UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS

DEPARTAMENTO DE ESTADISTICA

DISEÑO OPTIMO EN MODELOS LINEALES

Rafael Molina Soriano

Universidad de Granada

Facultad de Ciencias

Sección de Matemáticas

DISEÑO OPTIMO EN MODELOS LINEALES

30 de Diciembre de 1.983

DISEÑO OPTIMO EN MODELOS LINEALES

Memoria que para optar al grado de Doctor en Ciencias, Sección de Matemáticas, presenta el Licenciado Rafael Molina Soriano.

Directora de Tesis:
Profesora Dra. Dª Concepción Fernández Vivas

Vº Bº

Universidad de Granada . Facultad de Ciencias.

Departamento de Estadística.

" DISEÑO OPTIMO EN MODELOS LINEALES "

Realizado el acto público de la Defensa y Mantenimiento de esta Tesis Doctoral el día 29 de Marzo de 1.984, en la Universidad de Granada, ante - el tribunal formado por:

Presidente: Dr. Don Rafael Infante Macías

Vocales : Dr. Don Antonio Pascual Acosta

Dr. Don Ramón Gutiérrez Jaímez

Dra. Dª Concepción Fernández Vivas

Secretario: Dr. Don Eduardo Ramos Méndez

obtuvo la calificación de

SOBRESALIENTE CUM LAUDE

A la memoria de mi padre

y D. Ramiro Melendreras.

INDICE

INTRODUCCION

Capítulo 1

CRITERIOS DE OPTIMALIDAD

Capítulo 2

DISEÑO OPTIMO Y TEORIA DE JUEGOS

Capítulo 3

DISEÑO OPTIMO Y DUALIDAD

Capítulo 4

JUSTIFICACION DE HIPOTESIS EN DISEÑO OPTIMO

Apéndice 1

TEORIA DE LA APROXIMACION

INTRODUCCION

ORIGEN Y EVOLUCION DEL PROBLEMA.

Antes de comentar los resultados obtenidos resumiremos brevemente el origen y posterior desarrollo de la teoría de diseño óptimo en modelos lineales.

El problema con el que está relacionada la presente memoria se inicia con un artículo escrito por Smith en 1.918. La solución de dicho problema pareció no interesar a nadie, si exceptuamos las publicaciones de Wald(1.943) y Mood(1.946), hasta el comienzo de los años cincuenta. En esa década se realizan algunos trabajos sobre el tema, entre los que podemos citar los resultados de Elfving(1.952,1.955, 1.959), Chernoff(1.953,1.959), de la Garza(1.954,1.956), Ehrenfeld(1.955), Kempthorne(1.956), Guest(1.958) y Kiefer(1.958,1.959), a cuyo nombre están especialmente unidos los problemas de diseño óptimo. En la década siguiente, animados probablemente por los artículos de Kiefer(1.960,1.961,1.962a,1.962b) y las publicaciones conjuntas de Kiefer y Wolfowitz(1.960,1.964a,1.964b), que prueban la homogeneidad entre los criterios de optimalidad más conocidos por aquella época, numerosos autores dedican su trabajo al análisis de la optimalidad de un diseño; entre los resultados destacaremos los de Karlin y Studden (1.966) y Atwood(1.969) que suponen la utilización de la teoría de juegos como nueva herramienta matemática para la caracterización de los diseños óptimos. En la misma década Shah(1.960) propone un nuevo criterio de optimalidad. Al comienzo de los años setenta se formula rigurosamente el problema matemático, gracias principalmente a Fedorov y Malyutov(1.972), Whittle(1.973) y los artículos de Sibson(1.972, 1.974) y Silvey y Titterington(1.973,1.974), que suponen la utilización de ciertos aspectos de la teoría de la dualidad para resolver los problemas de diseño óptimo.

Por otra parte, al iniciarse la década de los cincuenta, G.E.-P. Box enfoca el problema no tanto desde el punto de vista del análisis de la teoría general sobre diseño óptimo, de difícil aplicación a modelos reales, sino pretendiendo más bien la solución de problemas -concretos, principalmente químicos. Aparecen entre otros los resultados de Box-Wilson(1.951), Box y Lucas(1.959), Box y Draper(1.959), -- Box y Hunter(1.965a,1.965b), Hill y Hunter(1.974), Hunter,Hill y Henson(1.969). Sin embargo, estas dos orientaciones que hacían dividir -las publicaciones en exclusivamente teóricas o exclusivamente aplicadas, fueron aunadas a comienzos de los setenta gracias principalmente a Wynn(1.970,1.972), quien elaboró métodos para lo obtención de soluciones óptimas en problemas generales. Un estudio más completo, recogiendo ambas orientaciones, se desarrollaba en la U.R.S.S., asociado-particularmente a los trabajos de Fedorov recogidos en un libro publicado en 1.972.

Casi mediada la década de los setenta, algunos autores dirigen su esfuerzo a la elaboración de nuevas formas de optimalidad para un-diseño y así surgen los criterios de optimalidad de Eccleston y Hedayat(1.974), Kiefer(1.974,1.975), el último de los cuales aún suscita-controversias como lo demuestra el artículo de Sinha y Mukerjee(1.982) Cheng(1.978), Magda(1.979) y Pukelsheim(1.980); mientras que otros se dirigen a la elaboración de algoritmos para la construcción de diseños óptimos como Wu y Wynn(1.978), Silvey, Titterington y Torsney(---1.978) y Silvey(1.980).

En 1.978, Ash y Hedayat publican una lista bibliográfica compuesta por 312 artículos relacionados con los problemas de diseño óptimo y que supone la mayor recopilación de material sobre este tema -hasta dicha fecha. En 1.980 ve la luz una monografía escrita por Silvey, en la que se recogen algunos avances en la teoría de diseño óptimo realizados desde la publicación del libro de Fedorov. Más reciente mente, Atkinson(1.982) ha realizado una síntesis sobre los desarrollos más importantes de los diseños de experimentos desde 1.975 a 1.980, -- analizando entre otros desarrollos los problemas de diseño óptimo.

En los últimos años, creemos apreciar una tendencia a aplicar- los resultados generales sobre diseño óptimo en áreas concretas de la- Estadística, especialmente al problema de diseño de pesadas en modelos químicos (ver Jacroux y otros 1.983) y al análisis de la optimalidad - de los diseños en bloques (para una reciente bibliografía ver el apartado 1.3.1. y la introducción de la sección 1.6. en la presente memoria).

Por último, destacaremos los resultados de Pulkesheim(1.980,--- 1.983) que sugieren la utilización de la teoría de la aproximaxión para la caracterización de la optimalidad de los diseños en bloques, optimalidad que hasta ahora sólo había sido abordada a través de los diseños exactos.

ORGANIZACION DE LA MEMORIA Y RESUMEN DE RESULTADOS.

De la anterior exposición sobre la historia del problema, obviamente se deduce que podíamos haber utilizado diversos enfoques para la presente memoria. Podíamos haber estudiado alguna parcela concreta en- la que se aplicase le teoría de los diseños óptimos o bien podíamos haber analizado un determinado criterio de optimalidad en una determinada situación.

Sin embargo, pensamos que sería más interesante abordar el problema de una forma global y suministrar a los interesados en el tema - un material lo suficientemente amplio como para que pudiesen comprender cualquiera de los diferentes aspectos de la teoría de diseño óptimo.

Con esta idea planteamos el primer capítulo. En él hemos pretendido describir los principales criterios de optimalidad con sus propie dades más relevantes. Como eran diversas las ópticas bajo las cuales - podíamos estudiar estos criterios, decidimos seguir el artículo de Hedayat(1.981) que proporciona un buen compendio de los criterios de optimalidad y sus principales propiedades. Sin embargo, creímos necesario además, introducir algunas precisiones sobre los diseños en bloques, teniendo en cuenta las propiedades que poseen las matrices de información que aparecen en este tipo de problemas; así elaboramos el apartado 1.3.1. y la primera parte de la sección 1.6. que nos sirven de introducción a los criterios que allí se enuncian y al mismo tiempo nos ayudan al análisis de los diseños óptimos en bloques.

En los dos capítulos siguientes, hemos estudiado los principales criterios de optimalidad y especialmente su caracterización con la ayuda de ciertos aspectos de la teoría de juegos y la teoría de la dualidad.

El capítulo segundo comienza describiendo brevemente los resultados sobre teoría de juegos que utilizamos para caracterizar los diseños óptimos. Nuestras aportaciones a este enfoque del problema son las siguientes:

- En la cuarta sección, un nuevo estudio de la D_s-optimalidad que debilita el material previo necesario para la caracterización de dicho criterio.

- En la sección quinta, obtenemos los diseños D_A-óptimos como las estrategias óptimas para un jugador en un juego bipersonal; - por otra parte, como este juego está basado en una transformación sobre el espacio de control y son varias las transformaciones que podemos realizar, probamos además la equivalencia de todas ellas.

En el capítulo tercero se utilizan algunos aspectos de la teoría de la dualidad para caracterizar los diseños óptimos con respecto a ciertos criterios de optimalidad. Resumimos los resultados más importantes ya conocidos y realizamos las siguientes aportaciones:

- En la sección segunda, obtenemos una demostración alternativa, más sencilla, para la cota sobre la D-optimalidad de un diseño propuesta por Atwood, probamos que dicha cota mejora la propuesta - por Kiefer y demostramos además que en general no es mejorable.

- En la primera parte de la tercera sección construímos diversos ejemplos para describir las diferentes formas que pueden tener las matrices asociadas a la D_s-optimalidad de un diseño.- En la segunda parte, proponemos dos demostraciones diferentes para los resultados obtenidos por Atwood sobre la D_s-optimalidad de un - diseño, estas dos demostraciones están basadas en las caracterizaciones de la D-optimalidad y D_s-optimalidad a través de la teoría - de la dualidad.

El capítulo cuarto no tiene bibliografía previa por ser completamente original. Lo que en él desarrollamos es la demostración de que la hipótesis de compacidad exigida al espacio de control no es debilitable si queremos garantizar la existencia de diseños óptimos con respecto a los criterios de optimalidad enunciados en el capítulo primero. Además demostramos que aquellos criterios generales que exigen la optimalidad de un diseño con respecto a más de un criterio, como por ejemplo la optimalidad universal, ni aún en el caso en que el espacio de control sea compacto podemos asegurar la existencia de diseños óptimos. Por último probamos, en el ejemplo correspondiente a la optimalidad universal, que la optimalidad de un diseño como distribución de probabilidad puede no ser alcanzada por un diseño basado en un número finito de observaciones.

La memoría termina con un apéndice sobre la teoría de la aproximación, describiendo sus resultados más importantes y redefiniendo los criterios de optimalidad a través de dicha teoría.

CAPITULO 1

CRITERIOS DE OPTIMALIDAD

1.0. RESUMEN.

En este capítulo, se analizan los criterios de optimalidad más utilizados para la elección de un diseño bajo el cual realizar un experimento. El capítulo comienza describiendo el problema para pasar a estudiar y relacionar, una vez justificados, los siguientes criterios de optimalidad: G-optimalidad, D-optimalidad, D_s-optimalidad, D_A-optimalidad, E-optimalidad, S-optimalidad, (M,S)-optimalidad, ϕ_p-optimalidad, optimalidad universal, criterios de tipo I y II y Schur-optimalidad.

1.1. EL PROBLEMA

1.1.1. PLANTEAMIENTO.

Supongamos que disponemos de un conjunto de condiciones, por nosotros controlables, bajo las cuales podemos realizar un experimento. Estamos interesados, mediante la realización de N observaciones, en clarificar, lo mejor posible, la relación existente entre las condiciones del experimento que controlamos y el valor medio del resultado experimental. Llamaremos al conjunto U, formado por las posibles condiciones controlables, el espacio de control, el dominio o el dominio de interés.

Así pues, estamos suponiendo que una variable aleatoria Y, que llamaremos variable de respuesta, está relacionada con una variable de control z por la expresión

$$E(Y_z) = f(z)$$

o bien

$$Y_z = f(z) + e_z$$

con $E(e_z) = 0$. Es decir, el valor observado de Y cuando el experimento se realiza bajo las condiciones que especifica z es una variable aleatoria Y_z cuya esperanza está funcionalmente relacionada con z. La función f puede ser llamada la superficie de respuesta, la función de respuesta o la función de regresión.

Normalmente se supone que f es parcialmente conocida; es decir $f(z) = f(z;\theta)$ en donde θ es un vector paramétrico desconocido cuya especificación determinaría completamente a f.

Nuestro problema podría considerarse entonces como la elección de condiciones bajo las cuales realizar el experimento, de forma que obtengamos máxima información, en algún sentido a precisar, sobre la función f.

EJEMPLO 1.1.1.1. Supongamos que un médico está interesado en conocer la respuesta de una determinada enfermedad a cierto tipo de compuesto. Para conocer dicha respuesta, el médico puede inyectar dicho compuesto con una concentración en agua tipificada entre cero y uno, de forma que si la concentración es u, entonces observa una variable aleatoria $N(\theta_0 + \theta_1 u, \sigma^2)$ con θ_0 y θ_1 desconocidos.

El problema puede plantearse en los siguientes términos: ¿qué concentraciones administrará a los enfermos- ¿qué elecciones de N puntos hará en el intervalo $[0,1]$? - de forma que obtenga máxima información, en algún sentido a precisar, sobre los paramétros desconocidos?.

EJEMPLO 1.1.1.2. Supongamos que estamos interesados en combinar y separar el efecto de dos tipos de abonos, T_1 y T_2, en el rendimiento de ciertas tierras de cultivo. Se supone que cualquier combinación de los abonos sólo afecta al rendimiento medio. Notemos por θ_1 el efecto del abono T_1, por θ_2 el efecto del abono T_2 y por $\theta_1 + \theta_2 + \theta_3$ la combinación de ambos. En cada unidad experimental podemos hacer lo siguiente:

a) No aplicar nada. b) Aplicar T_1. c) Aplicar T_2. d) Aplicar T_1 y T_2

Podemos codificar las cuatro situaciones escribiendo (0,0),(1,0), (0,1) y (1,1) para a),b),c) y d) respectivamente y reinterpretar la acción del experimentador en cada unidad experimental como la elección de dos variables de control z_1 y z_2, cada una tomando sólo los valores cero y uno. Es obvio que en nuestro problema, z puede ser considerado como un elemento del conjunto U, en donde

$$U = \{ (0,0) \; , \; (0,1) \; , \; (1,0), \; (1,1)\}$$

y escribir, supuesto $z = (z_1, z_2)^t$

$$E(Y_z) = \theta_0 + \theta_1 z_1 + \theta_2 z_2 + \theta_3 z_1 . z_2$$

Nuestro problema se convierte en ¿ qué elección de N puntos de U haremos, de forma que obtengamos máxima información sobre los parámetros desconocidos?.

1.1.2. FORMULACIÓN

Las hipótesis del problema son

H.1.: Para cada z de U podemos observar una variable aleatoria Y_z tal que

$$E(Y_z) = \sum_{i=1}^{k} f_i(z).\theta_i \qquad\qquad Var(Y_z) = \sigma^2$$

siendo $f = (f_1, f_2, \ldots, f_k)^t$ un vector de funciones completamente conocidas, definidas sobre U, $\theta = (\theta_1, \ldots, \theta_k)^t$ un vector paramétrico desconocido y σ^2 una cantidad conocida o desconocida.

El modelo anterior es conocido en la literatura estadística con el nombre de Modelo Lineal. Por otra parte, si llamamos X a f(U) y sus elementos los notamos $x = (x_1, x_2, \ldots, x_k)^t$ es obvio que la hipótesis anterior puede reformularse, - reformulación que adoptaremos en la presente memoria -, como

H.1.: Para cada x de X podemos observar una variable aleatoria Y_x tal que

$$E(Y_x) = \sum_i x_i \theta_i = x^t . \theta \qquad\qquad Var(Y_x) = \sigma^2$$

supuesto $x = (x_1, x_2, \ldots, x_k)^t$, $\theta = (\theta_1, \ldots, \theta_k)^t$ un vector paramétrico -- desconocido y σ^2 una cantidad conocida o desconocida.

H.2.: El espacio de control, $\mathcal{X}$, es un subconjunto compacto de R^k.

H.3.: Podemos realizar N observaciones incorreladas de la variable -- aleatoria Y_x en N puntos, no necesariamente distintos, de $\mathcal{X}$.

Una vez formuladas las hipótesis, el problema es: ¿ en qué N puntos de $\mathcal{X}$ realizaremos las observaciones de la variable Y_x , de forma - que disminuyamos, en algún sentido a precisar, el desconocimiento que- sobre el vector paramétrico θ tenemos?.

NOTACIONES: En la presente memoria, salvo mención expresa, adoptaremos las siguientes notaciones

a) B_k será la clase de matrices definidas no negativas de orden k.

b) $B_{k,0}$ denotará la clase de matrices definidas no negativas de- orden k tales que la suma de los vectores fila es cero y también la - de los vectores columna.

c) D notará la clase de diseños según Definición 1.2.1.1.

d) $C = \{M(d) : d \in D\}$

e) H designará a la clase de medidas de diseño según Def.A1.1.

f) $M = \{M(\eta) : \eta \in H\}$

1.2. PRINCIPALES CRITERIOS DE OPTIMALIDAD.

1.2.1. JUSTIFICACION DE ESTOS CRITERIOS.

Justificaremos estos criterios en función del tipo de información que sobre el vector paramétrico θ espera obtener el experimentador a- partir de la realización de N observaciones de la variable Y_x .

DEFINICION 1.2.1.1. Llamaremos diseño y lo notaremos d a cualquier -- conjunto de N puntos $x_{(1)}, \ldots, x_{(N)}$, no necesariamente distintos de $\mathcal{X}$.

Sea d un diseño, si observamos la variable Y_x en los puntos de d entonces, teniendo en cuenta las hipótesis de la sección anterior, podemos escribir

$$E(Y) = X.\theta \qquad\qquad cov(Y) = \sigma^2 I$$

en donde Y es el vector aleatorio de orden N×1 de las observaciones correspondientes a los puntos de d, X es una matriz de orden N×k cuya fila i-ésima es $x^t_{(i)}$ i=1,2,...,N , θ es un vector paramétrico desconocido de orden k×1 e I es la matriz identidad.

DEFINICION 1.2.1.2. La matriz $X^t.X$ recibe el nombre de matriz de información del diseño d y la notaremos M(d).

Consideremos ahora cuatro problemas estadísticos:

a) ESTIMACION DE CADA COMPONENTE DE θ.

Si M(d) es no singular, entonces podemos estimar cada componente de θ. Por el teorema de Gauss-Markov el mejor estimador lineal insesgado $\hat{\theta}$ de θ es:

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

y

$$cov(\hat{\theta}) = \sigma^2(X^t.X)^{-1}$$

Veamos algunas cuestiones sobre la matriz de información. Si llamamos c_j a la j-ésima columna de $X(X^t.X)^{-1}$ y si h_i es la i-ésima columna de X, entonces

$$\hat{\theta}_i = c_i^t.Y$$

con
$$var(\hat{\theta}_i) = \sigma^2(c_i^t.c_i)$$

Ahora bien, puesto que $(X^t.X)^{-1}.X^t.X = I_k$, podemos decir que $c_i^t h_j = \delta_{ij}$ siendo δ_{ij} la delta de Kronecker. Por otra parte, aplicando la desigualdad de Schwartz

$$(h_i^t.h_i)(c_i^t.c_i) \geq (h_i^t.c_i)^2 = 1$$

de modo que

$$\mathrm{var}(\hat{\theta}_i) \geq \sigma^2 / h_i^t . h_i \qquad (1)$$

Si fuera posible, nos gustaría seleccionar un diseño d que estimase cada uno de los parámetros con varianza mínima. Observemos que la igualdad en (1) sólo se da si $c_i = \gamma_i h_i$, en donde γ_i es una constante que será positiva puesto que $c_i^t h_i = 1$. Entonces

$$X.M^{-1}(d) = X.\ \mathrm{diag}(\gamma_1, \ldots, \gamma_k)$$

de modo que multiplicando la igualdad anterior por X^t obtenemos que la matriz de información es diagonal. Así pues, si pretendiéramos estimar el vector paramétrico θ de forma que la varianza de los estimadores fuese lo más pequeña posible, es obvio que nuestro mejor diseño sería aquel que su matriz de información fuese diagonal con elementos diagonales los mayores posibles. Claramente, en general no podemos asegurar que siempre exista un diseño que cumpla estas condiciones. De modo que podríamos buscar aproximaciones razonables tales como minimizar el promedio de las varianzas de cada uno de los parámetros estimados, minimizar la varianza generalizada, etc.

b) ESTIMACION DE FUNCIONES LINEALES DEL VECTOR PARAMETRICO.

b.1.) Funciones lineales en las que intervienen todas las componentes del vector paramétrico.

Sabemos que para que un diseño d pueda estimar $q^t.\theta$,- como es normal en la teoría de los modelos lineales entenderemos que $q^t.\theta$ es estimable si admite un estimador lineal insesgado-, en donde q es un vector columna de orden k×1, es condición necesaria y suficiente que la ecuación

$$M(d).Y = q$$

tenga solución en Y. Entonces, la varianza del mejor estimador lineal insesgado de $q^t.\theta$ es

$$\mathrm{var}(\widehat{q^t.\theta}) = \sigma^2 q^t.M^-(d).q$$

en donde $M^-(d)$ es cualquier inversa generalizada de $M(d)$.

Hemos de hacer notar que un diseño d puede estimar $q^t.\theta$ y no estimar el vector paramétrico θ .

Si queremos utilizar un diseño d para estimar $q^t.\theta$, existen varios criterios a la hora de la elección. Uno de ellos está basado en las siguientes desigualdades

$$\mu_{min} \leq \frac{q^t.M^-(d).q}{q^t.q} \leq \mu_{max}$$

siendo μ_{max} y μ_{min} el máximo y el mínimo respectivamente de los autovalores no nulos de $M^-(d)$. Así pues, tenemos las cotas siguientes para la varianza de $\widehat{q^t.\theta}$

$$q^t.q.\sigma^2\mu_{min} \leq \mathrm{var}(\widehat{q^t.\theta}) \leq \sigma^2.\mu_{max}.q^t.q$$

de modo que podría pensarse en utilizar diseños d que tuviesen el máximo autovalor no nulo de $M^-(d)$ lo más pequeño posible.

b.2.) Funciones lineales de un subvector paramétrico de θ.

Supongamos que realizamos la siguiente partición del vector paramétrico θ: $\theta^t = (\theta_1^t, \theta_2^t)$, en donde θ_1 tiene s<k componentes. Dado el diseño d, supuesto que $X^t = (X_1^t, X_2^t)$ siendo X_1^t una matriz de orden $s\times N$, definimos

$$M_S(d) = X_1^t X_1 - X_1^t X_2 (X_2^t X_2)^- . X_2^t X_1$$

y sabemos que θ_1 es estimable por el diseño d si y sólo si $M_S(d)$ es inversible. En este caso la matriz de covarianzas del mejor estimador lineal insesgado de θ_1 es

$$\mathrm{cov}(\hat{\theta}_1) = \sigma^2 M_S^{-1}(d)$$

NOTA 1: Obsérvese que en este punto podemos hacer el mismo estudio de θ_1 que el realizado anteriormente para θ.

Por otra parte, si nuestro interés se centra en la estimación de $q_1^t.\theta_1$, siendo q_1 un vector columna de orden $s\times 1$, sabemos que el diseño d estimará $q_1^t.\theta_1$ si y sólo si tiene solución en Y_1 el sistema

$$M_s(d).Y_1 = q_1$$

siendo entonces la varianza del mejor estimador lineal insesgado de $q_1^t.\theta_1$

$$\operatorname{var}(\widehat{q_1^t.\theta_1}) = \sigma^2.q_1^t.M_s^-(d).q_1$$

Observemos que al igual que en b.1.), un diseño puede estimar $q_1^t.\theta_1$ y sin embargo, no estimar el subvector paramétrico θ_1.

Si deseamos utilizar un diseño que estime $q_1^t.\theta_1$, hay varios criterios para la elección de los puntos $x_{(i)}$ que lo constituyen. Uno de ellos está basado en las desigualdades siguientes

$$\mu_{min} \leq \frac{q_1^t M_s^-(d).q_1}{q_1^t q_1} \leq \mu_{max}$$

siendo μ_{max} y μ_{min} el máximo y el mínimo respectivamente de los autovalores no nulos de $M_s^-(d)$. Esta desigualdad proporciona una cota para la varianza de $\widehat{q_1^t.\theta_1}$

$$\sigma^2\mu_{min}\, q_1^t q_1 \leq \operatorname{var}(\widehat{q_1^t\theta_1}) \leq \sigma^2\mu_{max}\, q_1^t q_1$$

De manera que una razonable elección del diseño d sería aquella en la que el mayor autovalor no nulo de $M_s^-(d)$ fuese lo más pequeño posible; siempre, claro está, que el diseño d estimase a $q_1^t.\theta_1$.

c) CONTRASTE DE HIPOTESIS

c.1.) Contrastes para todo el vector paramétrico θ.

Supuesto que Y es normal multivariante y queremos contrastar la hipótesis $H_0 : \theta = 0$, entonces si M(d) es no singular, el F contraste usual tiene una función de potencia que es monótona creciente en el parámetro λ, donde $\lambda = \sigma^{-2}\theta^t.M(d).\theta$.

De modo que podemos afirmar

$$\sigma^{-2}\bar{\mu}_{min}\,\theta^{t}.\theta \leq \lambda \leq \sigma^{-2}\bar{\mu}_{max}\,.\,\theta^{t}.\theta$$

siendo $\bar{\mu}_{max}$ y $\bar{\mu}_{min}$ el máximo y el mínimo respectivamente de los autovalores de M(d). Es obvio que puede interesarnos hacer máxima la potencia mínima, en este sentido podemos intentar maximizar $\bar{\mu}_{min}$.

c.2.) Contrastes sobre un subvector θ_1 del vector θ .

Supuesto que Y es normal multivariante, si queremos contrastar - la hipótesis $H_0 : \theta_1 = 0$, entonces cuando $M_s(d)$ es no singular, el F -- contraste usual tiene una función de potencia que es monótona creciente en λ_1 , en donde $\lambda_1 = \sigma^{-2}\theta_1^{t} M_s(d)\,\theta_1$.

De forma que podemos escribir las siguientes desigualdades

$$\sigma^{-2}\bar{\mu}_{min}\,\theta_1^{t}\theta_1 \leq \lambda_1 \leq \sigma^{-2}\bar{\mu}_{max}\,\theta_1^{t}\theta_1$$

en donde $\bar{\mu}_{max}$ y μ_{min} son el máximo y el mínimo respectivamente de los - autovalores de $M_s(d)$.

d) CONSTRUCCIÓN DE REGIONES DE CONFIANZA.

d.1.) Regiones de confianza para el vector paramétrico θ.

Si suponemos que Y es normal multivariante y M(d) es no singular, sabemos que la mejor región de confianza para θ es un elipsoide sólido con las siguientes propiedades:

- El volumen (volumen esperado si σ^2 es desconocido) de dicho elip-- soide es proporcional a la raíz cuadrada de $|M^{-1}(d)|$. Cabría, por tanto pensar como posible elección del diseño d aquella que hiciese mínimo $|M^{-1}(d)|$.
- Los semiejes (semiejes esperados si σ^2 es desconocido) del elipsoide son proporcionales a los autovalores de $M^{-1}(d)$. De modo análogo, - una elección del diseño razonable sería aquella en la que el máximo - autovalor de $M^{-1}(d)$ fuese mínimo.

d.2.) Regiones de confianza para el subvector paramétrico θ_1.

Si Y es normal multivariante y $M_s(d)$ es no singular, sabemos que la mejor región de confianza para θ_1 es un elipsoide s-dimensional -- con las siguientes propiedades:

- El volumen (volumen esperado si σ^2 es desconocido) de dicho elipsoide es proporcional a la raíz cuadrada de $|M_s^{-1}(d)|$. De forma que podemos pensar en utilizar un diseño que minimice dicho contenido.
- Los semiejes (semiejes esperados si σ^2 es desconocido) de dicho --- elipsoide son proporcionales a los autovalores de $M_s^{-1}(d)$. De modo que podríamos realizar las N observaciones de la variable Y_x en puntos tales que $M_s^{-1}(d)$ tuviese su máximo autovalor lo más pequeño posible.

1.2.2. DEFINICIONES Y PROPIEDADES.

Para definir y estudiar los criterios de optimalidad más conocidos supondremos que $C\subset B_k$, posteriormente los describiremos en el contexto $C\subset B_{k,0}$.

Por otra parte, hemos de hacer notar que para algunos criterios- de optimalidad no será suficiente imponer sólo que X sea compacto y - habrá que exigir que pueda generar R^k o algún subespacio, ya que sólo en este caso existirá solución óptima para determinados criterios de- optimalidad.

1.2.2.1. G-OPTIMALIDAD.

Supongamos que X es un subconjunto compacto de R^k que genera R^k. Sabemos que bajo estas hipótesis sólo los diseños d tales que su ma-- triz de información es no singular pueden estimar $x^t.\theta$ para cualquier- x perteneciente a X.

Por otra parte, si M(d) es no singular

$$\operatorname{var}(x^t.\hat{\theta}) = \sigma^2.x^t.M^{-1}(d).x$$

de modo que podemos definir el siguiente criterio de optimalidad:

DEFINICION 1.2.2.1. Un diseño d^0 es G-óptimo si y sólo si

a) $M(d^0)$ es no singular

b) $\min_{d\in D} \max_{x\in X} x^t.M^{-1}(d).x = \max_{x\in X} x^t.M^{-1}(d^0).x$

Obsérvese que la condición a) no es más que la imposición de que el diseño d^0 pueda estimar $x^t.\theta$ con x perteneciente a X.

Este criterio de optimalidad fue introducido por Smith (1.918) y el nombre de G-optimalidad, (G, por global), fue debido a Kiefer. Fedorov prefiere la utilización del término minimax, pues obviamente un diseño óptimo bajo este criterio minimiza el máximo de la varianza del mejor estimador lineal insesgado de $x^t.\theta$ con x perteneciente a X.

Nótese que las hipótesis impuestas a X nos garantizan la existencia de diseños G-óptimos.

1.2.2.2. D-OPTIMALIDAD.

Aquí volvemos a imponer que X sea un compacto de R^k que genera R^k

DEFINICION 1.2.2.2. Un diseño $d^0 \in D$ es D-óptimo si y sólo si

a) $M(d^0)$ es no singular

b) $|M^{-1}(d^0)| = \min_{d\in D} |M^{-1}(d)|$

Es obvio que bajo las hipótesis impuestas a X siempre existe un diseño D-óptimo.

El concepto de D-optimalidad, (D por determinante), fue introducido y estudiado por Wald (1.943) y aplicado por Mood (1.946).

Este criterio tiene las siguientes propiedades:

- Supuesto que las observaciones de la variable aleatoria Y_x provienen de una ley normal, el volumen, (volumen esperado si σ^2 es desconocido), del elipsoide sólido de estimación para θ, a un nivel de confianza dado, es proporcional, para un diseño d, a $|M^{-1}(d)|^{\frac{1}{2}}$. De manera que un diseño D-óptimo minimiza el contenido del elipsoide de estimación para θ.
- El diseño D-óptimo minimiza la varianza generalizada de los estima-

dores de los parámetros.

Este criterio, tal vez sobre el que más se ha escrito en la literatura de los diseños óptimos, debe su importancia, además de a las propiedades antes expuestas, a su relación con la G-optimalidad que ahora estudiaremos. Como ya enunciamos en el apéndice 1, llegado el momento de caracterizar las soluciones óptimas de los criterios de optimalidad, hemos de recurrir a la teoría de la aproximación. Es en este contexto donde se enuncia el teorema de equivalencia de Kiefer y Wolfowitz que establece la relación existente entre G-optimalidad y D-optimalidad. Antes de enunciarlo, recomendamos la lectura de la Nota 2 y las definiciones de G-optimalidad y D-optimalidad en el contexto de la teoría de la aproximación que se dan en el apéndice 1.

TEOREMA 1.2.2.1. Teorema de equivalencia. Kiefer-Wolfowitz (1.960).

Las siguientes proposiciones son equivalentes:

a) $\eta^0 \varepsilon H$ es D-óptimo.

b) $\eta^0 \varepsilon H$ es G-óptimo.

c) $\max\limits_{x \varepsilon X} d(x,\eta^0) = k$

siendo $d(x,\eta) = x^t . M^{-1}(\eta).x$ si $M(\eta)$ es no singular e ∞ si $M(\eta)$ es singular.

Varias son las demostraciones que de este teorema se han dado, la primera debida a Kiefer y Wolfowitz utiliza como herramienta fundamental el cálculo diferencial, en 1.966, Karlin-Studden proponen una nueva demostración basada en teoría de juegos que estudiaremos en el siguiente capítulo, por último, basado en consideraciones de tipo geométrico, que detallamos en el capítulo tercero de la presente memoria, Sibson en 1.972 propone un teorema de dualidad que tiene como corolario el teorema de Kiefer y Wolfowitz.

1.2.2.3. D_S- y D_A- OPTIMALIDAD.

Supongamos que realizamos la siguiente partición del vector para métrico θ , $\theta^t = (\theta_1^t, \theta_2^t)$ en donde θ_1 tiene s<k componentes. Sabemos que dado un diseño d, si realizamos la siguiente partición de M(d)

$$M(d) = \begin{pmatrix} M_{11}(d) & M_{12}(d) \\ M_{21}(d) & M_{22}(d) \end{pmatrix}$$

en la que $M_{11}(d)$ es una matriz de orden s y $M_{21}^t(d) = M_{12}(d)$, entonces la matriz de covarianzas del mejor estimador lineal insesgado de θ_1 es

$$cov(\hat{\theta}_1) = \sigma^2 M_s^{-1}(d)$$

de forma que podemos proponer un criterio de optimalidad análogo a la D-optimalidad cuando nuestro interés se centre en la estimación del subvector paramétrico θ_1. El criterio es el siguiente:

DEFINICION 1.2.2.3. Un diseño $d^0 \varepsilon D$ es D_s-óptimo si y sólo si

a) $M_s(d^0)$ es no singular

b) $|M_s^{-1}(d^0)| = \min_{d \varepsilon D} |M_s^{-1}(d)|$

Este criterio de optimalidad, propuesto y definido principalmente por Karlin y Studden (1.966) y Atwood (1.969), tiene propiedades análogas a la D-optimalidad.

Por otra parte, un cierto teorema de equivalencia para este criterio ha sido establecido, pero teniendo en cuenta la posibilidad de que diseños con matriz de información singular sean D_s-óptimos, esta caracterización resulta bastante complicada.

El teorema de equivalencia para este criterio, que comentamos en el capítulo siguiente, fue propuesto por Karlin y Studden (1.966), basándose en teoría de juegos. Sin embargo, fue necesaria una ligera modificación, debida a Atwood (1.969), por un error que contenía.

El teorema se enuncia en el contexto de la teoría de la aproximación, que describimos en el apéndice 1.

Sea $\eta \varepsilon H$ un diseño, una medida de diseño, definimos para cada matriz C de orden $s \times (k-s)$

$$d_s(x,\eta,C) = (x_1 - C.x_2)^t M_s^{-1}(\eta)(x_1 - C.x_2) \qquad \text{si } |M_s(\eta)| \neq 0$$
$$d_s(x,\eta,C) = \infty \qquad \text{si } |M_s(\eta)| = 0$$

supuesto que $x^t = (x_1^t, x_2^t)$ y x_1 tiene s componentes.

TEOREMA 1.2.2.2. (Atwood-Karlin-Studden). Existe una matriz C^0 de orden $s\times(k-s)$ tal que son equivalentes las proposiciones siguientes:

a) $\eta^0 \in H$ es D_s-óptimo

b) $\max_{x\in X} d_s(x,\eta^0;C^0) = \min_{\eta\in H} \max_{x\in X} d_s(x,\eta,C^0)$

c) $\max_{x\in X} d_s(x,\eta^0;C^0) = s$

Posteriormente en 1.973, Silvey y Titterington propusieron una -- nueva demostración del anterior teorema basada en teoría de la dualidad. Este teorema se analiza en el tercer capítulo de la presente memoria.

Supongamos ahora que nuestro interés se centra en la estimación $A^t.\theta$ en donde A es una matriz de orden $k\times s$ y rango $s \leq k$. Sabemos que - dado un diseño d si $A^t.M^-(d).A$ es no singular, la matriz de covarianzas del mejor estimador lineal insesgado de $A^t.\theta$, necesariamente -- único, es

$$\operatorname{cov}(\widehat{A^t.\theta}) = \sigma^2.A^t.M^-(d).A$$

de forma que Sibson (1.974) propone el siguiente criterio de optimalidad

DEFINICION 1.2.2.3. Un diseño $d^0 \in D$ es D_A-óptimo si y sólo si

a) $A^t.M^-(d^0).A$ es no singular

b) $|A^t.M^-(d^0).A| = \min_{d\in D} |A^t.M^-(d).A|$

Es obvio que este criterio es una extensión de la D- y D_s- optimalidad, puesto que estos dos criterios no corresponden más que a dos elecciones particulares de la matriz A, a saber $A = I_k$ y $A^t = (I_s : 0)$, respectivamente.

En el siguiente capítulo, propondremos una caracterización de este criterio de optimalidad basada en teoría de juegos. Por otra parte, Sibson en 1.974, como hizo el mismo autor en 1.972 y Silvey-Titterington en 1.973 para la D- y D_s- optimalidad respectivamente, propuso basándose en algunos conceptos de dualidad, una caracterización de este criterio. Este resultado por su proximidad con los anteriores no será reproducido aquí.

