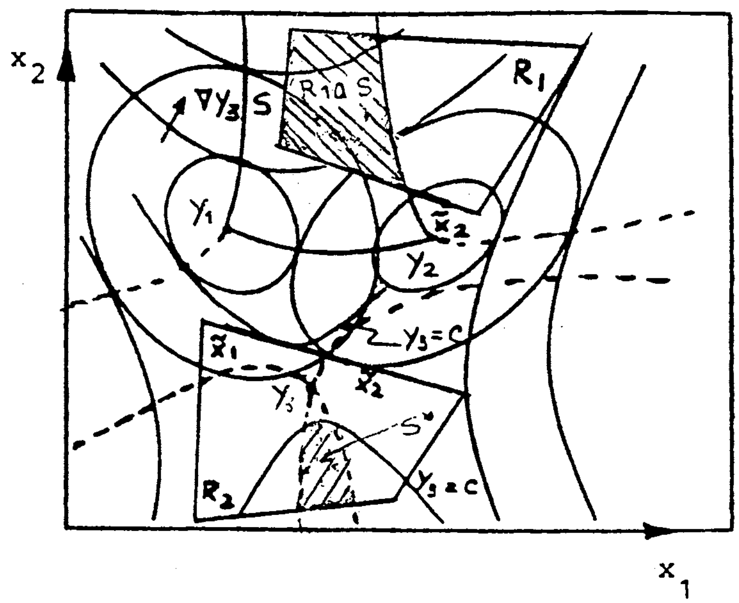


a. Maximum-maximum-maximum

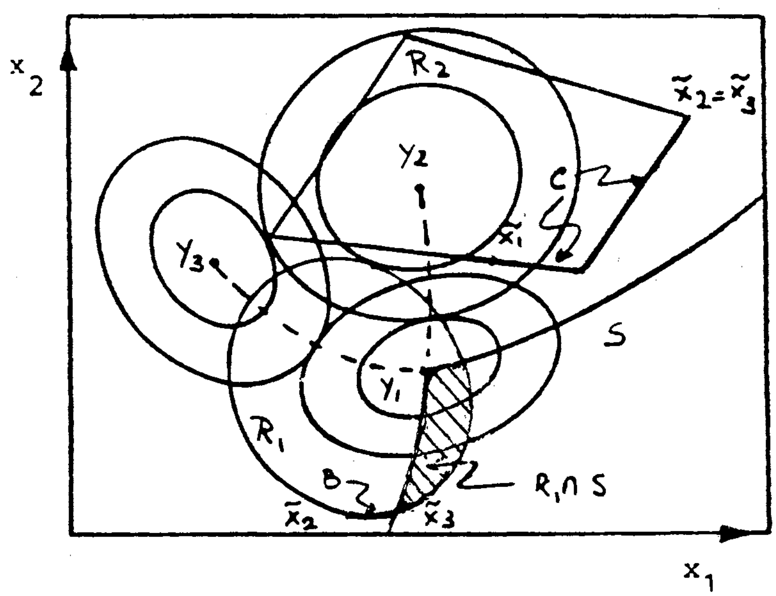


b. Maximum-maximum-minimum

Figure 9. Admissible Region for Three Responses with Constraints

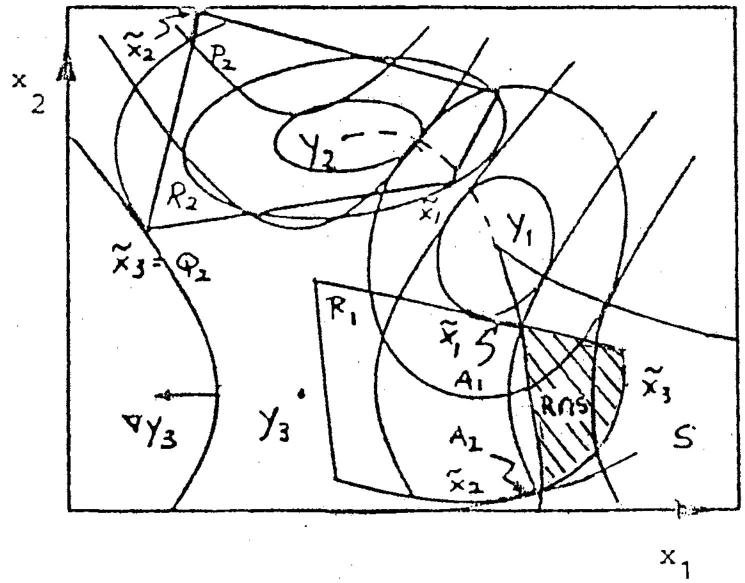


c. Maximum-maximum-saddle

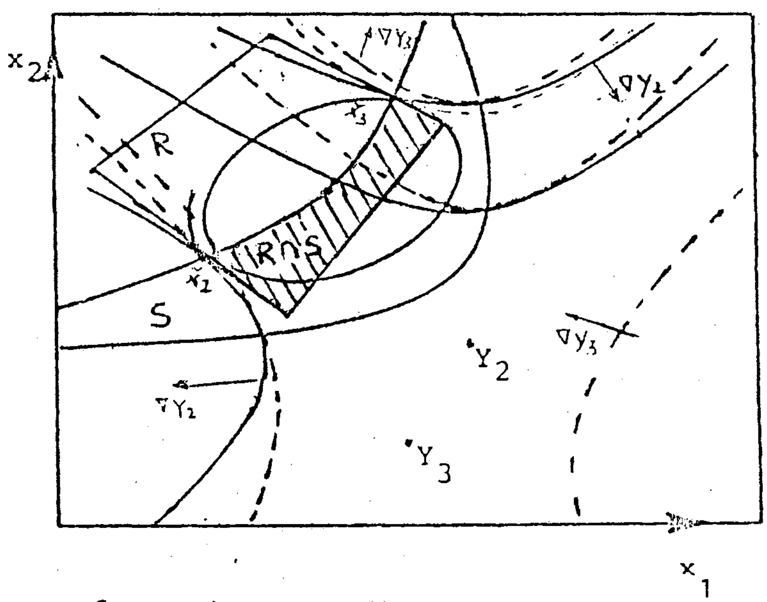


d. Maximum-minimum-minimum

Figure 9. (Continued)



e. Maximum-minimum-saddle



f. Maximum-saddle-saddle

Figure 9. (Continued)

<u>$S \cap R$ is an Empty Set</u>. By using previous lemmas, the following theorem can be developed. It characterizes the admissible region if $S \cap R$ is an empty set. Since the proof is obvious by using Lemma III.6, we shall not include it here. Illustrations are given in Figure 9.

Theorem III.8. Let S be an admissible region for unconstrained optimization. Let the feasible region R be a closed convex set. Let $R \cap S$ be an empty set. Let C, A_i, A_j^*, and S^* be the sets as defined in Lemma III.5, III.1, III.2, and III.4, respectively. Then, the feasible admissible region for $Y_i(x)$ is the union of C, all A_i, all A_j^*, and S^*.

<u>$p \geq 2$ and $k \geq 2$</u>

Since there are more than 2 responses and x belongs to p-dimensional space, in general, characterization of sets of admissible points cannot be shown geometrically. We can characterize them by algebraic notation or equations. The characterization is given in the following theorem.

Theorem III.9. Let $Y_i(x)$, i = 1, ..., k and $x \in E^p$ be second-order polynomial functions of x. Suppose the feasible region R is a compact set. Then the set of all the maximum points of $\sum_{i=1}^{k} \alpha_i Y_i(x)$, for every possible $\alpha_i > 0$, is a feasible admissible region for all $Y_i(x)$.

Proof. Take any $\alpha_i > 0$. Since R is a compact set, we can find a point $x^* \in R$, such that $\sum_{i=1}^{k}\alpha_i Y_i(x^*) \geq \sum_{i=1}^{k}\alpha_i Y_i(x^o)$, for all $x^o \in R$. By Theorem III.4, then the set of all x^* is an admissible region for all $Y_i(x)$.

For a special case that $R = \left\{x:\ \sum_{j=1}^{p} x_j^2 \leq r^2\right\}$, then the admissible region for all $Y_i(x)$ is

$$S_1 = \left\{x_c^*:\ x_c^* \text{ satisfies } \nabla\left[\sum_{i=1}^{k}\alpha_i Y_i(x^*_c)-uI\right] = 0 \text{ and } \left[\sum_{i=1}^{k}\alpha_i A_i - uI\right] \text{ is a negative definite matrix}\right\},$$

where u is a Lagrangian multiplier that satisfies the constraints; I is an identity matrix.

The Existence of Admissible Sets

Necessary Conditions

The following theorem describes the condition of α_i, for the global maximum point of a combined response to be an admissible point. The necessary condition is the first requirement for maximizing a combined response that leads to an admissible point.

Theorem III.10. Let $Y_i(x)$, i = 1, 2, .., k, be second-order polynomial functions of x. Then the global maximum of $\sum_{i=1}^{k}\alpha_i Y_i(x)$ at x^* is an admissible point for $Y_i(x)$ if and only if $\alpha_i > 0$, for all i.

Proof. If there exists a global maximum point of $\sum_{i=1}^{k}\alpha_i Y_i(x)$ at x^* in $R \subset E^p$, then

$$\sum_{i=1}^{k}\alpha_i Y_i(x^*) \geq \sum_{i=1}^{k}\alpha_i Y_i(x_o),\ \forall x_o \in R$$

First, we want to prove the "only if" part. Suppose x^* is not admissible for $Y_i(x)$; then there exists x_o such that

$$Y_i(x_o) \geq Y_i(x^*),\ \forall i \text{ and}$$

$$Y_i(x_o) > Y_i(x^*),\ \text{for at least one } i.$$

Since $\alpha_i > 0$, then

$$\sum_{i=1}^{k}\alpha_i Y_i(x_o) > \sum_{i=1}^{k}\alpha_i Y_i(x^*),$$

which is a contradiction.

Then we want to prove the "if" part. x^* is an admissible point for $Y_i(x)$. Let us consider any point $x_o \neq x^*$; then either

(1). $Y_i(x_o) \leq Y_i(x^*)$, $\forall i$ or

(2). $Y_i(x_o) \geq Y_i(x^*)$, for some i and

$Y_{i'}(x_o) < Y_{i'}(x^*)$, for $i \neq i'$.

Suppose $\alpha_i < 0$, for some i. For (1), it may happen that

$$\sum_{=1}^{k}\alpha_i Y_i(x_o) > \sum_{i=1}^{k}\alpha_i Y_i(x^*),$$

which is a contradiction. For (2), if the corresponding $Y_i(x)$ has $\alpha_i > 0$ and the corresponding $Y_{i'}$ has $\alpha_{i'} < 0$, then

$$\sum_{i=1}^{k}\alpha_i Y_i(x_o) > \sum_{i=1}^{k}\alpha_i Y_i(x^*),$$

which is a contradiction.

Suppose $\alpha_i \geq 0$, $\forall i$, and $\alpha_i = 0$, for at least one i. For (2), if the corresponding $Y_i(x)$ has $\alpha_i > 0$ and the corresponding $Y_{i'}(x)$ has $\alpha_{i'} = 0$, then

$$\sum_{i=1}^{k}\alpha_i Y_i(x_o) > \sum_{i=1}^{k}\alpha_i Y_i(x^*),$$

which is a contradiction.

