



Universidad Nacional
Autónoma de México.
Facultad de Ciencias.

"Un Enfoque Geométrico en el
Problema Numérico de Estimación
de Parámetros por el Método de
Mínimos Cuadrados"

Tesis que presenta:

José López Estrada

para obtener el Título de Matemático.

México, D. F.

1975.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

A mis Padres :

José López

Inés Estrada

A mis Hermanos y Amigos

A mi Maestro

Pablo Barrera

Índice

I	Introducción	1
	Sección 1. Estimación	1
	1.1 Estimación Lineal	3
	1.2 Estimación No-Lineal	6
	Sección 2. Papel del Residual en la Estimación	10
	2.1 Papel del Residual en la Estimación Lineal	11
	2.2 Papel del Residual en la Estimación No-Lineal	16
II	Estimación No-Lineal de un Parámetro	23
	Sección 1. Condiciones Suficientes para un Mínimo	23

	1.1	Curvas en el Plano	25
	1.2	Curvas Planas	29
	1.3	Curvas en General	35
	Sección 2. Método de Gauss		37
	2.1	Convergencia	39
III	Estimación No-Lineal de Varios Parámetros		46
	Sección 1. Condiciones Suficientes		47
	1.1	Superficies en el Espacio	51
	1.2	Curvas de Curvatura Mínima en una Superficie	59
	1.3	Interpretación de la Segunda Condición para Superficies	68
	1.4	Interpretación de la Segunda Condición para el Caso General	79
	Sección 2. Método de Gauss		83

2.1 Convergencia

85

Bibliografía

89

Capítulo I

Introducción.

En la primera parte de este trabajo — planteamos el problema central que ocupa nuestro interés, que es la estimación de parámetros utilizando el criterio de Gauss, mejor conocido con el nombre de mínimos cuadrados, e introducimos una estructura geométrica que nos — permitirá atacar sus dificultades numéricas.

1. Estimación.

Con frecuencia se presenta el problema de determinar los parámetros de una función — que se "ajuste" en lo mejor posible a una colección de puntos, que generalmente representan observaciones. Un criterio usual para este ajuste es el criterio de mínimos cuadrados; este criterio es el que emplearemos en el presente trabajo.

Ahora bien, si escribimos lo dicho en el párrafo anterior en lenguaje formal se tiene.

Se conoce la función

$$(1.1) \quad F(x; a_1, a_2, \dots, a_n)$$

que depende de los parámetros a_1, a_2, \dots, a_n — los cuales se desean determinar de tal manera que

$$(1.2) \quad \sum_{i=1}^m \gamma_i^2 = \sum_{i=1}^m (F(x_i; a_1, a_2, \dots, a_n) - y_i)^2 \\ = \| \bar{F}(\bar{a}) - \bar{y} \|^2$$

sea mínima, en donde x_i, y_i ($i=1, \dots, m$) son las componentes de los puntos P_i (ver figura 1.A).

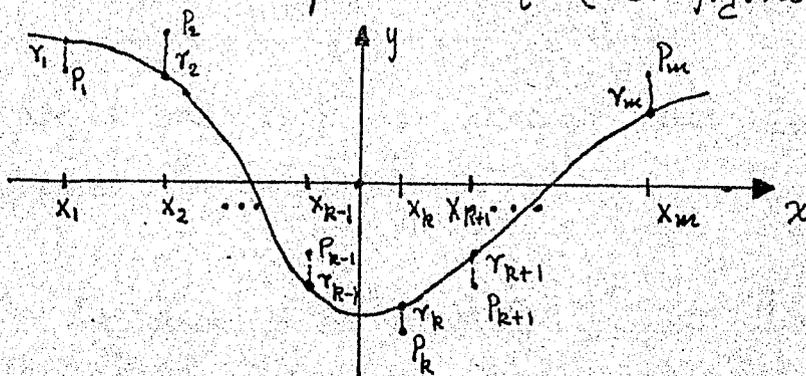


figura 1.A

Como podemos ver el problema de estimación consiste en determinar \bar{a}^* de tal manera que la función

$$(1.3) \quad H(\bar{a}) = \|\bar{F}(\bar{a}) - \bar{y}\|^2$$

alcanza un mínimo en este punto.

1.1 Estimación Lineal.

Con bastante frecuencia la función F en (1.1) se puede expresar en la forma lineal

$$(1.4) \quad \begin{aligned} F(x; a_1, a_2, \dots, a_n) &= a_1 \phi_1(x) + a_2 \phi_2(x) + \dots + \\ &\quad + a_n \phi_n(x) \\ &= \bar{a}^t \bar{\phi}(x) \end{aligned}$$

motivo por el cual esta es conocida como Estimación Lineal de parámetros.

En base a esto nuestra función H a minimizar ahora toma la forma vectorial

$$(1.5) \quad H(\bar{a}) = \|A\bar{a} - \bar{y}\|^2$$

en donde

$$(1.6) \quad A = \begin{bmatrix} \varphi^t(x_1) \\ \vdots \\ \varphi^t(x_m) \end{bmatrix} \quad \text{y} \quad \bar{y}^t = [y_1, y_2, \dots, y_m],$$

aquí, el problema de estimación lineal tiene la interpretación geométrica siguiente: determinar \bar{a}^* tal que $A\bar{a}^*$ sea el punto del subespacio generado por A más cercano a \bar{y} (observarse la figura 1.B).

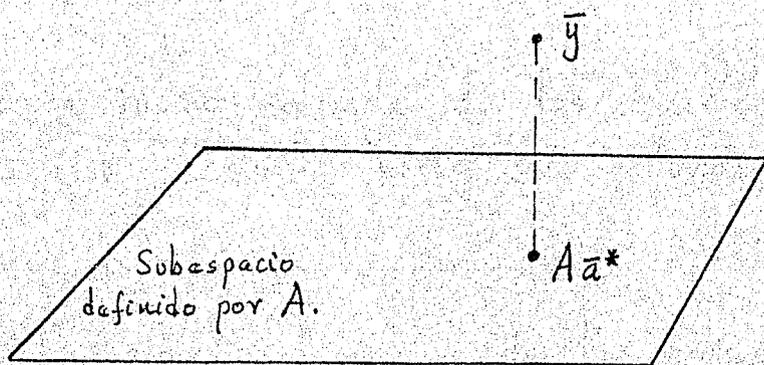


figura 1.B

Este problema de estimación lineal tradicionalmente se resolvía numéricamente a

través de las ecuaciones normales, esto es, si uno busca los \bar{a} que anulen el gradiente

$$(1.7) \quad H' = \nabla H = 2(A^t A \bar{a} - A^t \bar{y})$$

se obtiene

$$(1.8) \quad A^t A \bar{a} = A^t \bar{y}$$

si $A^t A$ es invertible, entonces como

$$(1.9) \quad H''(\bar{a}) = 2 A^t A$$

entonces la única solución de las ecuaciones normales (1.8) es el mínimo que estamos buscando.

La solución numérica de las ecuaciones normales da resultados aceptables solo en aquellos casos en que la condición* [1, 2, 3] de A

$$(1.10) \quad K(A) = \|A\| \|A^+\|$$

* $\|A\| = \max_{\|\bar{x}\|=1} \|A\bar{x}\|$, A^+ denota la pseudo inversa de A .

y la magnitud del residual

$$(1.11) \quad \bar{r} = A\bar{a} - \bar{y}$$

no son grandes (esto se precisa más adelante) ya que en caso contrario la solución de (1.8) — es bastante sensible a perturbaciones.

En la actualidad se prefiere usar métodos que toman en cuenta estas dificultades, tales como el método de "Gram-Schmit modificado" [3], y otros basados en la "descomposición singular" [1, 3, 11].

1.2 Estimación No-Lineal.

El problema de estimación de parámetros da lugar a la minimización de la función

$$H(\bar{a}) = \|\bar{F}(\bar{a}) - \bar{y}\|^2,$$

cuando la función \bar{F} es lineal se tiene un caso especial. El complemento de los problemas

de estimación lineal es conocido como estimación no-lineal, este conjunto de problemas es el tema de estudio de nuestro trabajo.

Como se puede ver este problema de estimación no-lineal consiste, geométricamente, en calcular las coordenadas del punto de la superficie definida por \bar{F} que está más cercano al punto \bar{y} , en general, esta superficie es de dimensión n (vease la figura 1.C).

Consideraremos solamente el caso en que hay un solo mínimo, ya que como veremos éste presenta serias dificultades.

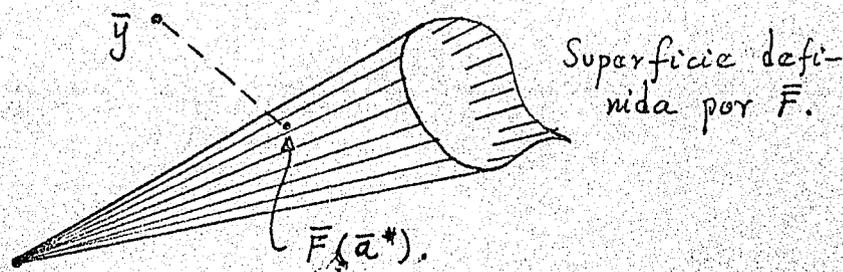


figura 1.C

En este caso calcular los caros del —

gradiente no es muy recomendable ya que solamente nos interesan aquellos en que la segunda derivada (el Hessiano) es positiva definida.

Si el problema que se tiene se ve como un problema de minimización de una función, entonces se puede aplicar a nuestro problema, determinar la solución numérica, los métodos iterativos que se conocen [5] y que usualmente consisten en generar una sucesión $\{\bar{a}_n\}$ tal que

$$\text{i)} \quad H(\bar{a}_{k+1}) < H(\bar{a}_k) ,$$

$$\text{ii)} \quad \bar{a}_{k+1} = \bar{f}(\bar{a}_k) ,$$

$$\text{iii)} \quad \lim_{n \rightarrow \infty} \bar{a}_n = \bar{a}^* .$$

donde \bar{a}^* es el mínimo buscado, y \bar{f} depende por lo general del método empleado [4,5].

Sin embargo es conocido que "los -

mejores resultados" son de aquellos métodos ite
rativos que aprovechan el hecho de que H
es una suma de cuadrados, estos métodos —
son del tipo siguientes:

$$(1.12) \quad \bar{a}_{k+1} = \bar{a}_k + \alpha_k \bar{h}_k$$

donde \bar{h}_k es el vector que minimiza

$$(1.13) \quad \|F_k \bar{h}_k - \bar{y}\|^2,$$

donde F_k es una aproximación de la función F
cerca de \bar{a}_k y α_k es tal que

$$(1.14) \quad H(\bar{a}_{k+1}) < H(\bar{a}_k).$$

El caso cuando F_k es la aproximación
lineal alrededor de \bar{a}_k es conocido con el nom
bre del "Método de Gauss", uno de los —
más populares, aunque no siempre da bue
nos resultados.

Entender las posibilidades y limita-

ciones del método de Gauss es el objetivo central de este trabajo, esto se habrá tratando de explotar al máximo los elementos geométricos que intervienen.

Los "otros" métodos que actualmente se utilizan, cuando el método de Gauss no funciona, están siendo objeto de revisión en base al espíritu de este trabajo y en un futuro próximo daremos a conocer nuestros resultados.

2. Papel del Residual en la Estimación.

En esta sección haremos ver que el residual $\bar{r} = \bar{F}(\bar{a}^*) - \bar{y}$ (donde \bar{a}^* es el mínimo) va a desempeñar un papel muy importante en la solución numérica ya que como veremos el error relativo en la solución dependerá en forma bastante significativa de la magnitud del residual.

En el caso lineal a través del caso particular más simple, se motiva el resultado — general que no se probará aquí [1,3], para no distraer la atención en los detalles técnicos, y en el caso no-lineal solo se analizará el caso más simple que nos permitirá ver que elementos geométricos utilizaremos para llevar a cabo nuestro objetivo central.

2.1 Papel del Residual en la Estimación Lineal.

El caso más sencillo de la estimación lineal es el que tiene como interpretación — geométrica determinar el punto sobre una recta en el plano más cercano a un punto. El proposito en esta sección es ver los elementos — geométricos que intervienen en la sensibilidad de la solución cuando la recta y el punto son perturbados (vease la figura 1.0). Así pues, buscamos una cota al error relativo de la solución.

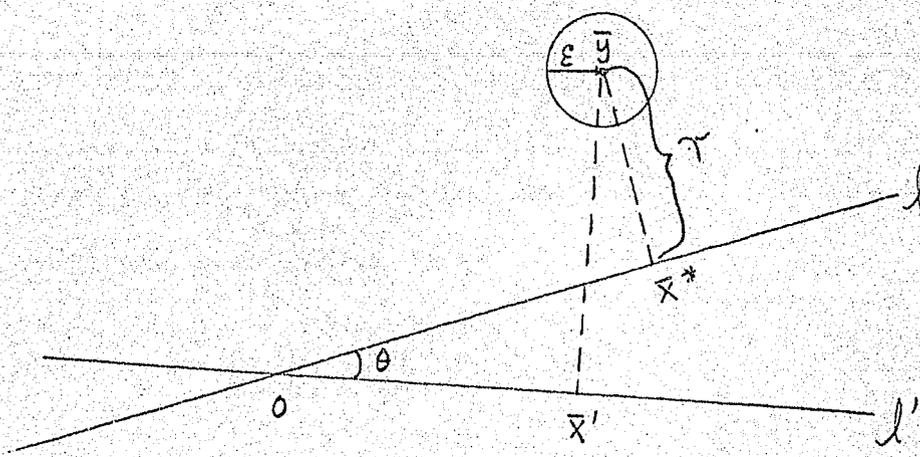


figura 1.D

Con respecto a la figura anterior, es de esperar que dicha cota dependa de $\theta, \epsilon, \tau, \bar{x}^*, \dots$. En símbolos, por hallar $E(\theta, \epsilon, \tau, \bar{x}^*, \dots)$ tal que

$$(1.15) \quad \frac{\|\bar{x}' - \bar{x}^*\|}{\|\bar{x}^*\|} \leq E(\theta, \epsilon, \tau, \bar{x}^*, \dots).$$

Observando que el triángulo $\bar{x}'\bar{c}\bar{x}^*$ de la figura 1.E es isósceles se tiene la relación

$$(1.16) \quad \text{sen } \theta = \frac{\delta}{\|\bar{y}\|},$$

por otra parte, del triángulo $\bar{x}^* \bar{y} \bar{o}$ de la misma figura se tiene que

$$(1.17) \quad \|\bar{y}\| = (\|\bar{x}^*\|^2 + r^2)^{\frac{1}{2}}.$$

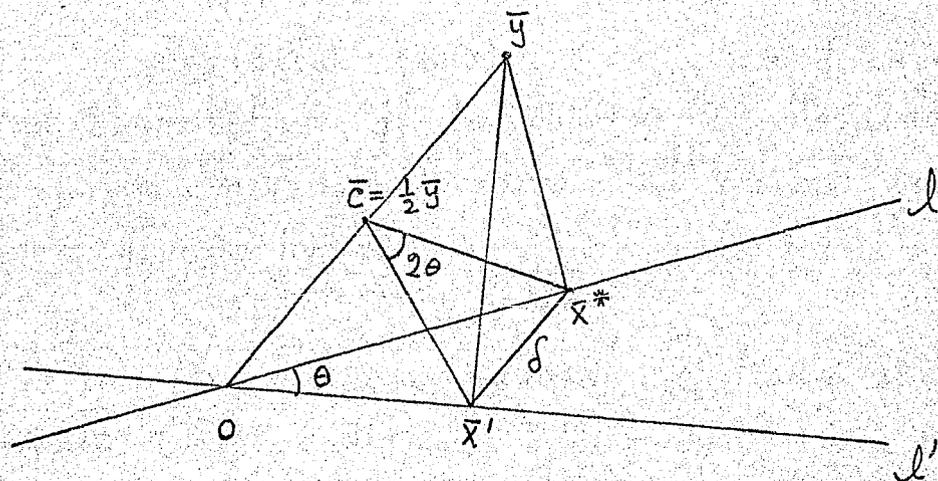


figura 1.E

Finalmente considerando la perturbación del punto \bar{y} (véase figura 1.D) y las relaciones (1.16), (1.17) se obtiene que

$$(1.18) \quad E(\theta, \varepsilon, \gamma, \bar{x}^*) = \operatorname{sen} \theta \left(1 + \frac{\gamma^2}{\|\bar{x}^*\|^2}\right)^{\frac{1}{2}} + \frac{\varepsilon}{\|\bar{x}^*\|}$$

satisface la desigualdad en (1.15). Para θ — "suficientemente pequeña" (1.18) toma la forma

$$(1.18') \quad E(\theta, \varepsilon, \gamma, \bar{x}^*) = \theta \left(1 + \frac{\gamma^2}{\|\bar{x}^*\|^2}\right)^{\frac{1}{2}} + \frac{\varepsilon}{\|\bar{x}^*\|}.$$

Las cotas (1.18) y (1.18') nos dicen que el error relativo de la solución \bar{x}^* depende — en forma muy significativa del residual (vea se la figura 1.F).