NOTA 2: Es obvio que para que existan diseños D_s-óptimos o D_A-óptimos será necesario imponer a X hipótesis de generar determinados subespacios de R^k.

1.2.2.4. L-OPTIMALIDAD.

Supongamos que deseamos estimar $c^t.\theta$ con $c \in C$, un conjunto compacto de R^k. Entonces, si el diseño d estima $c^t.\theta$, es decir si $c^t.M^-(d)c$ es positivo, podemos estar interesados en minimizar un determinado -- promedio sobre C de $c^t.M^-(d).c$. Si este promedio es con respecto a una distribución de probabilidad μ, entonces pretendemos minimizar

$$\int_C c^t.M^-(d).c\mu(dc)$$

pero

$$\int_C c^t.M^-(d).c\mu(dc) = tr(M^-(d).B)$$

en donde tr denota traza y

$$B = \int_C c.c^t\mu(dc)$$

Obsérvese que si B tiene rango s, entonces puede escribirse como $A.A^t$ en donde A tiene orden k×s y rango s, de manera que podemos escribir

$$tr(M^-(d).B) = tr(A^t.M^-(d).A)$$

que muestra una profunda relación con los criterios de optimalidad antes enunciados.

Por otra parte, este tipo de criterios, introducidos por Fedorov (1.972) tienen una forma funcional que es lineal no negativa en $M^{-}(d)$ de aquí su nombre de L-optimalidad.

Dentro de los criterios lineales existe uno que por su importancia se estudia separadamente. Es el siguiente:

DEFINICION 1.2.2.4. Un diseño $d^0 \in D$ es A-óptimo si y sólo si

a) $M(d^0)$ es no singular

b) $tr(M^{-1}(d^0)) = \min_{d \in D} tr(M^{-1}(d))$

A- aparece por promedio (average en inglés). Obsérvese que para que existan diseños A-óptimos es necesario que X genere R^k.

Por otra parte, en sentido estadístico, si d^0 es A-óptimo minimiza el promedio de las varianzas de los mejores estimadores lineales - insesgados de $\theta_1, \theta_2, \ldots, \theta_k$. Aunque este criterio se encuadra dentro de los criterios lineales su origen es anterior a ellos, concretamente - fue introducido y estudiado por Elfving (1.952) y Chernoff (1.953).

Dos son los resultados fundamentales que existen sobre la A-optimalidad, ambos enunciados por Fedorov (1.972) bajo la teoría de la -- aproximación. Son los siguientes:

TEOREMA 1.2.2.3. Las proposiciones siguientes son equivalentes:

a) $\eta^0 \in H$ minimiza $tr\, M^{-1}(\eta)$

b) $\eta^0 \in H$ minimiza $\max_{x \in X} x^t.M^{-2}(\eta).x$

c) $\max_{x \in X} x^t.M^{-2}(\eta^0).x = tr\, M^{-1}(\eta^0)$

El otro teorema es una condición suficiente para que un diseño - sea simultáneamente D- y A- óptimo.

TEOREMA 1.2.2.4. $\eta^0 \in H$ es D- y A- óptimo si

$$k.M^{-1}(\eta^0) = M^{-2}(\eta^0)$$

Cuando las distribuciones de probabilidad que generan los criterios lineales son degeneradas en un punto, el criterio de optimalidad recibe un nombre especial, es el siguiente:

DEFINICION 1.2.2.5. Un diseño d^0 es c-óptimo si y sólo si

a) $c^t.M^-(d^0).c$ es positivo

b) $c^t.M^-(d^0).c = \min_{d \in D} c^t.M^-(d).c$

El criterio fue establecido por Elfving (1.952) y su solución -- conduce a consideraciones geométricas muy interesantes que se describen en el siguiente capítulo.

1.2.2.5. E-OPTIMALIDAD.

DEFINICION 1.2.2.6. Un diseño $d^0 \in D$ es E-óptimo si y sólo si

a) $M(d^0)$ es no singular

b) $\min_{d \in D} \mu_k^{-1}(d) = \mu_k^{-1}(d^0)$

siendo $\mu_1(d) \geq \mu_2(d) \geq \ldots \geq \mu_k(d)$ los autovalores de $M(d)$.

Claramente, para la existencia de diseños E-óptimos es necesario que X genere R^k.

Este criterio de optimalidad fue considerado primero en contraste de hipótesis por Wald (1.943) y posteriormente por Ehrenfeld, que- en 1.955 llamando u al máximo valor de μ_k en el problema que consideramos define la eficiencia de un diseño como el cociente $\mu_k(d)/u$ y - aquel diseño d tal que su eficiencia sea igual a uno lo llama el más eficiente.

Kiefer utiliza el término E-optimalidad (E por autovalor,eigen-- value en inglés) con el que ha pasado a la literatura estadística.

Este criterio tiene las siguientes propiedades:

- En contraste de hipótesis, bajo normalidad de las observaciones, un diseño E-óptimo maximiza la potencia mínima del F contraste usual asociado a $H_0 : \theta = 0$, de tamaño α sobre el contorno $\theta^t.\theta = c$ para cada valor de α y c.

- En estimación puntual, un diseño E-óptimo minimiza la varianza máxima del mejor estimador lineal insesgado de $q^t.\theta$ sobre los vectores q de orden k con $q^t.q = 1$

- En estimación por intervalos, un diseño E-óptimo minimiza el mayor semieje del elipsoide de estimación para θ cuando las observaciones se suponen normales.

Los resultados más importantes sobre E-optimalidad no los reproduciremos aquí, puesto que este criterio presenta una conexión muy peculiar con la ϕ_p-optimalidad que describiremos en la sección cuarta presente capítulo, en consecuencia, su estudio será pospuesto a la introducción de dichos criterios.

NOTA 3: La definición de E-optimalidad se ha dado suponiendo que M(d) era no singular, es decir, nos hemos restringido a aquellos diseños d que estiman todos los parámetros. Sin embargo, si nuestro interés se centra sólo en un subvector paramétrico θ_1 de θ que tenga s componentes, entonces es evidente que sólo tendremos que exigirle al diseño óptimo que $M_s(d)$ sea no singular y definir la E-optimalidad en función de los autovalores de $M_s(d)$.

Tendríamos entonces el criterio de E_s-optimalidad y las propiedades anteriores son fácilmente formulables en función de θ_1. De forma análoga, si nuestro interés se centra en problemas estadísticos relacionados con la estimación de $A^t.\theta$ en donde A es una matriz de orden $k{\times}s$ y rango $s \leq k$, debemos exigirle sólo que $A^t.M^-(d).A$ sea no singular y definir lo que podríamos llamar E_A-optimalidad en función de los autovalores de $(A^t.M^-(d).A)^{-1}$ con una fácil reinterpretación de las propiedades enunciadas para la E-optimalidad.

Si analizamos los criterios de optimalidad propuestos en esta sección es obvio que todos ellos, salvo la G-optimalidad, pueden ser formulados como un problema de minimización de una función ϕ definida sobre las matrices de información de diseños.

Estas funciones son las siguientes:

D-optimalidad: $\phi_D(M(d)) = |M^{-1}(d)| = \prod_{i=1}^{k} \mu_i^{-1}(d)$

A-optimalidad: $\phi_A(M(d)) = tr(M^{-1}(d)) = \sum_{i=1}^{k} \mu_i^{-1}(d)$

E-optimalidad: $\phi_E(M(d)) = \mu_k^{-1}(d)$

en donde $\mu_1(d) \geq \mu_2(d) \geq \ldots \geq \mu_k(d)$ son los autovalores de M(d).

Obsérvese que no hemos escrito las funciones ϕ correspondientes a D_s-optimalidad, D_A-optimalidad, L-optimalidad,- salvo A-optimalidad -, E_s-optimalidad y E_A-optimalidad. La razón es clara, las funciones correspondientes a estos criterios son análogas a las anteriores pero formuladas en función de los autovalores de $M_s(d)$ o $(A^t.M^-(d).A)^{-1}$.

Nótese también que los criterios de optimalidad anteriores pueden ser formulados en función de la inversa de la matriz de información, sin embargo, este planteamiento no mejora la solución del problema, (Hedayat 1.981).

Por último, vamos a definir los criterios de optimalidad anteriores suponiendo que $C \subset B_{k,0}$. En este caso tendremos

D-optimalidad: $\phi_D(M(d)) = \prod_{i=1}^{k-1} \mu_i^{-1}(d)$

A-optimalidad: $\phi_A(M(d)) = \sum_{i=1}^{k-1} \mu_i^{-1}(d)$

E-optimalidad: $\phi_E(M(d)) = \mu_{k-1}^{-1}(d)$

Es obvio que los restantes criterios pueden ser fácilmente formulables en esta nueva situación.

Téngase en cuenta que al cumplirse $C \subset B_{k,0}$ el menor autovalor de M(d) es nulo. No vamos a justificar aquí los criterios de optimalidad así definidos, los estudiaremos en la próxima sección cuando analicemos la optimalidad de diseños en bloques.

1.3. S-OPTIMALIDAD Y (M,S)-OPTIMALIDAD.

1.3.1. OPTIMALIDAD DE DISEÑOS EN BOLQUES.

Consideremos la clase de diseños en bloques,(Graybill,1.961), -- D_{kvb}, para valores de k, v y b fijos con k>v, en donde k tratamientos-son aplicados en b bloques de v unidades experimentales cada uno, supuesto que cada tratamiento es aplicado r veces. Obsérvese que en este tipo de problemas los vectores del espacio de control X tienen una forma muy peculiar y además el conjunto de N observaciones que constituyen cada diseño ha de cumplir ciertas propiedades.

Frecuentemente, el experimentador está interesado exclusivamente en problemas estadísticos relacionados con las comparaciones de los - tratamientos. En el modelo usual $Y_{ij} = \alpha_i + \beta_j + e_{ij}$ en donde α_i es el -- efecto del i-ésimo tratamiento y β_j es el efecto debido al j-ésimo bloque, la matriz de información de los efectos de los tratamientos viene dada por

$$M(d) = r.I - \frac{1}{v}.N.N^t$$

en donde $N = (n_{ij})$ es la matriz de incidencia, siendo n_{ij} el número - de veces que el tratamiento i aparece en el bloque j. Sabemos que M(d) es definida no negativa, simétrica y con suma de los vectores fila, - columna, nula. De forma que las únicas combinaciones lineales estimables de los efectos de los tratamientos son los contrastes (es decir, combinaciones lineales de la forma $\Sigma c_i.\alpha_i$ con $\Sigma_i c_i = 0$).

Si queremos estimar todos los contrastes entre los tratamientos, hemos de considerar sólo los diseños conectados, es decir aquellos diseños d tales que M(d) tiene rango k-1.

Consideremos ahora los diseños en bloques conectados y sean los-autovalores no nulos de M(d), $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{k-1} > 0$, por brevedad de notación no escribiremos la dependencia de los autovalores del diseño d. Sea - $\{P_i^t.\alpha\}$ i=1,2,...,k-1 cualquier conjunto completo de k-1 contrastes - ortogonales normalizados. Si escribimos

$$P = (P_1, P_2, \ldots, P_{k-1}) \quad y \quad P^t.\alpha = \rho$$

se puede probar que $P^t.M(d).P$ es no singular, que tiene por autovalores a los autovalores no nulos de $M(d)$ y que la matriz de covarianzas del mejor estimador lineal insesgado de ρ es

$$\mathrm{cov}(\hat{\rho}) = \sigma^2 (P^t.M(d).P)^{-1}$$

de forma que la varianza generalizada de $\hat{\rho}$ viene dada por

$$\sigma^2 \prod_{i=1}^{k-1} \mu_i^{-1}$$

La hipótesis nula $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$, al ser $\{P_i^t.\alpha\}, i= 1,2,.,k-1$ una clase completa de contrastes, es equivalente a $H_0 : \rho_0 = \rho_1 = \ldots = \rho_{k-1} = 0$, tiene una función de potencia que es monótona creciente en β, siendo $\beta = \rho^t.P^t.M(d).P.\rho$.

Entonces la optimalidad de un diseño en bloques podría formularse como sigue:

- La varianza media del conjunto completo de k-1 contrastes ortogonales normalizados es proporcional a $\sum_{i=1}^{k-1} \mu_i^{-1}$ (Ehrenfeld,1.955;Kempthorne-1.956).
- Wald (1.943) afirma que en el contraste de H_0 no es posible maximizar la potencia para todos los valores de ρ, de modo que tendríamos que maximizar la potencia para $\rho^t.\rho$ = cte. Esto conduce a maximizar el autovalor μ_{k-1}.
- Wald (1.943) argumenta además que por consideraciones matemáticas sería a veces más simple minimizar $\prod_{i=1}^{k-1} \mu_i^{-1}$

Obsérvese que estos tres conceptos de optimalidad corresponden respectivamente a los criterios A-, E- y D- optimalidad.

Por otra parte, si pretendiésemos relacionar estos tres criterios de optimalidad en los diseños en bloques, bastaría observar que para cualquier diseño d:

$$tr(M(d)) = \sum_{i=1}^{k-1} \mu_i = k.r(1 - 1/v)$$

de modo que al ser la suma de los autovalores constante para todo diseño en bloques, aquel que tenga todos los autovalores iguales será simultáneamente A-, E- y D- óptimo.

Pero es sabido que si para k, v y b fijos existe un diseño en bloques incompletos balanceados, es decir un diseño en bloques d^0 que cumpla:

- Cada bloque contiene v unidades experimentales.
- Hay más tratamientos que unidades experimentales en cada bloque, $k>v$.
- Cada tratamiento aparece en r bloques.
- Cada par de tratamientos ocurren juntos en el mismo número de bloques ($\lambda = \sum_{j=1}^{b} n_{ij} . n_{i'j} \quad i \neq i'$).

Su matriz de información es

$$M(d^0) = a.I + b.J$$

en donde I es la matriz identidad de orden k y J es la matriz que tiene todos sus elementos iguales a uno y $a + k.b = 0$.

Ahora bien, una matriz A de orden k de la forma

$$A = u.I + v.J$$

tiene autovalores u y $u + k.v$ con orden de multiplicidad k-1 y 1,(ver Searle 1.982). En consecuencia, un diseño en bloques incompletos balanceados tiene todos sus autovalores no nulos iguales, de modo que si existe, es simultáneamente A-, E- y D- óptimo.

Supongamos que para nuestro problema de diseño óptimo no existe un diseño simétrico ($\mu_1 = \mu_2 = \ldots = \mu_{k-1}$). Entonces, si para cada diseño pudiésemos establecer una medida de la distancia al hipotético diseño simétrico; aquel que estuviese más próximo sería evidentemente un buen diseño. Esta idea, debida a Shah (1.960), es el origen de los criterios de optimalidad que analizaremos y estudiaremos en el siguiente apartado.

1.3.2. DEFINICIONES DE S-OPTUMALIDAD Y (M,S)-OPTIMALIDAD=

Aunque en su origen estos criterios de optimalidad fueron definidos suponiendo que $C \subset B_{k,0}$, aquí comenzaremos estudiándolos en el supuesto $C \subset B_k$ para posteriormente hacer una breve mención del caso anterior.

Supongamos que $C \subset B_k$, cuando $tr(M(d)) = A$ es una constante para todo d perteneciente a D, un diseño con todos los autovalores iguales es simultáneamente D-, A- y E- óptimo (este tipo de diseños recibe el nombre de simétricos). Desgraciadamente, los diseños simétricos no existen siempre, en consecuencia, en su ausencia, podríamos utilizar como óptimo aquel más próximo al hipotético diseño simétrico.

Así pues, siguiendo a Shah (1.960), si no existen diseños simétricos en D, utilizamos el diseño d que minimice la distancia euclídea entre $(\mu_1(d),\ldots,\mu_k(d))$ y $(A/k,\ldots,A/k)$ es decir

$$\sum_{i=1}^{k}(\mu_i(d) - A/k)^2$$

pero teniendo en cuenta que $\sum_i \mu_i(d) = A$, esto es equivalente a minimizar $\sum_i \mu_i^2(d) = tr(M^2(d))$. De forma que el criterio de optimalidad queda como sigue:

DEFINICION 1.3.2.1. Supongamos que $tr(M(d)) = A$, es una constante para todo $d \in D$. $d^0 \in D$ es S-óptimo si y sólo si

$$tr(M^2(d^0)) = \min_{d \in D} tr(M^2(d))$$

La expresión S-optimalidad fue utilizada por primera vez por Kiefer (1.974). En este artículo hace un breve pero muy interesante estudio sobre la caracterización de sus soluciones óptimas.

Si $C \subset B_{k,0}$ el criterio de optimalidad quedaría igual sin más que reemplazar k-1 por k,

Motivado por el criterio de Shah; Eccleston y Hedayat (1.974) --

proponen un procedimiento análogo cuando tr(M(d)) no es constante sobre D.

Sea $C^1 \subset C$ tal que las matrices de C^1 tienen traza máxima.

DEFINICION 1.3.2.2. Un diseño $d^0 \in D$ es (M,S)-óptimo si y sólo si

a) $M(d^0) \in C^1$

b) $tr(M^2(d^0)) = \min_{d \in D^1} tr(M^2(d))$

en donde $D^1 = \{ d \in D: M(d) \in C^1 \}$.

Si $C \subset B_{k,0}$ es obvio que para este criterio bastaría escribir $k-1$ por k, si se formula en función de los autovalores de M(d).

NOTA 3: Hemos de hacer notar que en el contexto de diseños en bloques, un diseño puede ser S- o (M,S)- óptimo sin ser un diseño conectado, - es decir, sin tener k-1 autovalores no nulos. Esto ha motivado la introducción de conceptos como diseños globalmente conectados,pseu-globalmente conectados y localmente conectados, desarrollados por Eccleston y Hedayat (1.974). En 1.978, Jacroux estudió algunas propiedades - de los diseños (M,S)-óptimos y ciertas relaciones con los diseños conectados. Posteriormente, en 1.980 Jacroux y Seely han establecido -- condiciones suficientes para la (M,S)-optimalidad de un diseño. Por - técnicas de construcción directa Roy (1.982) ha probado que para bloques de tamaño tres y si el número de tratamientos es $k \neq 3$(módulo 6) - existen los diseños propuestos por Jacroux y Seely independientemente del número de bloques.

Por último destaquemos la existencia de los diseños de grafo regular, introducidos por John y Michell (1.977), muy relacionados con la (M,S)-optimalidad y que estudiaremos cuando introduzcamos los criterios de optimalidad de tipo I y II.

1.4. Φ_p-OPTIMALIDAD.

En 1.974 Kiefer introduce una nueva familia de criterios de opti malidad. El escribe en 1.974 (en el contexto de la teoría de la aproximación que se estudia en el apéndice 1):

Sea ϕ una función que es real o ∞ en M. Un problema de la teoría de diseños óptimos es la caracterización de diseños η^0 que sean ϕ-óptimos, es decir, diseños para los cuales

$$\phi(M(\eta^0)) = \min_{\eta \in H} \phi(M(\eta))$$

Los ejemplos más conocidos de funciones ϕ corresponden a la D- , A- y E- optimalidad.

Ahora bien, puesto que la D- y A- optimalidad tienen teoremas de caracterización conocidos, la pregunta es la siguiente ¿ a qué tipo de funciones ϕ se les puede encontrar caracterizaciones análogas?.

Kiefer (1.974) estudia en su artículo numerosas funciones ϕ, nosotros nos restringiremos a una cierta subclase de dichas funciones.

Definimos para $\eta \in H$

$$\phi_p(M(\eta)) = [\frac{1}{k}.tr(M^{-p}(\eta))]^{1/p} = [\frac{1}{k}.\sum_{i=1}^{k} \mu_i^{-p}(\eta)]^{1/p} \qquad 0 < p < \infty$$

en donde $\mu_1(\eta) \geq \mu_2(\eta) \geq \ldots \geq \mu_k(\eta)$ son los autovalores de $M(\eta)$.

DEFINICION 1.4.1. Un diseño $\eta^0 \in H$ es ϕ_p-óptimo si y sólo si

$$\phi_p(M(\eta^0)) = \min_{\eta \in H} \phi_p(M(\eta))$$

El resultado más importante sobre estos criterios es el siguiente:

TEOREMA 1.4.1.

a) $\phi_1(M(\eta)) = \frac{1}{k}\sum_{i=1}^{k} \mu_i^{-1}(\eta)$

b) $\phi_0(M(\eta)) = \lim_{p \to 0} \phi_p(M(\eta)) = (\prod_{i=1}^{k} \mu_i^{-1}(\eta))^{1/K}$

c) $\phi_\infty(M(\eta)) = \lim_{p \to \infty} \phi_p(M(\eta)) = \mu_k^{-1}(\eta)$

Enunciemos a continuación el siguiente corolario.

COROLARIO 1.4.1.

a) ϕ_1-optimalidad $\Leftrightarrow$ A-optimalidad

b) ϕ_0-optimalidad $\Leftrightarrow$ D-optimalidad

c) ϕ_∞-optimalidad $\Leftrightarrow$ E-optimalidad

Observemos que al ser X compacto, si este conjunto genera R^k, entonces existirá solución óptima finita para estos criterios. De modo análogo, si nuestro interés se centra en la estimación de $A^t.\theta$, podríamos definir, siguiendo a Kiefer (1.974)

$$\phi_{p,A}(M(\eta)) = [\frac{1}{s}.tr(A^t.M^-(\eta).A)^p]^{1/p}$$

siendo s = rang(A). De modo análogo, para que exista solución óptima finita, para estos criterios de optimalidad, habrá que imponer que la ecuación $M(\eta).Y = A$ tenga solución en Y para algún $\eta \in H$. Notemos que aunque la función $\phi_{p,A}$ es bastante parecida a ϕ_p, el signo del exponente obliga a un estudio más detenido.

Por último, si nuestro interés no se centra en la caracterización de las soluciones óptimas para estos criterios de optimalidad y sólo pretendemos definir la función ϕ_p sobre el conjunto C, la definición será la siguiente

a) Si $C \subset B_k$

$$\phi_p(M(d)) = [\frac{1}{k}.tr(M^{-p}(d)]^{1/p} = [\frac{1}{k}\sum_{i=1}^{k}\mu_i^{-p}(d)]^{1/p} \qquad 0<p<\infty$$

en donde $\mu_1(d), \mu_2(d),\ldots,\mu_k(d)$ son los autovalores de M(d)

b) Si $C \subset B_{k,0}$

$$\phi_p(M(d)) = [\frac{1}{k-1}\sum_{i=1}^{k-1}\mu_i^{-p}(d)]^{1/p} \qquad 0<p<\infty$$

supuesto que $\mu_1(d),\mu_2(d),\ldots,\mu_{k-1}(d)$ y $\mu_k(d) = 0$ son los autovalores de M(d).

1.5. OPTIMALIDAD UNIVERSAL.

En 1.975 Kiefer definió un criterio de optimalidad muy general. Lo describiremos en un principio suponiendo que $C \subset B_{k,0}$. Posteriormente lo estudiaremos cuando $C \subset B_k$, comentando la modificación propuesta por Sinha y Mukerjee (1.982).

DEFINICION 1.5.1. Decimos que $d^0 \varepsilon D$ es un diseño universalmente óptimo si d^0 minimiza $\phi(M(d))$ para cualquier $\phi : B_{k,0} \rightarrow (-\infty, \infty]$ que cumpla

a) ϕ es convexa.

b) $\phi(b.M)$ es no creciente en el escalar $b \geq 0$, para cada $M \varepsilon B_{k,0}$

c) ϕ es invariante para cada permutación de filas y la misma de columnas.

Notemos que las funciones ϕ_p cumplen las propiedades anteriores. En consecuencia, si un diseño es universalmente óptimo, para $0 \leq p \leq \infty$ es ϕ_p-óptimo. Por tanto será A-, E- y D- óptimo.

Por otra parte es claro que no todos los problemas de diseño óptimo contienen un diseño universalmente óptimo.

Observemos que $-\mathrm{tr}M(d)$ cumple las condiciones enunciadas en la definición anterior, de modo que podemos escribir:

TEOREMA 1.5.1. Si $d^0 \varepsilon D$ es universalmente óptimo, entonces $\mathrm{tr}M(d^0)$ es máximo.

Estudiemos, ahora, muy brevemente, un tipo de matrices íntimamente relacionado con este criterio de optimalidad.

DEFINICION 1.5.2. Una matriz M es llamada completamente simétrica (c.s) si $M = \alpha.I_k + \beta.J_k$, en donde α y β son escalares, I_k es la matriz identidad de orden k y J_k es una matriz de orden k con todos sus elementos iguales a la unidad.

Las propiedades más importantes que poseen estas matrices, alguna de las cuales ya ha sido utilizada en la sección segunda del presente capítulo, son las siguientes:

- Los autovalores de M son α y $\alpha + k.\beta$ con multiplicidad $k-1$ y 1 respectivamente (Searle 1.982). En consecuencia, si $M \varepsilon B_{k,0}$ sus autovalores serán α y 0 con multiplicidad $k-1$ y 1 respectivamente.
- Por otra parte, teniendo en cuenta que si $A = A^t$ es una matriz de orden k que tiene un autovalor λ con multiplicidad m, entonces $A - \lambda I$ tiene rango $k-m$ (Searle 1.982). Podemos afirmar, por tanto, que si A tiene $k-1$ autovalores iguales, entonces es una matriz completamente simétrica.

La caracterización más importante sobre la optimalidad universal es la siguiente:

TEOREMA 1.5.2. Supongamos que $C \subset B_{k,0}$, contiene $M(d^0)$ tal que

a) $M(d^0)$ es completamente simétrica.

b) $tr(M(d^0)) = \max_{d \varepsilon D} tr(M(d))$ (1)

entonces d^0 es universalmente óptimo en D.

Podríamos preguntarnos ahora las condiciones bajo las cuales un diseño que sea óptimo para un criterio de optimalidad, que cumpla las condiciones de la definición 1.5.1., es universalmente óptimo. La condición suficiente es la siguiente:

TEOREMA 1.5.3. Supongamos que existe $M(d^0)$ que cumple (1). Sea ϕ una función que cumple las condiciones de la definición 1.5.1., si además ϕ es estrictamente convexa, entonces cada diseño d^1 que sea ϕ-óptimo es tal que $M(d^1) = M(d^0)$. En consecuencia d^1 es universalmente óptimo.

Como ya dijimos anteriormente, no todos los problemas de diseño poseen un diseño universalmente óptimo. Nos proponemos, en consecuencia dar respuesta a la siguiente pregunta ¿ Si d^* es ϕ^1-óptimo, bajo qué condiciones será ϕ^2-óptimo, siendo ϕ^1 y ϕ^2 criterios de optimalidad?.

TEOREMA 1.5.4. Si $\phi^1 \leq \phi^2$ sobre C y si $\phi^1(M(d^\bullet)) = \phi^2(M(d^\bullet))$,entonces si $d^\bullet$ es ϕ^1-óptimo también será ϕ^2-óptimo.

Observemos que para los criterios ϕ_p se cumple,teniendo en cuenta la desigualdad de Hölder, que si $p < q$ entonces $\phi_p(M(d)) \leq \phi_q(M(d))$, con la igualdad si y sólo si M(d) tiene todos los autovalores iguales.

Por tanto podemos escribir

COROLARIO 1.5.1. Si $M(d^\bullet)$ es completamente simétrica y $d^\bullet$ es ϕ_p-óptimo, entonces $d^\bullet$ es ϕ_q-óptimo para $q \geq p$.

Veamos ahora algunas condiciones suficientes, dadas por diferentes autores, para la caracterización de la optimalidad de un diseño teniendo en cuenta la forma funcional del criterio.

Obsérvese que los criterios de optimalidad ϕ_p pueden formularse equivalentemente como sigue:

$$\phi_p^\bullet(M(d)) = \sum_{i=1}^{k-1} \mu_i^{-p}(d) \qquad (\phi_p\text{-optimalidad})$$

$$\phi_0^\bullet(M(d)) = -\sum_{i=1}^{k-1} \log\mu_i(d) \qquad (\text{D-optimalidad})$$

$$\phi_\infty^\bullet(M(d)) = \mu_{k-1}^{-1}(d) \qquad (\text{E-optimalidad})$$

De manera que para $0 \leq p < \infty$, $\phi_p^\bullet(M(d))$ puede escribirse como

$$\phi_p^\bullet(M(d)) = \sum_{i=1}^{k-1} f(\mu_i(d))$$

siendo f una función convexa. Concretamente $f(x) = x^{-p}$ y $f(x) = -\log x$ para la ϕ_p-optimalidad y D-optimalidad, respectivamente.

Es obvio que la caracterización de las soluciones óptimas para este tipo de criterios de optimalidad pasa necesariamente por el conocimiento de los autovalores de la matriz de información. Sin embargo, teniendo en cuenta el lema siguiente, existe la posibilidad de caracterizar las soluciones óptimas sin conocer los autovalores.

LEMA 1.5.1. Si f es una función convexa en $[0,+\infty)$. Entonces

$$\sum_{i=1}^{k-1} f(\mu_i(d)) \geq \frac{k-1}{k} . \sum_{j=1}^{k} f\left(\frac{k}{k-1} m_{jj}(d)\right)$$

para cualquier $M(d) = (m_{ij}(d))$ de $B_{k,0}$ con la igualdad si y sólo si $\mu_1(d) = \mu_2(d) = \ldots = \mu_{k-1}(d)$.

En consecuencia es válido el resultado siguiente:

TEOREMA 1.5.5.

Si $\phi^{\bullet}(M(d)) = \sum_{i=1}^{k-1} f(\mu_i(d))$, siendo f convexa y si existe $d^{\bullet} \in D$ tal que

- $M(d^{\bullet})$ es completamente simétrica

- $d^{\bullet}$ minimiza $\sum_{j=1}^{k} f\left(\frac{k}{k-1} m_{jj}(d)\right)$

entonces $d^{\bullet}$ es $\phi^{\bullet}$-óptimo.

Teniendo en cuenta este teorema y la formulación equivalente de los criterios ϕ_p como criterios $\phi^{\bullet}_p$ se puede escribir

COROLARIO 1.5.2.

- Si $M(d^{\bullet})$ es c.s. y minimiza $\sum_j m_{jj}^{-p}(d)$ entonces $d^{\bullet}$ es un diseño ϕ_p-óptimo.

- Si $M(d^{\bullet})$ es c.s. y maximiza $\sum_j \log m_{jj}(d)$ entonces $d^{\bullet}$ es un diseño D-óptimo.

Es obvio que la formulación de la E-optimalidad como límite de criterios ϕ_p cuando p crece indefinidamente no tiene, en principio, resultado equivalente en el contexto de los criterios $\phi^{\bullet}_p$, puesto que al hacer tender p a infinito $\phi^{\bullet}_p$ se anula en los diseños no degenerados.

Sin embargo, se pueden normalizar los criterios $\phi^{\bullet}_p$ y reconvertirlos en criterios ϕ_p de forma que se obtenga el siguiente corolario del teorema 1.5.5.

COROLARIO 1.5.3.

- Si $M(d^0)$ es completamente simétrica.

- $\min_j m_{jj}(d^0) = \max_{d \in D} \min_j m_{jj}(d)$

entonces d^0 es E-óptimo.

Extendamos, ahora la definición de optimalidad universal al caso $C \subset B_k$.

La extensión que parece natural es la siguiente:

DEFINICION 1.5.3. Decimos que $d^0 \in D$ es universalmente óptimo si d^0 minimiza $\phi(M(d))$ para cualquier $\phi: B_k \to (-\infty, \infty]$ que cumpla

- ϕ es convexa

- $\phi(b.M)$ es no creciente en el escalar $b \geq 0$, para cada $M \in B_k$

- ϕ es invariante para cada permutación de filas y (la misma) de columnas.

Ahora bien, teniendo en cuenta que la modificación del teorema 1.5.1. para los modelos de rango máximo conduce a cierto tipo de problemas para algunos modelos, (Sinha y Mukerjee 1.982), Sinha y Mukerjee (1.982) proponen la siguiente definición de optimalidad universal para modelos de rango máximo:

DEFINICION 1.5.4. Decimos que $d^0 \in D$ es universalmente óptimo si d^0 minimiza $\phi(M(d))$ para cualquier $\phi: B_k \to (-\infty, \infty]$ que cumpla

- ϕ es convexa, $\phi(M) = \infty$ si $\mathrm{rang}(M) < k$

- $\phi(\alpha I_k + \beta J_k) \geq \phi(aI_k)$ si $a \geq \alpha + \beta$

- ϕ es invariante para cada permutación de filas y (la misma) de columnas.

Bajo esta definición el teorema 1.5.1. quedaría supuesto $C \subset B_k$, como sigue (Sinha y Mukerjee 1.982):

TEOREMA 1.5.6. Si C contiene $M(d^o)$ que es múltiplo de I_k y maximiza - $trM(d)$, entonces d^o es universalmente óptimo. (Aquí la optimalidad -- universal se entiende referida a la definición1.5.4.).

1.6. CRITERIOS DE OPTIMALIDAD DE TIPO I Y II.

Kiefer (1.958) generalizó la noción de diseños en bloques incompletos balanceados (B.B.I.D.) a diseños en bloques balanceados (B.B.D.) La modificación fue suponer que el tamaño de cada bloque podría ser - mayor o igual que el número de tratamientos, pero imponiendo que cada tratamiento apareciese en cada bloque $int[v/k]$ o $int[v/k] + 1$. Obsérvese que si $k>v$ coincide con la definición de B.B.I.D.. Kiefer(1.958) probó que estos diseños eran A-, E- y D- óptimos para los problemas - de diseño en los que cada fila de la matriz del diseño, X, tiene un- uno en los k primeros lugares y otro uno entre los b últimos elementos, siendo los restantes elementos de cada fila cero, y además las b últimas columnas de X contienen v unos cada una de ellas.

Como la propiedad de optimalidad se basaba en que la traza de la matriz de información era máxima y que dicha matriz era completamente simétrica, la optimalidad de estos diseños fue extendida por Kiefer - en 1.975 a optimalidad universal.

Es obvio que los diseños en bloques balanceados sólo existen para clases muy restringidas de valores de los parámetros k,v y b. En - 1.977 John y Mitchell consideran un tipo de diseños en bloques incompletos, llamados diseños de grafo regular, que es casi balanceado en- el sentido de que cada par de tratamientos aparecen juntos en λ_1 o λ_2 bloques siendo $\lambda_2 = \lambda_1 + 1$. Este tipo de diseños ya no es simétrico pero es el más próximo al hipotético diseño simétrico en el sentido de- Shah (1.960) o Eccleston y Hedayat (1.974).

Con la definición de diseños de grafo regular se conjetura por - John y Mitchell (1.977) la optimalidad de estos diseños con respecto- a una clase muy amplia de criterios de optimalidad, esto hace surgiramplia literatura sobre este tipo de diseños. No olvidemos que una vez caracterizada la optimalidad de los diseños simétricos, el paso siguiente es el estudio de los diseños más próximos a ellos. En 1.978 Cheng define los criterios de optimalidad de tipo I y II, que a continuación estudiaremos, y prueba que bajo determinadas condiciones una subclase de los diseños de grafo regular es óptima en sus criterios de optimalidad. Posteriormente, han surgido numerosos artículos sobre la optimalidad de los diseños de grafo regular, entre los que citaremos Cheng (1.980), Jacroux (1.980), John y Williams (1.981), Constantine (1.981) y Jacroux (1.982).

Vamos a definir a continuación los criterios de optimalidad de tipo I y II. Supondremos que $C \subset B_{k,0}$, posteriormente los estudiaremos si $C \subset B_k$.

Sea pues, $C \subset B_{k,0}$ y $\bar{t}_D = \max\limits_{d \in D} trM(d)$

DEFINICIÓN 1.6.1. Un diseño $d^0 \varepsilon D$ satisface un criterio de optimalidad de tipo I si minimiza

$$\phi_f(M(d)) = \sum_{i=1}^{k-1} f(\mu_i(d))$$

en donde f es una función real valuada definida en $[0,\bar{t}_D]$ tal que:

- f es continua, estrictamente convexa y estrictamente decreciente en $[0,\bar{t}_D]$. Incluímos la posibilidad $f(0) = \lim\limits_{x \to 0^+} f(x) = \infty$

- f es continuamente diferenciable en $(0,\bar{t}_D)$ y f' es estrictamente cóncava en $(0,\bar{t}_D)$, es decir $f' < 0$, $f'' > 0$ y $f''' < 0$ en $(0,\bar{t}_D)$.

DEFINICION 1.6.2. Un diseño $d^0 \varepsilon D$ satisface un criterio de optimalidad de tipo II si minimiza

$$\phi_f(M(d)) = \sum_{i=1}^{k-1} f(\mu_i(d))$$

en donde f es una función real valuada definida en $[0,\bar{t}_D]$ tal que:

- f es continua, estrictamente convexa y estrictamente decreciente en $[0,\bar{t}_D]$. Incluimos la posibilidad $f(0) = \lim_{x\to 0^+} f(x) = \infty$

- f es continuamente diferenciable en $(0,\bar{t}_D)$ y f'es estrictamente convexa en dicho intervalo, o sea, $f'<0$, $f''>0$, $f'''>0$ en $(0,\bar{t}_D)$

Definimos también un criterio de optimalidad generalizado de tipo I(II) como un punto límite (con la normalización necesaria para su existencia finita) de una sucesión de criterios de tipo I(II).

Observemos que la A-, D- y ϕ_p- optimalidad son criterios de optimalidad de tipo I que corresponden respectivamente a $f(x) = x^{-1}$, $-\log x$ y x^{-p}. Mientras que la E-optimalidad es un criterio generalizado de tipo I.