From the above theorem, if some $\alpha_i \leq 0$, then the maximization of $\sum_{i=1}^{k}\alpha_i Y_i(x)$ **does not always lead to** an admissible point. Thus, we can indicate that the **necessary condition** under which maximizing $\sum_{i=1}^{k}\alpha_i Y_i(x)$ always leads to an admissible point is

$$\alpha_i > 0, \text{ for all } i.$$

Necessary and Sufficient Conditions

The necessary and sufficient conditions under which optimizing a combined response leads to an admissible point will be considered for two cases, unconstrained and constrained optimization. The solution for maximizing $\sum_{i=1}^{k}\alpha_i Y_i(x)$ will lead to an admissible point if $\exists\ \alpha_i > 0$, $\forall i$, and $\exists\ x^*$ on $R \subset E^p$, a global maximum point of $\sum_{i=1}^{k}\alpha_i Y_i(x)$. If both requirements are satisfied, then the solution of maximizing a combined response guaranty to be an admissible point. Since the second requirement that x^* be a global maximum point, cannot be known before we perform the maximization, we need to simplify such a requirement.

Unconstrained Optimization. A global maximum point of a combined response exists if its Hessian matrix is negative definite. The following theorem connects the existence of $\alpha_i > 0$, $\forall i$, with the existence of the admissible region for all $Y_i(\mathbf{x})$.

Theorem III.11. Let $Y_i(\mathbf{x})$, $i = 1, 2, \ldots, k$, be second-order polynomial functions of $(\mathbf{x})$. If the admissible region of $Y_i(\mathbf{x})$ for unconstrained optimization exists, then there exist $\alpha_i > 0$ for all i, such that $\sum_{i=1}^{k}\alpha_i Y_i(\mathbf{x})$ has a maximum value at its stationary point.

Proof. Let $\mathbf{x}^* \in S$. Let us consider any point $\mathbf{x}_o \neq \mathbf{x}^*$; then either

(1). $Y_i(\mathbf{x}_o) \leq Y_i(\mathbf{x}^*)$, $\forall i$ or

(2). $Y_i(\mathbf{x}_o) \geq Y_i(\mathbf{x}^*)$, for some i and

$Y_{i'}(\mathbf{x}_o) < Y_{i'}(\mathbf{x}^*)$, for $i \neq i'$.

For (1), it is obvious that for every $\alpha_i > 0$, then

$$\sum_{i=1}^{k}\alpha_i Y_i(\mathbf{x}_o) \leq \sum_{i=1}^{k}\alpha_i Y_i(\mathbf{x}^*).$$

For (2), we can find a set of numbers $\alpha_i > 0$, for all i, such that

$$\sum_{i=1}^{k}\alpha_i Y_i(\mathbf{x}_o) \leq \sum_{i=1}^{k}\alpha_i Y_i(\mathbf{x}^*).$$

Since $Y_i(\mathbf{x})$ is second order polynomial functions, so is $\sum_{i=1}^{k}\alpha_i Y_i(\mathbf{x})$. Then $\mathbf{x}^*$ is a maximum point and also a stationary point of $\sum_{i=1}^{k}\alpha_i Y_i(\mathbf{x})$.

Moreover, we need to combine the conditions for the existence of the solution for maximizing a combined response with the conditions for the solution to be an admissible point for all $Y_i(x)$. We also need to connect the conditions with the surface of $Y_i(x)$.

$Y_i(x)$ can be written as $Y_i(x) = a_o + a_i'x + x'A_ix$, where $a_i' = (a_{1i}, a_{2i}, \ldots, a_{pi})$ and

$$A_i = \begin{bmatrix} a_{11i} & a_{12i}/2 & & a_{1pi}/2 \\ & a_{22i} & & a_{2pi}/2 \\ \text{symetric} & & & \\ & & & a_{ppi} \end{bmatrix}.$$

Then $\sum_{i=1}^{k}\alpha_i Y_i(x)$ can be written as

$$\sum_{i=1}^{k}\alpha_i Y_i(x) = b_o + b'x + x'Bx$$

where $b_o = \sum_{i=1}^{k}\alpha_i a_{oi}$,

$$b = \left(\sum_{i=1}^{k}\alpha_i a_{1i}, \sum_{i=1}^{k}\alpha_i a_{2i}, \ldots\ldots \sum_{i=1}^{k}\alpha_i a_{pi}\right)' \text{ and}$$

$$B = \left(\sum_{i=1}^{k}\alpha_i A_i\right).$$

The solution for maximizing $\sum_{i=1}^{k}\alpha_i Y_i(x)$ is given by

$$x^* = -0.5\left(\sum_{i=1}^{k}\alpha_i A_i\right)^{-1}\left(\sum_{i=1}^{k}\alpha_i a_{1i}, \ldots\ldots, \sum_{i=1}^{k}\alpha_i a_{pi}\right)'. \qquad (3.8)$$

The solution for maximization exists and is unique if $\sum_{i=1}^{k}\alpha_i A_i$ is a negative definite matrix.

By Theorem III.11, if S exists, then there exist $\alpha_i >$ 0, for all i, such that $\sum_{i=1}^{k} \alpha_i Y_i(x)$ has a maximum value at its stationary point x^*. Since x^* is a global maximum point of $\sum_{i=1}^{k} \alpha_i Y_i(x)$ and admissible for $Y_i(x)$; then $(\sum_{i=1}^{k} \alpha_i A_i)$ is a negative definite matrix. By Theorem III.1, the admissible region (S) exists if at least one of the Hessian matrices of $Y_i(x)$ is negative definite. Therefore, we can conclude that for unconstrained optimization, the necessary and sufficient conditions under which maximizing $\sum_{i=1}^{k} \alpha_i Y_i(x)$ leads to an admissible point are

$$\alpha_i > 0, \ \forall i, \text{ and } \mathbf{H}_{Y_i} \text{ is negative definite,} \qquad (3.9)$$

for at least one $\mathbf{H}_{Y_i}$.

<u>Constrained</u> <u>Optimization</u>. Let R be a compact set (closed and bounded). Let $\alpha_i > 0$, $\forall$i. From Theorem III.9, $x^* \in$ R, the maximum point of $\sum_{i=1}^{k} \alpha_i Y_i(x)$, belongs to the admissible region for $Y_i(x)$. It implies that if R is a compact set, then the necessary and sufficient conditions are $\alpha_i > 0$. However, if R is a closed convex set, but not bounded, then we need additional conditions that are described in the following theorem.

Theorem III.12. Let a feasible region R be a closed convex set. Let $\sum_{i=1}^{k} \alpha_i Y_i(x)$, $\alpha_i > 0$ for all i, be a continuous function on R. If at least one of the Hessian matrices of $Y_i(x)$ is negative definite, then there exist x^* $\in$ R, the maximum point of $\sum_{i=1}^{k} \alpha_i Y_i(x)$, which is admissible for $Y_i(x)$.

Proof. Let S be the admissible region for unconstrained optimization. By Theorem III.1, if at least one $Y_i(x)$ has a maximum value at its stationary point, then S exists. Then, by Theorem III.11, there exist $\alpha_i > 0$, such that $\sum_{i=1}^{k} \alpha_i Y_i(x)$ has a maximum point at $x^* \in S$. If we move farther from x^*, the value of $\sum_{i=1}^{k} \alpha_i Y_i(x)$ will decrease. There arises two cases: $R \cap S = \emptyset$ and $R \cap S \neq \emptyset$.

Let $R \cap S = \emptyset$. Suppose there does not exist a maximum point of $\sum_{i=1}^{k} \alpha_i Y_i(x)$, $\tilde{x} \in R$; then $\exists\, x^0 \in R$ such that

$$\sum_{i=1}^{k} \alpha_i Y_i(x^0) \longrightarrow \infty ,$$

which contradicts the fact that $x^* < \infty$ is a global maximum point of $\sum_{i=1}^{k} \alpha_i Y_i(x)$. Then we want to show that $\tilde{x} \in$ boundary of R. Maximizing $\sum_{i=1}^{k} \alpha_i Y_i(x)$ with constraints $g_j(x) \leq 0$, $j = 1, 2, \ldots, n$, is given by

$$\max_{x} f(x,\lambda,\theta) = \sum_{i=1}^{k} \alpha_i Y_i(x) - \sum_{j=1}^{n} \lambda_j (g_j(x) - \theta_j^2)$$

$$\partial f/\partial x = \partial \sum_{i=1}^{k} \alpha_i Y_i(x)/\partial x - \partial \sum_{j=1}^{n} \lambda_j\, g_j(x)/\partial x = 0,$$

$$\partial f/\partial \lambda = (g_j(x) - \theta_j^2) = 0,$$

$$\partial f/\partial \theta_j = 2\lambda_j \theta_j = 0. \tag{3.10}$$

For the last equation, either $\lambda_j^* = 0$ or $\theta_j^* = 0$ or both. If $\lambda_j^* = 0$ and $\theta_j^* \neq 0$, then the constraints $g_j(x) \leq 0$ are ignored. If $\lambda_j = 0$ and $\theta_j = 0$, $\forall j$, then the boundaries pass through the solution for unconstrained optimization. It implies that $S \cap R \not\subset \emptyset$, which is a contradiction. If $\lambda_j \neq 0$ and $\theta_j = 0$, then $g_j(x) = 0$. So that the solution x^* belongs to the boundary of R.

Let $R \cap S \neq \emptyset$. By Theorem III.11 and Corollary III.1, $\exists\ \alpha_i > 0$, $\forall i$, such that the value of $\sum_{i=1}^{k} \alpha_i Y_i(x)$ is maximum at $x^* \in S$. Then, x^* lies either on $R \cap S$ or on the boundary of R if $x^* \not\subset R \cap S$ (look at (1)).

By Theorem III.9 to III.12, we can conclude that the sufficient and necessary conditions under which maximizing $\sum_{i=1}^{k} \alpha_i Y_i(x)$ leads to an admissible point are

$\alpha_i > 0$, and $\exists\ x^*$, a global maximum of $\sum_{i=1}^{k} \alpha_i Y_i(x)$ or

$\alpha_i > 0$, $\forall i$ and the feasible region is a compact set or

$\alpha_i > 0$ and at least one H_{Y_i} is a negative definite matrix.

(3.11)

CHAPTER IV

CONVEX COMBINATION METHOD

The convex combination method is one way of combining several original responses into a single new response. It is an extension from the weighted sum linear combination of several responses. Any optimization will be done on this new response. Let us define

$$\alpha_i > 0, \sum_{i=1}^{k} \alpha_i = 1.$$

The transformation of $\mathbf{y}$ is denoted by $W = \sum_{i=1}^{k} \alpha_i Y_i(\mathbf{x})$. Then, the optimization of all $Y_i(\mathbf{x})$ is transformed into the optimization of W. Since W is a single response, the optimization becomes simpler than the optimization of vector $\mathbf{y}$. Also methods and computer software for optimizing one response are widely available. In this thesis $Y_i(\mathbf{x})$, i = 1, 2, .., k and $\mathbf{x} \in E^p$, are limited to second-order polynomial functions. The choice of α_i might be governed by engineers, based on the importance of the corresponding response. If the value of a certain $Y_i(\mathbf{x})$ is more important than the others, then the we assigned a higher corresponding value of α_i.

If the objective of optimization is that the higher value is better, we have a maximization problem. If the objective is that the lower value is better, we have a minimization problem. If the objective is getting close to a target value, then we deal with minimizing the deviation from the target value. Thus, this deviation is considered as one of the responses. The responses, as a vector of $\mathbf{y}(\mathbf{x})$, may be known or unknown functions.