Para el caso general se tiene el siguiente ([1, 3, 6, 11])

Teorema: Sea $\bar{r} = A\bar{a}^* - \bar{y}$, supongase que el $\operatorname{rango}(A) = n$ y que $\|A^+\| \|E\| < 1$ entonces si $\bar{x}^* = A\bar{a}^* \neq \bar{0}$

$$\frac{\|\bar{x}' - \bar{x}^*\|}{\|\bar{x}^*\|} \leq K(A, \alpha) \left[\left(1 + K(A, \alpha) \frac{\|\bar{r}\|}{\|A\| \|\bar{x}^*\|}\right) \alpha + \gamma \beta \right].$$

El teorema anterior confirma que el residual participa en forma muy significativa en la sensibilidad de la solución.

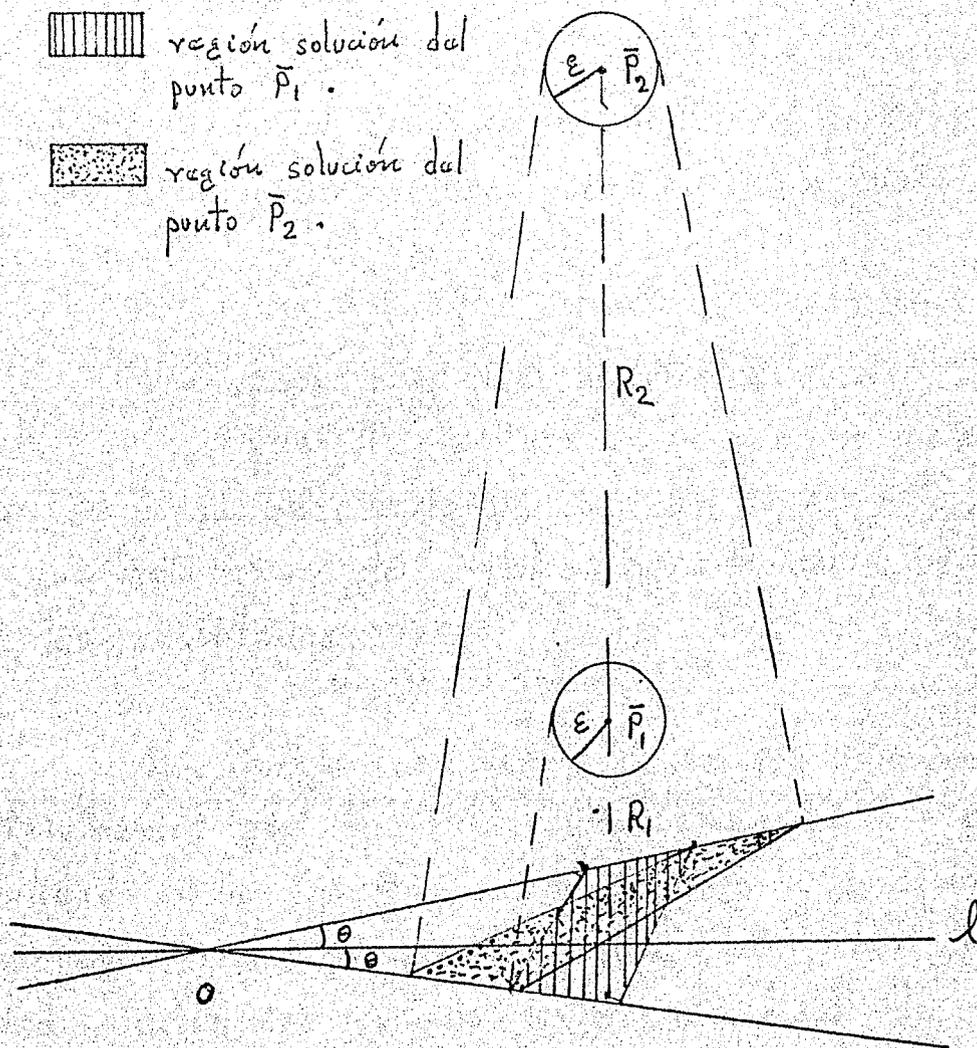


Figura 1.F. La figura ilustra la sensibilidad de la solución debida al residual.

2.2 Papel del Residual en la Estimación No - Lineal.

El problema de estimación no-lineal de un parámetro, en el que es muy fácil calcular la solución, corresponde al problema de hallar el punto sobre una circunferencia más cercano a un punto dado en el plano de la circunferencia.

Una razón más para discutir este caso particular es que el problema de la sección anterior se puede ver como un caso límite de este. Lo que trataremos de explotar para la unificación de los resultados.

En forma análoga a la sección anterior obtendremos que elementos intervienen en la sensibilidad de la solución cuando el punto y el centro son perturbados (vease la figura 1.6). Por claridad, dividimos en dos partes el cálculo de una cota para el error relativo de la solución.

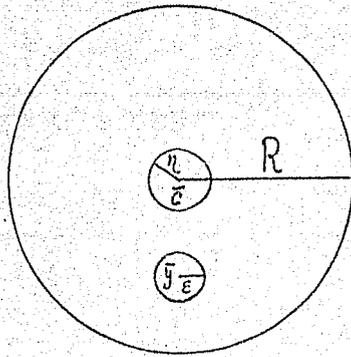


figura 1. G

La primera parte consiste en calcular la parte de la cota debida a la perturbación del centro de la circunferencia. (vease figura 1. H).

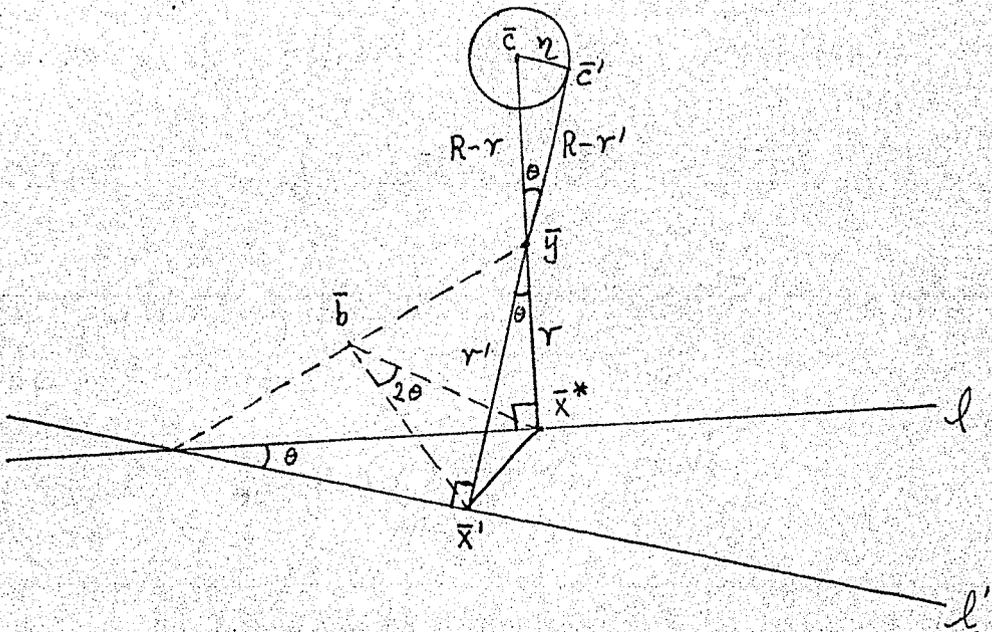


figura 1. H

Del triángulo isósceles $\bar{b}\bar{x}'\bar{x}^*$ de la figura 1.H tenemos la relación

$$(1.19) \quad \text{sen } \theta = \frac{\delta}{(\|\bar{x}^*\|^2 + r^2)^{\frac{1}{2}}}$$

también del triángulo formado por $\bar{c}\bar{y}\bar{c}'$ de la misma figura se tiene la relación

$$(1.20) \quad \text{sen } \theta = \frac{\eta}{R-r}$$

de estas dos últimas relaciones se tiene

$$(1.21) \quad \delta = \text{sen } \theta (\|\bar{x}^*\|^2 + r^2)^{\frac{1}{2}}$$

En particular, cuando $0 \leq \eta \ll R-r$ se tiene que

$$(1.22) \quad \delta \approx \theta (\|\bar{x}^*\|^2 + r^2)^{\frac{1}{2}}$$

La segunda parte de la discusión—
consiste en calcular la parte de la cota—

debida a la perturbación del punto.

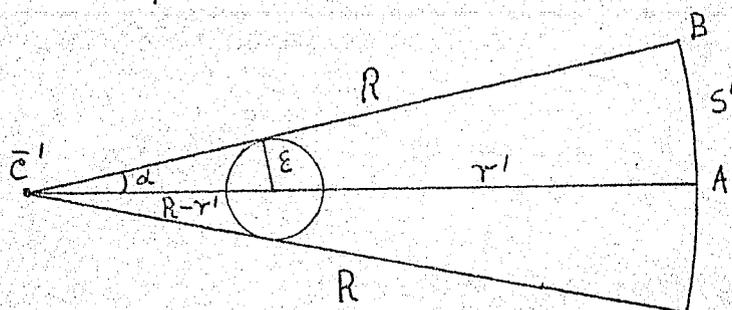


figura 1.I

Aproximando la cuerda AB por el ar
co S' (ver figura 1.I), una relación para S'
es

$$(1.23) \quad S' = \alpha R = \left(\text{arc sen } \frac{\epsilon}{R-r'} \right) R.$$

Cuando $0 \leq \epsilon \ll R-r'$ se tiene

$$(1.24) \quad S' \approx \left(\frac{\epsilon}{R-r'} \right) R = \frac{\epsilon}{1 - \frac{r'}{R}}.$$

Así de las relaciones (1.21) y (1.23)

se tiene que la cota E para el error relativo de \bar{x}^* es

$$(1.25) \quad E(\theta, \varepsilon, \bar{x}^*, r, R) = \operatorname{sen} \theta \left(1 + \frac{r^2}{\|\bar{x}^*\|^2}\right)^{\frac{1}{2}} + \frac{R}{\|\bar{x}^*\|} \operatorname{arc} \operatorname{sen} \frac{\varepsilon}{R-r},$$

cuando $r \ll R-r$ y $\varepsilon \ll R-r$ entonces la cota E es

$$(1.26) \quad E(\theta, \varepsilon, \bar{x}^*, r, R) = \theta \left(1 + \frac{r^2}{\|\bar{x}^*\|^2}\right)^{\frac{1}{2}} + \frac{\varepsilon}{\|\bar{x}^*\| \left(1 - \frac{r}{R}\right)}.$$

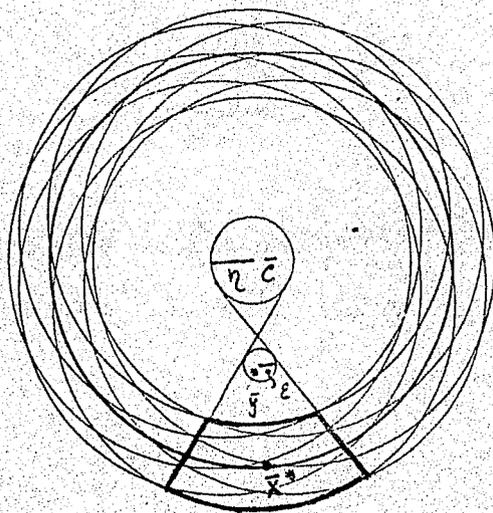
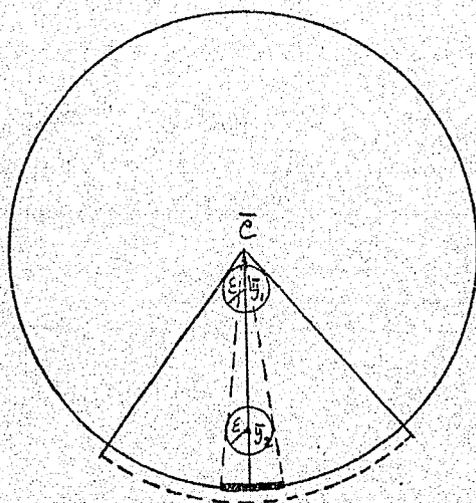


figura 1.J

Vemos que en la cota obtenida sigue participando en forma significativa el residual como en el caso lineal.

En la figura 1.J se da una ilustración geométrica de la región de incertidumbre para \bar{x}^* .



región solución para \bar{y}_1
 región solución para \bar{y}_2

figura 1.K

La cota (1.26) nos muestra que en ~~segundo~~ sumando, es importante que el

punto no este cerca del centro de la circunferencia, ya que, en este caso este sumando participa en forma bastante significativa — [12] (ver la figura 1.K).

Observación: Si tomamos el límite cuando R , el radio de la circunferencia, tiende a infinito en (1.26) obtenemos la cota que habíamos obtenido para el caso lineal.

Capítulo II

Estimación No-Lineal de un Parámetro.

Con el objeto de adquirir experiencia en la interpretación geométrica, de las condiciones suficientes para la existencia de un mínimo y la convergencia del método iterativo de Gauss, el presente capítulo lo dedicamos exclusivamente al problema de la estimación no-lineal de un solo parámetro, lo que corresponde a determinar el punto de una curva más cercano a un punto dado.

Primero establecamos condiciones de suficiencia para un mínimo, a continuación describimos el método iterativo de Gauss y finalmente establecamos su convergencia.

1. Condiciones Suficientes para un Mínimo.

Hemos visto en el capítulo anterior co-

no surge el problema de estimación no-lineal, el cual consiste en hallar el mínimo de una suma de cuadrados.

En lenguaje analítico la estimación de un solo parámetro da lugar a la función:

$$(2.1) \quad H(t) = \sum_{i=1}^m (y_i - \varphi(t, x_i))^2 \\ = \|\bar{r}(t)\|^2 = \|\bar{y} - \bar{x}(t)\|^2$$

$\bar{x}(t)$ es una curva en \mathbb{R}^m , y $\bar{r}(t)$ es el residual en t .

Si la función φ del problema es diferenciable, con respecto al parámetro, entonces las condiciones suficientes para que la función $H(t)$ tenga un mínimo local en t^* son:

$$(2.2) \quad \begin{cases} \dot{H}(t^*) = -2 \bar{r}^t \dot{\bar{x}} = 0 \\ \ddot{H}(t^*) = 2 (\dot{\bar{x}}^t \dot{\bar{x}} - \bar{r}^t \ddot{\bar{x}}) > 0 \end{cases} .$$

La primera condición en (2.2) nos dice que en el punto $\bar{x}(t^*)$ de la curva $\bar{x}(t)$, el residual $\bar{r}(t^*)$ y el vector tangente $\dot{\bar{x}}(t^*)$ son ortogonales.