NOTA 4: Obsérvese que, por la sección quinta de este capítulo, si existe un diseño simétrico que maximice tr(M(d)) sobre D, entonces es óptimo con respecto a una clase muy general de criterios que incluyen los generalizados de tipo I y II.

Por otra parte, si suponemos $C\subset B_k$, las definiciones de optimalidad son las mismas sin más que escribir

$$\phi_f(M(d)) = \sum_{i=1}^{k} f(\mu_i(d))$$

1.7. SCHUR-OPTIMALIDAD.

El concepto de Schur-optimalidad que estudiaremos a continuación fue introducido por Magda (1.979).

DEFINICION 1.7.1. Una matriz con elementos no negativos es llamada doblemente estocástica si la suma de los elementos de cada fila es uno y también la suma de los elementos de cada columna.

DEFINICION 1.7.2. Un vector x de orden k×1 se dice que es mayorante del vector y de orden k×1 si

$$x_{(1)} \geq y_{(1)}$$

$$x_{(1)} + x_{(2)} \geq y_{(1)} + y_{(2)}$$

....

$$\sum_{i=1}^{k-1} x_{(i)} \geq \sum_{i=1}^{k-1} y_{(i)}$$

$$\sum_{i=1}^{k} x_{(i)} = \sum_{i=1}^{k} y_{(i)}$$

en donde $x_{(i)}$ e $y_{(i)}$ son las componentes de x e y respectivamente en orden decreciente.

TEOREMA 1.7.1. El vector x es mayorante del vector y si y sólo si existe una matriz de orden $k \times k$ doblemente estocástica S tal que $y = S.x$.

Definamos ahora la Schur-convexidad:

DEFINICION 1.7.3. Sea I un intervalo de la recta real. Una función $\phi : I^k \to R$ es llamada Schur-convexa (Schur 1.923) si

$$\phi(S.x) \leq \phi(x)$$

para todo $x \in I^k$ y cualquier matriz doblemente estocástica S.

Observemos que una función Schur-convexa no es necesariamente convexa, por ejemplo la función

$$\phi(x_1, x_2) = |x_1 - x_2|^{1/2}$$

Veamos ahora algunas propiedades de las funciones Schur-convexas

TEOREMA 1.7.2. Cualquier función Schur-convexa es simétrica.

En cuanto a la relación de la Schur-convexidad y la convexidad tenemos el teorema siguiente:

TEOREMA 1.7.3. Si ϕ es convexa y simétrica entonces es Schur-convexa.

Definiremos a continuación la Schur-optimalidad.

Supongamos ahora que $C \subset B_{k,0}$, sea $I = [0, \bar{t}_D]$ y n el menor entero para el cual $\mu_{n+1}(d) = \mu_{n+2}(d) = \ldots = \mu_k(d) = 0$ para todo $d \in D$.

Sea $\sigma(M(d))$ el siguiente vector de I^n:

$$\sigma(M(d)) = (\mu_1(d),\ldots,\mu_n(d))^t$$

definido para cada $d\varepsilon D$. Sea ϕ una función Schur-convexa definida en I^n y no creciente en sus argumentos, escribamos

$$\phi(M(d)) = \phi(\sigma(M(d))$$

la Schur-convexidad queda establecida como sigue

DEFINICION 1.7.4. Un diseño $d^0\varepsilon D$ es Schur-óptimo si d^0 minimza $\phi(M(d))$ para todo $d\varepsilon D$ y toda función Schur-convexa ϕ no creciente en sus argumentos.

Es claro que tras esta definición no todos los problemas de diseño óptimo contienen un diseño Schur-óptimo.

Nota 5: Observemos que si $f:I\to R$ es convexa, entonces

$$\phi(x) = \sum_{i=1}^{n} f(x_i) \qquad\qquad x = (x_1,\ldots,x_n)^t$$

es Schur-convexa. Por tanto, la A-, ϕ_p- y D- optimalidad son ejemplos de funciones Schur-convexas.

Nota 6: Si definimos

$$E(x_1,\ldots,x_n) = -\min_{1\leq i\leq n}\{x_1,\ldots x_n\}$$

es evidentemente una función Schur-convexa que es equivalente a la E-optimalidad. En este caso, es obvio que este criterio no ha sido obtenido como ningún tipo de criterio límite.

CAPITULO 2

DISEÑO OPTIMO Y TEORIA DE JUEGOS.

2.0. RESUMEN.

En este capítulo se estudia la caracterización de la optimalidad de un diseño mediante la utilización de la teoría de juegos. Comienza con una breve introducción de los conceptos de la teoría de juegos, por nosotros utilizados. A continuación, se analiza brevemente el método gráfico de Elfving para la resolución de ciertos problemas de diseño óptimo en regresión lineal. En la tercera sección se estudia la D-optimalidad a través de la teoría de juegos y principalmente el resultado de Karlin-Studden que supone una nueva demostración del teorema de equivalencia de Kiefer-Wolfowitz. La cuarta sección está dedicada a la D_s-optimalidad, en ella disminuímos ligeramente el material previo que Karlin-Studden utilizan para la caracterización de dicho criterio y obtenemos los mismos resultados, en esta misma sección se analiza el error cometido por estos autores al enunciar los resultados por ellos obtenidos y la modificación final de Atwood, quien estableció definitivamente el teorema. Por último, en la sección quinta aportamos un nuevo estudio sobre la D_A-optimalidad bajo el punto de vista de la teoría de juegos, caracterizamos la D_A-optimalidad por diferentes extensiones y probamos la equivalencia de todas ellas.

En consecuencia, en este capítulo se da un tratamiento unificado a la D-, D_s- y D_A- optimalidad a través de la teoría de juegos.

2.1. ALGUNAS CONSIDERACIONES SOBRE TEORIA DE JUEGOS.

Sea f(x,y) el pago que el jugador I recibe del jugador II cuando aquel selecciona la estrategia $x \in X$ y éste la estrategia $y \in Y$. La función $f(x,y)$ recibe el nombre de función de pago y el triplete (f,X,Y) recibe el nombre de juego.

DEFINICION 2.1.1. El juego definido por (f,X,Y) se dice que tiene un valor v si

$$\sup_x \inf_y f(x,y) = \inf_y \sup_x f(x,y) = v \qquad (1)$$

DEFINICION 2.1.2. $x^0 \in X$ es una estrategia óptima para el jugador I si

$$\inf_y f(x^0,y) \geq \inf_y f(x,y) \qquad \text{para todo } x \in X$$

$y^0 \in Y$ es una estrategia óptima para el jugador II si

$$\sup_x f(x,y^0) \leq \sup_x f(x,y) \qquad \text{para todo } y \in Y$$

Sabemos que ni aun en el caso en que ambos jugadores tengan un número finito de estrategias, (juego finito), la ecuación (1) es válida. No obstante, siempre se cumple

$$\sup_x \inf_y f(x,y) \leq \inf_y \sup_x f(x,y)$$

Para estudiar bajo qué condiciones un juego tiene un valor, se introduce el concepto siguiente:

DEFINICION 2.1.3. $(x^0,y^0) \in X \times Y$ es un punto de silla para el juego definido por el triplete (f,X,Y) si

$$f(x,y^0) \leq f(x^0,y^0) \leq f(x^0,y) \qquad \text{para todo } (x,y) \in X \times Y$$

A partir de la definición de punto de silla podemos enunciar el teorema siguiente

TEOREMA 2.1.1. a) Si el juego (f,X,Y) tiene un punto de silla, (x^0,y^0) entonces tiene un valor $v = f(x^0,y^0)$ y x^0 e y^0 son estrategias óptimas para los jugadores I y II respectivamente.

b) Si el juego (f,X,Y) tiene estrategias óptimas x^0, y^0 para los jugadores I y II respectivamente y tiene un valor, entonces (x^0,y^0) - es un punto de silla y $v = f(x^0,y^0)$.

Como es sabido, los juegos en los que X e Y son finitos o f es - una función continua y X×Y es compacto tienen siempre estrategias óptimas para ambos jugadores pero no tienen necesariamente puntos de silla ni valor.

Si extendemos el juego (f,X,Y) a (f^1,Γ,Ξ) en donde

$$f^1(\xi,\eta) = \int_{X\times Y} f(x,y)\xi(dx)\eta(dy) \qquad \xi\varepsilon\Gamma,\eta\varepsilon\Xi$$

siendo Γ y Ξ las clases de distribuciones de probabilidad sobre X e Y respectivamente y llamamos a (f^1,Γ,Ξ) el juego extendido de (f,X,Y),- el teorema fundamental de la teoría de juegos afirma:

TEOREMA 2.1.2. El juego extendido de un juego finito siempre tiene un valor y un punto de silla.

Diversas son las extensiones de este teorema al caso en que no - siendo (f,X,Y) un juego finito, los conjuntos X e Y o la función f poseen alguna propiedad topológica.

Enunciaremos sólo el resultado que será utilizado en este capítulo.

TEOREMA 2.1.3. Si la función de pago es continua y X×Y es compacto, - entonces el juego extendido tiene un punto de silla y un valor. Además si la función de pago es cóncava en x y convexa en y, siendo X e Y convexos compactos, ambos jugadores tienen una estrategia óptima que es una distribución de probabilidad degenerada en un punto,(estrategia óptima pura).

2.2. C-OPTIMALIDAD, METODO GRAFICO DE ELFVING Y TEORIA DE JUEGOS.

Hay ciertos problemas de diseño de experimentos en los que el in terés del experimentador se centra en la estimación de funciones específicas de algunos parámetros desconocidos más que en la estimación de dichos parámetros. Veamos el ejemplo siguiente:

EJEMPLO 2.2.1. Supongamos que la respuesta Y de cierto tipo de pacientes a una determinada droga puede ser escrita como

$$Y_x = \alpha + \beta x + e$$

en donde x es el nivel de droga administrado, e es un error aleatorio de media nula y varianza σ^2 y α y β son parámetros desconocidos. Los niveles de droga han sido tipificados de forma que $x \epsilon [-1,1]$. Queremos saber en qué N puntos del intervalo $[-1,1]$ hemos de realizar las observaciones de la variable Y_x para obtener una estimación óptima de la pendiente.

Este tipo de problemas es un caso particular del siguiente problema de diseño óptimo.

Dado el modelo lineal

$$Y_x = \theta_1 x_1 + \ldots + \theta_k x_k + e = x^t \theta + e$$

en donde e tiene media cero y varianza σ^2, independiente de x, y $x \epsilon X$, compacto de R^k, ¿qué N puntos de X seleccionaremos de forma que estimemos óptimamente $\Psi = \Sigma c_i \theta_i = c^t \theta$?

Nota 1: Estimación óptima será entendida como estimación lineal insesgada de mínima varianza.

En nuestro ejemplo anterior, los elementos del modelo lineal son

$$X = \{(x_1, x_2)^t : x_1 = 1 \; , \; x_2 \epsilon [-1,1]\} \; ; \qquad (c_1, c_2) = (0,1)^t$$

Elfving (1.952) obtiene la siguiente solución gráfica para este tipo de problemas:

SOLUCION: Sea X^1 el conjunto convexo generado por los puntos de X y de X^-, el reflejado de X en el origen. Lanzamos un rayo del origen a través del punto c. El lugar z por el que el rayo penetra en X^1 es la solución óptima en el sentido siguiente: si z es una combinación convexa de puntos x_i de $\mathcal{X}$ o $-x_i$ de $\mathcal{X}^-$, con pesos w_i, asignamos Nw_i observaciones al nivel experimental x_i. La varianza del estimador de Ψ es- $\sigma^2 || c ||^2/N|| z ||^2$.

Si Nw_i no es exacto es obvio que la solución corresponde a un diseño aproximado y hemos de buscar la solución exacta más próxima.(Ver Apéndice 1).

Resolvamos el ejemplo 2.2.1. con la ayuda del método de Elfving,(figura 2.1.). Es obvio que el diseño óptimo para el ejemplo anterior consiste en realizar el 50% de las observaciones de la variable aleatoria Y_x en el punto $(1,1)^t$ y el restante en $(1,-1)^t$.

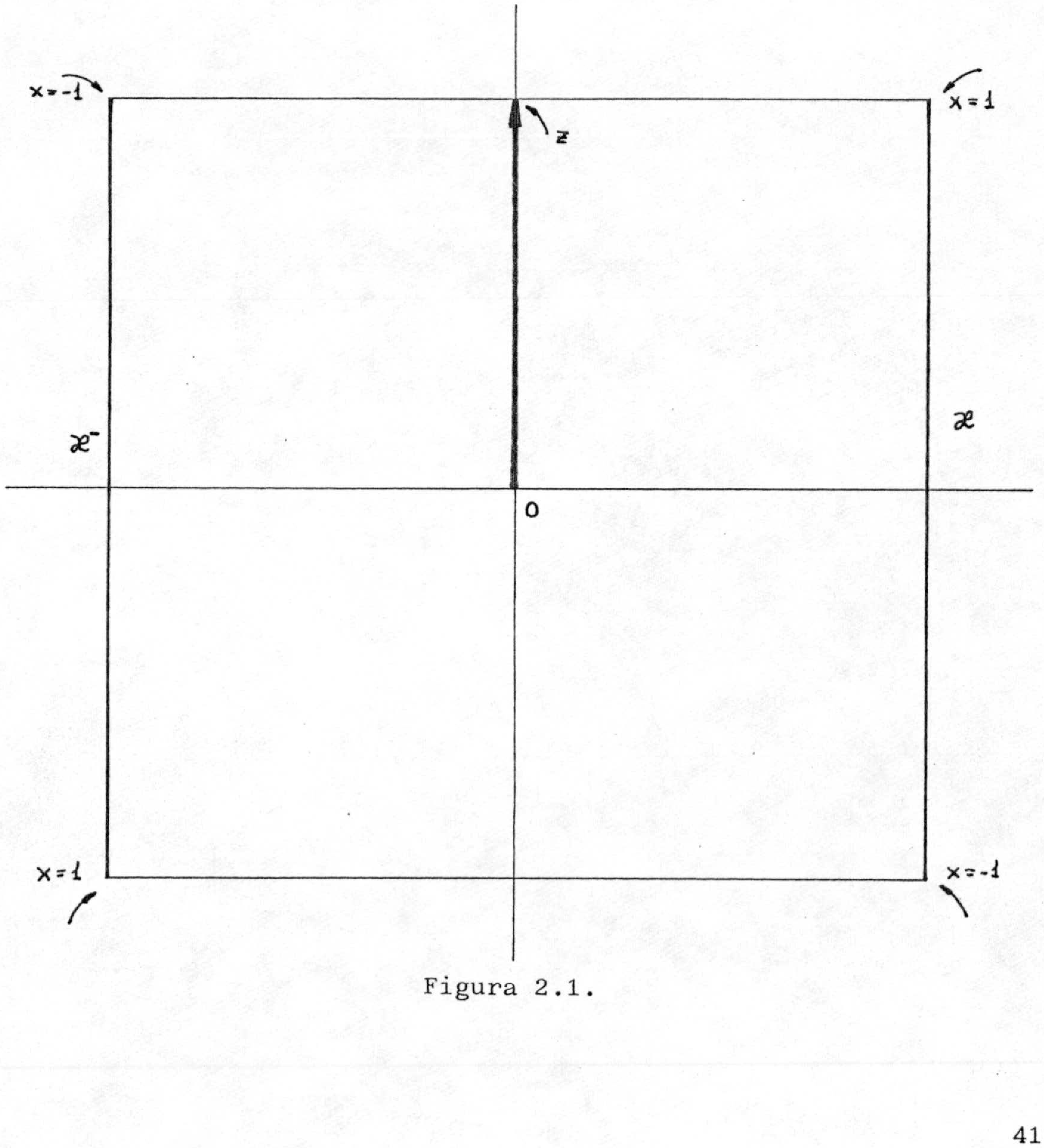

Figura 2.1.

Consideremos la siguiente variante del ejemplo anterior. Supongamos que bajo el mismo modelo, nuestro interés se centra en la estimación de $\alpha + \beta x_0$.

Vamos a analizar el caso en que x_0 es mayor que uno, los restantes casos son análogos en su estructura.

Es obvio que este problema correspondería a la estimación de la respuesta cuando por algún motivo no se puede administrar dicho nivel de droga.

La solución, como muestra la figura 2.2. es la siguiente: puesto que $z = (1/x_0,1) = \alpha(1,1) + (1-\alpha)(-1,1)$ con $\alpha= (x_0^{-1}+ 1)/2$, hemos de utilizar los puntos (1,1) y (1,-1) en la proporción $(x_0^{-1}+1)/2$ y $(1-x_0^{-1})/2$ respectivamente.

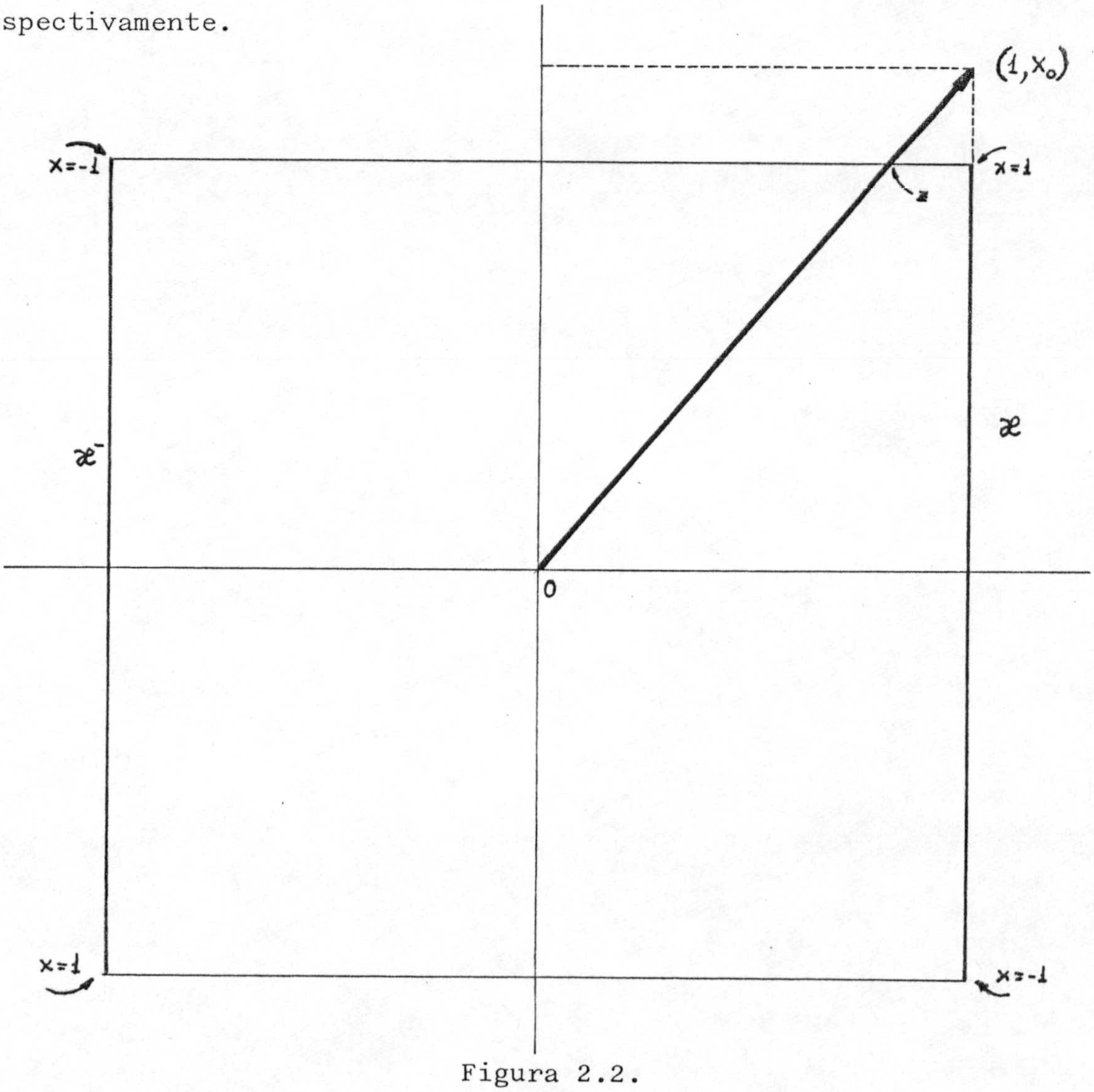

Figura 2.2.

La demostración de este teorema puede encontrarse en el artículo original de Elfving (1.952). Por otra parte, Karlin-Studden (1.966) enunciaron el resultado de Elfving, que probaron basándose en teoría de juegos, de la forma siguiente:

TEOREMA 2.2.1. Sea $R_+ = X$ y R_- la imagen simétrica de R_+, esto es, $R_- = \{-x\ ,\ x \in X\}$. Sea R la envolvente convexa de la unión de R_+ y R_-. Un diseño η^0 es c-óptimo si y sólo si existe una función real medible $\Psi(x)$ tal que $|\Psi(x)| \equiv 1$ de modo que $c^0 = \int \Psi(x) x \eta^0(dx)$ es

(a) Proporcional a c

(b) Un punto frontera de R.

Además h.c aparece en la frontera de R si y sólo si $h^2 = v_0^{-1}$ en donde $v_0 = \min_{\eta} d(c,\eta)$ siendo $d(x,\eta) = \sup_{d \in U^{\perp}, d \neq 0} [(x,d)^2/(d,M(\eta)d)]$ con $U = \{d\ :\ M(\eta)d = 0\}$.

2.3. D-OPTIMALIDAD Y TEORIA DE JUEGOS.

La idea de asignar pesos a los puntos de X para determinar el diseño c-óptimo, fue generalizada y utilizada por Kiefer para la caracterización de los mejores diseños con respecto a ciertos criterios de optimalidad.

Como puede verse en el Apéndice 1, llegado el momento de caracterizar las soluciones óptimas para ciertos criterios de optimalidad, Kiefer propone la utilización de la teoría de la aproximación y en consecuencia la consideración de los diseños como distribuciones de probabilidad sobre X, de manera que nuestro problema sea la búsqueda de distribuciones de probabilidad que optimicen cierta función Φ.

En esta línea, Kiefer y Wolfowitz (1.960) demuestran la equivalencia de dos criterios de optimalidad, a saber, la D-optimalidad y G-optimalidad utilizando cálculo diferencial.

Posteriormente, Karlin-Studden(1.966) obtienen los mismos resultados basándose en algunas consideraciones sobre teoría de juegos y proponen la caracterización de la D_s-optimalidad que ahora veremos.

2.4. D_s-OPTIMALIDAD Y TEORIA DE JUEGOS.

Como vimos en el capítulo anterior, el criterio de D_s-optimalidad para un diseño, (Karlin-Studden 1.966; Atwood 1.969), estaba motivado por el deseo de elección de N puntos de X, un compacto de R^k, que contuvieran máxima información sobre θ_1, supuesto que para cada $x \in X$ observamos la variable aleatoria Y_x tal que

$$Y_x = x_1^t \theta_1 + x_2^t \theta_2 + e$$

en donde $\theta^t = (\theta_1^t, \theta_2^t)$ es un vector paramétrico desconocido, e es una variable aleatoria de media nula y varianza σ^2; $x^t = (x_1^t, x_2^t)$ y x_1 tiene s componentes.

Es obvio que para que exista un diseño D_s-óptimo es necesario y suficiente que χ genere el subespacio s-dimensional de R^k formado por las s primeras componentes. Sin embargo, vamos a imponerle además que genere R^k; esto no es en modo alguno restrictivo, puesto que si no cumple esta segunda hipótesis podemos hacer una transformación en el modelo de forma que conservando las s primeras coordenadas, las restantes sean modificadas de manera que podamos escribir

$$Y_z = z^t \gamma + e$$

en donde γ tiene p componentes, con $k \geq p \geq s$, y $z \in Z$, un subconjunto compacto de R^p que genera R^p.

Nos proponemos caracterizar la D_s-optimalidad a través de la teoría de juegos.

Para cada diseño η sea

$$E_\eta[X.X^t] = M(\eta) = \begin{pmatrix} M_{11}(\eta) & M_{12}(\eta) \\ M_{21}(\eta) & M_{22}(\eta) \end{pmatrix}$$

en donde $M_{11}(\eta)$ es una matriz de orden $s \times s$ y $M_{12}^t(\eta) = M_{21}(\eta)$. Cuando esté claro en el contexto eliminaremos la dependencia de las submatrices del diseño η.

Sabemos que si $M(\eta)$ es no singular, la matriz de información de este diseño para θ_1 es $\sigma^{-2}M_s(\eta)$, con

$$M_s(\eta) = M_{11}(\eta) - M_{12}(\eta)M_{22}^{-1}(\eta)M_{21}(\eta)$$

El problema surge en la caracterización de dicha matriz de información cuando el diseño η tiene por submatriz $M_{22}(\eta)$ una matriz singular.

Veamos algunos resultados que nos ayudarán a definir $M_s(\eta)$ en este caso.

LEMA 2.4.1.(Karlin-Studden,1.966). Sea U el espacio vectorial de R^{k-s} generado por los vectores columna de $M_{21}(\eta)$ y V el subespacio vectorial de R^{k-s} generado por los vectores columna de $M_{22}(\eta)$, entonces

$$U \subset V$$

COROLARIO 2.4.1.(Karlin-Studden,1.966). Siempre existe una matriz X de orden (k-s)×s solución de la ecuación matricial

$$M_{22}(\eta).X = M_{21}(\eta)$$

LEMA 2.4.2.(Karlin-Studden,1.966). La matriz $X^tM_{22}(\eta)X$ es independiente de la solución de $M_{21}(\eta) = M_{22}(\eta)X$.

Se define entonces

$$M_s(\eta) = M_{11}(\eta) - X^tM_{22}(\eta)X$$

en donde X es cualquier solución de $M_{21}(\eta) = M_{22}(\eta)X$. Por el Lema2.4.2 la matriz anterior está bien definida independientemente de la solución particular utilizada.

La matriz $M_s(\eta)$ tiene las siguientes propiedades:

LEMA 2.4.3.(Karlin-Studden,1.966).

A) $M_s(\eta)$ es semidefinida positiva.

B) $M_s(\eta) - M_{11}(\eta)$ es semidefinida negativa

En el Apéndice 1, se define la estimabilidad de $A^t\theta$, por un diseño η, de la forma siguiente:

η estima $A^t\theta$ si tiene solución en Y el sistema $M(\eta)Y = A$

Tomemos $A = (I_s:0)$ y probemos el siguiente resultado personal

LEMA 2.4.4. $M_s(\eta)$ es inversible si y sólo si el sistema $M(\eta)Y = A$ admite solución en Y.

Demostración: Supongamos que $M_s(\eta)$ es no singular y sea

$$Y^t = (Y_{11}^t \quad Y_{21}^t)$$

con

$$Y_{11} = M_s^{-1}(\eta)$$
$$Y_{21} = - M_{22}^-(\eta)M_{22}(\eta)X(\eta)Y_{11}$$

siendo $X(\eta)$ cualquier solución de $M_{21}(\eta) = M_{22}(\eta)X$.

Es fácil probar que

$$M_{11}(\eta)Y_{11} + M_{12}(\eta)Y_{21} = I_s$$

$$M_{21}(\eta)Y_{11} + M_{22}(\eta)Y_{21} = 0$$

Por otra parte, sabemos que si $M(\eta)Y = A$ tiene solución en Y, la matriz de covarianzas de $\widehat{A^t\theta}$ es $\sigma^2 Y^t M(\eta)Y$ independientemente de la solución Y utilizada y además que esta matriz es definida positiva.(Searle,1.971).

Probemos que

$$Y^tM(\eta)YM_s(\eta) = I \qquad (1)$$

Puesto que

$$Y^tM(\eta) = (I_s:0)$$

el miembro izquierdo de la ecuación (1) se transforma en

$$Y_{11}^t(M_{11}(\eta) - X^t(\eta)M_{22}(\eta)X(\eta))$$

Ahora bien

$$M_{11}Y_{11} = I - M_{12}Y_{21}$$

y teniendo en cuenta que M_{11} es simétrica, podemos escribir

$$Y_{11}^t(M_{11}(\eta) - X^t(\eta)M_{22}(\eta)X(\eta)) = I - Y_{21}^tM_{21}(\eta) - Y_{11}^tX^t(\eta)M_{22}(\eta)X(\eta) =$$

$$= I - Y_{21}^tM_{22}(\eta)X(\eta) - Y_{11}^tM_{12}(\eta)X(\eta) = I - (Y_{21}^tM_{22} + Y_{11}^tM_{12})X(\eta)$$

pero

$$M_{21}Y_{11} + M_{22}Y_{21} = 0$$

de donde

$$Y_{11}^tM_s(\eta) = I$$

COROLARIO2.4.2. Si $M(\eta)$ estima θ_1, la matriz de covarianzas del mejor estimador lineal insesgado de θ_1 es $\sigma^2M_s^{-1}(\eta)$.

Sigamos estudiando algunos resultados necesarios para la caracterización de la D_s-optimalidad.

LEMA 2.4.5.(Karlin-Studden,1.966). Si $X(\eta)$ es solución de $M_{22}(\eta)X=M_{21}$ entonces

$$X^t(\eta)M_{22}X(\eta) - X^t(\eta)M_{22}D - D^tM_{22}X(\eta) + D^tM_{22}D$$

es semidefinida positiva para cualquier matriz D y es idénticamente nula si y sólo si D es solución de $M_{22}(\eta)D = M_{21}(\eta)$.

Probados estos lemas, Karlin y Studden enunciaron el siguiente resultado conocido, sobre matrices:

LEMA 2.4.6. Si C es una matriz semidefinida positiva de orden $s\times s$, entonces

$$\inf_{|P|=1} \operatorname{tr}P.C = s|C|^{1/s} \qquad (2)$$

en donde el ínfimo se extiende sobre el conjunto de todas las matri--

ces definidas positivas de orden $s \times s$ con determinante uno. Si $|C| > 0$, la igualdad ocurre en (2) si y sólo si P es proporcional a C^{-1}.

Nosotros pretendemos probar el siguiente resultado:

LEMA 2.4.7. Si C es una matriz semidefinida positiva y simétrica de orden s, entonces

$$\inf_{P \in U} \operatorname{tr} P.C = s|C|^{1/s}$$

en donde el ínfimo se extiende sobre la clase U de todas las matrices definidas positivas y simétricas P, de orden $s \times s$ y determinante uno. Si $|C| > 0$ la igualdad ocurre si y sólo si P es proporcional a C^{-1}.

Demostración: Es obvio, basándose en el lema anterior, que sólo hemos de probar que si $|C| = 0$ entonces

$$\inf_{P \in U} \operatorname{tr} P.C = 0$$

Por las propiedades de P y C es claro que $\operatorname{tr}(P.C) \geq 0$, (Bellman -- 1.965), de modo que

$$\inf_{P \in U} \operatorname{tr} P.C \geq 0$$

Por ser C semidefinida positiva y simétrica existe una matriz U-ortogonal tal que

$$U^t.C.U = \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} = D$$

en donde D_1 es una matriz diagonal de orden n, ($n < s$), tal que

$$D_1 = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_i & & \\ & & & \ddots & \\ & & & & \lambda_n \end{pmatrix} \qquad \lambda_i > 0 \ , \ i=1,2,.,n$$

En consecuencia $\operatorname{tr}(P.C) = \operatorname{tr}(U^t PUD)$

Sea $m \in N$, definimos

$$K_1 = \frac{1}{m} I_n \qquad K_2 = \begin{pmatrix} m^n & 0 \\ 0 & I_{s-n-1} \end{pmatrix}$$

es obvio que

$$U^t P_m U = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix} = K$$

tiene solución en $P_m \varepsilon U$. Por tanto

$$tr(P_m . C) = \frac{1}{m} \sum_{i=1}^{n} \lambda_i$$

de modo que

$$\inf_{P \varepsilon U} tr(P.C) = 0$$

Podemos comenzar ahora el análisis de las matrices $M_s(\eta)$. Puesto que X es compacto, se sigue del Lema.2.4.3. que

$$0 < \sup_{\eta} |M_{11}(\eta) - X^t(\eta) M_{22}(\eta) X(\eta)| = a < \infty$$

Sea $\{M(\eta_k)\}$ una sucesión de matrices de información tal que

$$\lim_{k \to \infty} |M_{11}(\eta_k) - X^t(\eta_k) M_{22}(\eta_k) X(\eta_k)| = a$$

Sin pérdida de generalidad, podemos suponer que $\{\eta_k\}$ converge débilmente a alguna distribución η^0,(Billingsley,1.979). Entonces

$$\lim_{k \to \infty} M(\eta_k) = M(\eta^0)$$

No es claro que las matrices $X(\eta_k)$ sean convergentes, sin embargo probaremos que $a = |M_s(\eta^0)|$.

En este punto, la demostración de Karlin-Studden(1.966) no es -- análoga a la que nosotros proponemos, puesto que la función que a continuación se define se hace, por nosotros, suponiendo que P es simétrica.

Sea

$$\Phi(P,D,\eta) = \mathrm{tr}\ P(M_{11} - D^t M_{21} - M_{12} D + D^t M_{22} D)$$

en donde D es una matriz arbitraria de orden $(k-s)\times s$ y P es una matriz definida positiva y simétrica con determinante a^{-1}. Denotemos por P dicha clase de matrices. Entonces

$$\Phi(P,D,\eta) = \mathrm{tr}\ P(M_{11} - X^t(\eta)M_{22}X(\eta) + X^t(\eta)M_{22}X(\eta) - D^t M_{21} - M_{12}D + D^t M_{22}D)$$

de modo que por el Lema 2.4.5.

$$\Phi(P,D,\eta) \geq \mathrm{tr}\ P(M_{11} - X^t(\eta)M_{22}X(\eta))$$

con la igualdad si y sólo si $M_{22}.D = M_{21}$. Por tanto,

$$\inf_{P\in P;D} \Phi(P,D,\eta) = \inf_{P\in P} \mathrm{tr}\ P(M_{11} - X^t(\eta)M_{22}X(\eta))$$

en consecuencia, por el Lema2.4.7.

$$\inf_{P\in P;D} \Phi(P,D,\eta) = |M_s(\eta)|^{1/s}/a^{1/s}$$

y

$$\sup_{\eta} \inf_{P\in P;D} \Phi(P,D,\eta) = s$$

finalmente, puesto que

$$\Phi(P,D,\eta^0) = \lim_{k\to\infty} \Phi(P,D,\eta_k)$$

y

$$\Phi(P,D,\eta_k) \geq s|M_s(\eta_k)|^{1/s}/a^{1/s}$$

tendremos

$$\Phi(P,D,\eta^0) \geq s$$

así pues

$$s|M_s(\eta^0)|^{1/s}/a^{1/s} = \inf_{P\in P} \Phi(P,D,\eta^0) \geq s$$

y

$$|M_s(\eta^0)| = a$$

El análisis anterior nos ayudará a demostrar el siguiente teorema:

TEOREMA 2.4.1. Sea P el conjunto de las matrices P definidas positivas y simétricas de orden s tal que $|P| = a^{-1}$ y sea D el conjunto de las matrices de orden $(k-s)\times s$. Entonces

$$\Phi(P,D,\eta) = \operatorname{tr} P(M_{11}(\eta) - D^t M_{21} - M_{12} D + D^t M_{22} D)$$

satisface la siguiente relación

$$\max_{\eta} \min_{P;D} \Phi(P,D,\eta) = \min_{P;D} \max_{\eta} \Phi(P,D,\eta)$$

además, η^0 determinado anteriormente cumple

$$\Phi(P,D,\eta^0) \geq s \qquad \text{para todo } (P,D) \in P\times D$$

si $(P_0, D_0) \in P\times D$ cumple

$$\Phi(P_0,D_0,\eta) \leq s \qquad \text{para todo } \eta \in H$$

entonces D_0 cumple $M_{21}(\eta^0) = M_{22}(\eta^0)D_0$ y P_0 es una matriz múltiplo de $(M_{11}(\eta^0) - D_0^t M_{22}(\eta^0)D_0)^{-1}$.

Demostración: Consideremos la sucesión de juegos siguiente con función de pago

$$\Psi_N(Q,E,\eta) = \operatorname{tr} (Q \quad E^t) M(\eta) \begin{pmatrix} Q \\ E \end{pmatrix}$$

en donde E denota una matriz arbitraria de orden $(k-s)\times s$ cuyos elementos están acotados en valor absoluto por N y Q es una matriz definida positiva y simétrica de orden s con sus elementos acotados en valor absoluto por N tal que $|Q| \geq a^{-1/2}$ y su autovalor menor es superior o igual a 1/N. La variable η recorre el conjunto de distribuciones de probabilidad sobre X; óbservese que en este juego, la variable η puede identificarse con la matriz de información $M(\eta)$.

Llamemos $Q_N\times E_N$ al espacio de estrategias para el jugador II que intenta minimizar su pérdida con respecto a la función de pago antes-definida.

Es claro que los conjuntos Q_N y E_N son convexos y compactos. Por otra parte, al ser X un conjunto compacto, el conjunto de distribuciones de probabilidad sobre X es compacto y convexo.

Teniendo en cuenta que Ψ_N es convexa en (Q,E) y lineal en η podemos afirmar la existencia de $(Q_N,E_N) \in Q_N \times E_N$ y $\eta_N \in H$ tales que

(3) $$\Psi_N(Q_N,E_N,\eta) \leq v_N \qquad \text{para todo } \eta \in H$$

y

(4) $$\Psi_N(Q,E,\eta_N) \geq v_N \qquad \text{para todo } (Q,E) \in Q_N \times E_N$$

para alguna constante v_N. Claramente, v_N decrece cuando N crece puesto que sólo $Q_N \times E_N$ aumenta. En consecuencia, $\{v_N\}$ está uniformemente - acotada.