Properties of the Maximum Point of W

In this thesis we only discuss maximization, since minimization of $Y(\mathbf{x})$ is equivalent to maximization of $(-Y(\mathbf{x}))$. If $\mathbf{y}(\mathbf{x})$ are known, maximizing $W = \sum_{i=1}^{k} \alpha_i Y_i$ by using differential method is as follows:

$$\partial W/\partial \mathbf{x} = \frac{\partial}{\partial(\mathbf{x})} \sum_{i=1}^{k} \alpha_i Y_i(\mathbf{x}) = \sum_{i=1}^{k} \alpha_i \partial Y_i/\partial \mathbf{x}. \quad (4.1)$$

The solution of $\mathbf{x}$, that satisfies

$$\frac{\partial}{\partial(\mathbf{x})} \sum_{i=1}^{k} \alpha_i Y_i(\mathbf{x}) = \sum_{i=1}^{k} \alpha_i \partial Y_i/\partial \mathbf{x} = 0$$

is the stationary point of W.

Let Y_i be a second-order polynomial equation

$$Y_i = a_{oi} + \mathbf{a}_i' \mathbf{x} + \mathbf{x}' A_i \mathbf{x} \quad (4.2)$$

$$\partial Y_i/\partial \mathbf{x} = \mathbf{a}_i + 2\mathbf{A}_i \mathbf{x}$$

$$\partial W/\partial \mathbf{x} = \sum_{i=}^{k} \alpha_i (\mathbf{a}_i + 2\mathbf{A}_i \mathbf{x}) = 0$$

$$\sum_{i=1}^{k} \alpha_i \mathbf{a}_i + 2 \left\{ \sum_{i=1}^{k} \alpha_i \mathbf{A}_i \right\} \mathbf{x} = 0 \quad (4.3)$$

Where $\mathbf{a}_i' = (a_{1i}. a_{2i}, \ldots, a_{pi})$, α_i is a scalar,

$$A_i = \begin{bmatrix} a_{11i} & a_{12i}/2 & \dots & a_{1pi}/2 \\ a_{12i}/2 & a_{22i} & \dots & a_{2pi}/2 \\ \cdot & \cdot & & \cdot \\ a_{1pi}/2 & a_{1pi}/2 & \dots & a_{ppi} \end{bmatrix}$$

and

$$\sum_{i=1}^{k} \alpha_i \mathbf{a}_i = \left[\sum_{i=1}^{k} \alpha_i a_{1i}, \dots, \sum_{i=1}^{k} \alpha_i a_{pi} \right]',$$

i = 1, 2,...., k. Then

$$\mathbf{x}_o = -0.5 \left[\sum_{i=1}^{k} \alpha_i \mathbf{A}_i \right]^{-1} \sum_{i=1}^{k} \alpha_i \mathbf{a}_i \qquad (4.4)$$

where $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$. The solution ($\mathbf{x}_o$) is unique if the matrix $\left(\sum_{i=1}^{k} \alpha_i \mathbf{A}_i \right)$ is nonsingular; otherwise there are an infinite number of solutions or the solution may not exist.

It is necessary to check the eigenvalues of the matrix $\sum_{i=1}^{k} \alpha_i \mathbf{A}_i$. If all the eigenvalues are greater than zero, then W has a minimum point. If all the eigenvalues are less than zero then W has a maximum point. If the eigenvalues have different signs, then W has a saddle point. If at least one of the eigenvalues equals zero, then W has an increasing, a decreasing, or a stationary ridge. For an increasing or decreasing ridge, there is no solution for unconstrained optimization.

If $y(\mathbf{x})$ are unknown, maximization is attempted by either (1) fitting a function to experimental data or (2) using some empirical search procedure to try to find the maximum point. Sometime we also follow both alternatives. When we fit a function to experimental data, then $\hat{y}(x)$ are estimators for $y(x)$, and $\hat{W} = \sum_{i=1}^{k} \alpha_i \hat{Y}_i$ is an estimator of W. Maximizing $\hat{W}$ is different from maximizing W. The solution for optimizing $\hat{W}$ is

$$\hat{x}_o = -0.5 \left[\sum_{i=1}^{k} \alpha_i \hat{A}_i \right]^{-1} \sum_{i=1}^{k} \alpha_i \hat{a}_i . \tag{4.5}$$

$$E(\hat{x}_o) = E\left\{ -0.5 \left[\sum_{i=1}^{k} \alpha_i \hat{A}_i \right]^{-1} \sum_{i=1}^{k} \alpha_i \hat{a}_i \right\}$$

$$\neq -0.5 \left[\sum_{i=1}^{k} \alpha_i A_i \right]^{-1} \sum_{i=1}^{k} \alpha_i a_i . \tag{4.6}$$

The following theorems related to the properties of the stationary point of $\hat{W}$, an estimator of W.

Theorem IV.1. Let W be a convex combination of $Y_i(x)$, $i = 1,\ldots, k$ and $x \in E^p$. The solution for maximizing $\hat{W}$, if it exists, converges in probability to the solution for maximizing W.

Proof. We know from the least square estimator that

$$E(\hat{a}_i) = a_i, \quad E(\hat{A}_i) = A_i .$$

If A is a nonsingular matrix, then the solution exists and it is unique. Let $\{\hat{a}_{in}\} = \sum_{m=1}^{n} \hat{a}_{ijm}/n$ and $\{\hat{A}_{in}\} = \sum_{m=1}^{n} \hat{A}_{ijlm}/n$ be sequences of random variables, where a_{ij} is the element of vector a_i and A_{ijl} is the element of matrix A_i. Suppose the estimated value and variance of $\hat{a}_i$ and $\hat{A}_i$ are finite.

By Chebychev's inequality, then

$$\hat{\mathbf{a}}_{in} \xrightarrow[n \to \infty]{p} \mathbf{a}_i \text{ and } \hat{\mathbf{A}}_{in} \xrightarrow[n \to \infty]{p} \mathbf{A}_i .$$

By Slutsky's theorem, then

$$\left[\hat{\mathbf{A}}_{in}\right]^{-1} \xrightarrow[n \to \infty]{p} [\mathbf{A}_i]^{-1} \text{ and}$$

$$\left[\hat{\mathbf{A}}_{in}\right]^{-1} \hat{\mathbf{a}}_i \xrightarrow[n \to \infty]{p} [\mathbf{A}_i]^{-1} \mathbf{a}_i \tag{4.7}$$

$$\Rightarrow \; E(\hat{\mathbf{x}}_o) \xrightarrow[n \to \infty]{p} -0.5 \left[\sum_{i=1}^{k} \alpha_i \mathbf{A}_i\right]^{-1} \sum_{i=1}^{k} \alpha_i \mathbf{a}_i = \mathbf{x}_o, \tag{4.8}$$

where $[\mathbf{A}]^{-1}$ is the inverse of $\mathbf{A}$.

One criterion for maximizing a combined response which leads to an admissible solution is that the gradient of the combined response belongs to the criterion cone of the gradient of the original responses. The following theorem shows that the gradient of $\sum_{i=1}^{k} \alpha_i Y_i$ belongs to the criterion cone of $\nabla Y_i(\mathbf{x})$, $i = 1, 2, \ldots, k$.

Theorem IV.2. Let W be a convex combination of $Y_i(\mathbf{x})$, $\mathbf{x} \in E^p$. If the null vector condition of the criterion cone of $\nabla \mathbf{Y}_i(\mathbf{x})$ does not hold, then ∇W belongs to the criterion cone of $\nabla \mathbf{Y}_i(\mathbf{x})$, $i = 1, 2, \ldots, k$.

Proof:

$$W = \sum_{i=1}^{k} \alpha_i Y_i(\mathbf{x}), \; \mathbf{x} \in E^p, \; \alpha_i > 0 \text{ and } \sum_{i=1}^{k} \alpha_i = 1$$

$$\nabla W = \left(\frac{\partial w}{\partial x_1}, \; \frac{\partial w}{\partial x_2}, \; \ldots\ldots\ldots, \; \frac{\partial w}{\partial x_p} \right)$$

$$\nabla Y_i(\mathbf{x}) = \left[\frac{\partial Y_i}{\partial x_1} , \frac{\partial Y_i}{\partial x_2} , \ldots\ldots\ldots , \frac{\partial Y_i}{\partial x_p} \right]$$

$$\frac{\partial W}{\partial \mathbf{x}} = \left[\sum_{i=1}^{k} \alpha_i \frac{\partial Y_i}{\partial x_1}, \sum_{i=1}^{k} \alpha_i \frac{\partial Y_i}{\partial x_2}, \ldots\ldots, \sum_{i=1}^{k} \alpha_i \frac{\partial Y_i}{\partial x_p} \right]. \quad (4.9)$$

At point $P(\mathbf{x}_o)$, let PA_i be $\nabla Y_i(\mathbf{x})$ and PQ be ∇W. It can be considered that A_i has coordinates

$$\left[\frac{\partial Y_i}{\partial x_1}, \frac{\partial Y_i}{\partial x_2}, \ldots\ldots\ldots, \frac{\partial Y_i}{\partial x_p} \right] \quad (4.10)$$

and Q has coordinates

$$\left[\sum_{i=1}^{k} \alpha_i \frac{\partial Y_i}{\partial x_1}, \sum_{i=1}^{k} \alpha_i \frac{\partial Y_i}{\partial x_2}, \ldots\ldots, \sum_{i=1}^{k} \alpha_i \frac{\partial Y_i}{\partial x_p} \right]. \quad (4.11)$$

By Definition II.1, then PQ belongs to the criterion cone of $\nabla Y_i(\mathbf{x})$.

When we follow the second alternative in searching for the maximum point, the steepest ascent method is usually applied. In cases with one response, the path of the steepest ascent is found as usual by calculating $\nabla Y(\mathbf{x})$. The experiment is continued by taking a point on the path of the steepest ascent as the center of the experiment. The experiment is repeated several times until we find a maximum point (Myers 1976).

Here we have k responses which have to be maximized simultaneously. From Theorem IV.2, we know that ∇W belongs to the criterion cone of $\nabla Y_i(\mathbf{x})$. This indicates that the solution for maximizing W leads to an admissible point, if the global maximum point exists and is finite. Now by using the steepest ascent method, the search for the maximum point will follow the path of the steepest ascent of W. Since we have k original responses, from which W is derived, we also have the path of the steepest ascent of each individual response. A convex combination of these paths yields the direction of simultaneous optimization. The following theorem shows that the path of the steepest ascent of W is similar to the combined paths from the original responses.

Theorem IV.3. Let W be a convex combination of $Y_i(\mathbf{x}) = \mathbf{a}_i'\mathbf{x}$, i = 1, 2, .., k. An experiment starts at several points and we calculate the path of the steepest ascent of W. The experiment is continued by adding some points along the path of the steepest ascent of W. Then the revised path of the steepest ascent of W, recalculated from revised $\mathbf{a}_i'$, equals its original path (without adding some points).

