

01168

10
20J

Universidad Nacional Autónoma de México

División de Estudios de Posgrado
de la Facultad de Ingeniería

**METODOLOGIA PARA EL ANALISIS
DE SUPERFICIE DE RESPUESTA**

T E S I S

que para recibir el grado de:

**MAESTRO EN INGENIERIA
(INVESTIGACION DE OPERACIONES)**

Presenta

DINO MON VÁSQUEZ

México, D.F. 1993

**TESIS CON
FALLA DE ORIGEN**



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

RESUMEN

La técnica de *Superficie de Respuesta* se presenta como un método que permite optimizar procesos físicos, químicos, biológicos o sociológicos donde se mide una variable cuantitativa que depende de uno o más factores cuantitativos. Se presentan las bases lógicas para postular modelos de respuesta. Se incluye el procedimiento que permite caracterizar la forma de la superficie de respuesta y encontrar los valores críticos (mínimos o máximos). Se discute el método de ascenso por pendiente máxima y su valor en la experimentación secuencial para determinar la región donde se encuentra el punto crítico. Se discuten previamente algunos diseños experimentales útiles en el ajuste de los modelos de Superficie de Respuesta.

Contenido

INTRODUCCION	iv
1 PRELIMINARES	1
1.1 CONCEPTOS BASICOS	3
1.2 EL AJUSTE	6
1.3 EL DISEÑO	14
1.3.1 Diseños para Estimar una Superficie de Primer Grado	15
1.3.2 Diseños Para Estimar Polinomios de Segundo Orden	31
1.3.3 Otros diseños más generales:	35
1.4 METODOS DE BUSQUEDA DEL OPTIMO	37
1.4.1 El Método Aleatorio	37
1.4.2 El Método del Factor Único	38
1.4.3 El Método de Experimento Único	40
1.4.4 Comentarios Sobre Los Diferentes Métodos	40
2 Cálculo de Condiciones de Operación Óptimas	43
2.1 DETERMINACIÓN DE PUNTO CRÍTICO	44
2.2 TRANSFORMACION CANONICA	49
2.3 METODO DE ASCENSO POR PENDIENTE MAXIMA	65
2.3.1 Ascenso Por Pendiente Máxima: (A.P.M)	65
2.4 MODIFICACION MYERS-KHURI (1979) PARA EL A.P.M	90
2.5 ALGUNAS VARIANTES AL METODO A.P.M.	95
2.5.1 Obtención de la Longitud del Paso t.	96

2.5.2 Localización del Centro del Nuevo Diseño, c.	98
3 Análisis De Cordilleras	100
3.1 EL ANALISIS DE COORDILLERAS	101
3.2 ANALISIS DE COORDILLERAS MODIFICADO	127
COMENTARIOS FINALES	142
APENDICE A	144
APENDICE B	146
Referencias	147

INTRODUCCION

En la industria, en la agricultura, en la administración y en otras muchas áreas, es frecuente encontrar situaciones en las que se requiere conocer la combinación de los niveles de un conjunto de factores que optimiza un resultado específico.

No siempre se necesita recurrir a la estadística para tomar una decisión; sin embargo, existen casos en los que por su misma complejidad, es imposible hacerlo sin un análisis estadístico, largo y minucioso, para obtener la combinación óptima de los factores. Por ejemplo: nadie, o casi nadie, pensaría en obtener una aleación para recubrir la cabeza del próximo cohete que será lanzado al espacio, sin antes haber realizado una serie de experimentos, tendientes a encontrar la combinación de materiales que proporcionan una mayor resistencia a las presiones a que estará expuesto.

A la serie de acciones (desde la elección del *diseño experimental*), para encontrar la combinación de factores donde se alcanza un máximo (o un mínimo, en su caso), se le conoce como *Metodología de Superficies de Respuesta*.

Esta técnica se desarrolló principalmente en el área de ingeniería química, conjugando técnicas de diseños experimentales y de optimización. Así el primer reporte aparece por Box y Wilson en 1951, quienes establecen el método de ascenso por pendiente máxima para encontrar un estimador de la combinación de factores óptima. Rápidamente crecieron las modalidades de la técnica original y sus aplicaciones prácticas, algunas de estas se presentan a continuación:

EN LA PRODUCCION DE CERAMICA: En esta rama industrial resulta de interés conocer la proporción de ingredientes, la temperatura y el tiempo de cocción necesario con los que se obtienen piezas con mayor dureza, brillantez o suavidad al tacto.

EN LA INDUSTRIA FARMACEUTICA: En este campo hay múltiples procesos en los que la metodología de superficies de respuesta tiene aplicación, por ejemplo: en la elaboración de un analgésico es necesario conocer qué proporción de las diferentes sustancias produce un efecto calmante más efectivo, o bien, minimiza los efectos colaterales no deseados.

EN LA INDUSTRIA AUTOMOVILISTICA: Aquí resulta de interés conocer, por ejemplo, el tiempo de elaboración necesario y la proporción en que se deben mezclar los materiales para obtener neumáticos con mayor resistencia al desgaste.

EN LA FRUTICULTURA: En el tratamiento para la comercialización y conservación de la fruta es importante conocer el tiempo de corte, la cantidad de cera para recubrir la superficie y la temperatura a la que debe mantenerse una determinada fruta para que el grado de maduración se obtenga en un tiempo máximo.

Estos ejemplos además de resaltar la relevancia práctica que tiene la Metodología de Superficies de Respuesta, permiten establecer, en general, el problema y los objetivos a cubrir con dicha metodología.

En cada uno de los ejemplos, el experimentador sabe que existe una relación entre una variable, que denomina la respuesta del proceso en estudio (en el caso del primer ejemplo; la dureza, brillantez o suavidad al tacto, podrían ser variables de respuesta al proceso producción de cerámica), y un conjunto (quizás numeroso) de variables, al que denomina conjunto de variables explicativas (en el caso del primer ejemplo, podrían ser la proporción de cada uno de los ingredientes, el tiempo y la temperatura de cocción para producir las piezas en el proceso), pero no conoce la expresión analítica de la relación. De aquí, que el primer objetivo del experimentador es aproximar de manera eficiente dicha relación.

En cada ejemplo, es notorio el hecho de que las variables en juego (respuesta y explicativas) se miden en escala continua. Por lo cual, el experimentador sólo conoce que el soporte¹ de cada una de las variables es toda la recta real.

En el desarrollo de este trabajo, se apreciará que la aproximación de la relación funcional entre la variable de respuesta y las variables explicativas depende de la elección de ciertos valores de estas últimas. Dichos valores son elegidos de entre la infinidad que componen cada uno de los soportes de las variables explicativas, es decir, la recta real para cada variable. Por lo cual, el experimentador debe construir un diseño de puntos o valores (diseño experimental) en base a cada eje descrito por cada variable explicativa, para luego hacer el ajuste o aproximación por medio de una relación funcional.

Aún más, si el experimentador después de realizar el ajuste desea encontrar el valor máximo o mínimo que la variable de respuesta adquiere en la relación, necesita pensar que esta aproximación depende en gran parte del diseño de puntos que construyó anteriormente. Por lo cual, si el experimentador encuentra un máximo o mínimo de la relación funcional, en base a cualquier procedimiento de cálculo, puede que éste sea

¹ El soporte es el conjunto de valores que puede tomar una variable específica

un óptimo local y no global, debido a que se encuentra dentro de una región de puntos previamente seleccionada. Por tanto, el objetivo final del experimentador después de haber decidido como hará el ajuste o aproximación de la relación funcional, es diseñar un método que de manera eficiente obtenga regiones de puntos en las variables explicativas, haga ajustes de la relación funcional y evalúe el punto óptimo en cada una de las regiones, comparando los óptimos locales hasta obtener el óptimo global.

Obviamente será necesario un criterio de terminación para el método. Este debe considerar la búsqueda del óptimo global a través de óptimos locales en diferentes regiones, de alguna manera exhaustiva.

De 1951 a 1971 no existía un texto sobre Superficie de Respuesta, aunque algunos textos de diseños como Cochran y Cox; Davies, Hicks, etc., incluyen un capítulo al respecto. En 1971 apareció el libro de R. Myers con un tratamiento bastante completo del tema. Box y Hunter (1954) establecen una región de confianza para la combinación óptima de factores basada en el teorema de Fieller. Box y Hunter (1957) Analizan el comportamiento de la varianza del predictor de la función de respuesta, con relación a los diseños y definen el concepto de diseños rotables. Myers y Khuri (1979) proporcionan un nuevo método de ascenso por pendiente máxima. Carter, Chinchilli, Campbell y Wampler (1984); y Carter, Chinchilli, Myers y Campbell (1986) proporcionan una nueva región de confianza. Notz (1988) propuso los diseños D-eficientes, lo que originó el concepto de diseños D-óptimos, abordado por Atkinson y Donev (1989) entre otros. Una excelente relación, muy detallada, de los trabajos que se han hecho sobre superficies de respuesta fue hecha por Myers, Khuri y Carter (1988).

Las bases lógicas y técnicas de la metodología de Superficie de Respuesta, así como la notación, son semejantes a la de los modelos lineales en general. Estas son el considerar una variable cuantitativa, Y , se le denomina la *respuesta* o la variable dependiente, dado que, depende de una serie de variables independientes (explicativas) cuantitativas, x_i y parámetros conocidos β_k , con $i = \overline{1, p}$, $k = \overline{1, l}$; p y l son arbitrarios. Esto es, se postula que

$$Y = \eta(\mathbf{x}, \boldsymbol{\beta}) + \epsilon$$

donde $\mathbf{x} = (x_1, x_2, \dots, x_p)$ y $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_l)$. Sin embargo, usualmente la función $\eta(\mathbf{x}, \boldsymbol{\beta})$ es: a) desconocida; b) conocida de forma, pero no sus parámetros; c) conocida,

²La notación $i = \overline{1, p}$, corresponde a $i = 1, \dots, p$.

pero muy complicada para manejarse analíticamente. Entonces se considera que la función η se puede aproximar satisfactoriamente con modelos sencillos, por lo menos en las regiones experimentales de interés. Región experimental o de diseño es el rango de valores de las variables x^i , restringido a los valores que toman las x en un diseño dado. Los valores de las x^i se determinan mediante un diseño determinado, buscando economía y precisión en la estimación de la función η aproximada.

En este trabajo se aborda uno de los aspectos más relevantes de la metodología de superficies de respuesta: El método de estimación del óptimo. Aunque cabe señalar que el diseño de la región experimental en cada etapa del análisis y la construcción de regiones de confianza del óptimo, son aspectos de vital importancia; para un mayor análisis de estos dos últimos puntos remitase a [27, 39].

El objetivo fundamental de este trabajo es brindar un panorama lo más completo posible de la metodología de superficies de respuesta, y dar los elementos que sirvan como apoyo en la resolución de problemas cuyo contexto lo permita. El trabajo se conforma de 3 capítulos:

En el primero se hace una recapitulación de los aspectos relevantes escritos sobre el tema. Se enlistan, sin ser exhaustivos, una serie de diseños para estimar funciones de primer y segundo grado, así como los diferentes métodos conocidos para estimar el óptimo.

En el segundo se estudia exhaustivamente el método de la Transformación Canónica que describe el problema de optimizar un polinomio de segundo orden dejando clara la necesidad de desarrollar técnicas de aplicación previas y posteriores a la Transformación Canónica como son El Método de Ascenso por Pendiente Máxima, que se presenta en el mismo capítulo y El Análisis de Cordilleras presentado en el Capítulo 3.

Es importante mencionar que el Método de Ascenso por Pendiente Máxima tiene como objetivo encontrar una mejor región experimental, en el sentido de que en ésta se encuentren las combinaciones de las variables independientes que maximizan las respuesta promedio estimada. En el capítulo 2, también se presentan algunas modificaciones recientes al método de ascenso por pendiente máxima realizadas con el fin de mejorar la precisión del método.

El Análisis de Cordilleras presentado en el capítulo 3, cuyo desarrollo es similar al del Método de Ascenso por Pendiente Máxima, también puede expresarse como un procedimiento para encontrar las condiciones operativas óptimas; el Análisis de

Coordilleras muchas veces se emplea cuando con la Transformación Canónica no se pueden determinar dichas condiciones.

Es importante recalcar que este trabajo intenta desarrollar, más que diferentes procedimientos, una metodología completa la cual puede resultar provechosa si se aplica correctamente.

Sin embargo también se elaboró cada capítulo de manera que pueda ser leído recurriendo lo menos posible a otros capítulos.

Si se desea profundizar en el estudio de los diseños 2^k y Análisis de Regresión Lineal Múltiple. Se recomienda para su estudio [21], [12] y [16].

Capítulo 1

PRELIMINARES

La metodología de superficie de respuesta (M.S.R.) comprende un conjunto de técnicas estadísticas y matemáticas cuyo objetivo principal, es organizar la información experimental producida en la observación de un proceso, construyendo un modelo empírico y explotando el mismo. Principalmente se ha desarrollado dentro de la operación industrial y en aquellas áreas en donde se permite la repetición de experimentos.

En general, la metodología de superficie de respuesta se utiliza en situaciones donde un gran número de variables explicativas presentes en un proceso tienen influencia sobre una o más características del proceso bajo estudio.

La característica (rendimiento, costo de producción, etc.) es llamada la *Respuesta* o variable de interés y se denotará por Y ; ésta es normalmente medida en una escala continua y frecuentemente representa la función más importante de los procesos, de aquí que no sea excluyente la posibilidad de tener más de una respuesta. Las variables que afectan a la respuesta o variables independientes denotadas por x_1, x_2, \dots, x_n también son cuantitativas y sus niveles están sujetos al control del científico o experimentador.

Para hacer más clara la comprensión de los diferentes conceptos que se han manipulado, considere el siguiente ejemplo:

Un laboratorio dedicado al desarrollo de nuevos productos considera la creación de cierto solvente para limpieza y desea probar su nivel máximo de abrasividad. La abrasión del solvente se mide frotando una toalla humedecida con una porción de solvente contra un azulejo de cerámica por 30 minutos, registrando la cantidad desgastada del azulejo (desgaste=grosor del azulejo antes de ser frotado menos el grosor del azulejo después de ser frotado). El principal ingrediente en el solvente es lima agria, y se conoce que la abrasividad del solvente es una función que crece y luego decrece con respecto al

porcentaje de lima agria. El científico desconoce cuál es la máxima abrasión causada al azulejo debido a la fricción y el porcentaje de lima agria en el solvente. Por otro lado el científico sabe que, demasiada lima agria en el solvente no sólo es costosa sino que también puede llegar a reducir el nivel de abrasividad del solvente. Para determinar el valor de abrasión máxima y el porcentaje de lima agria que lo produce en un tiempo fijo de 30 minutos de fricción, un programa de experimentos que consiste en exponer a fricción por 30 minutos azulejos a diferentes concentraciones de lima agria en el solvente es realizado. El valor de abrasión es registrado para cada una de las intensidades del solvente, y una ecuación matemática se desarrolla para expresar la abrasión como una función del porcentaje de lima agria. Se hacen algunas estimaciones para algunos parámetros desconocidos que estan en la relación funcional de la variable abrasión obteniendose el valor máximo de la misma y el correspondiente porcentaje de lima agria que lo produce.

Como se observa en el ejemplo, la variable de interés o de respuesta es la abrasividad y la variable independiente o explicativa de la respuesta es el porcentaje de lima agria. Algún especialista podría pensar que el porcentaje de lima agria en el solvente no sería el único factor que determine el nivel de la abrasividad, de hecho, en la práctica, dificilmente se podrían determinar todos los factores que influyen en la respuesta Y y aunque se pudieran especificar, es común cometer errores en su medición.

Por lo cual formular matemáticamente el problema de superficie de respuesta, es pensar en la relación funcional

$$Y = \eta(x_1, x_2, \dots, x_n) + \epsilon,$$

donde ϵ es un error aleatorio debido a las fallas en la medición y la incapacidad para determinar todos los factores que afectan a Y . η especifica la forma en que los factores x_1, x_2, \dots, x_n afectan la respuesta. η es conocida como la respuesta promedio.

Del ejemplo anterior es plausible ubicar las 3 áreas que comprende la Metodología de Superficie de Respuesta:

- a.- Definir el diseño experimental que resulte adecuado y confiable para medir la respuesta.
- b.- La determinación del modelo matemático que mejor se ajuste a la colección de datos del Diseño seleccionado en (a), mediante apropiadas pruebas de hipótesis concernientes a los parámetros desconocidos asociados al modelo y

- c.- Determinar el conjunto de factores óptimos del experimento que producen el máximo (mínimo) valor de respuesta.

Todo esto llevaría al experimentador a enfrentar las siguientes dificultades:

- i.- ¿Qué diseño utilizar?
ii.- ¿ Qué modelo ajustar?

[33] y [27] dejan plasmado en sus textos sobre Superficie de Respuesta, la relación existente entre las dos preguntas anteriores. *De acuerdo al modelo se propone un diseño.* Aunque el objetivo de este trabajo no es contestar (i) y (ii), en este capítulo se harán descripciones muy superficiales acerca de los diseños de experimentos que muy comúnmente se utilizan en la metodología de superficies de respuesta y en particular en este trabajo. De igual forma se hará una revisión general sobre el ajuste de modelos, en particular, los modelos lineales múltiples de regresión.

Otro de los objetivos de las técnicas que se plantean en Superficie de respuesta surge cuando las condiciones de optimalidad en (c) son inadecuadas o no se pueden determinar. Aprovechar la información dada por el diseño y ajuste resulta un aspecto de interés.

La parte central de este trabajo aborda la teoría necesaria para determinar condiciones de optimalidad y dar diagnósticos en caso de conclusiones no satisfactorias, además de exhibir algunos métodos recientes que se han propuesto como avances a la teoría existente y que se mostrarán en capítulos posteriores.

En este capítulo se intentan enfatizar las principales suposiciones que involucran las técnicas de superficie de respuesta como un enfoque en la solución de problemas. Al mismo tiempo, se espera que el lector adquiera una mejor comprensión de los objetivos reales de la metodología de superficies de respuesta y visualice la clase de problemas para los cuales la metodología puede aplicarse.

1.1 CONCEPTOS BASICOS

Se supone que el experimentador está interesado en un sistema que envuelve alguna respuesta Y que depende de variables independientes o explicativas $\xi_1, \xi_2, \dots, \xi_k$. Una distinción se puede hacer en este punto entre variables naturales (ξ^i) y codificadas o

variables de diseño (x^i), las primeras son las mediciones que recava el investigador mientras que las segundas son simples transformaciones lineales de las primeras y son las que realmente se utilizan dentro de los modelos que se proponen para explicar el experimento. En ([27], pág. 46) se describe el procedimiento para pasar de variables naturales a codificadas y viceversa. Además se supone que las ξ^i s pueden ser controladas por el experimentador con error pequeño. Como ejemplo de todo lo anterior, considere una situación en donde un ingeniero químico está interesado en el rendimiento, Y , de una reacción química. El rendimiento depende de la temperatura de reacción (ξ_1), la presión de la reacción (ξ_2), concentración de uno de los reactantes (ξ_3), etc. Se propone entonces un modelo de la forma $Y = \eta(x_1, x_2, x_3) + \epsilon$, donde cada x_i variable de diseño esta relacionada linealmente con la respectiva ξ_i , $i = \overline{1,3}$ y ϵ es un error aleatorio debido a las fallas en medición e incapacidad para determinar todos los factores que afectan a Y . En general se tiene,

$$Y = \eta(x_1, x_2, \dots, x_k) + \epsilon = \eta(\mathbf{x}) + \epsilon$$

donde la forma de η es desconocida y quizás extremadamente complicada, lo que lleva a la necesidad de aproximarla en alguna forma. De aquí que el éxito de la M.S.R. dependa de la aproximación de η por alguna ley en alguna región de las variables independientes. Si se considera a $\eta(\mathbf{x})$ como una función continua y suave, entonces es posible representar ésta localmente con cualquier grado de aproximación requerida, haciendo uso de la expansión en Series de Taylor alrededor de algún punto arbitrario \mathbf{x}_0 , esto es

$$\eta(\mathbf{x}) = \eta(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^t \nabla \eta(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^t H(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \dots, \quad (1.1.1)$$

donde $\nabla \eta$ es el gradiente o matriz de primeras derivadas y H es el Hessiano o la matriz de segundas derivadas de la función.

De esta forma si se decide aproximar con un hiperplano alrededor del punto \mathbf{x}_0 se tendría

$$\eta(\mathbf{x}) = \eta(\mathbf{x}_0) + \left(\frac{\partial \eta}{\partial x_1} \right)_{\mathbf{x}_0} x_1 + \left(\frac{\partial \eta}{\partial x_2} \right)_{\mathbf{x}_0} x_2 + \dots + \left(\frac{\partial \eta}{\partial x_k} \right)_{\mathbf{x}_0} x_k.$$

observe que el modelo puede ser escrito como

$$\eta(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1.1.2)$$

donde $\beta_i = \left(\frac{\partial \eta}{\partial x_i} \right)_{\mathbf{x}_0}$.

Este modelo es conocido como de primer orden y frecuentemente es usado cuando el experimentador está interesado en estudiar η en una vecindad cercana de x_1, x_2, \dots, x_k , esto es, donde se presenta sólo una pequeña curvatura de η .

Por otro lado, el experimentador podrá usar una aproximación de segundo orden para la función η tomando además los términos correspondientes al Hessiano en la ecuación (1.1.1), y en este caso se tendría una expresión del estilo

$$\eta(\mathbf{x}) = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \underbrace{\sum_i \sum_j}_{i < j} \beta_{ij} x_i x_j. \quad (1.1.3)$$

La fase de optimización del problema frecuentemente involucra la obtención de valores x_1, x_2, \dots, x_k que maximicen la respuesta. Esto es cierto particularmente en situaciones químicas donde la respuesta en cuestión es muchas veces algún tipo de variable de rendimiento la cual se espera lo más grande posible. Debe notarse que los procedimientos de superficie de respuesta no se usan principalmente con el propósito de permitirle al experimentador comprender el mecanismo que está detrás del sistema o proceso, aunque éstos, en algunos casos puedan llevarlo a este conocimiento. Más bien, su propósito es determinar cuáles son las condiciones de operación óptimas, o determinar una región del espacio total de las variables independientes (factores) en las cuales ciertas especificaciones de operación sean encontradas.

Aunque el objetivo eventual del análisis de superficies de respuesta es usualmente contestar ciertas preguntas concernientes a las condiciones de operación del sistema; es extremadamente importante que la decisión sea tomada, en función al diseño experimental de puntos utilizado, esto es, qué niveles de la variables independientes fueron considerados en el proceso experimental.

Los coeficientes β_0, β_1, \dots en los modelos dados en (1.1.2) y (1.1.3) son estimadores de los datos recolectados por el experimentador. Esta estimación puede ser desarrollada con mayor efectividad si se ha contestado convenientemente la pregunta de qué diseño experimental utilizar. Quedará claro en el siguiente capítulo, que una vez elegido el diseño experimental de puntos se apliquen técnicas de análisis de regresión para estimar los coeficientes β_i . Esto último debido a que la respuesta se aproxima mediante una superficie de cierto grado y ésta es una función lineal en los parámetros y por tanto

linealizable.

A veces, modelos de orden mayor que dos son usados, pero no es común. El lector puede percibir inmediatamente de lo anterior, el por qué se acuño el término *Superficie de Respuesta* para esta metodología. Si $k = 2$, y si se desea caracterizar una función de respuesta de segundo orden gráficamente, un procedimiento conveniente es graficar las curvas de nivel de la respuesta, y es visible que la función de respuesta para $k > 2$ se caracteriza por curvas de nivel que son superficies.

Hasta aquí se han hecho sólo 3 suposiciones fundamentales y en las cuales se basará este trabajo a lo largo de su desarrollo:

- 1.- Una estructura $\eta(\mathbf{x})$ existe y puede ser complicada y desconocida. Las variables que están involucradas en el modelo son cuantitativas y medidas en escala continua.
- 2.- La función η puede ser aproximada por un polinomio de orden menor en una región de interés, tales como los descritos en las ecuaciones (1.1.2) y (1.1.3).
- 3.- Las variables independientes x_1, x_2, \dots, x_k son controladas en el proceso de observación y medidas con error pequeño.

1.2 EL AJUSTE

Fundamentalmente, el problema de Superficie de Respuesta centra su interés en alguna respuesta, Y , la cual es una función de k variables explicativas, x_1, x_2, \dots, x_k , esto es

$$Y = f(x_1, \dots, x_k). \quad (1.2.1)$$

La forma real de f en la ecuación (1.2.1), frecuentemente se desconoce, pero se supone puede ser aproximada mediante una función polinomial de bajo orden.

Suponga que la función f en la ecuación (1.2.1) se aproxima por un modelo lineal en las variables x^s , aunque la discusión que sigue, puede ser alterada fácilmente para ser consistente con una aproximación de mayor orden. Suponga también, que n corridas experimentales se hacen sobre varias combinaciones de las x^s , dichas combinaciones

son determinadas por el experimentador. Los datos se escriben en la forma

$$\begin{array}{cccccc} Y_1 & x_{11} & x_{21} & \cdots & x_{k1} & \\ Y_2 & x_{12} & x_{22} & \cdots & x_{k2} & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Y_n & x_{1n} & x_{2n} & \cdots & x_{kn} & \end{array}$$

donde $n > k$. El plan de elección de niveles experimentales en las $x^{i's}$ (si existe en realidad un plan) es llamado el *diseño experimental*. El modelo supuesto por el experimentador puede ser escrito como

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \epsilon_i, \quad (1.2.2)$$

con $i = \overline{1, n}$, y donde ϵ_i es una variable aleatoria (v.a). Se supone que los errores, ϵ_i , son estocásticamente independientes de corrida a corrida, con media cero y varianza σ^2 .

El modelo de la ecuación (1.2.2), puede ser escrito convenientemente en forma matricial como

$$Y = X\beta + \epsilon, \quad (1.2.3)$$

aquí desde luego,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_k \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdots \\ \epsilon_n \end{bmatrix},$$

y

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}.$$

El modelo de la ecuación (1.2.3), se conoce como el *Modelo de regresión lineal general múltiple*.

Este modelo puede aplicarse fácilmente a modelos polinomiales de grado mayor a uno. Por ejemplo, suponga un modelo cuadrático en 2 variables, esto es, la respuesta para la i -ésima corrida que envuelve los niveles (x_{1i}, x_{2i}) es

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + \epsilon_i,$$

donde $i = \overline{1, n}$, con $n \geq 6$. De aquí, si se define

$$\begin{bmatrix} z_{1i} \\ z_{2i} \\ z_{3i} \\ z_{4i} \\ z_{5i} \end{bmatrix} = \begin{bmatrix} x_{1i} \\ x_{2i} \\ x_{1i}^2 \\ x_{2i}^2 \\ x_{1i}x_{2i} \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{bmatrix}$$

se tiene el modelo

$$Y_i = \alpha_0 + \alpha_1 z_{1i} + \alpha_2 z_{2i} + \alpha_3 z_{3i} + \alpha_4 z_{4i} + \alpha_5 z_{5i} + \epsilon_i$$

con $i = \overline{1, n}$, que es equivalente al modelo descrito en la ecuación (1.2.2) y en forma matricial al definido por la ecuación (1.2.3).

Como se puede observar, este procedimiento de definición, se generaliza a cualquier modelo polinomial de grado mayor a uno, convirtiendo el problema en uno de regresión lineal general múltiple. Se observa entonces, que para cada problema en donde se quiera ajustar un polinomio de alto grado, es necesario una transformación de los datos originales, similar a la realizada anteriormente.

Este procedimiento de transformación de un modelo polinomial a uno de regresión lineal múltiple es posible, debido a que los modelos polinomiales son *linealizables*, esto es, *son modelos lineales en sus parámetros*. También existen otros modelos reportados en la literatura, que son de igual manera linealizables, un listado de éstos viene dado en ([31], pág. 91).

Una vez supuesto el modelo lineal general múltiple de la ecuación (1.2.3), el foco de atención está dirigido al problema de estimación del vector de parámetros β . El método de mínimos cuadrados es un procedimiento de estimación útil, particularmente para modelos de la forma que presenta la ecuación (1.2.3). El método de mínimos cuadrados usa como estimador de β , el vector que resulta de la minimización de la función

$$L = \sum_{i=1}^n \epsilon_i^2 = \epsilon^t \epsilon,$$

llamada la suma de cuadrados de los errores o desviaciones entre la respuesta observada y la estimada, muy comúnmente denotada como SCE. L puede escribirse como

$$L = (Y - X\hat{\beta})^t (Y - X\hat{\beta})$$

$$\begin{aligned}
&= Y^t Y - (X\hat{\beta})^t Y - Y^t X\hat{\beta} + (X\hat{\beta})^t (X\hat{\beta}) \\
&= Y^t Y - \hat{\beta}^t X^t Y - Y^t X\hat{\beta} + \hat{\beta}^t X^t X\hat{\beta} \\
&= Y^t Y - 2\hat{\beta}^t X^t Y + \hat{\beta}^t X^t X\hat{\beta},
\end{aligned} \tag{1.2.4}$$

ya que $\hat{\beta}^t X^t Y$ es un real.

Se encuentra el vector $\hat{\beta}$ que hace que la función L alcance un mínimo, utilizando la proposición 1 del Apéndice A. Así

$$\frac{\partial L}{\partial \beta} = -2X^t Y + 2(X^t X)\beta,$$

igualando a cero y resolviendo para β , se tiene

$$(X^t X)\hat{\beta} = X^t Y, \tag{1.2.5}$$

dicho sistema se conoce como *ecuaciones normales*. Suponiendo que $X^t X$ es no singular, el estimador de mínimos cuadrados para β es

$$\hat{\beta} = (X^t X)^{-1} X^t Y.$$

La función L no puede ser maximizada por el valor de $\hat{\beta}$, dado que ésta es estrictamente convexa, por tanto $\hat{\beta}$ es el valor que hace a la función L mínima.

En cualquier proceso de estimación, es necesario obtener propiedades para el estimador, en vías de medir qué tan bueno es.

La primera de estas propiedades, consiste en mostrar que en promedio, la estimación coincide con el verdadero valor del parámetro. Esta propiedad se conoce como *Insesgamiento*. Para este caso

$$\begin{aligned}
E(\hat{\beta}) &= E\left[(X^t X)^{-1} X^t Y\right] \\
&= (X^t X)^{-1} X^t E(Y) \\
&= (X^t X)^{-1} X^t X\beta \\
&= \beta,
\end{aligned} \tag{1.2.6}$$

utilizando la proposición 2 inciso a) del Apéndice A. De aquí que, cada elemento del vector de estimadores sea insesgado.

Otra propiedad de los estimadores que frecuentemente es importante investigar en el desarrollo de experimentos para superficies de respuesta, es el efecto del diseño de puntos sobre la matriz de varianzas y covarianzas de $\hat{\beta}$, es decir, analizar la manera de obtener mayor precisión en la estimación de $\hat{\beta}$, utilizando un diseño de puntos que minimice la matriz de varianzas y covarianzas del mismo. La matriz de varianzas y covarianzas de $\hat{\beta}$ es

$$\begin{aligned}
 V(\hat{\beta}) &= V\left[(X^t X)^{-1} X^t Y\right] \\
 &= (X^t X)^{-1} X^t V(Y) (X^t X)^{-1} \\
 &= (X^t X)^{-1} X^t \sigma^2 I X (X^t X)^{-1} \\
 &= \sigma^2 (X^t X)^{-1}, \tag{1.2.7}
 \end{aligned}$$

utilizando la proposición 2 inciso c) del Apéndice A. Esto implica que las varianzas de los elementos del vector $\hat{\beta}$, están dadas por los elementos de la diagonal de $(X^t X)^{-1}$, cada uno multiplicado por σ^2 . Las covarianzas entre los elementos del vector $\hat{\beta}$, están dadas por los elementos que están fuera de la diagonal de la matriz $(X^t X)^{-1}$ multiplicados por σ^2 . Por lo cual, para minimizar $V(\hat{\beta})$, se requiere escoger un diseño de puntos x^s que haga que la matriz $(X^t X)$ tenga una estructura adecuada, esto último se mostrará en la siguiente sección.

Es posible calcular un estimador para la varianza del error, σ^2 , pidiendo una propiedad distribucional adicional para el vector de errores. Esto es, que ϵ tenga una distribución normal multivariada de dimensión n , con vector de medias cero y matriz de varianzas y covarianzas $\sigma^2 I$, como se supuso anteriormente. Así el estimador de la varianza del error está dado por la expresión: ([31], pág. 129)

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^t (Y - X\hat{\beta})}{n - p} = \frac{\epsilon^t \epsilon}{n - p},$$

donde n es el número de observaciones de la respuesta, Y , y p el número de parámetros en el modelo. Como se observa, el numerador de la expresión anterior es la suma de cuadrados del error, SCE, misma que es igual por la ecuación (1.2.4) a

$$\begin{aligned} SCE &= Y^t Y - 2\hat{\beta}^t X^t Y + \hat{\beta}^t X^t X \hat{\beta} \\ &= Y^t Y - \hat{\beta}^t X^t Y, \end{aligned}$$

dado que $X^t X \hat{\beta} = X^t Y$. Entonces el estimador de la varianza se escribe de manera simple como

$$\hat{\sigma}^2 = \frac{SCE}{n - p}. \quad (1.2.8)$$

Un problema cuando se realiza cualquier tipo de ajuste, es que no hay garantía de que el modelo que se está aproximando a los datos sea correcto, esto es, puede que se presente *carencia en el ajuste*. Por ejemplo, si se tiene una respuesta, Y , que depende de una variable explicativa y se desea hacer un ajuste lineal del tipo $Y = \beta_0 + \beta_1 x$, a datos que gráficamente son como los mostrados en la figura (1.2.1) a), el ajuste se visualizaría como el presentado en la figura (1.2.1) b). Se aprecia que un modelo lineal

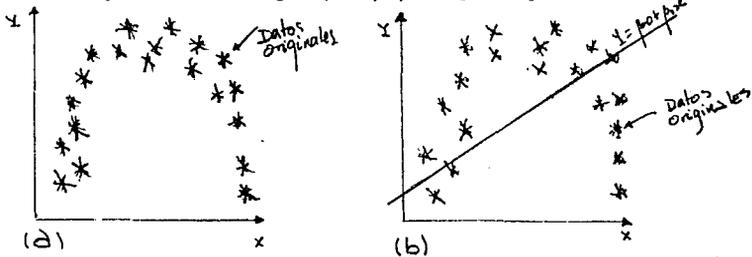


Figura 1.2.1:

no es adecuado para el tipo de datos que se tiene, es decir, hay carencia de ajuste. Esto último, debido a que para una o dos variables explicativas es posible graficar los datos en dos o tres dimensiones y observar qué tipo de modelo sería más conveniente ajustar. Cuando se tienen k variables explicativas, $k > 2$, una interpretación gráfica de los datos es imposible y por tanto no se logra medir visualmente la carencia de ajuste. Por ende, es necesario una prueba analítica que permita detectar la existencia o no de carencia de ajuste, en particular si se ajustan modelos de regresión lineal múltiple.

Se plantea entonces, un contraste estadístico de hipótesis del tipo

$$H_0^{(1)}: \text{No hay carencia de ajuste.} \quad - \text{vs} \quad - \quad H_1^{(1)}: \text{Hay carencia de ajuste.}$$

Obteniéndose como se verá mas adelante, una región crítica para la hipótesis $H_0^{(1)}$, que a un nivel de significancia α permite hacer el contraste.

Si al realizar la prueba anterior **no se rechaza la hipótesis nula** se prosigue con el análisis, ahora con el fin de verificar la significancia global del modelo. Esto es, probar que el vector de parámetros β , puede considerarse distinto del vector cero, ya que de no ser así el modelo ajustado sería una plano constante, lo cual no es razonable para un conjunto normal de datos.

Así se plantea el contraste

$$H_0^{(2)}: \beta = 0 \quad - \text{vs} - \quad H_1^{(2)}: \beta \neq 0$$

Se encuentra también una región crítica que a un nivel de significancia α lleva a rechazar la hipótesis nula.

El desarrollo de los dos contrastes de hipótesis anteriores, se basa en la suposición de que el vector de errores tiene una distribución normal multivariada, obteniéndose distribuciones normales multivariadas para el vector de observaciones, Y , y para el vector de parámetros estimados, $\hat{\beta}$, así como una distribución ji-cuadrada para el estimador de la varianza del error.

Obtenidas todas las propiedades distribucionales anteriores, se realizan cocientes de verosimilitudes generalizados, para desarrollar las pruebas de hipótesis, obteniéndose lo que se conoce en la literatura como la *Tabla de Análisis de Varianza* (ANOVA), que permite realizar el contraste. Los detalles completos del desarrollo de las pruebas pueden ser vistos en ([42], caps. 2 y 3; [16], cap. 2).

La tabla de Análisis de varianza es

TABLA ANOVA				
FUENTE	SUMA DE CUADRADOS	G.L. ¹	CUAD. MEDIO	F
REGRESION	$\hat{\beta}^t X^t Y = (1)$	p	$\frac{(1)}{p} = (a)$	
ERROR	$Y^t Y - \hat{\beta}^t X^t Y = (2)$	$n - p$	$\frac{(2)}{n-p} = (b)$	$F_{SG} = \frac{(a)}{(b)}$
CAR. DE AJUS.	$\sum_{i=1}^n n_i (\bar{Y}_i - \bar{Y})^2 = (3)$	$m - p$	$\frac{(3)}{m-p} = (c)$	
ERROR PURO	$\sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = (4)$	$n - m$	$\frac{(4)}{n-m} = (d)$	$F_{CA} = \frac{(c)}{(d)}$
TOTAL	$Y^t Y$	n		

donde p es el número de parámetros a estimar, n_i es el número de observaciones sobre la variable de respuesta en el vector $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})$, m el número de vectores x_i

distintos en donde se mide la respuesta, n el número total de observaciones y \bar{Y}_i es el promedio de las observaciones obtenidas en el vector x_i .

Si se observa el renglón correspondiente a la etiqueta ERROR en la tabla ANOVA, la suma de cuadrados coincide con el numerador en el estimador para la varianza del error, σ^2 , esto es, coincide con la suma de cuadrados del error, SCE, dada en la ecuación (1.2.8). De aquí la importancia de poder estimar la varianza del error para poder realizar las pruebas de hipótesis correspondientes.

Para realizar la prueba de carencia de ajuste $H_0^{(1)}$ -vs- $H_1^{(1)}$, se calcula F_{CA} de la tabla ANOVA y se compara con el cuantil de orden α (nivel de significancia) en tablas de una distribución F -de Fisher con $(m - p, n - m)$ grados de libertad, denotando este último valor como $F_{(m-p, n-m)}^\alpha$. Si $F_{CA} > F_{(m-p, n-m)}^\alpha$ se rechaza la hipótesis nula, esto es, hay carencia de ajuste y se debe buscar otro tipo de modelo para ser ajustado al conjunto de datos. En otro caso se continúa utilizando el modelo ajustado.

Para realizar la prueba de significancia global, $H_0^{(2)}$ -vs- $H_1^{(2)}$, se calcula F_{SG} de la tabla ANOVA y se compara con el cuantil de orden α (nivel de significancia) en tablas de una distribución F -de Fisher con $(p, n - p)$ grados de libertad, denotando este último valor como $F_{(p, n-p)}^\alpha$. Si $F_{SG} > F_{(p, n-p)}^\alpha$, se rechaza la hipótesis nula de que $\beta = 0$, por lo tanto el modelo es significativo. En caso contrario son necesarias técnicas de reducción de dimensionalidad², para graficar los datos y comprobar que el conjunto de puntos ciertamente forma una superficie plana constante, si no hay evidencia de este hecho, se debe ajustar otro tipo de modelo a los datos.

²Técnicas de reducción de dimensionalidad tales como: componentes principales, escalamiento multidimensional, bi-plots, etc., pueden ser vistas en [43].

1.3 EL DISEÑO

La elección del diseño, es decir, la combinación de factores en donde se toma la muestra, va a responder a los diferentes intereses del investigador, como por ejemplo: La economía, (tanto por el número de observaciones, como por el número de cálculos implicados) y la precisión de la estimación, es decir, que la varianza del estimador de η sea pequeña [37].

Entonces, es conveniente contar con criterios para elegir aquellos diseños que sean *económicos* y *precisos*. Con la finalidad de establecer dichos criterios se discuten las características del modelo lineal para ajustar un polinomio de grado r a los datos ($r = 1, 2, 3, \dots$); con m factores en estudio.