Tomemos ahora η tal que $M(\eta)$ sea inversible y llamemos U_N al conjunto de estrategias óptimas para el jugador II en el juego con función de pago Ψ_N. Es claro que $\{U_N\}$ está uniformemente acotado.

Sea, entonces (Q_0,E_0) un púnto límite de elementos de U_N y $\overline{\eta^0}$ un límite débil de elementos η_N.

Las desigualdades (3) y (4) conducen a

$$\Psi(Q_0,E_0,\eta) \leq v \qquad \text{para todo } \eta \in H$$

$$\Psi(Q,E,\overline{\eta^0}) \geq v \qquad \text{para todo } (Q,E) \in Q \times E$$

en donde $v = \lim v_N$, Q es el conjunto de matrices definidas positivas y simétricas de orden s con la condición $|Q| \geq a^{-\frac{1}{2}}$ y E es el conjunto de matrices de orden $(k-s) \times s$. Obviamente, Q_0 es una matriz definida positiva y simétrica de orden $s \times s$ que cumple $|Q_0| \geq a^{-1/2}$.

Al ser $|Q_0| \geq a^{-1/2}$ escribimos $D_0 = -E_0 Q_0^{-1}$ y $P_0 = Q_0 Q_0$ y en general - $P = Q.Q$, $D = -E.Q^{-1}$ para cualquier (Q,E) $Q \times E$. Entonces

(5) $$\Phi(P_0,D_0,\eta) = \Psi(Q_0,E_0,\eta) \leq v \qquad \text{para todo } \eta \in H$$

(6) $$\Phi(P,D,\overline{\eta^0}) = \Psi(Q,E,\overline{\eta^0}) \geq v \qquad \text{para todo } (Q,E) \in Q \times E$$

Podría ocurrir que $|P_0| > a^{-1}$ puesto que $|Q_0| \geq a^{-1/2}$. Sin embargo, teniendo en cuenta que $\Phi(\lambda P,D,\eta) = \lambda\Phi(P,D,\eta)$ para $\lambda > 0$, podemos suponer que $|P_0| = a^{-1}$ pues en caso contrario, bastaría multiplicar por λ conveniente ($0 < \lambda < 1$) y la relación (5) se sigue conservando.

Las desigualdades anteriores implican que

$$\min_{P;D} \max_{\eta} \Phi(P,D,\eta) \leq \max_{\eta} \min_{P;D} \Phi(P,D,\eta)$$

y puesto que siempre se verifica la desigualdad en sentido contrario, podemos escribir

$$\max_{\eta} \min_{P;D} \Phi(P,D,\eta) = \min_{P;D} \max_{\eta} \Phi(P,D,\eta)$$

Ahora bien, teniendo en cuenta los resultados anteriores a este teorema, se verifica $v = s$ y podemos reemplazar $\overline{\eta^0}$ por η^0 en la expresión (6). Puesto que (P_0,D_0) minimiza $\Phi(P,D,\eta^0)$ se obtiene $\Phi(P_0,D_0,\eta^0)=s$ y la caracterización de P_0 y D_0 según indica el teorema es obvia a partir de los resultados anteriores al teorema.

Cuando se pretende encontrar una caracterización de los diseños-D_s-óptimos análoga al teorema de equivalencia para la D-optimalidad, hemos de definir una función equivalente a $d(x,\eta)$, construída para la D-optimalidad.

Dicha función es la siguiente

$$d_s(x,D,\eta) = (x_1 - D^t x_2)^t (M_s(\eta))^{-1} (x_1 - D^t x_2) \qquad \text{si } |M_s(\eta)| \neq 0$$

$$d_s(x,D,\eta) = \infty \qquad \text{si } |M_s(\eta)| = 0$$

en donde $x^t = (x_1^t, x_2^t)$, x_1 tiene s componentes y $M_{22}(\eta)D = M_{21}(\eta)$. Sin embargo, esta función no fue definida en un principio así por Karlin-Studden (1.966), ellos no mostraban la dependencia de la matriz D elegida y en consecuencia escribían $d_s(x,\eta)$. Enunciaron entonces el teorema siguiente:

TEOREMA 2.4.2. Las proposiciones siguientes son equivalentes

a) η^0 es D_s-óptimo

b) η^0 minimiza $\max_{x\in X} d_s(x,\eta)$

c) $\max_{x\in X} d_s(x,\eta^0) = s$

En una publicación posterior, Atwood(1.969) demostró que aunque en el caso en que $M(\eta)$ sea no singular, D es única y tiene sentido escribir $d_s(x,\eta)$, en general dicha matriz D no es única, y los valores que toma $d_s(x,D,\eta)$ no son independientes de D,(ver ejemplo3.1. de dicho artículo; en el capítulo tercero de la presente memoria pueden encontrarse otros ejemplos).

Propuso, entonces, Atwood un nuevo teorema que modifica el anterior y cuya demostración se basa en el teorema 2.4.1.. El resultado es el siguiente:

TEOREMA 2.4.3. Existe una matriz D_0 de orden $(k-s)\times s$ tal que el conjunto de diseños que cumplen a),b) y c) coinciden.

a) η^0 maximiza $|M_s(\eta)|$

b) η^0 minimiza $\max_{x\in X} d_s(x,D_0,\eta)$

c) $\max_{x\in X} d_s(x,D_0,\eta^0) = s$

Enuncia entonces el corolario siguiente que puede obtenerse de la parte correcta del teorema 2.4.2.

COROLARIO 2.4.1. Si para alguna matriz X de orden $(k-s)\times s$

$$\max_{x\in X} d_s(x,X,\eta) = s$$

entonces η es D_s-óptimo.

Este último teorema y el corolario anterior han pasado a la literatura de diseño óptimo como debidos a Karlin-Studden-Atwood. Sin embargo, en su artículo de 1.969, Atwood los enunció como debidos exclusivamente a Karlin y Studden.

2.5. D_A-OPTIMALIDAD Y TEORIA DE JUEGOS.

Como ya dijimos en el capítulo anterior, Sibson(1.974) introdujo un criterio de optimalidad destinado a la elección de N puntos de X cuando el interés del experimentador se centra en la estimación de $A^t.\theta$, siendo A una matriz de orden $k\times s$ y rango $s < k$.

El criterio de optimalidad de Sibson, llamado D_A-optimalidad, consiste en la elección de N variables de control de forma que el determinante de la matriz de covarianzas del mejor estimador lineal insesgado de $A^t.\theta$ sea mínimo.

En esta sección, vamos a suponer que los vectores columna de A se pueden expresar como combinación lineal de vectores de X, para que el problema tenga contenido, pero además supondremos que los vectores de X generan R^k, esta hipótesis no es restrictiva pues si no se cumple podemos realizar una transformación lineal sobre el modelo de forma que se cumpla la condición impuesta.

El propósito de esta sección es obtener una caracterización de la D_A-optimalidad basada en teoría de juegos. Como veremos posteriormente, esta caracterización depende de la matriz que se utilice en una transformación que habremos de realizar; estableceremos, no obstante, la relación existente entre las diferentes transformaciones.

La matriz A es de orden $k\times s$ y rango s de forma que es posible encontrar una matriz a de orden $k\times(k-s)$ tal que $H = (\ A \quad a\ )$ sea invertible, a puede ser elegida de diferentes formas. La inversa de esta matriz la escribiremos

$$H^{-1} = \begin{pmatrix} B \\ b \end{pmatrix}$$

en donde B es una matriz de orden $s\times k$ y b tiene orden $(k-s)\times k$. Es obvio que $A^t.B^t = I$ y $A^t.b^t = 0$, sin más que tener en cuenta que $H^{-1}.H = I$

Si consideramos el modelo

$$Y_x = x^t.\theta + e$$

en donde e es un error aleatorio de media nula y varianza σ^2, θ es un vector paramétrico desconocido y x es una variable de control que suponemos perteneciente a X, un subconjunto compacto de R^k.

El modelo anterior puede escribirse

$$Y_x = x^t.(H^t)^{-1}.H^t.\theta + e$$

y llamando $H^t.\theta = \gamma$ y $z = H^{-1}.x$, podemos escribir

$$Y_z = z^t.\gamma + e$$

Es obvio que $z \in Z$, el transformado por H^{-1} de X, de forma que Z es compacto. Por otra parte, si $\gamma^t = (\gamma_1^t, \gamma_2^t)$ y γ_1 tiene s componentes se cumple $A^t.\theta = \gamma_1$. En consecuencia, la estimación de $A^t.\theta$ en el modelo inicial es equivalente a la estimación de γ_1 en este nuevo modelo y el diseño D_s-óptimo para el modelo dependiente de γ es el diseño D_A-óptimo para el modelo dependiente de θ.

Llamaremos ξ a las distribuciones de probabilidad sobre Z y η denotará a las distribuciones de probabilidad sobre X.

Si η es una distribución de probabilidad sobre X y ξ es su transformada sobre Z, es claro que

$$M(\xi) = E_\xi[Z.Z^t] = E_\eta[H^{-1}.X.X^t(H^t)^{-1}] = H^{-1}M(\eta)(H^{-1})^t$$

que escribiremos como

$$M(\xi) = \begin{pmatrix} BM(\eta)B^t & BM(\eta)b^t \\ bM(\eta)B^t & bM(\eta)b^t \end{pmatrix} = \begin{pmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) \end{pmatrix}$$

Utilizando, entonces los resultados obtenidos para la D_s-optimalidad en la sección anterior y concretamente el Teorema 2.4.3. podemos escribir:

TEOREMA 2.5.1. Existe D_0 de orden $(k-s)\times s$ tal que las tres proposiciones siguientes son equivalentes

a) ξ^0 es D_s-óptimo

b) ξ^0 minimiza $\max_{z\in Z} d_s(z,D_0,\xi)$

c) $\max_{z\in Z} d_s(z,D_0,\xi^0) = s$

Ahora bien,

$$d_s(z,D,\xi) = (z_1 - D^t.z_2)^t M_s^{-1}(\xi)(z_1 - D^t.z_2)$$

pero

$$z_1 = Bx \;,\; z_2 = bx \;\; y \;\; M_s^{-1}(\xi) = (BM(\eta)B^t - C^t.b^t.M(\eta)bC)^{-1}$$

en donde C es cualquier solución de $bM(\eta)B^t = bM(\eta)b^t.C$.

Definimos, en consecuencia

$$d_A^H(x,D,\eta) = (Bx - D^t.bx)^t (BM(\eta)B^t - C^t.b^t.M(\eta)bC)^{-1}.(Bx - D^t.bx)$$

y podemos enunciar

TEOREMA 2.5.2. Existe una matriz D_0 de orden $(k-s)\times s$ de forma que las proposiciones siguientes son equivalentes

a) η^0 es D_A-óptimo

b) η^0 minimiza $\max_{x\in X} d_A^H(x,D_0,\eta)$

c) $\max_{x\in X} d_A^H(x,D_0,\eta^0) = s$

Supongamos que tomamos una matriz a_1 diferente de a, de forma -- que $H_1 = (A \;\; a_1)$ sea inversible. La pregunta que intentaremos resolver ahora es la siguiente: ¿ qué relación existe entre las diferentes transformaciones que podemos realizar sobre el espacio de control X?.

Llamemos

$$H_1^{-1} = \begin{pmatrix} B_1 \\ b_1 \end{pmatrix}$$

y sea $y = H_1^{-1}.x$ la transformación que liga al nuevo espacio de control Y con X, además las distribuciones de probabilidad sobre Y serán notadas ρ.

Sea, entonces, η una distribución de probabilidad sobre χ, ξ su distribución de probabilidad transformada sobre Z y ρ la transformada sobre Y. Sean

$$M(\xi) = \begin{pmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) \end{pmatrix} \qquad M(\rho) = \begin{pmatrix} M_{11}(\rho) & M_{12}(\rho) \\ M_{21}(\rho) & M_{22}(\rho) \end{pmatrix}$$

Probemos el siguiente resultado:

LEMA 2.5.1. C es solución de $M_{22}(\xi).C = M_{21}(\xi)$ sii $C_1 = a_1^t.b_1^t.C - a_1^t.B^t$ es solución de $M_{22}(\rho).C = M_{21}(\rho)$

Demostración: Veamos la condición necesaria. Es obvio que

$$M(\xi) = \begin{pmatrix} B \\ b \end{pmatrix} M(\eta)(B^t b^t) = \begin{pmatrix} B \\ b \end{pmatrix} H_1 . H_1^{-1} . M(\eta) . (H_1^{-1})^t . H_1^t (B^t \ b^t)$$

que puede escribirse como

$$M(\xi) = \begin{pmatrix} I_s & Ba_1 \\ 0 & ba_1 \end{pmatrix} \begin{pmatrix} B_1 M(\eta) B_1^t & B_1 M(\eta) b_1^t \\ b_1 M(\eta) B_1^t & b_1 M(\eta) b_1^t \end{pmatrix} \begin{pmatrix} I_s & 0 \\ a_1^t B^t & a_1^t b^t \end{pmatrix}$$

en consecuencia, realizando la multiplicación por cajas se tiene la siguiente relación

$$M_{21}(\xi) = ba_1 b_1 M(\eta) B_1^t + ba_1 b_1 M(\eta) b_1^t a_1^t B^t$$

$$M_{22}(\xi) = ba_1 b_1 M(\eta) b_1^t a_1^t b^t$$

Por tanto, teniendo en cuenta el Lema 2.4.1., si C es solución de

$$M_{21}(\xi) = M_{22}(\xi).C$$

podemos escribir

$$(ba_1 b_1 M(\eta) B_1^t + ba_1 b_1 M(\eta) b_1^t a_1^t B^t) = ba_1 b_1 M(\eta) b_1^t a_1^t b^t . C$$

De donde

$$(ba_1) b_1 M(\eta) B_1^t = (ba_1)(b_1 M(\eta) b_1^t)(a_1^t b^t C - a_1^t B^t)$$

Pero $b.a_1$ es una matriz no singular y podremos escribir

$$b_1 M(\eta) B_1^t = (b_1 M(\eta) b_1^t)(a_1^t b^t . C - a_1^t B^t)$$

y teniendo en cuenta que $b_1 M(\eta) B_1^t = M_{21}(\rho)$ y $b_1 M(\eta) b_1^t = M_{22}(\rho)$ el resultado queda probado.

Recorriendo los pasos en sentido inverso tenemos probada la condición suficiente.

Veamos, con las mismas hipótesis del lema anterior, el resultado siguiente:

LEMA 2.5.2. Para todo $x \varepsilon X$ se cumple

$$d_A^H(x,C,\eta) = d_A^{H_1}(x,C_1,\eta)$$

Demostración:

$$d_A^H(x,C,\eta) = (Bx - C^t bx)^t M_A^{-1}(\eta)(Bx - C^t bx)$$

y teniendo en cuenta las propiedades de la traza de una matriz podemos escribir

$$d_A^H(x,C,\eta) = \operatorname{tr} M_A^{-1}(\eta)(I \quad -C^t)\begin{pmatrix} B \\ b \end{pmatrix} x . x^t (B^t \quad b^t)\begin{pmatrix} I \\ -C \end{pmatrix}$$

Probemos ahora que

$$(B^t \quad b^t)\begin{pmatrix} I \\ -C \end{pmatrix} = (B_1^t \quad b_1^t)\begin{pmatrix} I \\ -C_1 \end{pmatrix}$$

con lo cual el resultado quedará probado.

$$(B^t \quad b^t)\begin{pmatrix} I \\ -C \end{pmatrix} = (H_1^{-1})^t H_1^t . (B^t \quad b^t)\begin{pmatrix} I \\ -C \end{pmatrix} = (B_1^t \quad b_1^t)\begin{pmatrix} I \\ a_1^t B^t - a_1^t b^t C \end{pmatrix} = (B_1^t \quad b_1^t)\begin{pmatrix} I \\ -C_1 \end{pmatrix}$$

Observemos que a partir de este lema, podemos obtener el siguiente resultado que explica la relación existente entre las diferentes transformaciones que sobre X podemos realizar.

TEOREMA 2.5.3. Sea D una matriz de orden $(k-s)\times s$ y $D_1 = a_1^t b_1^t D - a_1^t B^t$. - Entonces para cualquier $x \in X$

$$d_A^H(x,D,\eta) = d_A^{H_1}(x,D_1,\eta)$$

En consecuencia, queda establecida la D_A-optimalidad de un diseño a través de la teoría de juegos independientemente de la transformación utilizada.

CAPITULO 3

DISEÑO OPTIMO Y DUALIDAD

3.0. RESUMEN

Nos proponemos estudiar los problemas de diseño óptimo bajo el punto de vista de la teoría de la dualidad. Este estudio proporcionará una interpretación geométrica clara de los criterios de D-optimalidad y D_s-optimalidad, y un tratamiento unificado de los teoremas de equivalencia respectivos, bajo la forma de teoremas de dualidad establecidos mediante el principio del lagrangiano fuerte.

El capítulo ha sido dividido en tres secciones. En la primera se da un resumen de los conceptos sobre dualidad, por nosotros utilizados, para una mejor comprensión del tema. La segunda sección ha sido dividida en dos partes; en la primera se analizan ejemplos que justifican intuitivamente el teorema de dualidad de Sibson(1.972), que por su interés intrínseco transcribimos, en la segunda parte, obtenemos, utilizando dicho teorema, la cota de Atwood(1.969) para el cociente de los determinantes de la matriz de información de un diseño arbitrario y el diseño D-óptimo, probando además que, en general, dicha cota no es mejorable. La tercera sección tiene una estructura análoga a la segunda; en la primera parte describimos algunos ejemplos y estudiamos el teorema de dualidad de Silvey y Titterington(1.973) para la D_s-optimalidad, en la segunda parte, con la ayuda de los teoremas de dualidad, debilitamos las hipótesis de Atwood(1.969) y proponemos dos nuevas demostraciones para los resultados por él obtenidos; esta tercera sección está también ilustrada con ejemplos para facilitar su comprensión.

3.1. ALGUNOS CONCEPTOS SOBRE DUALIDAD.

Sea f una función real valuada definida sobre un subconjunto X de R^n. Nos proponemos resolver el problema P(b):

$$\begin{array}{ll} \text{Maximizar } f(x) & \\ \text{s.a:} & \\ \quad g_1(x) = b_1 & \\ \quad g_2(x) = b_2 & \qquad x \in X \\ \quad \dots\dots\dots\dots & \\ \quad g_m(x) = b_m & \end{array}$$

donde $g_1, g_2, \dots, g_m$ son funciones real valuadas definidas sobre R^n y $b_1, b_2, \dots, b_m$ son números reales. Escribimos estas restricciones como $g(x) \equiv b$, donde g(x) es un vector de funciones m-dimensional y b un vector de m constantes.

El método clásico para resolver este problema es introducir los multiplicadores de Lagrange $\lambda_1, \lambda_2, \dots, \lambda_m$ y considerar los valores estacionarios del lagrangiano

$$f(x) + \sum_{i=1}^{m} \lambda_i \{b_i - g_i(x)\} = f(x) + \lambda^t \{b - g(x)\}$$

Bajo ciertas condiciones, $\bar{x}$, valor de x que resuelve P(b), aparece, para λ conveniente, como uno de estos valores estacionarios.

Este método depende de la validez del principio del lagrangiano débil que establece: Existe un vector λ tal que el lagrangiano es estacionario en cualquier $\bar{x}$ que resuelva P(b).

Sin embargo, para algunos problemas es válido un principio más potente, el principio del lagrangiano fuerte, que establece: Existe un vector λ tal que el lagrangiano es maximal en cualquier $\bar{x}$ que resuelva P(b). Condiciones suficientes para la validez de este principio se dan en Whittle(1.971).

Cuando el principio del lagrangiano fuerte es válido, se puede decir algo más sobre el λ apropiado. Dicho λ minimiza

$$\max_{x \in X} [f(x) + \lambda^t\{b - g(x)\}]$$

Así pues, en correspondencia con el problema primal de maximización, P(b), existe un problema de minimización llamado problema dual. Estos problemas, primal y dual, comparten un valor extremo común. La validez del principio del lagrangiano fuerte implica que podemos resolver P(b) maximizando el lagrangiano sobre X sin restricciones para cada λ y encontrar a continuación el λ que minimiza el máximo resultante.

Un problema que parece diferente del anterior es $P_1(b)$:

$$\begin{array}{ll} \text{Maximizar } f(x) & \\ \text{s.a:} & \\ \quad g_1(x) \leqslant b_1 & \\ \quad g_2(x) \leqslant b_2 & \quad x \in X \\ \quad \cdots\cdots\cdots\cdots & \\ \quad g_m(x) \leqslant b_m & \end{array}$$

es decir, un problema en el que las restricciones son desigualdades en lugar de igualdades. No obstante, la diferencia no es esencial, puesto que podemos introducir variables de holgura $s_1, s_2, \ldots, s_m$ y escribir $P_1(b)$ como:

$$\begin{array}{l} \text{Maximizar } f(x) \\ \text{s.a:} \\ \quad g(x) + s = b \\ \quad x \in X, \ s \geqslant 0 \end{array}$$

El problema $P_1(b)$ es análogo a P(b) con (x,s) en lugar de x y $X \times R^m_+$ en lugar de X, donde R^m_+ es el octante no negativo de R^m.

El principio del lagrangiano fuerte asegura ahora: existe un vector λ tal que el lagrangiano

$$f(x) + \lambda^t\{b - g(x) - s\}$$

es maximal en cualquier $(\bar{x},\bar{s})$ que resuelva $P_1(b)$. Cuando esto es así-el problema dual de $P_1(b)$ es encontrar λ que minimice

$$\max_{x\in X,\, s\in R^m_+}[f(x) + \lambda^t\{b - g(x) - s\}]$$

Podemos resolver $P_1(b)$ maximizando el lagrangiano para cada λ fijo y encontrar el λ que minimice dicho máximo.

3.2. D-OPTIMALIDAD Y DUALIDAD

3.2.1. ALGUNOS EJEMPLOS Y TEOREMA DE DUALIDAD DE SIBSON.

Sea X un subconjunto compacto del espacio euclídeo k-dimensional, cuyos elementos serán vectores columna que notaremos x. Llamaremos H- a la clase de todas las distribuciones de probabilidad sobre los conjuntos de Borel de X.

Cualquier elemento $\eta\in H$, ha sido llamado una medida de diseño o- simplemente diseño. Hemos definido también para cualquier $\eta\in H$

$$M(\eta) = E[X.X^t]$$

donde X es un vector aleatorio con distribución η y E denota valor esperado. Hemos llamado a $M(\eta)$, la matriz de información asociada al diseño η.

Dada una medida de diseño $\eta\in H$, si $M(\eta)$ es no singular, definíamos

$$d(x,\eta) = x^t M^{-1}(\eta)x$$

y

$$\bar{d}(\eta) = \max_{x\in X} d(x,\eta)$$

si $M(\eta)$ fuese singular entonces $\bar{d}(\eta)$ tomaba el valor ∞

Dijimos que un diseño η^0 era D-óptimo si

$$|M(\eta^0)| = \max_{\eta\in H} |M(\eta)|$$

y era G-óptimo si

$$\bar{d}(\eta^0) = \min_{\eta\in H} \bar{d}(\eta)$$

El resultado más importante probado sobre estos criterios es el- célebre teorema de equivalencia, que prueba lo siguiente:

Teorema 3.2.1.1.(Kiefer-Wolfowitz). Son equivalentes las afirmaciones

(a) η^0 es D-óptimo

(b) η^0 es G-óptimo

(c) $\bar{d}(\eta^0) = K$

Notemos ahora por $A(\eta)$ el elipsoide k-dimensional

$$u^t M^{-1}(\eta) u \leqslant k$$

el contenido de este elipsoide es proporcional a $|M(\eta)|^{\frac{1}{2}}$ de forma que el problema de diseño óptimo, puede considerarse equivalente al de encontrar un elipsoide $A(\eta)$ de contenido máximo. Por otra parte, el teorema 3.2.1.1. nos dice que $A(\eta^0)$ contiene a X, siendo además, de este tipo de elipsoides, el único que lo contiene.

Analicemos a continuación el ejemplo siguiente:

Ejemplo 3.2.1.1. Sea X el triángulo de vértices $(1,0)^t$, $(0,1)^t$, $(-1,0)^t$. Para cada $x = (x_1, x_2)^t$ de X se define una variable aleatoria Y_x tal que

$$E(Y_x) = \theta^t . x = \theta_1 x_1 + \theta_2 x_2$$

siendo $\theta = (\theta_1, \theta_2)^t$ un vector paramétrico desconocido. Estamos interesados en estimar el vector θ, para lo cual podemos realizar observaciones en puntos x de X, dichas observaciones son incorreladas y de varianza σ^2. Nos proponemos encontrar el diseño D-óptimo para este problema.

Consideremos el diseño η^1 que asigna probabilidad 1/3 a cada uno de los vértices del triángulo; este diseño correspondería a realizar la tercera parte de las observaciones en cada uno de los vértices de X. Entonces

$$M(\eta^1) = E_{\eta^1}[X.X^t] = \frac{1}{3}\begin{pmatrix}1\\0\end{pmatrix}(1\ 0) + \frac{1}{3}\begin{pmatrix}0\\1\end{pmatrix}(0\ 1) + \frac{1}{3}\begin{pmatrix}-1\\0\end{pmatrix}(-1\ 0) = \begin{pmatrix}2/3 & 0\\0 & 1/3\end{pmatrix}$$

de forma que

$$M^{-1}(\eta^1) = \begin{pmatrix}3/2 & 0\\0 & 3\end{pmatrix}$$

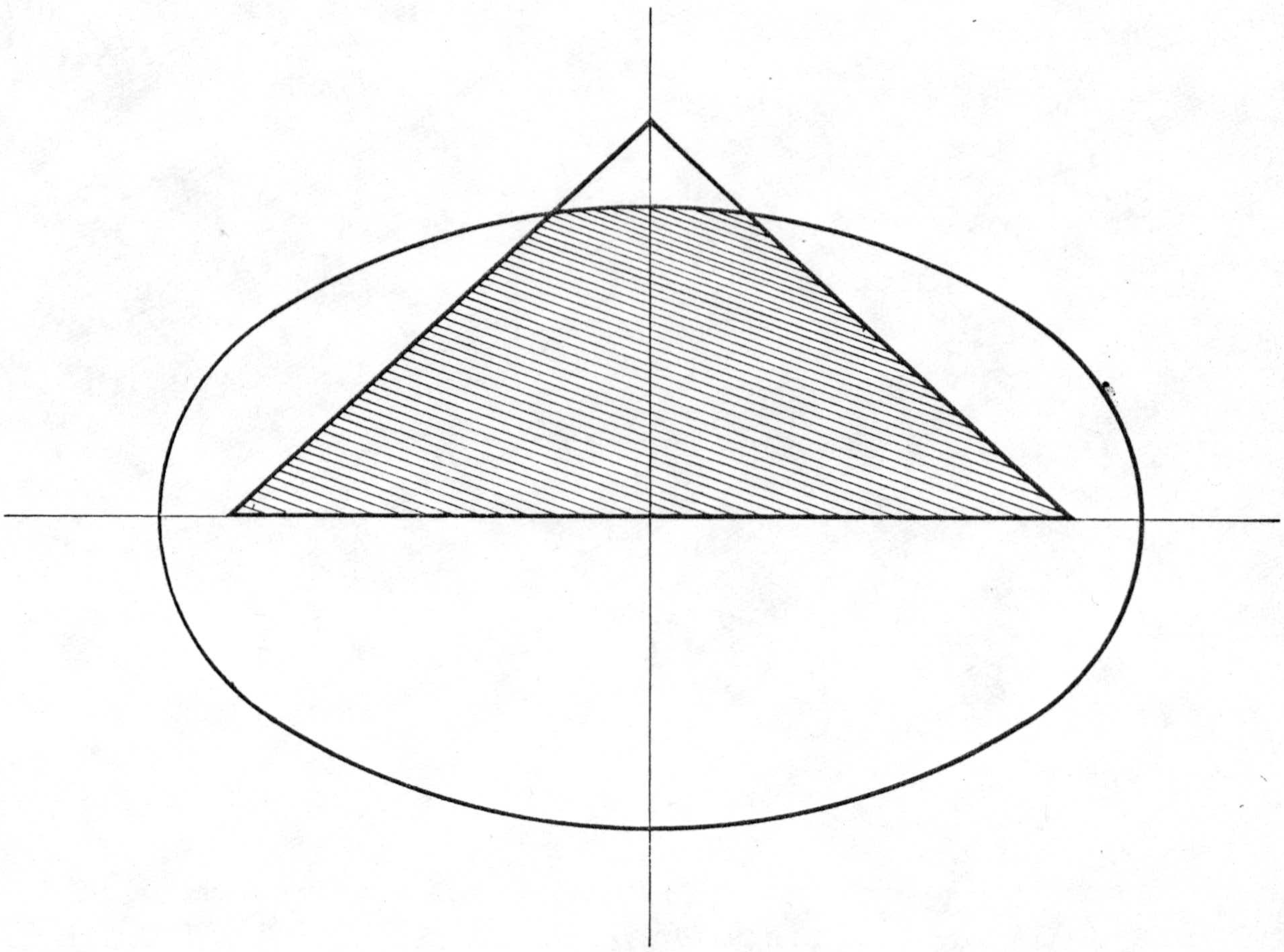

Figura 3.1.

y el elipsoide $u^t M^{-1}(\eta^1)u \leqslant 2$ no contiene a $\mathcal{X}$.(figura 3.1.).

Tomemos el diseño η^0 que asigna probabilidad 1/2 a los puntos --- $(1\ 0)^t$ y $(0\ 1)^t$ .Entonces

$$M(\eta^0) = E_{\eta^0}[X.X^t] = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

de forma que

$$M^{-1}(\eta^0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

y el elipsoide $A(\eta^0)$ puede escribirse

$$2x_1^2 + 2x_2^2 \leqslant 2$$

Este elipsoide contiene a $\mathcal{X}$, (figura 3.2.), por tanto, η^0 es el - diseño D-óptimo. Así pues, tendríamos que realizar el 50% de las observaciones en el punto $(1\ 0)^t$ y el 50% restante en el $(0\ 1)^t$.

Observemos también que $A(\eta^0)$ es el elipsoide de menor contenido -

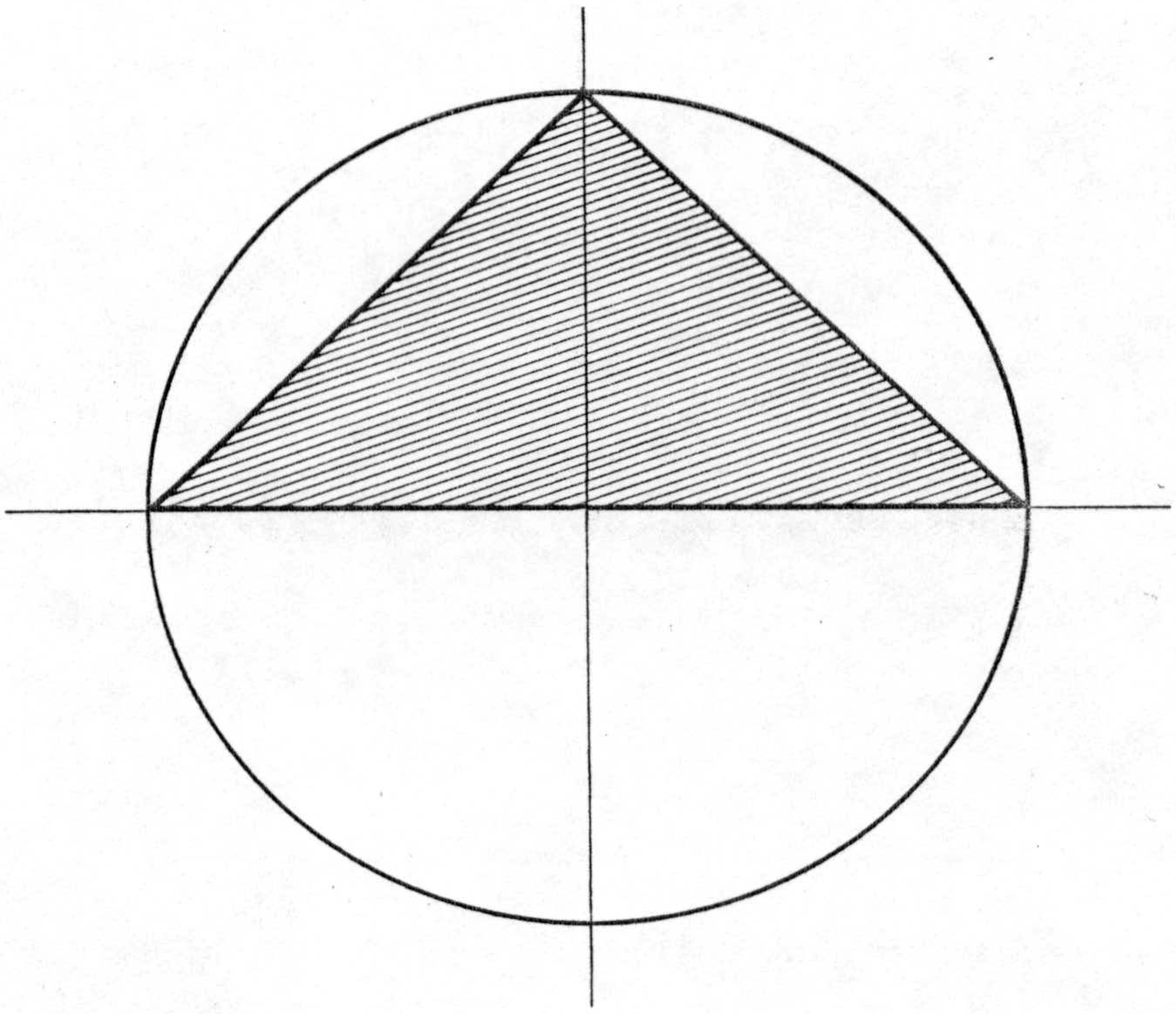

Figura 3.2.

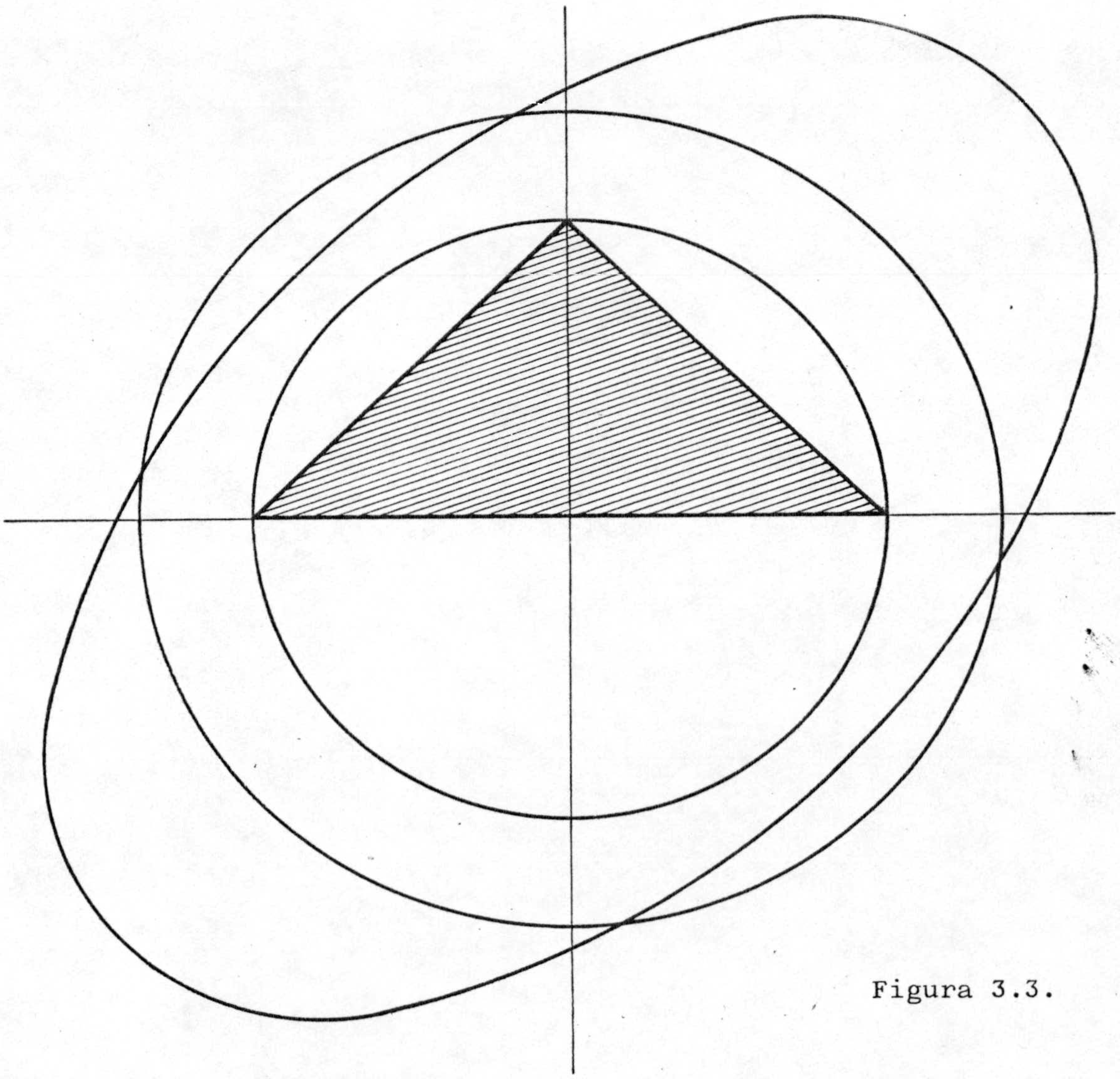

Figura 3.3.

de todos los que contienen a X, es decir, si A es una matriz definida positiva tal que $x^t Ax \leqslant 2$ para todo x de X, entonces $|A|^{-\frac{1}{2}} \geqslant |M(\eta^0)|^{\frac{1}{2}}$, o bien, $|A| \leqslant |M^{-1}(\eta^0)|$.(figura 3.3.).