Proof. Let P_i be starting points in p-dimensional space of the steepest ascent method. Let the original steepest ascent path of W be $(\omega_1^{o}, \omega_2^{o}, \ldots..\omega_p^{o})$. After adding some points along the steepest ascent path of W, the revised steepest ascent path of Y_i will be

$$(\mathbf{a}_{i1}^{*}, \mathbf{a}_{i2}^{*}, \ldots\ldots, \mathbf{a}_{ip}^{*}),\ i = 1, 2, \ldots, k.$$

The revised path of the steepest ascent of W recalculated from $(\mathbf{a}_{i1}^{*}, \mathbf{a}_{i2}^{*}, \ldots.., \mathbf{a}_{ip}^{*})$ is the combined path of $Y_i(\mathbf{x})$ denoted by

$$\left[\sum_{i=1}^{k}\alpha_i\mathbf{a}_{i1}^{*},\ \sum_{i=1}^{k}\alpha_i\mathbf{a}_{i2}^{*},\ \ldots\ldots.\sum_{i=1}^{k}\alpha_i\mathbf{a}_{ip}^{*} \right].$$

The revised steepest ascent path of W, calculated from revised W itself is (ω_1, ω_2, ...,ω_p), where

$$\omega = (\omega_0, \omega_1, .., \omega_p) = (\mathbf{X'X})^{-1}\mathbf{X'}W.$$

Since W is a convex combination of Y_i, then

$$\omega = (\mathbf{X'X})^{-1}\mathbf{X'}\left[\sum_{i=1}^{k} \alpha_i Y_{i1}, \ldots\ldots, \sum_{i=1}^{k} \alpha_i Y_{in}\right]'$$

$$= (\mathbf{X'X})^{-1}\mathbf{X'}[\mathbf{X}\ (\mathbf{a}_1^*\alpha_1) + \ldots + \mathbf{X}(\mathbf{a}_k^*\alpha_k)] \qquad (4.12)$$

and we can write

$$\omega_j = \sum_{i=1}^{k}\alpha_i a_{ij}^*. \qquad (4.13)$$

Where $\mathbf{a}_i^*$ is a (p+1)x1 vector whose elements are a_{ij}^*, ω is a (p+1)x1 vector, i = 1, 2,.., k and j = 0, 1, ..., p.

Johnson and Folks (1964) have proved that

$$(\omega_1^o, \omega_2^o, \ldots\ldots\omega_p^o) \propto (\omega_1, \omega_2, \ldots\ldots,\omega_p).$$

This implies that

$$\left[\sum_{i=1}^{k}\alpha_i \mathbf{a}_{i1}^*, \ldots, \sum_{i=1}^{k}\alpha_i \mathbf{a}_{ip}^*\right] \propto (\omega_1^o, \omega_2^o, \ldots, \omega_p^o).$$

Remark IV.1. From Theorem IV.3, the process of reaching W_{max} in the steepest ascent method, can be based on the steepest ascent path of W itself without predicting the coefficients of regression for individual $Y_i(\mathbf{x})$.

Searching for the maximum point of $W = \sum_{i=1}^{k} \alpha_i Y_i$, $\alpha_i > 0$ and $\sum_{i=1}^{k} \alpha_i = 1$, by using the steepest ascent method leads to an admissible point. Two reasons support this statement: $\alpha_i > 0$ and W has a global maximum point, which satisfy the necessary and sufficient conditions under which maximizing a combined response leads to an admissible solution.

Confidence Region about the Optimum Points of W

Let $\alpha_i > 0$ and $\sum_{i=1}^{k}\alpha_i = 1$. Let $W = \sum_{i=1}^{k}\alpha_i Y_i(x)$, i = 1, 2, .., k and $x \in E^p$. Let $Y_i(x)$ be a second-order polynomial function of x. $Y_i(x)$ can be written as

$$Y_i(x) = a_o + \mathbf{a}_i'\mathbf{x} + \mathbf{x}'\mathbf{A}_i\mathbf{x}$$

where $\mathbf{a}_i' = (a_{1i}, a_{2i}, \ldots., a_{pi})$ and

$$A_i = \begin{bmatrix} a_{11i} & a_{12i}/2 & & a_{1pi}/2 \\ & a_{22i} & & a_{2pi}/2 \\ \text{symetric} & & & \\ & & & a_{ppi} \end{bmatrix}.$$

Then W can be written as

$$W = b_o + \mathbf{b}'\mathbf{x} + \mathbf{x}'\mathbf{B}\mathbf{x}$$

where $b_o = \sum_{i=1}^{k}\alpha_i a_{oi}$, $\mathbf{b} = \left(\sum_{i=1}^{k}\alpha_i a_{1i}, \sum_{i=1}^{k}\alpha_i a_{2i}, \ldots. \sum_{i=1}^{k}\alpha_i a_{pi}\right)$

and $B = \left[\sum_{i=1}^{k}\alpha_i A_i\right]$. Let $W = (W_1, W_2, W_3, \ldots.. ,W_n)$, where n is the number of observations. Then the j^{th} observation of W for a fixed α_i is

$$W_j = \sum_{i=1}^{k} \alpha_i Y_{ij}, \; j = 1, 2, \ldots, n.$$

For a general linear model

$$\underset{n\times 1}{W} = \underset{n\times q}{\mathbf{X}} \; \underset{q\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$$\varepsilon \sim \mathbf{N}\,(0, \sigma^2 I)$$

Using the SAS program for regression analysis or for response surface analysis, we can estimate the coefficients of regression and the mean square error as

$$\hat{\beta} = (\mathbf{X'X})^{-1}\,\mathbf{X'}W \tag{4.14}$$

$$s^2 = (W'W - \hat{\beta}'\mathbf{X'X}\hat{\beta})/(n-q). \tag{4.15}$$

For this model,

$$\beta = (\beta_o, \beta_1, \ldots, \beta_p, \; \beta_{12}, \ldots \;, \beta_{p-1,p}, \ldots, \; \beta_{11}, \; \ldots, \beta_{pp}).$$

Then, we define $b_o = \beta_o$, $\mathbf{b} = (\beta_1, \beta_2, \ldots, \beta_p)$ and

$$B = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & & \beta_{1p}/2 \\ & \beta_{22} & & \beta_{2p}/2 \\ & & & \\ \text{symetric} & & & \\ & & & \beta_{pp} \end{bmatrix}.$$

Let x^* be the solution for maximizing W. For unconstrained optimization, then

$$\partial W/\partial \mathbf{x} = \mathbf{0} \rightarrow \mathbf{b} + 2\,\mathbf{Bx} = \mathbf{0}$$

$$x^* = -0.5\ B^{-1}b, \text{ and } W(x^*) = b_o - b'B^{-1}b/4. \qquad (4.16)$$

For constrained optimization with feasible region

$$R = \left\{x:\ \sum_{j=1}^{p} x_j^{\ 2} \le r^2,\ j = 1, 2, .., p\right\}$$

we need to maximize

$$f(\mathbf{x},u,\theta) = b_o + \mathbf{b'x} + \mathbf{x'Bx} - u\left[\sum_{j=1}^{p} x_j^{\ 2} - r^2 - \theta^2\right].$$

Then,

$$\partial f/\partial \mathbf{x} = \mathbf{b} + 2\ \mathbf{Bx} - 2\ u\ \mathbf{x} = \mathbf{0} \qquad (4.17)$$

$$\partial f/\partial u = \sum_{j=1}^{p} x_j^{\ 2} - r^2 - \theta^2 = 0 \qquad (4.18)$$

$$\partial f/\partial \theta = 2u\theta = 0. \qquad (4.19)$$

There are 3 possibilities:

(1). $u = 0$ and $\theta \neq 0$. The constraints can be ignored; then the solution is $\mathbf{x}^*$.

(2). Both u and $\theta = 0$. The boundary of R passes through $\mathbf{x}^*$.

(3). $u \neq 0$, $\theta = 0$. The solution will lie on the boundary of R.

If $\theta = 0$ and $u \neq 0$, then the solution is

$$\mathbf{x}_c^{\ *} = -0.5\ (\mathbf{B} - u\mathbf{I})^{-1}\ \mathbf{b}$$

and

$$W(\mathbf{x}_c^{\ *}) = b_o - 0.5\mathbf{b'}(\mathbf{B}-u\mathbf{I})^{-1}\mathbf{b} + 0.25\mathbf{b'}(\mathbf{B}-u\mathbf{I})^{-1}\mathbf{B}(\mathbf{B}-u\mathbf{I})^{-1}\mathbf{b}, \qquad (4.20)$$

where u = the Lagrangian multiplier. The value of u is chosen to be greater than the largest eigenvalue of **B** and satisfies $\sum_{j=1}^{p} x_j^{\ 2} \le r^2$.

As mentioned in Chapter II, Rao (1973) showed that for $g(\beta)$,

$$P\{\min_{\beta\in U} g(\beta) \le g(\beta) \le \max_{\beta\in U} g(\beta)\} \ge 1 - \alpha. \tag{4.21}$$

Where

$$U = \{\beta: (\hat{\beta}-\beta)'\mathbf{X}'\mathbf{X}(\hat{\beta}-\beta)/\ qs^2 \le F_{\alpha,q,(n-q)}\} \tag{4.22}$$

is the 100(1-α)% confidence region for β,

$s^2 = (W'W - \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta})/(n-q)$, and $\beta = (b_0, b_1, \ldots, b_p, b_{12}, \ldots, b_{p-1,p}, b_{11}, \ldots\ldots, b_{pp})$.

Carter et al (1984 and 1986) suggested a procedure for computing the confidence region about $g(\beta)$ and about eigenvalues of β, as mentioned in Chapter II. Once the elements of U have been determined, the confidence interval/region about $g(\beta)$ can be constructed (conservative confidence interval). By using this approach and evaluating Equation *(4.16)* and *(4.20)* we can construct the confidence region for x^*, $W(x^*)$, x_c^*, or $W(x_c^*)$ for a given $\alpha_i > 0$, $\sum_{i=1}^{k} \alpha_i = 1$, and a fixed u if $x^* \not\in R$.

Restrictions on α_i

If we impose restrictions on α_i, i = 1, 2, .., k, then the admissible region of $Y_i(x)$ with restrictions on α is a subset of the admissible region of $Y_i(x)$. We will evaluate both cases, without a constraint on x and with constraints on x.

For $p = 2$ and $k = 2$

Without a constraint on x. In this case for $i = 1, 2$ and $x = (x_1, x_2)$, the solution for maximizing W lies on the admissible tangent path of $Y_1(x)$ and $Y_2(x)$. Suppose we impose a restriction on α, such that

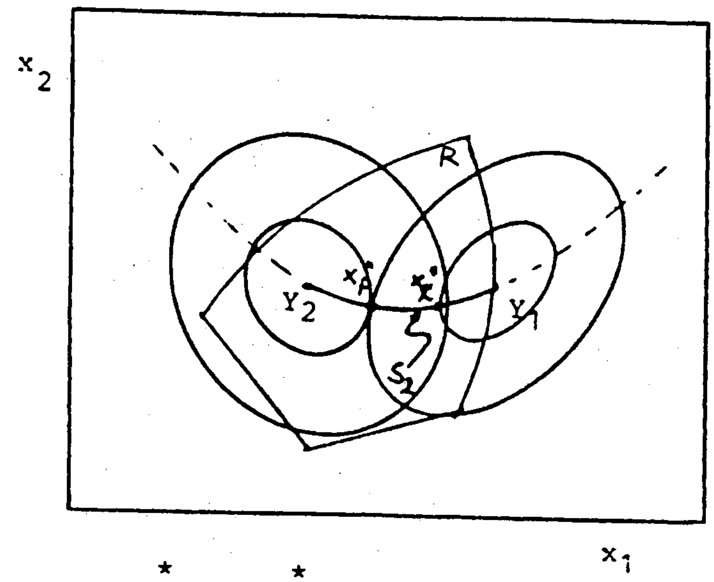
$$p \leq \alpha_1 \leq q,$$

where $p, q, > 0$ and $\alpha_2 = 1-\alpha_1$. Let x^* be the solution for maximizing W for any possible value of α_1. If x^* exists, it always lies on S, the admissible tangent path of $Y_1(x)$ and $Y_2(x)$. Let $x^*_q \in S$ be the maximum point of W for $\alpha_1 = q$. Let $x^*_p \in S$ be the maximum point of W for $\alpha_1 = p$. Then the admissible region for $Y_1(x)$ and $Y_2(x)$ with the restriction $p \leq \alpha_1 \leq q$, where $p, q, > 0$ and $\alpha_2 = 1-\alpha_1$ is the part of S between x^*_q and x^*_p.