El modelo lineal general es:

$$Y = X\beta + \epsilon \quad (1.3.1)$$

El estimador de mínimos cuadrados de β satisface la ecuación matricial $(X^t X)\hat{\beta} = X^t Y$ conocida como *sistema de ecuaciones normales*, y de esa manera el estimador del polinomio evaluado en (x_1, x_2, \dots, x_m) es

$$\hat{\eta}(x) = X\hat{\beta},$$

donde $X^t = (1, x_1, x_2, \dots, x_m, x_1^2, x_1 x_2, \dots, x_m^2, x_1^3, \dots, x_m^r)$.

El vector β , que siempre existe, es único si y sólo si la matriz $X^t X$ es invertible; esto es equivalente a que el rango de X sea igual a $k + 1$, por lo que se requiere de al menos $k + 1$ puntos para estimar a $\eta(x)$ de manera única.

De aquí se sigue que el diseño más económico es por su tamaño, el que tiene $k + 1$ puntos con X de rango completo; y, por los cálculos numéricos involucrados, el que proporciona una matriz $X^t X$ diagonal o con mayor número de ceros, lo que facilita su inversión.

Por otro lado, aplicando la proposición 2 inciso c) del Apéndice A, se obtiene que la magnitud de la varianza del estimador de η ,

$$V(\hat{\eta}) = X^t (X^t X)^{-1} X \sigma^2$$

determina la precisión de las estimaciones, y $V(\hat{\eta})$, por regla general es más pequeña cuando se tienen más datos. Por lo que claramente se ve que los conceptos de economía

y precisión son opuestos y debe buscarse un equilibrio entre los dos conceptos al elegir el diseño.

1.3.1 Diseños para Estimar una Superficie de Primer Grado

En este apartado, se enumeran únicamente aquellos diseños que tradicionalmente han sido utilizados para estimar el modelo lineal general (1.2.3) cuando $r = 1$, o sea cuando la superficie descrita por η es un hiperplano.

El modelo que describe el comportamiento de las observaciones, en este caso, es

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_m x_{mi} + \epsilon_i \quad (1.3.2)$$

el cual tiene $m + 1$ parámetros β_i (además de σ^2 , la varianza de las observaciones). Por lo tanto para estimar los parámetros de este modelo es necesario un diseño con al menos $m + 1$ puntos, distribuidos de forma tal que la matriz X sea no singular. Si además, se tiene que $X^t X$ es diagonal, el diseño es llamado **ortogonal** y la varianza del estimador $\hat{\eta}(x)$ se verá minimizada ([27], págs. 75-76). Si se desea estimar a σ^2 es necesario, cuando menos, $m + 2$ puntos.

Dado que $V(\hat{\beta}) = \sigma^2(X^t X)^{-1}$ por (1.2.7), una consecuencia del uso de diseños ortogonales, es que los efectos de las m variables explicativas, medidos en base a los valores de β_i , $i = \overline{1, m}$, pueden ser calculados de manera independiente. Esto es debido a que los elementos que están fuera de la diagonal en la matriz de varianzas y covarianzas del vector $\hat{\beta}$ son cero. Analíticamente, si

$$Y = X\beta + \epsilon$$

y

$$X\beta = [X_1, X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

es una partición de β en dos subconjuntos y de X en matrices de diseño correspondiente. Las ecuaciones normales dadas en la ecuación (1.2.5) resultan

$$\begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_1^t X_2 & X_2^t X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1^t Y \\ X_2^t Y \end{bmatrix},$$

que implica

$$\hat{\beta}_1 = (X_1^t X_1)^{-1} X_1^t Y \quad y \quad \hat{\beta}_2 = (X_2^t X_2)^{-1} X_2^t Y,$$

dado que $X_1^t X_2 = 0$. Mismas que corresponden a las soluciones que resultan de ajustar por separado, los modelos $Y = X_1 \beta_1 + \epsilon_1$ y $Y = X_2 \beta_2 + \epsilon_2$.

En general el uso de diseños ortogonales, permite realizar inferencias por separado de los vectores de parámetros β_1 y β_2 , ya sean pruebas de hipótesis del tipo $C_i \beta_i = \gamma_i$, con $i = 1, 2$, o intervalos de confianza sobre $\lambda_i^t \beta_i$, con $i = 1, 2$.

Por tanto, dado que los diseños ortogonales de primer orden son óptimos, en el sentido de que proveen mínima varianza (más precisión) para los estimadores de los β_i , $i = \overline{1, m}$ y facilidad en los cálculos. Se restringe el estudio en este trabajo a la clase de diseños ortogonales de primer orden con al menos $m + 1$ puntos. Tales como los diseños factoriales 2^m , factoriales fraccionales 2^{m-k} y diseños simplex.

A. Diseño Factorial 2^m :

La clase más elemental de experimentos factoriales son los llamados 2^m , esto es, m factores de interés cada uno medido en 2 niveles. Este tipo de diseños es extremadamente útil en casos donde la situación experimental se representa adecuadamente por una relación de primer orden. En el análisis de superficies de respuesta, se utilizan en las etapas preliminares y sólo para poder determinar ciertas direcciones a seguir.

Usualmente se aplican dos tipos de notaciones para hacer mención a los tratamientos generados por el diseño. Por ejemplo, en un diseño factorial 2^3 , con factores A, B y C o bien (x_1, x_2, x_3) , la notación (1) es la observación para la cual, los tres factores se encuentran en su nivel bajo. La presencia de una letra minúscula en la notación para una combinación de factores específica, indica que el factor en cuestión está en su nivel alto y la ausencia de una letra implica que el factor está en su nivel bajo. Se utiliza también el tipo de notación binaria, denotando como 0 el nivel bajo y como 1 el nivel alto de cualquier factor. Así

- (1) = 000 – A, B, C en nivel bajo
- a = 100 – A en nivel alto; B y C en su nivel bajo
- b = 010 – B en nivel alto; A y C en su nivel bajo
- c = 100 – C en nivel alto; A y B en su nivel bajo
- ab = 110 – A y B en nivel alto; C en su nivel bajo
- ac = 101 – A y C en nivel alto; B en su nivel bajo
- bc = 011 – B y C en nivel alto; A en su nivel bajo
- abc = 111 – A, B, C en nivel alto.

Cuando se ajusta una ecuación de primer orden a un experimento factorial 2^m , conviene **codificar** las variables explicativas, con un -1 representando el nivel bajo de la variable y $+1$ para representar el nivel alto. Esto, desde luego, corresponde a la transformación

$$x_i = 2 \left(\frac{\xi_i - \bar{\xi}_i}{d_i} \right),$$

donde ξ_i es el valor original que adquiere la variable ξ_i , $\bar{\xi}_i$ es el promedio de los valores de dicha variable y d_i la distancia entre el nivel bajo y alto de la variable ξ_i , para $i = \overline{1, m}$.

Codificar las variables reduce el esfuerzo computacional en los cálculos algebraicos, realizados al emplear el método de mínimos cuadrados para ajustar un modelo. Para más detalles sobre variables codificadas ver ([27], pág. 46).

Suponga, por ejemplo, que el experimento es un factorial 2^3 con los niveles fijos: $\xi_1 : 150, 200$; $\xi_2 : 8, 12$; $\xi_3 : 30, 40$. La siguiente es la matriz diseño (en términos de variables codificadas)

$$D = \begin{bmatrix} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \end{bmatrix} \begin{matrix} (1) \\ a \\ b \\ c \\ ab \\ ac \\ bc \\ abc \end{matrix} \quad (1.3.3)$$

la primer columna pertenece a la variable x_1 , la segunda a la variable x_2 y la tercera a la variable x_3 . Observe, que en el tercer renglón de la matriz anterior, sólo el factor B que corresponde a la variable original ξ_2 , esta en su nivel alto; entonces

$$x_2 = 2 \left(\frac{12 - 10}{4} \right) = +1,$$

y el valor de x_3 se obtendría como

$$x_3 = 2 \left(\frac{30 - 35}{10} \right) = -1.$$

Similarmente, x_1 toma el valor de -1 . Si la matriz de diseño D , de la ecuación (1.3.3), se utiliza para ajustar un modelo del tipo

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, \quad (1.3.4)$$

con $i = \overline{1, 8}$ con las suposiciones usuales sobre ϵ_i , se pueden aplicar los principios del modelo lineal general múltiple para estimar los parámetros β^i y, si es necesario, hacer inferencias sobre los mismos.

Dado que tendencias cuadráticas sobre cada eje en las variables explicativas, no pueden captarse teniendo tan sólo 2 niveles en cada variable; los términos x_1^2, x_2^2, \dots no pueden ser incluidos en el modelo. Por esta razón, al utilizarse los diseños 2^m se supone que el fenómeno en estudio tiene en realidad un comportamiento lineal. Sin embargo, como se verá más adelante, los términos correspondientes a interacciones o productos cruzados, tales como x_1x_2, x_1x_3, \dots , podrán incluirse en el modelo y sus parámetros serán estimados.

Para el modelo dado por la ecuación (1.3.4) se tiene como matriz X del modelo lineal general múltiple

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} D$$

y

$$X^t X = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}.$$

Se observa que $X^t X$ es diagonal y por tanto el diseño es ortogonal, particularidad de todos los diseños 2^m . Utilizando la notación definida previamente para las observaciones, los vectores Y y $X^t Y$ están dados por

$$Y = \begin{bmatrix} (1) \\ a \\ b \\ c \\ ab \\ ac \\ bc \\ abc \end{bmatrix}, \quad X^t Y = \begin{bmatrix} (1) + a + b + c + ab + ac + bc + abc \\ a + ab + ac + abc - (1) - b - c - bc \\ b + ab + bc + abc - (1) - a - c - ac \\ c + ac + bc + abc - (1) - a - b - ab \end{bmatrix}.$$

Así, los estimadores de $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_3$, se calculan de los correspondientes elementos de X^tY divididos por 8, por ejemplo

$$\hat{\beta}_1 = \frac{a + ab + ac + abc - (1) - b - c - bc}{8}. \quad (1.3.5)$$

Realmente la estimaciones de los coeficientes del modelo son *contrastes*³ divididos entre el número total de observaciones. En diseño de experimentos, los efectos se definen con base en totales de tratamientos, por ejemplo:

$$(A)_{i=0} = T_{000} + T_{010} + T_{001} + T_{011} = 000 + 010 + 001 + 011,$$

donde T_{000} , o bien, 000, representa el total o suma de todas las observaciones que reciben el tratamiento $000 = (1)$; T_{010} el total de todas las observaciones que reciben el tratamiento $010 = b$, y así sucesivamente. En lo anterior, se ha supuesto que se hacen repeticiones de observaciones sobre un mismo tratamiento.

$$(A)_{i=1} = T_{111} + T_{100} + T_{110} + T_{101} = 111 + 100 + 110 + 101.$$

Ahora con $(A)_0$ y $(A)_1$ se define el efecto del factor A como $(A) = (A)_1 - (A)_0$, el cual es un contraste. Obsérvese que $(A)_1 - (A)_0$, puede escribirse como

$$\begin{aligned} (A)_1 - (A)_0 &= 100 + 110 + 101 + 111 - 000 - 010 - 001 - 011 \\ &= a + ab + ac + abc - (1) - b - c - bc, \end{aligned}$$

que corresponde al numerador de la ecuación (1.3.5).

También observe que dada la naturaleza del diseño experimental, los coeficientes en esta regresión son no correlacionados, esto es

$$V(\hat{\beta}) = \sigma^2(X^tX)^{-1} = \frac{\sigma^2}{8}I.$$

Sólo para fijar ideas, considere el siguiente ejemplo

Ejemplo (1.3.1):

Se supone que tres variables x_1 , x_2 y x_3 , tienen influencia sobre alguna respuesta, de acuerdo al modelo

$$Y = \beta_0 + \sum_{i=1}^3 \beta_i x_i + \epsilon.$$

³Los contrastes son cierto tipo de funciones lineales a partir de las cuales se quiere hacer algún tipo de inferencia.

Se utilizó un diseño factorial 2^3 para ajustar el modelo. El vector de observaciones, así como la matriz X , están dados por

$$Y = \begin{bmatrix} (1) = 3.0 \\ a = 5.0 \\ b = 7.0 \\ c = 2.0 \\ ab = 12.0 \\ ac = 4.0 \\ bc = 5.0 \\ abc = 14.0 \end{bmatrix} \quad X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & +1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & +1 & -1 \\ 1 & +1 & -1 & +1 \\ 1 & -1 & +1 & +1 \\ 1 & +1 & +1 & +1 \end{bmatrix}$$

Los niveles de las variables en la matriz X , están desde luego codificados. Entonces

$$\hat{\beta}_0 = \frac{(1) + a + b + c + ab + ac + bc + abc}{8} = \frac{52}{8} = 6.5$$

$$\hat{\beta}_1 = \frac{a + ab + ac + abc - (1) - b - c - bc}{8} = \frac{16}{8} = 2,$$

y de igual manera para $\hat{\beta}_2$ y $\hat{\beta}_3$.

Inclusión de términos de interacción en el modelo:

El diseño factorial 2^m se puede utilizar en casos donde términos de interacción se incluyen en el modelo de regresión. Si se usa un diseño factorial completo, los coeficientes de interacción son ortogonales entre si y entre los coeficientes de los términos de primer orden. Por ejemplo, suponga que en los niveles de x_{1i} y x_{2i} , la medición de una respuesta Y_i es

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \epsilon_i. \quad (1.3.6)$$

Esto es, se incluye el término $\beta_{12} x_1 x_2$ que representa una desviación de la linealidad. En términos prácticos, se supone que el efecto de x_1 sobre la respuesta Y , depende del nivel de operación de x_2 . Para el caso de tres variables explicativas, términos como $x_1 x_2$, $x_1 x_3$, $x_2 x_3$ y $x_1 x_2 x_3$ podrían aparecer. De hecho, para m variables explicativas, habrán $2^m - m - 1$ de tales términos. El número $2^m - m - 1$ corresponde exactamente a los grados de libertad para la suma de cuadrados del error, que se presenta al usar un diseño 2^m de acuerdo a la tabla ANOVA presentada en la sección anterior. Esto último quiere decir, que se agotan los grados de libertad necesarios para estimar la varianza del error al estimar los parámetros correspondientes a interacciones. Por lo cual, si se desea estimar la varianza del error para poder realizar inferencias, es necesario tomar al

menos $r > 1$ observaciones por combinación de tratamientos (el mismo número en cada combinación, en orden de mantener la ortogonalidad), así el estimador de la varianza tendría $2^m(r - 1)$ grados de libertad.

Ejemplo (1.3.2):

Considere el modelo dado por la ecuación (1.3.6), en un experimento con dos observaciones por tratamiento. La matriz X y el correspondiente vector de observaciones están dados por:

$$Y = \begin{bmatrix} (1) = 2.0 \\ (1) = 2.3 \\ a = 5.7 \\ a = 4.8 \\ b = 4.5 \\ b = 4.2 \\ ab = 5.3 \\ ab = 5.5 \end{bmatrix} \quad X = \begin{bmatrix} 1 & -1 & -1 & +1 \\ 1 & -1 & -1 & +1 \\ 1 & +1 & -1 & -1 \\ 1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 \\ 1 & +1 & +1 & +1 \end{bmatrix}.$$

Permitiendo que (1) , a , b y ab denoten los totales o suma de las observaciones por tratamiento, por ejemplo, $(1) = 2.0 + 2.3 = 4.3$, el vector de estimadores se calcula como

$$\begin{aligned} \hat{\beta} &= (X^t X)^{-1} \begin{bmatrix} (1) + a + b + ab \\ a + ab - (1) - b \\ b + ab - (1) - a \\ (1) + ab - a - b \end{bmatrix} \\ &= \begin{bmatrix} 34.3/8 \\ 8.3/8 \\ 4.7/8 \\ -4.1/8 \end{bmatrix} \\ &= \begin{bmatrix} 4.29 \\ 1.04 \\ 0.59 \\ -0.51 \end{bmatrix}. \end{aligned}$$

y el estimador de σ^2 tendría 4 grados de libertad y se calcularía utilizando la ecuación (1.2.8).

Por último, se puede ver que $V(\hat{\eta})$ evaluada en el punto x , en un diseño factorial 2^m es

$$V(\hat{\eta}(x)) = \frac{\sigma^2(1 + \sum x_i^2)}{2^m},$$

la división entre 2^m implica que $V(\hat{\eta})$ es pequeña, esto es: los estimadores son precisos; sin embargo, al tener 2^m puntos el diseño resulta caro para m grande.

B. Diseño Factorial Fraccional, 2^{m-k} :

Un factorial fraccional es, como su nombre lo indica una fracción de un experimento factorial completo. Este diseño es particularmente útil, cuando la cantidad de experimentación requerida por un factorial completo es más de la que el experimentador puede realizar. Para una información completa acerca de diseños factoriales fraccionales ver [13, 14].

En este apartado se hace una breve revisión solamente en relación a fracciones de diseños factoriales 2^m , los cuales se utilizan en el análisis de superficies de respuesta para el ajuste de modelos de regresión, cuando m es muy grande y pocas o ninguna interacción se considera importante.

Como ejemplo, considere 4 factores en un experimento, el cual se supone puede ser modelado por

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \epsilon,$$

y en vez de realizar un experimento factorial 2^4 completo, se realizan las siguientes corridas: a,b,c,d,abc,abd,bcd,acd. La matriz X (con las variables x codificadas) está dada por

$$X = \begin{bmatrix} 1 & +1 & -1 & -1 & -1 \\ 1 & -1 & +1 & -1 & -1 \\ 1 & -1 & -1 & +1 & -1 \\ 1 & -1 & -1 & -1 & +1 \\ 1 & +1 & +1 & +1 & -1 \\ 1 & +1 & +1 & -1 & +1 \\ 1 & -1 & +1 & +1 & +1 \\ 1 & +1 & -1 & +1 & +1 \end{bmatrix}. \quad (1.3.7)$$

Para el modelo de primer orden postulado aquí, el diseño es ortogonal debido a que $X^tX = 8I_5$. Así los estimadores $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ y $\hat{\beta}_4$, obtenidos de la manera usual, son no correlacionados. El sacrificio al usar 8 puntos experimentales, en vez del experimento completo de $16 = 2^4$ puntos es muy simple de ilustrar. Suponga que se considera que la matriz X , se aumenta con los términos $x_1x_2, x_1x_3, \dots, x_1x_2x_3x_4$ correspondientes a

interacciones. La porción añadida sería:

x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4
-1	-1	-1	+1	+1	+1
-1	+1	+1	-1	-1	+1
+1	-1	+1	-1	+1	-1
+1	+1	-1	+1	-1	-1
+1	+1	-1	+1	-1	-1
+1	-1	+1	-1	+1	-1
-1	-1	-1	+1	+1	+1
-1	+1	+1	-1	-1	+1

$x_1x_2x_3$	$x_1x_2x_4$	$x_1x_3x_4$	$x_2x_3x_4$	$x_1x_2x_3x_4$
+1	+1	+1	-1	-1
+1	+1	-1	+1	-1
+1	-1	+1	+1	-1
-1	+1	+1	+1	-1
+1	-1	-1	-1	-1
-1	+1	-1	-1	-1
-1	-1	-1	+1	-1
-1	-1	+1	-1	-1

y se necesitaran al menos 16 puntos o renglones en la matriz X , para estimar todos los parámetros que surgen al hacer el aumento. Es claro, que con la matriz construída en base a 8 renglones, tal estimación es imposible de realizarse. Sin embargo, las columnas adicionales de la matriz X estan dadas para mostrar la consecuencia del uso de $1/2$ fracción de un factorial 2^4 completo. Se observa que el contraste calculado en la obtención del efecto del término x_1 , es idéntico, excepto por el signo, al que será usado para calcular el efecto de $x_2x_3x_4$. Esto es, los correspondientes elementos de X^tY son iguales excepto por su signo. Como resultado si el término $x_2x_3x_4$ se incluye en el modelo, los efectos de x_1 y $x_2x_3x_4$ serían *confusos* o *confundidos* uno con otro. Se dice entonces que x_1 y $x_2x_3x_4$ son *alias*, o bien, el alias de x_1 es $x_2x_3x_4$. En otras palabras, la función lineal que estima x_1 , también estima $x_2x_3x_4$. En términos prácticos, no tendría sentido incluir $x_2x_3x_4$ en el modelo si su efecto está captado al estimar el parámetro correspondiente a x_1 ; su efecto se confunde con el de x_1 . De igual manera, se observa que los demás términos de primer orden también tienen alias (el signo \equiv implica es *alias con*):

$$x_2 \equiv -x_1x_3x_4$$

$$x_3 \equiv -x_1x_2x_4$$

$$x_4 \equiv -x_1x_2x_3.$$

También puede verse que las interacciones de dos factores tienen alias:

$$x_1x_2 \equiv -x_3x_4$$

$$x_1x_3 \equiv -x_2x_4$$

$$x_1x_4 \equiv -x_2x_3.$$

$x_1x_2x_3x_4$ es alias con la primera columna de la matriz X , que representa el término constante β_0 .

Por lo anterior, pueden sacrificarse totalmente las interacciones de orden tres y cuatro, y considerarse sólo tres de las interacciones de 2 términos, debido a que tienen como alias interacciones del mismo orden; los parámetros de éstas últimas podrán ser estimados debido a que el diseño posee 8 puntos (cinco puntos para estimar los términos lineales y los 3 adicionales para estimar las 3 interacciones de orden dos). Las tres interacciones de orden dos, pueden ser consideradas por el experimentador, como poco importantes, por lo cual, pueden ser eliminadas del modelo.

Como se observa, el hecho de utilizar únicamente 1/2 fracción del factorial 2^4 produce una mezcla (alias) de los efectos, de aquí que, la eliminación de los términos no lineales no resulta problemática. Si de los miembros de un grupo alias sólo uno se considera como posibilidad y el otro inexistente, se puede estudiar el posible efecto de uno de ellos evaluando un estimador de su efecto y su suma de cuadrados.

Construcción de 1/2 fracción de un factorial 2^m :

Las variables x_1, x_2, \dots están referidas notacionalmente como A, B, \dots . Las interacciones son denotadas por $AB, AC, \dots, ABC, \dots$. Un contraste definido $A^{\gamma_1} B^{\gamma_2} C^{\gamma_3} \dots$ se selecciona, donde γ_i es cualquiera de los dos valores, 0 ó 1. Defina un nuevo conjunto de variables z_1, z_2, \dots, z_m , asociadas con las letras minúsculas a, b, c, \dots respectivamente. Para una combinación particular de tratamientos, z_j será 1 si su letra minúscula correspondiente aparece en la combinación de tratamientos y cero en otro caso. Por ejemplo, para $m = 4$, $z_1 = 1$, $z_2 = 1$ y todas las demás z^j en cero, la combinación de tratamientos es ab .

Se evalúa entonces

$$L = \gamma_1 z_1 + \gamma_2 z_2 + \dots + \gamma_m z_m,$$

para las 2^m combinaciones de tratamientos (1), a, b, \dots . Entonces para los 2^m valores

de L se define s como:

$$s = L(\text{módulo})2.$$

Esto es, evaluar $L/2$ y tomar s como el residuo (cualquiera 0 ó 1). En el procedimiento resultan 2^{m-1} combinaciones con $s = 0$ y un número igual con $s = 1$. Esto es, dos conjuntos de valores de s , los puntos que tienen valor de $s = 0$ y los que tienen valor $s = 1$. Cualquiera de los dos conjuntos produce la media fracción deseada del factorial 2^m . La media fracción que se utiliza para el experimento debe ser seleccionada aleatoriamente de este par.

Alias en la media fracción:

En la media fracción del factorial 2^m , cada efecto tiene un alias, los alias se determinan multiplicando el efecto en cuestión por un contraste seleccionado y definido previamente, con los exponentes reducidos módulo dos. Por ejemplo, en el factorial 2^4 definido anteriormente, y un contraste definido como $ABCD$ ($x_1x_2x_3x_4$), interacción de los cuatro factores, el efecto A es alias con BCD ; esto es

$$A(ABCD) = A^2BCD = BCD,$$

el exponente 2 sobre la letra A se reduce a cero al tomar 2(módulo)2. El efecto BCD se dice la interacción generalizada de A en el contraste definido por $ABCD$. En general, en los sistemas factoriales 2^m los alias de un efecto son interacciones generalizadas entre los efectos dada la definición de un contraste específico.

Con el siguiente ejemplo se intentarán fijar las ideas anteriores.

Ejemplo (1.3.3):

Considere el factorial 2^4 citado anteriormente. Se ilustrará el método de construcción que lleva a definir la matriz X , en (1.3.7). Con $x_1x_2x_3x_4$ ($ABCD$) como definición de contraste,

$$L = z_1 + z_2 + z_3 + z_4,$$

dado que $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1$. Los valores de s se dan en la siguiente tabla

Valores de s para 1/2 fracción de un diseño 2^4 con contraste $ABCD$	
combinación de tratamientos	s
(1)	0
a	1
b	1
c	1
d	1
ab	0
ac	0
ad	0
bc	0
bd	0
cd	0
abc	1
abd	1
acd	1
bcd	1
abcd	0

El diseño seleccionado consistió de las combinaciones de tratamientos que tienen $s = 1$, mismo que esta representado en la matriz X .

Si se seleccionan las combinaciones de tratamientos con $s = 0$, la misma estructura de alias resultaría. Sin embargo, un efecto tendría idénticamente el mismo contraste que su alias, en vez de su negativo.

Si se considera ahora, el problema de construir 1/2 fracción del factorial 2^4 , pero utilizando el contraste ACD. Se tiene

$$L = z_1 + z_3 + z_4,$$

dado que $\gamma_1 = \gamma_3 = \gamma_4 = 1$ y $\gamma_2 = 0$. Los valores de s se dan en la siguiente tabla.

Valores de s para 1/2 fracción de un diseño 2^4 con contraste ACD	
combinación de tratamientos	s
(1)	0
a	1
b	0
c	1
d	1
ab	1
ac	0
ad	0
bc	1
bd	1
cd	0
abc	0
abd	0
acd	1
bcd	0
abcd	1

otra vez, cualquier conjunto que tenga $s = 0$ o $s = 1$ es suficiente para dar la 1/2 fracción. En este ejemplo, los efectos principales A, C, D son alias con interacciones de orden dos, mientras que B es alias con $ABCD$.

Se observa que si se utiliza 1/2 fracción de un diseño factorial 2^4 para ajustar un modelo lineal, esto es, sin interacciones; se podrán estimar $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ y σ^2 . Esto, ya que para estimar el vector de parámetros se requiere solamente de 5 puntos y la matriz X cuenta con 8; los tres puntos sobrantes servirán para estimar la varianza del error.

Es importante revisar en general, la construcción de cualquier fracción de un factorial 2^m , para que dado un contexto, pueda saber qué fracción le conviene crear. Una revisión muy completa sobre la construcción de fracciones de factoriales viene dada en [13].

Usualmente los diseños factoriales 2^{m-k} no proporcionan estimadores de la varianza del error (este no fue el caso en el ejemplo dado). Lo más común es que se agreguen varias observaciones en el centro del diseño, esto es, en $\mathbf{x} = \mathbf{0}$, que permiten estimar el error puro con $n_2 - 1$ grados de libertad (donde $n_2 - 1$ es el número de repeticiones del experimento en el punto central), el cual se emplea para probar la hipótesis de carencia de ajuste.

Cuando no se incluyen las interacciones, o bien, algunos terminos en el modelo lineal general, éste puede ser escrito como

$$Y = X_1\beta_1 + \epsilon. \quad (1.3.8)$$

Sin embargo, la representación *verdadera* o más adecuada de la superficie de respuesta, es la que contiene todos los términos, esto es,

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon^*. \quad (1.3.9)$$

Si se estima el vector de parámetros en el modelo (1.3.8), se tiene que $\hat{\beta}_1 = (X_1^t X_1)^{-1} X_1^t Y$, pero resulta que este estimador es sesgado, esto, debido a que se elimina una parte del modelo *verdadero*, dado en la ecuación (1.3.9). De aquí, que para el casos de diseños fraccionales se tengan estimaciones sesgadas.

En ocasiones es útil conocer el tipo de sesgos que se presentan, para así utilizar diseños que los minimicen. Usualmente se tienen diseños de primer orden que minimicen el sesgo si el modelo *verdadero* es de segundo orden, o bien, diseños de segundo orden que minimicen sesgos si el modelo *verdadero* es de tercer orden. Estos diseños ya no son ortogonales. Por tanto, el utilizar diseños ortogonales no garantiza que el sesgo sea mínimo, pero si que la varianza de los estimadores se minimiza. A partir de estos dos criterios (sesgo y varianza) y algunos otros que existen en la literatura ([27], págs 122-139) el investigador debe elegir el diseño que más convenga al problema que intenta resolver. La elección del criterio podría estar dada en términos de costo. De diseños no ortogonales no se hará mención en el presente trabajo, debido a que su utilidad es muy limitada en problemas prácticos⁴. Una expresión para el sesgo del estimador es

$$\begin{aligned} E(\hat{\beta}_1) &= E[(X_1^t X_1)^{-1} X_1^t Y] \\ &= E[(X_1^t X_1)^{-1} X_1^t (X_1\beta_1 + X_2\beta_2 + \epsilon^*)] \\ &= \beta_1 + (X_1^t X_1)^{-1} X_1^t X_2\beta_2 \\ &= \beta_1 + A\beta_2. \end{aligned}$$

La matriz $A = (X_1^t X_1)^{-1} X_1^t X_2$ es llamada la *matriz alias*, y son sus elementos los que se busca minimizar escogiendo adecuadamente la estructura de la matriz X_1 , es decir el diseño.

⁴Si se desea vease [33]

Se puede verificar fácilmente que para un factorial fraccional 2^{m-k}

$$V(\hat{\eta}(\mathbf{x})) = \frac{\sigma^2(1 + \sum x_i^2)}{2^{m-k}},$$

es decir, se llega a un mejor equilibrio en términos de la precisión de las estimaciones y el número de puntos necesarios para obtenerla.

C. Diseño Simplex:

Este es el diseño más pequeño que reporta la literatura; consta de $m + 1$ puntos, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$, equidistantes al origen y equidistantes entre sí, de norma igual a uno. El diseño es insuficiente para estimar a σ^2 , por ello es necesario agregar, al menos, un punto más y por simetría se sugiere que este punto se localice en el centro del diseño.

Con este diseño la matriz $X^t X$ es diagonal; esto implica que los cálculos para obtener los estimadores de β y η son muy simples.

El diseño simplex es invariante a las rotaciones, ya que si se aplica una rotación a los puntos del diseño simplex, el resultado es otro diseño simplex.

Los puntos experimentales y el origen forman un ángulo θ constante para dos puntos cualquiera, donde $\cos \theta = -\frac{1}{m}$.

Para $m = 2$, el $\cos \theta = -\frac{1}{2}$ y por tanto $\theta = 120^\circ$. Así el simplex es un triángulo equilátero con centro en el origen. Para $m = 3$, los puntos del diseño son los vértices de un tetraedro. La matriz Diseño D , de la figura (1.3.1) es

$$D = \begin{bmatrix} \sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \\ -\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{2}{\sqrt{2}} \end{bmatrix} \quad (1.3.10)$$

Los renglones de la matriz diseño corresponden a las coordenadas en la figura en dos dimensiones. La matriz X es D aumentada con una columna de unos, esto es

$$X = \begin{bmatrix} 1 & \\ 1 & D \\ 1 & \end{bmatrix}, \quad (1.3.11)$$

donde

$$X^t X = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

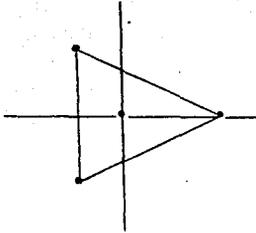


Figura 1.3.1: Diseño simplex $m = 2$

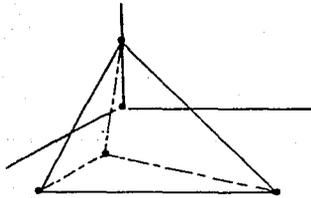


Figura 1.3.2: Diseño simplex $m = 3$

es una matriz diagonal y por lo tanto el diseño es ortogonal.

En general, La matriz X para un simplex m -dimensional se construye de manera muy simple. Se comienza con una matriz ortogonal O de orden $N \times N$, con $N = m + 1$, los elementos en la primera columna deben ser iguales. La matriz X es entonces

$$X = (O)(N)^{\frac{1}{2}},$$

y se observa fácilmente que

$$X^t X = N(O^t O) = N(I_N).$$

La matriz O puede obtenerse fácilmente de la siguiente manera: seleccione cualquier matriz no singular, Q , de dimensión $N \times N$, con elementos iguales en su primer columna. Utilizando la técnica de ortonormalización de Gram-Schmidt, las columnas de Q pueden transformarse linealmente en N nuevas columnas ortogonales entre si y de norma unitaria, en donde la primera columna consiste de elementos iguales a $\frac{1}{\sqrt{N}}$. Esta transformación lineal se realiza fácilmente con la ayuda del paquete MATLAB, o bien, utilizando el comando GS en el procedimiento PROC MATRIX del paquete SAS (1982). El conjunto de columnas ortonormalizadas forma la matriz O .

Por ejemplo, la matriz X de la ecuación (1.3.11) puede ser generada por la matriz ortonormal O , esto es, $X = \sqrt{3}O$, con

$$O = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{bmatrix}.$$

Por último, se puede ver fácilmente que para un diseño simplex con s puntos en el centro, $V(\hat{\eta})$ evaluada en el punto $\mathbf{x} = (x_1, x_2, \dots, x_m)$ es igual a:

$$V(\hat{\eta}) = \left[\frac{1}{m + s + 1} + \frac{m \sum x_i^2}{(m + 1)} \right] \sigma^2.$$

1.3.2 Diseños Para Estimar Polinomios de Segundo Orden

Después de revisar brevemente los diseños para ajustar un polinomio de primer grado, se enumeran los diseños utilizados para el ajuste de polinomios de segundo grado. Los

polinomios de segundo grado tienen más parámetros, por lo que se necesitan diseños con más puntos para estimarlos. Es común usar, en este caso, los diseños utilizados para el ajuste de un polinomio de primer grado, agregándole convenientemente algunos puntos más para tener el número necesario de observaciones.

El modelo lineal general con $r = 2$ es:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_m x_{mi} + \beta_{11} x_{1i}^2 + \beta_{12} x_{1i} x_{2i} + \cdots + \beta_{mm} x_{mi}^2 + \epsilon_i, \quad (1.3.12)$$

el cual tiene: un parámetro independiente, β_0 ; m parámetros en los términos de primer grado, β_i , $i = \overline{1, m}$; m parámetros en los términos cuadráticos, β_{ii} , $i = \overline{1, m}$; y por último, C_2^m parámetros en los términos cruzados, β_{ij} , ($1 \leq i < j \leq m$). En total hay

$$1 + m + m + C_2^m = C_0^m + C_1^m + C_1^m + C_2^m = C_1^{m+1} + C_2^{m+1} = C_2^{m+2},$$

parámetros. Entonces para estimar un polinomio de segundo grado se necesitan al menos C_2^{m+2} puntos (uno más por lo menos si se desea estimar también a σ^2) y, por supuesto, que X sea de rango completo.

A. Diseño Simplex Compuesto:

Este diseño se compone de los $m+1$ puntos de un diseño simplex, p_i , $i = \overline{1, m, m+1}$; más C_2^{m+1} puntos de la forma $k(p_i + p_j)$, donde $i \neq j$ y k es una constante de escalamiento. Por construcción, este diseño tiene el mínimo número de puntos, necesarios para estimar al vector β ; sin embargo la matriz X no es de rango completo. Para completar el rango de X es suficiente agregar al diseño un punto más en el centro. En las figuras (1.3.3) y (1.3.4) los puntos que se agregan al diseño simplex son los marcados con "o".

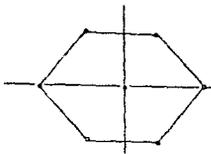


Figura 1.3.3: Diseño Simplex compuesto en dos factores

B. Diseño Central Compuesto:

Es el diseño más utilizado; está formado por los puntos de un diseño factorial 2^m , o de

⁵La notación C_n^m se refiere a las combinaciones de m en n , con $m \geq n$.

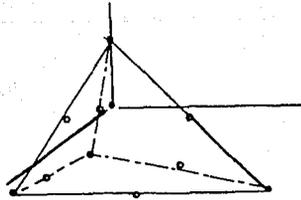


Figura 1.3.4: Diseño Simplex compuesto en tres factores

un factorial fraccionado 2^{m-k} , más los puntos de la forma $\pm\alpha e_i$, donde (e_1, e_2, \dots, e_m) es la base canónica de \mathbb{R}^m y α es una constante de escalamiento. Este diseño tiene $2^{m-k} + 2m$ puntos.

Los puntos que se agregan al diseño factorial, permiten estimar los términos de segundo orden de la superficie. El aumento hecho a la matriz diseño de un factorial 2^m es como sigue:

$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_m \\ 0 & 0 & 0 & \cdots & 0 \\ -\alpha & 0 & 0 & \cdots & 0 \\ +\alpha & 0 & 0 & \cdots & 0 \\ 0 & -\alpha & 0 & \cdots & 0 \\ 0 & +\alpha & 0 & \cdots & 0 \\ 0 & 0 & -\alpha & \cdots & 0 \\ 0 & 0 & +\alpha & \cdots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \cdots & -\alpha \\ 0 & 0 & 0 & \cdots & +\alpha \end{bmatrix},$$

donde el primer renglón representa el punto central y los demás renglones puntos auxiliares. El valor de α se selecciona en base a criterios útiles, que se mencionan más adelante.

Usualmente en el origen se hacen más de una observación. La figura (1.3.5) es una representación de los 15 puntos de un diseño central compuesto con $m = 3$. Los primeros 8 puntos corresponden a un diseño factorial 2^3 , el punto 9 es el centro del

diseño y los 6 restantes son puntos auxiliares.

$$D = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & +1 \\ -1 & +1 & -1 \\ -1 & +1 & +1 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \\ 0 & 0 & 0 \\ -\alpha & 0 & 0 \\ +\alpha & 0 & 0 \\ 0 & -\alpha & 0 \\ 0 & +\alpha & 0 \\ 0 & 0 & -\alpha \\ 0 & 0 & +\alpha \end{bmatrix}.$$

El primer criterio para la selección de α es que resulte un diseño ortogonal, esto es,

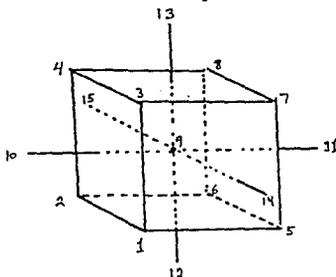


Figura 1.3.5: Diseño central compuesto para $m = 3$.

que $X'X$ sea diagonal. Donde X es la matriz diseño aumentada con una columna de unos. A continuación, se presentan algunos valores de α^n para crear diseños centrales

compuestos ortogonales en m variables

Valores de α , para diseños centrales compuestos ortogonales	
m	α
2	1
3	1.216
4	1.414
5	1.596
6	1.761
7	1.910
8	2.045

Otros criterios para seleccionar el valor de α , y en general para comparar diseños de segundo orden, son la eficiencia, medida en términos de varianza, en la estimación de los coeficientes del polinomio que representa la respuesta y los sesgos que se tienen si el modelo más adecuado, es en realidad un polinomio de tercer orden. Un estudio profundo sobre los diferentes criterios para elegir el escalara α se da en ([33], págs. 130-133; [27], págs 116-122); también en [6, 25].

1.3.3 Otros diseños más generales:

No es la finalidad de este trabajo hacer un listado exhaustivo de los diseños que se pueden encontrar en la literatura correspondiente al tema; sin embargo, es conveniente mencionar otros diseños frecuentemente mencionados, que se utilizan para estimar polinomios más generales.

A. Diseño Rotable:

Otra propiedad deseable de los diseños para estudiar superficies de respuesta es la *rotabilidad*. Este criterio es general para cualquier diseño, aunque su uso principal es con los diseños de segundo orden. Un diseño es *rotable* cuando la varianza de la respuesta estimada, $\hat{\eta}$, es una función únicamente de la distancia al centro del diseño y no de la dirección. La idea fue introducida por Box y Hunter (1957)[5].

Usualmente se tiene mayor precisión en el centro del diseño y en forma de círculos, esferas o hiper-esferas concéntricas, la precisión en la estimación de η va disminuyendo o se va aumentando la varianza, $V(\hat{\eta})$.

La determinación de las condiciones para rotabilidad en cualquier diseño de segundo orden, requiere considerar los *momentos* de los vectores que forman la matriz de diseño.

Esto es, suma de sus elementos elevados a la potencia p con $p = 1, 2$. Ya que aparecen en la expresión de $V(\hat{\eta})$. Usualmente la condición de rotabilidad se determina con los momentos de segundo orden de términos con productos cruzados $(\sum_i x_{1i}^2 x_{2i}^2)$.