El ejemplo anterior sugiere que, dado X subconjunto compacto de R^k, el problema siguiente:

$$\begin{array}{l} \text{Maximizar } \log|A| \\ \text{s.a:} \\ \qquad x^t Ax \leqslant k \quad \text{para todo } x \in X \\ \qquad A \text{ definida positiva} \end{array}$$

que llamaremos problema del elipsoide minimal, puede mantener alguna relación con el problema de diseño D-óptimo.

Dicha relación queda patente en el teorema siguiente:

Teorema 3.2.1.2.(Sibson 1.972). Si X es un subconjunto compacto de R^k, el problema de diseño D-óptimo con X como espacio de control, es el dual del problema del elipsoide minimal para X. Además, ambos problemas tienen un valor extremo común.

Demostración: Vamos a suponer, en un principio, que X es un conjunto finito y posteriormente extenderemos el problema al caso general. Hemos de hacer notar que para la existencia de diseños D-óptimos, X debe generar todo el espacio R^k pues en otro caso, no habría problema de diseño óptimo para este criterio puesto que no sería estimable todo el vector paramétrico de constantes desconocidas.

Así pues, supongamos que X es finito y está formado por los vectores $x_1, x_2, \ldots, x_N$. Nuestro problema será encontrar una matriz definida positiva A que maximice $\log|A|$, sujeto a las restricciones:

$$x_j^t A x_j \leqslant k \qquad j = 1,2,\ldots,N$$

Ahora bien, $\log|A|$ es una función cóncava sobre el cono convexo de las matrices definidas positivas y las restricciones son lineales en A. El principio del lagrangiano fuerte es válido y podemos introducir un N-vector λ de multiplicadores de Lagrange y un vector positivo o nulo s de variables de holgura de forma que existe λ tal que el lagrangiano

$$\log|A| + \sum_j \lambda_j (k - x_j^t A x_j - s_j)$$

es maximal en $(\bar{A},\bar{s})$, solución del problema del elipsoide minimal; por otra parte dicho λ minimiza el máximo respecto a A y s del lagrangiano.

Todos los elementos de λ han de ser no negativos puesto que si alguno fuese negativo el lagrangiano podría hacerse arbitrariamente grande y sabemos que para el λ apropiado el lagrangiano está acotado superiormente. Además, si λ_j es positivo, el término s_j ha de ser nulo para que sea máximo el óptimo. Así pues, nuestro problema se reduce a encontrar una matriz definida positiva que maximice

$$\log|A| - \sum_j \lambda_j x_j^t A x_j + k\Sigma\lambda_j$$

que podemos escribir

$$\log|A| - \mathrm{tr}AM(\lambda) + k\Sigma\lambda_j$$

con $M(\lambda) = \sum_j \lambda_j x_j x_j^t$

Ahora bien, $M(\lambda)$ ha de ser definida positiva pues en otro caso la expresión anterior puede hacerse arbitrariamente grande.

Sabemos también que si A y B son matrices semidefinidas positivas de orden p, se tiene

$$|A.B| \leqslant [\mathrm{tr}(A.B)/p]^p$$

con la igualdad si y sólo si $A.B = \mu.I$, siendo μ una constante arbitraria no negativa. De modo que, para que sea máximo en A

$$\log|A| - \mathrm{tr}A.M(\lambda) + k\Sigma\lambda_j$$

se escribirá

$$\log\mu^k |M^{-1}(\lambda)| - \mathrm{tr}\mu I + k\Sigma\lambda_j$$

o bien

$$\log|M^{-1}(\lambda)| + k\log\mu - k\mu + k\Sigma\lambda_j$$

pero $k\log\mu - k\mu$ es máximo en $\mu = 1$, por tanto se cumplirá $A = M^{-1}(\lambda)$

y la expresión anterior queda

$$\log|M^{-1}(\lambda)| + k(\Sigma\lambda_j - 1)$$

Veamos ahora que $\Sigma\lambda_j = 1$ para el valor de λ que minimiza la expresión anterior.

Si suponemos $\Sigma\lambda_j \neq 1$ escribimos entonces $\alpha_j = \lambda_j/\Sigma\lambda_j$ para todo j. Entonces

$$-\log|M(\alpha)| + k(\Sigma\alpha_j - 1) = -\log|M(\alpha)| = -\log|M(\alpha)| + k\log\Sigma\alpha_j <$$

$$-\log|M(\lambda)| + k(\Sigma\lambda_j - 1)$$

Puesto que $\log z < z - 1$, con la igualdad sólo si $z = 1$. De modo que λ verificará $\Sigma\lambda_j = 1$.

El problema del elipsoide minimal tiene, por tanto, un dual que se reduce a encontrar λ_j no negativos de suma la unidad que minimicen la cantidad $-\log|M(\lambda)|$. Este es evidentemente el problema de diseño D-óptimo.

Extendamos el resultado al caso de ser X un compacto arbitrario de R^k. Sabemos que existe un subconjunto finito de X que llamaremos X_1 en el que el problema de diseño D-óptimo es el mismo que en X. Dicho de otra forma, escribimos el diseño D-óptimo como una combinación lineal convexa de un número finito de diseños concentrados en un punto único, a ese conjunto de puntos lo llamaríamos X_1, y evidentemente el diseño D-óptimo de X coincide con el de X_1. Sea $\bar{\lambda}$ el diseño D-óptimo para X_1 y por tanto para X.

Supongamos ahora que el elipsoide mínimo para X_1 no contiene a X de modo que añadiendo un punto conveniente a X_1 encontraríamos un subconjunto finito de X para el que el elipsoide mínimo sería mayor que para X_1. Por la dualidad establecida para subconjuntos finitos esto implicaría la existencia de una medida de diseño λ sobre X tal que $-\log M(\lambda)$ fuese más pequeño que $-\log M(\bar{\lambda})$. Lo cual es obviamente una contradicción.

Observemos que el teorema de dualidad implica que η^0 es óptimo - si y sólo si $x^t M^{-1}(\eta^0)x \leqslant k$ para todo x de X y se da la igualdad en- los puntos soporte de η^0.

3.2.2. APORTACIONES A LA COTA DE ATWOOD SOBRE LA D-OPTIMALIDAD.

Nos proponemos, ahora, utilizar la teoría de la dualidad para establecer una nueva demostración para la cota de Atwood(1.969) que relaciona la proximidad de la matriz de información de un diseño η al - óptimo con la cercanía de $\bar{d}(\eta)$ a k, siendo k el número de parámetros- a estimar. Probaremos que esta cota mejora estrictamente la dada por- Kiefer(1.961) y que en general no admite mejora

El soporte intuitivo del resultado es el siguiente: aunque para- la matriz de información de un diseño η, $M(\eta)$, no será válido en general que $x^t M^{-1}(\eta)x \leqslant k$ para todo x de X, lo que sí será cierto es que- multiplicando por una constante conveniente α, se cumplirá que es válida la relación $x^t(\alpha M^{-1}(\eta))x \leqslant k$ para todo x de X; la existencia de- α está garantizada por ser X compacto. Entonces, $(\alpha^{-1}M(\eta))^{-1}$ es una- solución factible, no necesariamente óptima, del problema del elipsoide minimal y podremos entonces relacionar su determinante, y por tanto el de $M(\eta)$, con el determinante del diseño D-óptimo, pues ambos -- problemas comparten un valor extremo.

Estudiemos estas ideas en el ejemplo siguiente.

Ejemplo 3.2.2.1. Sea el modelo de regresión

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

donde e es un error aleatorio de media cero y varianza σ^2, $x^t = (x_1, x_2)$ es una variable de control que el experimentador puede tomar en el -- compacto $X = \{x = (x_1, x_2)^t : |x_i| < 1,\ i=1,2\}$. Estamos interesados en encontrar el diseño D-óptimo para lo cual podemos realizar observaciones de la variable aleatoria Y_x, que supondremos incorreladas.

Si consideramos el diseño η con la matriz de información

$$M(\eta) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

es decir, un diseño que asigna probabilidad 1/2 a los puntos $(1,0)^t$ y $(0,1)^t$, entonces el elipsoide

$$x^t \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} x \leqslant 2$$

no contiene a X.(figura3.4).

Sin embargo, si multiplicamos $M(\eta)$ por cuatro se tendrá que el - elipsoide

$$x^t(4M(\eta))^{-1}x = x^t \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} x \leqslant 2$$

contiene a X.(figura3.4).

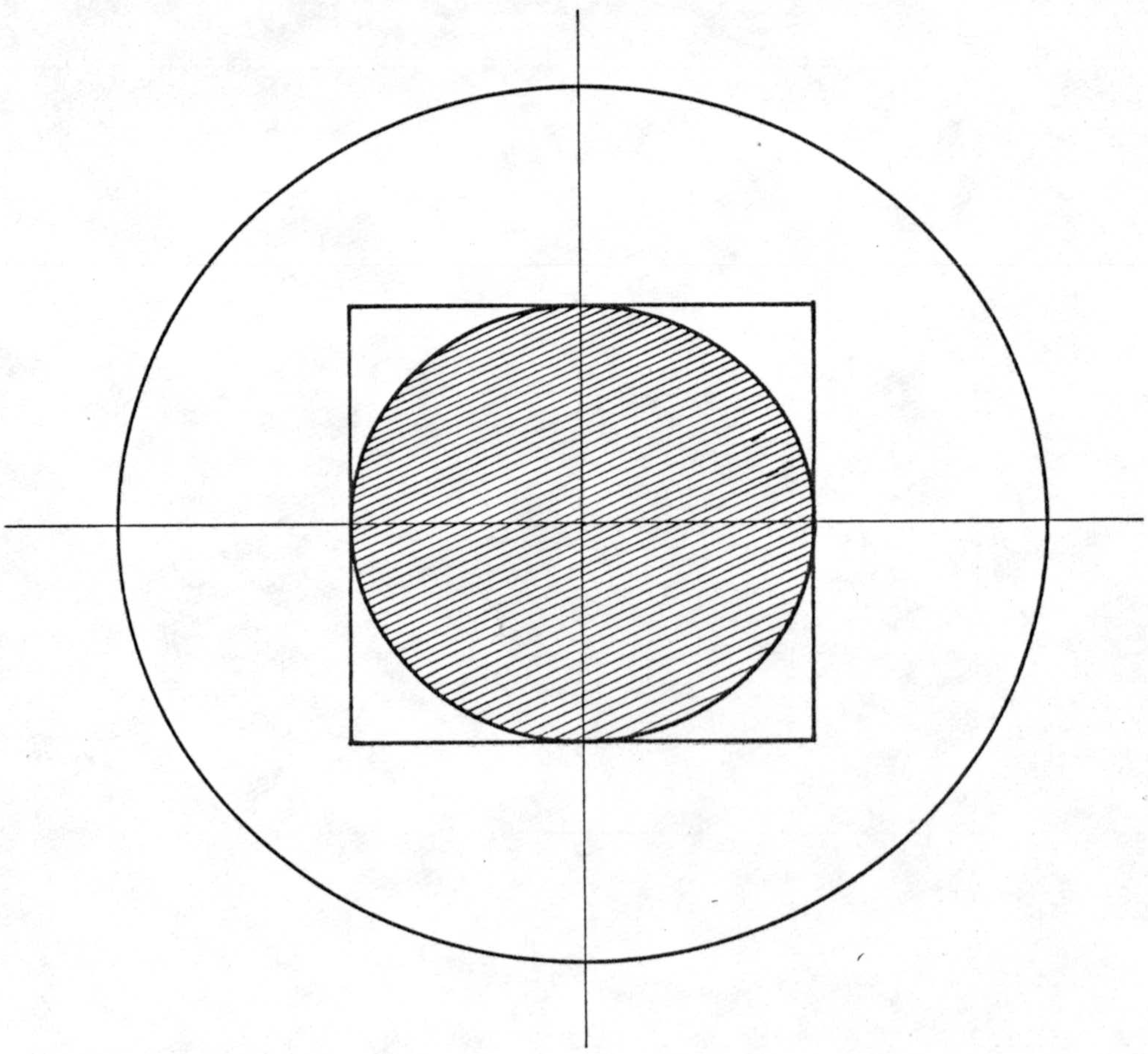

Figura 3.4.

De modo que por el teorema de dualidad de Sibson

$$4^{-2}|M^{-1}(\eta)| \leq |M^{-1}(\eta^0)|$$

siendo η^0 el diseño D-óptimo para el problema por nosotros considerado. Hemos podido, por tanto, relacionar los determinantes de las matrices de información de un diseño arbitrario y el óptimo. Esta relación es bastante pobre , pero utilizando una constante α apropiada vamos a obtener la cota de Atwood.

Teorema 3.2.2.1.(Atwood 1.969). Sea $\mathcal{X}$ un conjunto compacto incluido en R^k y η^0 un diseño D-óptimo para el problema de regresión asociado a $\mathcal{X}$. Entonces dado un diseño η con matriz de información no singular $M(\eta)$, se cumple

$$|M(\eta)|/|M(\eta^0)| \geq [k/\bar{d}(\eta)]^k$$

en donde $\bar{d}(\eta) = \max\limits_{x\in\mathcal{X}} x^t M^{-1}(\eta)x$

Demostración: Veamos como, con el teorema de Sibson, se puede obtener este resultado de una forma muy sencilla.

Para cualquier x de $\mathcal{X}$, dado un diseño η cuya matriz de información no singular es $M(\eta)$, se tiene:

$$x^t M^{-1}(\eta)x \leq \bar{d}(\eta)$$

$\bar{d}(\eta)$ es un número positivo, de forma que tendremos

$$x^t\left(\frac{k}{\bar{d}(\eta)}M^{-1}(\eta)\right)x \leq k$$

que podemos escribir

$$x^t\left(\frac{\bar{d}(\eta)}{k}M(\eta)\right)^{-1}x \leq k$$

ahora bien, la matriz de la forma cuadrática anterior es definida positiva. Además, dicha matriz es solución factible del problema del elipsoide minimal, de modo que tendremos

$$\left|\left(\frac{\bar{d}(\eta)}{k}M(\eta)\right)^{-1}\right| \leq |M^{-1}(\eta^0)|$$

que escribiremos

$$\left(\frac{k}{\bar{d}(\eta)}\right)^k |M(\eta)|^{-1} \leq |M(\eta^0)|^{-1}$$

o bien

$$\frac{|M(\eta)|}{|M(\eta^0)|} \geq \left(\frac{k}{\bar{d}(\eta)}\right)^k$$

con lo que queda demostrado el resultado de Atwood utilizando los resultados obtenidos por Sibson.

Kiefer en 1.960 estableció la siguiente desigualdad

$$\frac{|M(\eta)|}{|M(\eta^0)|} \geq \exp[k - \bar{d}(\eta)]$$

Nos preguntamos, ahora, si existe alguna relación entre la desigualdad propuesta por Kiefer y la propuesta por Atwood.

Vamos a demostrar ahora que es mejor la cota propuesta por Atwood

Proposición 3.2.2.1. Si $\bar{d}(\eta) \geq k$, entonces

$$\left(\frac{k}{\bar{d}(\eta)}\right)^k \geq \exp[k - \bar{d}(\eta)]$$

con la igualdad sólo si $\bar{d}(\eta) = k$

Demostración: Aplicando logaritmos a ambos miembros tendríamos que demostrar

$$k\log k - k\log\bar{d}(\eta) \geq k - \bar{d}(\eta)$$

o bien

$$\bar{d}(\eta) - k\log\bar{d}(\eta) \geq k - k\log k$$

con la igualdad sólo si $\bar{d}(\eta) = k$

Consideremos ahora la función siguiente

$$f(x) = x - \log x^k$$

de forma que podemos afirmar si $x \geq k$

$$x - k\log x \geq k - k\log k$$

y podemos afirmar que salvo si $\bar{d}(\eta) = k$

$$[k/\bar{d}(\eta)]^k > \exp[k - \bar{d}(\eta)]$$

Ahora bien, si $\bar{d}(\eta) = k$ entonces η es un diseño D-óptimo y coinciden ambas cotas, las dos toman el valor uno.

Recordando que para cualquier diseño η, que no sea D-óptimo, se cumple $\bar{d}(\eta) > k$, entonces teniendo en cuenta el resultado anterior, podemos afirmar que la cota propuesta por Atwood mejora estrictamente la dada por Kiefer.

Nos preguntamos ahora si la cota inferior propuesta por Atwood para el cociente de los determinantes de la matriz de información de un diseño arbitrario η y el determinante de un diseño, η^0, D-óptimo es mejorable en el sentido siguiente: ¿ Existe en cualquier problema de diseño D-óptimo, bajo las hipótesis usuales, un diseño η^1 tal que

$$|M(\eta^1)|/|M(\eta^0)| > [k/\bar{d}(\eta^1)]^k \text{ ?}$$

La respuesta es negativa como veremos a continuación

Ejemplo 3.2.2.2. Consideremos el siguiente modelo de regresión lineal

$$Y_x = \theta_1 x + e$$

en donde θ_1 es un parámetro desconocido a estimar, e un error aleatorio de media nula y varianza σ^2, independiente en cada observación, y x es una variable de control que supondremos perteneciente al intervalo unidad, es decir $X = [0,1]$. Nos proponemos encontrar el diseño que mejor estime θ_1 a través de una combinación lineal de las observaciones, es decir, el diseño D-óptimo.

Para cualquier diseño η

$$M(\eta) = E_\eta[X^2] = \int x^2 dF_\eta(x)$$

de forma que

$$M(\eta) = \sum_{j=1}^{N} \lambda_j x_j^2 \qquad \text{con } \lambda_j \geq 0,\ \Sigma\lambda_j = 1,\ x_j \in [0,1]$$

en donde hemos utilizado la representación de $M(\eta)$ como una combinación lineal convexa de un número finito de diseños puntuales.

Evidentemente

$$|M(\eta)| \leq 1$$

de forma que el diseño η^0 que asigna probabilidad uno al punto $x = 1$, es el D-óptimo, pues $|M(\eta^0)| = 1$. Este diseño consiste en realizar todas las observaciones en el punto $x = 1$.

Por otra parte, para cualquier diseño η con matriz de información no singular se tendrá

$$\bar{d}(\eta) = \sup_{x\in[0,1]} xM^{-1}(\eta)x = \sup_{x\in[0,1]} \frac{x^2}{(\Sigma\lambda_j x_j^2)} = \frac{1}{(\Sigma\lambda_j x_j^2)}$$

Ahora bien,

$$\frac{|M(\eta)|}{|M(\eta^0)|} = \sum_{j=1}^{N}\lambda_j x_j^2$$

y como $k = 1$

$$\left(\frac{k}{\bar{d}(\eta)}\right)^k = \sum_{j=1}^{N}\lambda_j x_j^2$$

de manera que para cualquier diseño η

$$\frac{|M(\eta)|}{|M(\eta^0)|} = \left(\frac{k}{\bar{d}(\eta)}\right)^k$$

por tanto podemos enunciar la siguiente proposición

Proposición 3.2.2.2. Para un problema general de diseño D-óptimo, la cota inferior $(k/\bar{d}(\eta))^k$ para el cociente de los determinantes de la matriz de información de un diseño arbitrario η y el óptimo, no es mejorable.

3.3. D_s-OPTIMALIDAD Y DUALIDAD.

3.3.1. EJEMPLOS Y TEOREMA DE DUALIDAD.

Si estamos interesados en la estimación de $s < k$ parámetros, siguiendo con el modelo de regresión propuesto en la sección 3.2., un criterio de optimalidad para este problema, análogo en su estructura al criterio D-óptimo para la estimación de los k parámetros, es lo que hemos definido en capítulos anteriores como D_s-optimalidad.

Si realizamos la partición de la matriz de información de un diseño η

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

en donde, por comodidad de notación, no escribiremos la dependencia de cada submatriz del diseño η considerado, siendo M_{11} una matriz de orden $s \times s$ y M_{12} la traspuesta de M_{21}, la D_s-optimalidad estudia, en una primera aproximación aquellos diseños que tienen por M_{11} una matriz definida positiva. Recordemos que para el estudio de la D-optimalidad nos ocupábamos de las matrices $M(\eta)$ que eran definidas positivas, sin embargo en este problema la matriz M_{22} puede ser singular. Esta posibilidad complica considerablemente el problema.

Para analizar, brevemente, la D_s-optimalidad seguimos Karlin-Studden(1.966).

Sea $D(\eta)$ la clase de matrices C de orden $s \times (k-s)$ que cumplen

$$C.M_{22} = M_{12}$$

estos autores probaron:

(a) La clase $D(\eta)$ es no vacía.

(b) $CM_{22}C^t$ es independiente de la solución C utilizada.

(c) $M_{11} - CM_{22}C^t$ es definida no negativa.

Se define entonces

$$M_s(\eta) = M_{11} - CM_{22}C^t$$

observemos que esta matriz $M_s(\eta)$ es independiente de la elección particular de C en $D(\eta)$.

Decimos entonces que η^0 es D_s-óptimo si

$$|M_s(\eta^0)| = \max_{\eta \in H} |M_s(\eta)|$$

Si quisiéramos tener un paralelismo exacto con el teorema 3.2.1.1-para este nuevo problema, estaríamos obligados a definir un criterio -de optimalidad que se llamase G_s-optimalidad. Esta definición puede hacerse pero carece de una fácil interpretación, sólo el teorema de dualidad de Silvey-Titterington(1.973) lo aclara un poco, por lo que prescindiremos de este concepto.

El resultado más similar al teorema 3.2.1.1. requiere las siguientes definiciones:

Particionamos x^t en (x_1^t, x_2^t) donde x_1 tiene s componentes. Sea D -cualquier matriz de orden s×(k-s) y $M_s(\eta)$ no singular. Definimos

$$d_s(x,\eta,D) = (x_1 - Dx_2)^t M_s^{-1}(\eta)(x_1 - Dx_2)$$

y

$$\bar{d}_s(\eta,D) = \max_{x \in X} d_s(x,\eta,D)$$

si $M_s(\eta)$ es singular, entonces, $\bar{d}_s(\eta,D)$ lo hacemos ∞

Tenemos según las definiciones anteriores el siguiente teorema

Teorema 3.3.1.1.(Atwood-Karlin-Studden). Existe una matriz C^0 tal que-las afirmaciones siguientes son equivalentes:

(a) η^0 es D_s-óptimo

(b) η^0 minimiza $\bar{d}_s(\eta,C^0)$

(c) $\bar{d}_s(\eta^0,C^0) = s$

Si estudiamos detenidamente los enunciados de los teoremas 3.2.1.1. y 3.3.1.1., encontramos bastante analogía entre ellos. Sin embargo, la demostración de Kiefer-Wolfowitz utiliza como herramienta fundamental-el cálculo diferencial, mientras que este último resultado se obtiene-de una forma mucho más complicada utilizando teoría de juegos.

En 1.972 Silvey sugería, en un comentario sobre los artículos de-

Laycock(1.972) y Wynn(1.972), la posibilidad de que ambos problemas tuvieran una estructura análoga, basándose en consideraciones de tipo geométrico. Sibson(1.972), como hemos visto, obtuvo el teorema de Kiefer y Wolfowitz como un corolario del teorema de dualidad por él propuesto. Como veremos ahora, Silvey y Titterington(1.973) obtuvieron un resultado análogo al de Sibson para la D_s-optimalidad.

Antes de enunciar el teorema de Silvey y Titterington, sigamos a ambos autores para estudiar la génesis de su resultado.

Se define para $u \in R^k$ la siguiente función, siendo η un diseño de H y D una matriz de orden $s \times (k-s)$.

$$\delta_s(u,\eta,D) = (u_1 - Du_2)^t M_s^{-1}(\eta)(u_1 - Du_2)$$

supuesto que $M_s(\eta)$ es no singular y que u^t ha sido dividido en (u_1^t, u_2^t) siendo s el número de componentes de u_1.

Evidentemente,

$$\{u \in R^k : \delta_s(u,\eta,D) \leq s\}$$

es un cilindro en R^k que notaremos $C(\eta,D)$.

Es sabido, Karlin-Studden(1.966), que

$$E_\eta[\delta_s(u,\eta,D)] \geq s$$

y se da la igualdad si y sólo si D pertenece a $D(\eta)$. Obviamente, si D no pertenece a $D(\eta)$ ha de existir $x \in X$ de forma que $\delta_s(x,\eta,D) > s$. Pero si $M_{12} = DM_{22}$, se da la igualdad en la expresión anterior y por tanto el cilindro $C(\eta,D)$ interseca a X.

Existe, evidentemente, la posibilidad de que un cilindro de la forma $C(\eta,C)$, con C perteneciente a $D(\eta)$ contenga a X. Claramente, el teorema 3.3.1.1. puede interpretarse diciendo que al menos existe uno y que tal cilindro maximiza $|M_s(\eta)|$.

Ahora bien, el contenido de la sección del cilindro $C(\eta,D)$ por el subespacio $u_2 = 0$ es proporcional a $|M_s(\eta)|^{\frac{1}{2}}$ y tenemos la interpre tación de que cualquier cilindro $C(\eta,C)$ que contenga a X, es el cilindro más grueso de los de esta forma.

Se introduce ahora un problema de minimización que llamaremos el problema del cilindro de mínimo contenido.

Sea A una matriz definida positiva de orden s y B una matriz de orden $s\times(k-s)$. El problema del cilindro de mínimo contenido es: encontrar el cilindro

$$(u_1 - Bu_2)^t A(u_1 - Bu_2) \leq s$$

que contega a X y que maximice $|A|$. Observemos que la sección de este cilindro por $u_2 = 0$ es el elipsoide

$$u_1^t Au_1 \leq s$$

cuyo contenido es proporcional a $|A|^{-\frac{1}{2}}$. Así pues, maximizar $|A|$ es lo mismo que minimizar el contenido de esta sección elipsoidal. De aquí el nombre del problema.

Veamos, entonces, el teorema de dualidad de Silvey-Titterington.

Teorema3.3.1.2.(Silvey-Titterington 1.973). Sea X cualquier subconjunto compacto de R^k que genera el subespacio s-dimensional formado por las s primeras coordenadas. Entonces, el problema de encontrar el diseño D_s-óptimo para X, es el dual del problema del cilindro de mínimo contenido para este mismo conjunto. Ambos problemas comparten un valor extremo.

Nos proponemos ahora clarificar la importancia de la matriz C^0 que aparece en el teorema 3.3.1.1.. Para ello vamos a estudiar varios ejemplos en los que la matriz C^0 aparece de forma diferente.

Ejemplo 3.3.1.1. Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde $x^t= (x_1,x_2)$ actúa como una variable de control que el experimentador, al realizar las observaciones de la variable Y_x, puede tomar-en el triángulo de vértices $(1,0)^t$, $(0,0)^t$ y $(1,1)^t$, θ_1 y θ_2 son parámetros desconocidos y e es un error de media nula y varianza σ^2. Estamos interesados en estimar θ_1, y nuestras elecciones de los puntos del espacio de control queremos que hagan dicha estimación D_1-óptima.

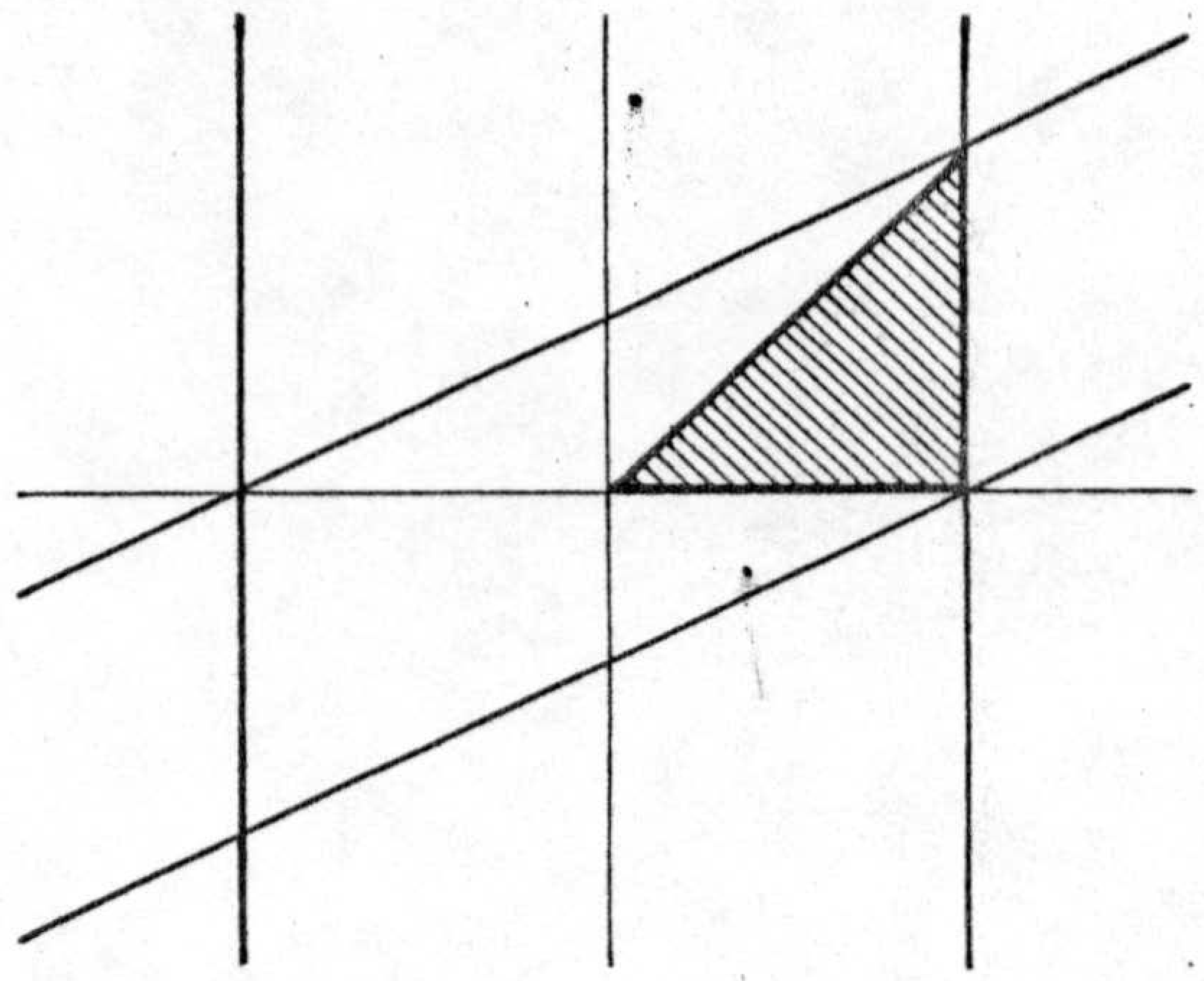

Figura 3.5.

Consideremos el diseño η^0 que realiza todas las observaciones de Y_x en el punto $(1,0)^t$. Su matriz de información será

$$M(\eta^0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

De forma que la ecuación $M_{12} = CM_{22}$ admite como solución en C cualquier número real. Si consideramos el cilindro $C(\eta^0,C)$, para este problema toma la forma

$$(u_1 - Cu_2)^2 \leq 1$$

de modo que si C pertenece al intervalo [0,2] el cilindro $C(\eta^0,C)$ contiene a X, (figura 3.5), pero si C no pertenece a dicho intervalo el cilindro anterior no contiene a X. (figura 3.6).

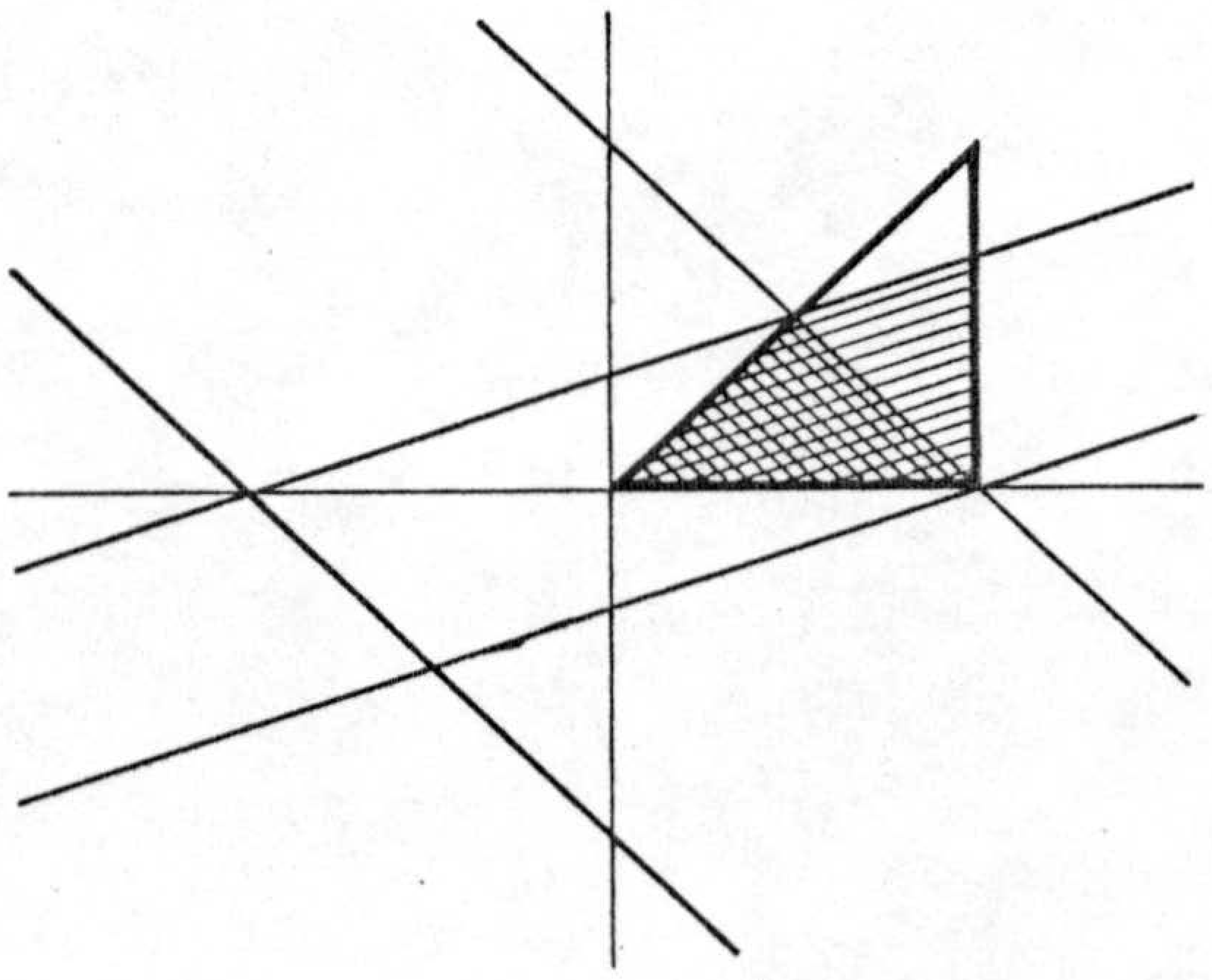

Figura 3.6.

Así pues, en este ejemplo vemos que C^0 no es única pero existe para el diseño η^0. En consecuencia, según este criterio la mejor forma de estimar θ_1 es realizar todas las observaciones de la variable aleatoria Y_x en el punto $(1,0)^t$.

Ejemplo 3.3.1.2. Supongamos ahora que considerando el mismo modelo lineal que en el ejemplo anterior y la estimación de θ_1, el espacio de control es el triángulo de vértices $(0,0)^t$, $(1,1)^t$ y $(1.-1)^t$.