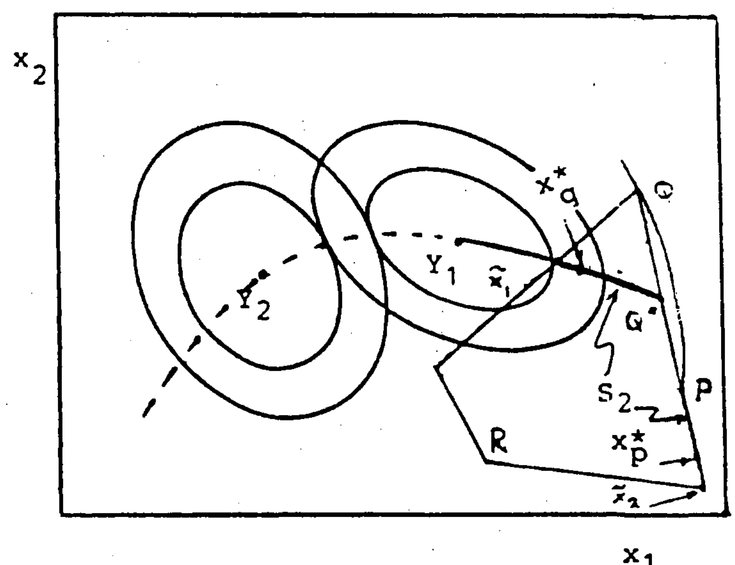
With constraints on x. Here, the admissible region for Y_i, denoted by set S_1, may lie in the interior of R or on the boundary of R or both. First, we take $\alpha_1 = q$; then we find $x^*_q \in R$, as the maximum point of W. Secondly, we take $\alpha_1 = p$; then we find $x^*_p \in R$ as the maximum point of W. Two possible situations will arise:

For situation 1, x^*_q and x^*_p lie on a connected subset of S_1; then the admissible region for $Y_i(x)$ is S_2, the part of S_1 between x^*_q and x^*_p.

For situation 2, Figure 10b shows that x^*_q and x^*_p that lie on disjoint subsets of S_1. Suppose $x^*_q \in (S \cap R)$ and $x^*_p \in A_j^*$; then the admissible region for $Y_i(x)$ is a part of S_1



a. x_p^* and x_q^* lie on joint subsets



b. x_p^* and x_q^* lie on disjoint subsets

Figure 10. Admissible Region for Y_1 and Y_2 with Constraints on α_i

from x^*_q to point Q^* union a part of A_j^* from point P to x^*_p (excluding point P), denoted by set S_2. Illustrations are given in Figure 10b.

For $p \geq 2$ and $k \geq 2$

Without any constraint on x. Let $W = \sum_{i=1}^{k} \alpha_i Y_i$ be a second-order polynomial function of x. It can be written as

$$W = b_o + \mathbf{b}'\mathbf{x} + \mathbf{x}'\mathbf{B}\mathbf{x}$$

Let $\mathbf{x}^*$ be the solution for maximizing W; then

$$\partial W / \partial \mathbf{x} = \mathbf{0} \Rightarrow \mathbf{b} + 2\ \mathbf{B}\mathbf{x} = \mathbf{0}$$

$$\mathbf{x}^* = -0.5\ \mathbf{B}^{-1}\mathbf{b}, \quad W(\mathbf{x}^*) = b_o - \mathbf{b}'\mathbf{B}^{-1}\mathbf{b}/4$$

where $b_o = \sum_{i=1}^{k} \alpha_i a_{oi}$,

$$\mathbf{b} = \left(\sum_{i=1}^{k} \alpha_i a_{1i},\ \sum_{i=1}^{k} \alpha_i a_{2i},\ \ldots\ \sum_{i=1}^{k} \alpha_i a_{ki} \right)' \text{ and}$$

$$\mathbf{B} = \left(\sum_{i=1}^{k} \alpha_i \mathbf{A}_i \right).$$

Let $0 < p_i \leq \alpha_i \leq q_i < 1$, i = 1, .., k-1, and $\sum_{i=1}^{k-1} q_i < 1$.

Let $\alpha_k = 1 - \sum_{i=1}^{k-1} \alpha_i$. Let set V be the set of every possible $\alpha_i > 0$; then

$$V = \left\{ \alpha\colon\ 0 < p_i \leq \alpha_i \leq q_i < 1,\ i = 1, 2, .., k-1, \right.$$
$$\left. \text{and } \sum_{i=1}^{k-1} q_i < 1,\ \alpha_k = 1 - \sum_{i=1}^{k-1} \alpha_i \right\}. \qquad (4.23)$$

Once we have determined the elements of set V, we can evaluate x^* and $W(x^*)$. So we can define

$$V_x = \Big\{ x^* : x^* = -0.5 \Big(\sum_{i=1}^{k} \alpha_i A_i\Big)^{-1} \Big(\sum_{i=1}^{k} \alpha_i a_{1i}, \ldots, \sum_{i=1}^{k} \alpha_i a_{pi}\Big)',$$

$$\alpha_i \in V \text{ and } \Big(\sum_{i=1}^{k} \alpha_i A_i\Big)^{-1} = \text{a negative definite matrix} \Big\}. \qquad (4.24)$$

The admissible region of Y_i is the set V_x.

<u>With constraints on x.</u> Let the feasible region R be the set $R = \Big\{ x : \sum_{j=1}^{p} x_j^2 \le r^2 \Big\}$. By the same way, we can evaluate the maximum point of W. If the solution lies in R then the maximum point of W is x^*. If not, the maximum point is $x_c^* \in R$ denoted by

$$x_c^* = -0.5 \Big(\sum_{i=1}^{k} \alpha_i A_i - uI\Big)^{-1} \Big(\sum_{i=1}^{k} \alpha_i a_{1i}, \ldots, \sum_{i=1}^{k} \alpha_i a_{pi}\Big)'$$

where u = a Lagrangian multiplier. The value of u is chosen to be greater than the largest eigenvalue of $\Big(\sum_{i=1}^{k} \alpha_i A_i - uI\Big)$ and satisfies $\sum_{j=1}^{p} x_j^2 \le r^2$. Then the admissible region V_{xc} is defined by

$$V_{xc} = V_x \cap R \cup V_{x_*}, \qquad (4.25)$$

where

$$V_{x_*} = \Big\{ x_c^* : x_c^* = -0.5 \Big(\sum_{i=1}^{k} \alpha_i A_i - uI\Big)^{-1} \Big(\sum_{i=1}^{k} \alpha_i a_{1i}, \ldots, \sum_{i=1}^{k} \alpha_i a_{pi}\Big)',$$

$$\alpha_i \in V, \text{ and } \Big(\sum_{i=1}^{k} \alpha_i A_i - uI\Big) = \text{a negative definite matrix} \Big\} \text{ or} \qquad (4.26)$$

$$V_{x_*} = \left\{x_c^*:\ x_c^* \in R \text{ is the maximum point of } \sum_{i=1}^{k} \alpha_i Y_i(x),\ \alpha_i \in V\right\}. \tag{4.27}$$

Numerical Examples

Several Kinds of Surface Combinations

There are four cases to be considered as follows:

(1). Both Y_1 and Y_2 have maximum points.

(2). Y_1 has a maximum point and Y_2 has a minimum point.

(3). Y_1 has a maximum point and Y_2 has a saddle point.

(4). Both Y_1 and Y_2 have saddle points.

Suppose both Y_1 and Y_2 have maximum points. Then let $Y_1 = 10 + 2x_1 + x_2 - x_1^2 + 2x_1x_2 - 3x_2^2$,

$Y_2 = 15 + x_1 - 0.5x_2 - 2x_1^2 - 3x_1x_2 - 2.5x_2^2$,

$W = 0.5\ (Y_1+Y_2) = 0.5\ (25+3x_1-0.5x_2-3x_1^2-x_1x_2-5.5x_2^2)$.

The stationary points are as follows:

Y_1 has a maximum point at (1.75, 0.75),

Y_2 has a maximum point at (0.59, -0.4545), and

W has a maximum point at (0.5, 0.0).

From the above equations, the tangent path of Y_1 and Y_2 can be computed as follows:

$$2 + 3x_1 + 2x_2 - 14x_1^2 + 12x_1x_2 + 28x_2^2 = 0.$$

By substituting the coordinate of W_{max} in the equation for the tangent path, it can be shown that the point (0.5, 0) lies on this path, since

$$2 + 2(0.5) - 14(0.5^2) + 0 = 0$$

and it also lies on the admissible tangent path of Y_1 and Y_2.

Suppose Y_1 has a maximum and Y_2 has a minimum point. Let $Y_1 = 10 + 2x_1 + x_2 - x_1^2 + 2x_1x_2 - 3x_2^2$,

$Y_2 = 15 + x_1 - 0.5x_2 + 2x_1^2 + 3x_1x_2 + 2.5x_2^2$,

$W_1 = 0.5\ (Y_1 + Y_2)$

$= 0.5\ (25 + 3x_1 + 0.5x_2 + x_1^2 + 5x_1x_2 - 0.5x_2^2)$, and

$W_2 = 0.9Y_1 + 0.1Y_2$

$= 2.4 + 1.9x_1 + 0.85x_2 - 0.7x_1^2 + 2.1x_1x_2 - 2.45x_2^2$.

The stationary points are as follows:

Y_1 has a maximum point at (1.75, 0.75),

Y_2 has a minimum point at (-0.59, 0.45),

W_1 has a saddle point at (-0.2039, 0.5185), and

W_2 has a maximum point at (4.53, 2.11).

The tangent path of Y_1 and Y_2 can be computed as

$$2 - x_1 - 12x_2 + 14x_1^2 - 14x_1x_2 - 28x_2^2 = 0.$$

By substitution, it can be shown that the coordinate of W_1 and W_2 both lie on the tangent path of Y_1 and Y_2. However, only W_{2max} lies on the admissible tangent path.

Suppose Y_1 has a maximum and Y_2 has a saddle point. Let $Y_1 = 10 + 2x_1 + x_2 - x_1^2 + 2x_1x_2 - 3x_2^2$,

$Y_2 = 15 + x_1 - 0.5x_2 - 2x_1^2 - 3x_1x_2 + 2.5x_2^2$, and

$W = 0.5\ (Y_1 + Y_2)$

$= 0.5\ (25 + 3x_1 + 0.5x_2 - 3x_1^2 - 5x_1x_2 - 0.5x_2^2)$.