Un ejemplo del diseño rotable, es el diseño central compuesto con $\alpha = \sqrt{2}$ para $m = 2$, el cual se toma de la tabla siguiente. En la tabla, F es el número de observaciones del factorial 2^m (es 2^m si el factorial es completo o 2^{m-k} en las fracciones de $\frac{1}{2^k}$). n_2 es el número de observaciones en las que se repite el centro del diseño y $T = 2m + n_2$ es el número de observaciones que no reciben tratamiento del factorial, son $2m$ puntos auxiliares o axiales. En la tabla $n_2 = 1$ en todos los casos. N total de observaciones, $N = F + T$.

VALORES DE α PARA DISEÑOS CENTRALES COMPUESTOS ROTABLES					
k		F	T	N	α
2		4	5	9	1.414
3		8	7	15	1.682
4		16	9	25	2.000
5		32	11	43	2.378
5	1/2 frac.	16	11	27	2.000
6		64	13	77	2.828
6	1/2 frac.	32	13	45	2.378
7		128	15	143	3.364
7	1/2 frac.	64	15	79	2.828
8		256	17	273	4.000
8	1/2 frac.	128	17	145	3.364

Un aumento de puntos centrales no afecta la rotabilidad y permite estimar el *error puro* y hacer pruebas de carencia de ajuste.

Otros tipos de diseños rotables son los simplex, simplex compuestos y factoriales.

B. Diseño Poligonal:

Los puntos de estos diseños son los vértices de poliedros regulares. Para $r = 2$ los puntos de los diseños poligonales forman un polígono regular; los diseños simplex, factorial 2^m y factorial 2^{m-k} caen en esta clasificación.

C. Diseños D-óptimos

Los D-óptimos son una clase de diseños con los que se maximiza el determinante de la matriz $X^t X$.

1.4 METODOS DE BUSQUEDA DEL OPTIMO

Como ya se mencionó, en la mayoría de los casos se utilizan aproximaciones locales a la función de respuesta η . Por ello, para estimar el o los puntos donde η alcanza un óptimo relativo (o absoluto) se necesita contar con un método que garantice un mínimo de confiabilidad en los resultados.

Por último, debe mencionarse que los métodos de búsqueda presentados solamente consideran el caso en que η presenta un máximo único y bien definido (o al menos, que se puede llegar a un máximo bien definido).

En esta parte, se revisarán los métodos más conocidos, los cuales son:

- El método aleatorio. (Cochran y Cox, 1980)
- El método de factor único. (Cochran y Cox, 1980)
- El método de un sólo experimento. (Cochran y Cox, 1980)
- El método de ascenso por pendiente máxima⁶. (Box y Wilson, 1951)

1.4.1 El Método Aleatorio

Este método consta de tres pasos:

- 1.- Se eligen al azar n puntos $\mathbf{x}_i' = (x_{i1}, x_{i2}, \dots, x_{im})$ cuyas coordenadas son variables aleatorias independientes con distribución uniforme: $x_{ij} \sim U(a_j, b_j)$; así, estos puntos resultan ser variables aleatorias multivariadas, independientes, con distribución uniforme en el conjunto $A = (a_1, b_1) \times (a_2, b_2) \times \dots \times (a_m, b_m)$, el cual se llama región de muestreo, o región de exploración.
- 2.- Se obtienen las observaciones $Y_i = \eta(\mathbf{x}_i) + \epsilon_i$, en cada uno de los puntos \mathbf{x}_i , antes seleccionados.
- 3.- Se escoge como $\hat{\mathbf{x}}_{op}$ al punto \mathbf{x}_i que presente mayor respuesta experimental.

Cuando \mathbf{x}_{op} no pertenece a A , $\hat{\mathbf{x}}_{op}$ resulta ser un estimador sesgado de \mathbf{x}_{op} , (esto es $E(\hat{\mathbf{x}}_{op}) \neq \mathbf{x}_{op}$), el sesgo es mayor conforme \mathbf{x}_{op} se encuentre más lejos de la región de muestreo.

⁶ Este método no se describe en esta sección, pues se tratará a fondo en el capítulo 2.

Cuando \mathbf{x}_{op} pertenece a A , para cada número natural N , se puede establecer una partición de la región de muestreo dada por los N^m hiper-rectángulos:

$$A(i_1, i_2, \dots, i_m) = (a_{1i_1}, a_{1(i_1+1)}) \times (a_{2i_2}, a_{2(i_2+1)}) \times \dots \times (a_{mi_m}, a_{m(i_m+1)}), \quad (1.4.1)$$

donde para cada $k = \overline{1, m}$; $0 \leq i_k \leq N - 1$ y $a_{ki_k} = a_k + i_k(b_k - a_k)/N$, y encontrar la probabilidad que alguno de los puntos \mathbf{x}_i esté en el mismo elemento de la partición que contiene a \mathbf{x}_{op} . (Como $\mathbf{x}_{op} \in A$), se sigue que \mathbf{x}_{op} pertenece a uno y sólo uno de los conjuntos de la partición).

Para calcular esta probabilidad recuerde que cada punto \mathbf{x}_i es un vector aleatorio distribuido uniformemente en A , y que cada conjunto $A(i_1, i_2, \dots, i_m)$ tiene la misma medida de Lesbegue.

Así el número de observaciones que caen en la misma celdilla que \mathbf{x}_{op} es una variable aleatoria binomial con parámetros n y $1/N^m$, por lo que, la probabilidad de que al menos una observación este en esta celdilla es igual a: ([12], pág. 406)

$$1 - \left(1 - \frac{1}{N^m}\right)^n. \quad (1.4.2)$$

Un resultado semejante se obtiene cuando en lugar de usar un hiper-rectángulo que contiene a \mathbf{x}_{op} , se usa una hiper-esfera de probabilidad $1/N^m$ con centro en \mathbf{x}_{op} .

Note que, independientemente del valor de esta probabilidad, un vector \mathbf{x}_i , en la misma celdilla que \mathbf{x}_{op} no necesariamente presenta la mayor respuesta observada y, por lo tanto, no necesariamente es el punto elegido como estimador de \mathbf{x}_{op} .

1.4.2 El Método del Factor Único

Con este método se propone descomponer el proceso multifactorial, en m procesos sencillos de búsqueda individual del óptimo de cada coordenadas; convirtiéndose en un proceso de búsqueda unidimensional.

El método consta de los siguientes pasos:

- 1.- Se ordenan los factores en orden de importancia, mismo que define el investigador, de acuerdo a su contribución en el resultado de la respuesta: así, x_1 es el factor que más contribuye al resultado; x_2 es el segundo en orden de contribución, etc. La ordenación se hace de acuerdo al criterio del investigador.

2.- Se elige un punto inicial, $P_0 = (x_{10}, x_{20}, \dots, x_{m0})$, ubicado donde, de acuerdo a la experiencia del investigador, es más probable que esté el óptimo.

3.- Se busca sucesivamente, el óptimo de cada coordenada y se reemplaza en P_0 . Este último paso se realiza de la siguiente manera:

- Se obtienen al menos tres observaciones de la forma:

$$Y_i = \eta(x_{1i}, x_{20}, \dots, x_{m0}) + \epsilon_i,$$

variando el nivel del primer factor y manteniendo constante el nivel de los demás factores.

- Con estas observaciones se estima la función cuadrática

$$\hat{Y}(t) = \hat{a}t^2 + \hat{b}t + \hat{c},$$

la cual es un estimador de la tendencia de η en dirección de la primera coordenada; si $\hat{a} < 0$, la función anterior alcanza el máximo en el punto $x_{11} = -\hat{b}/2\hat{a}$; este valor sustituye la primer coordenada de P_0 , resultando el nuevo punto $(x_{11}, x_{20}, \dots, x_{m0})$. (Cuando $\hat{a} = 0$, la ecuación anterior es una recta y no presenta ningún máximo relativo; cuando $\hat{a} > 0$, la ecuación presenta un mínimo como único punto crítico).

- A partir de este punto, se toman al menos tres observaciones de la forma $Y_i = \eta(x_{11}, x_{2i}, \dots, x_{m0}) + \epsilon_i$; variando ahora, únicamente, el segundo factor; con estas observaciones se estima el máximo en dirección de la segunda coordenada, y este valor sustituye a x_{20} . Así se continúa hasta tener el vector $P_1 = (x_{11}, x_{21}, \dots, x_{m1})$.

El proceso de búsqueda se detiene, cuando $\|P_i - P_{i+1}\|$ es menor a un valor preestablecido y $\hat{x}_{pp} = P_{i+1}$. Si esto no ocurre, se inicia nuevamente el proceso de búsqueda con el punto inicial P_{i+1} .

Este método no toma en cuenta las interacciones; es decir cuando no hay interacciones el método puede producir buenos resultados, de no ser así va a requerir de muchos pasos.

Note que con estos dos métodos no se determina la forma funcional de η . Si se desea estimarla es necesario hacerlo por otro camino.

1.4.3 El Método de Experimento Único

Con este método se considera que cerca del óptimo η se parece a un polinomio de segundo grado, por lo que, para buscar el punto óptimo de η es necesario ajustar a los datos un polinomio de este tipo.

El método consta de los dos pasos siguientes:

- 1.- Se elige un diseño para el ajuste de un polinomio de segundo grado y con las observaciones en estos puntos se estima dicho polinomio.
- 2.- El óptimo estimado resulta ser el óptimo de la respuesta estimada.

Las estimaciones así obtenidas pueden ser muy malas si el diseño está alejado del óptimo real; porque $V(\hat{\eta})$ crece conforme x se encuentre más lejos del diseño.

1.4.4 Comentarios Sobre Los Diferentes Métodos

El método aleatorio Se recomienda usar este método como un proceso exploratorio, tanto para determinar el punto de inicio del método de factor único, como para encontrar el centro del diseño en los métodos de un sólo experimento y de ascenso por pendiente máxima.

El método por sí sólo, proporciona para cada N , bajo ciertas hipótesis adicionales, una región de confianza para x_{op} ; pero, es necesario una muestra muy grande para que la confianza sea aceptable.

Por ejemplo: Con cuatro factores y A particionada con diez sub-divisiones por coordenada (i.e. $N = 10$ y $m = 4$); la probabilidad de que al menos uno de los puntos en la muestra esté en la misma celdilla de x_{op} (siempre y cuando $x_{op} \in A$) es:

$$1 - (1 - 1/N^m)^n = 1 - (1 - .1^4)^n = 1 - .99999^n.$$

para tener una confianza de 95%, el tamaño de la muestra debe cumplir la relación: $1 - .99999^n = .95$; y esto se da si:

$$n = \frac{\log(0.05)}{\log(.99999)} = 299571,$$

y para tener una confianza del 90%, $n = 230256$. Con un tamaño muestral así, el método resulta necesariamente caro.

Cabe aclarar que la evaluación de esta probabilidad supone que el error de observación tiene varianza pequeña, de modo que el punto \mathbf{x}_i con respuesta observada mayor realmente está próximo a \mathbf{x}_{op} . Si esto no ocurre la confiabilidad de la región es menor y de hecho no se puede calcular.

El método de un sólo experimento

Si el diseño se localiza cerca de \mathbf{x}_{op} las estimaciones obtenidas con este método coinciden con las estimaciones obtenidas con el método de ascenso por pendiente máxima; pero, si el diseño se encuentra lejos del óptimo, las estimaciones pueden diferir mucho, porque mientras $\hat{\mathbf{x}}_{op}$ este más lejos del diseño, menos confiable es $\hat{\eta}$ evaluada en $\hat{\mathbf{x}}_{op}$. Si se aumenta el área que abarca el diseño, es decir, los puntos se separan más; se puede evitar, parcialmente esto; sin embargo, cuando el área de diseño es grande la superficie de respuesta en esta región puede ser muy complicada y quizás una función cuadrática no sea lo más adecuado para describirla, en estos casos se necesita un diseño con más puntos para ajustar un polinomio de mayor grado.

El método de factor único

Este método es muy sencillo porque descompone el proceso en sus componentes elementales. Se ha visto en la práctica, que si los efectos de cada factor en lo individual, son independientes de los otros factores, el método da buenos resultados. Esto significa que la búsqueda realizada sobre cada factor, manteniendo a los otros fijos, va conduciendo paulatinamente al máximo de manera independiente. Pero, si los factores interactúan fuertemente entre sí, es necesario una muestra mayor, e inclusive, puede registrarse un óptimo sin que éste exista, principalmente si se está en una cordillera ascendente, como se ve en la figura (1.4.1)

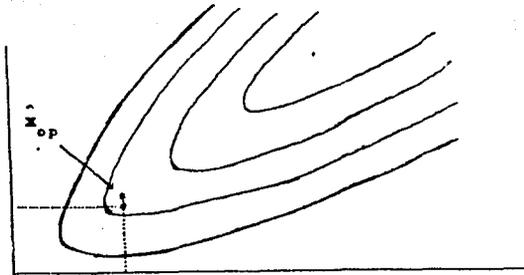


Figura 1.4.1: Curvas de nivel de una cordillera ascendente en dirección de $(x,y) \rightarrow (\infty,\infty)$. Se muestra la búsqueda del óptimo usando el método de factor único. Se busca el punto más alto manteniendo constante la segunda coordenada, éste se localiza cerca del lomo de la cordillera. En dirección de los ejes, el máximo se encuentra cerca del primer punto localizado.

Capítulo 2

Cálculo de Condiciones de Operación Óptimas

En el trabajo experimental, es frecuente que el investigador este interesado en buscar condiciones de experimentación que sean más deseables, las cuales dependan de algún criterio preseleccionado. Por ejemplo, vender un producto con gran pureza (alta concentración de componentes primarios), es un problema de mucha importancia para la industria química. Un investigador metalúrgico estará interesado en obtener el porcentaje de ciertas aleaciones que resulte en mínima corrosión. Como se mencionó en el capítulo anterior el enfoque para tratar estos problemas es el de los métodos de Superficie de Respuesta. Es muy común, que más de una respuesta sea de interés. Por ejemplo, el propósito del investigador puede ser maximizar una respuesta, Y_1 , mientras mantiene una segunda respuesta, Y_2 , tan pequeña como sea posible. En la practica, el problema de multirespuesta frecuentemente necesita un compromiso en la determinación final de las condiciones de operación¹ ([27], pág. 257).

En síntesis, el problema es determinar bajo qué niveles las variables independientes proporcionan una óptima respuesta, es decir, una respuesta máxima o mínima.

Esto puede lograrse con la ayuda de un modelo matemático que relacione las variables dependientes con las independientes. El modelo se construye utilizando un diseño de experimentos y de aquí se obtiene un óptimo estimado. Cabe señalar que existen métodos que no requieren de dicho ajuste como es el caso de las Tangentes Paralelas, el cual no se presenta en este trabajo, pero una muy buena referencia está dada en [9, 20].

¹cabe señalar que los problemas de multirespuesta no se abarcarán en este trabajo

El investigador tratará de obtener su ajuste usando una zona experimental donde se encuentre el óptimo. La elección de ésta, depende en gran medida del conocimiento previo que tenga el investigador acerca del problema.

La intención de este capítulo, es entonces, proporcionar métodos para localizar condiciones óptimas. Dos situaciones distintas son consideradas:

- 1.- El caso en el cual la región experimental usada, está dentro de una vecindad general del óptimo, y
- 2.- El óptimo real está fuera de la región experimental (región de diseño).

Desde luego, en la práctica, el experimentador espera que la segunda situación no ocurra. Sin embargo, un experimentador prudente podría planear su investigación de manera secuencial, es decir, dado que hay completa ignorancia de cuál sea la mejor región, comenzaría con una región muy alejada del óptimo y sistemáticamente trabajar para acercarse a las condiciones deseables. En la práctica, usualmente se inicia con diseños de primer orden y en la cercanía del óptimo se usa un diseño de segundo orden, rara vez, de tercer orden. Esto, ya que se considera que el óptimo será un máximo o mínimo absoluto.

2.1 DETERMINACIÓN DE PUNTO CRÍTICO

En esta sección se considera la situación en la cual el experimentador ha seleccionado una región de experimentación. Una región en la cual él cree sea factible encontrar el óptimo.

El caso más frecuente es considerar respuestas de tipo cuadrático en la cercanía de los puntos críticos. Lejos de los puntos críticos se consideran respuestas lineales (planos), entonces el proceso de obtención del punto crítico se aplicará a modelos de segundo grado, en el caso más general. Así, si se tienen k variables independientes, el modelo de segundo grado es

$$Y_i = \eta_i + \epsilon_i$$

donde

$$\eta_i = \beta_0 + \sum_{j=1}^k \beta_j x_j + \underbrace{\sum_{j=1}^k \sum_{m=1}^k \beta_{jm} x_{jm}}_{j < m} + \sum_{j=1}^k \beta_j x_j^2. \quad (2.1.1)$$

Un modelo de este tipo se puede aplicar por la siguiente razón: Cualquier superficie *sin picos* se puede aproximar mediante un polinomio en varias variables, según el TEOREMA DE TAYLOR.

Con los valores experimentales, se obtiene un estimador de η_i por el Método de Mínimos Cuadrados, el que se denotará como

$$\hat{\eta} = b_0 + \sum_{j=1}^k b_j x_j + \underbrace{\sum_j^k \sum_j^k}_{j < m} b_{jm} x_j x_m + \sum_j^k b_{jj} x_j^2, \quad (2.1.2)$$

donde b_i es el estimador por mínimos cuadrados de β_i .

El objetivo es maximizar o minimizar $\hat{\eta}$ en términos de x_1, \dots, x_k . Aquí se supone que el modelo no presenta carencia de ajuste y es significativo. Esto quiere decir que η está bien aproximada por $\hat{\eta}$. Puesto que η es lineal en los parámetros, se aplica la teoría del modelo de regresión lineal general en vías de verificar la calidad del ajuste.

El modelo (2.1.2), puede expresarse en términos matriciales como

$$\hat{\eta} = b_0 + \mathbf{x}^t \mathbf{b} + \mathbf{x}^t \mathbf{B} \mathbf{x}, \quad (2.1.3)$$

donde

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & \frac{b_{12}}{2} & \dots & \frac{b_{1k}}{2} \\ \frac{b_{12}}{2} & b_{22} & \dots & \frac{b_{2k}}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b_{1k}}{2} & \frac{b_{2k}}{2} & \dots & b_{kk} \end{bmatrix},$$

y \mathbf{x}^t denota el transpuesto del vector \mathbf{x} .

Dado que $\hat{\eta}$ es derivable sobre todo \mathbb{R}^k , si existe un máximo o un mínimo, ya sea local o global se encontrará entre los puntos tales que $\nabla \hat{\eta} = 0$.

Los puntos que satisfacen la última condición se conocen como puntos críticos o estacionarios de la superficie ajustada. El término estacionario se debe a que $\hat{\eta}$ no cambia cuando \mathbf{x} experimenta cambios pequeños. No todo punto estacionario está forzado a ser un óptimo (máximo o mínimo) y a los puntos que no son óptimos se les denomina puntos silla.

Para obtener los puntos estacionarios, hay que calcular $\nabla \hat{\eta} = \frac{\partial \hat{\eta}}{\partial \mathbf{x}}$ (que denota a $\nabla \hat{\eta}$ como un vector columna) y luego igualar a cero el gradiente. Usando la linealidad del

operador derivada se tiene

$$\frac{\partial \hat{\eta}}{\partial \mathbf{x}} = \frac{\partial b_0}{\partial \mathbf{x}} + \frac{\partial(\mathbf{x}'\mathbf{b})}{\partial \mathbf{x}} + \frac{\partial(\mathbf{x}'B\mathbf{x})}{\partial \mathbf{x}}. \quad (2.1.4)$$

Note que cada una de las derivadas tiene sentido puesto que sus argumentos son funciones escalares. Entonces utilizando la proposición 1 del Apéndice A en la ecuación (2.1.4) se tiene que

$$\frac{\partial \hat{\eta}}{\partial \mathbf{x}} = \mathbf{b} + 2B\mathbf{x}. \quad (2.1.5)$$

Igualando (2.1.5) a cero, se obtienen los puntos estacionarios, y entonces

$$\frac{\partial \hat{\eta}}{\partial \mathbf{x}} = \mathbf{b} + 2B\mathbf{x} = 0,$$

implica que

$$\mathbf{x}_0 = -\frac{B^{-1}\mathbf{b}}{2}, \quad (2.1.6)$$

siempre que exista B^{-1} . Si B^{-1} no existe, el problema no tiene solución o tiene una infinidad de soluciones. En este caso habría que verificar de acuerdo con el contexto del problema, si es correcto pensar en la inexistencia del óptimo o en su falta de unicidad. De no ser así, el error se deberá a la propuesta del modelo o al diseño usado para construirlo.

Para saber si se trata de un máximo, un mínimo o un punto silla, se puede estudiar la matriz de segundas derivadas o matriz Hessiana. Si la matriz Hessiana es definida positiva se tiene un mínimo, si es negativa definida se tiene un máximo y si es cero un punto silla. Esta manera de resolver el problema, podría presentar entre otras las siguientes dificultades.

- 1.- A medida que el número de variables crece, el cálculo del Hessiano es cada vez más difícil.
- 2.- No se puede analizar el comportamiento de la función $\hat{\eta}$ en las cercanías del punto \mathbf{x}_0 .

Con esto en mente, se discutirá una estrategia que permitirá caracterizar a \mathbf{x}_0 de manera más fácil. Esta estrategia es la llamada Transformación Canónica, que se estudiará en la siguiente sección.

Cuando x_0 es un óptimo muchas veces es de interés conocer la dirección de menor o mayor decremento ya que en una situación real el punto óptimo obtenido podría ser inaceptable por cuestiones presupuestales. En este caso, se buscaría algún punto que tuviera asociada una respuesta con aproximadamente el mismo valor y además que se considerara aceptable. En tal situación, es muy conocido que el negativo del gradiente de una función, evaluado en un punto, digamos, x^* , es la dirección de máximo decremento a partir del punto. Sin embargo, de la definición de derivada direccional en un punto, se sabe que si $\nabla f(x^*) \cdot d < 0$, entonces d es una dirección de decremento a partir de x^* . De esta forma, al evaluar $\nabla \hat{\eta}(x_0)$ y tomar el producto punto con el vector d de tal manera que sea menor que cero, se obtendrían direcciones de decremento a partir del óptimo, de tal forma que sería posible seleccionar un punto, $x_0 + \lambda d$, con $\lambda \geq 0$, que tenga asociada una respuesta aceptable.

Desde luego, si el punto estacionario x_0 es un máximo, en la medida en que se aleje de éste, la respuesta decrece. Si x_0 es un mínimo en la medida en que se aleje de éste la respuesta se incrementa. Sin embargo, en el caso de un punto silla, el experimentador podría conseguir un incremento o decremento en la respuesta cuando se aleje del punto estacionario, dependiendo de la dirección que elija. Por lo cual, en este caso es de interés conocer en que dirección se deben mover las coordenadas de las x^j para optimizar la respuesta. Es decir, maximizarla, si lo que se quiere es un máximo o minimizarla en caso contrario. Estas tres situaciones son ilustradas mediante contornos de gráficas en las figuras (2.1.1), (2.1.2) y (2.1.3).

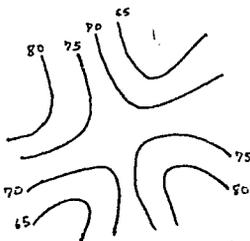


Figura 2.1.1: Punto silla en la superficie ajustada

Otra situaciones que pueden ocurrir en la práctica son las presentadas en las figuras (2.1.4) y (2.1.5).

La figura (2.1.4) muestra lo que se conoce como sistema de cordillera estacionaria,

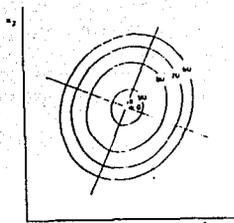


Figura 2.1.2: Punto máximo en una superficie ajustada

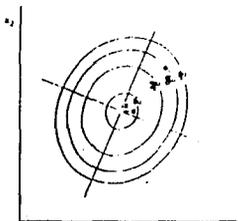


Figura 2.1.3: Punto mínimo en una superficie ajustada

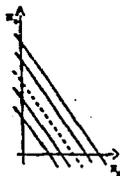


Figura 2.1.4: Sistema de Cordillera Estacionaria en una superficie ajustada

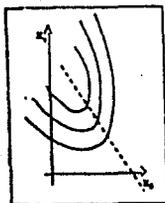


Figura 2.1.5: Sistema de Cordillera Ascendente

mientras que la figura (2.1.5) muestra uno de cordillera ascendente.

El primer caso determina que hay una región, más que un punto, en donde la respuesta se maximiza o se minimiza (aquí B^{-1} no existe). Esta región, proporciona condiciones alternativas para la optimalidad.

La cordillera ascendente indica la dirección en la cual el investigador debe mover la zona experimental, para obtener un óptimo que esté dentro de dicha región.

La detección de existencia de sistemas de cordilleras es una parte importante del análisis total de superficie de respuesta, por lo cual se discutirá a detalle en un capítulo posterior.

Como recomendación es conveniente que en el caso de trabajar con tres o menos variables independientes se grafiquen las curvas de nivel con valores cercanos a $\hat{\eta}(x_0)$, pues así se podría obtener información adicional.

2.2 TRANSFORMACION CANONICA

Esta transformación, consiste de una traslación seguida de una rotación, teniendo por objeto conocer las características de la respuesta ajustada. Para esto es necesario conocer previamente un punto estacionario. La traslación se realiza moviendo el origen del sistema al punto estacionario ubicado en $x_0 = -\frac{B^{-1}b}{2}$, obteniéndose así la siguiente expresión:

$$\hat{\eta} = \hat{\eta}_0 + z^t B z,$$

mediante el siguiente análisis.

De la ecuación (2.1.3) el valor estimado de la respuesta en el punto estacionario es

$$\hat{\eta} = b_0 + x_0^t b + x_0^t B x_0,$$

como $x_0 = -\frac{B^{-1}b}{2}$, se tiene

$$\begin{aligned}\hat{\eta}_0 &= b_0 + \left[-\frac{B^{-1}b}{2}\right]^t b + \left[-\frac{B^{-1}b}{2}\right]^t B \left[-\frac{B^{-1}b}{2}\right] \\ &= b_0 - \frac{b^t B^{-1}b}{2} + \frac{b^t B^{-1}b}{4} \\ &= b_0 - \frac{b^t B^{-1}b}{4},\end{aligned}$$

y nuevamente como $\mathbf{x}_0 = -\frac{B^{-1}\mathbf{b}}{2}$,

$$\hat{\eta}_0 = b_0 + \frac{\mathbf{x}_0^t \mathbf{b}}{2}. \quad (2.2.1)$$

Definiendo $\mathbf{z} = \mathbf{x} - \mathbf{x}_0$, o sea $\mathbf{x} = \mathbf{z} + \mathbf{x}_0$ (aquí las \mathbf{z} son las variables correspondientes a los ejes trasladados), en (2.1.3) se tiene

$$\begin{aligned} \hat{\eta} &= b_0 + (\mathbf{z}^t + \mathbf{x}_0^t)\mathbf{b} + (\mathbf{z}^t + \mathbf{x}_0^t)B(\mathbf{z} + \mathbf{x}_0) \\ &= b_0 + \mathbf{x}_0^t \mathbf{b} + \mathbf{x}_0^t B \mathbf{x}_0 + \mathbf{z}^t \mathbf{b} + \mathbf{z}^t B \mathbf{x}_0 + \mathbf{x}_0^t B \mathbf{z} + \mathbf{z}^t B \mathbf{z} \\ &= \hat{\eta}_0 + \mathbf{z}^t \mathbf{b} + \mathbf{z}^t B \mathbf{x}_0 + \mathbf{x}_0^t B \mathbf{z} + \mathbf{z}^t B \mathbf{z}, \end{aligned}$$

dado que $\mathbf{z}^t B \mathbf{x}_0$ y $\mathbf{x}_0^t B \mathbf{z}$ son equivalentes, puesto que uno es el transpuesto del otro, B es simétrica y el producto es un real, se obtiene que

$$\begin{aligned} \hat{\eta} &= \hat{\eta}_0 + \mathbf{z}^t \mathbf{b} + 2\mathbf{z}^t B \mathbf{x}_0 + \mathbf{z}^t B \mathbf{z} \\ &= \hat{\eta}_0 + \mathbf{z}^t (\mathbf{b} + 2B \mathbf{x}_0) + \mathbf{z}^t B \mathbf{z} \\ &= \hat{\eta}_0 + \mathbf{z}^t \left[\mathbf{b} - 2B \frac{B^{-1}\mathbf{b}}{2} \right] + \mathbf{z}^t B \mathbf{z} \\ &= \hat{\eta}_0 + \mathbf{z}^t (\mathbf{b} - \mathbf{b}) + \mathbf{z}^t B \mathbf{z} \\ &= \hat{\eta}_0 + \mathbf{z}^t B \mathbf{z}. \end{aligned} \quad (2.2.2)$$

Al hacer la traslación de los ejes al punto \mathbf{x}_0 , se transforma el modelo original en uno equivalente, que tiene una forma analítica más simple, en la que no aparecen términos lineales.

Lo que se requiere ahora es una rotación de ejes que elimine los productos cruzados de la forma cuadrática $\mathbf{z}^t B \mathbf{z}$. Esto se logra utilizando el Teorema de la Descomposición Espectral descrito en el Apéndice B. Así

$$\mathbf{z}^t B \mathbf{z} = \mathbf{z}^t U \Lambda U^t \mathbf{z} = \mathbf{w}^t \Lambda \mathbf{w} \quad \text{con } \mathbf{w} = U^t \mathbf{z}, \quad (2.2.3)$$

con U matriz ortogonal. Como Λ es una matriz diagonal de valores propios, entonces

$$\mathbf{w}^t \Lambda \mathbf{w} = \sum_{i=1}^k \delta_i w_i^2,$$

y se tiene que

$$\begin{aligned}\hat{\eta} &= \hat{\eta}_0 + \mathbf{z}' B \mathbf{z} \\ &= \hat{\eta}_0 + \sum_{i=1}^k \delta_i w_i^2.\end{aligned}\tag{2.2.4}$$

De donde es más fácil saber si en realidad el punto estacionario \mathbf{x}_0 corresponde a un máximo, a un mínimo o a un punto silla. Además, las variables w^s determinarán los ejes principales de las curvas de nivel de la función de respuesta.

Ahora se analizará la ecuación (2.2.4) por casos para poder especificar la naturaleza del punto estacionario. En lo sucesivo se supone que el punto estacionario está dentro de la región experimental a menos que se indique lo contrario. Recuerde que el efecto del punto crítico se encuentra encerrado únicamente en la función $\hat{\eta}_0$, por lo cual sólo será necesario analizar el término $\sum \delta_i w_i^2$.

CASO I:

Si en la ecuación (2.2.4) $\delta_i < 0$, para toda $i = \overline{1, k}$, entonces \mathbf{x}_0 corresponde a un valor máximo de $\hat{\eta}$, pues los términos $\delta_i w_i^2 < 0$, para toda $i = \overline{1, k}$ y la $\sum_{i=1}^k \delta_i w_i^2 < 0$ lo que haría que

$$\hat{\eta}_0 + \sum_{i=1}^k \delta_i w_i^2 < \hat{\eta}_0 = \hat{\eta}(\mathbf{x}_0).$$

Es decir, al alejarse de \mathbf{x}_0 , sin importar la magnitud del desplazamiento, se provoca un decremento en $\hat{\eta}$. Si $k = 2$ las curvas de nivel, en términos de las variables w^s , se verán como en la figura (2.2.1).

Si $|\delta_2|$ es considerablemente mayor que $|\delta_1|$, en la dirección del eje w_2 habrá un decremento más acelerado de la respuesta comparado con el descenso ocurrido al avanzar sobre w_1 . Esto provoca que las curvas de nivel se alarguen en la dirección de w_2 .

Cuando $k = 3$ las superficies de nivel presentan la forma que aparece en la figura (2.2.2).

Igualmente se pueden comparar las magnitudes $|\delta_i|$ dos a dos y ver hacia donde se alargan las superficies.

En general, $|\delta_i|$ mide la magnitud de decremento de la respuesta $\hat{\eta}$ en la dirección w_i .

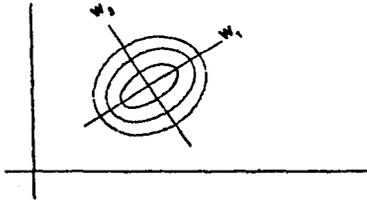


Figura 2.2.1: Mximo alargamiento a lo largo del eje w_2

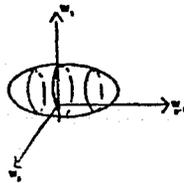


Figura 2.2.2:

La expresión (2.2.4) considerando este caso, presenta una manera muy sencilla para lograr un análisis de sensibilidad. Si existe al menos un sub-índice $j \in \{1, \dots, k\}$ tal que δ_j es aproximadamente cero y además $\delta_i < 0$, para toda $i = \overline{1, k}$, entonces \mathbf{x}_0 corresponde a un valor máximo $\hat{\eta}$, pero para propósitos prácticos no es único, sino que habrá una infinidad de puntos que dan aproximadamente el valor máximo $\hat{\eta}$, todos ellos en el hiperplano determinado por las w_j tales que sus correspondientes δ_j son casi cero. Cuando se presentan estos casos, se dice que hay una cordillera estacionaria.

Para $k = 2$ y en el límite, es decir, cuando $\delta_1 = 0$, $\delta_2 < 0$, la región de posibles máximos tiende a una recta, ver figura (2.1.4). Cuando se está cerca de dicho límite al desplazarse sobre el eje w_1 no hay cambios significativos en la respuesta $\hat{\eta}$ según la ecuación (2.2.4). Así pues, es posible encontrar valores en las \mathbf{x} que den valores cercanos al óptimo en la práctica. Esto puede hacerse ya que se conoce que $\mathbf{z} = U\mathbf{w}$ y $\mathbf{x} = \mathbf{z} + \mathbf{x}_0$.

Para $k = 3$ cuando $\delta_2 = \delta_3 = 0$, $\delta_1 < 0$, se tiene un plano, como se ve en la figura (2.2.3).

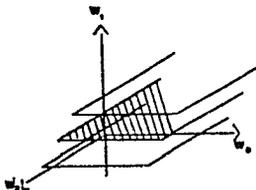


Figura 2.2.3:

Cuando $\delta_1, \delta_2 < 0$, $\delta_3 = 0$ la región de posibles máximos tiende a ser el eje w_3 encerrado por curvas de nivel en forma de cilindro, figura (2.2.4).

La importancia de hacer $\delta_j = 0$ radica en la facilidad para entender los casos en que δ_j es muy cercana a cero. Para realizar la Transformación Canónica se requiere que B^{-1} exista, que ocurre si y sólo si, todos los valores propios de B son diferentes de cero [19]

CASO II:

Si en la ecuación (2.2.4), $\delta_i > 0$, para toda $i = \overline{1, k}$ entonces \mathbf{x}_0 corresponde a un mínimo de $\hat{\eta}$, pues $\delta_i w_i^2 > 0$, para toda $i = \overline{1, k}$, lo cual implica que $\sum_{i=1}^k \delta_i w_i^2 > 0$ y

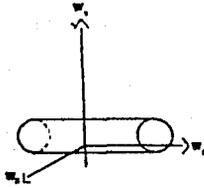


Figura 2.2.4:

por lo tanto

$$\hat{\eta}_0(\mathbf{x}_0) < \hat{\eta}_0(\mathbf{x}_0) + \sum_{i=1}^k \delta_i w_i^2,$$

con un análisis similar al del CASO I.

También se observa que si en la ecuación (2.2.3) al menos un sub-índice $j \in \{1, \dots, k\}$ es aproximadamente cero y $\delta_i > 0$, para toda $i = \overline{1, k}$, entonces se tiene una cordillera estacionaria de donde es posible escoger los puntos que más convengan.

CASO III:

Si en la ecuación (2.2.4) hay algun(os) valor(es) propio(s) positivo(s) y otro(s) negativo(s), entonces \mathbf{x}_0 es un punto silla. Esto quiere decir que a partir de \mathbf{x}_0 , existen direcciones en las que al desplazarse, la función $\hat{\eta}$ crece, pero también existen otras direcciones en las cuales $\hat{\eta}$ decrece.

Si fuera de interés, por ejemplo, obtener el punto que corresponde al máximo, se debería avanzar sobre las direcciones en las que $\hat{\eta}$ crece. Estas direcciones están dadas por las variables \mathbf{w}^s , gracias a la ecuación (2.2.3). Como se mencionó anteriormente los equivalentes de las variables \mathbf{w}^s en las \mathbf{x}^s se obtienen mediante las relaciones $\mathbf{x} + \mathbf{x}_0 = \mathbf{z} = U\mathbf{w}$.

Para terminar con este análisis, considere el caso donde el punto estacionario está alejado de la región experimental en que se ajustó el modelo. Se dice que se tiene una cordillera ascendente(descendente) cuando \mathbf{x}_0 corresponde a un valor máximo (mínimo) de $\hat{\eta}$, pero con la particularidad de que existe al menos un sub-índice $j \in \{1, \dots, k\}$ tal que δ_j es casi cero. Los casos más comunes de este tipo de cordilleras en forma gráfica son los presentados en la figura (2.2.5)

Donde lo que se recomienda es desplazarse sobre el eje w_i , siempre que $|\delta_i|$ tienda a

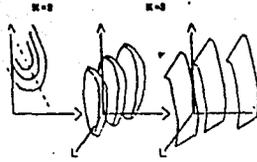


Figura 2.2.5:

ceros, para así poder determinar nuevas zonas de experimentación.

Cuando se está en este tipo de casos, debe volver a desarrollarse todo el procedimiento antes descrito: seleccionar la nueva región experimental, ajustar un nuevo modelo de segundo orden, obtener el punto estacionario correspondiente y finalmente bajo la Transformación Canónica determinar la naturaleza del punto estacionario usando la ecuación (2.2.4). Teóricamente, este algoritmo debe emplearse hasta que el punto estacionario que se obtenga sea un máximo o un mínimo, aunque en la práctica el costo y el tiempo limitan el número de experimentos posibles.

Para esclarecer el uso de la técnica se presenta el siguiente ejemplo.

Ejemplo (2.2.1): Myers [1971]

Un investigador que analiza la producción de pan de caja en cierta fábrica, está interesado en saber cuál será la máxima fuerza de sellado, Y (en gramos por pulgada), para el empaquetado de dicho producto. De esta manera podría garantizar el menor deterioro posible de la mercancía. El investigador conoce que las principales variables relacionadas con la fuerza de sellado en el proceso de empaque, son la temperatura de sellado, (ξ_1) , la temperatura de enfriamiento de la barra de pan, (ξ_2) , y el porcentaje de polietileno aditivo, (ξ_3) .

El investigador utilizando un procedimiento previo (como el presentado en la sección 2.3), conoce una región de puntos (diseño experimental) para las variables ξ_1, ξ_2, ξ_3 en donde posiblemente se encuentre el valor máximo de la fuerza de sellado, habiendo establecido una aproximación a la relación funcional entre dichas variables. La aproximación de la variable Y , en base a ξ_1, ξ_2, ξ_3 se hizo con un ajuste de un modelo de segundo orden, esto para que tenga sentido la posible existencia de un máximo.

El investigador decide utilizar la técnica de Transformación Canónica para obtener

el valor óptimo de la variable Y, o bien, obtener valores cercanos al óptimo. El análisis fue el siguiente:

Las variables originales fueron codificadas utilizando las fórmulas

$$x_1 = \frac{\xi_1 - 225}{30}, \quad x_2 = \frac{\xi_2 - 55}{9}, \quad x_3 = \frac{\xi_3 - 1.1}{0.6}.$$

Fueron usados en el diseño experimental cinco diferentes niveles para las variables. Los niveles para las variables codificadas y originales estan dadas en la tabla siguiente

	-1.682	-1.000	0.000	1.000	1.682
x_1	204.5	225	255	285	305.5
x_2	39.9	46	55	64	70.1
x_3	0.09	0.5	1.1	1.7	2.11

La matriz del diseño fue originada de un diseño central compuesto, diseños que se consideran adecuados para ajustar una superficie de segundo orden [33].