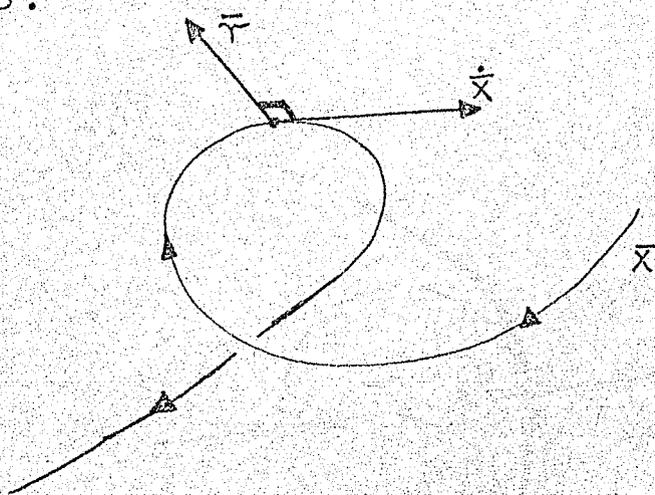


figura 2.A

Para darle una interpretación a la condición $\ddot{H}(t^*) > 0$, en la siguiente sección consideraremos casos particulares empezando con curvas en el plano.

1.1 Curvas en el Plano.

En estas condiciones suficientes las

* Estamos interesados en el caso de que solo hay un mínimo local.

podemos interpretar facilmente utilizando la circunferencia osculadora en el punto $\bar{x}(t^*)$, ya que el residual y la normal a la curva en $\bar{x}(t^*)$ son paralelos (vease la figura 2.B).

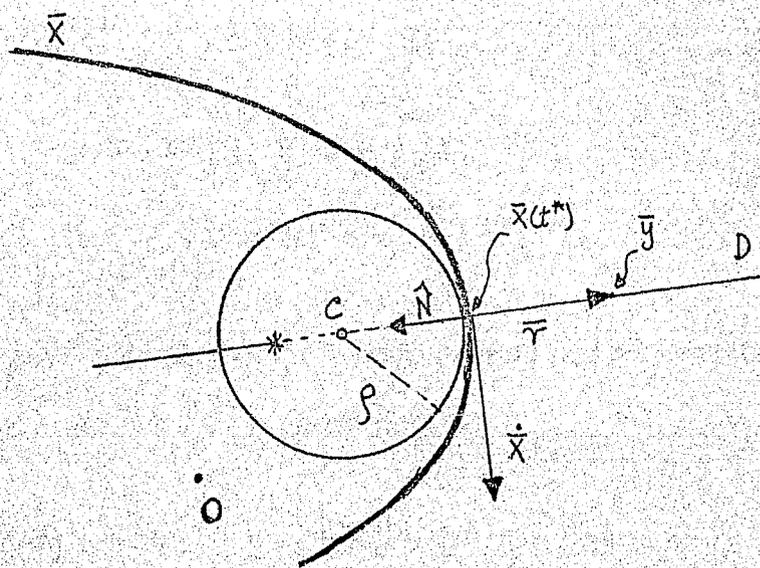


figura 2.B

¿ Para qué posiciones de \bar{y} el punto más cercano a la curva y a la circunferencia osculadora en $\bar{x}(t^*)$ coinciden? . De la figura 2.B es fácil concluir que \bar{y} debe estar en la semirrecta \overline{CD} .

Analíticamente \bar{y} se escribe como

$$\bar{X}(t^*) + (\rho - s)\hat{N}$$

con $s > 0$. Para escribir esta expresión en términos del residual, veamos que

$$\begin{aligned}\bar{r}(t^*) &= \bar{y} - \bar{X}(t^*) \\ &= (\rho - s)\hat{N}\end{aligned}$$

con $s > 0$, por un lado; por otro lado, podemos escribir a $\bar{r}(t^*)$ como

$$\bar{r}(t^*) = r \hat{r} \quad ; \quad \|\hat{r}\| = 1 \quad ; \quad r > 0.$$

Así tenemos que \bar{y} está en la semirecta \overline{cD} si

$$(2.5) \quad \begin{cases} \rho - r > 0 & \text{si } \hat{r} = \hat{N} \\ \rho + r > 0 & \text{si } \hat{r} = -\hat{N} \end{cases}.$$

La discusión anterior nos sugiere — buscar una expresión en la que \hat{H} sea una

función de $\rho \pm r$ en t^* .

Escribiendo $\dot{\vec{x}}$ y $\ddot{\vec{x}}$ en términos de los vectores tangente y normal principales a la curva tenemos:

$$(2.6) \quad \begin{cases} \dot{\vec{x}} = \dot{s} \hat{T} \\ \ddot{\vec{x}} = \frac{\dot{s}^2}{\rho} \hat{N} - \dot{s} \ddot{s} \hat{T}, \end{cases}$$

donde $\dot{s} = \|\dot{\vec{x}}\|$, lo que permite escribir a \ddot{H} como

$$(2.7) \quad \ddot{H} = 2 \dot{s}^2 \left(1 - \frac{1}{\rho} \bar{r} \hat{N} - \dot{s} \ddot{s} \bar{r} \hat{T} \right),$$

la cual evaluada en t^* toma la forma

$$(2.8) \quad \begin{aligned} \ddot{H}(t^*) &= 2 \dot{s}^2 \left(1 - \frac{r}{\rho} \hat{r} \hat{N} \right) \\ &= 2 \dot{s}^2 \frac{1}{\rho} (\rho - r \hat{r} \hat{N}), \end{aligned}$$

donde $\bar{r} = r \hat{r}$, ($r > 0$).

\ddot{H} es positiva cuando

$$(2.9) \quad \begin{cases} \rho - \tau > 0, & \text{si } \hat{\gamma}^t \hat{N} = 1 \\ \rho + \tau > 0, & \text{si } \hat{\gamma}^t \hat{N} = -1. \end{cases}$$

Lo que confirma la validez de nuestra conjetura. Sin embargo, estas condiciones son insuficientes para el caso general, en el cual $\hat{\gamma}^t \hat{N} \neq \pm 1$.

1.2 Curvas Planas.

Una de las ventajas del caso discutido en la sección anterior es que nos permite tener una expresión para \ddot{H} que sí es general, a saber

$$(2.10) \quad H(t^*) = 2\dot{s}^2 \left(1 - \frac{\tau}{\rho} \hat{\gamma}^t \hat{N}\right);$$

y que parece, en principio, más fácil de interpretar geoméricamente.

Supongamos que $\bar{x}(t)$ es una curva plana en \mathbb{R}^m .

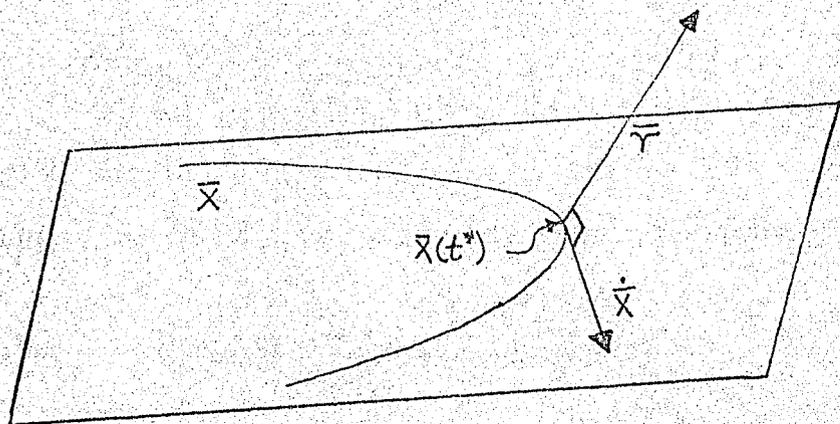


figura 2.c

Varemos que en este caso la condición de que $\ddot{H}(t^*)$ sea positiva tiene dos interpretaciones geométricas, que como se verá, serán de gran utilidad para este trabajo, y que nos permitirá en un futuro cercano construir nuevos algoritmos en el caso general ([7]).

Primera Interpretación.

La información geométrica que nos ofrece la primera condición se conserva igual que para las curvas en el plano (vease figura 2.C).

La expresión para la segunda condición se obtiene de la expresión (2.8), puesto que esta es independiente del número de observaciones

$$(2.10) \quad \ddot{H} = 2 \dot{S}^2 \left(1 - \frac{r}{\rho} \hat{r}^t \hat{N} \right) > 0 ,$$

la que tiene las siguientes dos interpretaciones.

Si vamos a $\hat{r}^t \hat{N}$ como la componente del vector residual sobre el vector normal principal. Esto nos conduce a descomponer el vector residual como:

$$\bar{r} = (\hat{r}^t \hat{N}) \hat{N} + \bar{r}_\perp$$

donde \bar{r}_\perp es ortogonal al plano Π de la curva (vease figura 2.D).

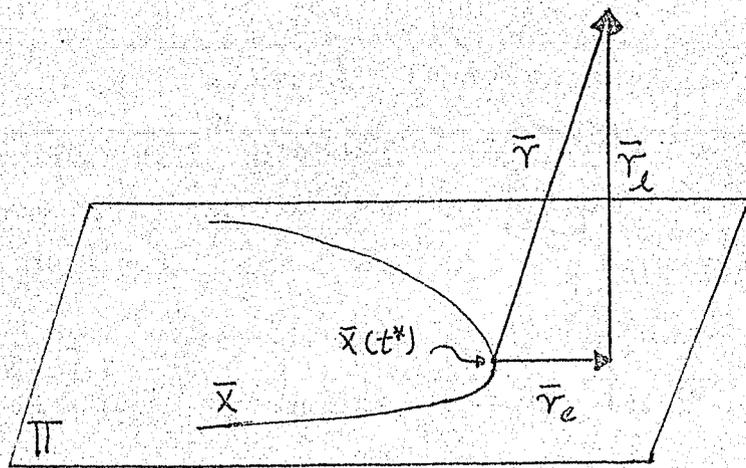


figura 2.D. En la figura \bar{r} es el vector residual, \bar{r}_l el vector residual lineal y \bar{r}_e el vector residual efectivo.

Ahora la expresión para la segunda derivada es

$$(2.11) \quad \ddot{H} = 2 \dot{s}^2 \left(1 - \frac{\tau_e}{\rho}\right) > 0 ,$$

τ_e es el residual efectivo y está dado por $\tau \bar{r}^t \bar{N}$.

Por lo tanto, \ddot{H} es positiva cuando

$$(2.12) \quad \rho - \tau_e > 0 .$$

La discusión anterior nos sugiere dividir nuestro problema en dos: uno lineal, que consiste en hallar la proyección de \bar{y} sobre el plano de la curva; y el otro no-lineal, que es calcular el punto de la curva más cercano a la proyección de \bar{y} .

Segunda Interpretación.

En esta interpretación lo que haremos será tratar de reducir el problema a una curva plana en el plano.

Esto lo lograremos de la manera siguiente: proyectamos la curva $\bar{x}(t)$ sobre el plano que contiene a los vectores $\bar{r}(t^*)$ y $\hat{T}(t^*)$ y que pasa por el punto $\bar{x}(t^*)$ (vease la figura 2.6).

Es fácil demostrar que la curvatura $K_{\bar{r}}$ de la curva proyectada en el punto $\bar{x}(t^*)$ está dada por

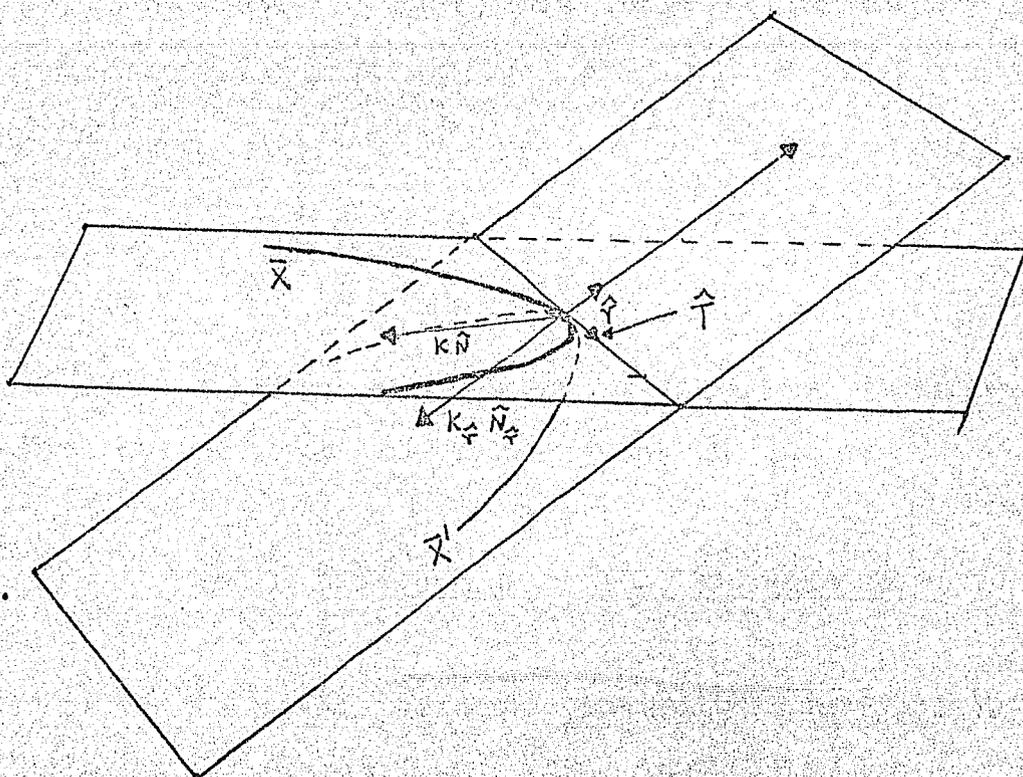


figura 2.6.

$$K_{\vec{r}} = K |\vec{r}^t \hat{N}|,$$

lo que permite escribir

$$(2.13) \quad \ddot{H}(\mathbf{t}^*) = \begin{cases} 2\dot{s}^2(1 - K_{\vec{r}} r), & \text{si } \vec{r}^t \hat{N} > 0 \\ 2\dot{s}^2(1 + K_{\vec{r}} r), & \text{si } \vec{r}^t \hat{N} < 0. \end{cases}$$

Por consiguiente $\ddot{H}(t^*)$ es positivo cuando

$$(2.14) \quad \begin{cases} 1 - K_{\tilde{r}} \tau > 0, & \text{si } \tilde{r}^t \tilde{N} > 0 \\ 1 + K_{\tilde{r}} \tau > 0, & \text{si } \tilde{r}^t \tilde{N} < 0. \end{cases}$$

Estas condiciones significan lo mismo que en el caso de curvas en el plano.

1.3 Curvas en General.

En esta sección generalizamos los resultados obtenidos en las secciones anteriores.

Las expresiones para las condiciones suficientes son las mismas sin cambio alguno, esto es,

$$(2.15) \quad \begin{cases} \dot{H} = -2\tilde{r}^t \dot{\tilde{x}} & y \\ \ddot{H} = 2\frac{\dot{s}^2}{s}(\rho - r_e) \\ \ddot{H} = 2\dot{s}^2(1 \pm K_{\tilde{r}}\tau). \end{cases}$$

Las interpretaciones geométricas son válidas si se considera el plano osculador en el punto $\bar{x}(t^*)$ (véase la figura 2.H).