The stationary points are as follows:

Y_1 has a maximum point at (1.75, 0.75),

Y_2 has a saddle point at (0.1207, 0.1724), and

W has a maximum point at (0.5, 0.0).

The equation of the tangent path of Y_1 and Y_2 is

$$2 + 3x_1 - 18x_2 - 14x_1^2 + 34x_1x_2 + 8x_2^2 = 0.$$

W_{max} (0.5 , 0.0) also lies on the admissible tangent path of Y_1 and Y_2.

Suppose both Y_1 and Y_2 have saddle points.

Let $Y_1 = 10 + 2x_1 + x_2 - 1.5x_1^2 + 4x_1x_2 - x_2^2$,

$Y_2 = 15 + x_1 - 0.5x_2 + x_1^2 - 3x_1x_2 - 2x_2^2$, and

$W = 0.5\ (Y_1 + Y_2)$

$= 0.5\ (25 + 3x_1 + 0.5x_2 - 0.5x_1^2 + x_1x_2 - 3x_2^2.$

The stationary points are:

Y_1 has a saddle point at (-0.80, -1.11),

Y_2 has a saddle point at (0.32, 0.12), and

W has a maximum point at (3.7, 0.7).

The equation of the tangent path of Y_1 and Y_2 is

$$2 + 10.5x_1 + 5x_2 - x_1^2 - 16x_1x_2 + 22x_2^2 = 0.$$

The maximum point of W also lies on the admissible tangent path of Y_1 and Y_2.

Confidence Region of $\underline{x}$

If the uncertainty of the optimum points is considered, then their confidence regions should be constructed. For this purpose, a numerical example is given for the optimum point of $\sum_{i=1}^{k} \alpha_i Y_i(x)$, i = 1, 2, 3, 4, $x \in E^p$ and fixed α_i, by using Carter's procedure.

Suppose we have 4 responses, as function of x_1 and x_2. The estimate of second-order polynomial equations are as follows:

$\hat{Y}_1 = 6.15 - 1.5283x_1 - 0.1586x_2 + 0.1381x_1^2 - 0.0094x_2^2$.

$\hat{Y}_2 = 17.495-1.4603x_1-1.4596x_2+0.5125x_1x_2-0.6263x_1^2-0.6288x_2^2$

$\hat{Y}_3 = 18 - 0.7292x_1 + 1.2593x_2 + x_1^2 + 0.75x_2^2$

$\hat{Y}_4 = 4.4775+1.1106x_1+0.1874x_2+0.0075x_1x_2+0.1031x_1^2-0.0019x_2^2$

Y_1 is moisture content in %; Y_2 is irregularity in "uster" units, Y_3 is cost of production, and Y_4 is yarn strength in grams per denier. Yarn manufacturer want to maximize Y_1, minimize Y_2, minimize Y_3 and maximize Y_4, for producing better yarn quality. The Textile engineer specifies α_i based on the importance of each response for a particular use of the yarn. Let

$$W = 0.35\ Y_1 + 0.15\ (-Y_2) + 0.30\ (-Y_3) + 0.20\ Y_4.$$

After data collection and analysis, then the estimate coefficients of regression for W are

$\hat{\beta}' = (-4.9763, 0.1250, -0.1772, -0.0754, -0.1371, -0.1343)$.

The **X'X** is defined as

$$\mathbf{X'X} = \begin{bmatrix} 12 & 0 & 0 & 0 & 8 & 8 \\ 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 8 & 0 & 0 & 0 & 12 & 4 \\ 8 & 0 & 0 & 0 & 4 & 12 \end{bmatrix}.$$

Let Λ be the diagonal matrix of eigenvalues of $\mathbf{X'X}$ and P be the correspond eigenvectors. From Equation (2.21), we can define

$$\beta = \hat{\beta} - P\Lambda^{-1/2}\mathbf{z},$$

where $\beta \in U$, as defined in Equation (4.22). For $\alpha = 0.10$, $q = 6$, $n = 12$, and $s^2 = 0.0013$, we compute

$$P\Lambda^{-1/2} = \begin{bmatrix} 0.1273 & 0 & 0 & 0 & 0 & -0.4835 \\ 0 & 0 & 0 & 0.3532 & 0 & 0 \\ 0 & 0 & 0.3536 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0.1073 & 0.25 & 0 & 0 & 0 & 0.2868 \\ 0.1073 & -0.25 & 0 & 0 & 0 & 0.2868 \end{bmatrix}.$$

Let $\theta_i = (-90, -45, 0, 45, 90)$, for i = 1, 2, 3, 4.

$\theta_5 = (-150, -105, -60, -15, 30, 75, 120, 165)$.

Based on the above combinations for θ, we have 5000 sets of values for β, for constructing the simultaneous confidence region for x_1 and x_2, the coordinates of the maximum point of W. The plot of x_1 versus x_2 is shown in Figure 11. It seems that we need many more values for β, to get a "good" shape of the plot.

Restrictions on α

Based on data in sub (2), suppose we restrict α_i as follows:

$$0.32 \le \alpha_1 \le 0.38,\ 0.12 \le \alpha_2 \le 0.18,$$

$$0.26 \le \alpha_3 \le 0.34,\ \alpha_4 = 1- (\alpha_1 + \alpha_2 + \alpha_3).$$

Let $\alpha_1 = (0.32, 0.34, 0.35, 0.36, 0.38)$,

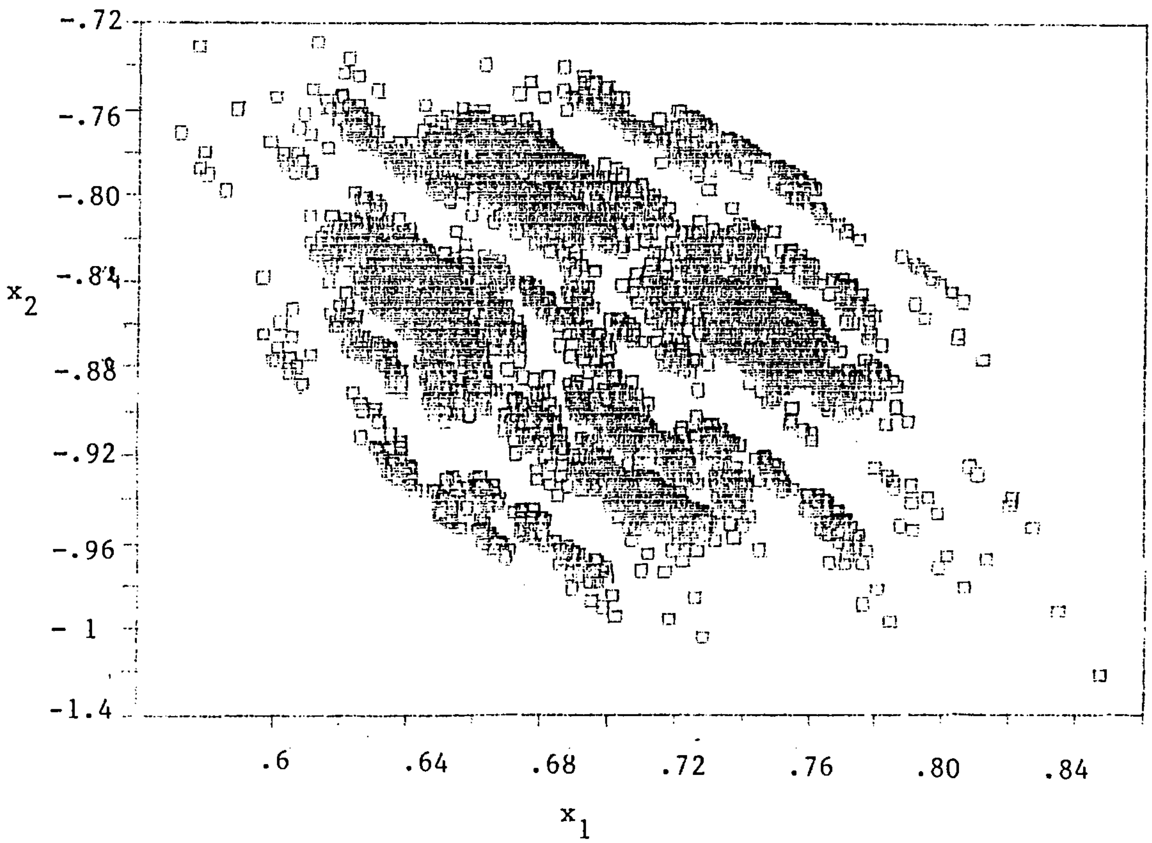


Figure 11. Plot of Optimum Points for Fixed α_i in 90% Confidence Region

α_2 = (0.12, 0.14, 0.15, 0.16, 0.18), and

α_3 = (0.26, 0.28, 0.30, 0.32, 0.34).

From the above combinations, we have 125 sets of values for β, so that we have 125 maximum points of W. Figure 12 shows the plot of x_1 versus x_2. If we develop very large number of values for β, we can construct the region of x, for certain restrictions on α.

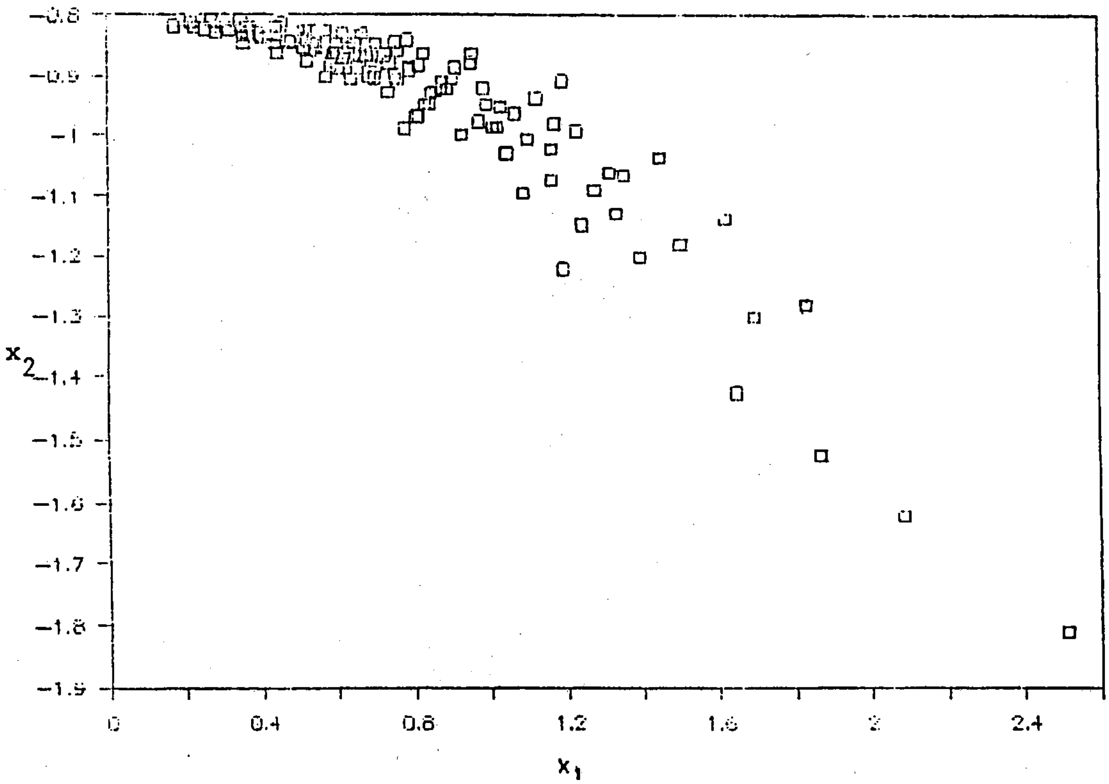


Figure 12. Plot of Optimum Points for Given Intervals of α_i

CHAPTER V

COMPARISONS OF THE ADMISSIBILITY OF SOME COMBINED RESPONSE METHODS

Convex Combination Method

Let $Y_i(x)$, $i = 1, 2, \ldots, k$ and $x \in E^p$ be second-order polynomial functions. Let W be a convex combination of $Y_i(x)$, denoted by

$$W = \sum_{i=1}^{k} \alpha_i Y_i(x), \quad x \in E^p$$

where $\alpha_i > 0$ and $\sum_{i=1}^{k} \alpha_i = 1$.