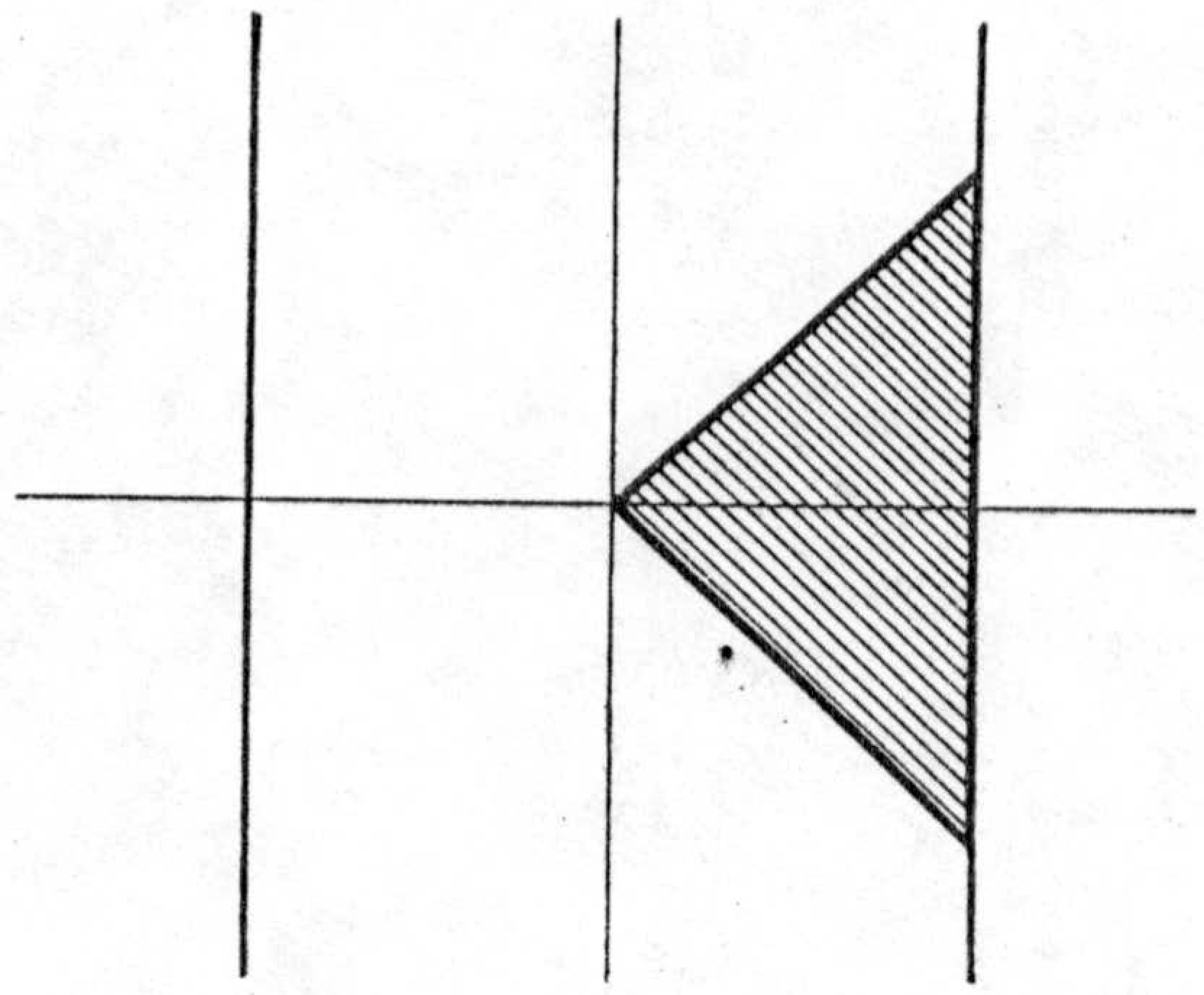

Figura 3.7.

Sigue pareciendo lógica la consideración del diseño η^0 del problema anterior como D_1-óptimo.

En este caso, la familia de cilindros $C(\eta^0,C)$, en donde C es un número real, se escribirá $(u_1 - Cu_2)^2 \leq 1$; sólo contendrá a X cuando C = 0, (figura 3.7.), pues para cualquier real no nulo su cilindro asociado nunca contendrá a X. (figura 3.8).

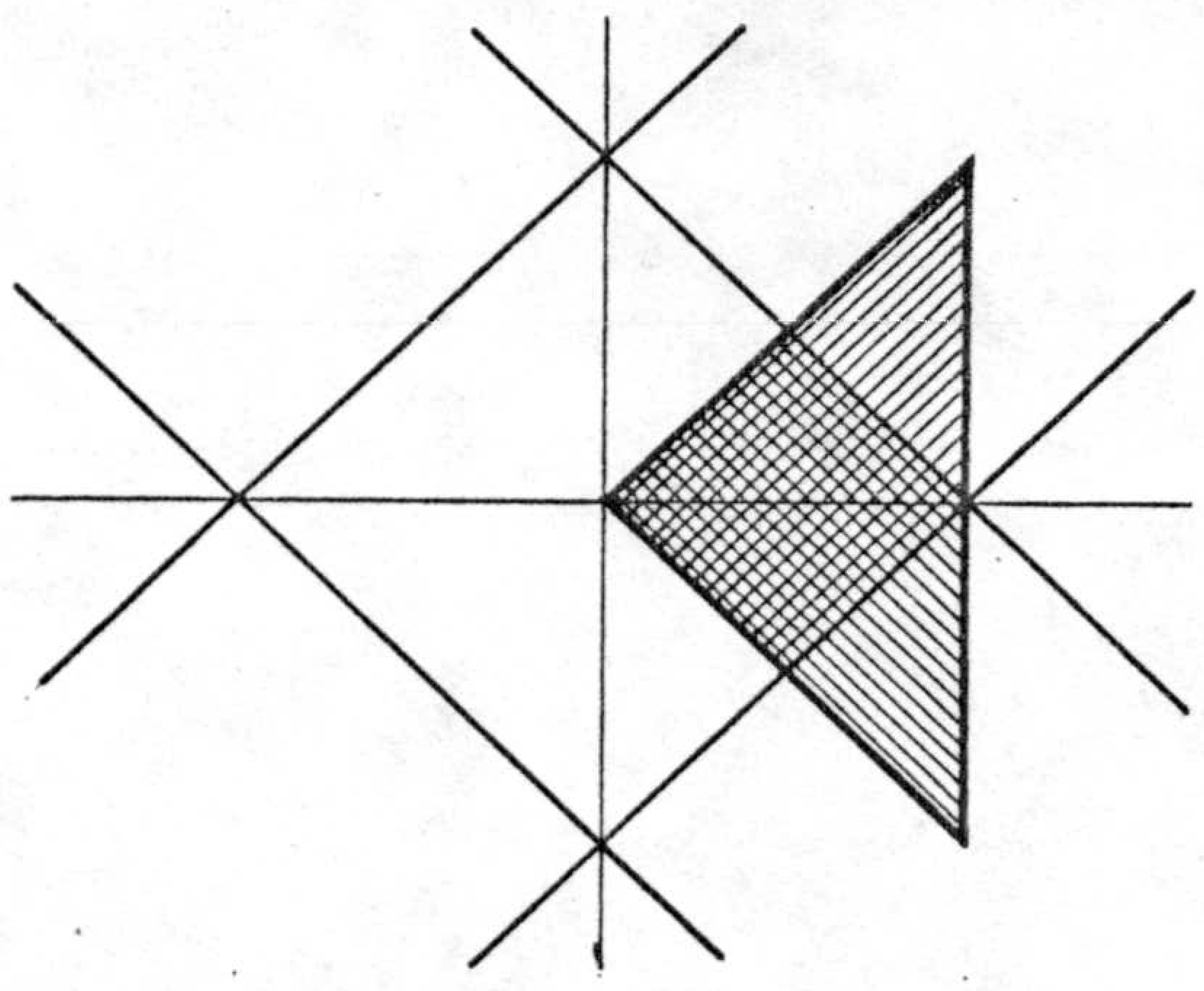

Figura 3.8.

Así pues, en este problema, aunque $D(\eta^0)$ coincide con R, sólo el cero es tal que $C(\eta^0,0)$ contiene a X. Sin embargo, esto es suficiente para poder afirmar que η^0 es D_1-óptimo. Por otra parte, observando el recinto X, es obvio que la solución del problema del cilindro de mínimo contenido es

$$u_1^2 \leq 1$$

que es la elección de η^0 como óptimo

Ejemplo 3.3.1.3.. Consideremos el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde θ_1 y θ_2 son parámetros desconocidos y x es una variable de control que suponemos perteneciente al cuadrilátero de vértices $(0,0)^t$,

$(1,0)^t$, $(3,1)^t$ y $(3,3)^t$. Estamos interesados al igual que en problemas anteriores en encontrar el diseño D_1-óptimo.

Podríamos pensar nuevamente que el diseño que asigna probabilidad uno al punto $(1,0)^t$ es el D_1-óptimo. En este caso, como en los problemas anteriores, tendríamos que hallar un número real, si existiese, tal que el cilindro

$$(u_1 - Cu_2)^2 \leq 1$$

contuviese a X.

Sin embargo, si observamos la figura 3.9., la mejor elección de c es aquella tal que la recta de ecuación

$$u_1 - Cu_2 = 1$$

pasa por el punto (3,1), es decir, C = 2. Pero, el cilindro

$$(u_1 - 2u_2)^2 \leq 1$$

no contiene a X.

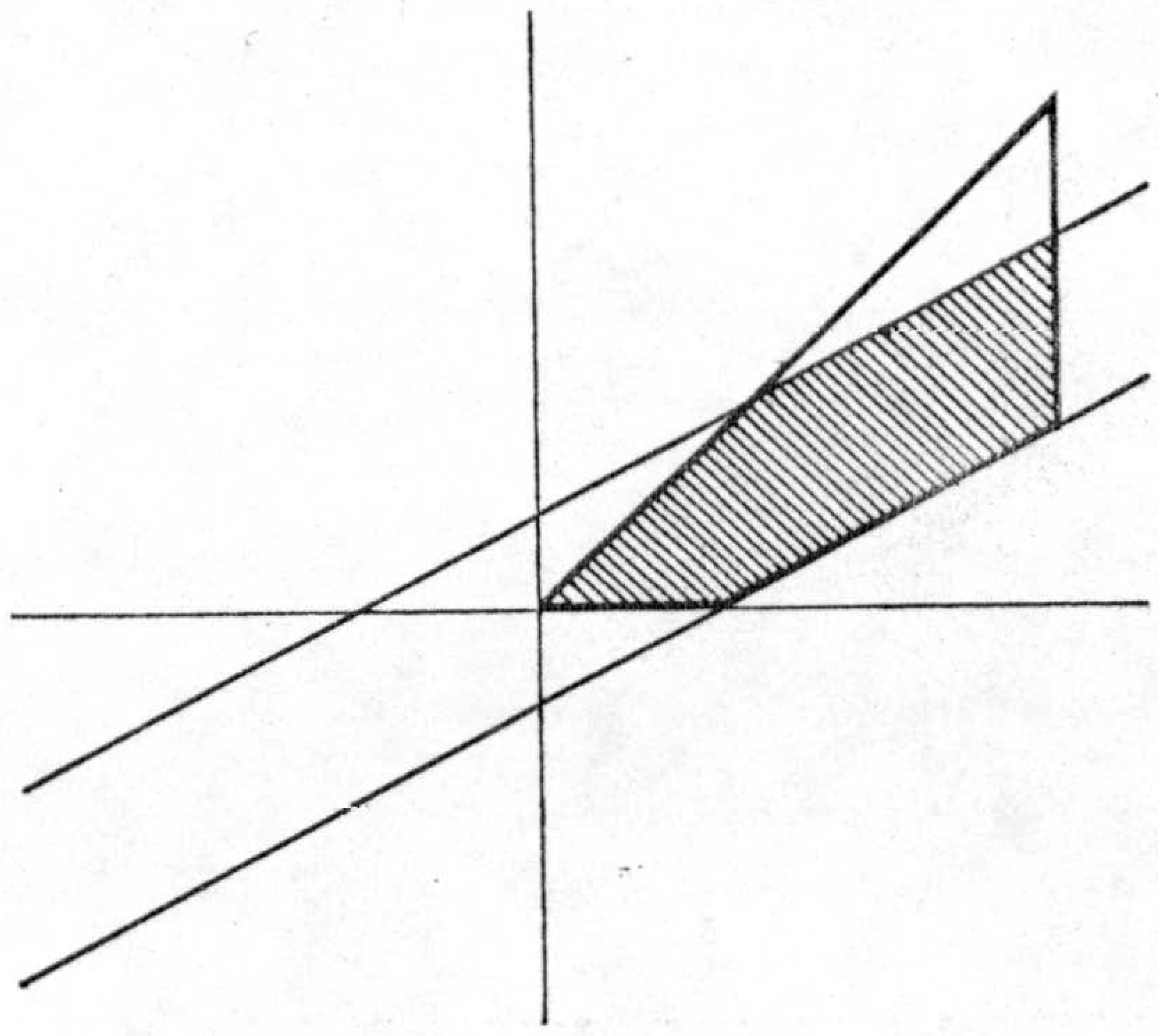

Figura 3.9.

El teorema de dualidad sugiere, sin más que observar el recinto X, que la mejor forma de diseñar el experimento es realizar observaciones en los puntos $(3,1)^t$ y $(3,3)^t$. Veamos si es cierto.

Si consideramos el diseño η_α que asigna probabilidad α al punto $(3,1)^t$ y probabilidad $1-\alpha$ al punto $(3,3)^t$, entonces

$$M(\eta_\alpha) = \begin{pmatrix} 9 & 9 - 6\alpha \\ 9 - 6\alpha & 9 - 8\alpha \end{pmatrix}$$

de forma que para cualquier $\alpha \epsilon [0,1]$ sólo existe una solución de la ecuación

$$M_{12} = CM_{22}$$

esta solución es

$$C = \frac{9 - 6\alpha}{9 - 8\alpha}$$

Por tanto, la matriz $M_1(\eta_\alpha)$ podrá escribirse

$$M_1(\eta_\alpha) = 9 - \frac{9 - 6\alpha}{9 - 8\alpha}(9 - 8\alpha)\frac{9 - 6\alpha}{9 - 8\alpha} = 36\frac{\alpha(1 - \alpha)}{9 - 8\alpha}$$

puede probarse fácilmente que el determinante de $M_1(\eta_\alpha)$ es máximo si- $\alpha = 3/4$. Entonces, para el diseño $\eta_{\frac{3}{4}}$, la solución de

$$M_{12} = CM_{22}$$

es

$$C = \frac{3}{2}$$

Para saber si el diseño $\eta_{\frac{3}{4}}$ es D_1-óptimo hemos de comprobar que - el cilindro

$$(u_1 - \frac{3}{2}u_2)\frac{4}{9}(u_1 - \frac{3}{2}u_2) \leq 1$$

contiene a X.(figura 3.10). Esto es evidente; por tanto, el diseño $\eta_{\frac{3}{4}}$ es D_1-óptimo pues el cilindro $C(\eta_{\frac{3}{4}},\frac{3}{2})$ contiene a X. Así pues, la mejor forma de diseñar el experimento es realizar el 25% de las observaciones en el punto $(3,3)^t$ y el 75% en $(3,1)^t$.

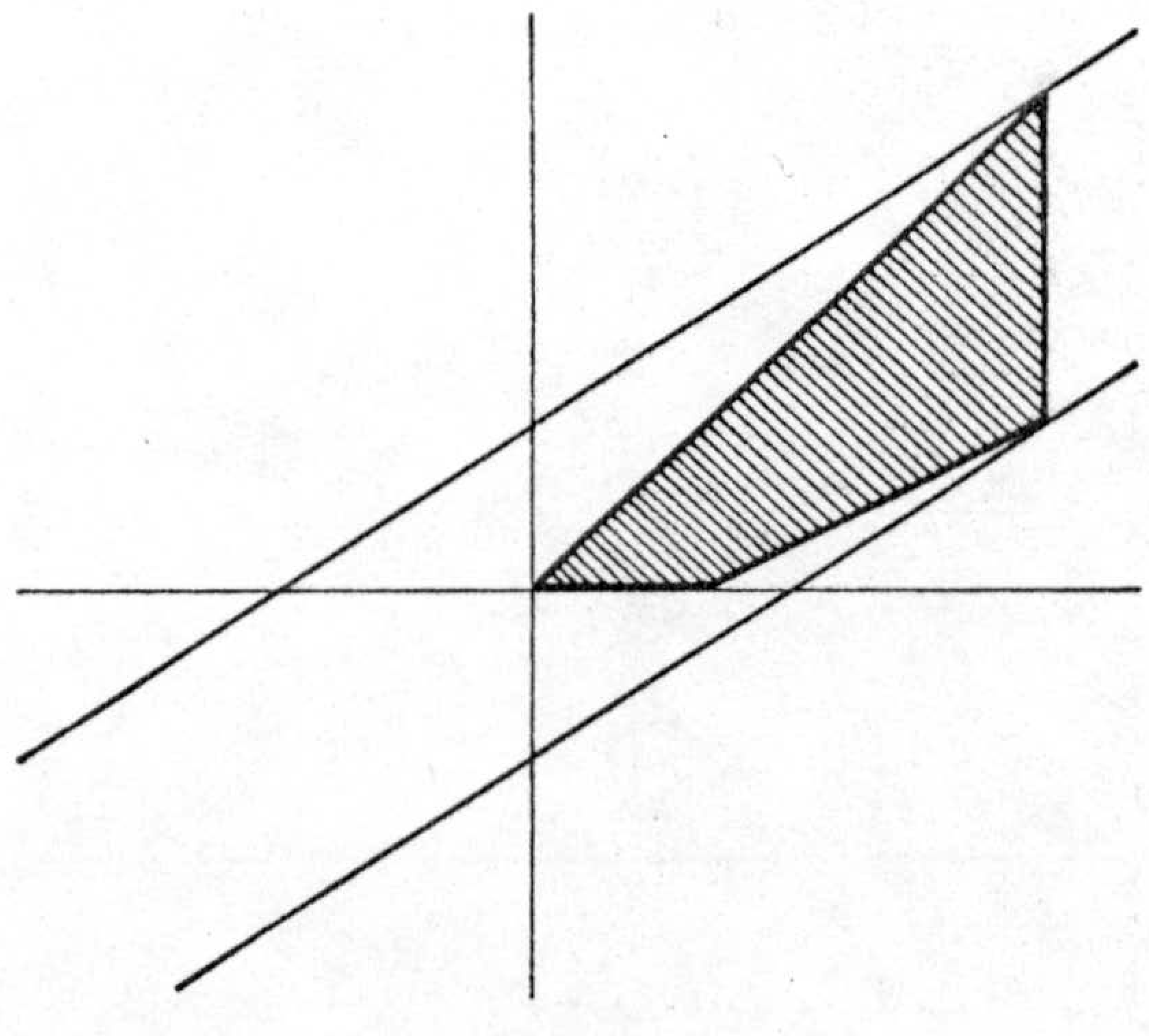

Figura 3.10.

Observemos cómo, en este ejemplo, el único elemento de $D(\eta_{\frac{3}{4}})$ nos ha bastado para asegurar que $\eta_{3/4}$ es D_1-óptimo. Sin embargo, aunque la clase $D(\eta^0)$ es R, no hemos podido encontrar ningún elemento C de dicha clase tal que su cilindro asociado $C(\eta^0,C)$ contenga a X. Por último, resaltemos que si diseñamos el experimento según la medida $\eta_{\frac{3}{4}}$, la varianza del mejor estimador lineal insesgado de θ_1, para un número fijo de observaciones es proporcional a 4/9, mientras que si el diseño se hace según la medida η^0 dicha varianza es proporcional a 1. Con lo cual el diseño $\eta_{3/4}$ representa una notable mejoría con respecto a η^0.

Ejemplo 3.3.1.4. Consideremos el siguiente modelo de regresión

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + e$$

en donde $x^t = (x_1, x_2, x_3)$ es un vector de control que el experimentador puede tomar en el tetraedro de vértices $(1,0,0)^t$, $(0,1,0)^t$, $(0,0,1)^t$ y $(0,0,0)^t$, e es un error aleatorio de media nula y varianza σ^2, independiente para cada observación.

Supongamos que estamos interesados en la estimación de (θ_1, θ_2). Evidentemente, hay varias formas de estimar este subvector paramétrico pero una que parece razonable es realizar las observaciones de forma que la varianza generalizada del mejor estimador lineal insesgado de (θ_1, θ_2) sea mínima. Es decir, estamos intentando encontrar el diseño D_2-óptimo. Por otra parte, parece lógico pensar que la mejor forma de diseñar el experimento es realizar el 50% de las observaciones en el vértice $(1,0,0)^t$ y el otro 50% en el $(0,1,0)^t$.

Veamos si este diseño, que llamaremos η^0, es D_2-óptimo. La matriz de información asociada al diseño η^0 es

$$M(\eta^0) = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

de forma que el conjunto de soluciones de $M_{12} = CM_{22}$ está formado por las matrices $C = (\alpha\ \beta)^t$ que cumplen

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} . 0$$

es decir, cualquier vector de R^2 pertenece a $D(\eta^0)$.

Por otra parte

$$M_2(\eta^0) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

y tomando $C = (0,0)^t$, su cilindro asociado

$$2u_1^2 + 2u_2^2 \leq 2$$

contiene a X. (figura 3.11).

De forma que η^0 es un diseño D_2-óptimo. Este resultado era evidente sin más que observar el recinto X y tener en cuenta el teorema de dualidad.

Por último, para completar el cuadro de situaciones de la clase $D(\eta)$, vamos a encontrar en este mismo problema un diseño η^1 tal que $D(\eta^1)$ está formada por un único elemento y el diseño η^1 no es D_2-óptimo.

Sea el diseño η^1 que asigna probabilidad 1/3 a los puntos $(1,0,0)^t$ $(0,1,0)^t$ y $(0,0,1)^t$. Entonces

$$M(\eta^1) = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}$$

Para este diseño existe una única solución C de

$$M_{12} = CM_{22}$$

pues en este caso la ecuación anterior se convierte en

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \frac{1}{3}$$

de forma que $C = (0,0)^t$

Entonces la matriz $M_2(\eta^1)$ es

$$M_2(\eta^1) = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}$$

y el cilindro asociado, en este caso único, queda

$$3u_1^2 + 3u_2^2 \leq 2$$

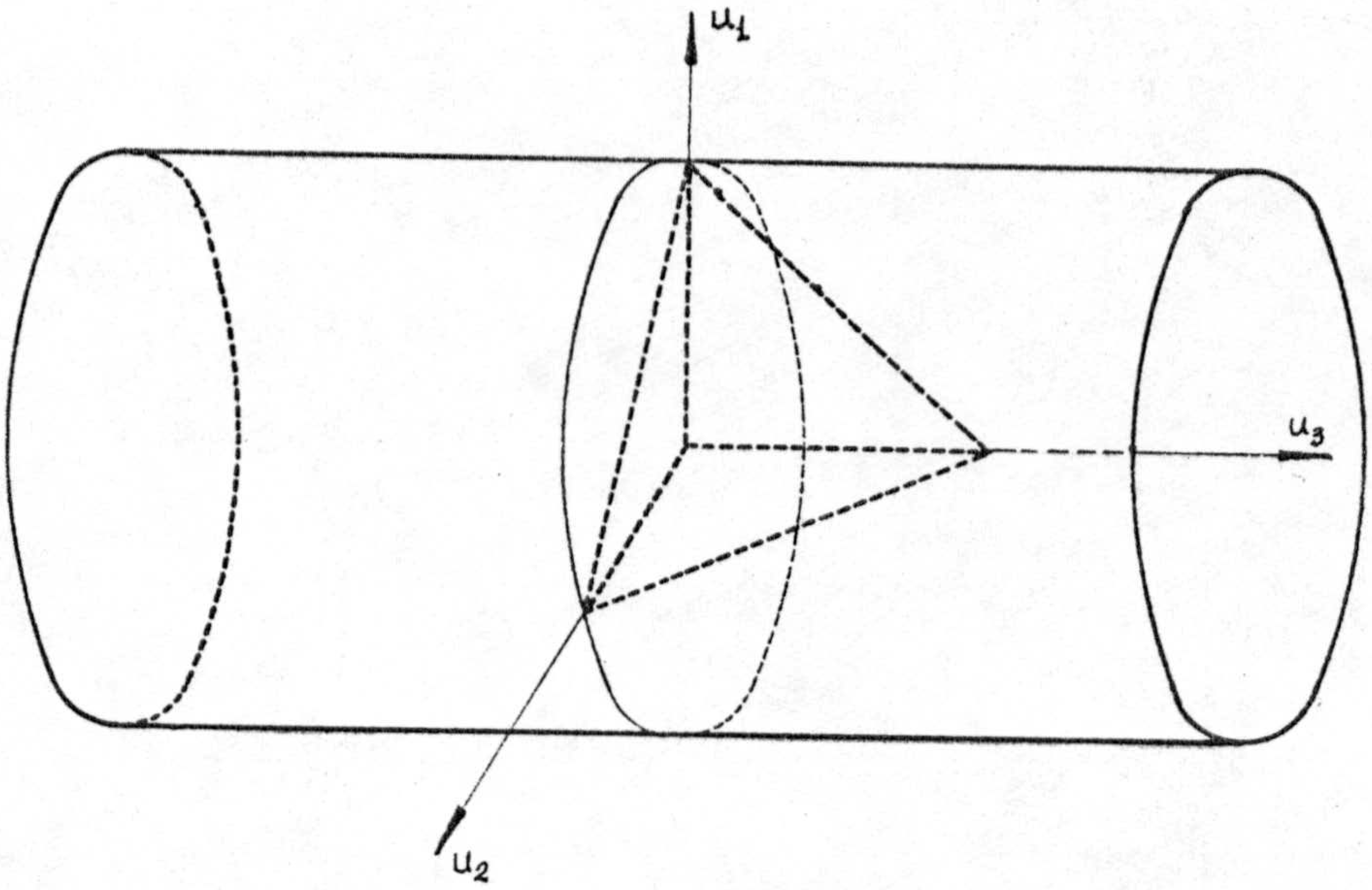

Figura 3.11.

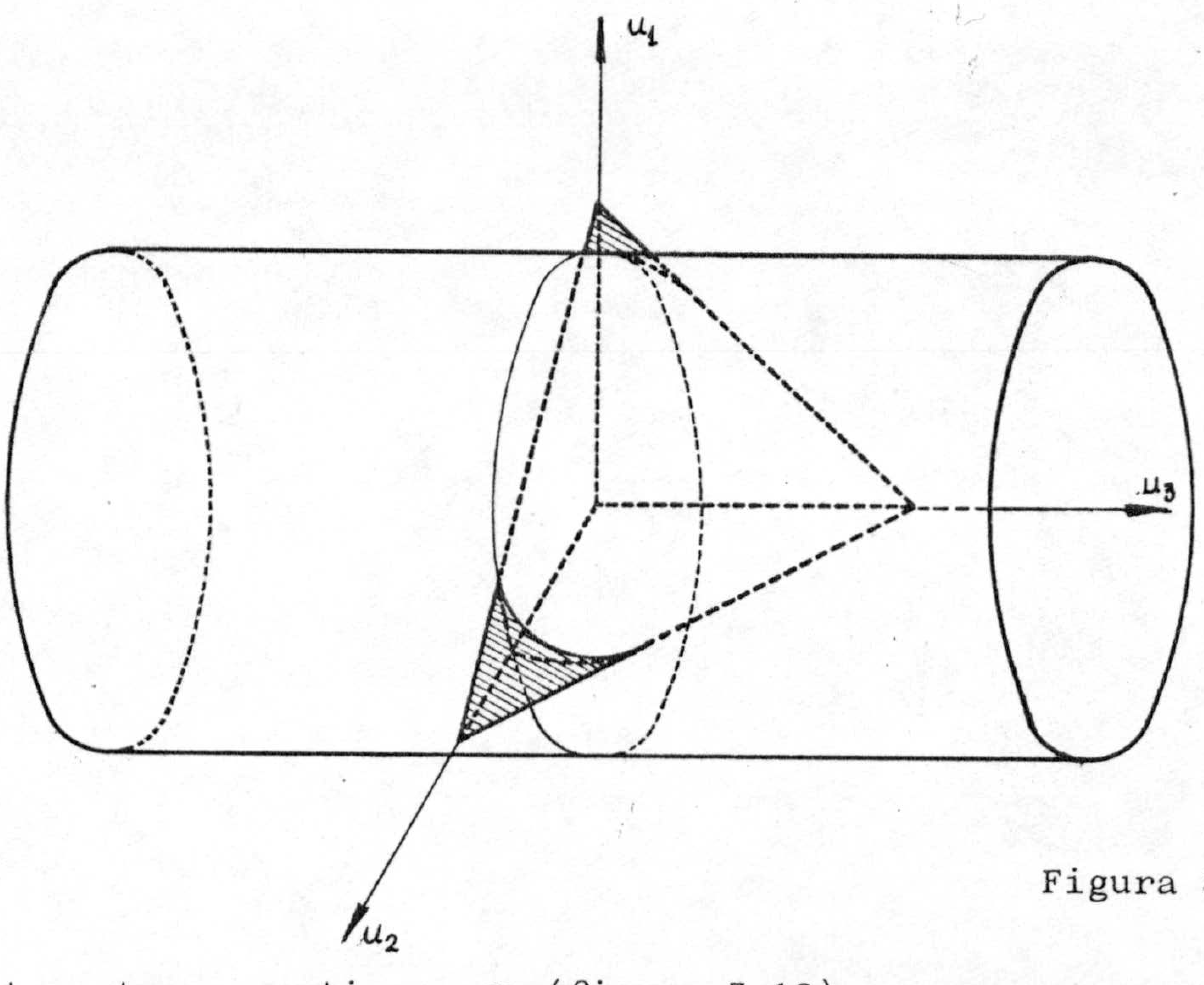

Figura 3.12.

que evidentemente no contiene a X.(figura 3.12).

De forma que η^1 no puede ser D_2-óptimo. Esto es claro, por otraparte, puesto que el determinante de la matriz de información del diseño η^1 es 1/9, mientras que para el diseño η^0 dicho determinante toma el valor 4^{-1}.

3.3.2. APORTACIONES SOBRE LA COTA DE ATWOOD PARA LA D_s-OPTIMALIDAD.

Recordemos, como hemos dicho en 3.3.1., que si $M(\eta)$ es no singular, entonces el sistema

$$M_{12} = CM_{22}$$

tiene solución única en C. Por ello, en este caso podemos escribir la cantidad $d_s(x,\eta)$ en lugar de $d_s(x,\eta,C)$, al no existir ambigüedad posible con respecto a C.

Atwood(1.969), en su teorema 4.3, probó el resultado siguiente

Teorema 3.3.2.1. Si $M(\eta)$ es no singular, entonces

$$\frac{|M_s(\eta)|}{|M_s(\eta^0)|} \geq \left[\frac{s}{\bar{d}_s(\eta)}\right]^s$$

siendo η^0 un diseño D_s-óptimo.

Probaremos, mediante dos demostraciones diferentes, que el resultado es válido independientemente de que $M(\eta)$ sea o no singular. Por otra parte, la demostración del teorema de Atwood es muy complicada, no así las propuestas por nosotros basadas en los teoremas de dualidad.

Para probar el teorema anterior utilizando el teorema de Sibson necesitamos los lemas siguientes:

Lema 3.3.2.1.(Silvey 1.980). Sea M el siguiente conjunto

$$M = \{M(\eta):\ \eta \in H\}$$

Entonces, cada elemento de M puede ser expresado como una combinación lineal convexa

$$\sum_{i=1}^{l} \lambda_i x^i x^{i\,t}$$

con $x^i \in X$, $i = 1,2,\ldots,l$ y $l \leq \frac{1}{2}k(k+1) + 1$

Lema 3.3.2.2. Si $M(\eta)$ puede escribirse como

$$M(\eta) = \sum_{i=1}^{l} \lambda_i x^i x^{i\,t}$$

con $x_i \in X, \Sigma\lambda_i = 1$ y $\lambda_i \geq 0$. Entonces

$$M_s(\eta) = \sum_{i=1}^{l} \lambda_i (x_1^i - Cx_2^i)(x_1^i - Cx_2^i)^t$$

en donde, x^i ha sido escrito como $(x_1^{i\,t}, x_2^{i\,t})^t$ y x_1^i tiene s componentes, siendo C cualquier solución de $M_{12} = CM_{22}$.

La demostración es una consecuencia inmediata del corolario 6.1- de Karlin-Studden(1.966),

Lema 3.3.2.3. Bajo las mismas hipótesis que el lema anterior, para -- cualquier matriz D de orden s×(k-s), la matriz

$$\sum_{i=1}^{l} \lambda_i (x_1^i - Dx_2^i)(x_1^i - Dx_2^i)^t - M_s(\eta)$$

es definida no negativa.

La demostración es inmediata por el lema 6.4 de Karlin-Studden,- (1.966)

Pasemos a demostrar el teorema siguiente, utilizando los lemas anteriores.

Teorema 3.3.2.2. Sea η un diseño arbitrario. Entonces

$$\frac{|M_s(\eta)|}{|M_s(\eta^0)|} \geq \left[\frac{s}{\bar{d}_s(\eta, C)}\right]^s$$

en donde C pertenece a $D(\eta)$ y η^0 es cualquier diseño D_s-óptimo.

Demostración: Por el Lema 3.3.2.1. $M(\eta)$ y $M(\eta^0)$ pueden representarse como

$$M(\eta) = \sum_{i=1}^{l^1} \lambda_i x^i x^{i\,t}$$

$$M(\eta^0) = \sum_{j=1}^{l^0} \lambda_j x^j x^{j\,t}$$

Por otra parte, teniendo en cuenta el Lema 3.3.2.2. podemos escribir

$$M_s(\eta) = \sum_{i=1}^{l^1} \lambda_i (x_1 - Cx_2)(x_1 - Cx_2)^t$$

en donde C es cualquier solución de

$$M_{12}(\eta) = CM_{22}(\eta)$$

por el mismo lema tendremos que

$$M_s(\eta^0) = \sum_{j=1}^{l^0} \lambda_j (x_1^j - C(\eta^0)x_2^j)(x_1^j - C(\eta^0)x_2^j)^t$$

siendo $C(\eta^0)$ cualquier solución de

$$M_{12}(\eta^0) = C(\eta^0)M_{22}(\eta^0)$$

Por otra parte, si llamamos $M^C(\eta^0)$ a la siguiente matriz

$$\sum_{j=1}^{l^0} \lambda_j (x_1^j - Cx_2^j)(x_1^j - Cx_2^j)^t$$

se tendrá, por el Lema 3.3.2.3.

$$|M^C(\eta^0)| \geq |M_s(\eta^0)|$$

Sea, ahora, el siguiente conjunto

$$X_1 = \{ (x_1 - Cx_2) : x = (x_1^t, x_2^t)^t \varepsilon X \text{ y } x_1 \text{ tiene s componentes} \}$$

evidentemente, X_1 es un subconjunto compacto de R^s por ser compacto X.

Si consideramos ahora el problema de diseño D-óptimo, s-dimensional, para X_1 y a las distribuciones de probabilidad sobre X_1 las notamos ξ, entonces $M_s(\eta)$ puede ser considerada como la matriz de información de un diseño ξ^1 sobre X_1 y escribir

$$M_s(\eta) = M(\xi^1)$$

por otra parte, puede darse una interpretación análoga a $M^C(\eta^0)$ y escribir, para un diseño ξ^2

$$M^C(\eta^0) = M(\xi^2)$$

es claro, además que

$$|M(\xi^2)| \leq |M(\xi^0)|$$

siendo ξ^0 un diseño D-óptimo para el problema asociado a X_1.

Podemos escribir, en consecuencia, la siguiente cadena de desigualdades

$$\frac{|M_s(\eta)|}{|M_s(\eta^0)|} \geq \frac{|M_s(\eta)|}{|M^C(\eta^0)|} \geq \frac{|M(\xi^1)|}{|M(\xi^0)|}$$

Ahora bien, por el teorema 3.2.2.1.

$$\frac{|M(\xi^1)|}{|M(\xi^0)|} \geq \left[\frac{s}{\bar{d}(\xi^1)}\right]^s$$

pero

$$\bar{d}(\xi^1) = \bar{d}_s(\eta, C)$$

con lo cual el resultado queda probado.

Observación: Podría ocurrir que X_1 no generase el subespacio correspondiente a los s primeros parámetros. Pero en este caso sabemos que

$$|M_s(\eta)| = 0$$

y por tanto, $\bar{d}_s(\eta,C)$ vale infinito con lo cual la desigualdad es trivialmente cierta.

Definimos para cualquier diseño

$$\bar{d}_s(\eta) = \inf_{C\in D(\eta)} \bar{d}_s(\eta, C)$$

Vamos a probar utilizando primero los lemas anteriores, en consecuencia utilizando el teorema de dualidad de Sibson, y posteriormente mediante el teorema de dualidad de Silvey-Titterington el resultado siguiente, que supone una debilitación de las hipótesis del teorema 3.3.2.1..

Teorema 3.3.2.3. Sea η un diseño arbitrario. Entonces

$$\frac{|M_s(\eta)|}{|M_s(\eta^0)|} \geq \left[\frac{s}{\bar{d}_s(\eta)}\right]^s$$

Demostración:

(a) Utilizando el teorema anterior es inmediata.

(b) Utilizando el teorema de dualidad para la D_s-óptimalidad de Sil--vey-Titterington la demostración sería la siguiente:

Para cualquier diseño η y cualquier matriz C de orden $s\times(k-s)$ -- se tiene

$$d(x,\eta;C) \leq \bar{d}_s(\eta,C)$$

en consecuencia, para todo $x \in X$ podemos afirmar que

$$(x_1 - Cx_2)^t\left(\frac{s}{\bar{d}_s(\eta,C)} M_s^{-1}(\eta)\right)(x_1 - C\,x_2) \leq s$$

de modo que por el Teorema 3.2.1.2.

$$|M_s(\eta)|\left[\frac{s}{\bar{d}_s(\eta,C)}\right]^{-s} \geq |M_s(\eta^0)|$$

que puede escribirse

$$\frac{|M_s(\eta)|}{|M_s(\eta^0)|} \geq \left[\frac{s}{\bar{d}_s(\eta,C)}\right]^s$$

En consecuencia, tomando ínfimo en la clase $D(\eta)$ se obtiene el -resultado pedido.

CAPÍTULO 4

JUSTIFICACIÓN DE HIPÓTESIS EN DISEÑO ÓPTIMO.

4.0. RESUMEN.

En este capítulo probaremos que la hipótesis de compacidad exigida al espacio de control X no es debilitable si queremos garantizar la existencia de al menos un diseño óptimo para cada criterio de optimalidad enunciado en el primer capítulo, salvo para la optimalidad universal y Schur —optimalidad. Demostraremos que estos dos criterios pueden no admitir diseños óptimos ni aun en el caso en que el espacio de control sea compacto.