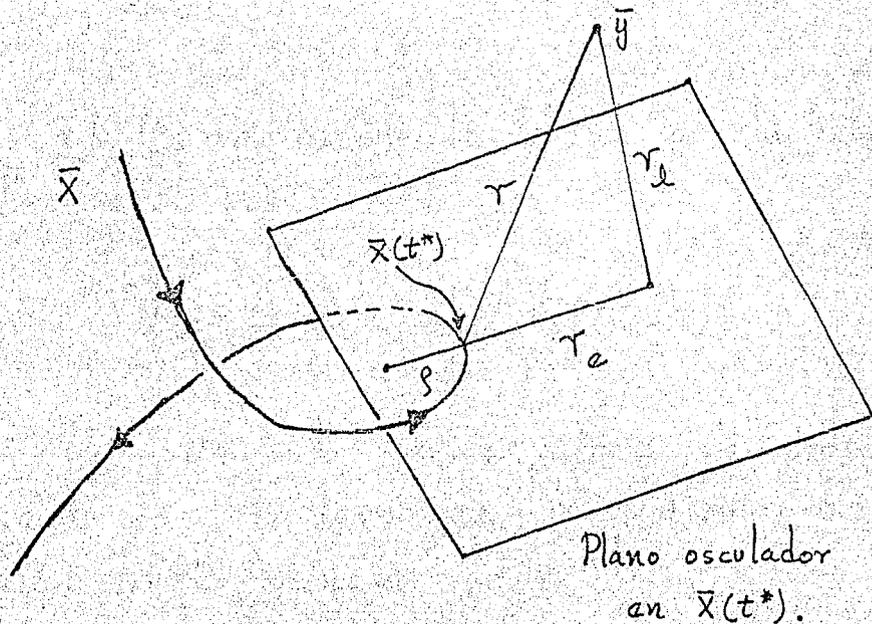


figura 2.H.

O bien el plano definido por los vectores residual unitario y tangente unitario en el punto (véase la figura 2.G).

2. Método de Gauss.

El método de Gauss obtiene una corrección a una aproximación t_0 dada, de la siguiente forma:

i) Se aproxima la curva por la recta tangente en el punto $\bar{x}(t_0)$

$$\bar{x}(t) \simeq \bar{x}(t_0) + \dot{\bar{x}}(t_0)(t - t_0).$$

ii) De esta forma ahora se procede como si se tratara de un problema de estimación lineal minimizando

$$\| \bar{y} - \bar{x}(t_0) - \dot{\bar{x}}(t_0) \alpha \|^2.$$

La α^* que minimiza esta dada por

$$\alpha^* = - (\dot{\bar{x}}^t \dot{\bar{x}})^{-1} (\bar{y} - \bar{x}(t_0))^t \dot{\bar{x}}$$

que es la solución de la ecuación normal para este caso, como pueda verse fácilmente.

iii) Así se obtiene la corrección

$$t_1 = t_0 + \alpha^*$$

$$= t_0 - (\dot{\bar{X}}(t_0)^t \dot{\bar{X}}(t_0))^{-1} (\bar{y} - \bar{X}(t_0))^t \dot{\bar{X}}(t_0).$$

(vease la figura 2.I).

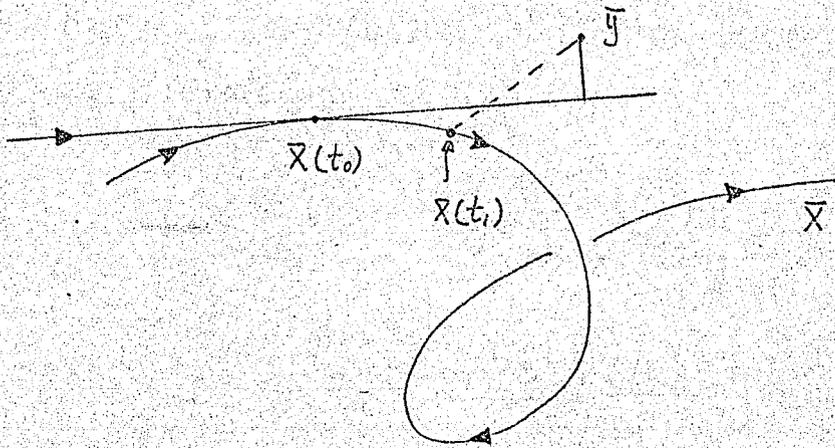


figura 2.I

Estamos interesados en detectar los elamentos geométricos que intervienen en la convergencia del método de Gauss, ya que hasta ahora las demostraciones de convergen

cia son indirectas, via algunos otros teoremas más generales y la mayoría suponen para la convergencia del método que el residual debe ser pequeño.

2.1 Convergencia.

Para estudiar la convergencia del método de Gauss, escribimos de otra forma la regla de iteración

$$(2.16) \quad t_{n+1} = t_n - \frac{(\bar{y} - \bar{x}(t_n))^t \dot{\bar{x}}(t_n)}{\dot{\bar{x}}^t(t_n) \dot{\bar{x}}(t_n)}$$

$$= t_n - \frac{\bar{r}^t(t_n) \hat{T}(t_n)}{\|\dot{\bar{x}}(t_n)\|},$$

donde $\bar{r}(t_n) = \bar{y} - \bar{x}(t_n)$ y $\hat{T}(t_n)$ es el vector tangente unitario a la curva en el punto $\bar{x}(t_n)$.

Desde el punto de vista de la convergencia lo que necesitamos es encontrar —

condiciones bajo las cuales la función

$$(2.17) \quad g(t) = t - \frac{\bar{r}^t(t) \hat{J}(t)}{\|\dot{\bar{x}}(t)\|},$$

tiene un punto fijo t^* , es decir,

$$(2.18) \quad g(t^*) = t^*.$$

Como las condiciones que buscamos son de carácter local, entonces como es conocido bastará encontrar bajo que condiciones $|g'(t^*)| < 1$ [8], para que el método de Buss converja localmente, sin embargo procedemos explícitamente a estimar

$$(2.19) \quad t_{n+1} - t^* = f(t_n - t^*),$$

para entender geométricamente las condiciones para la convergencia.

A partir de la función recursiva (2.16), desarrollando esta en series de Tay-

lor alrededor del punto solución se tiene

$$(2.20) \quad t_{n+1} = t^* - \frac{1}{f} \bar{r}^{\hat{t}_n} \hat{N} (t_n - t^*) + o(t_n - t^*)^2$$

asi

$$(2.21) \quad t_{n+1} - t^* = \frac{r}{f} (\bar{r}^{\hat{t}_n} \hat{N}) (t_n - t^*) + o(t_n - t^*)^2,$$

la que nos indica que la sucesión $\{t_n\}$ converge a t^* siempre y cuando

$$(2.22) \quad |r(\bar{r}^{\hat{t}_n} \hat{N})| = |r| = r < f,$$

ademas de revelar que la convergencia es monótona si el punto está situado en la parte convexa de la curva y es oscilante si está situado en la parte cóncava de la curva.

La interpretación geométrica, la hacemos por casos:

a) Si la curva está en el plano, la interpretación es ilustrada en la figura 2.J.

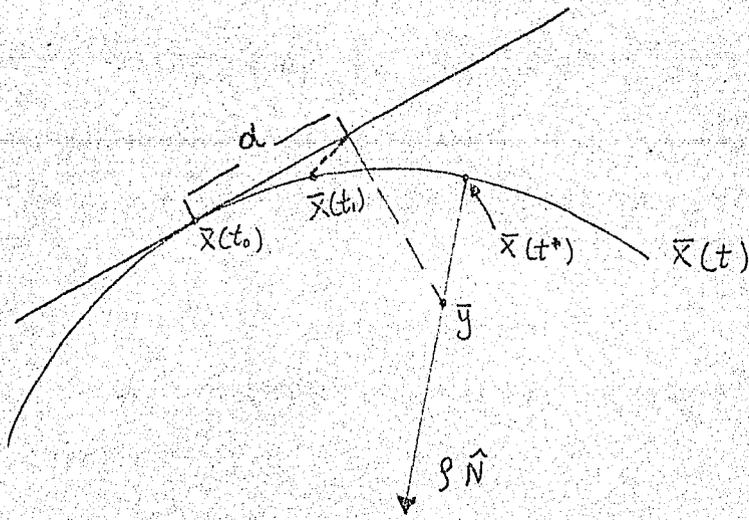


figura 2. J (a) Convergencia mono
tona.

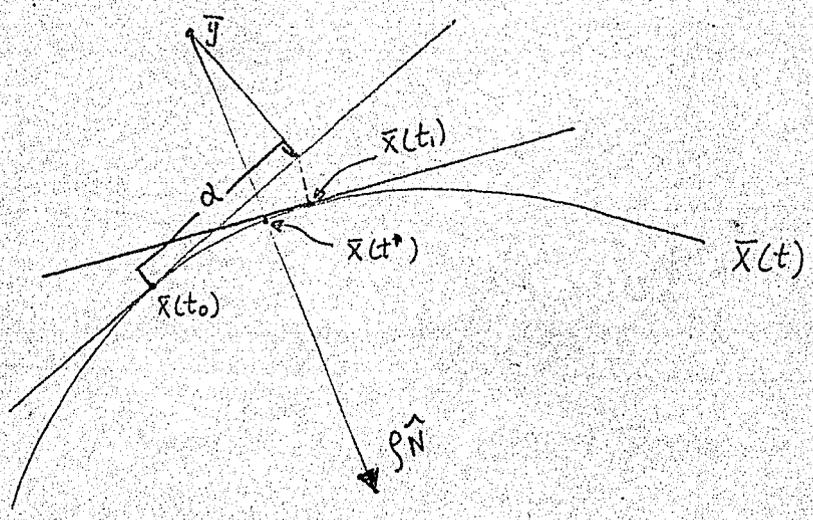


figura 2. J (b) Convergencia os
cilante.

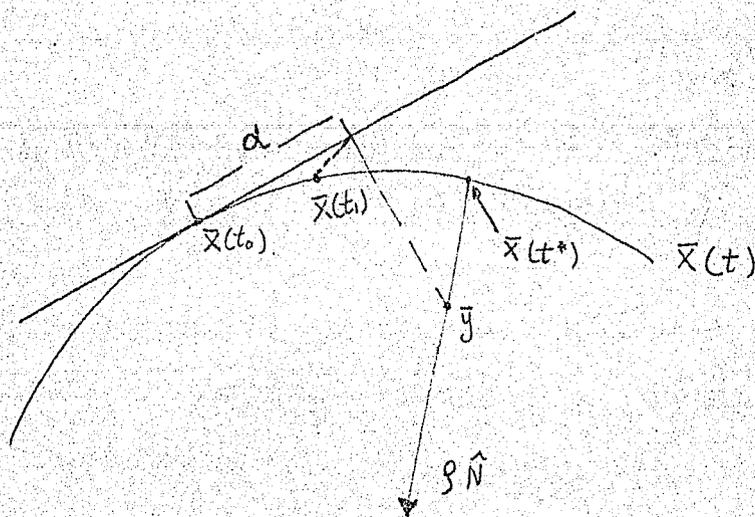


figura 2.J (a) Convergencia monotonica.

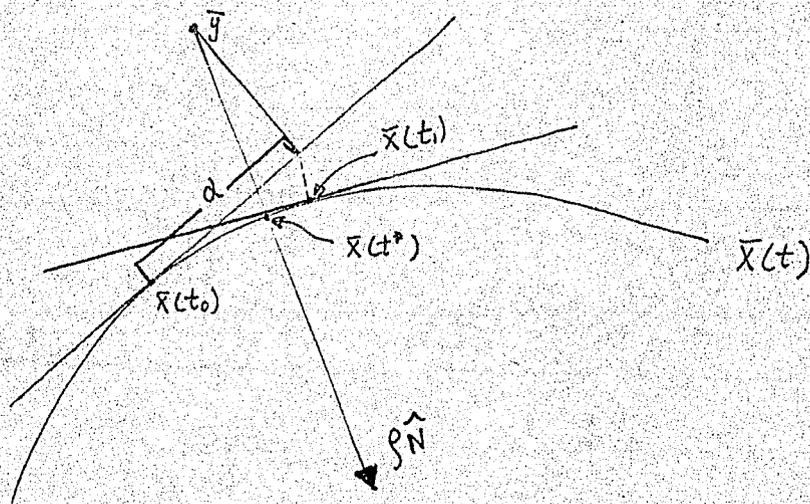


figura 2.J (b) Convergencia oscilante.

b) Cuando la curva es plana en el espacio, entonces la igualdad en (2.21) nos señala que $\{\tau_n\}$ converge siempre que

¿i) Si el vector residual lo descomponemos en el residual lineal y el residual efectivo entonces

$$(2.23) \quad |r_e(\hat{y}^t \hat{N})| = |r_e| = r_e \leq \rho.$$

Lo que se ilustra en la figura 2.K

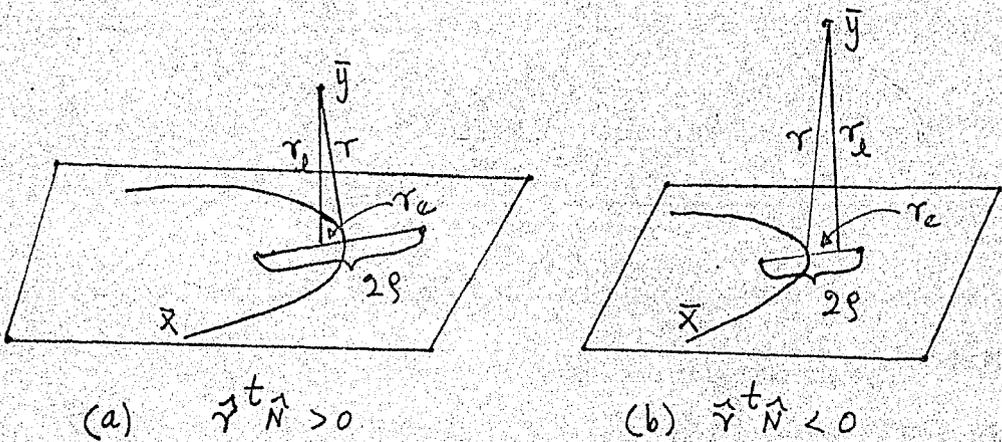
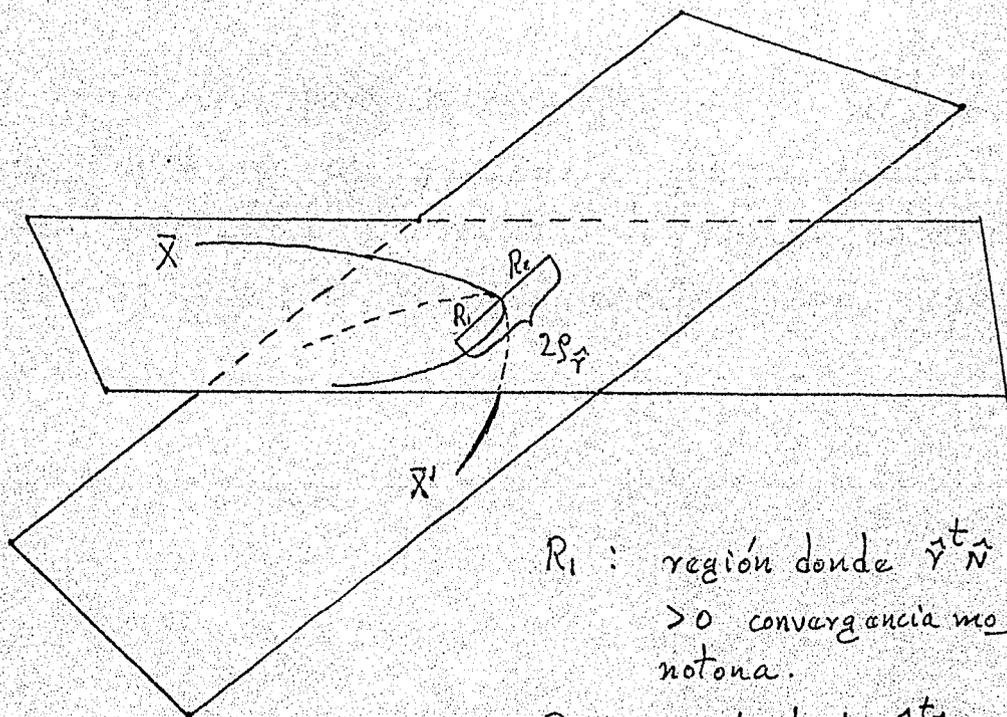


figura 2.K

ii) Si trabajamos la proyección de la curva en el plano definido por los vectores tangente unitario y residual unitario se tiene

$$(2.24) \quad |\gamma(\hat{\gamma}^t \hat{N})| = |\gamma| = \gamma < \rho_{\hat{\gamma}}.$$

Lo que se ilustra en la figura 2.L.