La matriz de diseño X (en términos de variables codificadas) y el correspondiente vector Y de observaciones son:

$$D = \begin{bmatrix} x_1 & x_2 & x_3 \\ -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \\ -1.682 & 0 & 0 \\ +1.682 & 0 & 0 \\ 0 & -1.682 & 0 \\ 0 & +1.682 & 0 \\ 0 & 0 & -1.682 \\ 0 & 0 & +1.682 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 6.6 \\ 6.9 \\ 7.9 \\ 6.1 \\ 9.2 \\ 6.8 \\ 10.4 \\ 7.3 \\ 9.8 \\ 5.0 \\ 6.9 \\ 6.3 \\ 4.0 \\ 8.6 \\ 10.1 \\ 9.9 \\ 12.2 \\ 9.7 \\ 9.7 \\ 9.6 \end{bmatrix}.$$

Para el modelo de segundo orden

$$\hat{\eta} = b_0 + \sum_{i=1}^3 b_i x_i + \sum_{i=1}^3 b_{ii} x_i^2 + \sum_{i=1}^3 \sum_{j=1}^3 b_{ij} x_i x_j,$$

matriz X es

$$X = \begin{bmatrix} 1 & x_1 & x_2 & x_3 & x_1^2 & x_2^2 & x_3^2 & x_1 x_2 & x_1 x_3 & x_2 x_3 \\ 1 & & & & +1 & +1 & +1 & +1 & +1 & +1 \\ 1 & & & & +1 & +1 & +1 & -1 & -1 & +1 \\ 1 & & & & +1 & +1 & +1 & -1 & +1 & -1 \\ 1 & & & & +1 & +1 & +1 & +1 & -1 & -1 \\ 1 & & & & +1 & +1 & +1 & +1 & -1 & -1 \\ 1 & & & & +1 & +1 & +1 & -1 & +1 & -1 \\ 1 & & & & +1 & +1 & +1 & -1 & -1 & +1 \\ 1 & & & & +1 & +1 & +1 & +1 & +1 & +1 \\ 1 & & D & & 2.828 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 2.828 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 2.828 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 2.828 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 2.828 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 2.828 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & & & & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

estimando por mínimos cuadrados los parámetros, esto es resolviendo las ecuaciones normales dadas por

$$(X^t X) \hat{\beta} = X^t Y,$$

con

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_{11} \\ \beta_{22} \\ \beta_{33} \\ \beta_{12} \\ \beta_{13} \\ \beta_{23} \end{bmatrix},$$

se tiene que

$$\hat{\eta} = 10.1657 - 1.1038x_1 + 0.0872x_2 + 1.0206x_3 \\ - 0.7602x_1^2 - 1.430x_2^2 - 1.14x_3^2 \\ - 0.35x_1x_2 - 0.5x_1x_3 + 0.1500x_2x_3.$$

Se presenta la tabla de Análisis de Varianza (incluyendo el renglón correspondiente a la Carencia de Ajuste).

Fuente	Suma de cuadrados	Grados de Libertad	Cuadrado medio	F
regresión	70.3056	9	7.817	
error	11.8644	10	1.1864	6.58
carencia de ajuste	6.9044	5	1.3809	
error puro	4.9600	5	.9920	1.39
total	82.1700	19		

Consultando tablas de F , a un nivel de significancia $\alpha = 0.5$, se obtiene que $F_{m-p, n-m}^{\alpha} = F_{8,5}^{0.5} = 5.05$. Como $F_{CA} = 1.39 < 5.05$, se concluye que no hay carencia de ajuste. Dado que no se presenta la falta de ajuste, se procede a calcular el cuantil de significancia global, que es $F_{p, n-p}^{\alpha} = F_{9,10}^{0.5} = 3.02$. Como $F_{SG} = 6.58 > 3.02$, se concluye que el modelo es significativo, y en resumen el modelo ajustado es adecuado.

Se calcula el punto estacionario, utilizando la ecuación

$$\mathbf{x}_0 = -\frac{B^{-1}\mathbf{b}}{2}$$

llegando a que

$$\mathbf{x}_0 = \begin{bmatrix} x_{1,0} \\ x_{2,0} \\ x_{3,0} \end{bmatrix} = \begin{bmatrix} -1.0098 \\ 0.2602 \\ 0.6808 \end{bmatrix}.$$

Decodificando las variables para obtener el punto estacionario en las variables originales se tiene

$$\xi_0 = \begin{bmatrix} \xi_{10} \\ \xi_{20} \\ \xi_{30} \end{bmatrix} = \begin{bmatrix} 194.70 \\ 57.34 \\ 1.51 \end{bmatrix},$$

con respuesta estimada $\hat{\eta}_0 = 11.80$.

Para calcular los valores propios de la matriz B , δ_1 , δ_2 , δ_3 , se utiliza la ecuación

$$\begin{vmatrix} (-0.7602 - \delta) & -0.1750 & -0.2500 \\ -0.1750 & (-1.043 - \delta) & 0.0750 \\ -0.2500 & 0.0750 & (-1.1491 - \delta) \end{vmatrix} = 0$$

de donde los valores propios son

$$\delta_1 = -0.5630 \quad \delta_2 = -1.2712 \quad \delta_3 = -1.1172$$

dando lugar a la forma canónica

$$\hat{\eta} = 11.08 - 0.5630w_1^2 - 1.2712w_2^2 - 1.1172w_3^2.$$

Puesto que todas las δ_i son negativas el punto estacionario es un máximo. Note que este punto estacionario está dentro de la región experimental y que las curvas de nivel de la respuesta se alargan en dirección del eje w_1 que corresponde al valor propio más pequeño. La respuesta es igualmente sensitiva en cambios a lo largo de los ejes w_2 y w_3 ya que $|\delta_2| \approx |\delta_3|$.

Interpretando el resultado, se necesita una temperatura de sellado de $195^\circ C$, que la barra de pan se enfrie a una temperatura de $57^\circ C$ y añadir 1.5% de polietileno aditivo para obtener una fuerza de sellado aproximadamente máxima de 11.80 gramos por pulgada en el empaquetado.

Quizás sea de interés conocer la relación entre las variables w^s y las variables x^s . Recuerdese que $z = Uw$.

Para encontrar la columna $U_{(i)}$ de la matriz U , debe resolverse el sistema de ecuaciones $(B - \delta_i I_k)U_{(i)} = 0$. u_{ij} denotará el elemento en el i -ésimo renglón y en la j -ésima columna de U . Para $\delta_1 = -.5630$ se tiene que

$$\begin{bmatrix} (b_{11} + 0.5630) & \frac{b_{12}}{2} & \frac{b_{13}}{2} \\ \frac{b_{12}}{2} & (b_{22} + .5630) & \frac{b_{23}}{2} \\ \frac{b_{13}}{2} & \frac{b_{23}}{2} & (b_{33} + .5630) \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{21} \\ u_{31} \end{bmatrix} = 0$$

al resolver el sistema las soluciones deben ser normalizadas, por ejemplo, para $U_{(1)}$ debe ocurrir que

$$u_{11}^2 + u_{21}^2 + u_{31}^2 = 1.$$

La forma de resolver estas ecuaciones es dar un valor arbitrario a una de las incógnitas reduciendo la dimensión del sistema. Se puede sustituir $u'_{31} = 1$ en u_{31} y obtener las soluciones $u'_{11} = -2.0791$, $u'_{21} = .9142$.

Es fácil verificar que siempre que se de un valor a u_{31} diferente de cero se pueden encontrar u_{21} y u_{11} que sean solución al sistema. Esto se debe a que las ecuaciones que se desean resolver determinan un subespacio de dimensión uno generado por un vector con tercera componente distinta de cero. De aquí que las soluciones no sean únicas. Para la solución u'_{31} , u'_{21} , u'_{11} , los valores u_{31} , u_{21} , u_{11} se obtienen dividiendo u'_{31} , u'_{21} y u'_{11} entre $\sqrt{u'_{31} + u'_{21} + u'_{11}} = \sqrt{6.1584}$, dando como resultado

$$u_{11} = -.8378, \quad u_{21} = 0.3684, \quad u_{31} = 0.4030.$$

Se aplicó el mismo procedimiento para δ_2 y δ_3 resultando la matriz ortogonal de vectores propios igual a

$$U = \begin{bmatrix} -0.8378 & 0.4535 & -0.3052 \\ 0.3684 & 0.0552 & -0.9281 \\ 0.4030 & 0.8895 & 0.2132 \end{bmatrix}$$

Finalmente usando la ecuación (2.2.3)

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -0.8378 & 0.4535 & -0.3052 \\ 0.3684 & 0.0552 & -0.9281 \\ 0.4030 & 0.8895 & 0.2132 \end{bmatrix} \begin{bmatrix} (x_1 + 1.0098) \\ (x_2 - 0.2602) \\ (x_3 - 0.6808) \end{bmatrix}.$$

Si x_0 representa condiciones de costo muy elevadas se pueden encontrar vectores x que den casi la misma respuesta y no sean tan caros. Esto se logra de la siguiente manera.

Al observar la ecuación:

$$\hat{\eta} = 11.08 - 0.5630w_1^2 - 1.27w_2^2 - 1.1172w_3^2.$$

Se nota que al desplazarse sobre alguno de los ejes w_1 , w_2 ó w_3 a partir del punto $(0,0,0)$, la respuesta decrece menos sobre la dirección de w_1 . Así que se dan valores a w_1 cercanos a cero y se consideran vectores de la forma $(w_1, 0, 0)$ con sus correspondientes (x_1, x_2, x_3) , los vectores con entradas x^i son máximos aproximados. Dentro de estos

últimos se seleccionan los más adecuados. La siguiente tabla presenta algunos máximos aproximados:

w_1	.1	.2	.3	.4	.5
x_1	-1.093	-1.1773	-1.2611	-1.3449	-1.4287
x_2	.2970	.3338	.3707	.40756	.4444
x_3	.7211	.7611	.8017	.8420	.8823
$\hat{\eta}$	11.074	11.057	11.029	10.989	10.939

Ejemplo (2.2.2): Myers[1971]

En un proceso químico, un experimentador desarrolla una reacción para convertir 1,2-PROPANODIOL en 2,5-DIMETIL PIPERACINA. El 1,2-PROPANODIOL es usado como anticongelante en cervezas, como solvente farmacéutico, sustituto del epilenglicol y gliserol en la manufactura de resinas sintéticas e inhibe la fermentación; es un producto de muy bajo costo. El 2,5 DIMETILPIPERACINA es un solvente farmacéutico de gran efectividad, lo cual lo hace un producto muy costoso. Se sabe que es posible obtener este último como subproducto del primero, en una reacción a la que habría que adicionar elementos de muy bajo costo. De aquí que el experimentador realice la reacción de conversión y desee obtener la mayor cantidad en gramos de 2,5 DIMETIL PIPERACINA.

El experimentador conoce que la reacción se ve afectada en su desarrollo por diferentes factores, tales como: la cantidad de amonio, NH_3 (en gramos), la temperatura, T (en $^{\circ}C$), la cantidad de agua, H_2O (en gramos), y la presión de hidrógeno, P (en psi). Por lo cual, desea determinar las condiciones en base a estos factores para las cuales se da la máxima conversión.

Si el experimentador define la respuesta, Y , como la cantidad en gramos obtenida de 2,5-DIMETIL PIPERAZINE en la reacción, y aplica la metodología de superficie de respuesta como en el ejemplo anterior, para obtener una región de puntos en donde posiblemente se encuentre el óptimo de la variable Y ; previa selección de un modelo de segundo orden para Y , en base a NH_3 , T , H_2O y P como aproximación para la relación funcional. El experimentador puede utilizar el método de la transformación canónica para obtener un óptimo o un valor cercano al óptimo de Y . El Análisis es el siguiente:

El diseño consiste de 25 puntos en las variables codificadas

$$x_1 = \frac{NH_3-102}{51}, \quad x_3 = \frac{H_2O-300}{200},$$

$$x_2 = \frac{T-250}{20}, \quad x_4 = \frac{P-850}{350}.$$

Como en el ejemplo anterior, el tipo de diseño para este modelo de segundo orden fue un diseño central compuesto. La matriz de diseño y el correspondiente vector de observaciones de respuesta estan dados como sigue

$$D = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ -1 & -1 & -1 & -1 \\ +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 \\ -1 & +1 & +1 & -1 \\ +1 & +1 & +1 & -1 \\ -1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 \\ -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & +1 \\ -1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 \\ 0 & 0 & 0 & 0 \\ -1.4 & 0 & 0 & 0 \\ +1.4 & 0 & 0 & 0 \\ 0 & -1.4 & 0 & 0 \\ 0 & +1.4 & 0 & 0 \\ 0 & 0 & -1.4 & 0 \\ 0 & 0 & +1.4 & 0 \\ 0 & 0 & 0 & -1.4 \\ 0 & 0 & 0 & +1.4 \end{bmatrix} \quad Y = \begin{bmatrix} 58.2 \\ 23.4 \\ 21.9 \\ 21.8 \\ 14.3 \\ 6.3 \\ 4.5 \\ 21.8 \\ 46.7 \\ 53.2 \\ 23.7 \\ 40.3 \\ 7.5 \\ 13.3 \\ 49.3 \\ 20.1 \\ 32.8 \\ 31.1 \\ 28.1 \\ 17.5 \\ 49.7 \\ 49.9 \\ 34.2 \\ 31.1 \\ 43.1 \end{bmatrix}$$

Los niveles para las variables codificadas y originales estan dados en la tabla siguiente

FACTOR	-1.4	-1.0	0	1.0	1.4
AMONIO (g)	30.6	51	102	153	173.4
TEMP. (°C)	222	230	250	270	278
AGUA (g)	20	100	300	500	580
PRES. DE HID. (psi)	360	500	850	1200	1340

La superficie de segundo orden ajustada por el método de mínimos cuadrados está dada por

$$\begin{aligned}\hat{\eta} = & 40.198 - 1.511x_1 + 1.284x_2 - 8.739x_3 + 4.955x_4 \\ & - 6.3324x_1^2 - 4.292x_2^2 + 0.020x_3^2 - 2.506x_4^2 + 2.194x_1x_2 \\ & - 0.144x_1x_3 + 1.581x_1x_4 + 8.006x_2x_3 + 2.806x_2x_4 + 0.294x_3x_4.\end{aligned}$$

Bajo la suposición de que el modelo es significativo, es decir, se rechaza la hipótesis nula de que el vector de parámetros \mathbf{b} es cero, se prosigue con el análisis.

De la ecuación (2.1.6), las coordenadas del punto estacionario son

$$x_0 = \begin{bmatrix} x_{1,0} = 0.265 \\ x_{2,0} = 1.034 \\ x_{3,0} = 0.291 \\ x_{4,0} = 1.668 \end{bmatrix},$$

en las variables codificadas

$$\xi_0 = \begin{bmatrix} \xi_{1,0} \\ \xi_{2,0} \\ \xi_{3,0} \\ \xi_{4,0} \end{bmatrix} = \begin{bmatrix} 115.51 \\ 270.68 \\ 358.2 \\ 1433.8 \end{bmatrix},$$

con respuesta $\hat{\eta}_0 = 43.53$.

Obteniendo los valores propios $\delta_1, \delta_2, \delta_3, \delta_4$ de la matriz

$$B = \begin{pmatrix} -6.332 & 1.0969 & -0.0720 & 0.7905 \\ 1.0969 & -4.292 & 4.0030 & 1.4030 \\ -0.0720 & 4.0030 & 0.020 & 0.1470 \\ 0.7905 & 1.4030 & 0.1470 & -2.506 \end{pmatrix},$$

de la forma canónica, se encuentra que

$$\hat{\eta} = 43.53 - 7.55w_1^2 - 6.01w_2^2 - 2.16w_3^2 + 2.60w_4^2 \quad (2.2.5)$$

lo cual dice que se tiene un punto silla, puesto que hay valores propios positivos y negativos. Observe el punto estacionario está fuera de la región de diseño, por lo que se requiere dar una mejor región experimental para ajustar nuevamente un modelo de segundo orden y obtener un punto estacionario que este dentro de la región experimental.

O bien, utilizar la ecuación (2.2.3) para encontrar posibles condiciones de operación factibles, esto es, que estén dentro de la región experimental, el procedimiento es como sigue:

La transformación que relaciona las x 's con las w 's está dada por la ecuación (2.2.3) y así

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \begin{bmatrix} 0.5977 & -0.7025 & 0.3756 & 0.0908 \\ -0.7688 & -0.4568 & 0.2858 & 0.3445 \\ 0.2151 & -0.1374 & -0.3071 & 0.9168 \\ 0.0741 & 0.5282 & 0.8264 & 0.1803 \end{bmatrix} \begin{bmatrix} x_1 - 0.265 \\ x_2 - 1.034 \\ x_3 - 0.291 \\ x_4 - 1.668 \end{bmatrix}$$

La matriz U se obtiene de manera similar al problema anterior.

La ecuación (2.2.5), muestra que un decremento en la estimación del rendimiento ocurre moviéndose del punto estacionario a lo largo de los ejes w_1 , w_2 y w_3 . Sin embargo, el rendimiento estimado aumenta desplazándose a lo largo del eje w_4 . Parece entonces razonable que el experimentador use la ecuación (2.2.3) para encontrar posibles condiciones de operación factibles, esto es, condiciones de alto rendimiento de 2,5-DIMETIL PIPERAZINE, dentro de la región experimental. Obsérvese que eventuales condiciones de operación deben ser tales que, en las variables codificadas, $x_3 \geq -1.5$, dado que un x_3 de -1.5 representa la falta de agua en el proceso. La ecuación (2.2.4) puede ser usada como en el ejemplo anterior, para calcular las condiciones x_1 , x_2 , x_3 y x_4 tomando w_1 , w_2 y w_3 como cero y dando valores (positivos y negativos) para w_4 . Los resultados se dan en la tabla siguiente

w_4	1.0	1.5	2.0	-1.0	-1.5	-2.0	-2.5
x_1	0.339	0.376	0.413	1.91	0.154	0.117	0.080
x_2	1.562	1.826	2.090	0.506	0.242	-0.022	-0.287
x_3	1.117	1.531	1.944	-0.535	-0.949	-1.362	-1.775
x_4	1.848	1.938	2.028	1.488	1.398	1.307	1.217

Note que para valores positivos de w_4 , los valores de las x 's están fuera de la región del experimento diseñado. Parecen entonces más deseables, las condiciones que resulten de tomar $w_4 \approx -2.2$, esto es, $x_1 = 0.102$, $x_2 = -0.128$, $x_3 = -1.50$ (falta de agua), y $x_4 = 1.27$. Cualquier otro punto que resulte de tomar, $w_1 = 0$, $w_2 = 0$, $w_3 = 0$ y $|w_4| > 2.2$ representaría un experimento imposible, ya que requeriría de gramos negativos de agua.

2.3 METODO DE ASCENSO POR PENDIENTE MAXIMA

En la sección anterior se consideraba que el investigador en base a su experiencia determinaba una región de experimentación adecuada, y en base a ésta, ajustaba un modelo de segundo orden para realizar posteriormente un Análisis Canónico (A.C.) en base a la Transformación Canónica.

Al hacer el análisis Canónico es importante verificar si el óptimo se encuentra o no dentro de la región de experimentación escogida. En caso de que el punto este fuera, se debe analizar el tipo de coordillera para determinar una mejor zona. Otro caso en el que se necesita localizar una mejor región de experimentación es cuando el ajuste indica que el punto crítico corresponde a un punto silla.

Bajo estas circunstancias se aplica nuevamente el procedimiento completo de ajuste y análisis de alguna superficie de segundo orden.

Debido a que, en general, en el Análisis Canónico se utilizan diseños que representan altos costos y no pocas dificultades técnicas, resulta poco práctico que el experimentador utilice el análisis secuencialmente.

De esta manera, se requiere desarrollar una técnica que minimice costos y que sea útil para ubicar alguna región donde posiblemente se encuentre el óptimo y sólo entonces aplicar el (A.C). Un procedimiento con estas características es el Método de Ascenso por Pendiente Máxima (A.P.M.).

2.3.1 Ascenso Por Pendiente Máxima: (A.P.M)

Suponga que se desea maximizar una función que depende de k variables independientes, ξ_i , $i = \overline{1, k}$, pero que se desconoce la forma analítica de la relación entre las variables independientes y la variable dependiente.

El primer paso para optimizar una superficie es proponer alguna región inicial de experimentación. Esta servirá como base para encontrar otra(s) que sean mejores, en el sentido de que contengan al óptimo.

Como también se busca disminuir los costos y los cálculos relacionados con las estimaciones de los parámetros, se propone iniciar con un diseño factorial 2^k o bien un diseño factorial fraccional. Dado que no se tiene la seguridad de que la región de

experimentación inicial determinada por un diseño de este tipo contenga al óptimo, es conveniente pensar que la superficie es monótona sobre esta región, por lo que se recomienda que en el ajuste se utilice como la media una función lineal, es decir, una función de la forma

$$\beta_0 + \sum_{i=1}^k \beta_i x_i,$$

donde x_i es la variable codificada correspondiente a ξ_i .

En los diseños factoriales 2^k , la variable codificada toma únicamente uno de dos posibles valores -1 ó 1 , siendo el centro del experimento el punto $(0, 0 \dots, 0)$, es decir, la codificación se hace de manera que la media de cada x_i sea igual a cero.

Como ejemplo, para dos variables a dos niveles cada una, se tiene lo siguiente:

ξ_1	10	20
ξ_2	2	4

Si ξ_1, ξ_2 toman los valores de la tabla anterior la codificación debe hacerse como sigue

$$x_1 = \frac{\xi_1 - \frac{(10+20)}{2}}{\frac{(20-10)}{2}} = \frac{\xi_1 - 15}{5}$$

$$x_2 = \frac{\xi_2 - \frac{(2+4)}{2}}{\frac{(4-2)}{2}} = \frac{\xi_2 - 3}{1}.$$

Note que x_1 toma los valores -1 y 1 cuando ξ_1 vale 10 y 20 respectivamente. Además el promedio de los valores de x_1 vale cero. Análogamente, la variable x_2 vale -1 y 1 cuando ξ_2 vale 2 y 4 .

Sea

$$\hat{\eta} = b_0 + \sum_{i=1}^k b_i x_i, \quad (2.3.1)$$

el modelo ajustado con b_i , para $i = \overline{1, k}$, los estimadores de mínimos cuadrados de los parámetros β_i^s , $i = \overline{1, k}$.

Para analizar el comportamiento de $\hat{\eta}$ se requiere encontrar los valores de (x_1, \dots, x_k) que maximicen $b_0 + \sum_{i=1}^k b_i x_i$, sujeto a la restricción $\sum_{i=1}^k x_i^2 = R^2$, esto es, estar en una hipersfera de radio R . Los vectores (x_1, x_2, \dots, x_k) obtenidos, determinan un camino

de ascenso máximo. Entonces, el problema es maximizar (para el método de descenso por pendiente máxima se debe minimizar) la función

$$\hat{\eta} = \beta_0 + \sum_{i=1}^k b_i x_i,$$

sujeto a

$$\sum_{i=1}^k x_i^2 = R^2.$$

La restricción se establece por razones muy simples:

1. Maximizar $\hat{\eta}$ sin ninguna condición adicional no produce soluciones satisfactorias. Cuando el modelo es un plano horizontal todos los vectores \mathbf{x} son óptimos. Cuando el plano tiene cierta inclinación $\hat{\eta}$ es no acotada.
2. Maximizar $\hat{\eta}$ restringiendo a la región de diseño puede resultar analíticamente muy complicado.
3. El valor de R es variable de tal manera que se puedan considerar todos los valores de \mathbf{x} que determinan las región experimental.

Observe que los valores originales de las variables codificadas, provenientes de los datos $\xi^{i/s}$ sólo sirven para hacer el primer ajuste de $\hat{\eta}$.

Usando el método de Multiplicadores de Lagrange para el problema de maximización restringida, se tiene la función lagrangeana asociada

$$\mathcal{L} = b_0 + \sum_{i=1}^k b_i x_i - \mu \left(\sum_{i=1}^k x_i^2 - R^2 \right).$$

Derivando parcialmente respecto a las variables $x^{i/s}$ se tiene que

$$\frac{\partial \mathcal{L}}{\partial x_j} = b_j - 2\mu x_j.$$

La derivada parcial respecto a μ es

$$\frac{\partial \mathcal{L}}{\partial \mu} = - \left(\sum_{i=1}^k x_i^2 - R^2 \right).$$

Igualando $\partial L/\partial x_j$ con cero y despejando x_j resulta

$$x_j = \frac{b_j}{2\mu} \quad j = \overline{1, k}. \quad (2.3.2)$$

elevando al cuadrado ambos lados de la igualdad y sumando sobre $j = \overline{1, k}$ se tiene

$$\sum_{j=1}^k x_j^2 = \frac{\sum_{j=1}^k b_j^2}{4\mu^2}.$$

Usando la restricción del problema se obtiene

$$R^2 = \frac{\sum_{j=1}^k b_j^2}{4\mu^2}$$

y despejando μ resulta que

$$\mu = \frac{\pm \sqrt{\sum_{j=1}^k b_j^2}}{2R}.$$

Esta última expresión indica que son dos los posibles puntos críticos. Uno correspondiente a un máximo y el otro a un mínimo, puesto que $\tilde{\eta}$ es continua y se optimiza sobre un conjunto cerrado y acotado (compacto).

Si $\mu = \frac{\sqrt{\sum_{j=1}^k b_j^2}}{2R}$ entonces,

$$x_j = \frac{Rb_j}{\sqrt{\sum_{i=1}^k b_i^2}}, \quad j = \overline{1, k}$$

sustituyendo en la función objetivo se tiene

$$\begin{aligned} \tilde{\eta}_0 &= b_0 + \sum_{i=1}^k b_i x_i \\ &= b_0 + \sum_{i=1}^k b_i \left[\frac{Rb_i}{\sqrt{\sum_{j=1}^k b_j^2}} \right] \\ &= b_0 + R \sqrt{\sum_{j=1}^k b_j^2}. \end{aligned}$$

Si $\mu = -\frac{\sqrt{\sum_{j=1}^k b_j^2}}{2R}$ entonces

$$x_j = -\frac{Rb_j}{\sqrt{\sum_{i=1}^k b_i^2}} \quad j = \overline{1, k}$$

y

$$\hat{\eta}_0 = b_0 - R \sqrt{\sum_{j=1}^k b_j^2}.$$

por lo tanto

$$\hat{\eta}_0 = b_0 + \gamma R \sqrt{\sum_{j=1}^k b_j^2}, \quad \text{con } \gamma = \begin{cases} 1 & \text{si } \mu = \frac{\sqrt{\sum_{j=1}^k b_j^2}}{2R} \\ -1 & \text{si } \mu = -\frac{\sqrt{\sum_{j=1}^k b_j^2}}{2R} \end{cases} \quad (2.3.3)$$

De aquí que el máximo se alcanza si $\mu > 0$ y el mínimo si $\mu < 0$.

En la mayoría de los casos el experimentador quiere conocer cuál es el máximo aumento en la respuesta si se da un incremento Δ_j en alguna variable ξ_j . Este incremento determina un aumento Δx_j en la correspondiente variable codificada x_j . Si se sustituye Δx_j en la expresión (2.3.2), se obtiene un valor para μ . Este valor se puede substituir en $x_i = b_i/2\mu$, $i = \overline{1, k}$, con $i \neq j$, dando lugar a aumentos en las variables independientes, $\Delta x_1, \Delta x_2, \dots, \Delta x_k$, $i \neq j$, que con Δx_j determinan respuesta máxima sujeto a que los vectores x esten en la esfera de radio $R' = \sum_{i=1}^k (\Delta x_i)^2$. Los aumentos Δx_i , $i = \overline{1, k}$, $i \neq j$, determinan aumentos Δ_i , ($i \neq j$) en las variables originales.

Se evalúa la respuesta, Y , en forma experimental en los puntos ξ^* , $\xi^* + \Delta$, $\xi^* + 2\Delta$, $\xi^* + 3\Delta$, etc, que esten fuera de la región del diseño donde ξ^* es el vector cuyas componentes son los correspondientes valores de $\xi_1, \xi_2, \dots, \xi_k$ cuando $x_1 = x_2 = \dots = x_k = 0$ y Δ el vector de incrementos $\Delta = (\Delta_1, \dots, \Delta_k)$ o *paso* de la ruta. Las evaluaciones terminan hasta notar alguna curvatura en la función, es decir, cuando Y presenta un incremento seguido de un decremento. En este momento habría que ubicar una vecindad en las variables ξ^* donde se dió la caída de Y para considerar a esta región como candidata para aplicar nuevamente todo el procedimiento hasta aquí descrito.

Este método secuencial se detiene cuando al comparar la máxima respuesta Y obtenida en las dos últimas regiones utilizadas, el cambio en Y es nulo, o casi cero,

o bien negativo. En este instante se espera tener la región adecuada para ajustar un modelo de segundo orden y realizar el Análisis Canónico.

Tradicionalmente, este es el método más utilizado para localizar regiones en las que se encuentre el óptimo de una función bajo estudio. Sin embargo, existen algunas desventajas al aplicar esta herramienta, dentro de las cuales se tienen las siguientes:

1. El punto óptimo localizado en la región obtenida, no necesariamente es un óptimo global, sino que puede en algún caso, ser local.
2. El análisis depende fuertemente de las escalas de medición de las variables; al respecto se dará el ejemplo (2.3.2). Una discusión más profunda sobre el tema puede encontrarse en [3], en particular el comentario hecho por N.L. Johnson.
3. En la medida que se procede con el método se va perdiendo precisión, al grado que es necesario hacer pruebas de hipótesis para saber si los términos de primer orden siguen dominando.

Para fijar ideas sobre el desarrollo del algoritmo se brinda el siguiente ejemplo, el cual no tiene valor práctico, pero esclarece los pasos del método de ascenso por pendiente máxima.

Ejemplo (2.3.1): Myers [1971]

$\xi_1, \xi_2, \xi_3, \xi_4$ son las variables que determinan una cierta respuesta. Los niveles alto y bajo de cada una están dadas en la siguiente tabla

ξ_1	10	15
ξ_2	1	2
ξ_3	25	35
ξ_4	75	85

Además se decide trabajar con un $1/2$ fracción de un diseño 2^4 factorial. Las observa-

ciones de las 8 unidades experimentales (u.e) aleatorizadas son:

TRATAMIENTO	Y
0000	62.0
1100	69.0
0011	57.0
1010	64.5
1001	61.8
0110	64.7
0101	62.2
1111	66.3

Empleando variables codificadas, se ajusta un modelo de la forma

$$\hat{\eta} = b_0 + \sum_{j=1}^k b_j x_j.$$

Los estimadores obtenidos por el método de mínimos cuadrados para el modelo anterior, son

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 63.44 \\ 1.9625 \\ 2.1125 \\ -0.3125 \\ -1.612 \end{bmatrix}.$$

Si se quiere escoger μ de modo que un cambio de una unidad en el primer factor, A , defina un incremento a lo largo del camino de ascenso, entonces en la variable codificada $x_1 = \xi_1 - 12.5/2.5$, un incremento de una unidad en ξ_1 genera una correspondiente variable x'_1 dada por

$$x'_1 = \frac{(\xi_1 + 1) - 12.5}{2.5} = \frac{\xi_1 - 12.5}{2.5} + \frac{1}{2.5} = x_1 + \Delta x_1,$$

donde $\Delta x_1 = 0.4$.

Es decir, si ξ_1 aumenta en uno, x_1 aumenta en 0.4 unidades. En general esto prueba que si ξ_1 se incrementa en Δ_1 , x_1 se incrementa en $\frac{\Delta_1}{2.5}$. Como b_1 es positivo, la primera coordenada a lo largo del camino es (0.4). Entonces para $\Delta_1 = 1$ sustituyendo Δx_1 por x_1 en $x_1 = \frac{b_1}{2\mu}$ se tiene que $\mu = \frac{1.9625}{(0.4)(2)} = 2.453$. El hecho de que $\mu > 0$ implica que se está determinando una ruta de ascenso máximo. En este caso $\mu > 0$, porque tanto Δ

como b_i son positivos. El que $\Delta_1 > 0$, no necesariamente implica que $\mu > 0$. Es decir Δ_1 positiva no determina que se avance en la ruta de ascenso. Conociendo μ es fácil encontrar los valores de x_2, x_3, x_4 de respuesta máxima. Simplemente se substituye y se despeja en $x_i = b_i/2\mu$ con $i = 2, 3, 4$.

Antes de presentar un camino de ascenso máximo, considere las siguientes observaciones:

El centro del experimento, en las variables decodificadas, es $\xi_1^o = 12.5, \xi_2^o = 1.5, \xi_3^o = 30, \xi_4^o = 80$, que es el punto de inicio de la ruta. Recuerde que son los valores $\xi^{i's}$ donde las correspondientes $x^{i's}$ se hacen cero.

Como $x_2 = \frac{\xi_2 - 1.5}{0.5}$ entonces

$$\begin{aligned}\xi_2 &= 0.5x_2 + 1.5 \\ &= 0.5x_2 + \xi_2^o \\ &= \Delta\xi_2 + \xi_2^o\end{aligned}$$

dado que $x_2 = \frac{b_2}{2\mu}$ se tiene que $\Delta\xi_2 = \frac{(0.5)b_2}{2\mu}$, lo cual significa que si x_2 cambia $\frac{b_2}{2\mu}$, entonces el correspondiente cambio en ξ_2 será $\frac{(0.5)b_2}{2\mu}$. Análogamente, para decodificar x_3 y x_4 hay que multiplicar por 5 (en este caso coincide en ambas variables), que es el correspondiente divisor en la codificación.

En general, ya se vió que un incremento Δ_1 en ξ_1 corresponderá a un incremento $\frac{\Delta_1}{2.5}$ en x_1 , que puede expresarse como $\Delta_1 C_1$. Dado que $\mu = \frac{1.9625}{2\Delta_1}(2.5)$ entonces

$$x_2 = \frac{b_2}{2\mu} = \frac{2\Delta_1 b_2}{2(1.9625)(2.5)},$$

que se puede reescribir como $x_2 = \Delta_1 C_2$. De manera análoga se puede decir que $x_3 = \Delta_1 C_3$ y $x_4 = \Delta_1 C_4$.

Si el incremento para ξ_1 fuese $k\Delta_1$ (esto es, se da el k -ésimo paso en la ruta) entonces x_1, x_2, x_3 y x_4 dependerán de k , de manera que $x_1 = k\Delta_1 C_1, x_2 = k\Delta_1 C_2, x_3 = k\Delta_1 C_3$, y $x_4 = k\Delta_1 C_4$.

Entonces $\Delta\xi_1 = k\Delta_1, \Delta\xi_2 = k\Delta_1 C_2(0.5), \Delta\xi_3 = k\Delta_1 C_3(5.0)$ y $\Delta\xi_4 = k\Delta_1 C_4(5.0)$, que pueden reescribirse como:

$$\Delta\xi_1 = (k-1)\Delta_1 + \Delta_1$$

$$\Delta\xi_2 = (k-1)\Delta_1C_2(0.5) + \Delta_1C_2(0.5)$$

$$\Delta\xi_3 = (k-1)\Delta_1C_3(5.0) + \Delta_1C_3(5.0)$$

$$\Delta\xi_4 = (k-1)\Delta_1C_4(5.0) + \Delta_1C_4(5.0)$$

Entonces el camino de ascenso máximo se puede dar de la siguiente manera:

- 1.- Se parte del centro del experimento $(\xi_1^0, \xi_2^0, \xi_3^0, \xi_4^0) = (12.5, 1.5, 30, 80,)$.
- 2.- Después se suma Δ_1 a ξ_1^0 , $\Delta_1C_2(0.5)$ a ξ_2^0 , $\Delta_1C_3(5.0)$ a ξ_3^0 y $\Delta_1C_4(5.0)$ a ξ_4^0 . Se obtiene un nuevo punto $(\xi_1', \xi_2', \xi_3', \xi_4')$, al cual se le agrega el vector $(\Delta_1, \Delta_1C_2(0.5), \Delta_1C_3(5.0), \Delta_1C_4(5.0))$ y así sucesivamente.
- 3.- Hay que evaluar la función objetivo en cada uno de los puntos que se vayan construyendo siempre y cuando estén fuera de la región experimental. Se procede de esta forma hasta que la función, después de haber mostrado un comportamiento creciente, muestre algún decremento.

La tabla siguiente muestra un camino de ascenso máximo para el modelo ajustado en el que fueron necesarios 12 puntos, para finalmente localizar una región donde posiblemente está el punto que corresponde al valor óptimo de la función.

corrida		ξ_1	ξ_2	ξ_3	ξ_4	Y
	base	12.5	1.5	30	80	
	Δ	1.0	$\frac{b_2}{2\mu}(0.5) = 0.215$	$\frac{b_3}{2\mu}(5.0) = -0.319$	$\frac{b_4}{2\mu}(5.0) = -1.643$	
	base+ Δ	13.5	1.715	29.681	78.357	
	base+2 Δ	14.5	1.930	29.362	76.714	
	base+3 Δ	15.5	2.145	29.043	75.071	
9	base+4 Δ	16.5	2.360	28.724	73.428	74.0
10	base+6 Δ	18.5	2.790	28.086	70.142	77.0
11	base+8 Δ	20.5	3.220	27.448	66.856	81.0
12	base+9 Δ	21.5	3.435	27.129	65.213	78.7

Las corridas 9,10,11 y 12 son coordenadas fuera de la región experimental original, donde observaciones adicionales sobre la respuesta fueron hechas.

Note que el incremento Δ es positivo para las variables 1 y 2, pero negativo para las variables 3 y 4, como se esperaba. Las corridas 9,10,11 y 12 indican un incremento

seguido de un decremento en la respuesta. Esto implica que ya no hay más ventajas en seguir iterando sobre el presente camino. Podría pensarse entonces, en un experimento factorial cuyo centro fuera el punto 11 de la tabla pues entre los elementos 11 y 12 se nota una curvatura de la función.

Esencialmente, se espera que el segundo experimento y el camino de pendiente máxima resultante, llevaría al experimentador a una vecindad del óptimo. El cambio en la respuesta en el ensayo 11 es de 63.44 (promedio de las observaciones experimentales iniciales) a 81 (valor obtenido en la corrida 11). Se espera que en un experimento posterior el incremento no fuese tan grande; si esto sucede se procede a ajustar un modelo de segundo orden y hacer el Análisis Canónico.

A continuación se da un ejemplo que ilustrará como es que el método de A.P.M. depende fuertemente de la escala de medición de las variables. La manera en que se desarrolla es delineando dos rutas diferentes, una obtenida tras haber trabajado con variables codificadas y la otra con variables decodificadas.

Ejemplo (2.3.2): Myers [1971]

Suponga que la respuesta depende de dos variables independientes ξ_1 y ξ_2 y que se han considerado los siguientes ensayos experimentales

corrida	ξ_1	ξ_2
1	10	60
2	10	80
3	20	60
4	20	80

Codificando las variables

$$x_1 = \frac{\xi_1 - 15}{5}, \quad x_2 = \frac{\xi_2 - 70}{10}.$$

El origen del diseño es $x_1 = 0$ y $x_2 = 0$, que en las variables decodificadas corresponde a $\xi_1 = 15$ y $\xi_2 = 70$.

Se supone que la ecuación ajustada está dada por

$$\hat{\eta} = 1 + 10x_1 + 2x_2.$$

y que se utilizó un diseño factorial 2^2 para dicho ajuste.

Si el primer incremento es calculado sobre la base de un cambio de 5 unidades en ξ_1 (que corresponde a una unidad en x_1), entonces la ruta de ascenso está dada en la

siguiente tabla:

	x_1	x_2	ξ_1	ξ_2
base	0	0	15	70
Δ	1	0.2	5	2
base+ Δ	1	0.2	20	72
base+2 Δ	2	0.4	25	74
base+3 Δ	3	0.6	30	76
base+4 Δ	4	0.8	35	78

Si x_1 se incrementa en 1 se tiene que $1 = \frac{10}{2\mu}$, lo cual implica que $\mu = 5$. Por tanto $x_2 = \frac{2}{2\mu} = \frac{1}{5} = 0.2$, y en ξ_2 corresponde a un incremento de $(0.2)(10) = 2$. Esto explica los valores Δ en la tabla.

Usando la ecuación ajustada en términos de las variables decodificadas, se tiene que

$$\begin{aligned}\hat{\eta} &= 1 + 10 \left[\frac{\xi_1 - 15}{5} \right] + 2 \left[\frac{\xi_2 - 70}{10} \right] \\ &= -43 + 2\xi_1 + 0.2\xi_2.\end{aligned}$$

Nuevamente el método consiste en encontrar las coordenadas ξ_1, ξ_2 que maximizan la respuesta sobre una esfera con centro en el origen y de radio R . Por lo tanto, se requiere maximizar la respuesta estimada, sujeta a la restricción

$$(\xi_1 - 15)^2 + (\xi_2 - 70)^2 = R^2.$$

El problema se resuelve por medio de los multiplicadores de lagrange. Obteniendo las derivadas parciales con respecto a ξ_1 y ξ_2 del lagrangeano con respecto a ξ_1 y ξ_2 resulta que

$$\xi_1 - 15 = \frac{2}{\mu} \quad \text{y} \quad \xi_2 - 70 = \frac{0.2}{2\mu}.$$

Se considera un incremento de 5 unidades en ξ_1 que da lugar a un incremento de 0.5 en ξ_2 después de despejar μ , esto es

$$20 - 15 = \frac{1}{\mu} \quad \text{que implica} \quad \mu = \frac{1}{5},$$

substituyendo para ξ_2 se tiene

$$\xi_2 = 70 + \frac{5(0.2)}{2} = 70.5,$$

que implica

$$\Delta\xi_2 = 70.5 - 70 = 0.5.$$

La ruta de ascenso por pendiente máxima esta dada en la tabla que sigue

	ξ_1	ξ_2
base	15	70
Δ	5	0.5
base+ Δ	20	70.5
base+2 Δ	25	71.0
base+3 Δ	30	71.5
base+4 Δ	35	72.0

En este momento es natural preguntar qué camino de A.P.M resulta más conveniente.