El análisis se hará, para cada criterio, proponiendo dos ejemplos, en uno de ellos se supondrá que X es cerrado y no acotado y en el otro que el espacio de control es acotado y no cerrado; en ambos casos se demostrará que no existe diseño óptimo y en consecuencia habremos probado que la condición de ser cerrado y acotado, que caracteriza la compacidad de un subconjunto de R^k con la topología usual no es debilitable.

4.1. INTRODUCCIÓN.

Consideremos el siguiente modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k + e = x^t \theta + e$$

en donde e es un error aleatorio de media nula y varianza σ^2, $\theta^t=(\theta_1,\ldots,\theta_k)$ es un vector paramétrico desconocido y $x=(x_1,\ldots,x_k)^t$ es una variable de control que suponemos perteneciente a un subconjunto compacto, X, de R^k.

El propósito de todo criterio de optimalidad es guiar nuestra elección de N puntos de X, $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$, en los que observar la variable aleatoria Y_x, de forma que si estas observaciones son incorreladas obtengamos máxima información, en algún sentido a precisar, sobre el vector paramétrico θ.

Con este fin definíamos en el capítulo primero un diseño d como un conjunto de N puntos $x_{(1)}, \ldots, x_{(N)}$, no necesariamente distintos, de X y llamábamos matriz de información del diseño d y la notábamos M(d) a $\sum_i x_{(i)} x_{(i)}^t$. Pero como decimos en el Apéndice 1, llegado el momento de caracterizar la optimalidad de un diseño, hemos de recurrir a lo que hemos llamado teoría de la aproximación y definir un diseño η (medida de diseño), como una distribución de probabilidad sobre X y su matriz de información como

$$M(\eta) = E_\eta [X.X^t]$$

En consecuencia si d es un diseño basado en las observaciones $x_{(1)}, \ldots, x_{(N)}$ y η es la medida de diseño que asigna probabilidad 1/N a dichos puntos, tendremos

$$M(\eta) = \frac{1}{N}\sum_i x_{(i)} x_{(i)}^t = \frac{1}{N}M(d)$$

de forma que una vez caracterizada la optimalidad de un diseño η^0 buscaremos el diseño basado en N puntos más próximo.

Para los ejemplos que estudiaremos a continuación utilizaremos preferentemente las medidas de diseño η para la caracterización de las soluciones óptimas y una vez caracterizadas dichas soluciones, buscaremos el diseño basado en N puntos de X más próximo.

4.2. JUSTIFICACIÓN DE HIPÓTESIS EN LOS CRITERIOS MÁS CONOCIDOS.

4.2.1. D-OPTIMALIDAD.

Teniendo en cuenta que D-optimalidad y G-optimalidad son criterios equivalentes, en el contexto de la teoría de la aproximación, va mos a justificar las hipótesis para ambos criterios simultáneamente.- Podría pensarse, no obstante, que al no darse la equivalencia de ambos criterios, cuando trabajamos con diseños basados en N puntos de X, la - simplificación no será válida. Sin embargo, probaremos que la medida- de diseño que es D-óptima para el problema que propondremos puede obtenerse como un diseño basado en N puntos de X.

4.2.1.1. X ACOTADO Y NO CERRADO.

Consideremos el modelo lineal

$$Y_x = \theta_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media cero y varianza σ^2, $\theta^t=(\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x=(1,x_2)^t$ es una variable de - control que supondremos perteneciente al espacio de control X, en donde

$$X = \{ x = (x_1,x_2)^t \epsilon R^2 : x_1 = 1 , 0 < x_2 < 1 \}$$

Supongamos que pretendemos encontrar el diseño D-óptimo en base- a la realización de dos observaciones incorreladas de la variable --- aleatoria Y_x.

Para encontrar dicho diseño vamos a recurrir a la teoría de la - aproximación y posteriormente buscaremos de los diseños basados en -- dos observaciones el que más se aproxima.

Demostraremos que no existe diseño D-óptimo para este espacio de control, probando que para el espacio de control

$$X_1 = \{ x = (x_1,x_2)^t \epsilon R^2 : x_1 = 1 , 0 \leq x_2 \leq 1 \}$$

el diseño D-óptimo cumple las siguientes propiedades: i) asigna sólo- probabilidades a los puntos $(1,0)^t$ y $(1,1)^t$ de X_1, (en consecuencia - no será un diseño sobre X); ii) es aproximable por una sucesión de diseños sobre X.

Sea η^0 el diseño que asigna probabilidad 1/2 al punto $(1,0)^t$ y al punto $(1,1)^t$, entonces

$$M(\eta^0) = E_{\eta^0}[X.X^t] = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

de modo que

$$M^{-1}(\eta^0) = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$$

y en consecuencia por el teorema de equivalencia de Kiefer-Wolfowitz, η^0 es un diseño D-óptimo para el espacio de control X_1.

Sabemos también por el teorema de equivalencia que todos los diseños D-óptimos tienen la misma matriz de información. Observemos,por otra parte, que cualquier diseño η sobre X_1 puede ser considerado como una distribución de probabilidad sobre $I = [0,1]$ y

$$E_\eta[X.X^t] = \begin{pmatrix} 1 & E_\eta[X] \\ E_\eta[X] & E_\eta[X^2] \end{pmatrix}$$

Además, si η es un diseño sobre el intervalo $(0,1)$, se tendrá

$$E_\eta[X^2] < E_\eta[X]$$

y en consecuencia nunca será D-óptimo para el espacio de control X_1.

Probaremos ahora que existe una sucesión de diseños $\{\eta_n\}_{n\in N}$ sobre X que convergen al diseño D-óptimo sobre X_1.

Sea, para $n\geq 2$, el diseño η_n que asigna probabilidad 1/2 a los puntos $(1,1/n)^t$ y $(1,1-1/n)^t$. Claramente η_n es un diseño sobre X y

$$E_{\eta_n}[X.X^t] = \begin{pmatrix} 1 & 1/2 \\ 1/2 & \tfrac{1}{2}(1+2/n^2-2/n) \end{pmatrix}$$

de modo que

$$\lim_{n\to\infty} M(\eta_n) = M(\eta^0)$$

En consecuencia, si η es una distribución de probabilidad sobre X se tiene

$$|M(\eta)| < 1/4$$

$$\sup_{\eta \in H} |M(\eta)| = 1/4$$

Por tanto, no existen diseños D-óptimos para el espacio de control X.

Obviamente, si queremos hallar el diseño d^0 basado en dos observaciones que sea D-óptimo dicho diseño realizará una observación en el punto $(1,0)^t$ y otra en el punto $(1,1)^t$ y claramente es D-óptimo y G-óptimo para el espacio de control X_1, de forma que para el espacio de control X no existe diseño D-óptimo ni G-óptimo.

4.2.1.2. X CERRADO Y NO ACOTADO.

Supongamos que para el mismo modelo del apartado 4.2.1.1. el espacio de control es

$$X = \{ (x_1,x_2)^t : x_1 = 1 \ , \ x_2 \geq 0 \}$$

y queremos hallar los diseños D-óptimos y G-óptimos para este espacio de control, basado en una muestra de tamaño dos. Entonces, si d es un diseño que realiza las observaciones en los puntos $(1,x)^t$ y $(1,y)^t$ se tendrá

$$M(d) = \begin{pmatrix} 2 & x + y \\ x+y & (x^2 + y^2) \end{pmatrix}$$

de modo que

$$|M(d)| = (x - y)^2$$

por tanto

$$\sup_d |M(d)| = \infty$$

y en consecuencia no existe diseño D-óptimo.

Por otra parte, si $x \neq y$

$$M^{-1}(d) = \frac{1}{(x-y)^2} \times \begin{pmatrix} x^2+y^2 & -(x+y) \\ -(x+y) & 2 \end{pmatrix}$$

de modo que

$$\sup_{z \in X} z^t M^{-1}(d) z = \sup_{u \geq 0} \left[(x-u)^2 + (y-u)^2 \right] / (x-y)^2 = \infty$$

Por tanto, tampoco existe diseño G-óptimo.

4.2.2. D_S - y D_A - OPTIMALIDAD.

Vamos a probar que para estos dos criterios tampoco es debilitable la hipótesis de compacidad exigida al espacio de control $\mathcal{X}$. Por otra parte, teniendo en cuenta, según vimos en el capítulo segundo de la presente memoria, que la D_A-optimalidad puede ser reformulada como D_S-optimalidad sin más que realizar una transformación lineal sobre el espacio de control, justificaremos las hipótesis sólo para la D_S-optimalidad.

4.2.2.1. $\mathcal{X}$ ACOTADO Y NO CERRADO.

Consideremos el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media cero y varianza σ^2; $\theta^t = (\theta_1, \theta_2)$ es un vector paramétrico desconocido y $x = (x_1, x_2)^t$ es una variable de control que suponemos perteneciente al interior del recinto limitado por el triángulo de vértices $(1,1)^t$, $(0,0)^t$ y $(1,-1)^t$ y el círculo unidad (figura 4.1.).

Nos proponemos encontrar el diseño d_1^0 basado en dos observaciones de la variable Y_x, que estime mejor θ_1, en consecuencia pretendemos encontrar el diseño D_1-óptimo.

Para que un diseño d estime θ_1, ha de realizar las observaciones de la variable aleatoria Y_x en dos puntos distintos no proporcionales o si son proporcionales serán necesariamente de la forma $(\alpha,0)^t$ y $(\beta,0)^t$

con α y β positivos y menores que uno, pudiendo ocurrir $\alpha = \beta$.

Así pues, si el diseño d estima a θ_1 y realiza las observaciones de la variable Y_x en los puntos distintos y no proporcionales $x^t_{(1)} = (x_1, x_2)$ y $x^t_{(2)} = (y_1, y_2)$, entonces

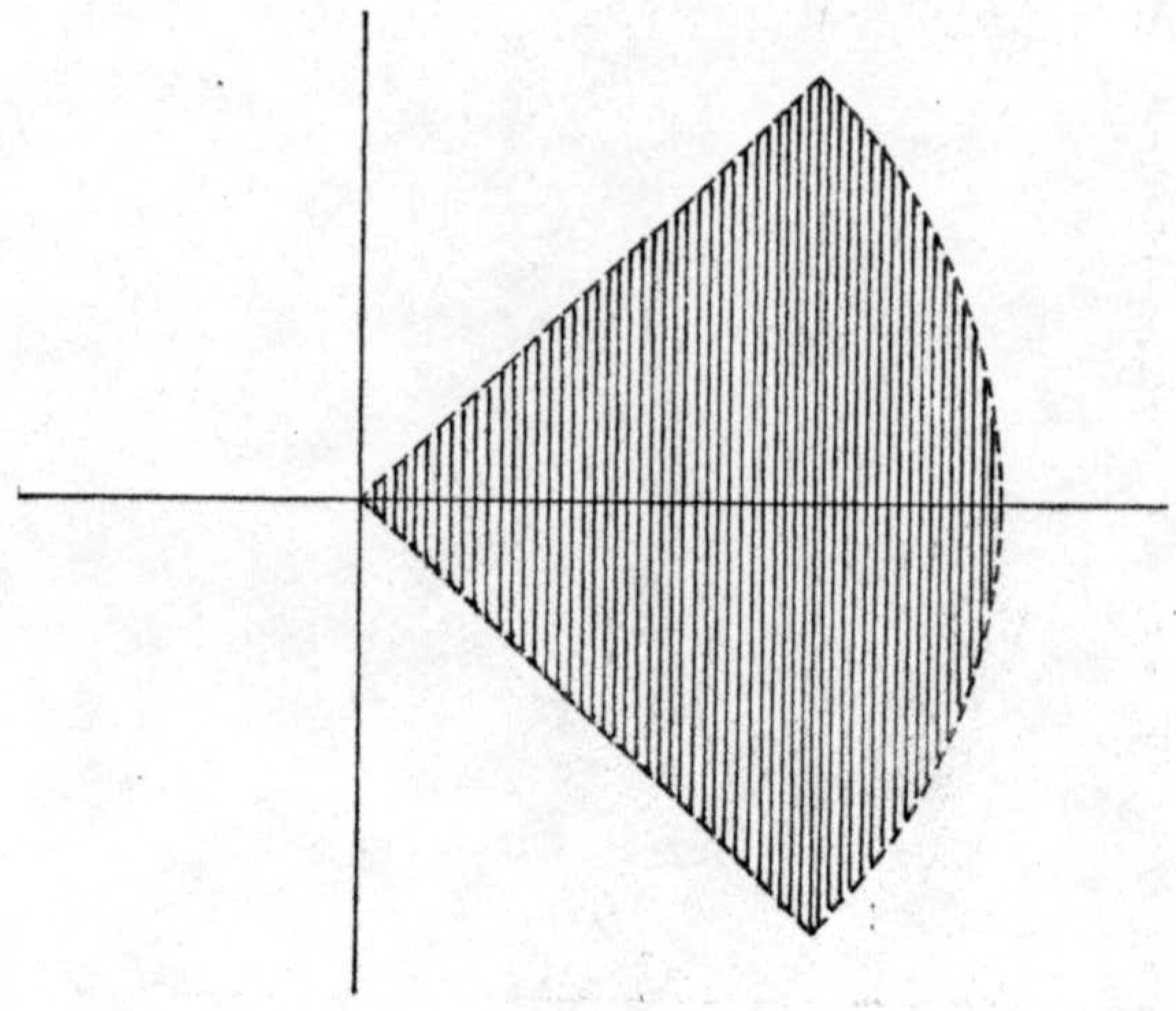

Figura 4.1.

$$M(d) = \begin{pmatrix} x_1^2+y_1^2 & x_1x_2+y_1y_2 \\ x_1x_2+y_1y_2 & x_2^2+y_2^2 \end{pmatrix}$$

de modo que

$$M_1(d) = x_1^2 + y_1^2 - \frac{(x_1x_2 + y_1y_2)^2}{x_2^2 + y_2^2} = \frac{(x_1y_2 - x_2y_1)^2}{x_2^2 + y_2^2}$$

Ahora bien, teniendo en cuenta la desigualdad de Cauchy-Schwartz

$$(x_1y_2 - x_2y_1)^2 \leq (x_1^2 + y_1^2)(x_2^2 + y_2^2)$$

tendremos

$$M_1(d) \leq x_1^2 + y_1^2 < 2$$

Por otra parte, si d realiza las observaciones de la variable Y_x en los puntos $x^t_{(1)} = (\alpha, 0)$ y $x^t_{(2)} = (\beta, 0)$ con α y β pertenecientes al intervalo $(0,1)$, entonces

$$M_1(d) = \alpha^2 + \beta^2 < 2$$

Si llamamos d_n al diseño que realiza las dos observaciones de la variable Y_x en el punto $(1 - 1/n, 0)^t$, es claro que

$$\lim_{n\to\infty} M_1(d_n) = 2$$

En consecuencia si D es la clase de todos los diseños basados en dos observaciones incorreladas de la variable Y_x

$$\sup_{d \in D} |M(d)| = 2$$

y por tanto no existe diseño D_1-óptimo para el problema que estamos considerando.

4.2.2.2. X CERRADO Y NO ACOTADO.

Consideremos el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; $\theta^t=(\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x^t = (x_1,x_2)$ es una variable de control que suponemos perteneciente a X, con

$$X = \{ (x_1,x_2)^t : x_1 \geq 0 \quad , \quad x_2 \varepsilon [0,1] \}$$

Pretendemos probar que no existe diseño d, basado en dos observaciones incorreladas de la variable aleatoria Y_x que sea el mejor estimador de θ_1, es decir, que sea D_1-óptimo.

Llamemos d_n al diseño que realiza las dos observaciones de la variable aleatoria Y_x en el punto $(n,0)^t$. Entonces

$$M(d_n) = \begin{pmatrix} 2n^2 & 0 \\ 0 & 0 \end{pmatrix}$$

de modo que

$$M_1(d_n) = 2n^2$$

En consecuencia si D es la clase de diseños que estamos considerando

$$\sup_{d \in D} |M_1(d)| = \infty$$

y, por tanto, no existe diseño D_1-óptimo.

4.2.3. L-OPTIMALIDAD.

Dentro de los criterios lineales vamos a estudiar sólo los contraejemplos correspondientes a la A-optimalidad y c-optimalidad, puesto que los restantes casos son análogos a los anteriores.

4.2.3.1. A-OPTIMALIDAD.

4.2.3.1.1. X ACOTADO Y NO CERRADO.

Consideremos el siguiente modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; $\theta^t = (\theta_1, \theta_2)$ es un vector paramétrico desconocido y $x^t = (x_1, x_2)$ es una variable de control que suponemos perteneciente a X, con

$$X = \{ (x_1, x_2)^t \in R^2 : x_1 = 1 \ , \ |x_2| < 1 \}$$

Pretendemos encontrar el diseño A-óptimo para este problema, basado en una muestra de tamaño dos, supuesto que las observaciones son incorreladas.

Resolveremos este problema como en el apartado 4.2.1.1., es decir, hallaremos el diseño A-óptimo, en el contexto de la teoría de la aproximación, para el espacio de control

$$X_1 = \{ (x_1, x_2)^t \in R^2 : x_1 = 1 \ , |x_2| \leq 1 \}$$

a continuación encontraremos el diseño basado en una muestra de tamaño dos más próximo, probando que dicho diseño realiza observaciones en puntos de X_1 que no pertenecen a X; sin embargo, podremos encontrar una sucesión de diseños sobre X que lo aproximen y en consecuencia no habrá diseño A-óptimo para X.

Así pues, sea η un diseño sobre X_1, sabemos que su matriz de información puede escribirse como

$$M(\eta) = \sum_{i=1}^{n} p_i \begin{pmatrix} 1 \\ x_i \end{pmatrix} (1 \ x_i)$$

en donde $x_i \varepsilon [-1,1]$, $p_i \geq 0$ y $\sum_i p_i = 1$.

Obviamente, toda matriz $M(\eta)$ puede representarse como un punto de R^3 con coordenadas

$$(1, \sum p_i x_i, \sum p_i x_i^2) = \sum p_i (1, x_i, x_i^2)$$

en consecuencia, el conjunto

$$M = \{M(\eta) : \eta \varepsilon H\}$$

es la envolvente convexa de

$$A = \{(1,x,x^2): |x| \leq 1\}$$

Claramente, todo punto de co(A), $(1,x_1,x_2)$, puede representarse de la forma

$$(1,x_1,x_2) = \lambda(1,x,x^2) + (1-\lambda)(1,-x,x^2)$$

en donde $x \varepsilon [0,1]$.

Por tanto, para cada $\eta \varepsilon H$

$$M(\eta) = \begin{pmatrix} 1 & \lambda x + (1-\lambda)(-x) \\ \lambda x + (1-\lambda)(-x) & x^2 \end{pmatrix}$$

Ahora bien, puesto que pretendemos encontrar el diseño A-óptimo, hemos de imponer que $M(\eta)$ sea no singular, lo que en este caso es equivalente a decir que la distribución η sobre $\mathcal{X}$, no es degenerada en un punto.

Además

$$\text{tr } M^{-1}(\eta) = \frac{1+x^2}{x^2 - (\lambda x + (1-\lambda)(-x))^2} = \frac{1+x^2}{x^2(4\lambda(1-\lambda))}$$

Por otra parte, teniendo en cuenta que la función $f(\lambda) = \lambda(1-\lambda)$, alcanza su máximo en el intervalo $I = [0,1]$ en $\lambda = ½$, y que la función $g(x) = (x^2+1)/x^2$ es decreciente para $x > 0$, podemos afirmar que el diseño η^0 que asigna probabilidad 1/2 a los puntos $(1,1)^t$ y $(1,-1)^t$ es A-óptimo y su matriz de información es la matriz unidad de orden dos.

Ahora bien, ningún diseño η sobre X, tiene matriz de información igual a $M(\eta^0)$, y en consecuencia, no puede ser A-óptimo para el espacio de control X_1, de modo que para cualquier diseño η sobre X

$$\operatorname{tr} M^{-1}(\eta) > 2$$

Veamos que si H_1 es la clase de diseños sobre X, entonces

$$\inf_{\eta \in H_1} \operatorname{tr}(M^{-1}(\eta)) = 2$$

y por consiguiente no existirá diseño A-óptimo.

Sea η_n el diseño que asigna probabilidad 1/2 a los puntos $(1,1-\frac{1}{n})^t$ y $(1,-(1-\frac{1}{n}))^t$ de X. Entonces

$$M(\eta_n) = \begin{pmatrix} 1 & 0 \\ 0 & (1-1/n)^2 \end{pmatrix}$$

de modo que

$$\operatorname{tr} M^{-1}(\eta_n) = 1 + (1-1/n)^{-2}$$

en consecuencia

$$\lim_{n\to\infty} \operatorname{tr}(M^{-1}(\eta_n)) = 2$$

y no existe diseño A-óptimo.

Si queremos encontrar el diseño A-óptimo basado en una muestra - de tamaño dos, es claro que dicho diseño realiza una observación en - el punto $(1,1)^t$ y otra en $(1,-1)^t$ para el espacio de control X_1. Sin embargo, no existe diseño A-óptimo basado en una muestra de tamaño dos- para X, pues para cualquier diseño d sobre X

$$\operatorname{tr} M^{-1}(d) > 4$$

y los diseños d_n que realizan una observación en el punto $(1,1-1/n)^t$ y otra en $(1,-(1-1/n))^t$ cumplen que

$$\lim_{n\to\infty} tr(M^{-1}(d_n)) = 4$$

4.2.3.1.2. X CERRADO Y NO ACOTADO.

Consideremos el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media cero y varianza σ^2; $\theta^t=(\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x^t = (x_1,x_2)$ es una variable de control, que suponemos perteneciente al siguiente cono convexo

$$X = \{ (x_1,x_2)^t : x_1 \geq x_2 \geq 0 \}$$

Claramente, X es cerrado pero no acotado. Pretendemos encontrar el diseño A-óptimo basado en una muestra de tamaño dos, supuesto que las observaciones son incorreladas.

Sea d_n el diseño que realiza una observación en el punto $(n,0)^t$ y otra en el punto $(n,n)^t$ del cono X. Entonces

$$M(d_n) = \begin{pmatrix} 2n^2 & n^2 \\ n^2 & n^2 \end{pmatrix}$$

por tanto

$$tr(M^{-1}(d_n)) = 3/n^2$$

en consecuencia

$$\lim_{n\to\infty} tr(M^{-1}(d_n)) = 0$$

Así pues, si D es la clase de diseños que consideramos

$$\inf_{d\in D} tr(M^{-1}(d_n)) = 0$$

Pero obviamente, para todo diseño d sobre X que estime θ, es decir, que M(d) sea no singular $tr(M^{-1}(d))>0$. En consecuencia, no existe diseño A-óptimo.

4.2.3.2. c-OPTIMALIDAD.

4.2.3.2.1. X ACOTADO Y NO CERRADO.

Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; $\theta^t=(\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x^t=(x_1,x_2)$ es una variable de control que suponemos perteneciente a

$$X = \{(x_1,x_2)^t : x_1^2 + x_2^2 < 1 \ , \ x_1 > 0 \ , \ x_2 > 0\}$$

Claramente, el recinto X está acotado pero no es cerrado. Consideremos el problema estadístico de la mejor estimación de $\theta_1 + \theta_2$ basado en la realización de dos observaciones incorreladas de la variable Y_x.

Si el diseño d realiza las observaciones de la variable Y_x en los puntos $x_{(1)}^t=(x_1,x_2)$ y $x_{(2)}^t=(y_1,y_2)$, entonces

$$M(d) = \begin{pmatrix} x_1^2+y_1^2 & x_1x_2+y_1y_2 \\ x_1x_2+y_1y_2 & x_2^2+y_2^2 \end{pmatrix}$$

Ahora bien, teniendo en cuenta,(Whittle,1.973), que

$$c^tM^{-1}(d)c = \sup_{h^t=(h_1,h_2)} [2c^th - h^tM(d)h]$$

podemos escribir

$$(1,1)M^{-1}(d)\begin{pmatrix}1\\1\end{pmatrix} = \sup_h [2h_1 + 2h_2 - (h_1x_1+h_2x_2)^2 - (h_1y_1 + h_2y_2)^2] \qquad (2)$$

Por otra parte puesto que

$$(3) \qquad (h_1x_1 + h_2x_2)^2 \leq (h_1^2 + h_2^2)(x_1^2 + x_2^2) < h_1^2 + h_2^2$$

$$(4) \qquad (h_1y_1 + h_2y_2)^2 \leq (h_1^2 + h_2^2)(y_1^2 + y_2^2) < h_1^2 + h_2^2$$

podemos afirmar para la expresión (2)

$$(5) \qquad (1,1)M^{-1}(d)\begin{pmatrix}1\\1\end{pmatrix} \geq \sup_{h\in R^2} [2h_1(1 - h_1) + 2h_2(1 - h_2)] = 1$$

Por otra parte, se da la igualdad en la primera inecuación de (3) y (4) si y sólo si $x_{(1)}^t = \lambda(h_1,h_2)$ y $x_{(2)}^t = \mu(h_1,h_2)$. Además, el supremo se alcanza en (5) para $h_1 = h_2 = 1/2$. De modo que si llamamos D a la clase de diseños que consideramos y D_c denota la clase de diseños que realizan las observaciones de la variable aleatoria Y_x en puntos de la forma $\alpha(1/2,1/2)^t$ siendo $\alpha^2/4 + \alpha^2/4 < 1$, es decir, $0 < \alpha < \sqrt{2}$, se cumplirá

$$\inf_{d\in D} (1,1)M^{-1}(d)\begin{pmatrix}1\\1\end{pmatrix} = \inf_{d\in D_c} (1,1)M^{-1}(d)\begin{pmatrix}1\\1\end{pmatrix}$$

Sea ahora, d un diseño que realiza las observaciones de la variable Y_x en los puntos $(\alpha/2,\alpha/2)^t$ y $(\beta/2,\beta/2)^t$ con α y β pertenecientes a $(0,\sqrt{2})$, tendremos

$$(1\ 1)M^{-1}(d)\begin{pmatrix}1\\1\end{pmatrix} = \sup_{h\in R^2} [(h_1 + h_2)(2 - \frac{\alpha^2+\beta^2}{4}(h_1 + h_2))] = \sup_{\lambda\in R} \lambda(2-\frac{\alpha^2+\beta^2}{4}\lambda)$$

Observemos que $\alpha^2 + \beta^2 < 4$ y en consecuencia para $0 < p < 1$ podemos escribir

$$(1,1)M^{-1}(d)\begin{pmatrix}1\\1\end{pmatrix} = \sup_{\lambda\in R} [\lambda(2 - p\lambda)] = 1/p > 1$$

de modo que

$$\inf_{d\in D} (1,1)M^{-1}(d)\begin{pmatrix}1\\1\end{pmatrix} = 1$$

y no es alcanzable bajo ningún diseño d, así pues, no existe diseño c-óptimo para el espacio de control X.

4.2.3..2.2. X CERRADO Y NO ACOTADO.

Consideremos el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

siendo e un error aleatorio de media nula y varianza σ^2, $\theta^t = (\theta_1, \theta_2)$ un vector paramétrico desconocido y $x^t = (x_1, x_2)$ una variable de control que suponemos perteneciente a

$$X = \{(x_1, x_2)^t : x_i \geq 0 \quad i=1,2\}$$

El problema estadístico que pretendemos resolver es la estimación de $\theta_1 + \theta_2$ de la mejor forma posible basándonos en la realización de -- dos observaciones incorreladas de la variable Y_x. Vamos a probar que- si llamamos D a la clase de los diseños que están basados en una muestra de tamaño dos se cumplirá

$$\inf_{d \in D} (1,1) M^{-1}(d) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0$$

y en consecuencia no existirá diseño c-óptimo para estimar $\theta_1 + \theta_2$.

Sea d_n el diseño que realiza una observación en el punto $(n,0)^t$ y otra en $(n,n)^t$ siendo n un entero positivo. Entonces

$$d((1,1), d_n) = (1,1) M^{-1}(d_n) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1/n^2$$

por tanto

$$\lim_{n \to \infty} d((1,1), d_n) = 0$$

y en consecuencia no existe diseño c-óptimo.

4.2.4. E-OPTIMALIDAD.

4.2.4.1. X ACOTADO Y NO CERRADO.

Sea el modelo lineal

$$Y_x = \theta x + e$$

en el que e es un error aleatorio de media nula y varianza σ^2, $\theta \in R$ - es un vector paramétrico desconocido y x es una variable de control - que suponemos perteneciente al intervalo (0,1).

Pretendemos saber en qué N puntos del intervalo (0,1) hemos de - observar la variable Y para encontrar el diseño E-óptimo, supuesto -- que dichas observaciones son incorreladas.

Para este problema, el diseño E-óptimo si existiese sería aquel- que maximizase el mínimo autovalor de M(d), pero al ser un problema - unidimensional la E-optimalidad es equivalente a la D- y A- optimalidad.

En consecuencia, nuestro problema es

$$\max \sum_{i=1}^{N} x_i^2 \qquad x_i \in (0,1)$$

claramente

$$\sup_{x_i \in (0,1)} \sum_i x_i^2 = N$$

pero dicho supremo no es alcanzable y por tanto, no existen diseños - E-,D-,A- óptimos para el espacio de control (0,1). Por otra parte, la sucesión de diseños que realizan todas las observaciones en el punto- (1 - 1/n) tiende al diseño E-óptimo.

4.2.4.2. X CERRADO Y NO ACOTADO.

Si consideramos el mismo modelo que en el apartado anterior y suponemos que el espacio de control es

$$X = \{ x : x \geq 0 \}$$

Claramente, este recinto no está acotado y es cerrado. Si estu-- diamos

$$\sup_{x_i \in X} (\sum_i x_i^2)$$

es obvio que dicho supremo es infinito y en consecuencia no existe -

diseño E-óptimo, puesto que para todo diseño $\sum_i x_i^2$ es finito.

4.3. S-OPTIMALIDAD Y (M,S)-OPTIMALIDAD.

4.3.1. S-OPTIMALIDAD.

4.3.1.1. $\mathcal{X}$ ACOTADO Y NO CERRADO.

Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; $\theta^t=(\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x = (x_1,x_2)^t$ es una variable de control que suponemos perteneciente a $\mathcal{X}$, en donde

$$\mathcal{X} = \{ (x_1,x_2)^t : x_1^2+ x_2^2 = 1 \ , \ x_1 > 0 \ , x_2 > 0 \}$$

Pretendemos encontrar el diseño S-óptimo para dicho modelo, basado en la realización de dos observaciones incorreladas de la variable Y_x.

Sea d el diseño que realiza dichas observaciones en los puntos - $x_{(1)}^t=(x_1,x_2)$ y $x_{(2)}^t=(y_1,y_2)$ de $\mathcal{X}$. Entonces

$$M(d) = \begin{pmatrix} x_1^2+ y_1^2 & x_1x_2+ y_1y_2 \\ x_1x_2+ y_1y_2 & x_2^2+ y_2^2 \end{pmatrix}$$

de modo que

$$\operatorname{tr} M(d) = x_1^2+ y_1^2+ x_2^2+ y_2^2= 2 \qquad \text{para todo } d$$

Si $\mathcal{X}$ contuviese a los puntos $(1,0)^t$ y $(0,1)^t$ entonces el diseño d° que realiza una observación en el punto $(1,0)^t$ y otra en $(0,1)^t$ tendría matriz de información, M(d°), la matriz unidad de orden dos. En consecuencia, al tener los dos autovalores iguales sería simultáneamente - A-, E- ,D- y S- óptimo. Claramente, este diseño no es realizable en - nuestro espacio $\mathcal{X}$. Sin embargo, si tomamos el diseño d_n que realiza - una observación en el punto $(1 - 1/n, \sqrt{2/n - 1/n^2})^t$ y otra en el ele-- mento del espacio de control $(\sqrt{2/n - 1/n^2}, 1 - 1/n)^t$ para $n \geq 2$; es claro que d_n es un diseño sobre X y

$$M(d_n) = \begin{pmatrix} 1 & 2(2/n - 1/n^2)^{\frac{1}{2}}(1 - 1/n) \\ 2(2/n - 1/n^2)^{\frac{1}{2}}(1 - 1/n) & 1 \end{pmatrix}$$

y los autovalores de $M(d_n)$ son las soluciones de la siguiente ecuación $(1 - \lambda)^2 = 4(2/n - 1/n^2)(1 - 1/n)^2$. Es decir

$$\lambda_1(d_n) = 1 + 2\sqrt{(2/n - 1/n^2)}(1 - 1/n)$$

$$\lambda_2(d_n) = 1 - 2\sqrt{(2/n - 1/n^2)}(1 - 1/n)$$

de modo que

$$\lim_{n\to\infty} [\,(\lambda_1(d_n) - 1)^2 + (\lambda_2(d_n) - 1)^2] = 0$$

y por tanto, no existe diseño S-óptimo.

4.3.1.2. X CERRADO Y NO ACOTADO.

Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k + e$$

en donde e es un error aleatorio que supondremos de media nula y varianza σ^2; $\theta^t = (\theta_1, \ldots, \theta_k)$ es un vector paramétrico desconocido y x^t denota al vector de coordenadas $(x_1, \ldots, x_k)$ que pertenece al espacio de control X que es cerrado y no acotado.

Es claro que $x_1^2 + x_2^2 + \ldots + x_k^2$ no está acotado sobre X y en consecuencia no pueden existir diseños S-óptimos para problemas lineales con este espacio de control.

4.3.2. (M,S)-OPTIMALIDAD.

4.3.2.1. X ACOTADO Y NO CERRADO.

Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; $\theta^t = (\theta_1, \theta_2)$

es un vector paramétrico desconocido y $x^t = (x_1, x_2)$ es una variable de control que suponemos perteneciente al recinto X, en donde

$$X = \{(x_1, x_2)^t : x_1^2 + x_2^2 < 1 \ , \ x_1 > 0 \ , \ x_2 > 0\}$$

Obviamente X está acotado pero no es cerrado. Pretendemos encontrar el diseño (M,S)-óptimo para dos observaciones incorreladas de la variable aleatoria Y_x.

Sea d el diseño que realiza dichas observaciones en los puntos de coordenadas $(x_1, x_2)^t$ y $(y_1, y_2)^t$. Entonces

$$M(d) = \begin{pmatrix} x_1^2 + y_1^2 & x_1 x_2 + y_1 y_2 \\ x_1 x_2 + y_1 y_2 & x_2^2 + y_2^2 \end{pmatrix}$$

y

$$\operatorname{tr} M(d) = x_1^2 + y_1^2 + x_2^2 + y_2^2 < 2$$

Sin embargo, si llamamos D a la clase de diseños que consideramos

$$\sup_{d \in D} \operatorname{tr}(M(d)) = 2$$

y en consecuenecia no existe diseño (M,S)-óptimo.

Obsérvese, por otra parte, que si tomamos el espacio de control

$$X = \{x_1^2 + x_2^2 \leq 1 \ , \ 0 \leq x_1 < 1 \ , \ 0 \leq x_2 < 1\}$$

entonces existe

$$\max_{d \in D} \operatorname{tr}(M(d))$$

pero, sin embargo el diseño (M,S)-óptimo no es alcanzable en nuestro espacio de control.

4.3.2.2. X CERRADO Y NO ACOTADO.

Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media cero y varianza σ^2; $\theta^t=(\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x^t=(x_1,x_2)$ es una variable de control que suponemos perteneciente a

$$X = \{(x_1,x_2)^t : x_1 \geq 0 \;,\; x_2 \geq 0\}$$

Obviamente, X es cerrado y no acotado. Por otra parte, si llamamos D a la clase de diseños basados en dos observaciones incorreladas de la variable aleatoria Y_x, es claro que

$$\sup_{d \in D} tr(M(d)) = \infty$$

y por consiguiente, no pueden existir diseños (M,S)-óptimos.

4.4. ϕ_p-OPTIMALIDAD.

4.4.1. X ACOTADO Y NO CERRADO.

Retomemos el ejemplo correspondiente a la S-optimalidad. Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; $\theta^t=(\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x^t = (x_1,x_2)$ es una variable de control que suponemos perteneciente al espacio X, con

$$X = \{(x_1,x_2)^t \in R^2 : x_1^2 + x_2^2 = 1 \;,\; x_1 > 0 \;,\; x_2 > 0\}$$

Queremos encontrar el diseño ϕ_p-óptimo para este modelo lineal basado en un diseño de tamaño dos.

Claramente X es acotado pero no es cerrado. Llamemos D a la clase de diseños por nosotros considerados y sea d un diseño de dicha clase que realiza las observaciones de la variable aleatoria Y_x en los puntos $(x_1,x_2)^t$ y $(y_1,y_2)^t$. Entonces

$$|M(d)| = (x_1 y_2 - x_2 y_1)^2$$

Observemos que $|M(d)| = 0$ si y sólo si el diseño d realiza las --

dos observaciones en el mismo punto. Por otra parte, si llamamos $\lambda_1(d)$ y $\lambda_2(d)$ a los autovalores de $M(d)$ con $\lambda_1(d) \geq \lambda_2(d)$ se tendrá

$$\lambda_1(d) = 1 + \sqrt{1 - |M(d)|}$$

$$\lambda_2(d) = 1 - \sqrt{1 - |M(d)|}$$

Tengamos en cuenta que

$$(x_1 y_2 - x_2 y_1)^2 \leq (x_1^2 + x_2^2)(y_1^2 + y_2^2) = 1$$

con la igualdad si y sólo si

$$\alpha(x_1, x_2) = (y_2, -y_1)$$

que obviamente no tiene solución en X. Por tanto para cualquier $d \in D$,

$$\lambda_1(d) > 1 > \lambda_2(d)$$

Al estudiar la ϕ_p-optimalidad de un diseño hemos de tener en cuenta que minimizar la función $(\frac{1}{2}\lambda_1^{-p}(d) + \frac{1}{2}\lambda_2^{-p}(d))^{1/p}$ es equivalente minimizar $\left(\frac{1}{2}\lambda_1^{-p}(d) + \frac{1}{2}\lambda_2^{-p}(d)\right)$.