R_1 : región donde $\hat{\gamma}^t \hat{N} > 0$ convergencia monótona.

R_2 : región donde $\hat{\gamma}^t \hat{N} < 0$ convergencia oscilante.

figura 2.L

e) Cuando la curva es arbitraria y no cae en los incisos anteriores, los resultados en inciso b son válidos al considerar el plano osculador ó el plano definido por los vectores residual unitario y tangente unitario.

En este capítulo hemos obtenido una información importante, que nos servirá para la parte posterior de nuestro trabajo, esta es:

Que para la convergencia del método de Gauss no es necesario que el residual sea pequeño, en realidad lo que es necesario es que el residual afectivo sea pequeño.

Capítulo III

Estimación No-Linear de Varios Parámetros.

En este capítulo estudiamos la estimación no-lineal de parámetros en su caso general, es decir, el ajuste de una función que depende de varios parámetros.

Procediendo en la misma dirección del capítulo anterior, primero interpretamos las condiciones suficientes para un mínimo y después los requisitos que se deben tener para la convergencia del método de Gauss, aprovechando al máximo la experiencia obtenida en el capítulo anterior.

1. Condiciones Suficientes.

En el capítulo uno planteamos el problema de estimación no-lineal de parámetros en su forma general ahí encontramos que había que minimizar una suma de cuadrados,

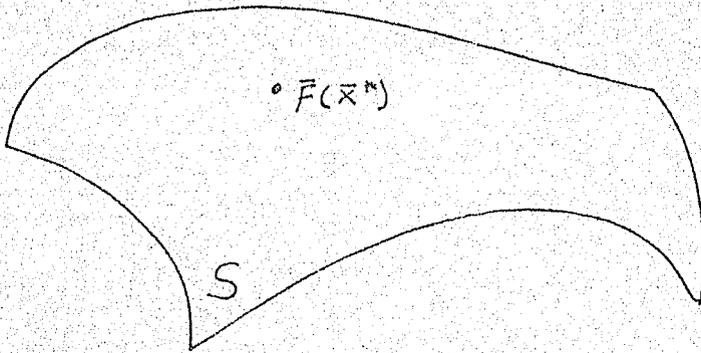
$$(3.1) \quad H(\bar{x}) = \sum_{i=1}^m (f_i(\bar{x}) - y_i)^2 \\ = \| \bar{F}(\bar{x}) - \bar{y} \|^2 ; \quad \bar{x} \text{ en } \mathbb{R}^n ,$$

\bar{x}^* es solución del problema si

$$H(\bar{x}^*) = \text{Min}_{\bar{x} \text{ en } \mathbb{R}^n} \| \bar{F}(\bar{x}) - \bar{y} \|^2 .$$

Como ya se vio en el capítulo uno este problema en términos geométricos consiste en determinar un punto $\bar{F}(\bar{x}^*)$ sobre la "superficie" generada por \bar{F} , que está más cerca de un punto \bar{y} en \mathbb{R}^m (vease la figura 3.A).

\bar{y}



\bar{y} punto en el espacio de dimensión m .

S "superficie" de dimensión n .

$\bar{F}(\bar{x}^*)$ punto en la "superficie" más cercano a \bar{y} .

figura 3.A

Para derivar condiciones suficientes para un mínimo, necesitamos que \bar{F} sea suficientemente diferenciable y que \bar{F}' sea de rango máximo, es decir n .

Las expresiones para la primera y -

segunda derivadas de H son:

$$(3.2) \quad H'(\bar{x}) = \bar{F}'(\bar{x})^t (\bar{F}(\bar{x}) - \bar{y}),$$

$$(3.3) \quad H''(\bar{x}) = 2 \left(\bar{F}'(\bar{x})^t \bar{F}'(\bar{x}) + \sum_{i=1}^m (f_i - y_i) f_i''(\bar{x}) \right),$$

donde

$$\bar{F}(\bar{x}) = \begin{bmatrix} f_1(\bar{x}) \\ f_2(\bar{x}) \\ \vdots \\ f_m(\bar{x}) \end{bmatrix} \quad y$$

$f_i''(\bar{x})$ es el Hessiano de f_i y está dado por

$$f_i''(\bar{x}) = \left[\frac{\partial^2 f_i}{\partial x_j \partial x_k}(\bar{x}) \right] \quad (j, k = 1, \dots, n).$$

La primera condición $H'(\bar{x}^*) = \bar{0}$ en el punto \bar{x}^* nos dice que, el vector residual $\bar{r}(\bar{x}^*) = \bar{y} - \bar{F}(\bar{x}^*)$ es ortogonal a cada columna de \bar{F}' . Es decir, $\bar{r}(\bar{x}^*)$ es ortogonal al subespacio generado por las colum-

nas de $F'(\bar{x}^*)$, este subespacio es conocido como el "espacio tangente" a la superficie - en el punto $\bar{F}(\bar{x}^*)$ (vease la figura 3.B).

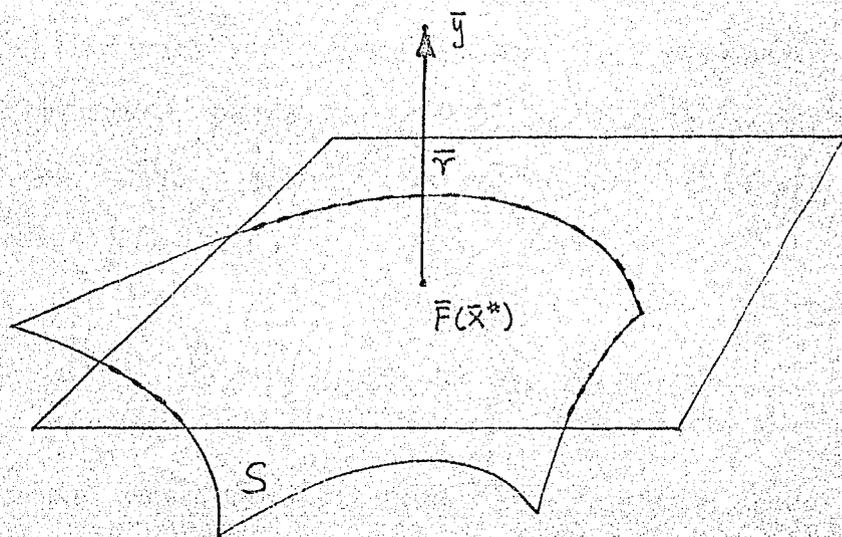


figura 3.B

No se ve como interpretar la segunda condición $H''(\bar{x}^*)$ (sea definida positiva) a primera vista. Sin embargo, como en el capítulo anterior observamos que la segunda condición se expresaba en términos de la curvatura, es natural tratar de buscar una in

interpretación análoga, pero el camino no es corto, ya que: ¿Cuál es la curvatura de una superficie en un punto dado?

Procedamos a interpretar la segunda condición, primero para el caso de superficies en \mathbb{R}^3 , con estos elementos se podrá interpretar el caso general.

1.1 Superficies en el Espacio.

Como consideramos que la segunda condición, debe poder interpretarse en términos de la curvatura, es importante ver cuál de los conceptos que se utilizan en la literatura de Geometría Diferencial nos es más útil [9].

Primero veamos que por un punto dado de una superficie, siempre se pueden construir curvas con curvatura arbitrariamente grande. Para ello veamos los siguientes casos particulares pero genéricos.

a) Considérese una esfera, cortémos la por planos de tal manera que cada vez obtengamos casquetes más chicos.

Es claro que las circunferencias que se obtienen con los cortes tienen radios cada vez más pequeños, es decir con curvatura cada vez más grande (véase la figura 3.c).

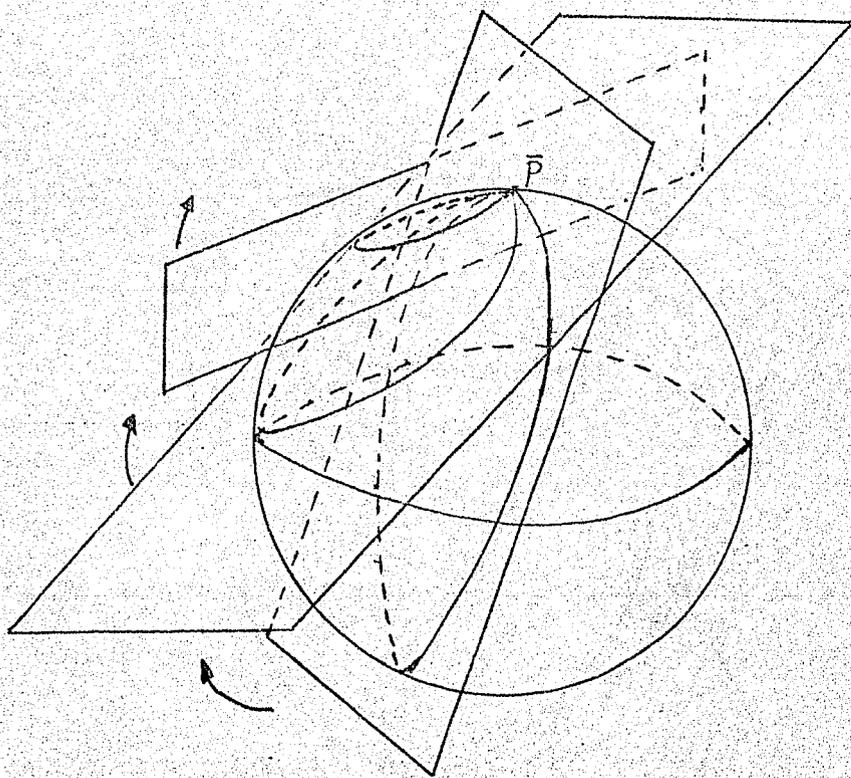


figura 3.c

b) Considérase un paraboloide elíptico, cortemoslo por planos de tal manera que cada vez - obtengamos casquetes más chicos (vease la figura 3.D). Las elipses que se obtienen, tienen cada vez su semi eje mayor (a) más chico y como el radio de curvatura para una elipse [10] es proporcional al producto ae^2 , donde e es la excentricidad y $0 < e < 1$, entonces en particular la curvatura se hace cada vez más grande.

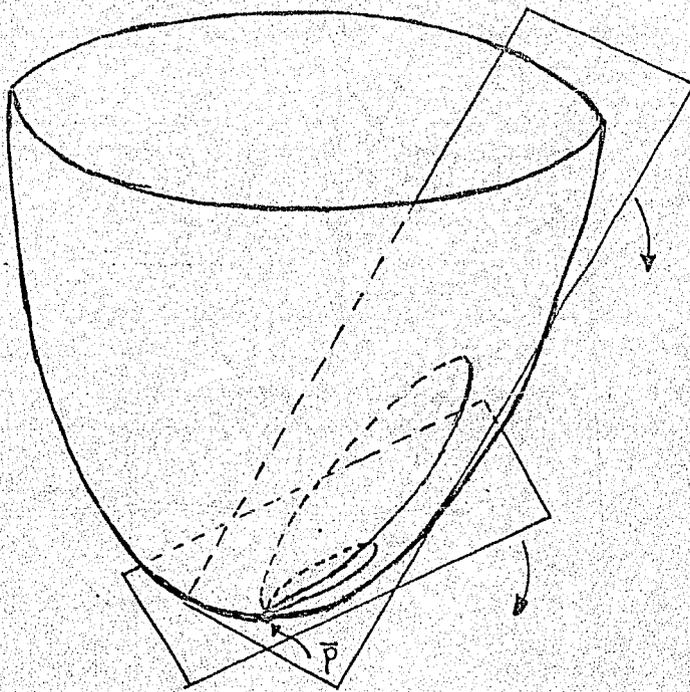


figura 3.D

c) Considerárese un paraboloide hiperbólico, procediendo en forma análoga a los ejemplos anteriores. A las hipérbolas que se obtienen en este caso se les puede aplicar el argumento del ejemplo anterior para concluir que la curvatura se hace cada vez más grande (vease la figura - 3.E).

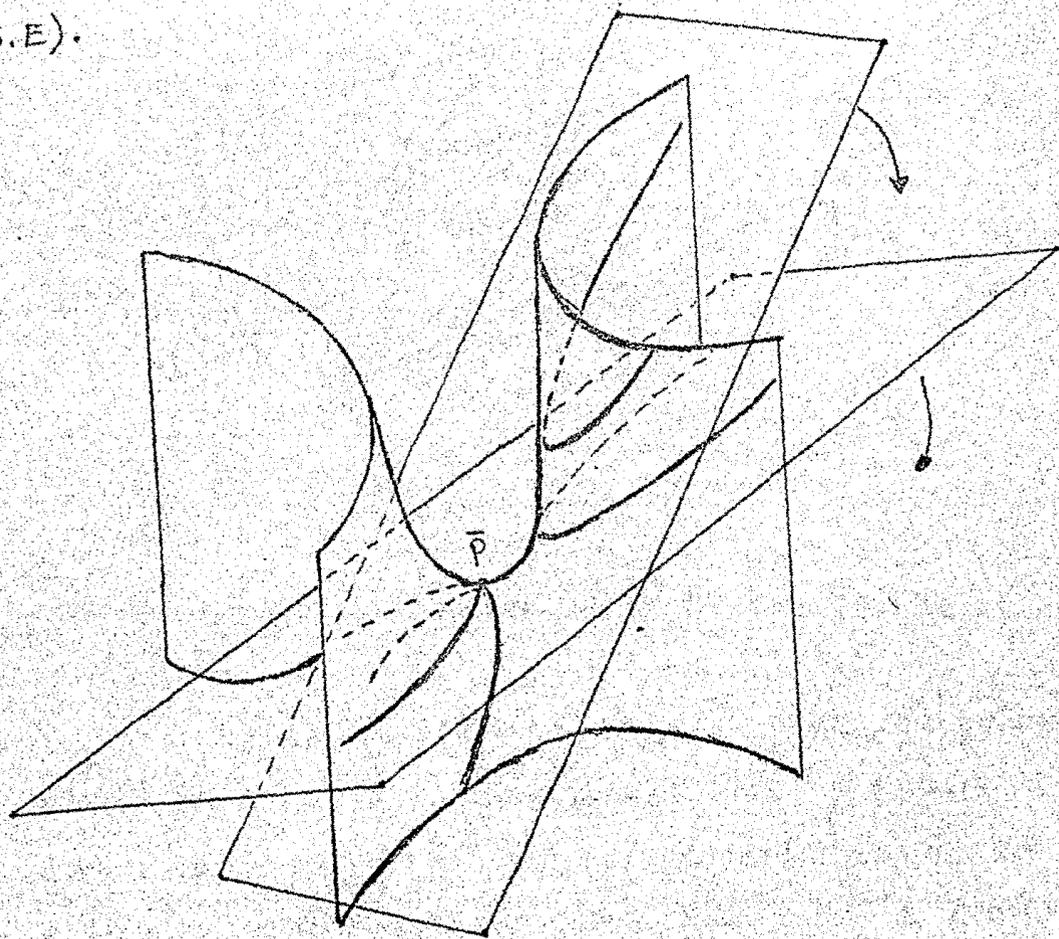


figura 3.E

d) Considérese por último, un paraboloide cilíndrico, las curvas que se obtienen estas son parábolas con curvaturas cada vez más grandes (vease la figura 3.F).