Alternativamente podría quedarse con la ruta que presente el primer decremento. Sin embargo no se debe olvidar que para la otra ruta, aunque se realice un mayor número de ensayos se puede obtener un valor más grande de Y .

También se debe pensar en qué tan cerca se está de obtener una región experimental buena, quizá cuando la cercanía sea obvia convenga tomar lo menos ensayos posibles, pero cuando se piense que se está lejos, quizá convenga tomar un mayor número de ensayos.

Volviendo al ejemplo, se puede ver que en una ruta el incremento Δ_2 del paso es mayor que en la otra. Tal vez la de incremento más grande obtenga mayores niveles de Y , pero debido a la longitud del paso, la ruta puede brincarse el óptimo. Podría ser conveniente quedarse con la ruta de incremento más pequeño que sería la de menor descenso.

Para ninguna ruta de ascenso máximo existe la seguridad de que se presente un decaimiento en Y , de tal manera que se continuaría experimentando indefinidamente. En estos casos lo que se sugiere es fijar un nivel de respuesta que se considere adecuado, es decir, detener la ruta hasta que Y supere cierto valor.

También como se indicó anteriormente, se espera que el método A.P.M se haga menos efectivo a medida que se siguen haciendo nuevos grupos de experimentos (iterando). Por lo cual es imperativo que el experimentador haga una pausa entre grupos de experimentos para asegurarse de que los modelos de primer orden que ajusta en cada paso siguen dominando. Esto es probar la carencia de ajuste y significancia del modelo. El

siguiente ejemplo enfatiza este hecho en cada grupo de experimentos, aunque no se expresan explícitamente las tablas del análisis de varianza, sino hasta el último paso, debido a que éstas se registran en el libro de donde se tomó el ejemplo. Básicamente se recalca de manera muy detallada el método de optimización aplicando A.P.M y partiendo de modelos de primer orden hasta llegar a un modelo final de segundo orden y aplicar el Análisis Canónico.

Ejemplo (2.3.3): Khuri y Cornell [1987]

Un ingeniero químico, observa que la temperatura, ξ_1 , y el período de tiempo, ξ_2 , de una cierta reacción, afectan al porcentaje de rendimiento de la misma. El ingeniero está interesado en determinar si es posible obtener un incremento en el porcentaje de rendimiento de la reacción. Por lo cual, decide aplicar el método A.P.M. para encontrar una región en donde posiblemente se de el mayor valor de porcentaje de rendimiento de la reacción y estimar el mismo. Así podría comparar el valor obtenido con sus registros anteriores y observar si realmente es posible obtener el aumento deseado.

Decide realizar una serie de experimentos variando la temperatura y el tiempo de la reacción, manteniendo todos los demás factores que intervienen fijos. El conjunto inicial de experimentos consisten de probar dos niveles iniciales de temperatura (70° y $90^\circ C$) y dos niveles de tiempo (30 y 90 segundos). Es claro entonces, que las variables explicativas son la temperatura y tiempo. La respuesta de interés, Y , es el porcentaje de rendimiento, el cual se mide en términos de la cantidad de materia residual liberada durante la reacción resultando en una medida de la pureza del producto final.

La combinación de temperatura-tiempo, debe ser tal, que permita obtener un porcentaje de pureza mayor al rango de valores obtenidos en reacciones anteriores, que varía entre 55% y 75%.

Para el conjunto de experimentos iniciales, el modelo de dos variables a ser ajustado es

$$Y = \alpha_0 + \alpha_1 \xi_1 + \alpha_2 \xi_2. \quad (2.3.4)$$

Cada una de las cuatro combinaciones (temperatura-tiempo), $70^\circ C - 30$ seg, $70^\circ C - 90$ seg, $90^\circ C - 30$ seg, $90^\circ C - 90$ seg, se repite 2 veces y se registra el porcentaje de rendimiento para cada uno de los ocho ensayos. Los valores de rendimiento medidos asociados con cada combinación de temperatura-tiempo se enlistan en la tabla siguiente con los valores de las variables codificadas x_1 y x_2 , las cuales estan definidas como

$$x_1 = \frac{\xi_1 - 80}{10}, \quad x_2 = \frac{\xi_2 - 60}{30}.$$

DISEÑO I				
VAR. ORIGINALES		VAR. CODIFICADAS		% DE RENDIMIENTO
Temp. (°C)	Tiempo (seg.)	x_1	x_2	Y
70	30	-1	-1	49.8, 48.1
90	30	+1	-1	57.3, 52.3
70	90	-1	+1	65.7, 69.4
90	90	+1	+1	73.1, 77.8

El modelo expresado en variables codificadas queda como

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \quad (2.3.5)$$

donde los parámetros β_0 , β_1 y β_2 son funciones de α_0 , α_1 y α_2 en la ecuación (2.3.4), esto último debido a la codificación. Así

$$\beta_0 = \alpha_0 + 80\alpha_1 + 60\alpha_2, \quad \beta_1 = 10\alpha_1, \quad \beta_2 = 30\alpha_2.$$

El término sobrante, ϵ , en la ecuación (2.3.5), representa un error aleatorio en los valores de rendimiento.

La figura (2.3.1), muestra la localización de los 4 puntos del diseño dibujados relativos a toda la región de posibles combinaciones de temperatura-tiempo. También se presentan las formas verdaderas de las curvas de nivel de la superficie de porcentaje de rendimiento. La forma de la superficie es conocida por el experimentador.

En forma matricial el modelo queda expresado como

$$Y = X\beta + \epsilon,$$

$$\begin{bmatrix} 49.8 \\ 48.1 \\ 57.3 \\ 52.3 \\ 65.7 \\ 69.4 \\ 73.1 \\ 77.8 \end{bmatrix} = \begin{bmatrix} +1 & -1 & -1 \\ +1 & -1 & -1 \\ +1 & +1 & -1 \\ +1 & +1 & -1 \\ +1 & -1 & +1 \\ +1 & -1 & +1 \\ +1 & +1 & +1 \\ +1 & +1 & +1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}.$$

Aplicando el método de mínimos cuadrados para estimar β_0 , β_1 y β_2 se tiene que

$$\hat{\beta} = \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 61.6875 \\ 3.4375 \\ 9.8125 \end{bmatrix}$$

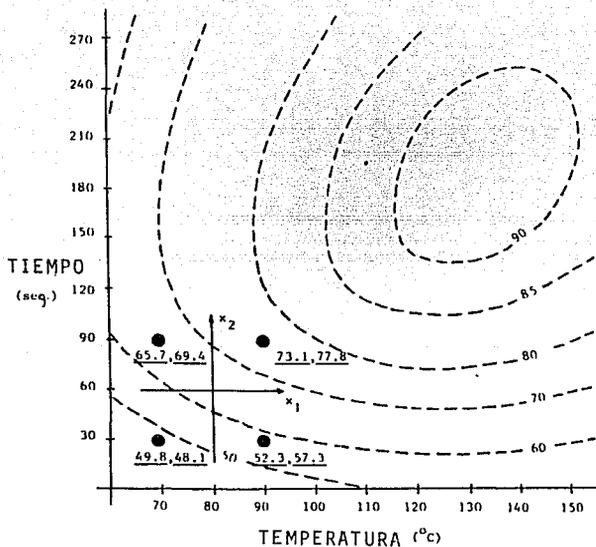


Figura 2.3.1: Los cuatro niveles iniciales de tiempo y temperatura y los porcentajes de rendimientos observados, sobre las curvas de nivel de la superficie de porcentajes de rendimientos hipotética

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA
BIBLIOTECA CENTRAL

Entonces

$$\hat{\eta} = 61.6875 + 3.4375x_1 + 9.8125x_2. \quad (2.3.6)$$

equivalentemente en las variables originales se tiene

$$\hat{\eta} = 14.5625 + 0.34375\xi_1 + 0.32708\xi_2. \quad (2.3.7)$$

Se verifica fácilmente utilizando los resultados de la sección (1.2) que el modelo no presenta carencia de ajuste y además es significativo, todo a un nivel de significancia de 0.5.

La superficie ajustada en las variables codificadas es

$$\hat{\eta} = 61.6875 + 3.4375x_1 + 9.8125x_2.$$

Dado que b_2 es positivo, significa que la superficie incrementa de valor si incrementa el valor de x_2 (o para valores de tiempo mayores de 60 seg.). Ahora, para encontrar el valor de μ que corresponde a un cambio arbitrario en el tiempo (ξ_2), de $\Delta_2 = 45$ seg. En la variable codificada x_2 , el cambio de 45 seg. corresponde a un cambio de $\frac{45}{30} = 1.5$ unidades, ya que

$$x_2' = \frac{(\xi_2 + \Delta_2) - 60}{30} = \frac{\xi_2 - 60}{30} + \frac{\Delta_2}{30} = x_2 + \Delta x_2.$$

De la ecuación (2.3.2) sustituyendo Δx_2 por x_2 en $x_2 = \frac{b_2}{2\mu}$, se tiene que

$$\mu = \frac{b_2}{2x_2} = \frac{9.8125}{2(1.5)} = 3.27.$$

El valor de x_1 , que corresponde al cambio de 1.5 unidades en x_2 es

$$x_1 = \frac{b_1}{2\mu} = \frac{3.4375}{2(3.27)} = 0.53,$$

y el cambio incremental en la temperatura es $\Delta_1 = 0.53(10) = 5.3^\circ C$. De aquí que el primer punto en el camino de ascenso por pendiente máxima, está localizado en las coordenadas $(x_1, x_2) = (0.53, 1.5)$, que corresponde a $(\xi_1, \xi_2) = (85.3, 105)$, un punto fuera de la región experimental original. Para comprobar si un incremento ocurre en la respuesta se utiliza la ecuación (2.3.6) y se compara la respuesta estimada en este último punto con la respuesta valuada en el centro del diseño, esto es, $\hat{\eta}(0.53, 1.5) - \hat{\eta}(0, 0) =$

78.2281 - 61.6875 = 16.54. Este hecho dice que se está determinando una ruta de ascenso máximo, lo cual también es plausible por el hecho de que $\mu > 0$, como se hizo notar en el ejemplo anterior.

Aunque la estimación $\hat{\eta}(0.53, 1.5)$ es una extrapolación fuera de la región, se cree que valores adicionales de altas respuestas serán observados si se corren experimentos extras en la dirección especificada por la ecuación (2.3.2). Usualmente se toman los incrementos de tal forma que el primer punto encontrado esté dentro de la región de diseño y se define un camino de ascenso por pendiente máxima, esto es que $\mu > 0$, luego se varían los incrementos hasta caer fuera de la región de diseño y allí se observan verdaderos valores de la respuesta hasta ver donde cambia la curvatura de la función de respuesta.

Experimentos adicionales se realizan a lo largo del camino A.P.M en puntos que corresponden a incrementos de distancias $1.5\Delta_i$, $2\Delta_i$, $3\Delta_i$ y $4\Delta_i$, ($i = 1, 2$). La siguiente tabla, lista las coordenadas de los puntos a lo largo del camino de ascenso hasta el punto que corresponde al incremento $4\Delta_i$, y los correspondientes valores de porcentajes de rendimiento observados.

	ξ_1 (Temp.)	ξ_2 (Tiempo)	Y
base	80.0	60.0	
Δ_i	5.3	45.0	
base+ Δ_i	85.3	105	74.3
base+ $1.5\Delta_i$	87.95	127.5	78.6
base+ $2\Delta_i$	90.6	150	83.2
base+ $3\Delta_i$	95.9	195	84.7
base+ $4\Delta_i$	101.2	240	80.1

Los valores de porcentaje de rendimiento incrementan hasta un valor de 84.7% que corresponde a las coordenadas (95.9°C, 195 seg.), respectivamente, y entonces el valor cae hasta 80.1% en (101.2°C, 240 seg.). En este momento se puede pensar que la temperatura, 101.2°C es muy alta o la longitud del tiempo, 240 seg. es muy grande, por consiguiente experimentos adicionales a lo largo del camino para valores grandes de ξ_1 y ξ_2 no son útiles. Observe que el cambio en la respuesta en este diseño fue de 61.6875 a 80.1, es decir, un cambio promedio de 18.41, el cual se espera disminuya en experimentos posteriores.

Observando un decremento en la respuesta en el punto $\xi_1 = 101.2^\circ C$ y $\xi_2 = 240$ seg., la decisión es hacer un segundo grupo de experimentos y nuevamente ajustar

un modelo de primer orden. El nuevo diseño estará anclado o tendrá como origen el punto $\xi_1 = 95.9^\circ C$ y $\xi_2 = 195$ seg., que es el último valor de incremento observado en la respuesta. Para este diseño, la amplitud del conjunto de niveles de temperatura y tiempo se mantienen igual que en el conjunto de experimentos iniciales, sólo que en este segundo grupo las variables codificadas se definen como

$$x_1 = \frac{\xi_1 - 95.9}{10}, \quad x_2 = \frac{\xi_2 - 195}{30}.$$

Cuando $x_1 = -1$ y $x_1 = +1$, los niveles de temperatura serían $85.9^\circ C$ y $105.9^\circ C$, y $x_2 = -1, +1$ corresponden a 165 seg. y 225 seg., en el tiempo respectivamente. Los cuatro nuevos puntos del diseño, el punto central y los valores del porcentaje de rendimiento en cada uno de los puntos se listan en la tabla siguiente. Dos repeticiones de los valores de rendimientos se observaron en cada una de las cuatro combinaciones del diseño factorial con una segunda observación en el punto central.

DISEÑO II				
x_1	x_2	ξ_1	ξ_2	% de REND.(Y)
-1	-1	85.9	165	82.9, 81.4
+1	-1	105.9	165	87.4, 89.5
-1	+1	85.9	225	74.6, 77.0
+1	+1	105.9	225	84.5, 83.1
0	0	95.9	195	84.7, 81.9

El modelo (2.3.5) se considera como adecuado para el ajuste. La superficie ajustada en las variables codificadas es

$$\hat{\eta} = 82.70 + 3.575x_1 - 2.75x_2,$$

y nuevamente se puede verificar utilizando los resultados de la sección (1.2), que el modelo no es carente de ajuste y es significativo.

Dado que b_1 es positivo, el valor del rendimiento se incrementa si incrementa el valor de x_1 (o para valores de más de $95.9^\circ C$). Para encontrar el valor de μ correspondiente a un cambio arbitrario en la temperatura (ξ_1) de $\Delta_1 = 10^\circ C$, es necesario observar el cambio en la variable codificada x_1 . El cambio de $10^\circ C$ corresponde a un cambio de una unidad ya que

$$x'_1 = \frac{(\xi_1 + \Delta_1) - 95.9}{10} = x_1 + \frac{\Delta_1}{10} = x_1 + \Delta x_1.$$

De la ecuación (2.3.2) sustituyendo Δx_1 por x_1 en $x_1 = \frac{b_1}{2\mu}$, se tiene que

$$\mu = \frac{b_1}{2x_1} = \frac{3.575}{2} = 1.7875.$$

El valor de x_2 , que corresponde al cambio de 1 unidad ($\frac{\Delta_1}{10}$) en x_1 es

$$x_2 = \frac{b_2}{2\mu} = \frac{-2.750}{2(1.7875)} = 0.7724$$

y el cambio incremental en la temperatura es $\Delta_2 = (0.77)30 = 23.1$. Así se tiene entonces el siguiente camino de A.P.M

	ξ_1 (Temp.)	ξ_2 (Tiempo)	Y
base	95.9	195	
Δ_i	10	-23.1	
base+ Δ_i	105.9	171.9	89.0
base+2 Δ_i	115.9	148.8	90.2(*)
base+3 Δ_i	125.9	125.7	87.4
base+4 Δ_i	135.9	102.6	82.6

Se hace notar que se observó el valor de respuesta 89.0 correspondiente al punto (105.9, 171.9) aunque no era necesario, debido a que está dentro de la región de diseño.

Se observa que la respuesta aumenta hasta el valor 90.2 en el punto (115.9,148.8) y luego decrece a 87.4 en el punto (125.9,125.7). De aquí que sea coherente hacer otro grupo de experimentos pero ahora con centro en (115.9,148.8). Observe que el cambio en la respuesta fue de 82.70 a 90.2, es decir, un cambio promedio de 7.5 unidades y como se esperaba menor que en el primer grupo de experimentos.

Habiendo observado entonces un decremento en la respuesta en el punto $\xi_1 = 125.9$ y $\xi_2 = 125.7$, se decide hacer un tercer grupo de experimentos y nuevamente ajustar un modelo de primer orden. Pero en este punto se decide escoger una fracción de $\frac{3}{4}$ de un diseño factorial 2^2 como región de diseño, para reducir la cantidad de cálculos requeridos por un factorial 2^2 completo; la construcción de cualquier fracción de un factorial, y en particular de la fracción $\frac{3}{4}$ de un 2^m , viene realizada en [14]. Esta elección representa una desviación del método convencional de A.P.M, pero, se reduce la cantidad de esfuerzo. Así la fracción $\frac{3}{4}$ del diseño factorial 2^2 constará de los puntos (105.9,171.9);

(125.9,125.7); (125.9,171.9) con centro en (115.9,148.8). Además se observó 2 veces el valor de la respuesta en cada uno de los puntos anteriores y luego se ajustó el modelo de primer orden, el cual se omite en este punto. Con este modelo se encontró un camino de A.P.M cuyo resultado se resume en la siguiente tabla

	ξ_1 (Temp.)	ξ_2 (Tiempo)	Y
base	115.9	148.8	91.0(*)
Δ_i	10	23.1	
base+ Δ_i	125.9	171.9	93.0
base+2 Δ_i	135.9	195	96.2←
base+3 Δ_i	145.9	218.1	92.9

Nótese que el experimentador decidió observar nuevamente el rendimiento en el punto (115.9,148.8) correspondiente al centro del diseño 3 y al paso 2 en la ruta del diseño 2. Este hecho debido a que le parecía extraño el alto rendimiento obtenido en el punto (115.9,148.8) en el paso 2 en la ruta del diseño 2. Este dato lo pudo corroborar y se refleja en los asteriscos de las dos últimas tablas. Dado que hay que fijarse sólo en los valores de respuesta que estan fuera de la región de diseño; se observa un cambio de 96.2 en la respuesta del punto (135.9,195) a 92.9 en el punto (145.9,218.1). Por lo cual parece razonable realizar otro experimento, pero ahora con centro en (135.9,195). Observe que el cambio promedio en la respuesta en este punto es de 91.0 a 96.2 que corresponde a 5.2 unidades. Recuerde que en este procedimiento iterativo se espera que el cambio promedio en la respuesta de un grupo de experimentos a otros sea cada vez menor, de tal forma que se llegue a un punto donde sea obvio que los modelos de primer orden dejen de dominar.

Ahora bien, el centro del nuevo experimento es (135.9,195) y se decide hacer un diseño factorial 2^2 con dobles repeticiones en cada punto, en vías de estimar el error puro y poder hacer la prueba de carencia de ajuste. Las variables codificadas para el nuevo centro son

$$x_1 = \frac{\xi_1 - 135.9}{10}, \quad x_2 = \frac{\xi_2 - 195}{23.1},$$

así para $x_1 = -1, +1$, los valores de temperatura son 125.9 y 145.9°C, y para $x_2 = -1, +1$ son 171.9 y 218.1 seg. respectivamente. El modelo ajustado es

$$\hat{\eta} = 93.983 + 0.025x_1 - 0.375x_2$$

la tabla de análisis de varianza para este último modelo es

FUENTE	g. lib.	SUM. DE CUADRADOS	CUAD. MEDIO	F
REGRESION	2	0.5650	0.2825	0.04
RESIDUAL	3	22.1833	7.3944	
TOTAL	5	22.7483		

Es claro por el valor de F, que el modelo ajustado no explica una cantidad significativa de la variación total del porcentaje de rendimiento. De hecho, una gráfica de los residuales ($Y_u - \hat{\eta}_u$) en los puntos del diseño, que se muestra en la figura (2.3.2), indica que el modelo ajustado subestima el porcentaje de rendimiento en el centro, pero sobreestima el porcentaje de rendimiento en los cuatro puntos del diseño factorial.

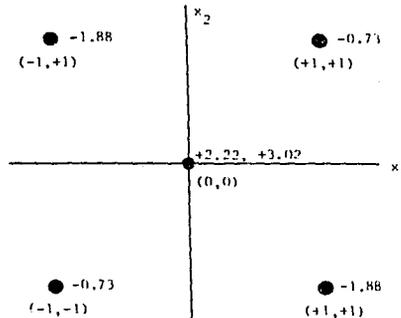


Figura 2.3.2: Residuales, $Y_u - \hat{Y}_u$, en los puntos (x_1, x_2) del diseño 3.

Este es un indicativo de que quizás no es adecuado ajustar un plano, en una región donde la superficie es más bien curva, en particular una montaña. La figura (2.3.3) muestra un resumen gráfico, de los pasos realizados en el A.P.M.

Aquí es entonces el momento de completar el diseño y estimar una función cuadrática para aplicar el Análisis Canónico. Así pues, utilizando los puntos obtenidos en el diseño 3 y agregando 4 puntos más (obtenidos de 4 experimentos adicionales) para crear un

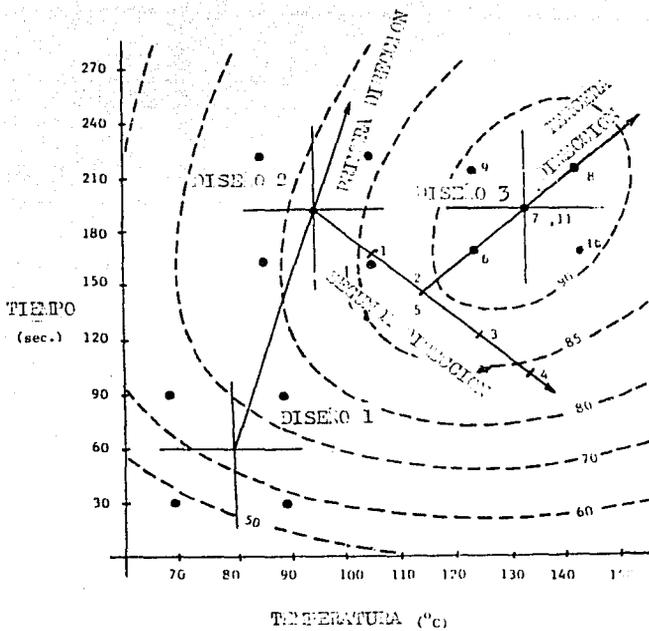


Figura 2.3.3: Sucesión de experimentos realizados (pasos 1-4) a lo largo de la segunda ruta de ascenso máximo. Puntos 5-8 representan experimentos realizados en una tercera dirección. Puntos 6-11 comprenden el tercer diseño

diseño central compuesto rotable, se tiene

VAR. CODIFICADAS		% de RENDIMIENTO	VAR. ORIGINALES	
x_1	x_2	Y	Temp.(°C)	Tiempo (seg.)
-1	-1	93.6	125.9	171.9
-1	+1	91.7	125.9	218.1
+1	-1	92.5	145.9	171.9
+1	+1	92.9	145.9	218.1
0	0	96.2, 97.0	135.9	195.0
$-\sqrt{2}$	0	92.7	121.75	195.0
$\sqrt{2}$	0	92.8	150.04	195.0
0	$-\sqrt{2}$	93.4	135.9	162.3
0	$\sqrt{2}$	92.7	135.9	227.7

Los nuevos puntos $(x_1, x_2) = (\pm\sqrt{2}, 0)$ y $(x_1, x_2) = (0, \pm\sqrt{2})$, junto con los cuatro puntos del diseño factorial $(x_1, x_2) = (\pm 1, \pm 1)$ y el punto central, forman el diseño compuesto central rotable. Obsérvese que se hicieron dos repeticiones en el origen.

Utilizando estos datos, el modelo ajustado en las variables codificadas es

$$\hat{\eta} = 96.60 + 0.0302x_1 - 0.3112x_2 - 1.9813x_1^2 - 1.8313x_2^2 + 0.5750x_1x_2.$$

Este modelo expresa al porcentaje de rendimiento como una función de las variables codificadas

$$x_1 = \frac{\text{temp.} - 135.9}{10} \quad x_2 = \frac{\text{tiempo} - 195}{23.1}.$$

en forma matricial $\hat{\eta}$ de acuerdo a (2.1.3) puede escribirse como

$$\begin{aligned} \hat{\eta} &= 96.60 + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^t \begin{bmatrix} 0.0302 & \\ & -0.3112 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^t \begin{bmatrix} -1.9813 & 0.2875 \\ 0.2875 & -1.8313 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= b_0 + \mathbf{x}^t + \mathbf{x}^t B \mathbf{x}. \end{aligned}$$

Se verifica, utilizando los resultados de la sección (1.2), que no hay carencia de ajuste y que el modelo es significativo. Pasando entonces a obtener el punto estacionario \mathbf{x}_0 que está dado por la ecuación (2.1.6), esto es,

$$\begin{aligned} \mathbf{x}_0 &= -\frac{B^{-1}\mathbf{b}}{2} = -\frac{1}{2} \begin{bmatrix} -1.51650 & -0.08109 \\ -0.08109 & -0.5588 \end{bmatrix} \begin{bmatrix} 0.03018 \\ -0.31125 \end{bmatrix} \\ &= \begin{bmatrix} -0.00486 \\ -0.08568 \end{bmatrix}. \end{aligned}$$

y la respuesta estimada en el punto x_0 , es igual a $\hat{\eta}_0 = 96.61$. Ahora la transformación Canónica expresa a $\hat{\eta}$ como

$$\hat{\eta} = \hat{\eta}_0 + \sum_{i=1}^k \delta_i w_i^2.$$

Para obtener esta ecuación canónica, se encuentran los valores propios (δ_1, δ_2) de la matriz B , resolviendo el determinante de la ecuación $|B - \delta I| = 0$, esto es,

$$\begin{vmatrix} (-1.9813 - \delta) & 0.2875 \\ 0.2875 & (-1.8313 - \delta) \end{vmatrix} = 0$$

y desarrollando se tiene

$$\delta^2 + 3.8125\delta + 3.5455 = 0.$$

Las raíces de esta ecuación son $\delta_1 = -1.6091$ y $\delta_2 = -2.2034$. De aquí que la ecuación canónica es

$$\hat{\eta} = 96.6133 - 1.6091w_1^2 - 2.2034w_2^2.$$

Dado que ambas δ_1 y δ_2 son negativas, el punto estacionario es máximo y la superficie se maximiza allí. La superficie estimada cae más rápido si se mueve en dirección del eje w_2 . Obviamente es posible encontrar los valores de la transformación en términos de las variables originales, con el método descrito en la sección 2.2.

Dado que el valor de la respuesta que se obtuvo mediante el análisis es 96.61%, el experimentador verifica la posibilidad de obtener mayores porcentajes de rendimientos con tan sólo variar la temperatura y el tiempo de la reacción a los niveles definidos por el punto estacionario.

Aunque el método de A.P.M tiene muchas desventajas, es probablemente el más utilizado para la búsqueda de puntos óptimos o para la búsqueda de regiones óptimas. Un ejemplo extremadamente interesante e informativo del uso de este método se da en [3]

Existen otros procedimientos de búsqueda de óptimos los cuales son usados en los trabajos de Metodología de Superficie de Respuesta. Uno de los más populares entre estos es el Método de tangentes paralelas. Una buena exposición sobre este procedimiento es dada en [9, 24].

Para finalizar esta sección, a manera de resumen, se presentan los pasos para aplicar el método de A.P.M:

1. Estimar con base en los datos de un diseño la tendencia local de $\eta(\mathbf{x})$ mediante la función lineal $\hat{\eta}(\mathbf{x})$.
2. Hacer las pruebas correspondientes para determinar si el ajuste lineal es satisfactorio (es decir, si se está lejos del óptimo); si las pruebas son positivas, se pasa al punto 3; en caso contrario, se pasa al punto 5.
3. Encontrar el vector de dirección de máxima pendiente en $\hat{\eta}(\mathbf{x})$; para ello se maximiza la función $\eta(\mathbf{x})$ sujeta a la condición de que $\|\mathbf{x}\| = 1$. Al vector encontrado se le denota por Δ . (obsérvece que la condición de que la norma del vector \mathbf{x} sea igual a la unidad, es simplemente por efectos algorítmicos y puede sustituirse en general, igual a R).
4. Tomar en $t\Delta$, $2t\Delta$, $3t\Delta$, \dots , $it\Delta$, \dots nuevas observaciones de tal manera que si $Y_i < Y_{i+1}$, se toma la observación Y_{i+2} , en caso contrario se detiene el proceso porque se considera que en el punto $it\Delta$ la función de respuesta $\eta(\mathbf{x})$ en dirección de Δ cambia de ser una función creciente a ser decreciente, entonces con centro en el punto $it\Delta$, que se refiere como c ($c = it\Delta$), se elige un nuevo diseño. Con los datos de esta nueva muestra se reinicia el proceso desde el paso 1. Se hace la observación de que Box y Wilson no dan un criterio específico para elegir el escalar t . En la sección (2.5) se proponen dos criterios para elegirlo.
5. Completar el diseño y estimar una función cuadrática $\eta(\mathbf{x})$. El punto donde esta función alcanza el óptimo se denota por $\hat{\mathbf{x}}_{op}$ y es el estimador del óptimo real.

Después de Box y Wilson (1951), no hubo cambios significativos en los procesos de estimación del óptimo hasta que Myers y Khuri (1979) propusieron una modificación en el paso número 4, en el que introducen un nuevo criterio para determinar el punto c donde $\eta(\mathbf{x})$ cambia de ascendente a descendente, sobre la trayectoria de ascenso. Esta modificación se describe a continuación.

2.4 MODIFICACION MYERS-KHURI (1979) PARA EL A.P.M

Como se ha hecho notar en la sección anterior, el método de ascenso por pendiente máxima es un procedimiento secuencial que sirve para determinar una posible mejor región de experimentación.

Con esto se quiere decir que la zona de operación que el procedimiento arroje, debe proporcionar niveles altos en la respuesta y caracterizarse por contener al punto máximo de Y , si éste existe.

Es claro entonces que se pueden determinar familias de regiones con esta cualidad; por ejemplo se puede pensar en la familia de esferas con radio R y con centro en el punto que corresponde al óptimo. Es por ello que se habla de una mejor región y no de la mejor región.

También es posible que la zona experimental obtenida con ascenso por pendiente máxima, no contenga al óptimo de Y aún cuando éste exista. Esto entre otras cosas, puede deberse a la regla de desición que hasta el momento se ha adoptado para detener el procedimiento secuencial, ya que

1. Si $Y_i > Y_{i-1}$, se obtiene el punto $i + 1$ de la ruta de A.P.M, pero
2. Si $Y_i < Y_{i-1}$, se determina una nueva región de experimentación que contenga (entre otros) a los puntos $(i - 1, i)$.

Aunque parece una regla *lógica* para detener el procedimiento, tiene, entre sus defectos, el no considerar las variaciones en Y debidas al error experimental. De esta manera podría suceder que $Y_i < Y_{i-1}$ sin que en realidad se hubiese pasado cerca de un máximo. Por lo cual en esta sección se esbozará la modificación de Myers y Khuri, que considera las fluctuaciones en la respuesta Y debidas al error experimental, dando así una mejor forma de realizar el proceso de A.P.M.

Sea $Y = \eta(x_1, x_2, \dots, x_k) + \epsilon$, la respuesta que se desea optimizar, donde

$$\eta(x_1, x_2, \dots, x_k) + \epsilon$$

es una función continua y el error ϵ cumple con los supuestos: $\epsilon_i \sim N(0, \sigma^2)$, $COV(\epsilon_i, \epsilon_j) = 0$, para toda $i \neq j$. Con esto $Y \sim N(\eta, \sigma^2)$.

Suponga nuevamente que las variables x_i han sido codificadas y por otro lado que η puede aproximarse localmente por una función de la forma $\beta_0 + \sum_{i=1}^k \beta_i x_i$. De esta manera $\hat{\eta}$ está dada por $b_0 + \sum_{i=1}^k b_i x_i$, donde b_i es el estimador de mínimos cuadrados de β_i .

La Ruta de A.P.M está dada, para diferentes valores de R , por aquéllos puntos que maximizan $\hat{\eta}$ sujeta a $\sum_{i=1}^k x_i^2 = R^2$.

Ahora bien, se sabe que para un radio dado, el óptimo, llamado \mathbf{x}^* , es:

$$\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_k^*) = \left(\frac{b_1}{2\mu}, \frac{b_2}{2\mu}, \dots, \frac{b_k}{2\mu} \right),$$

donde μ es el multiplicador de lagrange que debe ser igual a $\frac{\sqrt{\sum b_i^2}}{2R}$, para que \mathbf{x}^* sea máximo (el negativo de la expresión, para que \mathbf{x}^* sea mínimo).

Por lo tanto

$$\mathbf{x}^* = \left(\frac{b_1 R}{\sqrt{\sum b_i^2}}, \frac{b_2 R}{\sqrt{\sum b_i^2}}, \dots, \frac{b_k R}{\sqrt{\sum b_i^2}} \right) = (\lambda_1 t, \lambda_2 t, \dots, \lambda_k t),$$

con $\lambda_i = \frac{b_i}{\sqrt{\sum b_i^2}}$ y $t = \sqrt{\sum x_i^2} = R$. Es decir la i -ésima componente de \mathbf{x}^* , puede ser escrita como $x_i^* = \lambda_i t$, $i = \overline{1, k}$, con λ el vector normalizado de \mathbf{b} y $t = R$ es la norma del vector \mathbf{x} .

Las expresiones $x_i^* = \lambda_i t$, $i = \overline{1, k}$, se conocen como ecuaciones paramétricas.

Si t_i es la norma del i -ésimo punto en la ruta, se denotará por $Y(t_i)$ a la respuesta obtenida con radio t_i , ($t_0 = 0$). Recuerde que la ruta se puede obtener dando diferentes valores a t_i o bien dando valores a μ y obteniendo los respectivos valores de t_i .

Sea t_{n_0} el primer valor para el cual $Y(t_{n_0}) > Y(t_{n_0+1})$. Para determinar si ese decremento en la respuesta es real o es causado por el error experimental (las fluctuaciones en la varianza) se propone realizar la prueba de hipótesis.

$H_0^{(1)}$: $\eta(x)$ sigue creciendo en dirección de la trayectoria de ascenso.

-vs-

$H_1^{(1)}$: $\eta(x)$ ya alcanzó el máximo en dirección de la trayectoria de ascenso.

Hipótesis que pueden ser escritas como

$$H_0^{(1)}: \eta(t) \geq m_0^{(1)} \text{ -vs- } H_1^{(1)}: \eta(t) \leq m_1^{(1)}$$

donde

$$\begin{aligned} t &> t_{n_0} \\ m_0^{(1)} &= \eta(t_{n_0}) \\ m_1^{(1)} &= m_0^{(1)} + \Delta \quad \text{con } \Delta < 0, \end{aligned}$$

De aquí que, $H_0^{(1)}$ y $H_1^{(1)}$ pueden ser reescritas nuevamente como

$$H_0^{(1)}: \eta(t) \geq \eta(t_{n_0}) \quad \text{con } t > t_{n_0}.$$

-vs-

$$H_1^{(1)}: \eta(t_{n_0}) \leq \eta(t_{n_0}) - \delta \quad \text{con } \delta = -\Delta > 0.$$

definiendo $S_i^{(1)} = Y(t_{n_0+i}) - Y(t_{n_0})$, $i = 1, 2, \dots$ y luego de un largo proceso algebraico y suponiendo condiciones específicas, descritas en [34], Myers y Khuri encuentran las regiones de rechazo y aceptación para la prueba. La regla de decisión obtenida es

$$\text{NO RECHAZAR } H_0^{(1)} \quad \text{si } S_i^{(1)} \geq c \quad (c > 0) \quad (2.4.1)$$

$$\text{RECHAZAR } H_0^{(1)} \quad \text{si } S_i^{(1)} \leq a \quad (a < 0) \quad (2.4.2)$$

$$\text{OBSERVAR } S_{i+1}^{(1)} \quad \text{si } a < S_i^{(1)} < c \quad (2.4.3)$$

En resumen, se ha propuesto una prueba de hipótesis secuencial donde la zona de preferencia por aceptación está determinada por (2.4.1), la zona de preferencia por rechazo por (2.4.2) y la zona de indiferencia por (2.4.3).

Si $H_0^{(1)}$ se acepta, entonces quiere decir que el decremento en la función Y se debe al error experimental, en este caso se debe seguir avanzando en la ruta A.P.M. Pero cuando $H_0^{(1)}$ se rechaza, la ruta debe abandonarse para empezar una nueva.

Cuando se cae en el caso (2.4.3) se debe seguir muestreando hasta que se de el caso (2.4.1) o bien el (2.4.2).

Si $H_0^{(1)}$ no se rechaza, después de seguir avanzando por la ruta A.P.M. se puede encontrar un punto t_{n_1} , con el cual suceda que $Y(t_{n_1+1}) < Y(t_{n_1})$, nuevamente se tiene que contrastar la hipótesis

$$H_0^{(2)}: \eta(t) \geq m_0^{(2)} \quad \text{-vs-} \quad H_1^{(2)}: \eta(t) \leq m_1^{(2)},$$

con $t > t_{n_1}$, $m_0^{(2)} = \eta(t_{n_1})$, $m_1^{(2)} = m_0^{(2)} + \Delta$ ($\Delta < 0$). La regla de decisión será en esta ocasión:

$$\text{NO RECHAZAR } H_0^{(2)} \quad \text{si } S_i^{(2)} \geq c \quad (c > 0)$$

$$\text{RECHAZAR } H_0^{(2)} \quad \text{si } S_i^{(2)} \leq a \quad (a < 0)$$

$$\text{OBSERVAR } S_{i+1}^{(2)} \quad \text{si } a < S_i^{(2)} < c$$

que es análoga a la anterior.

La problemática es entonces, cómo determinar los valores de a y c . Myers y Khuri (1979) demuestran que a y c son tales que $a = -c = a_0$ y a_0 cumple que

$$\Phi\left(\frac{a_0}{\sqrt{2}\hat{\sigma}}\right) = \frac{1}{2k'},$$

donde $\hat{\sigma}$ es la varianza estimada de σ en el modelo, Φ es la función de distribución de una densidad Normal(0, 1). La constante k' se determina pensando cuántas observaciones serían necesarias para que se incremente lo suficiente la respuesta si es creciente, esto en vías de reducir el error tipo I (rechazar H_0 , cuando H_0 es cierta). Se fija con base a la experiencia que del problema posea el experimentador.

El procedimiento para probar $H_0^{(1)}$ queda como:

$$\text{NO RECHAZAR } H_0^{(1)} \quad \text{si } Y(t_{n_0+i}) \geq Y(t_{n_0}) + a_0$$

$$\text{RECHAZAR } H_0^{(1)} \quad \text{si } Y(t_{n_0+i}) \leq Y(t_{n_0}) - a_0$$

$$\text{OBSERVAR } Y(t_{n_0+i+1}) \quad \text{si } Y(t_{n_0}) - a_0 \leq Y(t_{n_0+i}) \leq Y(t_{n_0}) + a_0.$$

Con este nuevo procedimiento, el experimento no se detiene debido a fluctuaciones falsas de Y . La efectividad del método será notablemente mayor comparada con la del procedimiento con detención como se verá en el siguiente ejemplo.

Ejemplo (2.4.1): Continuación del Ejemplo (2.3.1)

Lo primero que se necesita para aplicar el procedimiento de A.P.M. con la modificación Myers-Khuri (1979) es calcular el valor de a_0 . Esto se hará con un valor de $k' = 25$, considerado como coherente por el experimentador.

Con los datos del ejemplo (2.3.1) se tiene que

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\eta}_i)^2}{n-1} = \frac{1.4838}{7} = .211974,$$

y $\hat{\sigma} = .4604035$. Ahora

$$\Phi \left(\frac{a_0}{2^{1/2}(.4604035)} \right) = \frac{1}{50} = 0.02.$$

Entonces el cuantil .02 de una distribución Normal con media cero y varianza uno es -2.06. Así que

$$\frac{a_0}{2^{1/2}(.4604035)} = -2.06,$$

por tanto $a_0 = -1.3412843$. La prueba se expresa como:

NO RECHAZAR $H_0^{(1)}$ si $Y(t_{n_0+i}) \geq Y(t_{n_0}) + 1.3412843$

RECHAZAR $H_0^{(1)}$ si $Y(t_{n_0+i}) \leq Y(t_{n_0}) - 1.3412843$

OBSERVAR $Y(t_{n_0+i+1})$ si $Y(t_{n_0}) - 1.3412843 \leq Y(t_{n_0+i}) \leq Y(t_{n_0}) + 1.3412843$.