Steuer (1986) has proved that if $\sum_{i=1}^{k} \alpha_i Y_i(x)$, $\alpha_i > 0$, has a global maximum at x^*; then x^* is admissible for Y_i. However, if $\sum_{i=1}^{k} \alpha_i Y_i(x)$ has a saddle point at x^o then this point is inadmissible for $Y_i(x)$.

Let W satisfies the above conditions. Let a feasible region R be a compact set. We shall describe the method with zero error. Let x_o be the maximum point of W on R, despite the nature of the surface of W or Y_i. Since R is a compact set, we can find a point $x_o \in R$, for which $W(x_o)$ has the highest value on R. Then by Theorem III.9, the maximum point of W, $x_o \in R$, is admissible for all $Y_i(x)$. So that the solution for maximizing W always leads to an admissible point.

Harrington's Method

Harrington (1965) maximized the estimated response $\hat{Y}_i(x)$. We shall describe the method with zero error. Let the feasible region R be a compact set. The desirability functions are

$$d_i = \exp(-|Z_i|^n),\ n > 0\ ,\ i = 1, 2, \ldots, k$$

where $$Z_i = \frac{Y_i(x) - (Y_i^* + Y_{i*})/2}{Y_i^* - Y_{i*}}\ ,\ i = 1, 2, \ldots, k,$$

$$D = \left(\prod_{i=1}^{k} d_i\right)^{1/k}\ ,\ i = 1, 2, \ldots, k,$$

$$\log D = \frac{1}{k} \sum_{i=1}^{k} \alpha_i (-|\{Y_i(x) - a_i\}/b_i|^n),$$

where Y_i^* = the upper specification limit, Y_{i*} = the lower specification limit, and

$a_i = (Y_i^* + Y_{i*})/2$ and $b_i = Y_i^* - Y_{i*}$.

Maximizing D is equivalent to maximizing log D. If n is an odd number, then we can rewrite

$$\log D = \sum_{i=1}^{k} \alpha_i \beta_i (\{Y_i(x) - a_i\}/b_i)^n \qquad (5.1)$$

where $$\beta_i = \begin{cases} -1/k, & \text{if } \{Y_i(x) - a_i\} > 0 \\ 1/k, & \text{if } \{Y_i(x) - a_i\} < 0. \end{cases}$$

Thus, sometimes β_i may be positive and sometimes negative, so that the global maximum of D may not be admissible for all $Y_i(x)$. If we consider a_i as a target value that lie in the center of the specification limit, then the global maximum of D is admissible for $|Y_i(x) - a_i|$, since the objective is to minimize the deviation from the target value. Therefore, maximizing D using Harrington's method may not always lead to an admissible solution for $Y_i(x)$.

Derringer- Suich's Method

In this method, two cases will arise: one-sided and two-sided desirability functions (Derringer and Suich 1980). Let the feasible region be a compact set. For one-sided cases, the desirability functions are given (again with zero error) by

$$d_i = \begin{cases} 0 & Y_i < Y_{i*} \\ \left[\dfrac{Y_i - Y_{i*}}{Y_i^* - Y_{i*}} \right]^r & Y_{i*} \le Y_i \le Y_i^* \\ 1 & Y_i > Y_i^* \end{cases}$$

For two-sided cases, the desirability functions are given by

$$d_i = \begin{cases} \left[\dfrac{Y_i - Y_{i*}}{c_i - Y_{i*}} \right]^s & Y_{i*} \le Y_i \le c_i \\ \left[\dfrac{Y_i - Y_i^*}{c_i - Y_i^*} \right]^t & c_i < Y_i \le Y_i^* \\ 0 & Y_i < Y_{i*} \quad \text{or} \quad Y_i > Y_i^* \end{cases}$$

and

$$D = \left(\prod_{i=1}^{k} d_i\right)^{1/k}, \quad i = 1, 2, .., k,$$

$$\log D = \frac{1}{k}\sum_{i=1}^{k}\alpha_i \log d_i .$$

Suppose s = t = 1; then for two-sided cases

$$\log D = \sum_{i=1}^{k} \log\{(c_i - Y_{i*})^{-1}(Y_i(x) - Y_{i*})\}, \text{ if } Y_{i*} \le Y_i(x) \le c_i$$

$$= \sum_{i=1}^{k} \log\{(c_i - Y_i^*)^{-1}(Y_i(x) - Y_i^*)\}, \text{ if } c_i \le Y_i(x) \le Y_i^* . \tag{5.2}$$

Equation (5.2) implies that for $c_i \le Y_i(x) \le Y_i^*$, if $Y_i(x)$ increases, then log D will decrease. We can rewrite log D as

$$\log D = \begin{cases} \alpha_i \sum_{i=1}^{k} \alpha_i Y_i(x), \ \alpha_i > 0 \text{ for } Y_{i*} \le Y_i \le c_i \\ \alpha_i \sum_{i=1}^{k} \beta_i Y_i(x), \ \beta_i < 0 \text{ for } c_i \le Y_i^* \le Y_i . \end{cases} \tag{5.3}$$

Therefore, maximizing D using Derringer-Suich's method may not always lead to an admissible solution for all $Y_i(x)$. Yet, if we consider c_i as a target value, then the global maximum of D is admissible for $|Y_i(x) - c_i|$.

Khuri-Conlon's Method

Optimization by Khuri and Conlon (1981) is done by minimizing

$$\rho[\mathbf{y(x)}, \phi] = [\{\mathbf{y(x)} - \phi\}'\{\mathrm{var}(\mathbf{y(x)})\}^{-1}\{\mathbf{y(x)} - \phi\}]^{1/2} \text{ or}$$

$$[\rho\{\mathbf{y(x)}, \phi\}]^2 = [\{\mathbf{y(x)} - \phi\}'\{\mathrm{var}(\mathbf{y(x)})\}^{-1}\{\mathbf{y(x)} - \phi\}]$$

Again, we describe the optimization without error. Let the feasible region be a compact set. Suppose vector ϕ is treated as a vector of constant. Let $x_o \in R$ be a global minimum of ρ. Then for all $x^* \in R$,

$$\{y(x_o)-\phi\}'\{\text{var } y(x_o)\}^{-1}\{y(x_o)-\phi\} \leq$$

$$\{y(x^*)-\phi\}'\{\text{var } y(x^*)\}^{-1}\{y(x^*)-\phi\}. \tag{5.4}$$

In this method, $Y_i(x)$ are assumed to be linearly independent, hence the var$\{y(x)\}$ can be assumed to be a diagonal matrix.

$$\text{var}\{y(x)\} = \text{diag}\{V_{11}(x), V_{22}(x), \ldots, V_{kk}(x)\}. \tag{5.5}$$

Then the Inequality (5.4) can be written as

$$\sum_{i=1}^{k}(Y_i(x_o)-\phi)^2/V_{ii}(x_o) \leq \sum_{i=1}^{k}(Y_i(x^*)-\phi)^2/V_{ii}(x^*). \tag{5.6}$$

Suppose x_o is admissible for $Y_i(x)$; then there does not exist x^* such that

$$Y_i(x^*) \geq Y_i(x_o), \text{ for all i and}$$

$$Y_i(x^*) > Y_i(x_o), \text{ for at least one i.}$$

Let us consider any $x^* \neq x_o$. Then either

$$(Y_i(x^*)-\phi) \leq (Y_i(x_o)-\phi), \forall i \text{ or}$$

$$(Y_i(x^*)-\phi) > (Y_i(x_o)-\phi), \text{ for some i and}$$

$$(Y_{i'}(x^*)-\phi) \leq (Y_{i'}(x_o)-\phi), \text{ for } i \neq i'. \tag{5.7}$$

Since $(Y_i(x)-\phi) \leq 0$, then either

$$(Y_i(x^*)-\phi)^2 \geq (Y_i(x_o)-\phi)^2, \forall i \text{ or}$$

$$(Y_i(x^*)-\phi)^2 < (Y_i(x_o)-\phi)^2, \text{ for some i and}$$

$$(Y_{i'}(x^*)-\phi)^2 \geq (Y_{i'}(x_o)-\phi)^2, \text{ for } i \neq i'.$$

If $V_{ii}(x_o) \leq V_{ii}(x^*)$, for all i, then it may happen that

$$(Y_i(x^*)-\phi)^2/V_{ii}(x^*) \leq (Y_i(x_o)-\phi)^2/V_{ii}(x_o) \tag{5.8}$$

Let α_i have the same value, $\forall i$; then it may happen that

$$\sum_{i=1}^{k}(Y_i(x^*)-\phi)^2/V_{ii}(x^*) \leq \sum_{i=1}^{k}(Y_i(x_o)-\phi)^2/V_{ii}(x_o), \tag{5.9}$$

which contradicts Inequality (5.6). Therefore, simultaneous optimization using Khuri-Conlon's method may not always lead to an admissible solution for $Y_i(x)$.

Numerical Comparisons

Comparisons between the desirability functions versus the convex combination method apply a set of data, developed by computer, based on second order polynomial regression. We consider three responses as follows:

$Y_1 = 10 + 2x_1 + x_2 - x_1^2 + 2x_1x_2 - 3x_2^2 + \epsilon_1$ (maximum)

$Y_2 = 15 + x_1 - 0.5x_2 - 2x_1^2 - 3x_1x_2 + 2.5x_2^2 + \epsilon_2$ (saddle)

$Y_3 = 12 + x_1 - 0.5x_2 + 2x_1^2 + 3x_1x_2 + 2.5x_2^2 + \epsilon_3$ (minimum)

ϵ_1 is assumed to be distributed as N(0, 0.001)

ϵ_2 is assumed to be distributed as N(0, 0.0064)

ϵ_3 is assumed to be distributed as N(0, 0.0225).

The specification limits of Y_i are $Y_1 \geq 4$, $5 \leq Y_2 \leq 15$, with $c_2 = 10$, and $20 \leq Y_3 \leq 30$ with $c_3 = 25$.

For optimizing W we can choose several sets of $\alpha_i > 0$, as long as W has a global maximum point. In this example we choose 3 sets of α_i to produce W_1, W_2, and W_3. Where $W_1 = 0.60Y_1 + 0.25Y_2 + 0.15Y_3$, $W_2 = 0.55Y_1 + 0.30Y_2 + 0.15Y_3$ and $W_3 = 0.5718834Y_1 + 0.2792704Y_2 + 0.1488462Y_3$. All W has global maximum points. The solution of optimizing W_3 is similar to the result of Derringer's method. Any choice of $\alpha_i > 0$ can be used in the optimization of W, as long as W has maximum point. This choice will determine the value of Y_i. The solution of optimization is shown in Table I.