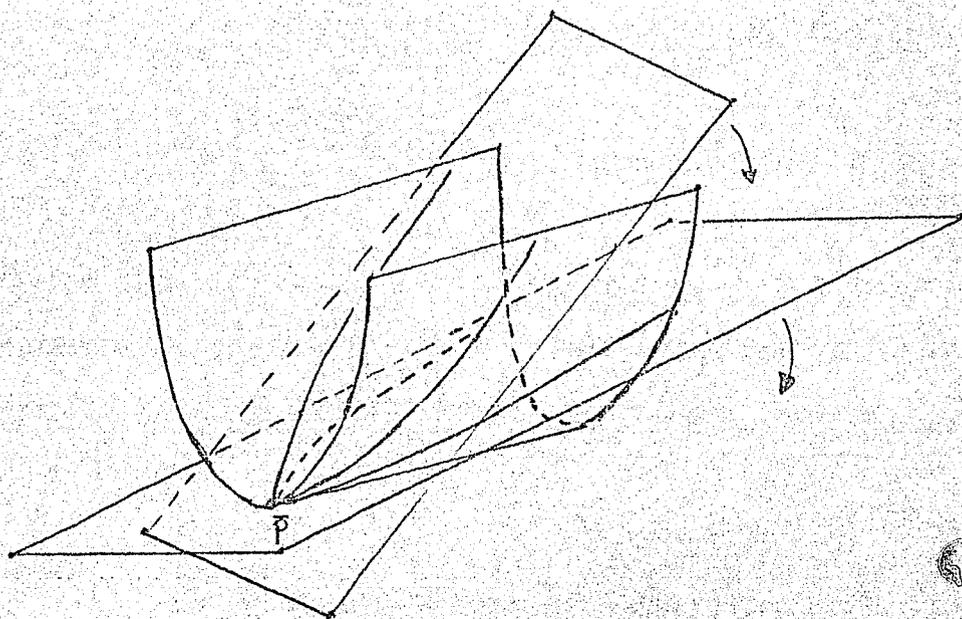


figura 3.F

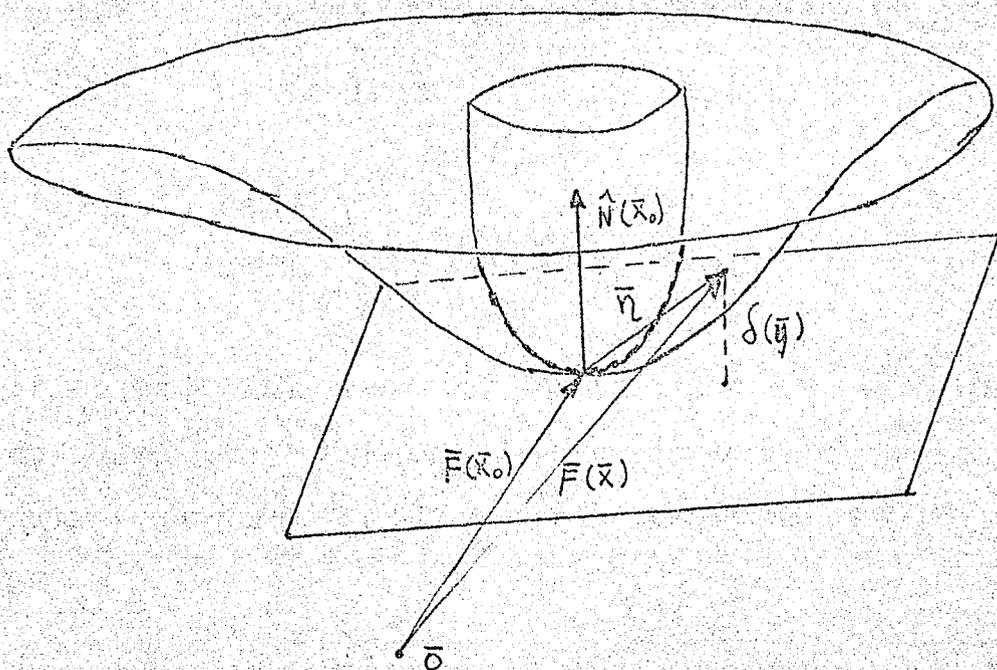
Esta técnica se aplica para cualquier superficie, ya que localmente las podemos aproximar por una de las superficies de los ejemplos anteriores, como se verá a continuación.

Si F es suficientemente diferenciable entonces alrededor de \bar{x}_0 , tenemos

$$\begin{aligned} \bar{F}(\bar{x}) = & \bar{F}(\bar{x}_0) + \bar{F}'(\bar{x}_0) \bar{y} + \frac{1}{2} \left(\bar{F}_{x_1 x_1}(\bar{x}_0) y_1^2 + \right. \\ & \left. + 2 \bar{F}_{x_1 x_2}(\bar{x}_0) y_1 y_2 + \bar{F}_{x_2 x_2}(\bar{x}_0) y_2^2 \right) + \dots, \end{aligned}$$

donde $\bar{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \bar{x} - \bar{x}_0$. El carácter de la aproximación de segundo grado se puede ver claramente tomando el plano tangente como referencia y midiendo la desviación de la superficie con respecto a este.

La desviación de la superficie con respecto a su plano tangente en el punto $\bar{F}(\bar{x}_0)$, está dada por la componente del vector $\bar{F}(\bar{x}) - \bar{F}(\bar{x}_0)$ en dirección de su vector normal $\hat{N}(\bar{x}_0)$ (vease la figura 3.6).



$$\tilde{\eta} = \bar{F}(x) - \bar{F}(x_0).$$

figura 3.6

Analíticamente la desviación $\delta(\bar{y})$ es
ta dada por

$$\delta(\bar{y}) = (\bar{F}(x) - \bar{F}(x_0))^{\perp} \hat{N}(x_0).$$

Con esta expresión la función S está dada por la siguiente igualdad excepto por términos de orden cúbico en y_1 y y_2 ,

$$(3.4) \quad S(\bar{y}) \approx \frac{1}{2} \bar{y}^t \begin{bmatrix} \bar{F}_{x_1 x_1}^t \hat{N} & \bar{F}_{x_1 x_2}^t \hat{N} \\ \bar{F}_{x_2 x_1}^t \hat{N} & \bar{F}_{x_2 x_2}^t \hat{N} \end{bmatrix} \bar{y},$$

el segundo miembro de esta expresión es conocido como la segunda forma diferencial de la superficie en $\bar{F}(\bar{x}_0)$ y la escribimos en la forma

$$(3.5) \quad \bar{y}^t S_{II} \bar{y},$$

donde

$$S_{II} = \begin{bmatrix} \bar{F}_{x_1 x_1}^t \hat{N} & \bar{F}_{x_1 x_2}^t \hat{N} \\ \bar{F}_{x_1 x_2}^t \hat{N} & \bar{F}_{x_2 x_2}^t \hat{N} \end{bmatrix}.$$

El determinante de S_{II} caracteriza el tipo de paraboloides que más se parece a la superficie en el punto $\bar{F}(\bar{x}_0)$ [9].

Observamos que si $\hat{N}(\bar{x}_0)^t = [n_1, n_2, n_3]$ y $\bar{F}^t = [f_1, f_2, f_3]$ entonces

$$S_{II}(\bar{x}) = \sum_{i=1}^3 n_i \begin{bmatrix} \frac{\partial^2 f_i}{\partial \bar{x}_1^2} & \frac{\partial^2 f_i}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_i}{\partial x_1 \partial x_2} & \frac{\partial^2 f_i}{\partial \bar{x}_2^2} \end{bmatrix} ,$$

esta representación de la matriz S_{II} , nos será de mucha utilidad más adelante.

1.2 Curvas de Curvatura Mínima en una Superficie.

Del análisis de la sección anterior se desprende de manera natural la pregunta:

¿Cuál es la curvatura mínima que puede tener una curva que pase por un punto dado de una superficie?

En base al método utilizado en la sección precedente es claro que dichas curvas se deben obtener mediante cortes transversales a la superficie, es decir mediante cortes normales al plano tangente (vease la figura 3.H).

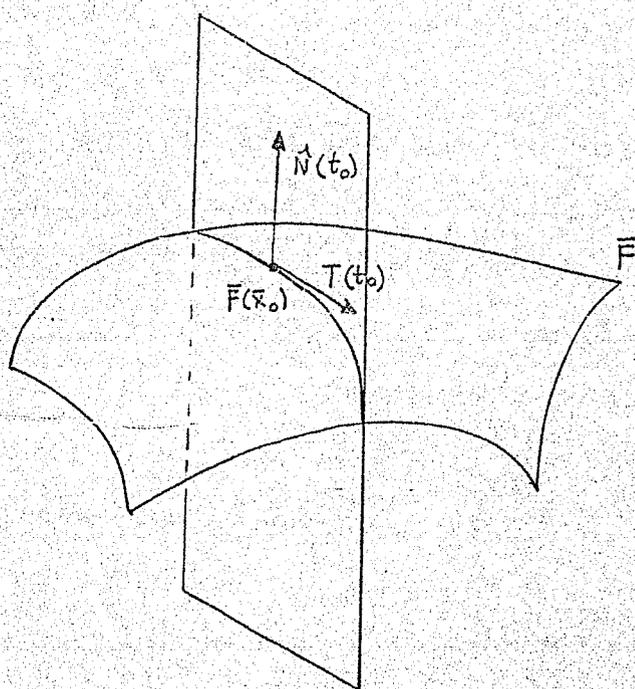


figura 3.H

Para confirmar nuestras sospechas procedemos a analizar uno de dichos cortes.

Tomemos un plano $\Pi_{\bar{T}}$, normal a la superficie en el punto $\bar{F}(\bar{x}_0)$, sea \bar{T} el vector tangente unitario a la curva L_n definida por la intersección del plano y la superficie, esta curva entre las curvas sobre la superficie que pasan por $\bar{F}(\bar{x}_0)$, debe tener en cierto sentido curvatura mínima. Para convencernos de lo anterior consideremos un plano Π_s que contenga a \bar{T} y no contenga a $\hat{N}(\bar{x}_0)$, establezcamos que la curva definida por este nuevo plano y la superficie tiene mayor curvatura que aquella obtenida con el corte transversal. Es decir, queremos demostrar que la curva L_n , obtenida con el corte transversal, es la curva con curvatura mínima de todas aquellas que tienen como tangente en $\bar{F}(\bar{x}_0)$ a \bar{T} (véase la figura 3.I).

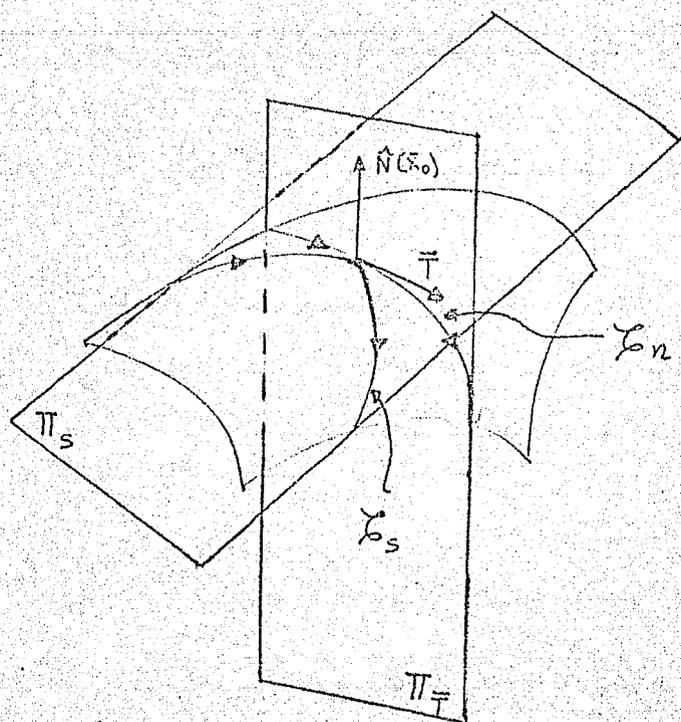


figura 3.I

Observase que la figura anterior se parece a la 2.6 del capítulo precedente. En efecto estamos en un caso semejante y podemos "pensar" que localmente la curva definida por el corta transversal Γ_n es la proyección de la curva Γ_s sobre el plano Π_T .

Nota: Lo anterior no quiere decir que la curva \mathcal{E}_n sea la proyección de \mathcal{E}_s , si no que, solamente "muy cerca" de $\bar{F}(\bar{x}_0)$, la curva proyección de \mathcal{E}_s se "confunde" con \mathcal{E}_n .

Procediendo como en el final de la sección 1.2 del capítulo anterior, se puede establecer directamente la relación

$$(3.6) \quad K_{\mathcal{E}_s} \cos \theta = K_{\mathcal{E}_n}$$

en el punto $\bar{F}(\bar{x}_0)$. Lo que demuestra la validez de nuestra afirmación.

Observamos que hay tantos cortes transversales a la superficie como vectores tangentes unitarios en el punto $\bar{F}(\bar{x}_0)$; por lo que debemos investigar como varía esta curvatura mínima al variar el vector tangente \bar{T} . Bueno, pero ¿qué tiene que ver todo esto con nuestro problema?, recordemos en primer lugar que la primera condición suficiente nos dice que el residual es ortogonal al plano —

tangente, así que es natural que estas curvas — con curvatura mínima juegan un papel importante en la interpretación de la segunda condición suficiente.

Lo anterior sugiere que busquemos una expresión explícita para la curvatura de las curvas obtenidas por cortes transversales, para — ello consideremos

$$(3.7) \quad \begin{cases} \bar{w}(t) = \bar{F}(\bar{x}(t)) \\ \dot{\bar{w}}(t) = \bar{F}'(\bar{x}(t)) \dot{\bar{x}}(t) \\ \ddot{\bar{w}}(t) = \bar{F}''(\bar{x}(t)) \ddot{\bar{x}}(t) + \begin{bmatrix} \dot{\bar{x}}^t(t) f_1''(\bar{x}) \dot{\bar{x}}(t) \\ \dot{\bar{x}}^t(t) f_2''(\bar{x}) \dot{\bar{x}}(t) \\ \dot{\bar{x}}^t(t) f_3''(\bar{x}) \dot{\bar{x}}(t) \end{bmatrix} \end{cases}$$

donde

$$f_i'' = \begin{bmatrix} \frac{\partial^2 f_i}{\partial x_1 \partial x_1} & \frac{\partial^2 f_i}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_i}{\partial x_1 \partial x_2} & \frac{\partial^2 f_i}{\partial x_2 \partial x_2} \end{bmatrix} ;$$

si \hat{N}_w y \hat{T}_w son los vectores normal y tangente principales a la curva $\bar{w}(t)$ entonces [9]

$$(3.8) \quad \ddot{\bar{w}} = K_w \hat{N}_w \dot{s}_w^2 - \dot{s}_w \ddot{s}_w \hat{T}_w$$

de donde , si $\hat{N}_s = \hat{N}(\bar{x}_0)$ y

$$\hat{N}_s^t \hat{N}_w = \cos \theta$$

entonces utilizando (3.7) y (3.8) tenemos

$$(3.9) \quad \hat{N}_s^t \ddot{\bar{w}} = K_w \dot{s}_w^2 \cos \theta \quad y$$

$$(3.10) \quad \hat{N}_s^t \ddot{\bar{w}} = \hat{N}_s^t \begin{bmatrix} \dot{\bar{x}}^t f_1'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_2'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_3'' \dot{\bar{x}} \end{bmatrix}$$

$$= [n_1, n_2, n_3] \begin{bmatrix} \dot{\bar{x}}^t f_1'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_2'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_3'' \dot{\bar{x}} \end{bmatrix}$$

$$= \dot{\bar{x}}^t \left(\sum_{i=1}^3 n_i \begin{bmatrix} \frac{\partial^2 f_i}{\partial x_1 \partial x_1} & \frac{\partial^2 f_i}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_i}{\partial x_1 \partial x_2} & \frac{\partial^2 f_i}{\partial x_2 \partial x_2} \end{bmatrix} \right) \dot{\bar{x}}$$

si \hat{N}_w y \hat{T}_w son los vectores normal y tangente principales a la curva $\bar{w}(t)$ entonces [9]

$$(3.8) \quad \ddot{\bar{w}} = K_w \hat{N}_w \dot{S}_w^2 - \dot{S}_w \ddot{S}_w \bar{T}_w$$

de donde , si $\hat{N}_s = \hat{N}(\bar{x}_0)$ y

$$\hat{N}_s^t \hat{N}_w = \cos \theta$$

entonces utilizando (3.7) y (3.8) tenemos

$$(3.9) \quad \hat{N}_s^t \ddot{\bar{w}} = K_w \dot{S}_w^2 \cos \theta \quad y$$

$$(3.10) \quad \hat{N}_s^t \ddot{\bar{w}} = \hat{N}_s^t \begin{bmatrix} \dot{\bar{x}}^t f_1'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_2'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_3'' \dot{\bar{x}} \end{bmatrix}$$

$$= [n_1, n_2, n_3] \begin{bmatrix} \dot{\bar{x}}^t f_1'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_2'' \dot{\bar{x}} \\ \dot{\bar{x}}^t f_3'' \dot{\bar{x}} \end{bmatrix}$$

$$= \dot{\bar{x}}^t \left(\sum_{i=1}^3 n_i \begin{bmatrix} \frac{\partial^2 f_i}{\partial x_1 \partial x_1} & \frac{\partial^2 f_i}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f_i}{\partial x_1 \partial x_2} & \frac{\partial^2 f_i}{\partial x_2 \partial x_2} \end{bmatrix} \right) \dot{\bar{x}}$$

$$= \dot{\bar{x}}^t S_{II} \dot{\bar{x}}$$

de donde

$$\dot{\bar{x}}^t S_{II} \dot{\bar{x}} = K_w \dot{S}_w^2 \cos \theta$$

pero

$$\dot{S}_w^2 = \|\bar{F}' \dot{\bar{x}}\|^2 = \dot{\bar{x}}^t (\bar{F}'^t \bar{F}') \dot{\bar{x}} = \dot{\bar{x}}^t S_I \dot{\bar{x}}$$

donde $S_I = \bar{F}'^t \bar{F}'$, se reconoce inmediatamente, como la primera forma fundamental de la superficie [9].