Considérese que se continuó observando Y sobre más puntos en la ruta de A.P.M. para el ejemplo (2.3.1) y que de la corrida 11 a 21 se obtuvieron los valores de respuesta dados en la tabla siguiente

corrida	Y
11	81.0
12	79.7
13	80.1
14	82.4
15	83.3
16	85.1
17	86.3
18	86.9
19	87.3
20	87.5
21	85.3

En el ejemplo (2.3.1) se observó un decaimiento en la respuesta en el punto 11, por lo que $Y(t_{n_0}) = 81.0$ y $Y(t_{n_0+1}) = 79.7$. Aplicando el método modificado

$$Y(t_{n_0}) - 1.3412843 = 79.658716$$

$$Y(t_{n_0}) + 1.3412843 = 82.341284$$

y $79.6587 < 79.7 < 82.3412$. Por tanto, se observa $Y(t_{n_0+2}) = 80.1$.

Como $79.6587 < 80.1 < 82.3412$, se observa $Y(t_{n_0+3}) = 82.4$ que lleva a no rechazar $H_0^{(1)}$ (porque $82.4 > 82.341284$) y a seguir observando valores en la ruta de A.P.M hasta que se de otra caída en el valor de la respuesta. Se ve que hasta la corrida 20 se presenta esta situación, de donde $Y(t_{n_1}) = 87.5$, $Y(t_{n_1+1}) = 85.3$.

Ahora se realiza la prueba

$$\text{NO RECHAZAR } H_0^{(2)} \quad \text{si } Y(t_{n_1+i}) \geq Y(t_{n_1}) + 1.3412843$$

$$\text{RECHAZAR } H_0^{(2)} \quad \text{si } Y(t_{n_1+i}) \leq Y(t_{n_1}) - 1.3412843$$

$$\text{OBSERVAR } Y(t_{n_1+i+1}) \quad \text{si } Y(t_{n_1}) - 1.3412843 \leq Y(t_{n_1+i}) \leq Y(t_{n_1}) + 1.3412843.$$

donde

$$Y(t_{n_1}) + 1.341284 = 88.841284$$

$$Y(t_{n_1}) - 1.341284 = 86.158716$$

Como $85.3 < 86.158716$ se Rechaza $H_0^{(2)}$ y se debe pensar en construir un nuevo experimento tomando como origen el ensayo 20 y obviamente ajustar un modelo de primer orden, validar sus supuestos y encontrar otro camino ascendente aplicando la prueba de hipótesis anterior.

Se observa una ganancia en la respuesta de 4.3 unidades del método modificado con respecto al método original.

2.5 ALGUNAS VARIANTES AL METODO A.P.M.

Uno de los aspectos que está sujeto a la subjetividad del investigador, es la distancia que existe entre dos puntos diferentes del diseño. Si estos se encuentran muy alejados, se corre el riesgo de no detectar cambios importantes en la función de respuesta η ;

si por el contrario, los puntos se encuentran muy cerca, se corre el riesgo de que su tendencia quede completamente confundida con la variación aleatoria del proceso. Este hecho despierta la inquietud de explorar nuevos criterios para determinar la longitud del paso adecuada.

Recuerde que la longitud del *paso* es una constante que determina el espaciamiento entre dos observaciones subsecuentes en dirección de la pendiente máxima de $\hat{\eta}$, luego que las pruebas realizadas dan evidencia suficiente de que el óptimo aún se encuentra *lejos* y que, un plano se ajusta bien a los datos.

En este apartado se presenta una propuesta hecha por [39, 40] para escoger la longitud del *paso* con la cual se tenga poca probabilidad de confundir la tendencia de η con el error de las observaciones.

2.5.1 Obtención de la Longitud del Paso t .

Suponga que la tendencia local de η , esta suficientemente estimada en una región circundante al diseño mediante un polinomio de primer grado; suponga también, que el vector Δ indica la dirección de máxima pendiente de $\hat{\eta}$; entonces *cerca* del diseño se cumple la relación:

$$\eta(it\Delta) < \eta((i+1)t\Delta).$$

Dado que las observaciones tienen un error aleatorio, que Y_i resulte ser mayor que Y_{i+1} , no necesariamente implica que la tendencia de η ya cambió, de ascendente a descendente; este hecho como ya se hizo notar, podría deberse únicamente a la aleatoriedad del proceso. Es por ésto, que tiene sentido pedir que la longitud del paso sea tal que permita detectar la tendencia de η con poca influencia del error de observación.

Se desea encontrar la magnitud del paso tal que si la tendencia de η en dirección de Δ es creciente, la probabilidad de que Y_i resulte ser mayor o igual a Y_{i+1} no sea mayor o igual que un valor prefijado.

Como

$$P(Y_i \geq Y_{i+1}) = P[\eta(it\Delta) + \epsilon_i \geq \eta((i+1)t\Delta) + \epsilon_i] \quad (2.5.1)$$

$$= P[\eta((i+1)t\Delta) - \eta(it\Delta) \leq \epsilon_i - \epsilon_{i+1}] \quad (2.5.2)$$

con $\epsilon_i \sim N(0, \sigma^2)$ independientes, se sigue que: $\epsilon_i - \epsilon_{i+1} \sim N(0, 2\sigma^2)$; y

$$P(Y_i \geq Y_{i+1}) = \alpha \quad (2.5.3)$$

cuando

$$\eta((i+1)t\Delta) - \eta(it\Delta) = z_\alpha \sigma \sqrt{2}. \quad (2.5.4)$$

Por ejemplo: si la probabilidad de detectar equivocadamente un cambio de ascenso a descenso, es igual a .025, la constante z_α en (2.5.4) es 1.96; si dicha probabilidad es igual a .0013, entonces z_α es 3.

Dado que tanto σ como η son desconocidos, se puede utilizar el hecho de que $SCE/\sigma^2 = (n-2)\hat{\sigma}^2/\sigma^2 \sim \chi^2$ con $n-2$ grados de libertad y que por lo tanto $(\epsilon_i - \epsilon_{i+1})/\sqrt{2}\hat{\sigma} \sim T$ - student con $n-2$ grados de libertad; la demostración de estos dos resultados puede verse en ([31], pág. 135; [32], cap. 6). De esto se sigue que (2.5.4) se sustituye por

$$\eta((i+1)t\Delta) - \eta(it\Delta) = T_\alpha \hat{\sigma} \sqrt{2}. \quad (2.5.5)$$

con esta ecuación se puede encontrar un estimador de la longitud del paso t , al sustituir η por $\hat{\eta}$ y despejar a \hat{t} de la ecuación resultante, con lo que se llega a

$$\hat{t} = \frac{T_\alpha \hat{\sigma} \sqrt{2}}{b_1 x_1 + \dots + b_m x_m} \quad (2.5.6)$$

donde \mathbf{b} es el vector estimado por mínimos cuadrados de β .

Otra variante para elegir el escalar t , es considerar los resultados debidos a O'Reilly. Este autor considera que: si $Y_i \sim N(x_i^t; \beta, \sigma^2)$ con x_i^t el i -ésimo renglón de una matriz X de rango completo, entonces se dice que una v.a. Y tiene una distribución estimable si existe una función $h(Y_1, Y_2, \dots, Y_n)$ con la misma distribución de Y .

La extrapolación en un punto \mathbf{x}_0 es válida si $Y(\mathbf{x}_0) \sim N(\mathbf{x}_0^t; \beta, \sigma^2)$ tiene una distribución estimable.

O'Reilly [38], muestra que la región donde es válido extrapolar con un modelo de regresión lineal en r factores, esto es

$$Y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi} + \beta_{11} x_{1i}^2 + \dots + \beta_{mm} x_{mi}^r + \epsilon_i,$$

está dada por:

$$S = \{x \in \mathbb{R}^m \mid V(\hat{\eta}) \leq \sigma^2\} = \{x \in \mathbb{R}^m \mid w^t (w^t w)^{-1} w \leq 1\},$$

con $w^t = (1, x_1, \dots, x_m^t)$, que es el conjunto en donde la distribución de $\eta(x)$ es estimable insesgadamente.

Entonces se puede suponer que entre dos elementos de S no existen fluctuaciones importantes que pasen inadvertidas al modelo y de esta manera si t es solución de la ecuación

$$(1, t\Delta^t)(w^t w)^{-1} \begin{pmatrix} 1 \\ t\Delta \end{pmatrix} = 1, \quad (2.5.7)$$

se minimiza el riesgo de confundir la tendencia de η con el error aleatorio manteniéndose dentro de la región de extrapolación.

Se considera que un punto x esta cerca del diseño, si $x \in S$.

2.5.2 Localización del Centro del Nuevo Diseño, c.

Para estimar el punto donde η alcanza el óptimo en dirección de Δ , se propone utilizar un criterio semejante al que se usa para encontrar el máximo de cada factor en el método del factor único. De esta manera se utiliza toda la información disponible en dirección de Δ , y no únicamente la información de las dos últimas observaciones, como ocurre con el método original de A.P.M. donde se considera que hubo un cambio en la tendencia de η si la última observación es menor o igual a la penúltima.

A diferencia del método del factor único donde se maximiza cada factor en el orden indicado por el investigador ; aquí se busca el máximo en la dirección óptima estimada; la cual no depende, directamente, del criterio del investigador, y si de la información contenida en los datos. Además de esta manera Δ y el subespacio ortogonal a Δ , tienen poca interacción.

Esto sugiere que se puede obtener el estimador del óptimo con pocos *datos*.

Específicamente se propone en $t\Delta, 2t\Delta, 3t\Delta, \dots, kt\Delta$ se tomen k nuevas observaciones, Y_1, Y_2, \dots, Y_k ; con k un número fijo.

Con estas observaciones se estima la función

$$f(u) = \beta_0 + \beta_1 u + \beta_2 u^2, \quad (2.5.8)$$

que define la tendencia cuadrática de $\eta(x)$ en dirección de Δ y la variable u determina el factor real que multiplica al vector $t\Delta$.

Si $\hat{\beta}_2 < 0$, entonces en el valor de $u_0 = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$ es en donde $\hat{f}(u)$ alcanza el máximo;

así, el máximo estimado de $\eta(\mathbf{x})$ en dirección de Δ es en el punto $c = u_0 t \Delta = -\frac{\hat{\beta}_1}{2\hat{\beta}_2} t \Delta$.

Por consiguiente c resulta ser el centro del nuevo diseño.

Si $\hat{\beta}_2 > 0$, se toma un nuevo valor para el escalar t (igual a kt), y se estima nuevamente la tendencia cuadrática de $\eta(\mathbf{x})$ en dirección de Δ .

Otras modificaciones son presentadas en el trabajo de Perez [39]. En donde se propone incluir toda la información disponible de los diseños anteriores al estimar nuevamente la tendencia lineal local de $\eta(\mathbf{x})$, al terminar el paso 4 del algoritmo de Ascenso por Pendiente Máxima. Esto permite tener estimadores de menor varianza.

También se propone *agrandar* cada nuevo diseño por un factor proporcional a la magnitud del vector en donde se manifestó el cambio de ascendente a descendente en la trayectoria de ascenso.

Por último, este autor propone modificar el paso 5, condicionando la terminación del algoritmo a que el óptimo estimado se encuentre *cerca* del diseño, con el criterio de una métrica particular. Esto es que $\hat{\mathbf{x}}_{op} \in S^2$.

Note que en cada nuevo proceso de búsqueda, el conjunto S es recalculado.

Para sondear las bondades de los diferentes métodos de búsqueda que hasta el momento se han descrito, en [39], se llevó a cabo una serie de simulaciones cuyas particularidades y resultados se especifican en el capítulo 3, página 33 de su trabajo. Cabe destacar que la Modificación Myers-Khuri (1979) con las modificaciones de Pérez (1991), resulta hasta el momento, el mejor algoritmo de A.P.M. en la búsqueda de puntos y regiones óptimas.

²Para un mayor alcance sobre las 3 últimas modificaciones, remítase al trabajo de [39]

Capítulo 3

Análisis De Cordilleras

En el capítulo precedente se han estudiado tres procedimientos que forman parte de la metodología empleada para optimizar superficies de respuestas. Unos son complementarios de otros.

Se observa que los multiplicadores de Lagrange son herramienta necesaria para desarrollar el Método de Ascenso por Pendiente Máxima y que este último debe aplicarse antes que el Análisis Canónico para así obtener mejores resultados. El Análisis de Cordilleras no será la excepción, aquí también se necesitará utilizar la técnica de los Multiplicadores de Lagrange.

El Análisis de Cordilleras (A.CO.) es similar al Ascenso por Pendiente Máxima, ya que también se busca maximizar la respuesta ajustada restringida a esferas de radio variante. Sin embargo, el A.CO. por lo general se aplica post Análisis Canónico y la función objetivo es de segundo orden.

El A.CO. se utilizará cuando el Análisis Canónico muestre que la superficie de respuesta ajustada tiene una Cordillera Ascendente o una Descendente, o bien que la superficie sea una silla de montar. Por ejemplo, en caso de una cordillera ascendente el experimentador desearía aumentar el Análisis Canónico tomando observaciones a lo largo de una cordillera en su búsqueda del óptimo. Es decir, este método permite localizar óptimos que con el Análisis Canónico no se puedan determinar, aunque también puede ser aplicado directamente.

El término Análisis de Cordilleras fue acuñado por Hoerl [23], pero su trabajo más que un análisis completo es una exposición, la cual no refleja didácticamente los pasos a seguir en el análisis y mucho menos la naturaleza teórica del método. Un desarrollo riguroso acerca de la teoría de la técnica de Análisis de Cordilleras se presenta en el

trabajo de Drapper [15].

3.1 EL ANALISIS DE COORDILLERAS

Suponga que después de aplicar el Análisis Canónico se localizó un punto crítico que resultado estar alejado de la zona de experimentación o bien resultó ser un punto silla. El problema sigue siendo el optimizar $\hat{\eta}$ (superficie de respuesta ajustada) sobre toda la región de experimentación.

Para comprender mejor el sistema, la idea del A.CO. es ir optimizando la función alejándose poco a poco del origen y obteniendo puntos con respuesta óptima. Esto se puede lograr si se maximiza $\hat{\eta}$, restringiendola a esferas de radio R , en donde R toma valores cada vez mas grandes. Entonces se desea maximizar la respuesta

$$\hat{\eta} = b_0 + \sum_{i=1}^k b_i x_i + \underbrace{\sum_i^k \sum_j^k b_{ij} x_i x_j}_{i < j}, \quad (3.1.1)$$

sujeta a

$$\sum_{i=1}^k x_i^2 = R^2.$$

Aunque después de aplicar el Análisis Canónico la función $\hat{\eta}$ tiene la forma

$$\hat{\eta} = \hat{\eta}_0 + \sum_{i=1}^k \delta_i w_i^2,$$

con ésta, ya no es posible obtener más información. Por lo que se debe trabajar, más bien, con la expresión (3.1.1).

Para resolver el problema se utiliza el método de los multiplicadores de lagrange. Por lo cual hay que maximizar

$$\hat{\eta} = f(x_1, x_2, \dots, x_k) = b_0 + \sum_{i=1}^k b_i x_i + \underbrace{\sum_i^k \sum_j^k b_{ij} x_i x_j}_{i < j},$$

sujeto a

$$g(x_1, x_2, \dots, x_k) = \sum_{i=1}^k x_i^2 - R^2.$$

Sea \mathcal{L} la lagrangeana, así

$$\mathcal{L} = b_0 + \sum_{i=1}^k b_i x_i + \underbrace{\sum_i \sum_j b_{ij} x_i x_j}_{i < j} - \lambda \left[\sum_{i=1}^k x_i^2 - R^2 \right].$$

Usando notación matricial, \mathcal{L} se expresa de la siguiente manera

$$\mathcal{L} = b_0 + \mathbf{x}^t B \mathbf{x} + \mathbf{x}^t \mathbf{b} - \lambda (\mathbf{x}^t \mathbf{x} - R^2),$$

luego

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 2B\mathbf{x} + \mathbf{b} - 2\lambda \mathbf{x},$$

que de manera explícita es

$$\begin{bmatrix} \partial \mathcal{L} / \partial x_1 \\ \partial \mathcal{L} / \partial x_2 \\ \vdots \\ \partial \mathcal{L} / \partial x_k \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} + 2 \begin{bmatrix} b_{11} & b_{12}/2 & \cdots & b_{1k}/2 \\ & b_{22} & \cdots & b_{2k}/2 \\ & & \ddots & \vdots \\ & & & b_{kk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} - 2\lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix},$$

además

$$\begin{bmatrix} \partial^2 \mathcal{L} / \partial x_1^2 \\ \partial^2 \mathcal{L} / \partial x_2^2 \\ \vdots \\ \partial^2 \mathcal{L} / \partial x_k^2 \end{bmatrix} = \begin{bmatrix} 2b_{11} - 2\lambda \\ 2b_{22} - 2\lambda \\ \vdots \\ 2b_{kk} - 2\lambda \end{bmatrix},$$

es claro que $\frac{\partial^2 \mathcal{L}}{\partial x_i \partial x_j} = b_{ij}$, ($i \neq j$), de donde

$$H(\mathbf{x}) = 2(B - \lambda I),$$

es la matriz Hessiana o de segundas derivadas, con

$$B = \begin{bmatrix} b_{11} & b_{12}/2 & \cdots & b_{1k}/2 \\ & b_{22} & \cdots & b_{2k}/2 \\ & & \ddots & \vdots \\ & & & b_{kk} \end{bmatrix},$$

y los puntos estacionarios son aquellos que satisfacen simultáneamente las ecuaciones

$$\begin{cases} (B - \lambda I)\mathbf{x} = -\frac{1}{2}\mathbf{b} & \text{con } \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \\ \sum_{i=1}^k x_i^2 - R^2 = 0. \end{cases} \quad (3.1.2)$$

Los puntos estacionarios se determinan fácilmente, si se resuelve el sistema $(B - \lambda I)x = -\frac{b}{2}$, asignando valores a λ . Posteriormente se calcula x y finalmente se obtiene R . Esto da lugar al siguiente Algoritmo.

ALGORITMO PARA EL CALCULO DE PUNTOS ESTACIONARIOS

- [1] Considérese R como variable y sea λ un valor fijo.
- [2] Sustitúyase el valor de λ en la expresión $(B - \lambda I)x = -\frac{b}{2}$. Así se obtendrá un punto estacionario x_0 .
- [3] Calcúlese $R = \sqrt{x_0^t x_0} = \sqrt{\sum_{i=1}^k x_i^2}$. Para saber a qué radio corresponde la restricción.
- [4] Evalúese $\hat{\eta}$ con (3.1.1).

Nótese que el punto que se obtiene depende del valor que tome λ . Puede ocurrir que el punto estacionario sea un óptimo global o local de la función $\hat{\eta}$ o bien, que sea un punto silla según el valor de λ . Por lo que se debe desarrollar un criterio que permita caracterizar a los puntos obtenidos con el algoritmo.

A continuación se presentan una serie de resultados que permitirán proporcionar un algoritmo para la búsqueda de mínimos y máximos absolutos.

Proposición 3.1.1 Si $R_1 = R_2$, $\lambda_1 > \lambda_2$ entonces $\hat{\eta}_1 > \hat{\eta}_2$.

Demostración:

Se sabe que

$$(B - \lambda_1 I)x_1 = -\frac{b}{2}, \quad (3.1.3)$$

$$(B - \lambda_2 I)x_2 = -\frac{b}{2}, \quad (3.1.4)$$

y

$$x_1^t x_1 = x_2^t x_2 = R^2.$$

También

$$\hat{\eta}_1 = x_1^t B x_1 + x_1^t b + b_0, \quad (3.1.5)$$

$$\hat{\eta}_2 = x_2^t B x_2 + x_2^t b + b_0, \quad (3.1.6)$$

multiplicando (3.1.3) y (3.1.4) por \mathbf{x}_1^t y \mathbf{x}_2^t respectivamente y restando se obtiene

$$\mathbf{x}_1^t B \mathbf{x}_1 - \mathbf{x}_2^t B \mathbf{x}_2 + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)^t \mathbf{b} = (\lambda_1 - \lambda_2) R^2, \quad (3.1.7)$$

sustrayendo (3.1.6) de (3.1.5) y utilizando (3.1.7), se obtiene

$$\begin{aligned} \hat{\eta}_1 - \hat{\eta}_2 &= \mathbf{x}_1^t B \mathbf{x}_1 - \mathbf{x}_2^t B \mathbf{x}_2 + (\mathbf{x}_1 - \mathbf{x}_2)^t \mathbf{b} \\ &= \mathbf{x}_1^t B \mathbf{x}_1 - \mathbf{x}_2^t B \mathbf{x}_2 + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)^t \mathbf{b} + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)^t \mathbf{b} \\ &= (\lambda_1 - \lambda_2) R^2 + \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)^t \mathbf{b}. \end{aligned} \quad (3.1.8)$$

Multiplicando (3.1.3) por \mathbf{x}_2^t y (3.1.4) por \mathbf{x}_1^t y tomando la correspondiente resta se obtiene

$$(\lambda_1 - \lambda_2) \mathbf{x}_1^t \mathbf{x}_2 = \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)^t \mathbf{b} \quad (3.1.9)$$

en donde se utilizó el hecho de que $\mathbf{x}_2^t B \mathbf{x}_1 = \mathbf{x}_1^t B \mathbf{x}_2$ y $\mathbf{x}_2^t \mathbf{x}_1 = \mathbf{x}_1^t \mathbf{x}_2$, dado que un número real es igual a su transpuesto. Sustituyendo (3.1.9) en (3.1.8) se tiene que

$$\begin{aligned} \hat{\eta}_1 - \hat{\eta}_2 &= (\lambda_1 - \lambda_2) R^2 - (\lambda_1 - \lambda_2) \mathbf{x}_1^t \mathbf{x}_2 \\ &= (\lambda_1 - \lambda_2)(R^2 - \mathbf{x}_1^t \mathbf{x}_2), \end{aligned}$$

pero

$$R^2 - \mathbf{x}_1^t \mathbf{x}_2 = \|\mathbf{x}_1\| \cdot \|\mathbf{x}_2\| - (\mathbf{x}_1^t \cdot \mathbf{x}_2) > 0,$$

ya que por la desigualdad de Cauchy-Schwartz $\mathbf{x}_1^t \cdot \mathbf{x}_2 < \|\mathbf{x}_1\| \cdot \|\mathbf{x}_2\|$. Por tanto, $\hat{\eta}_1 > \hat{\eta}_2$. □

Proposición 3.1.2 Si $R_1 = R_2$, $H(\mathbf{x}_1)$ es definida positiva y $H(\mathbf{x}_2)$ es indefnida. Entonces $\hat{\eta}_1 < \hat{\eta}_2$.

Una matriz H de $k \times k$ se dice indefnida, si existe al menos un vector $\mathbf{z}_0 \in \mathbb{R}^k$, tal que $\mathbf{z}_0^t H \mathbf{z}_0 > 0$ y un vector \mathbf{z}_1 tal que $\mathbf{z}_1^t H \mathbf{z}_1 < 0$.

Demostración:

Puesto que $H(\mathbf{x}_1)$ es definida positiva para todo \mathbf{u} vector de $k \times 1$, $\mathbf{u} \neq 0$, se tiene que

$$\mathbf{u}^t H(\mathbf{x}_1) \mathbf{u} > 0, \quad (3.1.10)$$

puesto que $H(\mathbf{x}_2)$ es indefinida, existe al menos un vector $\mathbf{q} \neq 0$ de $k \times 1$ tal que

$$\mathbf{q}'H(\mathbf{x}_2)\mathbf{q} < 0. \quad (3.1.11)$$

Como $H(\mathbf{x}_1) = (B - \lambda_1 I)$, entonces $\mathbf{u}'B\mathbf{u} - \lambda_1 \mathbf{u}'\mathbf{u} > 0$, en particular para el vector \mathbf{q} resulta que

$$\mathbf{q}'B\mathbf{q} - \lambda_1 \mathbf{q}'\mathbf{q} > 0$$

por lo que

$$\mathbf{q}'B\mathbf{q} - \lambda_2 \mathbf{q}'\mathbf{q} < 0.$$

Por lo anterior se tiene que

$$\lambda_1 \mathbf{q}'\mathbf{q} < \mathbf{q}'B\mathbf{q} < \lambda_2 \mathbf{q}'\mathbf{q}.$$

Se puede cancelar $\mathbf{q}'\mathbf{q}$, puesto que es un número real positivo, entonces $\lambda_1 < \lambda_2$. Utilizando la Proposición (3.1.1) se concluye que $\hat{\eta}_1 < \hat{\eta}_2$.

□

Proposición 3.1.3 Si $\lambda_v > \delta_i$ (i -ésimo valor propio de la matriz B), para toda $i = 1, \dots, k$, entonces $\hat{\eta}$ alcanza un máximo local en \mathbf{x}_v . De manera análoga, si $\lambda_v < \delta_i$ (i -ésimo valor propio de la matriz B), para toda $i = 1, \dots, k$, entonces $\hat{\eta}$ alcanza un mínimo local en \mathbf{x}_v .

Demostración:

Se desea probar que la matriz $H(\mathbf{x}_1)$ es definida positiva (negativa) para que se alcance un mínimo (máximo) local de $\hat{\eta}$.

Sea \mathbf{u} vector de $k \times 1$, $\mathbf{u} \neq 0$. Para $H(\mathbf{x}_1)$ se tiene que

$$\mathbf{u}'H(\mathbf{x}_1)\mathbf{u} = \mathbf{u}'(B - \lambda_1 I)\mathbf{u} = \mathbf{u}'B\mathbf{u} - \lambda_1 \mathbf{u}'\mathbf{u}.$$

Considérese $\mathbf{u}'B\mathbf{u}$, del capítulo 2 se sabe que existe una transformación ortogonal $\mathbf{u}' = \mathbf{v}'T'$, de tal manera que

$$\mathbf{u}'B\mathbf{u} = \mathbf{v}'[\text{diag}(\delta_1, \dots, \delta_k)]\mathbf{v},$$

donde $\text{diag}(\delta_1, \dots, \delta_k)$ denota una matriz diagonal cuyos elementos son las raíces características de B .

Como T es una matriz ortogonal

$$\lambda_1 \mathbf{u}' \mathbf{u} = \lambda_1 \mathbf{v}' T' T \mathbf{v} = \lambda_1 \mathbf{v}' \mathbf{v}.$$

Entonces

$$\begin{aligned} \mathbf{u}' H(\mathbf{x}_1) \mathbf{u} &= \mathbf{u}' (B - \lambda_1 I) \mathbf{u} = \mathbf{u}' B \mathbf{u} - \lambda_1 \mathbf{u}' I \mathbf{u} \\ &= \mathbf{v}' \text{diag}(\delta_1, \dots, \delta_k) \mathbf{v} - \lambda_1 \mathbf{v}' I \mathbf{v} \\ &= \mathbf{v}' \text{diag}(\delta_1 - \lambda_1, \delta_2 - \lambda_1, \dots, \delta_k - \lambda_1) \mathbf{v} \\ &= \sum_{i=1}^k (\delta_i - \lambda_1) v_i^2. \end{aligned}$$

Si $\mathbf{u} \neq 0$, entonces $\mathbf{v} \neq 0$ (ya que $\mathbf{u}' = \mathbf{v}' T'$). Ahora bien, si $\lambda_1 > \delta_i$ para toda $i = 1, \dots, k$, entonces, $\sum_{i=1}^k (\delta_i - \lambda_1) v_i^2$ es menor que cero y por lo tanto $H(\mathbf{x}_1)$ es definida negativa (se alcanza un máximo local en $\hat{\eta}$). Si $\lambda_1 < \delta_i$, para toda $i = 1, \dots, k$, entonces $\sum_{i=1}^k (\delta_i - \lambda_1) v_i^2$ es mayor que cero de donde $H(\mathbf{x}_1)$ es definida positiva, es decir, se alcanza un mínimo local de $\hat{\eta}$.

□

Proposición 3.1.4 *Supóngase que en la medida que R se incrementa, se traza un lugar geométrico de puntos estacionarios del mismo tipo (máximos, mínimos o puntos silla), entonces se cumple una de las siguientes afirmaciones:*

[i] $\hat{\eta}$ decrece monótonamente.

[ii] $\hat{\eta}$ crece monótonamente.

[iii] $\hat{\eta}$ llega a un máximo y posteriormente decrece monótonamente.

[iv] $\hat{\eta}$ llega a un mínimo y posteriormente crece monótonamente.

Demostración:

Se desea encontrar una expresión adecuada para $\frac{\partial \hat{\eta}}{\partial R}$ con

$$\hat{\eta} = \mathbf{b}_0 + \mathbf{x}' B \mathbf{x} + \mathbf{x}' \mathbf{b}.$$

El vector \mathbf{x} tiene que cumplir con la ecuación

$$(B - \lambda I)\mathbf{x} = -\frac{\mathbf{b}}{2},$$

ya que pertenece a un lugar geométrico de puntos estacionarios.

Suponga que ocurre un pequeño cambio $\Delta\lambda$ en λ ; por la ecuación anterior, éste inducirá pequeños cambios $\Delta\mathbf{x}$ en \mathbf{x} y por tanto un pequeño cambio ΔR en R .

Multiplicando la ecuación anterior por \mathbf{x}' , se obtiene

$$\mathbf{x}'B\mathbf{x} = \lambda\mathbf{x}'\mathbf{x} - \frac{\mathbf{x}'\mathbf{b}}{2}.$$

Sustituyendo esta última expresión en $\hat{\eta}$, se llega a que

$$\hat{\eta} = b_0 + \lambda\mathbf{x}'\mathbf{x} - \frac{\mathbf{x}'\mathbf{b}}{2} + \mathbf{x}'\mathbf{b} = b_0 + \lambda\mathbf{x}'\mathbf{x} + \frac{\mathbf{x}'\mathbf{b}}{2}, \quad (3.1.12)$$

en donde $\hat{\eta}$ está en función de λ y de \mathbf{x} .

Dado que un cambio infinitesimal en λ provoca un cambio infinitesimal en \mathbf{x} , la respuesta estimada, $\hat{\eta}$, también cambia infinitesimalmente debido a la ecuación (3.1.12). Análiticamente se tiene que

$$\begin{aligned} \hat{\eta} + \Delta\hat{\eta} &= b_0 + (\lambda + \Delta\lambda)(\mathbf{x} + \Delta\mathbf{x})'(\mathbf{x} + \Delta\mathbf{x}) + \frac{(\mathbf{x} + \Delta\mathbf{x})'\mathbf{b}}{2} \\ &= b_0 + (\lambda + \Delta\lambda)(\mathbf{x}' + (\Delta\mathbf{x})')(\mathbf{x} + \Delta\mathbf{x}) + \frac{\mathbf{x}'\mathbf{b}}{2} + \frac{(\Delta\mathbf{x})'\mathbf{b}}{2} \\ &= b_0 + \lambda\mathbf{x}'\mathbf{x} + \lambda(\Delta\mathbf{x})'\mathbf{x} + (\Delta\lambda)\mathbf{x}'\mathbf{x} + (\Delta\lambda)(\Delta\mathbf{x})'\mathbf{x} \\ &\quad + \lambda\mathbf{x}'(\Delta\mathbf{x}) + \lambda(\Delta\mathbf{x})'(\Delta\mathbf{x}) + (\Delta\lambda)\mathbf{x}'(\Delta\mathbf{x}) \\ &\quad + (\Delta\lambda)(\Delta\mathbf{x})'(\Delta\mathbf{x}) + \frac{\mathbf{x}'\mathbf{b}}{2} + \frac{(\Delta\mathbf{x})'\mathbf{b}}{2}. \end{aligned} \quad (3.1.13)$$

en donde $\Delta\lambda$, $\Delta\mathbf{x}$ y $\Delta\hat{\eta}$ denotan incrementos en λ , \mathbf{x} y $\hat{\eta}$ respectivamente. sustrayendo la ecuación (3.1.12) de (3.1.13) se llega a que

$$\Delta\hat{\eta} = 2\lambda\mathbf{x}'(\Delta\mathbf{x}) + (\Delta\lambda)\mathbf{x}'\mathbf{x} + \frac{(\Delta\mathbf{x})'\mathbf{b}}{2} + Q_2,$$

con

$$Q_2 = (\Delta\lambda)(\Delta\mathbf{x})'\mathbf{x} + \lambda(\Delta\mathbf{x})'(\Delta\mathbf{x}) + (\Delta\lambda)\mathbf{x}'(\Delta\mathbf{x}) + (\Delta\lambda)(\Delta\mathbf{x})'(\Delta\mathbf{x}),$$

que denota términos de segundo orden en $\Delta\lambda$ y $\Delta\mathbf{x}$.

La ecuación (3.1.9) de la Proposición (3.1.1) se obtuvo para cualquiera \mathbf{x}_1 (con el correspondiente λ_1) y \mathbf{x}_2 (con el correspondiente λ_2) puntos críticos de \mathcal{L} , entonces

$$(\lambda_2 - \lambda_1)\mathbf{x}_1^t\mathbf{x}_2 = \frac{(\mathbf{x}_1 - \mathbf{x}_2)^t\mathbf{b}}{2}.$$

Haciendo $\lambda_2 = \lambda + \Delta\lambda$, $\lambda_1 = \lambda$, $\mathbf{x}_2 = \mathbf{x} + \Delta\mathbf{x}$, $\mathbf{x}_1 = \mathbf{x}$ y sustituyendo se tiene que

$$(\Delta\lambda)\mathbf{x}^t(\mathbf{x} + \Delta\mathbf{x}) = -\frac{(\Delta\mathbf{x})^t\mathbf{b}}{2},$$

que implica

$$\Delta\lambda\mathbf{x}^t\mathbf{x} + (\Delta\lambda)\mathbf{x}^t(\Delta\mathbf{x}) + \frac{(\Delta\mathbf{x})^t\mathbf{b}}{2} = 0,$$

y de aquí que

$$\Delta\lambda\mathbf{x}^t\mathbf{x} + \frac{(\Delta\mathbf{x})^t\mathbf{b}}{2} = Q_2'$$

con $Q_2' = -(\Delta\lambda)\mathbf{x}^t(\Delta\mathbf{x})$. Por lo tanto,

$$\Delta\hat{\eta} = 2\lambda\mathbf{x}^t\Delta\mathbf{x} + Q_2' + Q_2 = 2\lambda\mathbf{x}^t\Delta\mathbf{x} + Q_2''.$$

con $Q_2'' = Q_2' + Q_2$.

Sin embargo, es de interés encontrar una expresión para $\Delta\hat{\eta}$ que sólo dependa de R , por lo que si se considera $\Delta\mathbf{x}$ tal que sea paralelo a \mathbf{x} , es decir el ángulo entre los dos vectores es cero¹, resulta que

$$\mathbf{x}^t(\Delta\mathbf{x}) = \mathbf{x} \cdot (\Delta\mathbf{x}) = \|\mathbf{x}\| \cdot \|\Delta\mathbf{x}\| = R(\Delta R)$$

y así

$$\Delta\hat{\eta} = 2\lambda R(\Delta R) + Q_2''$$

por tanto

$$\frac{\Delta\hat{\eta}}{\Delta R} = 2\lambda R + \frac{Q_2''}{\Delta R}.$$

¹recuerde que para dos vectores \mathbf{a} y \mathbf{b} el $\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$, con θ el ángulo entre los dos vectores

Haciendo tender a cero todos los incrementos se obtiene

$$\frac{\partial \hat{\eta}}{\partial R} = 2\lambda R.$$

Considere ahora, el lugar geométrico con puntos estacionarios del mismo tipo (máximos o mínimos absolutos, máximos o mínimos locales, o bien puntos silla).

La función $\hat{\eta}(R)$ tiene a lo más dos puntos críticos, en $\lambda = 0$ o $R = 0$.

Si $R = 0$, estamos en el origen y el lugar geométrico es de mínimos o máximos absolutos. Esto último debido a que el problema de optimizar $\hat{\eta}$ sujeto a $\sum_{i=1}^k x_i^2 = 0$, sólo tiene un punto que cumple con la restricción, ese punto es máximo o mínimo.

Cuando $R \neq 0$, $\hat{\eta}$ es estacionaria con respecto a R sólo si $\lambda = 0$, pero si $\lambda = 0$, la ecuación (3.1.2), da como solución el centro de la superficie de segundo orden, esto es

$$B\mathbf{x} = -\frac{\mathbf{b}}{2},$$

y por tanto

$$\mathbf{x} = -\frac{B^{-1}\mathbf{b}}{2}.$$

O bien, el punto en el cual

$$\frac{\partial \hat{\eta}}{\partial x_i} = 0 \quad \forall i = 1, \dots, k,$$

que coincide con el punto estacionario sin restricciones al cual se denota por \mathbf{x}_0 .

Considérese un lugar geométrico que pase por el punto estacionario sin restricciones \mathbf{x}_0 y sea $R_0 = \mathbf{x}_0' \mathbf{x}_0$. Si $R < R_0$, a la función $\hat{\eta}(R)$, sólo le queda crecer o decrecer. (Sólo tiene un punto crítico, en $R = R_0$). Lo mismo ocurre si $R > R_0$. Quizá uno de los puntos críticos a considerar sea $R = 0$, que haría que la derivada se anulara, pero este punto está en el extremo inferior del dominio de $\hat{\eta}(R)$.

Cuando el lugar geométrico no pasa por el punto estacionario sin restricciones, λ siempre es diferente de cero y la función $\hat{\eta}(R)$, sólo crece o decrece ya que no tiene puntos críticos. (Excepto quizá por $R = 0$).

De aquí que *i*), *ii*), *iii*), *iv*) de la proposición (3.1.4) se cumplen y por lo mencionado anteriormente, si *iii*) y *iv*) ocurren, entonces el lugar geométrico pasa por el centro del

sistema.

□

Las cuatro proposiciones anteriores, dejan ver claramente la que a continuación se muestra y en la que se basa realmente el A.CO.

Proposición 3.1.5 *Considere nuevamente la matriz B . Sean $\delta_1, \delta_2, \dots, \delta_k$ sus k valores propios (estos son reales puesto que la matriz B es simétrica de entradas reales) y suponga que $\delta_{i+1} > \delta_i$ con $i = 1, \dots, k-1$. Sean λ_0, λ_1 tales que*

$$\lambda_0 \leq \delta_1 \leq \delta_2, \dots, \leq \delta_k \leq \lambda_1.$$

Si \mathbf{x}_0 satisface que $(B - \lambda_1 I)\mathbf{x}_1 = -\mathbf{b}/2$ y \mathbf{x}_1 satisface $(B - \lambda_0 I)\mathbf{x}_0 = -\mathbf{b}/2$, entonces \mathbf{x}_0 es un mínimo global de $\hat{\eta}$, \mathbf{x}_1 es un máximo global de $\hat{\eta}$ restringido a los radios $R_0 = \mathbf{x}_0^t \mathbf{x}_0$ y $R_1 = \mathbf{x}_1^t \mathbf{x}_1$, respectivamente.

Demostración:

Este resultado es claro a partir de los cuatro demostrados anteriormente. Lo único que resta por argumentar, es por qué los máximos (mínimos) locales de la proposición (3.1.3) son en este resultado globales.

Para esto, divídase el rango de λ (pensada como variable), llamado de $-\infty$ a ∞ , en $k+1$ intervalos utilizando los valores característicos de la matriz B , suponiendo que ya están ordenados, es decir, $\delta_1 \leq \delta_2 \leq \dots \leq \delta_k$.

Estos intervalos son de la forma

$$(-\infty, \delta_1), (\delta_1, \delta_2), \dots, (\delta_{k-1}, \delta_k), (\delta_k, \infty).$$

Además

$$\begin{aligned}(B - \lambda I_k) &= U \text{diag}(\delta_1, \delta_2, \dots, \delta_k) U^t - \lambda I_k \\ &= U \text{diag}(\delta_1, \delta_2, \dots, \delta_k) U^t - \lambda U U^t \\ &= U \text{diag}(\delta_1 - \lambda, \delta_2 - \lambda, \dots, \delta_k - \lambda) U^t\end{aligned}$$

dado que U es ortogonal.

Ahora, si λ tiende a δ_i , entonces el determinante $|(B - \lambda I)|$ tiende a cero, de donde \mathbf{x} tiende a $\pm\infty$, de tal manera que R tiende a ∞ .

Si λ tiende a $\pm\infty$ entonces el determinante $|(B - \lambda I)|$ tiende a ∞ , por lo que x tiende a cero, de tal forma que R tiende a cero.