The comparison of Khuri and Conlon's versus the convex combination method applies an example from Khuri and Conlon's paper (1981). In that paper they gave the solutions of optimization for both cases: if the vector of individual optima is treated as a vector of constant and if it is treated as a vector of random variable.

There are 4 responses:

$\hat{Y}_1 = 1.526 - 0.575x_1 - 0.524x_2 + 0.318x_1x_2 - 0.171x_1^2 - 0.098x_2^2$ (sadd)

$\hat{Y}_2 = 0.660 - 0.092x_1 - 0.010x_2 - 0.070x_1x_2 - 0.096x_1^2 - 0.058x_2^2$ (max)

$\hat{Y}_3 = 1.776 - 0.250x_1 - 0.078x_2 + 0.010x_1x_2 - 0.156x_1^2 - 0.079x_2^2$ (max)

$\hat{Y}_4 = 0.468 + 0.131x_1 + 0.073x_2 - 0.083x_1x_2 + 0.026x_1^2 + 0.024x_2^2$. (sadd)

The solution of optimization and the values of Y_i at the optimum point are shown in Table II. All W have global maximum points. The solution of maximizing W_4 is similar to that of of minimizing ρ if ϕ is treated as a random vector, but the solution of maximizing W_3 is slightly different from that of minimizing ρ, if ϕ is treated as a vector of constant.

TABLE I

COMPARISONS AMONG CONVEX COMBINATIONS, HARRINGTON'S, AND DERRINGER-SUICH'S METHODS

Methods	x_1	x_2	Y_1	Y_2	Y_3
Harrington, n=3	1.2670	1.4426	9.825	12.069	26.411
Derringer, r=s=1	1.4436	1.1167	11.446	10.000	25.000
C.C. method, W_1	1.6585	1.1811	11.518	8.174	27.940
W_2	1.3451	1.1411	11.236	10.825	24.186
W_3	1.4436	1.1167	11.446	10.000	25.000

TABLE II

COMPARISONS BETWEEN KHURI-CONLON'S AND CONVEX COMBINATION METHODS

Methods	x_1	x_2	Y_1	Y_2	Y_3	Y_4
Khuri, ϕ as r.v	-.57	-1.29	2.54	0.55	1.84	0.29
Khuri, ϕ as const.	-.46	-1.38	2.47	0.55	1.83	0.31
C.C method, W_1	-1.40	-1.86	3.46	0.24	1.72	0.07
W_2	-.41	-1.16	2.36	0.58	1.84	0.33
W_3	-.46	-1.31	2.47	0.55	1.83	0.31
W_4	-.57	-1.29	2.54	0.55	1.84	0.29

Where $W_1 = 0.25 \sum_{i=1}^{4} Y_i$, $W_2 = 0.15Y_1 + 0.15Y_2 + 0.1Y_3 + 0.6Y_4$,

$W_3 = 0.1578Y_1 + 0.1342Y_2 + 0.0995Y_3 + 0.6985Y_4$, and

$W_4 = 0.163Y_1 + 0.153Y_2 + 0.103Y_3 + 0.581Y_4$.

CHAPTER VI

SUMMARY AND CONCLUSIONS

This thesis discusses the simultaneous optimization of several responses. The discussions focus on the characterization of sets of admissible points and determination of the existence of the admissible region.

The responses are limited to second order polynomial functions of **x**. We observed several kinds of surfaces of the original responses and the combination of the surfaces in forming a combined response.

Several lemmas and theorems is developed for characterizing the sets of admissible points for both constrained and unconstrained optimization. If the number of responses is less than or equal to three and the control variables **x** lie in 2-dimensional space, the characterization is well defined and can be shown by graphs; otherwise by algebraic notations.

The admissible region will exist if $\alpha_i > 0$ and at least one of the Hessian matrices of $Y_i(\mathbf{x})$ is negative definite. If the feasible region is a compact set (closed and bounded) the condition for the existence of the admissible point is only $\alpha_i > 0$.

For unconstrained optimization, the admissible region for two responses in 2-dimensional space is a particular part of their tangent path. For three responses in 2-dimensional space, the admissible region is the closed region bounded by the admissible tangent paths.

For constrained optimization, the admissible region may lie in the interior of feasible region R, that is $R \cap S$, and on some parts of its boundary or only on some parts of its boundary. The last position happens if R does not intersect S, the admissible region for unconstrained optimization. If at least one response has a saddle point, the subset of admissible region may lie in the interior of R, but outside $R \cap S$. If the number of responses or control variables is greater than two, the admissible region is S_1 the set of all $\mathbf{x}^*$ denoted by

(1) for unconsctrained optimization

$$S_1 = \{\ \mathbf{x}^*\colon \mathbf{x}^* \text{ satisfies } \nabla(\Sigma\alpha_i Y_i(\mathbf{x}^*)) = 0,\ \alpha_i > 0 \text{ and } \Sigma\alpha_i \mathbf{A}_i \text{ is a negative definite matrix } \},$$

where $\mathbf{A}_i$ is the Hessian matrix of $Y_i(\mathbf{x})$.

(2) for consctrained optimization with $R = \{\ \mathbf{x}\colon \Sigma \mathbf{x}_j^2 \le r^2\ \}$,

$$S_1 = \{\ \mathbf{x}^*\colon \mathbf{x}^* \text{ satisfies } \nabla(\Sigma\alpha_i Y_i(\mathbf{x}^*) - uI) = 0,\ \alpha_i > 0, \text{ and } (\Sigma\alpha_i \mathbf{A}_i - uI) \text{ is a negative definite matrix } \}.$$

If the uncertainty of the optimum points is considered, then their confidence regions can be constructed by using Carter's procedure. For this purpose, a numerical example is given in this thesis for the optimum point of W =

$\sum_{i=1}^{k} \alpha_i Y_i(x)$, i = 1, 2, 3, 4, $x \in E^p$ and fixed α_i.

We also compare the admissibility of the solutions of four combined response functions: convex combination, Harrington's, Derringer-Suich's, and Khuri-Conlon's methods. In these comparisons the feasible region is a compact set. It can be proven that the convex combination method always leads to an admissible solution, but the other three methods do not always lead to an admissible solution.

We recommend using the convex combination method in searching for the optimum point. If the functions of responses of interest are not known, the steepest ascent method can be used for maximizing the convex combination function of the original responses. In experiments, the feasible region is usually a compact set; then the solution for optimizing the covex combunation function will always lead to an admissible point. Also it can be proven that the solution converges in probability to the true value.

It will be interesting for future research to discuss simultaneous optimization for general functions of vector **x**, the confidence region of the admissible region and to investigate the rate of convergence.

BIBLIOGRAPHY

Baker, T. B. (1986). Quality Engineering by Design: Taguchi's Philosophy. *Quality Progress 19, 32-42.*

Bertsimas, D. and Tsitsiklis, J. (1992). Simulated Annealing. in: National Research Council eds.: Probability and Algorithms. *National Academy Press, Washington D.C.*

Box, George. (1988). Signal-to-Noise-Ratio, Performance Criteria, and Transformations. *Technometrics (with discussions) 30, 1-30.*

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

Byrne, D. M. and Taguchi, S. (1987). The Taguchi Approach to Parameter Design. *Quality Progress 20, 19-26.*

Carter, W.H. Jr., Chinchilli, V. M., Campbell, E. D. and Wampler, G. L. (1984). Confidence Interval About the Response at the Stationary Point of a Response Surface with an Application to Pre Chemical Cancer Therapy. *Biometrics 40, 1125-1130.*

Carter, W. H. Jr., Chinchilli, V. M., Myers, R. H., and Campbell, E. D. (1986). Confidence Intervals and an Improved Ridge Analysis of Response Surface. *Technometrics 28, 339-346.*

Carter, W. H. Jr., Chinchilli, V. M., and Campbell, E. D. (1990). A Large-Sample Confidence Region Useful in Characterizing the Stationary Point of a Quadratic Response Surface. *Technometrics 32, 425-444.*

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

Folks, J. L. and Antle, C. E. (1965). Optimum Allocation of Sampling units to strata when there are R Responses of Interest. *Journal of the American*

Statistical Association 60, 225-233.

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

Johnson, C. H. and Folks, J. L. (1964). A Property of the Method of Steepest Ascent. *Ann. Math. Statist. 35, 435-437.*

Karlin, S. (1959). Mathematical Methods and Theory in Games, Programming, and Economics. *Addision-Wesley, Reading, Mass.*

Khuri, A. I. and Conlon, M. (1981). Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions. *Technometrics 23, 363-374.*

Kirkpatrick, S., Gelatt, C. D. Jr., and Vecci, M. P. (1983). Optimization of Simulated Annealing. *Science 220, 671-679.*

Leon, R. N., Shoemaker, A. C., and Kacker, R. N. (1987). Performance Measures Independent of Adjustment: An Explanation and Extension of Taguchi's Signal-to-Noise Ratios. *Technometrics (with discussions) 29, 253-285.*

Mockus, J. B. and Mockus, L. J. (1991). Bayesian Approach to Global Optimization of Multiobjective and Constrained Problems. *JOTA 70, 157-172.*

Myers, R. H. (1976). Response Surface Methodology. *Edward Brothers, Ann Arbor, MI.*

Rao, C. R. (1973). Linear Statistical Inference and its Application. *Wiley, New York.*

Ross, P. J. (1988). Taguchi Techniques for Quality Engineering. *McGraw-Hill Book Co, New York.*

Steuer, R. E. (1986). Multiple Criteria Optimization Theory, Computation and Application. *John Wiley & Son, New York.*

Wismer, D. A. and Chattergy, R. (1979). Introduction to Non Linear Optimization A Problem Solving Approach. *North Holland, New York.*

Vos, H. J. (1990). Simultaneous Optimization of Decisions Using a Linear Utility Function. *Journal of Educational Statistics 15, 309-340.*

VITA

Astini Salihima

Candidate for the degree of

Doctor Philosophy

Thesis: THE STUDY OF SIMULTANEOUS OPTIMIZATION FOR SEVERAL RESPONSES

Major Field: Statistics

Biographical:

Personal Data: Born in Marabahan, South Borneo, Indonesia, October 8, 1940, the daughter of Mathan and Maswinah Kambri.

Education: Graduated from Banjarmasin High School, Banjarmasin, Indonesia, in 1959; graduated from Textile College, Bandung, Indonesia, in 1963; received Bachelor of Science degree in Textile Technology from Textile Technology Institute, Bandung, in 1972; received Master of Science degree in Applied Statistics from Bogor Agriculture Institute, Indonesia, in 1979; received Master of Science degree in Statistics from Iowa State University in 1989; completed requirements for the Doctor of Philosophy degree in statistics at Oklahoma State University in May 1993.

Professional Experience: Researcher, Institute of Research and Development for Textile Industries Bandung, Indonesia, 1984-1986; research associate, Textile Technology Institute, Bandung, 1980-1984; under graduate instructor, Textile Technology Institute, Bandung, 1974-1986; under graduate teaching assistant, Textile Technology Institute, Bandung, 1963-1973;

Professional Organization: Ikatan Ahli Tekstil Seluruh Indonesia (Indonesian Textile Expert Association), American Society for Quality Control, American Statistical Association.