En conclusión

$$(3.11) \quad K_w \cos \theta = \frac{\dot{\bar{x}}^t S_{II} \dot{\bar{x}}}{\dot{\bar{x}}^t S_I \dot{\bar{x}}}$$

(véase la figura 3.J).

Esta última relación nos demuestra — analíticamente el resultado, que con cortes — normales se obtienen curvas con curvatura mí

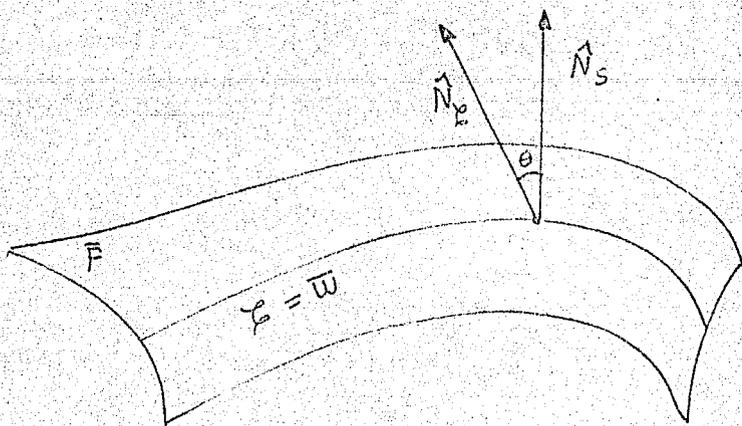


figura 3.J.

nima. En este caso llamaremos K_n a K_w —
claramente obtenemos

$$(3.12) \quad K_n = \frac{|\dot{\bar{x}}^t S_{II} \dot{\bar{x}}|}{\dot{\bar{x}}^t S_I \dot{\bar{x}}}$$

Definimos $K_n(\dot{\bar{x}})$ como

$$(3.13) \quad K_n(\dot{\bar{x}}) = \frac{\dot{\bar{x}}^t S_{II} \dot{\bar{x}}}{\dot{\bar{x}}^t S_I \dot{\bar{x}}}$$

claramente K_n es independiente de la norma
de $\dot{\bar{x}}$ y su valor absoluto nos dá el valor

de la curvatura del corte transversal en $\bar{F}(\bar{x}_0)$ y con tangente $\bar{F}'(\bar{x}_0) \hat{x}$.

Procedamos ahora a la interpretación para la segunda condición para un mínimo en el caso de superficies en \mathbb{R}^3 .

1.3 Interpretación de la Segunda Condición para Superficies.

En este caso la función H a minimizar es

$$(3.14) \quad H(\bar{x}) = \|\bar{F}(\bar{x}) - \bar{y}\|^2$$

donde \bar{F} es regular, la expresión de nuestro interés es la de la segunda derivada de la función H y esta es

$$(3.15) \quad \ddot{H}(\bar{x}) = 2 \left(\bar{x}^t \bar{F}'^t \bar{F}' \bar{x} + \sum_{i=1}^3 (f_i - y_i) f_i'' \right)$$

de la primera condición obtuvimos que $\bar{y}(\bar{x}^*) = \bar{y} - \bar{F}(\bar{x}^*)$ es ortogonal al plano tangente en el punto $F(\bar{x}^*)$, luego si $\bar{r} = \gamma \bar{N}_s$ entonces la segunda derivada toma en \bar{x}^* la forma

$$(3.16) \quad H''(\bar{x}^*) = 2(S_I(\bar{x}^*) - \gamma S_{II}(\bar{x}^*)),$$

la segunda condición de suficiencia para que \bar{x}^* sea un mínimo de H es que (3.16) sea positiva definida, para lo que es necesario que

$$(3.17) \quad \begin{aligned} \bar{u}^t \ddot{H} \bar{u} &= 2\bar{u}^t (S_I - \gamma S_{II}) \bar{u} \\ &= 2\bar{u}^t S_I \bar{u} \left(1 - \gamma \frac{|\bar{u}^t S_{II} \bar{u}|}{\bar{u}^t S_I \bar{u}}\right) \\ &= 2\bar{u}^t S_I \bar{u} (1 - \gamma K_n(\bar{u})) \end{aligned}$$

sea mayor que cero si $\bar{u} \neq \bar{0}$, $\bar{u}^t S_I \bar{u}$ es siempre positiva ya que $S_I = \bar{F}'^t \bar{F}'$ y \bar{F}' es de rango n , por consiguiente para que $H''(\bar{x}^*)$ sea positiva es necesario que

$$1 - r K_n(\bar{u}) > 0$$

para toda $\bar{u} \neq \bar{0}$.

Esta condición es muy fácil de interpretar geométricamente, ya que si tomamos un corte transversal que contenga al vector $F'(x^*)\bar{u} = \bar{t}$ (véase la figura 3.K).

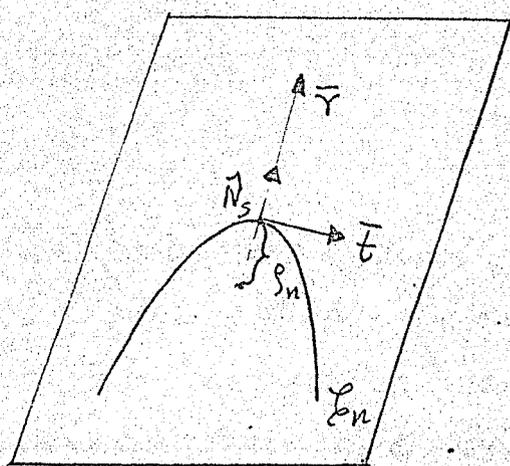


figura 3.K

Entonces para la curva C_n definida por el corte, la condición nos dice lo mismo que en el caso de curvas planas del capítulo ante

$$1 - \gamma K_n(\bar{u}) > 0$$

para toda $\bar{u} \neq \bar{0}$.

Esta condición es muy fácil de interpretar geoméricamente, ya que si tomamos un corte transversal que contenga al vector $F'(x^*)\bar{u} = \bar{F}$ (vease la figura 3.K).

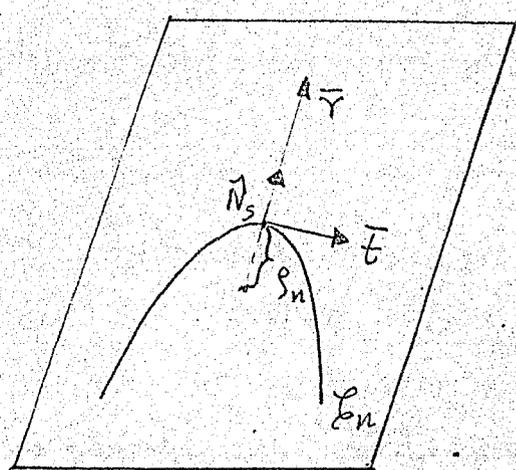


figura 3.K

Entonces para la curva C_n definida por el corte, la condición nos dice lo mismo que en el caso de curvas planas del capítulo anterior.

rior. Es decir, que $\bar{F}(x^*)$ es el punto más cercano a \bar{y} si $\bar{F}(x^k)$ es el punto más cercano a \bar{y} para cada ξ_n .

Localmente, la superficie \bar{F} se parece a un paraboloides, por lo que veremos en que se traduce la condición anterior.

Paraboloides Elíptico.

En este caso todas las curvas ξ_n tienen su concavidad hacia el mismo lado, entre las cuales hay una con curvatura mínima y otra con curvatura máxima (véase la figura 3. L). Por consiguiente la magnitud del residual tiene que ser menor que el radio de curvatura más chico si \bar{y} está del lado convexo del paraboloides y arbitrario en el caso contrario.

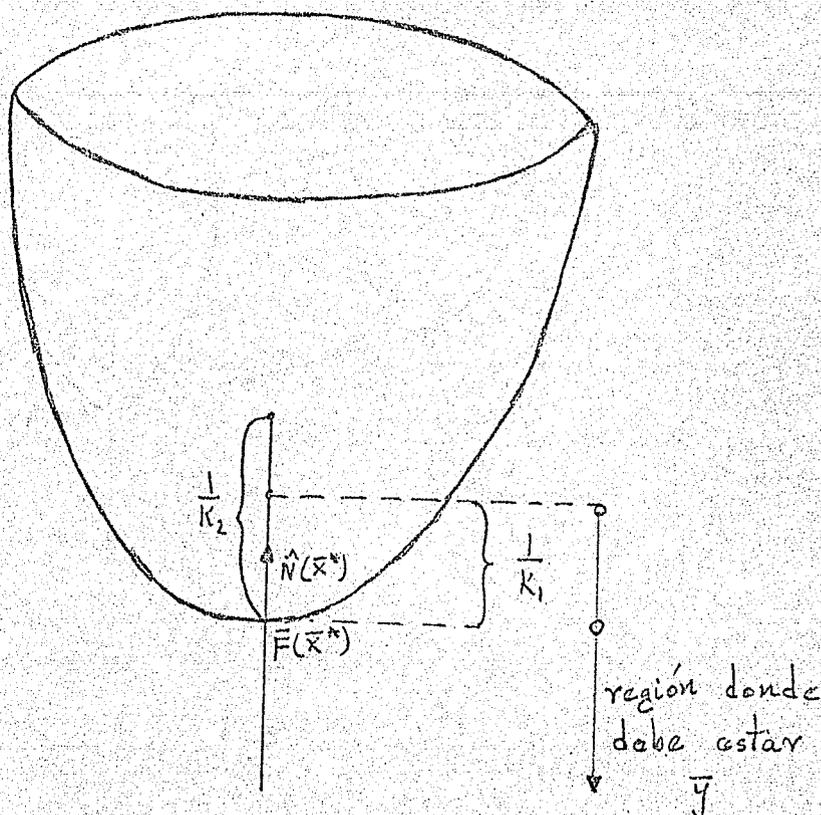
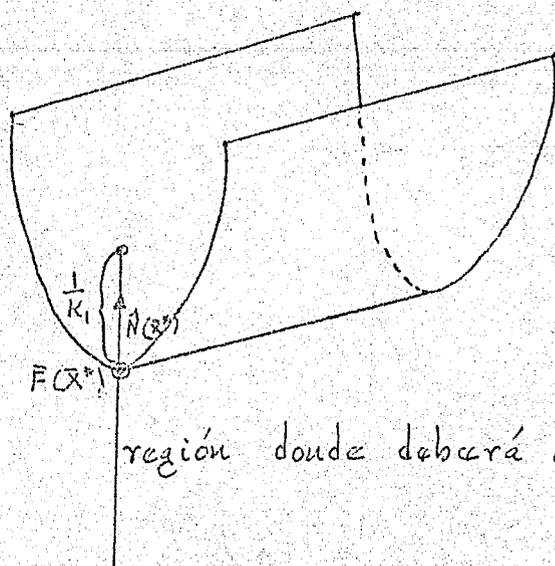


figura 3.2

Paraboloide Cilíndrico.

Este caso es el mismo que el anterior, solo que la curvatura mínima es cero (vease la figura 3.M).



región donde deberá estar el punto \bar{y} .

figura 3.M

Paraboloide Hiperbólico.

En este caso las curvas C_n se pueden agrupar en dos tipos: las que tienen la concavidad hacia donde está la normal a la superficie y las que no. Cada grupo tiene una curva con curvatura máxima.

Por lo que \bar{y} deberá estar en el seg-

mento de recta comprendido entre los centros de curvatura (véase la figura 3.N).

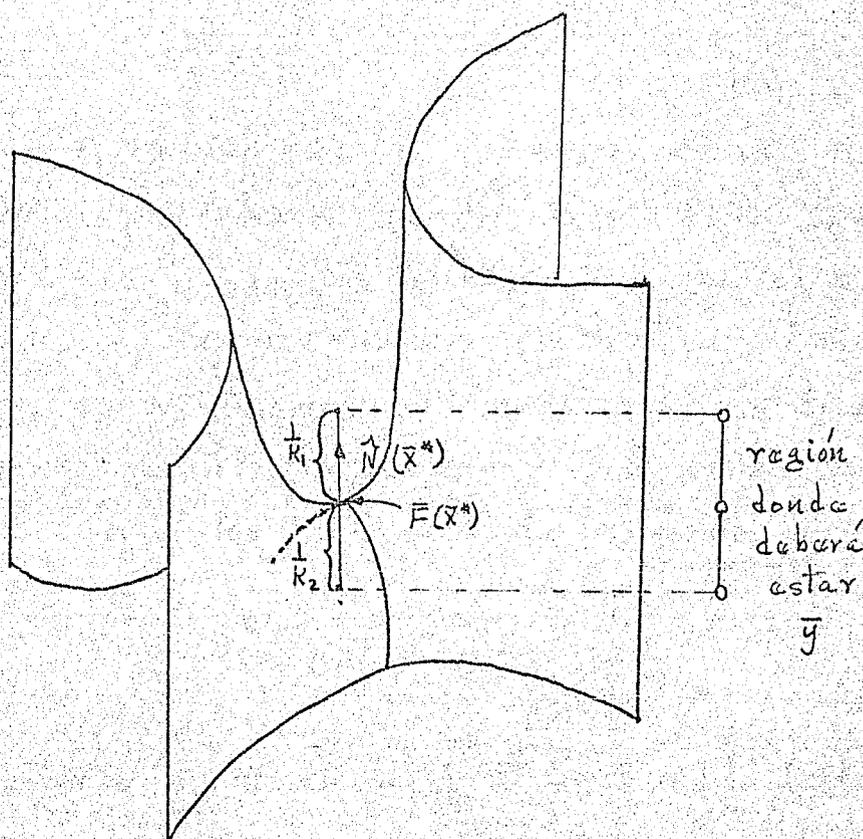


figura 3.N

De la discusión anterior se desprende — que lo que nos interesa son los máximos y mínimos de

$$K_n(\bar{u}) = \frac{\bar{u}^t S_{II} \bar{u}}{\bar{u}^t S_I \bar{u}} .$$

Como $K_n(\bar{u})$ no depende de la norma de \bar{u} , es conveniente normalizar de tal forma que sea fácil optimizar $K_n(\bar{u})$; luego hay que convertir a $K_n(\bar{u})$ en una forma cuadrática. Esto se logra si utilizamos al denominador $\bar{u}^t S_I \bar{u}$ como norma de \bar{u} , de donde el problema se reduce a calcular los extremos de la forma cuadrática

$$\bar{v}^t S_{II} \bar{v}$$

en el conjunto $\{\bar{v} \text{ en } \mathbb{R}^2 \mid \bar{v}^t S_I \bar{v} = 1\}$ (este problema es muy conocido).