□

Ahora se demostrará que $\frac{\partial R^2}{\partial \lambda^2}$ es positiva para toda $R \neq 0$. Esto permitirá elaborar una gráfica donde R será la ordenada y λ la abscisa.

Proposición 3.1.6 $\frac{\partial R^2}{\partial \lambda^2}$ es positiva para toda $R \neq 0$.

Demostración:

Se sabe que

$$(B - \lambda I)x = -\frac{b}{2}, \quad (3.1.14)$$

$$x^t x = R^2. \quad (3.1.15)$$

Derivando (3.1.14) con respecto a λ se tiene que

$$\frac{\partial(B - \lambda I)x}{\partial \lambda} = \frac{\partial(-b/2)}{\partial \lambda} = 0.$$

Por la regla de la derivada de un producto se tiene que

$$\frac{\partial(B - \lambda I)x}{\partial \lambda} = -x + (B - \lambda I) \frac{\partial x}{\partial \lambda}$$

que implica

$$(B - \lambda I) \frac{\partial x}{\partial \lambda} = x. \quad (3.1.16)$$

Haciendo lo mismo con la ecuación (3.1.15) se tiene

$$x^t \frac{\partial x}{\partial \lambda} + \frac{\partial x^t}{\partial \lambda} x = 2R \frac{\partial R}{\partial \lambda}.$$

Como $x^t \frac{\partial x}{\partial \lambda} = \frac{\partial x^t}{\partial \lambda} x$, se tiene que

$$x^t \frac{\partial x}{\partial \lambda} = R \frac{\partial R}{\partial \lambda}. \quad (3.1.17)$$

Tomando la segunda derivada de la ecuación (3.1.14) respecto a λ , se obtiene

$$-\frac{\partial \mathbf{x}}{\partial \lambda} + (B - \lambda I) \frac{\partial^2 \mathbf{x}}{\partial \lambda^2} = \frac{\partial \mathbf{x}}{\partial \lambda}$$

que implica

$$(B - \lambda I) \frac{\partial^2 \mathbf{x}}{\partial \lambda^2} = 2 \frac{\partial \mathbf{x}}{\partial \lambda}. \quad (3.1.18)$$

Diferenciando por segunda vez (3.1.15) queda

$$\mathbf{x}^t \frac{\partial^2 \mathbf{x}}{\partial \lambda^2} + \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) = R \frac{\partial^2 R}{\partial \lambda^2} + \left(\frac{\partial R}{\partial \lambda} \right)^2. \quad (3.1.19)$$

Si se multiplica (3.1.16) por $\left(\frac{\partial^2 \mathbf{x}}{\partial \lambda^2} \right)^t$ la ecuación quedaría como

$$\left(\frac{\partial^2 \mathbf{x}}{\partial \lambda^2} \right)^t (B - \lambda I) \frac{\partial \mathbf{x}}{\partial \lambda} = \left(\frac{\partial^2 \mathbf{x}}{\partial \lambda^2} \right)^t \mathbf{x}$$

que se cumple si y sólo si ocurre

$$\left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t (B - \lambda I) \frac{\partial^2 \mathbf{x}}{\partial \lambda^2} = \mathbf{x}^t \frac{\partial^2 \mathbf{x}}{\partial \lambda^2}, \quad (3.1.20)$$

puesto que $(B - \lambda I)$ es simétrica.

Si la ecuación (3.1.18) se multiplica por $\left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t$ se obtiene

$$\left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t (B - \lambda I) \frac{\partial^2 \mathbf{x}}{\partial \lambda^2} = 2 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right). \quad (3.1.21)$$

Si se restan (3.1.20) y (3.1.21) se obtiene que

$$\mathbf{x}^t \frac{\partial^2 \mathbf{x}}{\partial \lambda^2} - 2 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) = 0. \quad (3.1.22)$$

La ecuación (3.1.19) se puede reexpresar como

$$\mathbf{x}^t \frac{\partial^2 \mathbf{x}}{\partial \lambda^2} - 2 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) + 3 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) = R \frac{\partial^2 R}{\partial \lambda^2} + \left(\frac{\partial R}{\partial \lambda} \right)^2,$$

y sustituyendo (3.1.22) en la reexpresión de (3.1.19) resulta que

$$R \frac{\partial^2 R}{\partial \lambda^2} = 3 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) - \left(\frac{\partial R}{\partial \lambda} \right)^2. \quad (3.1.23)$$

Ahora,

$$\begin{aligned} \frac{\partial R}{\partial \lambda} &= \frac{\partial (\mathbf{x}^t \mathbf{x})^{\frac{1}{2}}}{\partial \lambda} \\ &= \frac{1}{2} (\mathbf{x}^t \mathbf{x})^{-\frac{1}{2}} \left[\left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \mathbf{x} + \mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda} \right] \\ &= \frac{1}{2} (\mathbf{x}^t \mathbf{x})^{-\frac{1}{2}} \left(2 \mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda} \right) \\ &= \frac{\mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda}}{(\mathbf{x}^t \mathbf{x})^{1/2}}. \end{aligned} \quad (3.1.24)$$

Multiplicando (3.1.23) por R^2 se obtiene

$$R^3 \frac{\partial^2 R}{\partial \lambda^2} = 3R^2 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) - R^2 \left(\frac{\partial R}{\partial \lambda} \right)^2 \quad (3.1.25)$$

sustituyendo (3.1.24) en (3.1.25) se tiene

$$\begin{aligned} R^3 \frac{\partial^2 R}{\partial \lambda^2} &= 2R^2 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) + R^2 \left[\left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) - \frac{(\mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda})^2}{(X^t X)^{1/2}} \right] \\ &= 2R^2 \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) + \left[\mathbf{x}^t \mathbf{x} \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right) - \left(\mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda} \right)^2 \right]. \end{aligned} \quad (3.1.26)$$

En (3.1.26) se puede ver que del lado derecho, el primer término es no negativo y sólo se hace cero si $R = 0$ o bien $\frac{\partial \mathbf{x}}{\partial \lambda} = 0$.

El segundo término del lado derecho en (3.1.26) es no negativo debido a que

$$\left| \mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda} \right| \leq \|\mathbf{x}\| \cdot \left\| \frac{\partial \mathbf{x}}{\partial \lambda} \right\|$$

por la desigualdad de Cauchy-Schwartz [19]. Entonces

$$\left| \left(\mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda} \right)^2 \right| \leq \|\mathbf{x}\|^2 \left\| \frac{\partial \mathbf{x}}{\partial \lambda} \right\|^2.$$

Esta última expresión se puede reescribir como

$$\left(\mathbf{x}^t \frac{\partial \mathbf{x}}{\partial \lambda} \right)^2 \leq \mathbf{x}^t \mathbf{x} \left(\frac{\partial \mathbf{x}}{\partial \lambda} \right)^t \frac{\partial \mathbf{x}}{\partial \lambda}$$

que es como aparece en (3.1.26).

Ese término sólo se hace cero si $\mathbf{x} = 0$, es decir $R = 0$ o bien que $\frac{\partial \mathbf{x}}{\partial \lambda} = 0$. (El paralelismo ente \mathbf{x} y $\frac{\partial \mathbf{x}}{\partial \lambda}$ no se da porque $\frac{\partial \mathbf{x}}{\partial \lambda}$ es un vector tangente a la curva en $\mathbf{x}(\lambda)$).

Si $\frac{\partial \mathbf{x}}{\partial \lambda} = 0$, la ecuación (3.1.16) dice que $\mathbf{x} = 0$ y entonces $R = 0$. Entonces por el análisis hecho en este punto y la demostración de la Proposición (3.1.5), se concluye que, $\frac{\partial^2 R}{\partial \lambda^2}$ es positiva excepto cuando $R = 0$.

Cabe señalar que si $\frac{\partial R}{\partial \lambda} = 0$, esto no implica que $\frac{\partial \mathbf{x}}{\partial \lambda} = 0$, ya que si el producto escalar de dos vectores es cero no necesariamente alguno de ellos tiene que ser el vector cero, ver ecuación (3.1.17).

□

La gráfica de R (tomada como ordenada) contra λ (tomada como abscisa), se comporta de la siguiente manera:

Si λ tiende a $-\infty$, entonces R tiende a cero, en cambio R se incrementa monótonamente con concavidad hacia arriba, tendiendo a infinito cuando λ tiende a δ_1 (el menor valor característico).

Entre dos raíces características, R viene de infinito en δ_i , pasa a través de un punto estacionario (será un mínimo local debido a la concavidad) y otra vez se encamina hacia infinito cuando se acerca a δ_{i+1} .

Finalmente para λ mayor a δ_k , pero muy cercano a ella, R toma valores en infinito y en la medida en que λ crece R tiende a cero, ver figura (3.1.1)

Cada valor de R proporciona a lo más $2k$ correspondientes valores de λ , es decir $2k$ puntos estacionarios, ya que en los intervalos $(-\infty, \delta_1)$, (δ_k, ∞) habría una λ y en (δ_i, δ_{i+1}) podría haber dos. En total quedarían $2 + 2(k-1) = 2k$.

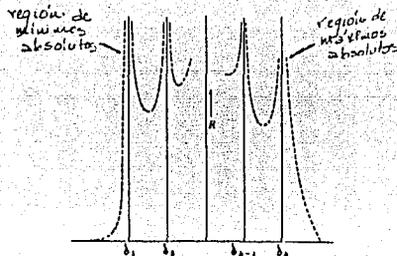


Figura 3.1.1:

El número de λ^a y por lo tanto de x^a estacionarias que corresponde a un valor de R , será menor si algunas de las curvas en la figura (3.1.1) tienen su mínimo por arriba del valor de R considerado.

Numérese estas λ^a como $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_h$, donde $h \leq 2k$. Se desea el máximo global de $\hat{\eta}$ sujeto a $\sum_{i=1}^k x_i^2 = R^2$ (R fija).

Los únicos puntos estacionarios en esta esfera son generados por $\delta_1, \delta_2, \dots, \delta_h$ al sustituirlos en $(B - \lambda I)x = -\frac{b}{2}$. Dichos puntos estacionarios son denotados por x_1, x_2, \dots, x_h .

El problema de maximización aquí siempre tiene solución ya que $\hat{\eta}$ es continua y la restricción determina un conjunto compacto. Ahora, por el Teorema de Lagrange en el caso de una sola restricción en forma de ecuación, dice que el máximo se alcanzará en x_i , para alguna $i = 1, \dots, h$.

Utilizando la proposición (3.1.1), $\hat{\eta}(x_h) > \hat{\eta}(x_i)$ para toda $i = 1, \dots, h-1$. Entonces, dentro de los puntos estacionarios, x_h es el de respuesta máxima, $\hat{\eta}$ alcanza un máximo absoluto restringiéndose a una esfera de radio R en x_h .

Puesto que $\lambda_h > \delta_k$, entonces x_h es el máximo local de la proposición (3.1.3).

De manera análoga se puede concluir que x_1 , es un punto donde se alcanza respuesta mínima.

PROCEDIMIENTO DE BUSQUEDA DE MAXIMOS O MINIMOS ABSOLUTOS.

Se desea conocer el lugar geométrico de los máximos (mínimos) absolutos de $\hat{\eta}$ cuando R se incrementa. Se debe entonces:

- [1] Conocer todos los valores propios de la matriz B . (Se supone que ellos son $\delta_1, \dots, \delta_k$

y además que $\delta_1 \leq \delta_2 \leq \dots \leq \delta_k$).

[2] Sustituir en la expresión $(B - \lambda I)x = -\frac{b}{2}$ valores de λ tales que $\lambda > \delta_k$. Esto garantizará que $H(x)$ sea una matriz negativa definida y por lo tanto el punto estacionario será un máximo absoluto de $\hat{\eta}$ sobre la esfera de radio $R(\lambda)$. (Para el caso de mínimos hay que tomar $\lambda < \delta_1$).

[3] Hacer $k + 1$ gráficas en \mathbb{R}^2 , donde la abscisa es R , la ordenada para la i -ésima gráfica ($i = 1, \dots, k$) es x_i y la ordenada para la $k + 1$ es $\hat{\eta}$.

Con unos cuantos valores de λ será posible determinar aproximadamente el lugar geométrico que podrá ser utilizado como una ruta de ascenso máximo.

La idea del apartado 3) de este procedimiento es que con unos cuantos puntos y aplicando las conclusiones de la proposición (3.1.4) se puedan determinar condiciones de optimalidad para valores de R que sean de interés.

Si la R que se tomó proporciona puntos fuera de la región de experimentación, la información obtenida es poco confiable.

Si se escoge λ entre dos valores característicos se puede caer en dos diferentes lugares geométricos dependiendo si se está a la derecha o a la izquierda del valor de λ para la cual R es estacionaria. Si se quiere maximizar o minimizar $\hat{\eta}$ estos valores de λ no son de interés.

Se concluye esta sección con una breve explicación del objetivo del A.CO. en términos del problema de superficie de respuesta, de los procedimientos aquí expuestos y un ejemplo de los mismos.

El A.CO. es un método que permite transformar un problema de optimización cuya representación gráfica estaría en \mathbb{R}^{k+1} , $k \geq 2$, a uno en \mathbb{R}^2 .

El procedimiento de búsqueda de máximos y mínimos absolutos y las formas que adopta la gráfica $\hat{\eta}$ contra R son las principales consecuencias del A.CO.

Cabe señalar que se toma como variable independiente a λ , es decir al multiplicador de Lagrange, lo cual parece ser contradictorio al tipo de gráficas que se están proponiendo, ya que R depende de λ . Esta elección es sólo por facilidad en los cálculos. Es claro que los puntos óptimos que se obtengan dependen básicamente de la esfera a la que se esté restringiendo la función $\hat{\eta}$. Por esto, R es variable independiente en las gráficas contra $\hat{\eta}$, x_1, x_2, \dots, x_k .

Se centra la atención en la gráfica que se elabora con $\hat{\eta}$, pues no se debe olvidar que la respuesta es la que interesa optimizar. Gracias a la proposición (3.1.4) se sabe que la curva que describe $\hat{\eta}$ cuando cambia R será

- a) Monótona creciente,
- b) Monótona decreciente,
- c) Pasará por un máximo y luego decrecerá,
- d) Pasará por un mínimo y luego crecerá.

Suponga que esta curva se construyó considerando puntos de respuesta máxima, que aún no se ha realizado Análisis Canónico y que es de interés ver si $\hat{\eta}$ tiene máximo o no.

Si se presenta el caso c), $\hat{\eta}$ tiene un máximo global. Entonces lo que habría que observar es si el valor de R correspondiente a este máximo global, es tal que el punto \mathbf{x} donde $\hat{\eta}$ alcanza el máximo, permite que \mathbf{x} esté dentro de la zona experimental.

Si cae dentro se tienen las condiciones de respuesta máxima, en caso contrario se presenta una cordillera ascendente.

Si ocurre a) lo más que se puede hacer es fijar un cierto nivel $\hat{\eta}$ y buscar las condiciones \mathbf{x}_0 que determinen ese nivel (con apoyo de las gráficas x_i vs. R , $i = 1, \dots, k$). Debe tenerse en mente que para el nivel que se fije en $\hat{\eta}$, \mathbf{x}_0 debe estar dentro de la Zona inicial experimental. Nótese que extrapolar puede llevar a conclusiones falsas. Algo análogo se haría en el caso d).

Si ocurriese b) se tendría que $\hat{\eta}$ tiene un máximo global en $R = 0$, es decir en el origen del diseño experimental. Puesto que en un punto máximo global se satisface la ecuación

$$B\mathbf{x} = -\frac{\mathbf{b}}{2},$$

entonces $\mathbf{b} = 0$, es decir, \mathbf{x}_0 máximo es igual al vector cero, el estimador de mínimos cuadrados de los términos lineales vale exactamente cero. Esto por supuesto difícilmente ocurre en la práctica e implica que un modelo de segundo orden no es adecuado.

Nótese que la exposición anterior se hizo pensando que no se había realizado Análisis Canónico y que las conclusiones obtenidas sobre el sistema fueron muy similares a las obtenidas en el Análisis Canónico.

Sin embargo en la introducción se mencionó que el A.CO. por lo general se aplica post-A.C. Se ilustrará esto cuando la gráfica $\hat{\eta}$ vs R es construida utilizando un lugar geométrico de puntos máximos y que el interés sobre la respuesta es maximizarla. En este caso se recurre al A. CO. cuando el A.C. ha presentado un punto silla, un punto mínimo o algún tipo de cordillera (ascendente o descendente).

Si ocurriera a) y d) nuevamente se debería seleccionar un nivel adecuado de $\hat{\eta}$ (adecuado en el sentido de que produzca valores grandes de $\hat{\eta}$) y se tendría que volver a experimentar considerando en la zona del diseño puntos x^* que den respuesta estimada alta.

Si ocurriera c) el radio R donde $\hat{\eta}$ alcanza un máximo absoluto sería tal que la esfera de radio R contendría a la zona del diseño. En este caso se podría tratar de maximizar $\hat{\eta}$ para valores de R sin salirse de la región experimental, buscar las condiciones x que dieran este máximo y tomar estas condiciones como las correspondientes a un nuevo origen experimental.

Todo esto se podría replantear diciendo que en vez de maximizar $\hat{\eta}$ se desea optimizar $\hat{\eta}$.

Antes del Análisis Canónico se tendrían que ir analizando simultáneamente las gráficas del lugar geométrico de máximos restringidos y la del lugar geométrico de mínimos restringidos.

Si por ejemplo, la gráfica de máximos correspondiera al caso c), como sólo hay un punto estacionario sin restricciones, la de mínimos correspondería al caso b) y c), después habría que ver si la R que produce la $\hat{\eta}$ máxima permite que el punto óptimo este dentro de la región experimental.

Si las gráficas se hacen Post Análisis Canónico y se obtiene un punto silla o alguna cordillera, las gráficas serían como las de la figura (3.1.2).

Ejemplo (3.1.1):[18]

Un ingeniero químico está interesado en obtener el mayor conocimiento posible, acerca de la relación existente entre el rendimiento de MERCAPTOBENZOTIOL (MBT)² en una reacción y el tiempo y temperatura de la misma. Ajusta una superficie de segundo orden para aproximar la relación, tomando como variable de respuesta el rendimiento

²Su fórmula química $C_7H_6NS_2$, y se utiliza como acelerador en pegamento para la vulcanización de ruedas. Es muy utilizado comercialmente.

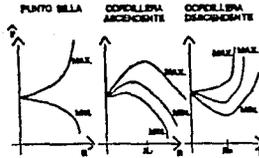


Figura 3.1.2:

de la relación existente entre el rendimiento de MERCAPTOBENZOTIOL (MBT)² en una reacción y el tiempo y temperatura de la misma. Ajusta una superficie de segundo orden para aproximar la relación, tomando como variable de respuesta el rendimiento y como variables explicativas el tiempo y temperatura.

La ecuación que resulta después del ajuste es

$$\hat{\eta} = 82.17 - 1.01x_1 - 8.61x_2 + 1.40x_1^2 - 8.76x_2^2 - 7.20x_1x_2. \quad (3.1.27)$$

donde

$$x_1 = \frac{\text{tiempo(hr)}-12}{8}, \quad x_2 = \frac{\text{temperatura}(\text{°C})-250}{30}$$

Los contornos de respuesta constante se muestran en la figura (3.1.3). Los puntos en la figura representan localizaciones de puntos reales del diseño. Obteniendo las derivadas parciales $\frac{\partial \hat{\eta}}{\partial x_i}$, e igualando a cero, para $i = 1, 2$; o bien, resolviendo la ecuación $x = -\frac{B^{-1}b}{2}$, puede verse fácilmente que el punto estacionario del sistema esta localizado en $x_0^t = (-0.439, -0.311)$. La respuesta estimada en el punto estacionario es $\hat{\eta}_0 = 83.73$.

Los valores propios de B son las raíces de la ecuación

$$\begin{vmatrix} 1.40 - \delta & -3.60 \\ -3.60 & -8.76 - \delta \end{vmatrix} = 0,$$

resultando $\delta_1 = -9.9063$ y $\delta_2 = 2.5463$, y dado que son positivos y negativos se tiene un punto silla. Suponga que el experimentador está interesado en obtener el valor máximo de rendimiento de MBT. Aplicando el A.CO., se deben escoger valores de

²Su fórmula química $C_7H_8NS_2$, y se utiliza como acelerador en pegamento para la vulcanización de ruedas. Es muy utilizado comercialmente.

resultando $\delta_1 = -9.9063$ y $\delta_2 = 2.5463$, y dado que son positivos y negativos se tiene un punto silla. Suponga que el experimentador está interesado en obtener el valor máximo de rendimiento de MBT. Aplicando el A.CO., se deben escoger valores de λ mayores de 2.5463 y substituirlos en la ecuación (3.1.2). Se necesita resolver entonces el sistema

$$\begin{bmatrix} 1.40 - \lambda & -3.60 \\ -3.60 & -8.76 - \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.505 \\ 4.305 \end{bmatrix}$$

así

$$x_1 = \frac{11.0742 - 0.505\lambda}{\lambda^2 + 7.36\lambda - 25.224},$$

$$x_2 = \frac{7.845 - 4.305\lambda}{\lambda^2 + 7.36\lambda - 25.224}.$$

Estos valores dan la localización de puntos sobre la coordillera de máximos absolutos para $\hat{\eta}$. Para un punto particular, R se calcula como

$$R^2 = x_1^2 + x_2^2,$$

y la respuesta estimada en el punto se da utilizando la función de respuesta de la ecuación (3.1.27).

Como un ejemplo del cálculo de los valores de las x^a , considere el valor de $\lambda = 4.0 > 2.5463$. Las coordenadas correspondientes estan dadas por

$$x_1 = \frac{11.072 - 0.505(4.0)}{(4.0)^2 + 7.36(4.0) - 25.224} = 0.44,$$

$$x_2 = \frac{7.845 - (4.305)(4.0)}{(4.0)^2 + 7.36(4.0) - 25.224} = -0.46.$$

Así, la coordenada (0.44, -0.46) representa un punto sobre la región de máximos absolutos de $\hat{\eta}$. El radio R se calcula como

$$R = \sqrt{(0.44)^2 + (-0.46)^2} = 0.64,$$

y en el punto en cuestión, $\hat{\eta} = 85.602$.

Valores de $\lambda > 2.5463$, fueron usados para obtener la gráfica de $\hat{\eta}$ contra R que se muestra en la figura (3.1.4) para la coordillera de máximos absolutos. La figura (3.1.5)

muestra una gráfica de x_1 y x_2 contra R . Note que en el valor $R = 1$, localizado en la frontera de la región de experimentación, el rendimiento máximo estimado de MBT es 88%, que corresponde a $x_1 \simeq 0.8$ y $x_2 \simeq -0.6$. Por tanto, el experimentador debe escoger estos niveles de tiempo y temperatura para conseguir el máximo rendimiento, ya que estos valores están en el límite de la región de diseño. La recomendación hecha por el experimentador a los talleres que reparan llantas, es que las máquinas de vulcanización funcionen a los niveles de temperatura y tiempo encontrados (obviamente decodificados) para mejores resultados del pegamento.

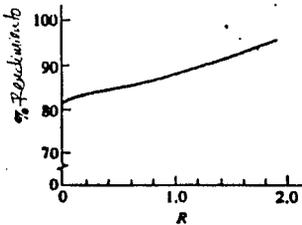


Figura 3.1.4: Rendimiento de MBT como una función de R para la cordillera de máximos absolutos.

Se presenta un problema en donde se ilustra la dificultad para optimizar diferentes respuestas al mismo tiempo.

Ejemplo (3.1.2):

Considere las superficies ajustadas de segundo orden

$$f(x_1, x_2) = 55.84 + 7.31x_1 + 26.75x_2 - 3.03x_1^2 + 6.96x_2^2 + 2.69x_1x_2$$

$$g(x_1, x_2) = 85.72 + 21.85x_1 + 8.52x_2 - 9.2x_1^2 + -5.18x_2^2 + 6.26x_1x_2,$$

ahora se aplicará a cada una de ellas la técnica planteada en esta sección. Primero se hará con $f(x_1, x_2)$, donde

$$f(x_1, x_2) = b_0 + \mathbf{x}^t \mathbf{b} + \mathbf{x}^t B \mathbf{x},$$

con

$$b_0 = 55.84, \quad \mathbf{b} = \begin{pmatrix} 7.31 \\ 26.65 \end{pmatrix}, \quad B = \begin{pmatrix} 3.03 & 1.345 \\ 1.345 & -6.96 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

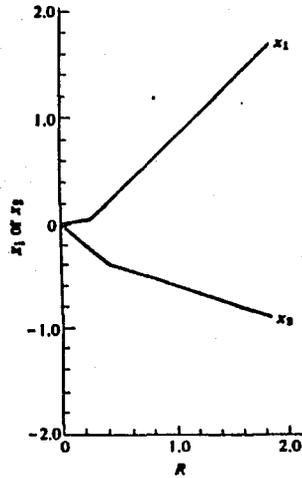


Figura 3.1.5: Gráficas de x_1 y x_2 como función de R para la cordillera de máximos absolutos

Se tiene que

$$B^{-1} = \begin{pmatrix} -0.0361 & -0.0697 \\ -0.0697 & -0.1571 \end{pmatrix}.$$

El cálculo de las raíces características se hace resolviendo la ecuación

$$\begin{vmatrix} -3.03 - \delta & 1.345 \\ 1.345 & -6.96 - \delta \end{vmatrix} = 0,$$

equivalentemente a

$$\delta^2 + 9.99\delta + 19.2798 = 0,$$

cuyas raíces son

$$\delta_1 = -2.6138, \quad \delta_2 = -7.3762.$$

Por los signos de δ_1 y δ_2 , x_0 es un máximo global. Para conocer puntos con respuesta máxima en esferas de radio R , se resuelve el sistema

$$\begin{pmatrix} -3.03 - \lambda & 1.345 \\ 1.345 & -6.96 - \lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -3.655 \\ -13.325 \end{pmatrix},$$

para valores de λ mayores que δ_1 .

Las soluciones de x_1 y x_2 para cualquier λ son

$$x_1 = \frac{343.3609 + 3.655\lambda}{\lambda^2 + 9.99\lambda + 19.2798},$$
$$x_2 = \frac{45.2906 + 13.325\lambda}{\lambda^2 + 9.99\lambda + 19.2798}.$$

Después de obtener x_1 y x_2 se evalúan R y f . La tabla siguiente muestra valores de

x_1 , x_2 , y f para λ entre -1 y 20 .

TABLA (3.1.1)				
x_1	x_2	λ	R	f
0.1881	0.5036	20.0	0.54	69.02
0.2031	0.5341	18.5	0.57	69.74
0.2207	0.5685	17.0	0.61	70.54
0.2414	0.6077	15.5	0.65	71.45
0.2662	0.6528	14.0	0.70	72.47
0.2964	0.7052	12.5	0.76	73.64
0.3340	0.7669	11.0	0.84	74.98
0.3819	0.8407	9.5	0.92	76.54
0.4449	0.9307	8.0	1.03	78.38
0.5307	1.0430	6.5	1.17	80.58
0.6541	1.1877	5.0	1.36	83.25
0.8445	1.3825	3.5	1.62	86.53
1.1713	1.6630	2.0	2.03	90.56
1.8426	2.1184	0.5	2.81	94.74
2.2490	2.3491	0.0	3.25	95.36
3.8588	3.1065	-1.0	4.95	86.80

$$g(x_1, x_2) = c_0 + \mathbf{x}'\mathbf{c} + \mathbf{x}'C\mathbf{x},$$

donde

$$c_0 = 85.72, \quad \mathbf{c} = \begin{pmatrix} 21.85 \\ 8.59 \end{pmatrix}, \quad C = \begin{pmatrix} -9.2 & -3.13 \\ -3.13 & -5.18 \end{pmatrix},$$

y se tiene que

$$C^{-1} = \begin{pmatrix} -0.1368 & .0826 \\ .0826 & -0.243 \end{pmatrix}.$$

El punto estacionario de g es

$$\mathbf{x}_1 = \begin{pmatrix} 1.1397 \\ 0.1404 \end{pmatrix}.$$

El determinante de $C - \delta I$ es igual a

$$\delta^2 + 14.38\delta + 37.8591,$$

cuyas raíces son $\delta_3 = -0.6814$ y $\delta_4 = -13.6885$.

Puntos de respuesta máxima para una esfera de radio R se encuentran resolviendo el sistema

$$\begin{pmatrix} -9.2 - \lambda & -3.13 \\ -3.13 & -5.18 - \lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -10.925 \\ -4.295 \end{pmatrix},$$

para valores de λ mayores que δ_3 .

Las soluciones del sistema son

$$x_1 = \frac{43.14815 + 10.925\lambda}{\lambda^2 + 14.38\lambda + 37.8591},$$

$$x_2 = \frac{5.3187 + 4.295\lambda}{\lambda^2 + 14.38\lambda + 37.8591}$$

La tabla siguiente muestra la misma información que la tabla (3.1.1), pero para la función g . Los valores de λ estan entre -0.5 y 20 .

TABLA (3.1.2)				
x_1	x_2	λ	R	g
0.3607	0.1257	20.0	0.38	93.12
0.3796	0.1312	18.5	0.40	93.41
0.4006	0.1371	17.0	0.42	93.73
0.4241	0.1435	15.5	0.45	94.08
0.4506	0.1504	14.0	0.48	94.45
0.4807	0.1578	12.5	0.51	94.85
0.5152	0.1658	11.0	0.54	95.28
0.5551	0.1742	9.5	0.58	95.75
0.6019	0.1829	8.0	0.63	96.25
0.6577	0.1915	6.5	0.68	96.78
0.7255	0.1988	5.0	0.75	97.33
0.8103	0.2026	3.5	0.84	97.88
0.9204	0.1970	2.0	0.94	98.39
1.0731	0.1648	0.5	1.09	98.74
1.1397	0.1405	0.0	1.15	98.77
1.2188	0.1926	-0.5	1.22	98.73

Nótese que si la respuesta en f crece los valores de x_2 también crecen, pero si g crece los correspondientes valores de x_2 decrecen.

La gráfica f contra R con sus correspondientes gráficas x_1 contra R , x_2 contra R y la gráfica de g contra R con sus correspondientes gráficas x_1 contra R , x_2 contra R , se presentan en las figuras (3.1.6) y (3.1.7) respectivamente. Si se intenta maximizar f y g en forma simultánea se observa que valores de respuesta alta para estas funciones se alcanzan en vectores (x_1, x_2) diferentes.

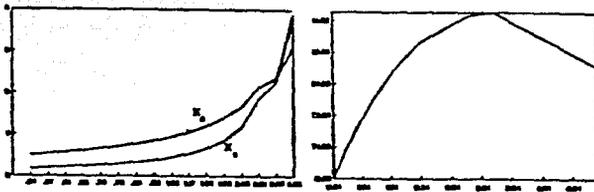


Figura 3.1.6:

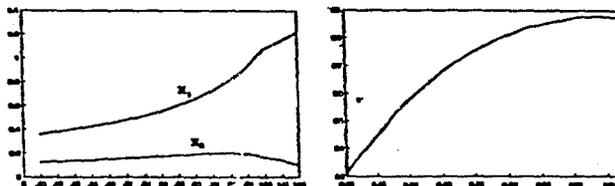


Figura 3.1.7:

3.2 ANALISIS DE COORDILLERAS MODIFICADO

El método de A.CO. usual en ningún momento considera las dificultades que puedan ocurrir debido a grandes fluctuaciones en la varianza de la respuesta estimada sobre una esfera de radio dado.

Por ello, en general el método usual puede producir estimaciones muy malas de la respuesta máxima (mínima) y de las condiciones sobre las variables independientes que determinan esta respuesta. Dichas dificultades se presentan principalmente en diseños que no son rotables. (Un diseño rotable es aquel en el que la varianza de la función estimada depende exclusivamente del radio de la esfera).

La idea de esta sección es presentar modificaciones al procedimiento que usualmente sigue el A.CO. para obtener las x^s que dan origen a las $\hat{\eta}^s$ óptimas [26].

Se estudiarán dos casos:

- I.- Cuando los niveles de varianza fija determinan curvas elipsoidales y
- II.- Cuando el diseño se ha realizado de manera no sistemática, es decir, cuando no se ha seguido un plan experimental.

En el segundo caso se logra reducir la varianza de $\hat{\eta}$, restringiendo únicamente una parte de ella, la parte que contribuye en mayor grado a que ésta crezca y que como se verá es la correspondiente al valor propio más pequeño de la matriz $x^t x$.

CASO I:

Se tiene un modelo de la forma

$$Y = \eta + \epsilon \quad \text{donde} \quad E(\epsilon) = 0 \quad \text{y} \quad E(\epsilon^t \epsilon) = \sigma^2 I,$$

con esperanza

$$\eta = \beta_0 + \sum_{i=1}^k \beta_j x_j + \sum_{j=1}^{k-1} \sum_{m>j}^k \beta_{jm} x_j x_m + \sum_{j=1}^k \beta_{jj} x_j^2.$$

Las varianzas que aquí se manejan están codificadas y el centro del diseño es el punto $(0, 0, \dots, 0)$. La función ajustada es

$$\hat{\eta} = b_0 + x^t b + x^t B x,$$

Con (x^t, b, B) como se especificarán anteriormente.

η se puede escribir como $\eta = \mathbf{z}^t \gamma$, entonces $\hat{\eta} = \mathbf{z}^t \hat{\gamma}$, donde $\hat{\gamma}$ es el estimador por mínimos cuadrados de γ y además

$$\begin{aligned}\mathbf{z}^t &= [1, x_1, \dots, x_k, x_1^2, \dots, x_k^2, x_1 x_2, \dots, x_{k-1} x_k], \\ \gamma^t &= [\beta_0, \beta_1, \dots, \beta_k, \beta_{11}, \beta_{12}, \dots, \beta_{k-1} k].\end{aligned}$$

Puesto que el modelo es lineal con respecto a los parámetros

$$\text{VAR}(\hat{\gamma}) = \sigma^2 (X^t X)^{-1},$$

con X la matriz de diseño. Por lo que

$$\begin{aligned}\text{VAR}(\hat{\eta}) &= \sigma^2 \mathbf{z}^t (X^t X)^{-1} \mathbf{z}, \\ X^t &= [z_1, z_2, \dots, z_n],\end{aligned}$$

donde n es el número de corridas experimentales ($n \geq k$) y

$$\mathbf{z}_i^t = [1, x_{1i}, \dots, x_{ki}, x_{1i}^2, \dots, x_{ki}^2, x_{1i} x_{2i}, \dots, x_{k-1i} x_{ki}],$$

para $i = 1, \dots, n$.

Este último vector denota la i -ésima corrida experimental en las variables independientes.

Con el procedimiento usual de A.C.O., lo que se hace es maximizar $\hat{\eta}$ sujeta a la restricción $\mathbf{x}^t \mathbf{x} = R^2$ (\mathbf{x} vector k -dimensional).

Se desea añadir la condición adicional de que la varianza sea constante y más o menos pequeña en toda esfera de radio R .

Cuando el diseño es rotable, esto no tiene mayor problema, ya que como antes se mencionó la varianza sólo depende del radio de la esfera y bastará con ir restringiendo $\hat{\eta}$ a esferas de radio suficientemente pequeño.

Pero si el diseño no es rotable, la varianza puede comportarse de manera no uniforme sobre la esfera y como consecuencia, acarreará malas estimaciones del máximo de $\hat{\eta}$ y dependerá de las condiciones de \mathbf{x} para obtener dicho máximo.

Se debe intentar minimizar entonces el papel de la $\text{VAR}(\hat{\eta})$ en la creación de incertidumbre en la estimación. La manera en que se hace es optimizando $\hat{\eta}$ sobre curvas de varianza constante. Es decir se optimiza $\hat{\eta}$ sujeta a

$$\mathbf{z}^t (X^t X)^{-1} \mathbf{z} = l^2,$$

donde l^2 es una constante relativamente pequeña.

Se dará un tratamiento para diseños elipsoidales, no sin antes mencionar que en la práctica, cuando el plan de diseño es modificado o simplemente se carece de algún plan, es difícil obtener diseños rotables o elipsoidales.

Se dice que un diseño es elipsoidal si $\text{VAR}(\hat{\eta}) = \psi(x^t A x)$ donde ψ es una función de $x^t A x$, A es definida positiva y simétrica.

Los diseños rotables son casos especiales de los diseños elipsoidales cuando A es la matriz identidad.

Usar el A.CO. estándar cuando el diseño es elipsoidal pero no rotable, puede llevar a salir rápidamente de la región experimental usada para calcular los estimadores de los parámetros sobre todo cuando esta región es muy angosta, produciendo graves trastornos en la varianza de la respuesta. Aún más, al estar fuera de la región experimental difícilmente se podría conocer esta varianza.

Se propone entonces, optimizar $\hat{\eta}$ sujeta a que $\text{VAR}(\hat{\eta}) = \sigma^2 l^2$. Si el diseño es elipsoidal, esto es equivalente a optimizar $\hat{\eta}$ sujeta a que $x^t A x = a^2$, donde a^2 es una constante. Para el caso en el que se trabaje con un diseño rotable, el método modificado de A.CO. coincide con el método estándar.

Este problema puede reducirse al método de A.CO. estándar de la siguiente manera: En el capítulo 2 se mencionó que existe una matriz ortogonal P tal que

$$A = P^t \text{diag}(\lambda_i) P,$$

donde P es una matriz formada por vectores propios de A , cada uno correspondiente a un valor propio de A .

$\text{diag}(\lambda_i)$ es una matriz diagonal formada con los valores propios de A .

Sean

$$w = \text{diag}(\sqrt{\lambda_i}) P x,$$

$$c = \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right) P b,$$

$$C = \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right) P B P^t \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right),$$

entonces

$$\hat{\eta} = b_0 + w^t c + w^t C w,$$

ya que

$$\begin{aligned}
 & b_0 + \mathbf{w}'\mathbf{c} + \mathbf{w}'\mathbf{C}\mathbf{w} \\
 &= b_0 + \mathbf{x}'P^t \text{diag}(\sqrt{\lambda_i}) \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right) P\mathbf{b} \\
 &\quad + \mathbf{x}'P^t \text{diag}(\sqrt{\lambda_i}) \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right) PBP^t \text{diag}\left(\frac{1}{\sqrt{\lambda_i}}\right) \text{diag}(\sqrt{\lambda_i}) P\mathbf{x} \\
 &= b_0 + \mathbf{x}'P^t P\mathbf{b} + \mathbf{x}'P^t PBP^t P\mathbf{x} \\
 &= b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'B\mathbf{x} \\
 &= \hat{\eta}.
 \end{aligned}$$

También se tiene que

$$\begin{aligned}
 \mathbf{w}'\mathbf{w} &= \mathbf{x}'P^t \text{diag}(\sqrt{\lambda_i}) \text{diag}(\sqrt{\lambda_i}) P\mathbf{x} \\
 &= \mathbf{x}'P^t \text{diag}(\lambda_i) P\mathbf{x} \\
 &= \mathbf{x}'A\mathbf{x} = a^2.
 \end{aligned}$$

Obtenidas estas expresiones no hay más que aplicar el Método de Análisis de Co-ordilleras usual a $\hat{\eta} = b_0 + \mathbf{w}'\mathbf{c} + \mathbf{w}'\mathbf{C}\mathbf{w}$ sujeta a $\mathbf{w}'\mathbf{w} = a^2$.

Ejemplo 3.2.1:[26]

Considérese un modelo ajustado en dos variables x_1, x_2 . Las variables del diseño y los correspondientes valores observados de Y se dan en la tabla (3.2.1).

x_1	x_2	Y
$-3/4\sqrt{2}$	$-5/4\sqrt{2}$	16.79
$-5/4\sqrt{2}$	$-3/4\sqrt{2}$	17.34
$5/4\sqrt{2}$	$3/4\sqrt{2}$	14.88
$3/4\sqrt{2}$	$5/4\sqrt{2}$	16.43
0	0	15.00
1	1	16.00
-1	-1	18.00
-1/4	1/4	16.50
1/4	-1/4	15.00

La matriz de varianzas y covarianzas del vector de estimadores para los coeficientes

de regresión es $\sigma^2 (X^T X)^{-1}$, donde

$$(X^T X)^{-1} = \begin{bmatrix} 1 & 0 & 0 & -4.25 & -4.25 & 7.5 \\ & 1.06 & -.94 & 0 & 0 & 0 \\ & & 1.06 & 0 & 0 & 0 \\ & & & 24.8 & 22.8 & -43.8 \\ & & & & 24.8 & -43.8 \\ & & & & & 81.3 \end{bmatrix}. \quad (3.2.1)$$

El modelo ajustado es

$$\hat{\eta} = 15 - 1.99x_1 + .99x_2 + 2.97x_1^2 + 3.97x_2^2 - 4.94x_1x_2, \quad (3.2.2)$$

y puede verse fácilmente que el diseño utilizado no es rotable. Sin embargo, la $\text{VAR}(\hat{\eta})$ tiene la forma:

$$\text{VAR}(\hat{\eta}) = \sigma^2(1 - .87\rho^2 + .34375\rho^4), \quad (3.2.3)$$

donde $\rho^2 = 8.51x_1^2 - 15x_1x_2 + 8.5x_2^2$.