Ahora bien, la función $f(x) = x^{-p}$ $0 < p < \infty$ es convexa en el intervalo $(0,\infty)$ de modo que para cualquier diseño $d \in D$

$$\left(\frac{1}{2}\lambda_1(d) + \frac{1}{2}\lambda_2(d)\right)^{-p} = 1 < \frac{1}{2}\left(1 + \sqrt{1 - |M(d)|}\right)^{-p} + \frac{1}{2}\left(1 - \sqrt{1 - |M(d)|}\right)^{-p}$$

Por otra parte, si tomamos el diseño d_n que realiza las observaciones en los puntos

$$x_{(1)}^t = (1/n, \sqrt{1-1/n^2}) \quad y \quad x_{(2)}^t = (\sqrt{1-1/n^2}, 1/n)$$

tendremos que

$$\lambda_1(d_n) = 1 + \sqrt{1 - (1 - 2/n^2)^2} \quad y \quad \lambda_2(d_n) = 1 - \sqrt{1 - (1 - 2/n^2)^2}$$

de modo que

$$\inf_{d \in D} \left[\tfrac{1}{2}\lambda_1^{-p}(d) + \tfrac{1}{2}\lambda_2^{-p}(d)\right] = 1$$

y no existe diseño ϕ_p-óptimo para $0 < p < \infty$.

Obsérvese que si hubiésemos podido realizar una observación en el punto $(1,0)^t$ y otra en $(0,1)^t$ dicho diseño hubiese sido ϕ_p-óptimo para todo p.

4.4.2. X CERRADO Y NO ACOTADO.

Consideremos el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; $\theta^t = (\theta_1, \theta_2)$ es un vector paramétrico desconocido y $x^t = (x_1, x_2)$ es una variable de control que supondremos perteneciente a X, en donde X es el cuadrante no negativo de R^2.

Claramente el espacio X es cerrado y no acotado. Pretendemos encontrar el diseño ϕ_p-óptimo basado en dos observaciones incorreladas de la variable aleatoria Y_x.

Sea d_n el diseño que realiza una observación en el punto $(n,0)^t$ y otra en $(0,n)^t$ en donde n es un número natural. Entonces

$$\text{tr } M^{-p}(d_n) = 2n^{-2p}$$

y en consecuencia

$$\lim_{n\to\infty} \left(\tfrac{1}{2}\text{tr } M^{-p}(d_n)\right)^{p^{-1}} = \lim_{n\to\infty} n^{-2} = 0$$

de modo que no existe diseño ϕ_p-óptimo.

4.5. OPTIMALIDAD UNIVERSAL.

Cuando en el capítulo primero estudiamos la optimalidad universal en el caso en el que las matrices de información fuesen definidas no negativas, comentamos la definición propuesta por Sinha-Mukerjee.- Puede comprobarse, con facilidad, que los criterios de D-, A- optimalidad están incluidos en cualquiera de las dos definiciones de optimalidad universal.

Por consiguiente, los ejemplos dados en la sección 4.2. correspondientes a ambos criterios nos pueden servir para demostrar que sin compacidad no es seguro que exista optimalidad universal en ninguna de las dos definiciones dadas.

Nos proponemos demostrar, sin embargo, que ni aun en el caso en que el espacio de control sea compacto se puede asegurar que existe un diseño universalmente óptimo.

El ejemplo que analizaremos a continuación nos servirá además de para probar que no existe diseño universalmente óptimo, puesto que demostraremos que la A- y D- optimalidad no se alcanzan el el mismo diseño, para comprobar que la optimalidad de un diseño en la teoría de la aproximación puede no ser alcanzada por un diseño basado en un número finito de observaciones.

Sea el modelo lineal

$$Y_x = \theta_1 x_1 + \theta_2 x_2 + e$$

en donde e es un error aleatorio de media cero y varianza σ^2, $\theta^t = (\theta_1,\theta_2)$ es un vector paramétrico desconocido y $x^t = (x_1,x_2)$ es una variable de control que suponemos perteneciente al recinto X, en donde

$$X = \{(x_1,x_2)^t : x_1 = 1 \ , \ x_2 \in [0,1]\}$$

Pretendemos probar que los diseños D- y A- óptimos para este modelo, supuesto que el experimentador puede realizar diez observaciones incorreladas de la variable Y_x, no coinciden. La demostración se hará buscando los diseños óptimos para estos criterios en el contexto de la teoría de la aproximación, para analizar posteriormente la proximidad a estos diseños de un diseño basado en diez observaciones de la variable Y_x.

En el apartado 4.2.1.1. hemos probado que el diseño η^0 que asigna probabilidad 1/2 a los puntos $(1,0)^t$ y $(1,1)^t$ es D-óptimo. En nuestro caso dicho diseño equivaldría a realizar cinco observaciones de la variable aleatoria Y_x en el punto $x^t = (1,0)$ y otras cinco en el -- punto $(1,1)^t$.

Para analizar la A-optimalidad procederemos como sigue: sea η un diseño sobre X, sabemos que su matriz de información puede- escribirse como

$$M(\eta) = \sum_{i=1}^{n} p_i \begin{pmatrix} 1 \\ z_i \end{pmatrix} (1 \;\; z_i)$$

en donde $z_i \varepsilon [0,1]$, $p_i \geq 0$ y $\sum p_i = 1$

Obviamente, toda matriz $M(\eta)$ puede representarse como un punto - de R^3 con coordenadas

$$(1, \sum_i p_i z_i, \sum_i p_i z_i^2) = \sum_i p_i (1, z_i, z_i^2) \qquad z_i \varepsilon [0,1]$$

y en consecuencia, el conjunto M de las matrices de información es la- envolvente convexa de

$$A = \{(1,z,z^2) : 0 \leq z \leq 1\}$$

Veamos, ahora, que todo punto de la envolvente convexa de A, que notaremos co(A), puede expresarse como una combinación lineal convexa del punto $(1,0,0)$ y $(1,z,z^2)$ con $0 \leq z \leq 1$.

Sabemos que

$$co(A) = \{(1,y,u) : y^2 \leq u \leq y \leq 1\}$$

Sea, entonces, $(1,y_0,u_0) \varepsilon co(A)$, consideremos la recta que une - dicho punto con el punto $(1,0,0)$. Supongamos que $y_0 > 0$, pues en caso- contrario la representación es obvia. Las ecuaciones de dicha recta - son, supuesto que las coordenadas de los puntos de R^3 son (x_1,y,u),

$$\frac{y-0}{0-y_0} = \frac{u-0}{0-u_0}$$

$$x_1 = 1$$

es decir

$$x_1 = 1$$

$$u = (u_0/y_0)y$$

Si calculamos su intersección con A tendremos que

$$y^2 = (u_0/y_0)y$$

de donde $y = u_0/y_0$.

Obviamente,

$$(1,y_0,u_0) = (1-\lambda)(1,0,0) + \lambda(1,u_0/y_0,u_0^2/y_0^2)$$

tiene solución en $\lambda \varepsilon [0,1]$.

En consecuencia para todo diseño η, con λ y z apropiados se tiene

$$M(\eta) = \begin{pmatrix} 1 & \lambda z \\ \lambda z & \lambda z^2 \end{pmatrix}$$

Por tanto

$$tr(M^{-1}(\eta)) = \frac{1}{\lambda(1-\lambda)} \times \frac{1+\lambda z^2}{z^2}$$

Ahora bien, para λ fijo $tr(M^{-1}(\eta))$ es decreciente en $z \varepsilon [0,1]$, de modo que hallar el diseño A-óptimo es equivalente a hallar el diseño de la forma

$$M(\eta) = \begin{pmatrix} 1 & \lambda \\ \lambda & \lambda \end{pmatrix} \qquad 0 < \lambda < 1$$

que minimiza $tr(M^{-1}(\eta))$. Pero, puesto que $tr(M^{-1}(\eta)) = (1+\lambda)/\lambda(1-\lambda)$ y el mínimo de dicha función se alcanza en $\lambda = \sqrt{2} - 1$, podemos afirmar que el diseño A-óptimo, η^1, es aquel que asigna probabilidad $\sqrt{2} - 1$ a $(1,1)^t$ y probabilidad $2 - \sqrt{2}$ a $(1,0)^t$.

Su matriz de información tiene la forma

$$M(\eta^1) = \begin{pmatrix} 1 & \sqrt{2} - 1 \\ \sqrt{2} - 1 & \sqrt{2} - 1 \end{pmatrix}$$

Obviamente, ningún diseño basado en un número finito de observaciones puede tener como matriz de información $M(\eta^1)$ y en consecuencia la optimalidad bajo la teoría de la aproximación no es alcanzable por diseños finitos.

Por otra parte, si consideramos en nuestro problema original el diseño d^1 que realiza cuatro observaciones en el punto $(1,1)^t$ y seis en $(1,0)^t$ entonces

$$tr(M^{-1}(d^1)) = 7/12$$

Ahora bien, llamando d^0 al diseño D-óptimo se tendrá

$$tr(M^{-1}(d^0)) = 3/5$$

de modo que

$$tr(M^{-1}(d^0)) > tr(M^{-1}(d^1))$$

y en consecuencia, d^0 no puede ser A-óptimo. Por tanto, no existe diseño universalmente óptimo para este problema.

4.6. CRITERIOS DE OPTIMALIDAD DE TIPO I Y II.

4.6.1. X ACOTADO Y NO CERRADO.

Es obvio que la búsqueda de ejemplos de la no existencia de diseños óptimos para este tipo de criterios no puede pasar por el estudio de todas y cada una de las funciones que pueden considerarse como criterios de optimalidad de tipo I y II.

El ejemplo que proponemos es unidimensional y f notará cualquier función que cumpla las condiciones impuestas por Cheng(1.978) para un criterio de optimalidad de tipo I y II.

Sea el modelo lineal

$$Y_x = \theta x + e$$

en donde e es un error aleatorio de media nula y varianza σ^2; θ es un- parámetro desconocido y x es una variable de control que suponemos perteneciente al intervalo $[0,1)$.

Para cualquier diseño d basado en N puntos, $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$, -- del espacio de control es claro que

$$M(d) = \sum_i x^2_{(i)} < N$$

En consecuencia, si D es la clase de tales diseños

$$\sup_{d \in D} trM(d) = N$$

Obsérvese que este supremo no es máximo como exige Cheng en la - formulación de sus criterios de optimalidad, sin embargo, supondremos que podemos escribir supremo en lugar de máximo.

Sea f cualquier función que cumpla las condiciones correspondientes a un criterio de optimalidad de tipo I o II. Es claro que en cualquiera de los casos f es estrictamente decreciente y continua, por -- tanto

$$\phi_f(M(d)) = f(\sum_i x^2_{(i)}) < f(N)$$

de forma que

$$\inf_{d \in D} \phi_f(M(d)) = f(N)$$

y por tanto,no existe diseño óptimo con respecto a la función f.

Obsérvese que el argumento ha sido utilizado para funciones f -- que cumplen las condiciones de los criterios de tipo I o II, pues en- ambos casos son continuas y estrictamente decrecientes en $[0,N]$.

4.6.2. X CERRADO Y NO ACOTADO.

Es claro que el mismo modelo del apartado anterior nos puede servir para este estudio supuesto que el espacio de control es el conjun to

$$X = \{x \in R : x \geq 0\}$$

Es obvio que para esta nueva situación no existe diseño que sea óptimo para estos criterios de optimalidad.

4.7. SCHUR-OPTIMALIDAD.

Teniendo en cuenta, como vimos en el primer capítulo, que la D-, A-, E- y ϕ_p- optimalidad son funciones Schur-convexas, los ejemplos correspondientes a estos criterios nos pueden servir aquí para estudiar el caso en que X no sea compacto. Por otra parte, aun subsistiendo la hipótesis de compacidad de X, el análisis realizado en la sección 4.5. para la optimalidad universal también es válido aquí, puesto que la A- y D- optimalidad son funciones, como hemos dicho antes, Schur-convexas.

APÉNDICE 1

TEORÍA DE LA APROXIMACIÓN

A.1.0. RESUMEN.

Cuando consideraciones de tipo geométrico nos lleven a definir - y utilizar un determinado criterio de optimalidad, nuestro problema - se habrá convertido en uno de optimización. Pero teniendo en cuenta - la naturaleza discreta de los diseños basados en un número finito de- observaciones, la caracterización de las soluciones óptimas puede resultar complicada.

Nuestra situación puede considerarse parecida a aquella en la -- que se pretende minimizar una función definida sobre los enteros. Es- sabido, sin embargo, que una simplificación del problema es extender- la definición de la función a los reales, encontrar la solución óptima sobre la recta real y una vez hallado el óptimo del problema extendido, buscar la solución en los enteros más próxima.

La anterior idea fue adaptada al problema de diseño óptimo por - Kiefer y recibió el nombre de teoría de la aproximación para diseños- en regresión lineal.

La teoría de la aproximación, que ahora describiremos, se ha utilizado principalmente para los criterios de optimalidad propuestos en las secciones segunda y cuarta del primer capítulo, sin ninguna res-- tricción, salvo claro está que el ser definidas no negativas, para -- las matrices de información. Nosotros nos restringiremos aquí a estos- criterios. En consecuencia, en este apéndice ϕ notará genéricamente,- cualquier criterio de optimalidad que cumpla las condiciones expues-- tas.

A.1.1. ELEMENTOS DE LA TEORIA DE LA APROXIMACION

A.1.1.1. PROPIEDADES Y DEFINICIONES.

Para cualquier función ϕ ,antes citada, es obvio que se cumple la siguiente propiedad: Dada a, constante positiva

(1) $$\phi(aM(d)) = h\phi(M(d))$$

en donde h es una constante positiva independiente del diseño d, generalmente depende del número de observaciones, de modo que si d^0 miniza $\phi(aM(d))$ también minimizará $\phi(M(d))$.

Nota 1: Es obvio que las funciones Φ que estamos citando pueden definirse sobre el conjunto de matrices definidas no negativas, de formaque tiene sentido hablar de $\Phi(aM(d))$ cuando a es una constante positiva

Sea, entonces, un diseño d, éste contendrá un número n de vectores distintos $x_1, x_2, \ldots, x_n$ que se observarán $r_1, r_2, \ldots, r_n$ veces respectivamente con $\Sigma r_i = N$. Es obvio que

$$M(d) = X^t.X = \sum_{j=1}^{N} x_j x_j^t = \sum_{i=1}^{n} r_i x_i x_i^t$$

Por otra parte, si definimos sobre X, el espacio de control, ladistribución de probabilidad $\eta_N(d)$ que asigna probabilidad $p_i = r_i/N$ al punto x_i, i=1,2,...,N y llamamos $M(\eta_N(d))$ a $E_{\eta_N(d)}[X.X^t]$ en donde X es un vector aleatorio con distribución $\eta_N(d)$, entonces

$$M(\eta_N(d)) = E_{\eta_N(d)}[X.X^t] = \sum_{i=1}^{n} p_i x_i x_i^t = N^{-1}M(d)$$

y teniendo en cuenta (1) podemos reinterpretar el encontrar un diseño d^0 que sea ϕ-óptimo como " encontrar $\eta_N(d^0)$ perteneciente a H_N, el conjunto de distribuciones de probabilidad sobre X que provienen de diseños basados en N observaciones, que minimice $\phi(M(\eta_N(d)))$".

Este nuevo planteamiento no mejora, por si sólo la resolución del problema. Sin embargo, podemos extender la definición de $M(\eta_N(d))$

al conjunto H formado por todas las distribuciones de probabilidad sobre X, como sigue:

Si $\eta \varepsilon H$ y X es un vector aleatorio con distribución η, entonces

$$M(\eta) = E_{\eta}[X.X^{t}]$$

extendiendo la definición de ϕ a la clase de matrices

$$\mathcal{M} = \{M(\eta) \ : \eta \varepsilon H \}$$

Nuestro problema se ha transformado, por tanto, en encontrar una distribución η^{0} sobre X que minimice $\phi(M(\eta))$ sobre H. Este problema-- es evidentemente mucho más tratable que el anterior, entre otras razones por poseer $\mathcal{M}$ propiedades de las que carece $\mathcal{M}_{N}$, en donde

$$\mathcal{M}_{N} = \{M(\eta_{N}(d)) \ : \ \eta_{N}(d) \varepsilon H_{N}\}$$

Una vez hallada η^{0} que sea solución óptima para el problema ampliado, buscaremos un diseño d^{0} que tenga una distribución de probabilidad asociada que esté próxima a η^{0} y evidentemente este diseño estará próximo al óptimo.

Damos ahora unas definiciones que son usuales en la teoría de la aproximación.

Definición A.1.1. Cualquier distribución de probabilidad η de H recibirá el nombre de medida de diseño.

Definición A.1.2. Para cualquier medida de diseño η

$$M(\eta) = E_{\eta}[X.X^{t}]$$

recibe el nombre de matriz de información de la medida de diseño η.

Es obvio que según las definiciones anteriores los conjuntos H y $\mathcal{M}$ recibirán los nombres de conjunto de medidas de diseños y conjunto de matrices de información, respectivamente.

Nota 2: En toda la literatura sobre optimalidad en modelos lineales,- el término diseño se ha aplicado indistintamente a los diseños defini dos en el capítulo primero, definición 1.2.1., y las medidas de diseño que hemos definido anteriormente. Este hecho que en principio puede pensarse que ha de inducirnos a errores no presenta complicaciones por dos causas fundamentalmente:

(1) Los diseños d y las medidas de diseño η, aparecen en contextos diferentes, más claramente, los primeros se insertan en el problema de diseño óptimo sin la utilización de ningún tipo de herramienta matemática adicional, mientras que las medidas de diseño sólo aparecen en el contexto de la teoría de la aproximación.

(2) Es aceptada universalmente la notación con las letras η y ξ a las medidas de diseño, de forma que la utilización de estos símbolos implica que trabajamos bajo la teoría de la aproximación.

En la presente memoria el término diseño es aplicado indistintamente a los diseños según la definición 1.2.1. y a las medidas de diseños de la definición A.1.1.. Sólo en caso de posible ambigüedad serán utilizados los términos diseño y medida de diseño según estas definiciones.

A.1.1.2. EL CONJUNTO DE LAS MATRICES DE INFORMACIÓN.

i) Es obvio que cada elemento de M, el conjunto de las matrices de información, es una matriz simétrica, definida no negativa, de orden k que puede representarse como un punto de $R^{\frac{1}{2}k(k+1)}$

ii) Por otra parte, siguiendo Sibson (1.974), al ser X compacto

$$M(X) = \{x.x^t : x \varepsilon X\}$$

es compacto, de modo que su envolvente convexa, co(M(X)), también lo es. Ahora bien, para cada elemento de M habrá un elemento de co(M(X)) arbitrariamente próximo a él, de forma que por las propiedades del conjunto co(M(X)) se tendrá:

$$M = co(M(\mathcal{X}))$$

iii) Teniendo en cuenta el teorema de Caratheodory, cada elemento de M podrá ser expresado como una combinación lineal convexa

$$\sum_{i=1}^{l} \lambda_i x_i x_i^t$$

con $x_i \varepsilon X$ y $l \leq \frac{1}{2}k(k+1)+1$. Por el mismo teorema, si una matriz es punto frontera de M entonces $l \leq \frac{1}{2}k(k+1)$.

iv) Desde el punto de vista práctico esta situación es extremadamente ventajosa puesto que si la solución óptima para ϕ se alcanza en M^0, sabemos que M^0 puede ser expresada siempre como $M(\eta^0)$ siendo η^0 un diseño (una distribución de probabilidad) cuyo soporte tiene a lo sumo $\frac{1}{2}k(k+1)+1$ puntos. Es decir, siempre hay un diseño discreto que alcanza la solución óptima.

v) Por otra parte, no olvidemos que nuestro último proyecto es aproximar un diseño basado en N observaciones al diseño óptimo y esto es evidentemente más fácil si la solución es un diseño con soporte discreto

A.1.2. LOS CRITERIOS DE OPTIMALIDAD BAJO LA TEORÍA DE LA APROXIMACIÓN

a) G-OPTIMALIDAD.

Supongamos que X, el espacio de control, es un subconjunto de R^k que genera R^k, es decir, no incluído en ningún subespacio propio.

Para cada $x \varepsilon X$ y $\eta \varepsilon H$ definimos

$$\begin{aligned} d(x,\eta) &= x^t M^{-1}(\eta)\, x && \text{si } M(\eta) \text{ es no singular} \\ &= \infty && \text{si } M(\eta) \text{ es singular} \end{aligned}$$

Definición A.1.3. $\eta^0 \varepsilon H$ es G-óptimo si y sólo si

$$\max_{x\varepsilon X} d(x,\eta^0) = \min_{\eta\varepsilon H} \max_{x\varepsilon X} d(x,\eta)$$

Es obvio que esta definición extiende el concepto de G-optimalidad, dado en el capítulo primero, a todas las medidas de diseños y que el diseño $\eta_N(d^0)$ más próximo al óptimo será la mejor solución para el problema de diseño basado en N observaciones de variables de control.

Nota 3: Obsérvese que

$$d(x,\eta) = x^t M^{-1}(\eta)x$$

puede considerarse, salvo una constante, como la extensión de la definición de la varianza del mejor estimador lineal insesgado de $x^t\theta$ bajo cualquier diseño η.

b) D-OPTIMALIDAD.

Minimizar $|M^{-1}(d)|$ es claramente equivalente a minimizar el determinante de $M^{-1}(\eta_N(d))$, de forma que la definición de D-optimalidad en el contexto de la teoría de la aproximación queda

Definición A.1.4. Un diseño $\eta^0 \varepsilon$ H es D-óptimo si y sólo si

$$|M^{-1}(\eta^0)| = \min_{\eta \varepsilon H} |M^{-1}(\eta)|$$

c) D_A - y D_s - OPTIMALIDAD.

Supongamos que estamos interesados en la estimación de algunas combinaciones lineales de los parámetros desconocidos, a saber $A^t\theta$, en donde A^t es una matriz de orden $s \times k$ y rango $s<k$. Puesto que $s<k$, en la teoría sobre modelos lineales algunos diseños d con matriz de información singular pueden estimar $A^t\theta$. Esto ha de ser tenido en cuenta en la teoría de la aproximación. Por analogía con la teoría de los mínimos cuadrados podemos decir que una medida de diseño η nos permite estimar $A^t\theta$ si y sólo si $A = M(\eta)Y$ tiene solución en Y, obsérvese que cuando una medida de diseño proviene de un diseño basado en N observaciones esta definición es equivalente a la clásica en modelos lineales. Sea M_A el subconjunto de M formado por todas las matrices que cumplen la propiedad anterior; M_A contiene todas las matrices $M(\eta)$ --

que son no singulares y algunas singulares.

De nuevo por analogía con los resultados en la teoría de los mínimos cuadrados, la matriz de covarianzas del estimador por mínimos cuadrados de $A^t\theta$ que proviene de una medida de diseño η con matriz de información $M(\eta)\in M_A$ es proporcional a $A^t\bar{M}(\eta)A$, en donde $M^-(\eta)$ es cualquier inversa generalizada de $M(\eta)$, obsérvese el paralelismo con la matriz $A^tM^-(d)A$ en los modelos lineales. Por otra parte, $A^tM^-(\eta)A$ es una matriz definida no negativa de orden $s\times s$. Teniendo en cuenta que $A^tM^-(\eta)A$ es definida positiva si y sólo si $A = M(\eta)Y$ tiene solución en Y, podemos definir la D_A-optimalidad en el contexto de la teoría de la aproximación como sigue:

Definición A.1.5. Un diseño $\eta^0\in H$ es D_A-óptimo si y sólo si

1) $A^tM^-(\eta^0)A$ es no singular

2) $|A^tM^-(\eta^0)A| = \min\limits_{\eta\in H} |A^tM^-(\eta)A|$

La D_s-optimalidad, en el contexto de la teoría de la aproximación no es más que la reformulación de la D_A-optimalidad tomando $A^t = (I_s\ 0)$. En este caso, si

$$M(\eta) = \begin{pmatrix} M_{11}(\eta) & M_{12}(\eta) \\ M_{21}(\eta) & M_{22}(\eta) \end{pmatrix}$$

en donde $M_{11}(\eta)$ es de orden $s\times s$, puede probarse que si $M(\eta)\in M_A$

$$A^tM^-(\eta)A = (M_{11}(\eta) - M_{12}(\eta)M_{22}^-(\eta)M_{21}(\eta))^{-1}$$

(Karlin-Studden 1.966), de modo que tenemos la siguiente definición

Definición A.1.6. Un diseño $\eta^0\in H$ es D_s-óptimo si y sólo si

$$|M_{11}(\eta^0) - M_{12}(\eta^0)M_{22}^-(\eta^0)M_{21}(\eta^0)| = \max_{\eta\in H} |M_{11}(\eta) - M_{12}(\eta)M_{22}^-(\eta)M_{21}(\eta)|$$

d) L-OPTIMALIDAD.

Los criterios de optimalidad lineales se formulaban en función - de $tr(A^t M^-(d)A)$, de manera que, bajo la teoría de la aproximación la- L-optimalidad se expresa en función de la minimización de $tr(A^t M^-(\eta)A)$ En particular

Definición A.1.7. Un diseño $\eta^0 \in H$ es A-óptimo si y sólo si

1) $M(\eta^0)$ es no singular

2) $tr(M^{-1}(\eta^0)) = \min_{\eta \in H} tr(M^{-1}(\eta))$

Obsérvese que por paralelismo con la teoría de los modelos linea- les, el exigir que $M(\eta)$ sea no singular puede considerarse, hablando- libremente, equivalente a que la medida de diseño η (el diseño η) es- time el vector paramétrico θ.

e) E-OPTIMALIDAD.

Sean $\mu_1(\eta) \geq \mu_2(\eta) \geq \ldots \geq \mu_k(\eta)$ los autovalores de $M(\eta)$, enton- ces bajo la teoría de la aproximación tenemos:

Definición A.1.8. Un diseño $\eta^0 \in H$ es E-óptimo si y sólo si

$$\mu_k^{-1}(\eta^0) = \min_{\eta \in H} \mu_k^{-1}(\eta)$$

f) ϕ_p-OPTIMALIDAD.

Este criterio fue formulado, de hecho, por Kiefer (1.974) en el- contexto de la teoría de la aproximación.

Definición A.1.9. Un diseño $\eta^0 \in H$ es ϕ_p-óptimo si y sólo si

$$\left[\frac{1}{k} tr(M^{-p}(\eta^0))\right]^{\frac{1}{p}} = \min_{\eta \in H} \left[\frac{1}{k} tr(M^{-p}(\eta))\right]^{\frac{1}{p}} \qquad p \geq 0.$$

BIBLIOGRAFIA

-ASH, A. and HEDAYAT, A.(1978). An introduction to design optimality - with an overview of the literature. Commun. Statist. A7,14,1295-1.325

-ATKINSON, A.C.(1.982). Developments in the design of experiments. Int. Statistc. Rew.,50,161-177.

-ATWOOD, C.L.(1.969). Optimal an efficient designs of experiments. Ann. Math. Statist.,40,1570-1602.

-BELLMAN, R.(1.965). Introducción al análisis matricial. Ed: Reverté,S.A.

-BILLINGSLEY, P.(1.979). Probability and Measure. New York: Wiley.

-BOX, G.E.P. and DRAPER, N.R.(1.959). A basis for the selection of a - response surface design. J, Amer. Statist.,54,622-654.

-BOX, G.E.P. and HUNTER, W.G.(1.965a). Sequential design of experiments for non linear models. Proc. IBM Sc. Comp. Symp.1.963, IBM, New York.

-BOX, G.E.P. and HUNTER, W.G.(1.965b). The experimental study of physical mechanisms. Technometrics,7,23-42.

-BOX, G.E.P. and LUCAS, H.L.(1.959). Design of experiments in non-linear situations. Biometrika,46,77-90.

-BOX, G.E.P. and WILSON, K.B.(1.951). On the experimental attainment - of optimum conditions. J. R. Statist. Soc. B.,13,1-45.

-BRADLEY, R.A. and HEDAYAT, A.(1.977). Optimal designs for two non-interactive treatments. Technometrics,19,53-57.

-CHENG, C.-S.(1.978). Optimality of certain asymmetrical experimental-designs. Ann. Statist,6,1239-1261.

-CHENG, C.-S.(1.980). On the E-optimality of some block designs. J. R. Statist. Soc. B.,42,199-204.

-CHERNOFF, H.(1.953). Locally optimal designs for estimating parame--- ters. Ann. Math. Statist. 24,586-602.

-CHERNOFF, H.(1.959). Sequential design of experiments. Ann. Math. Statist. 30,755-770.

-CONSTANTINE, G.M.(1.981). Some E-optimal block designs. Ann. Statist, 9,886-892.

-de la GARZA, A.(1.954). Spacing of information in polynomial regres-- sion. Ann. Math. Statist., 25, 123-130.

-de la GARZA, A.(1.956). Quadratic extrapolation and a related test of hypothesis. J. Amer. Statist. Assoc., 27, 644-649.

-ECCLESTON , J.A. and HEDAYAT, A.(1.974). On the theory of connected - designs: characterization and optimality. Ann. Statist,2,1238-1255.

-EHRENFELD, S.(1.955). On the efficiency of experimental designs. Ann. Math. Statist., 26, 247-255.

-ELFVING,G.(1.952). Optimum allocation in linear regression theory. -- Ann. Math. Statist., 23, 255-262

-ELFVING,G.(1.955). Geometric allocation theory. Skand. Aktuarietidskr. 37, 170-190.

-ELFVING,G.(1.959). Design of linear experiments. Cramer Festschrift - Volume. New York: Wiley, 58-74.

-FEDOROV, V.V.(1.972). Theory of optimal experiments. New York: Academic Press.

-FEDOROV, V.V. and MALYUTOV, M.B.(1.972). Optimal design in regression experiments. Math. Operationsforsch. Statist., 14, 237-324.

-GRAYBILL, F.A.(1.961). An introduction to linear statistical models -- (Volume I). McGraw-Hill. Series in Probability and Statistics.

-HEDAYAT, A.(1.981). Study of optimality criteria in design of experiments. In Statistics and Related Topics.(Csörgo, ed). North-Holland.

-HILL, W.J. and HUNTER, W.G.(1.974). Design of experiments for subsets of parameters. Technometrics, 16, 425-434.

-HUNTER, W.G., HILL, W.J. and HENSON, T.L.(1.979). Designing experi--- ments for precise estimation of some or all of the constants in a mechanistic model. Can. Journ. Chem. Eng., 47, 76-80.

-JACROUX, M.(1.978). On the properties of proper (M,S)-optimal block-designs. Ann. Statist., 6, 1.302-1.309.

-JACROUX, M.(1.980). On the E-optimality of regular graph designs. J. R. Statist. Soc. B., 42, 205-209.

-JACROUX, M. (1.982). Some E-optimal designs for the one-way and two-way elimination of heterogeneity. J. R. Statist. Soc. B,44, 253-261.

-JACROUX, M. and SEELY, J.(1.980). Some sufficient conditions for establishing (M,S)-optimality. J. Statist. Plann. Inference, 4, 3-11.

-JACROUX, M., WONG, C.C. and MASARO, J.C.(1.983). On the optimality - of chemical balance weighing designs. J. Statist. Plann. Inference., 8, 231-240.

-JOHN, J.A. and MITCHELL, T.J.(1.977). Optimal incomplete block de--- signs. J. Roy. Statist. Soc. B., 39, 39-43.

-JOHN, J.A. and Williams, E.R.(1.982). Conjectures for optimal block-designs. J. R. Statist. Soc. B., 44, 221-225.

-KARLIN, S. and STUDDEN, W.J.(1.966). Optimal experimental designs. - Ann. Math. Statist., 37, 783-815.

-KEMPTHORNE, O.(1.956). The efficiency factor on an incomplete block-design. Ann. Math. Statist., 27, 846-849.

-KIEFER, J.(1.958). On the nonrandomized optimality and randomized - non-optimality of symmetrical designs. Ann. Math. Statist.29,675-699.

-KIEFER, J.(1.959). Optimum experimental designs. J. R. Statist. Soc. B., 21, 272-319.

-KIEFER, J.(1.960). Optimum experimental designs V, with applications to systematic and rotatable designs. Proc. Fourth Berkeley Symp.Math. Statist. Prob.Los Angeles and Berkeley: Univ. of California Press, - 381-405.

-KIEFER, J.(1.961). Optimum designs in regression problems, II. Ann.- Math. Statist., 32, 298-325.

-KIEFER, J.(1.962a). An extremum result. Canad. J. Math., 14, 597-601.

-KIEFER, J.(1.962b). Two more criteria equivalent to D-optimality of designs. Ann. Math. Statist., 33, 792-796.

-KIEFER, J.(1.974). General Equivalence theory for optimum designs -- (approximate theory). Ann. Statist., 2, 849-879.

-KIEFER, J.(1.975). Constructions and optimality of generalized Youden designs. A Survey of Statistical Design and Linear Models (J. N. Srivastava, ed.). Amsterdam: North-Holland Publishing Cp., 333-353.

-KIEFER, J. and WOLFOWITZ, J.(1.960). The equivalence of two extremum problems. Canad. J. Math., 12, 363-366.

-KIEFER, J. and WOLFOWITZ, J.(1.964a). Optimum extrapolation and interpolation designs, I. Ann. Inst. Statist. Math., 16, 79-108.

-KIEFER, J, and WOLFOWITZ, J.(1.964b). Optimum extrapolation and interpolation designs,II. Ann. Inst. Statist. Math., 16, 295-303.

-LAYCOCK, P.J.(1.972). Convex loss applied to design in regression -- problems. J. R. Statist. Soc. B., 34, 148-170.

-MAGDA, C. G.(1.979). On E-optimality and Schur-optimality. Ph. D. -- Thesis. University of Illinois at Chicago Circle.

-MOOD, A.M.(1.946). On Hotelling's weighing problem. Ann. Math. Statist., 17, 432-446.

-PUKELSHEIM, F.(1.980). On linear regression designs which maximize - information. J. Statist. Plann. Inference, 4, 339-364.

-PUKELSHEIM, F.(1.983). On optimality properties of simple block designs in the approximate design theory. J. Statist. Plann. Inference. 8, 193-208.

-ROY, B.K.(1.982). Construction of (M,S)-optimal design for block size 3. J. Statist. Plann. Inference., 7, 35-37.

-SEARLE, S.R.(1.971). Linear Models. New York: Wiley.

-SEARLE, S.R.(1.982). Matrix algebra useful for Statistics. New York: Wiley.

-SHAH, K.R.(1.960). Optimality criteria for incomplete block designs. Ann. Math. Statist., 31, 791-794.

-SIBSON, R.(1.972). Contribution to discussion of " Results in the -- theory and construction of D-optimum experimental designs" by H.P. - -WYNN. J. R. Statist. Soc. B., 34, 181-183.

-SIBSON, R.(1.974). D_A-optimality and duality. Progress in Statistics. Colloq. Math. Soc. Janos Bolyai., 9, 677-692.

-SILVEY, S.D.(1.972). Contribution to discussion of " Results in the- theory and construction of D-optimum experimental designs" by H. P.- -WYNN. J. R. Statist. Soc. B., 34, 174-175.

-SILVEY, S.D.(1.980). Optimal Design. Monographs on applied Probability and Statistics. London: Chapman and Hall.

-SILVEY, S.D. and TITTERINGTON, D.M.(1.973). A geometric approach to- optimal design theory. Biometrika., 60, 21-32.

-SILVEY, S.D. and TITTERINGTON, D.M.(1.974). A Lagrangian approach to optimal design. Biometrika., 61, 299-302.

-SILVEY, S.D., TITTERINGTON, D.M. and TORSNEY, B.(1.978). An algorithm for optimal designs on a finite design space. Commun Statist. A7, 14, 1.379-1389.

-SINHA, B.K. and MUKERJEE, R.(1.982). A note on the unoversal optimality criterion for full rank models. J. Statist. Plann. Inference.,- 7, 97-100.

-SMITH, K.(1.918). On the standard desviations of adjusted and interpolated values of an observerd polynomial function and its constants and the guidance they give towards a proper choice of the distribu-- tion of observations. Biometrika., 12, 1-85.

-WALD, A.(1.943). On the efficient design of statistical investiga--- tions. Ann. Math. Statist., 14, 134-140.

-WHITTLE, P.(1.971). Optimization under constraints. New York: Wiley

-WHITTLE, P.(1.973). Some general points in the theory optimal experimental design. J. R. Statist. Soc. B., 35, 123-130.

-WU, C.-F. and WYNN, H.P.(1.978) The convergence of general steplength algorithms for regular optimum design criteria. Ann. Statist. 6, 1273 1285.

-WYNN, H.P.(1.970). The sequential generation of D-optimal experimen-- tal designs. Ann. Math. Statist., 41, 1655-1664.

-WYNN, H.P.(1.972). Results in the theory and construction of D-opti-- mun experimental designs. J. R. Statist. Soc. B., 34, 133-147.

-WYNN, H.P.(1.973). Simple conditions for optimal design algorithms. A Survey of Statistical Design and Linear Models (J. N. Srivastava, ed). Amsterdam: North-Holland Publishing Co., 571-579.