Aplicando multiplicadores de Lagrange se obtiene que las soluciones están dadas por las soluciones del problema de eigenvalores generalizado

$$(3.18) \quad S_{II} \bar{v} = \lambda S_I \bar{v}$$

si λ_1, λ_2 son los eigenvalores y \bar{v}_1, \bar{v}_2 los correspondientes eigenvectores, entonces

$$\bar{v}_i^t S_I \bar{v}_j = \delta_{ij} \quad (i, j = 1, 2).$$

Además las curvaturas máxima y mínima están dadas por

$$K_{\max} = \max \{ |\lambda_1|, |\lambda_2| \}$$

$$K_{\min} = \min \{ |\lambda_1|, |\lambda_2| \}$$

si $\bar{u} = \eta_1 \bar{v}_1 + \eta_2 \bar{v}_2$ entonces

$$K_n(\bar{u}) = \frac{\lambda_1 \eta_1^2 + \lambda_2 \eta_2^2}{\eta_1^2 + \eta_2^2}.$$

Sustituyendo ahora esta expresión en $\bar{u}^t H''(\bar{x}^*) \bar{u}$ obtenemos

$$(3.19) \quad \bar{u}^t H''(\bar{x}^*) \bar{u} = \bar{u}_1^t S_I \bar{u} \left(1 - \gamma \frac{\lambda_1 u_1^2 + \lambda_2 u_2^2}{u_1^2 + u_2^2}\right) \\ = (1 - \gamma \lambda_1) u_1^2 + (1 - \gamma \lambda_2) u_2^2 .$$

Finalmente concluimos que $H''(\bar{x}^*)$ es positiva definida si

$$(3.20) \quad \begin{cases} 1 - \gamma \lambda_1 > 0 \\ 1 - \gamma \lambda_2 > 0 \end{cases} .$$

Ahora estamos en posibilidad de autan-
dar mejor el servicio que nos presta la fórmula

$$K_n(\bar{u}) = \frac{\bar{u}^t S_{II} \bar{u}}{\bar{u}^t S_I \bar{u}} .$$

ya que, no solo nos dá la magnitud de la cur-
vatura si no que nos indica el cambio de con-
cavidad de las secciones transversales, lo -
que se manifiesta cuando λ_1, λ_2 tienen -
signos contrarios.

Si $0 < \lambda_1 < \lambda_2$ entonces

$$\left. \begin{array}{l} 1 - r\lambda_1 > 0 \\ 1 - r\lambda_2 > 0 \end{array} \right\} \Rightarrow \left\{ r < \frac{1}{\lambda_2} \right.$$

este caso corresponde al del paraboloides elíptico.

Si $\lambda_1 < 0 < \lambda_2$ entonces

$$\left. \begin{array}{l} 1 - r\lambda_1 > 0 \\ 1 - r\lambda_2 > 0 \end{array} \right\} \Rightarrow \left\{ \frac{1}{\lambda_1} < r < \frac{1}{\lambda_2} \right.$$

este caso corresponde al del paraboloides hiperbólico.

Las otras posibilidades para λ_1, λ_2 son análogas.

Así hemos logrado finalmente interpretar geométricamente la condición $\bar{u}^t H''(x^*) \bar{u} > 0$, pero más aun, en el camino descubrimos una herramienta analítica que utilizaremos para el caso general.

1.4 Interpretación de la Segunda Condición para el Caso General.

Recordemos que la expresión para la segunda derivada es

$$H''(\bar{x}) = 2(\bar{F}'(\bar{x})^t \bar{F}'(\bar{x}) + \sum_{i=1}^m (f_i(\bar{x}) - y_i) f_i''(\bar{x}))$$

donde

$$f_i''(\bar{x}) = \left[\frac{\partial^2 f_i(\bar{x})}{\partial x_j \partial x_k} \right]; \quad j, k = 1, \dots, n.$$

Notemos $S_I = \bar{F}'(\bar{x})^t \bar{F}'(\bar{x})$ es positiva definida, ya que la "superficie" que estamos considerando es regular.

Así lo que necesitamos para aplicar la técnica es ver si el segundo término de $H''(\bar{x})$

$$(3.21) \quad \sum_{i=1}^m (f_i(\bar{x}) - y_i) f_i''(\bar{x})$$

es análogo a la segunda forma fundamental. Recordemos que para superficies la segunda-

forma fundamental esta dada por

$$\sum_{i=1}^3 n_i f_i''(\bar{x})$$

donde $\vec{N}_s = [n_1, n_2, n_3]$ es el vector normal unitario a la superficie.

El problema que se nos presenta ahora es que por lo general en esta clase de problemas $n < m$, lo que quiere decir que no hay una única dirección normal a la "superficie" generada por \bar{F} , pues esta es de dimensión n .

En este momento es cuando la experiencia del capítulo anterior nos es útil, ya que ahí vimos como interpretar la segunda condición exclusivamente en términos de los vectores tangente y residual.

Uniendo a lo anterior el hecho que el vector residual nos determina una dirección normal a la "superficie" en $\bar{F}(\bar{x}^*)$ es natural definir

$$\hat{N}_s = \hat{N}_s(\bar{x}^*) = \frac{\bar{r}}{\|\bar{r}\|} = \hat{r}$$

donde $\bar{r} = \bar{y} - \bar{F}(\bar{x}^*)$. Lo que a su vez nos permite definir

$$S_{II}(\hat{r}) = \sum_{i=1}^m \gamma_i f_i''(\bar{x}^*)$$

en $\bar{F}(\bar{x}^*)$, donde $\hat{r} = [\gamma_1, \dots, \gamma_m]$.

Con los elementos anteriores escribimos

$$(3.22) \quad H''(\bar{x}^*) = 2(S_I(\bar{x}^*) - \|\bar{r}\| S_{II}(\hat{r}))$$

ahora si, procedemos en forma análoga al caso de superficies en \mathbb{R}^3 .

Sean $\{\lambda_i, \bar{v}_i\}$ $i=1, \dots, n$ tales que

$$\begin{cases} S_{II}(\hat{r}) \bar{v}_i = \lambda_i S_I \bar{v}_i \\ \bar{v}_i S_I \bar{v}_j = \delta_{ij}, \end{cases}$$

sea \bar{u} en \mathbb{R}^n con.

$$(3.23) \quad \bar{u} = \sum_{i=1}^n \mu_i \bar{v}_i$$

sustituyendo en $\bar{u}^t H''(\bar{x}^*) \bar{u}$, obtenemos

$$(3.24) \quad \bar{u}^t H''(\bar{x}^*) \bar{u} = 2 \sum_{i=1}^n \mu_i^2 (1 - \|\bar{r}\| \lambda_i)$$

por consiguiente $H''(\bar{x}^*)$ es positiva definida si

$$(3.25) \quad 1 - \|\bar{r}\| \lambda_i > 0 \quad \text{para toda } i=1, \dots, n.$$

Para interpretar consideremos los siguientes casos:

i) Si $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ entonces

$$\|\bar{r}\| < \frac{1}{\lambda_n} \leq \frac{1}{\lambda_i}.$$

ii) Si $\lambda_n \leq \lambda_{n-1} \leq \dots \leq \lambda_1 < 0$ entonces

$$1 - \|\bar{r}\| \lambda_i$$

es siempre positivo, y entonces $\|\bar{r}\|$ puede ser arbitraria. Estos dos casos corresponden al caso elíptico.

(iii) Si $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 0 < \lambda_{s+1} \leq \dots \leq \lambda_n$ entonces

$$\|\bar{r}\| < \frac{1}{\lambda_n}$$

este corresponde al caso hiperbólico.

Quando algunas λ_i sean cero, se tendrán ~~situaciones análogas a las anteriores.~~

2. Método de Gauss.

En forma análoga al capítulo anterior el método de Gauss obtiene una corrección a una aproximación \bar{x}_0 dada, de la siguiente forma:

i) Se aproxima a la "superficie" $\bar{F}(\bar{x})$ por el subespacio lineal tangente en el punto $\bar{F}(\bar{x}_0)$, esto es

$$\bar{F}(\bar{x}) \approx \bar{F}(\bar{x}_0) + \bar{F}'(\bar{x}_0)(\bar{x} - \bar{x}_0).$$

ii) De esta forma ahora se procede — como si se tratara de un problema de estimación lineal minimizando

$$\| \bar{y} - \bar{F}(\bar{x}_0) - \bar{F}'(\bar{x}_0) \bar{\alpha} \|^2$$

la $\bar{\alpha}^*$ que minimiza esta dada por

$$\bar{\alpha}^* = \left(\bar{F}'(\bar{x}_0)^t \bar{F}'(\bar{x}_0) \right)^{-1} \bar{F}'(\bar{x}_0)^t (\bar{y} - \bar{F}(\bar{x}_0))$$

que como se puede reconocer, son las ecuaciones normales.

iii) Así se obtiene una corrección

$$\begin{aligned} \bar{x}_1 &= \bar{x}_0 + \bar{\alpha}^* \\ &= \bar{x}_0 + \left(\bar{F}'(\bar{x}_0)^t \bar{F}'(\bar{x}_0) \right)^{-1} \bar{F}'(\bar{x}_0)^t (\bar{y} - \bar{F}(\bar{x}_0)) \end{aligned}$$

en la figura 3.0 se ilustra esto para el caso $n=2$ y $m=3$.

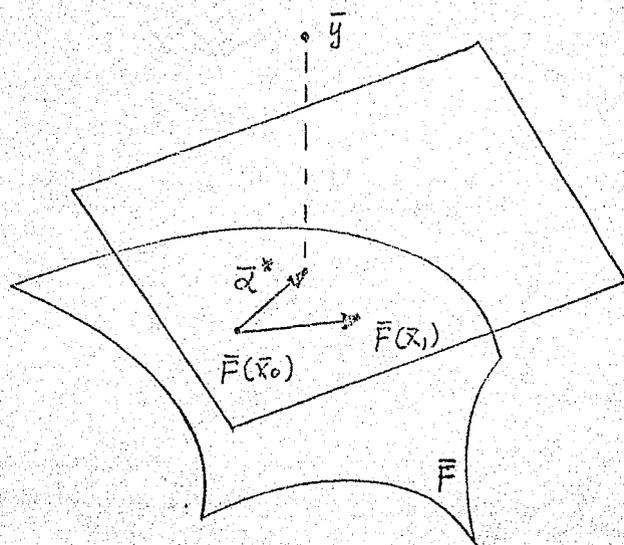


figura 3.0

2.1 Convergencia.

Ahora tenemos la fórmula de iteración

$$(3.26) \quad \bar{x}_{n+1} = \bar{b}(\bar{x}_n) = \bar{x}_n + (\bar{F}'^t(\bar{x}_n) \bar{F}'(\bar{x}_n))^{-1} \bar{F}'^t(\bar{x}_n) \cdot (\bar{y} - \bar{F}(\bar{x}_n)),$$

lo que necesitamos encontrar es un punto fijo \bar{x}^* , esto es

$$(3.27) \quad \bar{G}(\bar{x}^*) = \bar{x}^*$$

mediante los mismos argumentos del capítulo anterior, procedemos a estimar

$$(3.28) \quad \bar{x}_{n+1} - \bar{x}^* = \bar{J}(\bar{x}_n - \bar{x}^*).$$

Desarrollando en series de Taylor la función (3.26) alrededor del punto \bar{x}^* tenemos

$$(3.29) \quad \bar{x}_{n+1} = \bar{x}^* + \bar{G}'(\bar{x}^*)(\bar{x}_n - \bar{x}^*) + o(\|\bar{x}_n - \bar{x}^*\|^2)$$

luego

$$\bar{x}_{n+1} - \bar{x}^* = \bar{G}'(\bar{x}^*)(\bar{x}_n - \bar{x}^*) + o(\|\bar{x}_n - \bar{x}^*\|^2)$$

pero \bar{G}' tiene expresión

$$(3.30) \quad \begin{aligned} \bar{G}'(\bar{x}^*) &= (\bar{F}'(\bar{x}^*)^t \bar{F}'(\bar{x}^*))^{-1} \bar{F}''(\bar{x}^*) (\bar{y} - \bar{F}(\bar{x}^*)) \\ &= S_I^{-1} \left(\sum_{i=1}^m (y_i - f_i(\bar{x}^*)) f_i''(\bar{x}^*) \right) \end{aligned}$$

$$= \|\bar{\gamma}\| S_I^{-1} S_{II}(\bar{\gamma})$$

de esto se sigue que

$$(3.31) \quad \bar{\epsilon}_{n+1} = \|\bar{\gamma}\| S_I^{-1} S_{II}(\bar{\gamma}) \bar{\epsilon}_n + o(\|\bar{\epsilon}_n\|^2)$$

donde $\bar{\epsilon}_n = \bar{x}_n - \bar{x}^*$.

Ahora la sucesión $\{\bar{x}_n\}$ converge si los eigenvalores de

$$\|\bar{\gamma}\| S_I^{-1} S_{II}(\bar{\gamma})$$

son en valor absoluto menor que la unidad []. Los eigenvalores son $\|\bar{\gamma}\| \lambda_i$ de donde

$$\|\bar{\gamma}\| < \frac{1}{\lambda_{\max}}$$

donde $\lambda_{\max} = \max \{|\lambda_i| \text{ con } i=1, \dots, n\}$.

Terminamos el presente trabajo observando que la convergencia del método de Gauss depende de las curvaturas principales

λ_i , asociadas a la dirección del residual
y no solo que el residual sea pequeño.

Bibliografía.

- [1] Pablo Barrera y Jesús López,
"Ecuaciones Lineales Algebraicas vía
Descomposición Singular", Com. Tec.
CIMAS, Vol. I, Serie A, No. 2, 1974.
- [2] Pablo Barrera, Jesús López, Humberto -
Madrid y Ma. Eugenia Reyes,
"La Inversa Generalizada. Un Enfoque
Geométrico y un Método Iterativo para
Calcularla", Com. Tec. CIMAS, Vol. I,
Serie A, No. 7, 1974.
- [3] Charles L. Lawson y Richard J.
Hanson, "Solving Least Squares -
Problems", Ed. Prentice-Hall, 1974.
- [4] Jonathan Bard, "Nonlinear Param-
eter Estimation", Ed. Academic Press,
1974.

- [5] J. Kowalik y M.R. Osborne ,
"Methods for Constrained Optimization
Problems", Ed. Elsevier , 1968.
- [6] Björck , Åke . "Solving Linear Least
Squares Problems by Gram - Schmit -
Orthogonalización" , BIT , 7 , 1-21 , 1967.
- [7] Pablo Barrera , "Nuevos Algoritmos -
para la Estimación No-Lineal de -
Parámetros" , Por aparecer.
- [8] Ortega y Reinbolt , "Iterative -
Solution of Nonlinear Equations in
Several Variables" , Ed. Ac. Press -
1970.
- [9] Stoker J.J. , "Differential -
Geometry" , Ed. Wiley - Interscience
1969 .

[10] J. Dennis Lawrence, "A Catalog of Special Plane Curves", Ed. Dover 1972.

[11] Wedin, Per-Ake, "Perturbation Theory of Pseudo-Inverses", BIT, 13, 217-232, 1973.

[12] Wedin, Per-Ake, "The Nonlinear Least Squares Problems from a Numerical Point of View", Tec. Rep. Univ. of Umea. Sweden.