De aquí que, el diseño es elipsoidal, con respecto a la forma cuadrática de la matriz A , donde

$$A = \begin{pmatrix} 8.5 & -7.5 \\ -7.5 & 8.5 \end{pmatrix}.$$

Se sigue que

$$\text{diag}(\lambda_i) = \begin{bmatrix} 1 & 0 \\ 0 & 16 \end{bmatrix} \quad \text{y} \quad P = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},$$

de tal manera que

$$(w_1, w_2)^t = \frac{1}{\sqrt{2}} [x_1 + x_2, -4x_1 + 4x_2], \quad (3.2.4)$$

que se calculó utilizando la expresión $w = \text{diag}(\sqrt{\lambda_i}) P x$.

La respuesta estimada evaluando (3.2.1) en términos de las w^s queda como

$$\hat{\eta} = 15 - .707w_1 + .527w_2 + w_1^2 + .371w_2^2 + .25w_1w_2, \quad (3.2.5)$$

que es una forma cuadrática en C donde

$$C = \begin{bmatrix} 1 & .125 \\ .125 & .371 \end{bmatrix}.$$

Después de estos cambios, hay que aplicar el A.CO. usual. Es decir considerar la función F con

$$F = \hat{\eta} - \mu(w_1^2 + w_2^2 - a^2).$$

Los puntos estacionarios satisfacen la ecuación

$$\frac{\partial F}{\partial w} = 0.$$

Las soluciones estan dadas por

$$w_1 = \frac{.164086 - .3534\mu}{\mu^2 - 1.37\mu + .355375}, \quad (3.2.6)$$

$$w_2 = \frac{-.3076875 + .2635\mu}{\mu^2 - 1.37\mu + .355375}, \quad (3.2.7)$$

que surgieron del sistema de ecuaciones dado por

$$(C - \mu I)w = \frac{-1}{2} \begin{bmatrix} -.707 \\ .527 \end{bmatrix}.$$

Recuérdese que si se desea obtener máximos de $\hat{\eta}$, entonces μ debe escogerse más grande que el mayor valor característico de la matriz C .

Los valores propios de C son: .34707 y 1.02393. Se escogieron diez diferentes valores de μ todos mayores que 1.02393. Para cada una de éstos se encontró el correspondiente máximo absoluto (w_1, w_2) sobre el círculo

$$w_1^2 + w_2^2 = a^2, \quad (3.2.8)$$

utilizando las ecuaciones (3.2.5) y (3.2.6). Se obtuvieron además las x_1 y x_2 de (3.2.3). Las a^2 y $\sigma^{-2}\text{VAR}(\hat{\eta})$ se calcularon de (3.2.7) y (3.2.2) respectivamente. La respuesta estimada en términos de las w_i está dada por (3.1.4). Los resultados se presentan en la tabla (3.2.2)

μ	4.0	3.5	3.0	2.5	2.0
w_1	-0.1149729	-0.1374639	-0.17009938	-0.2264607	-.3365082
w_2	0.0686493	0.0787206	0.092098	0.110478	.1359339
x_1	-0.0934384	-0.1111232	-0.1371986	-0.1796697	-0.2619904
x_2	-0.0691660	-0.0832899	-0.1046354	-0.1406104	-0.2139496
$\hat{\eta}$	15.130456	15.157161	15.197874	15.267882	15.418205
a	0.133908	0.1584084	0.1942185	0.2519686	0.3629266
$\sigma^{-2}\text{VAR}(\hat{\eta})$	0.9844206	0.9782599	0.9674833	0.9458335	0.8907125

μ	1.50	1.40	1.30	1.25	1.20
w_1	-0.6671172	-0.8354416	-1.1231169	-1.3608769	-1.7320725
w_2	0.1595308	0.1545867	0.1325192	0.1062461	0.0566838
x_1	-0.499949	-0.6181052	-0.8176316	-0.9811117	-1.234844
x_2	-0.4435444	-0.5634478	-0.7707765	-0.9435516	-1.214801
$\hat{\eta}$	15.983603	16.346664	17.094578	17.838156	19.231169
a	0.6859266	0.84962325	1.130908	1.365018	1.733
$\sigma^{-2}\text{VAR}(\hat{\eta})$	0.6644111	0.5474945	0.4431951	0.5630631	1.4726584

Ahora se maximizará la función de respuesta en términos de las variables originales sujeta a la condición $x_1^2 + x_2^2 = R^2$ para verificar de manera experimental que el método modificado produce mejores resultados que el método usual. Considérese entonces

$$F^* = \hat{\eta} - \theta(x_1^2 + x_2^2 - R^2).$$

La solución de $\frac{\partial F^*}{\partial x} = 0$ está dada por las ecuaciones

$$x_1 = \frac{2.7275 - 0.995\theta}{\theta^2 - 6.94\theta + 5.69},$$

$$x_2 = \frac{.9875 + .495\theta}{\theta^2 - 6.94\theta + 5.69}.$$

Los valores propios de la matriz B

$$\begin{bmatrix} 2.97 & -2.47 \\ -2.47 & 3.97 \end{bmatrix},$$

son .95 y 5.99, entonces valores de θ mayores que 5.99 deben ser considerados. La tabla (3.2.3) presenta las respuestas máximas, los correspondientes valores de R , los de $\hat{\eta}$ y aquéllos de $\sigma^{-2}\text{VAR}(\hat{\eta})$ para 5 diferentes valores de θ .

θ	9.5	9.0	8.5	8.0	7.5
x_1	-0.2240	-0.25701	-0.302374	-0.36926	-0.478766
x_2	0.18960	0.224618	0.274142	0.349153	0.475227
$\hat{\eta}$	16.135	16.4155	16.85252	17.6063	19.1245
a	0.29354	0.341336	0.408147	0.50819	0.674580
$\sigma^{-2}\text{VAR}(\hat{\eta})$	0.44641	0.560246	1.104137	03.24800	12.8516

Nótese que el valor más grande de $\hat{\eta}$ en la tabla (3.2.3) es de 19.124547 obtenido en el punto (-.4787664, .4752275), que cae fuera de la región experimental. También note el rápido incremento en $\sigma^{-2}VAR(\hat{\eta})$ de .4464172 a 12.851665. Este resultado es causado por la magnitud de los elementos de $(X^tX)^{-1}$.

En el procedimiento modificado, el valor más grande de $\hat{\eta}$ es de 19.231169 en el punto (-1.2348435, -1.2148018), ver tabla (3.2.2), que se encuentra cercano a la periferia de la región experimental con una varianza de 1.4726584, la cual es muy pequeña en relación con la anterior.

CASO II:

En muchas ocasiones el diseño experimental sufre algunas modificaciones, es frecuente que se le agreguen puntos que son de interés o bien que el ajuste de una superficie de segundo orden se realice sin llevar a cabo un plan de diseño específico. El costo que se paga en estos casos es que la $VAR(\hat{\eta})$, por lo general crece mucho. Aquí se desarrolla un procedimiento con el cual se podrá controlar el término de $VAR(\hat{\eta})$ que contribuye más a que ésta sea grande.

Si $\zeta_1, \zeta_2, \dots, \zeta_p$ con $p = (k+1)(k+2)/2$, son los valores propios de X^tX y utilizando que $VAR(\hat{\eta}) = \sigma^2 \mathbf{z}^t (X^tX)^{-1} \mathbf{z}$, se puede probar que

$$\frac{\sigma^2 \mathbf{z}^t \mathbf{z}}{\max_i \zeta_i} < VAR(\hat{\eta}) < \frac{\sigma^2 \mathbf{z}^t \mathbf{z}}{\min_i \zeta_i}.$$

Esta doble desigualdad refleja la dificultad que se tiene para acotar la $VAR(\hat{\eta})$ cuando X^tX tiene uno o más valores propios pequeños.

Sea $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p]$ $\mathbf{v}_i \in \mathbb{R}^p$ $i = 1, \dots, p$, una matriz ortogonal de vectores propios de X^tX . Entonces

$$X^tX = V \text{diag}(\zeta_i) V^t,$$

implica que

$$(X^tX)^{-1} = V \text{diag} \left(\frac{1}{\zeta_i} \right) V^t.$$

Los valores propios de $(X^tX)^{-1}$ son $\frac{1}{\zeta_i}$, para toda $i = \overline{1, p}$, siempre que $\zeta_i \neq 0$.

$VAR(\hat{\eta})$

$$\begin{aligned}
&= \sigma^2 \mathbf{z}' V \text{diag} \left(\frac{1}{\zeta_i} \right) V' \mathbf{z} \\
&= \sigma^2 (\mathbf{z}' \mathbf{v}_1, \mathbf{z}' \mathbf{v}_2, \dots, \mathbf{z}' \mathbf{v}_p) \text{diag} \left(\frac{1}{\zeta_i} \right) (\mathbf{v}_1' \mathbf{z}, \mathbf{v}_2' \mathbf{z}, \dots, \mathbf{v}_p' \mathbf{z}) \\
&= \sigma^2 \left(\frac{\mathbf{z}' \mathbf{v}_1}{\zeta_1}, \frac{\mathbf{z}' \mathbf{v}_2}{\zeta_2}, \dots, \frac{\mathbf{z}' \mathbf{v}_p}{\zeta_p} \right) (\mathbf{z}' \mathbf{v}_1, \mathbf{z}' \mathbf{v}_2, \dots, \mathbf{z}' \mathbf{v}_p)' \\
&= \sigma^2 \frac{\sum_{i=1}^p (\mathbf{z}' \mathbf{v}_i)^2}{\zeta_i}.
\end{aligned}$$

Se tiene que

$$\sigma^2 \frac{\sum_{i=1}^p (\mathbf{z}' \mathbf{v}_i)^2}{\max_i \zeta_i} \leq \sigma^2 \frac{\sum_{i=1}^p (\mathbf{z}' \mathbf{v}_i)^2}{\zeta_i} \leq \sigma^2 \frac{\sum_{i=1}^p (\mathbf{z}' \mathbf{v}_i)^2}{\min_i \zeta_i}$$

pero

$$\begin{aligned}
&\sigma^2 \frac{\sum_{i=1}^p (\mathbf{z}' \mathbf{v}_i)^2}{\max_i \zeta_i} \\
&= \sigma^2 \left(\frac{\mathbf{z}' \mathbf{v}_1}{\max_i \zeta_i}, \frac{\mathbf{z}' \mathbf{v}_2}{\max_i \zeta_i}, \dots, \frac{\mathbf{z}' \mathbf{v}_p}{\max_i \zeta_i} \right) (\mathbf{z}' \mathbf{v}_1, \mathbf{z}' \mathbf{v}_2, \dots, \mathbf{z}' \mathbf{v}_p)' \\
&= \mathbf{z}' V \text{diag} \left(\frac{1}{\max \zeta_i} \right) V' \mathbf{z} \\
&= \frac{1}{\max_i \zeta_i} \mathbf{z}' V V' \mathbf{z} \\
&= \frac{\mathbf{z}' \mathbf{z}}{\max_i \zeta_i}.
\end{aligned}$$

Análogamente se tiene que

$$\sigma^2 \frac{\sum_{i=1}^p (\mathbf{z}' \mathbf{v}_i)^2}{\min_i \zeta_i} = \frac{\mathbf{z}' \mathbf{z}}{\min_i \zeta_i},$$

de donde se desprende la desigualdad que se había propuesto para la varianza de $\hat{\eta}$.

Las componentes de \mathbf{v}_i se denotan como

$$v_{0i}, v_{1i}, \dots, v_{ki}, v_{11i}, v_{22i}, \dots, v_{kk_i}, v_{12i}, v_{13i}, \dots, v_{k-1k_i}$$

para $i = \overline{1, p}$. Entonces

$$\mathbf{z}' \mathbf{v}_i = v_{0i} + \mathbf{x}' \boldsymbol{\tau}_i + \mathbf{x}' \mathbf{T}_i \mathbf{x},$$

con $i = \overline{1, p}$. Donde $\tau_i^t = [v_{1i}, v_{2i}, \dots, v_{ki}]$ y

$$T_i = \begin{bmatrix} v_{11i} & v_{12i}/2 & \cdots & v_{1ki}/2 \\ & v_{22i} & \cdots & v_{2ki}/2 \\ & & \cdots & \vdots \\ & & \cdots & v_{kki} \end{bmatrix},$$

que implica que

$$\text{VAR}(\hat{\eta}) = \frac{\sigma^2 \sum_{i=1}^p (v_{0i} + \mathbf{x}^t \tau_i + \mathbf{x}^t T_i \mathbf{x})^2}{\zeta_i}.$$

De esta última igualdad se observa que si se trata de optimizar $\hat{\eta}$ sujeta a curvas de nivel de varianza constante se tendrán dificultades al aplicar la técnica de multiplicadores de lagrange ya que aparecerán términos no lineales en las variables independientes.

Se adoptará entonces una solución un tanto comprometida, que recupera el espíritu de A.CO. y también considera un control sobre la varianza de $\hat{\eta}$.

Nótese, que valores pequeños de ζ_i contribuyen a producir varianzas grandes en la respuesta. Entonces se va a restringir la parte de $\text{VAR}(\hat{\eta})$ correspondiente al valor propio más pequeño de $X^t X$, denotado por ζ_{\min} . De esta manera la varianza de la respuesta se controlará en la región en la que se aplique el A.CO.

Sean v_{0m}, τ_m, T_m los valores de v_0, τ, T correspondientes al valor propio más pequeño, ζ_{\min} , respectivamente.

Se optimizará $\hat{\eta}$ sujeta a que

$$v_{0m} + \mathbf{x}^t \tau_m + \mathbf{x}^t T_m \mathbf{x} = r,$$

donde r es una constante positiva lo suficientemente pequeña como para anular el efecto del valor propio más pequeño.

Cuando T_m es indefinida, las curvas de nivel de la forma cuadrática (3.2.1) no son ni elipses ni esferas. De esta manera la restricción sólo controla a la varianza, pero no tendría que considerar \mathbf{x} en la región de diseño, entonces se añade la restricción adicional $\mathbf{x}^t \mathbf{x} \leq R^2$, donde R es lo suficientemente grande para contener al diseño experimental.

Utilizando el método de multiplicadores de lagrange

$$\mathcal{L} = \hat{\eta} - \mu(v_{0m} + \mathbf{x}^t \tau_m + \mathbf{x}^t T_m \mathbf{x} - r) - \lambda(\mathbf{x}^t \mathbf{x} - R^2),$$

donde r es algun valor deseable para $(v_{0m} + \mathbf{x}'\tau_m + \mathbf{x}'T_m\mathbf{x} - r)$.

Se obtiene $\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0$, entonces

$$(B - \mu T_m - \lambda I)\mathbf{x} = 1/2(\mu\tau_m - \mathbf{b}).$$

Lo que debe hacerse es escoger valores de μ y λ apropiados para asegurar que se obtenga una respuesta óptima restringida.

Para un valor fijo de μ la matriz de segundas derivadas parciales $2(B - \mu T_m - \lambda I)$ es definida negativa si λ es mayor que todas las raíces características de $(B - \mu T_m)$, las λ^s con esas propiedades proporcionarán un máximo de $\hat{\eta}$, es decir cada punto \mathbf{x} que se obtenga dará respuesta óptima para un nivel

$$r = v_{0m} + \mathbf{x}'\tau_m + \mathbf{x}'T_m\mathbf{x}$$

y por supuesto para un radio fijo $R = (\mathbf{x}'\mathbf{x})^{1/2}$.

Ejemplo (3.2.2): [26]

Se elaboró un experimento para una respuesta que depende de tres variables. El primer intento fue construir y usar un diseño compuesto central. Sin embargo, dificultades experimentales impidieron el uso del diseño como se había planeado inicialmente y el que finalmente se utilizó condujo a problemas de mal condicionamiento con respecto al ajuste de una superficie de segundo orden. Los datos se presentan en la tabla (3.2.4).

TABLA (3.2.4)			
x_1	x_2	x_3	Y
-1.020	-1.402	-0.998	13.5977
0.900	0.478	-0.818	12.7838
0.870	-1.282	0.882	16.2780
-0.950	0.458	0.972	14.1678
-0.930	-1.242	-0.868	9.2461
0.750	0.498	-0.618	17.0167
0.830	-1.092	0.732	13.4253
-0.950	0.378	0.832	16.0967
1.950	-0.462	0.002	14.5438
-2.150	-0.402	-0.038	20.9534
-0.550	0.058	-.518	11.0411
-0.450	1.378	0.182	21.2088
0.150	1.208	0.082	25.5514
0.100	1.768	-0.008	33.3793
1.450	-0.342	0.182	15.4341

En este ejemplo, el valor propio más pequeño de $X^t X$ es $\zeta_{\min} = 0.321$.

$$\begin{aligned} v_{0m} + x^t \tau_m + x^t T_m x \\ = -0.2935 + 0.0469x_1 + .48081x_2 + .4071x_3 \\ + (x_1, x_2, x_3) \begin{pmatrix} .1129 & .0095 & .2709 \\ .0095 & -.1382 & -.0148 \\ .2709 & -.0148 & .6453 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \end{aligned}$$

Puesto que T_m es indefinida (sus valores propios son .759, 0.0008, -.1398) se añade la restricción

$$x_1^2 + x_2^2 + x_3^2 \leq R^2.$$

Para tener una idea de que valores de r son razonables, se calculó $(\sigma^2[v_{0m} + x^t \tau_m x + x^t T_m x]^2) / \zeta_{\min}$ para cada uno de los 15 puntos del diseño. (Esta es la porción de la estimación de $\text{VAR}(\hat{\eta})$ correspondiente a $\zeta_{\min} = 0.0321$). El valor más grande que se encontró fue 0.66, correspondiente a un valor de 0.087 para $v_{0m} + x^t \tau_m x + x^t T_m x$.

Con esta información r no debería ser mucho más grande que 0.09 y R no excedera 2.0 esta cota para R fue obtenida al evaluar la ecuación $x_1^2 + x_2^2 + x_3^2 = R^2$ sobre cada punto del diseño experimental. El valor .09 sirve como una guía para lo que podrían considerarse valores deseables de r .

El método de A.CO. sin modificación se aplicó. La tabla (2.3.5) proporciona valores de $\hat{\eta}$ para diferentes x que son soluciones de la ecuación

$$(B - \lambda I)x = -\frac{b}{2},$$

TABLA (2.3.5)

R	2.0	1.9749	1.944	1.744	1.4836	1.3945
$\hat{\eta}$	57.176	56.272	55.147	48.332	40.197	37.640
$\text{VAR}(\hat{\eta})$	405.44	385.73	360.783	231.696	118.03	91.108
$\sigma^{-2}\text{VAR}(\hat{\eta})$	145.90	138.668	129.834	83.38	42.475	32.787
x_1	1.265	1.249	1.227	1.085	.899	.835
x_2	0.936	0.9272	0.916	0.845	0.749	0.715
x_3	1.232	1.2168	1.197	1.074	0.912	.857
r	2.079	2.025	1.958	1.5523	1.0845	.941
r^2 / ζ_{\min}	134.73	127.815	119.371	75.163	36.641	27.631

R	1.146	0.9378	0.698	0.3785	0.1396
$\hat{\eta}$	31.085	26.229	21.365	16.021	12.796
$\text{VAR}(\hat{\eta})$	39.750	17.057	5.846	4.320	7.202
$\sigma^{-2}\text{VAR}(\hat{\eta})$	14.305	6.138	2.104	1.554	2.592
x_1	0.6598	0.5146	0.3518	0.1519	0.0371
x_2	0.6179	0.5308	0.4223	0.2551	0.1027
x_3	0.7048	0.5770	0.4305	0.2347	0.0869
r	0.5866	0.3372	0.1041	0.1237	0.2406
r^2/ζ_{\min}	10.71	3.543	0.337	0.477	1.804

Aquí, por supuesto, los valores de λ fueron escogidos mayores que el valor propio más grande de B para maximizar $\hat{\eta}$ sobre una esfera de radio R . Para ilustrar la precisión de $\hat{\eta}$ en lo que ordinariamente se podrían considerar condiciones de operación recomendables, también se proporcionan valores estimados de $\sigma^{-2}\text{VAR}(\hat{\eta})$ en cada punto \mathbf{x} .

Posteriormente se encontraron las condiciones que maximizan $\hat{\eta}$ sujeta a las restricciones del A.CO. modificado. Se tuvo especial atención en $R \leq 2.0$ y $r \leq 0.09$, aunque con el objeto de generar más resultados, se consideraron valores de r mayores, pero no mucho mayores que 0.09. Los resultados se presentan en la tabla (2.3.6).

En dicha tabla aparece la varianza estimada de $\hat{\eta}$ para hacer posible una comparación con el análisis de cordilleras usual.

R	2.098	2.000	1.9345	1.9039	1.85	1.795
$\hat{\eta}$	49.14	46.26	44.11	42.83	40.22	37.042
$\text{VAR}(\hat{\eta})$	18.34	14.828	11.607	9.754	6.474	3.627
$\sigma^{-2}\text{VAR}(\hat{\eta})$	6.602	5.336	4.177	3.510	2.330	1.305
x_1	1.060	0.965	0.820	0.750	0.6011	0.405
x_2	1.810	1.7515	1.752	1.750	1.751	1.752
x_3	-0.04	-0.0297	0.015	0.012	0.000	0.000
r	0.183	0.1650	0.1457	0.1259	0.086	0.047

R	1.623	1.53	1.162	0.848
$\hat{\eta}$	37.19	35.242	27.78	22.42
$\text{VAR}(\hat{\eta})$	8.745	8.12	4.542	3.2567
$\sigma^{-2}\text{VAR}(\hat{\eta})$	3.147	2.922	1.635	1.17
x_1	0.7845	0.7734	0.563	0.410
x_2	1.421	1.320	1.015	0.7367
x_3	0.011	0.0186	0.0632	0.0972
r	0.139	0.1357	0.0744	0.0062

Comparando entre las dos tablas se puede observar que la restricción extra sobre la porción de mayor varianza, mejora la precisión de $\hat{\eta}$ en el óptimo estimado.

Para un radio específico, el análisis de coordilleras sin modificación proporciona una respuesta $\hat{\eta}$ mayor que la obtenida con el método modificado, sin embargo, al comparar la r y la $\text{VAR}(\hat{\eta})$ de ambos procedimientos se observa una gran mejoría del segundo método con respecto al primero.

Aunque no es fácil decidir entre una respuesta mayor contra una varianza pequeña se aprecia una superioridad de las coordenadas proporcionadas por el método modificado que aparecen en la tabla (2.3.6).

Por ejemplo, difícilmente se podría esgoger cualquier condición de operación en la tabla (2.3.5) que indique $\hat{\eta} \geq 50$ debido a las varianzas tan grandes que presenta la predicción. Por otro lado la tabla (2.3.6) revela que en un radio aproximadamente igual a 2, se obtiene $\hat{\eta} = 49.143$ y $\sigma^{-2}\text{VAR}(\hat{\eta}) = 6.602$.

Es interesante notar cuál es el impacto de la componente r de la $\text{VAR}(\hat{\eta})$ en el A. CO. El último renglón de la tabla (2.3.5) proporciona $r^2/\zeta_{\min} = r^2/0.0321$, el cual debería ser comparado con el renglón etiquetado con $\text{VAR}(\hat{\eta})/\sigma^2$. Se puede ver en este ejemplo que r^2/η_{\min} determina un mayor valor de $\text{VAR}(\hat{\zeta})/\sigma^2$.

El minimizar el valor propio más pequeño de X^tX ayuda a prevenir el problema de multicolinealidad en Regresión ([31], pág. 189). $\text{VAR}(\hat{\eta})$ será grande cuando ζ_{\min} es cercano a cero, pero si alguno de los valores propios de (X^tX) es cercano a cero existen dificultades para determinar la inversa de esta matriz y la de la matriz X . Esto se debe a que alguna de las columnas de X es casi un múltiplo escalar de la otra. Algunas veces, aparece una matriz de diseño con estas características, si es que este no es planeado. Precisamente el objeto de esta sección es abordar el A. CO. cuando el

diseño presenta dificultades como las que aparecen en multicolinealidad.

COMENTARIOS FINALES

Este trabajo se ubica en un área de reciente explotación. Se trata de la metodología aplicada en el análisis de superficies de respuesta. Existe una gran cantidad de contribuciones sobre el tema, tanto, que en dos periodos distintos, diferentes autores se han dado a la tarea de recopilar toda la información existente. La primera relación detallada de los trabajos escritos sobre el tratamiento de superficies de respuesta fue publicada por Hill y Hunter en (1966) y una segunda revisión bibliográfica es dada en (1988) por Myers, Khuri y Carter.

Se observa que tales contribuciones, en su inmensa mayoría, proponen métodos aislados que dan solución a las partes más importantes que integran el análisis total de superficies de respuesta. En consecuencia, no se prevé la aparición de pequeños problemas que impiden una secuencia lógica en el desarrollo integral del análisis. Obteniendo como resultado, una metodología particularizable y en muchos de los casos limitada más que una de aplicación general y algorítmica. Un ejemplo de esto último, es la elección de la longitud del paso de la ruta, t , en el método de ascenso por pendiente máxima. Hasta 1990 no se contaba con un criterio que permitiera definirlo, por ende su elección quedaba al juicio y subjetividad del investigador. Tal situación, planteaba una problemática en la aplicación algorítmica de la metodología que se tenía hasta el momento.

El año de (1991), culmina con lo que se podría denominar la *Implementación Algorítmica de la Metodología para el Análisis de Superficies de Respuesta*, esto, debido a los resultados obtenidos por Pérez en el mismo año. Dicho autor obtiene, [39], numerosos resultados de aplicabilidad a la metodología ya existente, entre los cuales se resaltan de manera especial: un nuevo criterio para elegir la longitud del paso, t , en la ruta de ascenso por pendiente máxima, una nueva manera de encontrar el centro de los subsecuentes diseños utilizados en el desarrollo secuencial del método de ascenso por pendiente máxima, y la determinación de un criterio para finalizar la búsqueda del óptimo; dicho criterio está relacionado al hecho de que el óptimo pertenezca o no al conjunto de extrapolación S descrito en la sección (2.5); esto equivale a pedir que el óptimo se encuentre cercano al diseño. La puesta en marcha de los resultados anteriores, hace posible el desarrollo de programas computacionales que permiten realizar el análisis completo de superficies de respuesta y aplicarlo a problemas reales. Así, la experiencia y juicios del experimentador quedan circunscritas a las partes menos

técnicas del análisis.

Este trabajo ha logrado reunir los nuevos avances obtenidos para el análisis de superficies de respuesta con la metodología ya existente, dando así la posibilidad de tener en un sólo texto todos los elementos necesarios para el desarrollo algorítmico del análisis.

Punto importante, ha sido la presentación de las diferentes partes que conforman la metodología, buscando ese sutil equilibrio entre lo teórico y lo práctico. Lo anterior, reflejado en la gran cantidad de resultados presentados y en sus demostraciones, logrando que estas últimas sean lo más accesible a personas no versadas en el tema; asimismo los numerosos ejemplos de interés práctico, planteados con el objeto de esclarecer cada parte del análisis.

En resumen, este trabajo es base para comprender las direcciones a seguir en futuros trabajos de investigación acerca de la metodología de superficies de respuesta, tales como:

- 1.- La optimización simultánea o el problema de multirespuesta. En esta área los estudios apuntan sobre todo a la búsqueda de diseños adecuados, ya que un diseño eficiente para una respuesta puede no ser eficiente para otras. La experimentación secuencial propia del problema de superficie de respuesta resulta un reto en el caso de multirespuestas: la decisión para hacer un cambio de diseño experimental no sólo depende de una sino de varias respuestas.
- 2.- La optimización de una respuesta sujeta a restricciones de tipo lineal. Se ha observado que en mucho de los casos, este problema puede ser reducido a uno de complementariedad lineal. Por lo cual, es necesario el desarrollo de algoritmos eficientes que permitan la obtención de la solución básica factible complementaria que resuelve el problema, o bien, desarrollar métodos que obtengan los óptimos con restricciones a partir de los óptimos no restringidos, los cuales se calculan con la metodología expuesta en este trabajo.
- 3.- La optimización de respuestas que dependen de variables explicativas de tipo categórico o numerable. Si este es el caso, no es posible dar una expansión en series de Taylor para la respuesta y por ende modelos polinomiales no necesariamente serían buenas aproximaciones para la misma. De aquí la necesidad de crear nueva teoría al respecto.

APENDICE A.

Proposición 1 Sean \mathbf{x} y \mathbf{b} vectores de dimensión k , B una matriz simétrica de $k \times k$ y b_0 una constante. Los siguientes resultados se verifican

- 1) $\frac{\partial b_0}{\partial \mathbf{x}} = 0.$
- 2) $\frac{\partial \mathbf{x}^t \mathbf{b}}{\partial \mathbf{x}} = \mathbf{b}.$
- 3) $\frac{\partial \mathbf{x}^t B \mathbf{x}}{\partial \mathbf{x}} = 2B\mathbf{x}.$

Demostración:

$$\begin{aligned}
 1) \quad \frac{\partial b_0}{\partial x_i} &= 0 \quad \forall i = \overline{1, k}, \text{ entonces } \frac{\partial b_0}{\partial \mathbf{x}} = 0. \\
 2) \quad \frac{\partial (\mathbf{x}^t \mathbf{b})}{\partial x_i} &= \frac{\sum_{i=1}^k x_i b_i}{\partial x_i} = b_i, \text{ entonces, } \frac{\partial (\mathbf{x}^t \mathbf{b})}{\partial \mathbf{x}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \mathbf{b}. \\
 3) \quad \frac{\partial (\mathbf{x}^t B \mathbf{x})}{\partial x_m} &= \frac{\partial (\sum_{i=1}^k \sum_{j=1}^k b_{ij} x_i x_j)}{\partial x_m} \\
 &= \frac{\partial (\sum_{j=1}^k b_{mj} x_m x_j + \sum_{i \neq m}^k a_{ij} x_i x_j)}{\partial x_m} \\
 &= 2b_{mm} x_m + \sum_{j \neq m}^k b_{mj} x_j + \sum_{i \neq m}^k b_{im} x_i \\
 &= 2b_{mm} x_m + 2 \sum_{i \neq m}^k b_{mi} x_i \text{ por simetría de } B \\
 &= 2 \sum_{i=1}^k b_{mi} x_i
 \end{aligned}$$

el cual es el término m -ésimo de $2B\mathbf{x}$, por lo tanto

$$\frac{\partial (\mathbf{x}^t B \mathbf{x})}{\partial \mathbf{x}} = 2B\mathbf{x}.$$

□

Proposición 2 Sea z un vector de dimensión k , A una matriz de $q \times k$ y B una matriz de $1 \times m$. Los siguientes resultados se verifican

a) $E[AzB] = AE[z]B.$

b) $V[z] = E[zz^t] - E[z]E[z^t].$

c) $V[Az] = AV[z]A^t.$

La notación E se refiere a la esperanza y la V , se refiere a la matriz de varianzas y covarianzas.

Demostración:

La demostración de los resultados es muy simple y puede verse en ([43], págs 5-15; [28]).

APENDICE B.

TEOREMA 1 (DESCOMPOSICION ESPECTRAL) *Cualquier matriz simétrica $B_{k \times k}$ puede escribirse como*

$$B = U \Lambda U^t = \sum_{i=1}^k \delta_i U_{(i)} U_{(i)}^t,$$

donde $\Lambda = \text{diag}(\delta_1, \delta_2, \dots, \delta_k)$; $\delta_1, \dots, \delta_k$ son los valores propios de la matriz B y U es una matriz ortogonal cuyas columnas son los vectores propios de B .

Demostración:

Primero se detallará como calcular los valores propios de la matriz, con los procedimientos conocidos de álgebra lineal. Por definición, se tiene que $U_{(i)}$ es el vector propio asociado al valor propio δ_i de la matriz B , si $BU_{(i)} = \delta_i U_{(i)}$. La manera de calcular los vectores y valores propios, es resolver el polinomio característico en δ , $P(\delta)$, que resulta de calcular el determinante de la matriz $(B - \delta I_{k \times k})$ e igualar a cero. Dado que B es una matriz simétrica, del álgebra lineal se sabe que todos los valores propios son reales. Ahora para cada valor propio δ_i se busca el $U_{(i)}$ que resuelva la ecuación característica $(B - \delta_i I)U_{(i)} = 0$, y se elige $U_{(i)}$ de norma unitaria y así para cada $i = 1, \dots, k$. Además se pide que $U_{(i)}^t \cdot U_{(j)} = 0$, para toda $i \neq j$, $i, j = 1, \dots, k$, lo cual se logra mediante el proceso de Gram-Schmit. Una vez hecho esto, se definen las matrices de valores y vectores propios como

$$\Lambda = \text{diag}(\delta_1, \dots, \delta_k) \text{ y } U = [U_{(1)}, U_{(2)}, \dots, U_{(k)}].$$

Entonces

$$\begin{aligned} BU &= [BU_{(1)}, BU_{(2)}, \dots, BU_{(k)}] \\ &= [\delta_1 U_{(1)}, \delta_2 U_{(2)}, \dots, \delta_k U_{(k)}], \end{aligned}$$

por definición. Ahora

$$\begin{aligned} U^t B U &= \begin{bmatrix} U_{(1)}^t \\ U_{(2)}^t \\ \vdots \\ U_{(k)}^t \end{bmatrix} [\delta_1 U_{(1)}, \delta_2 U_{(2)}, \dots, \delta_k U_{(k)}], \\ &= \text{diag}(\delta_1, \delta_2, \dots, \delta_k), \end{aligned}$$

dado que $U_{(i)}^t \cdot U_{(j)} = 0$ para toda $i \neq j$ y $U_{(i)}$ es un vector de norma unitaria, por lo cual $U_{(1)}^t U_{(1)} = 1$.

Luego entonces $U^t B U = \Lambda$, que implica

$$B = U \Lambda U^t \text{ con } U^t U = I.$$

□

Referencias

- [1] Anderson V.L., R.A Mclean (1974). *Design of experiments: A realistic approach*. New York: Marcel Dekker.
- [2] Atkinson A.C. y A.N. Donev (1989). *The construction of exact D-optimum experimental design with application to blocking response surface design*. *Biometrika*, **76**, No 3, 515-526.
- [3] Box G.E.P., K.G. Wilson (1951). *On the experimental attainment of optimum conditions*. *J.R. Statist. Soc.*, **B13**, 1-45.
- [4] Box G.E.P., J.S. Hunter (1954). *A confidence region for the solution of a set of simultaneous equations with application to experimental design*. *Biometrika*, **41**, 190- 199.
- [5] Box G.E.P., J.S. Hunter (1957). *Multi-Factor experimental design for exploring response surfaces*. *Ann. Math. Statist.*, **28**, 195-241.
- [6] Box, G.E.P., N.R. Drapper (1959). *A basis for the selection of response surface design*. *Journal of American Statistical Association*, **54**, 622-654.
- [7] Box G.E.P., W.G. Hunter y J.S. Hunter (1978). *Statistics for experiments*. New York: John Wiley.
- [8] Brooks S.H., M.R. Mickey (1961). *Optimum estimation of gradient direction in steepest ascent experiments*. *Biometrics*, **17**, 49-56.
- [9] Buehler R.J, B.V. Shah y O.Kemphorne (1964). *Methods of parallel tangents*. *Chemical Engineering Progress Symposium series*, **60**, 1-7.

- [10] Carter Jr. W.H., V.M. Chinchilli, E.D. Campbell, G.L. Wampler (1984). *Confidence interval about the response at the stationary point of a response surface, with an application to preclinical cancer therapy*. Biometrics, **40**, 1125-1130.
- [11] Carter Jr. W.H., V.M. Chinchilli, R.H. Myers, E.D. Campbell (1986). *Confidence interval and an improved ridge analysis of response surfaces*. Technometrics, **28**, No 4, 339-346.
- [12] Cochran W.G., G.M. Cox (1980). *Diseños Experimentales*. Editorial Trillas, México.
- [13] Davies O.L. (1950). *The construction and uses of fractional factorial designs in industrial research*. Biometrics, **6**, 233-249.
- [14] Davies O.L. (1956). *The Design and Analysis of industrial Experiments*. New York: Hafner Press.
- [15] Drapper N.R. (1963). *Ridge analysis of response surfaces*. Technometrics, **5**, No 4, 469-479.
- [16] Drapper N.R., H. Smith (1981). *Applied regression analysis*. 2nda Ed., New York: John Wiley.
- [17] Fieller E.C. (1944). *A fundamental formula in the statistics of biological assay, and some applications*. Quarterly Journal of Pharmacology, **XVII**, No 2, 117-123.
- [18] Frankel S.A (1961). *Statistical design of experiments for process development of MBT*. Rubber Age, **89**, 453.
- [19] Friedberg S.H., A.J. Insel y L.E. Spence (1982). *Algebra lineal*. Publicaciones Cultural S.A., 1a ed.
- [20] Harkins, A. (1964). *The use of parallel tangents in optimization*. Chemical Engineering Progress Symposium series, No 50, **60**, 35-40.
- [21] Hicks C.R. (1982). *Fundamental concepts of design of experiments*. 3era Ed., New York: Holt Rinehart and Winston.
- [22] Hill W.J., W.G. Hunter (1966). *A review of response surface methodology: A Literature Survey*. Technometrics, **8**, 571-589.

- [23] Hoerl A.E. (1964). *Ridge analysis*. Chemical Ingeneering Progress Symposium Series, **60**, No 50, 67-77.
- [24] Huerta G.J.G, A.L.M. Hernandez (1991). *Optimización: Elementos de teoría en superficie de respuesta*. Tesis de licenciatura en Actuaría, Fac. Ciencias, U.N.A.M.
- [25] Karson M.J, A.R. y R.J. Harder (1969). *Minimum bias estimation and experimental design for response surface*. Technometrics, **11**, 461-475.
- [26] Khuri A.I., R.H. Myers (1979). *Modified ridge analysis*. Technometrics, **4**, 469-579.
- [27] Khuri A.I., J.A. Cornell (1987). *Response Surfaces: (Designs and Analyses)*. Marcel Dekker, Inc.
- [28] Mardia K.V, J.T. Kent, J.M. Bibby (1979). *Multivariate analysis*. Academic Press.
- [29] Marsden J.E., A.J. Tromba (1981). *Cálculo vectorial*. Fondo Educativo Interamericano S.A.
- [30] Méndez I. (1977). *Metodología de superficie de respuesta*. Comunicaciones Técnicas, I.I.M.A.S., U.N.A.M.
- [31] Montgomery D.C, E.A. Peck (1992). *Introduction to linear regression Analysis*. John Wiley and Sons, 2da edición.
- [32] Mood A. M, F.A Graybill, D. Boes (1974). *Introduccion to the theory of statistics*. Mc Graw Hill.
- [33] Myers R.H. (1971). *Response Surface Methodology*. Boston: Allyn and Bacon Inc.
- [34] Myers R.H., A.I. Khuri (1979). *A new procedure for steepest ascent*. Commun. Statist. Theor. Meth., **A8**, 1359-1376.
- [35] Myers R.H., A.I.Khuri, Jr. W.H. Carter (1989). *Response surface methodology: 1966-1988*. Technometrics, **31**, No 2, 137-157.
- [36] Noble B. (1969). *Applied linear algebra*. Prentice Hall, Inc.

- [37] Notz W. (1988). *Minimal point second order Design*. Journal of Statistical Planning and Inference, **6**, 47-58.
- [38] O'reilly F.J. (1975). *On a criterion for extrapolation in Normal Regression*. Ann. Statist., **3**, No 1, 219-222.
- [39] Pérez S.B.R (1991). *Regiones de predictibilidad en superficies de respuesta*. Tesis Doctoral de Matemáticas, Fac. Ciencias, U.N.A.M.
- [40] Pérez S.B.R, F.J. O'reilly (1993). *Sobre el ascenso por pendiente máxima en superficies de respuesta*. Reportes de Investigación, I.I.M.A.S., U.N.A.M, **3**, No 26.
- [41] SAS Institute (1982). *SAS user guide: Statistics, Cary, NC, SAS Institute*.
- [42] Searle S.R. (1971). *Linear Models*. John Wiley and Sons, New York.
- [43] Seber G.A.F. (1984). *Multivariate Observations*. John Wiley and Sons.
- [44] Umland A.W., W.W. Smith (1959). *The use of lagrange multipliers with response surfaces*. Technometrics, **1**, 289.