

"MINIMOS CUADRADOS NO LINEALES:
Los Métodos
Levenberg Marquardt Morrison"

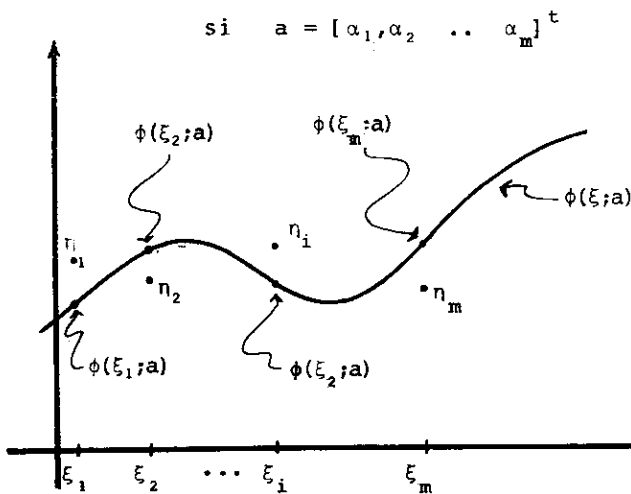
José Guerrero *

I PRESENTACION GENERAL.

Uno de los problemas que con mayor frecuencia se presentan en el trabajo diario de quienes se dedican a las aplicaciones en matemáticas y que es resuelto diariamente un gran número de ocasiones en los centros de cómputo, es el siguiente: dado un conjunto de datos (ξ_i, η_i) , $i = 1, 2, \dots, m$, los cuales en general provienen de algún tipo de observación experimental, se trata de construir un modelo matemático, el cual bajo cierto criterio nos dé una representación global lo más adecuada posible del fenómeno descrito parcialmente por los datos. La fuentes de este tipo de problemas van desde la Economía, la Psicología o la Sociología hasta la Química, la Biología, la Física y la Ingeniería. Se trata pues de un campo rico en posibilidades, y esto no sólo desde el punto de vista práctico, según trataremos de hacer notar en el presente trabajo.

* Profesor del Departamento de Matemáticas de la Facultad de Ciencias de la U.N.A.M. 1980.

Al tipo de problemas mencionados se les conoce como problemas de "ajuste" y la razón de esto es que en general junto con los datos se propone un modelo $\phi(\xi; \alpha_1, \alpha_2, \dots, \alpha_n)$ que depende de los parámetros $\alpha_1, \dots, \alpha_n$ y de lo que se trata entonces es de determinar aquellos $\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*$ de tal manera que $\phi(\xi; \alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)$ se ajuste lo mejor posible a los datos. Gráficamente se tiene:



Existen diversos criterios de ajuste. Uno de los más conocidos y que es el que aquí trataremos es el conocido como ajuste en el sentido de mínimos cuadrados. Este criterio consiste en determinar aquel vector

$$a^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*]^t$$

con la propiedad de que a^* minimice la suma

tros $\alpha_1, \dots, \alpha_n$. En caso contrario, $\phi(\xi; a)$ será no-lineal.

En la sección siguiente haremos notar que en el caso de un modelo lineal, el problema

$$(1) \quad \min \left[\frac{1}{2} \sum_{i=1}^m (\eta_i - \phi(\xi_i; a))^2 \right]$$

puede traducirse en el de resolver un sistema de ecuaciones algebraicas lineales de la forma

$$Aa = C$$

donde A es una matriz simétrica y a es el vector parámetros $\alpha_1, \alpha_2, \dots, \alpha_n$. Si por el contrario se trata de un modelo no-lineal, algunas veces (1) se intenta resolver tomando los ceros del gradiente de $\frac{1}{2} \sum_{i=1}^m (\eta_i - \phi(\xi_i; a))^2$, lo que lleva directamente al problema de resolver un sistema de ecuaciones no-lineales. Sin embargo, la solución de (1) puede también abordarse de manera distinta, y de hecho existen diversos tipos de métodos mediante los cuales se calculan soluciones aproximadas de él en la práctica. Esos métodos consisten esencialmente en un proceso iterativo mediante el cual a partir de una propuesta de aproximación inicial a_0 se genera una sucesión $\{a_n\}$ la que bajo adecuadas condiciones se espera, que converja a la solución a^* . Ejemplo clásicos de este tipo de métodos son los conocidos como: Método del Descenso más Rápido y Método de Gauss-Newton, los que discutiremos

con algún detalle para posteriormente introducir y desarrollar las ideas fundamentales de una cierta familia de métodos conocidos con el nombre colectivo de Levenberg-Marquardt-Morrison que son el resultado de una combinación de los dos anteriores, y que presentan la característica de que en cada iteración se resuelve un problema de ajuste lineal

Haremos también mención de algunas de las dificultades más frecuentes con respecto al cálculo de soluciones aproximadas del problema (1), independientemente del método usado, - tales como: el que las funciones involucradas en el modelo resulten difíciles de manejar, o bien que se tengan dimensiones demasiado grandes debido a un gran número de observaciones y /o de parámetros. Finalmente, presentaremos algunos resultados en los que se garantiza la convergencia de los métodos

L M M.

II.- EL PROBLEMA DE MINIMOS CUADRADOS

Según hemos visto, el problema de ajuste en el sentido de mínimos cuadrados puede plantearse como sigue: dado un conjunto de datos

$$(\xi_i, \eta_i), \quad i = 1, 2, \dots, m$$

y un modelo

$$\phi(\xi; \alpha_1, \alpha_2, \dots, \alpha_n)$$

que depende de los parámetros $\alpha_1, \alpha_2, \dots, \alpha_n$, se trata de determinar el vector

$$a^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*]^t$$

con la propiedad de que a^* minimice la suma

$$\frac{1}{2} \sum_{i=1}^m (\eta_i - \phi(\xi_i; a))^2,$$

$$a = [\alpha_1, \alpha_2, \dots, \alpha_n]^t.$$

Dado que esta suma depende sólo de a , puesto que los (ξ_i, η_i) están fijos, en lo sucesivo escribiermos

$$r(a) = \frac{1}{2} \sum_{i=1}^m (\eta_i - \phi(\xi_i; a))^2$$

Adoptada esta convención nuestro problema original lo podemos reescribir como

$$(2) \quad \min_a r(a) , a \in \Omega \subset \mathbb{R}^n$$

A continuación presentamos algunos ejemplos mediante los cuales se ilustran cuestiones de interés con respecto a este tipo de problemas.

1) Supongamos dado el conjunto (ξ_i, η_i) , $i = 1, \dots, m$ de datos correspondientes a un fenómeno cuyo comportamiento es del tipo, digamos,

$$\eta = \alpha_1 + \alpha_2 \xi + \alpha_3 \xi^2 = \phi(\xi; \alpha_1, \alpha_2, \alpha_3) = \phi(\xi; a)$$

El problema en este caso consiste en determinar $\alpha_1^*, \alpha_2^*, \alpha_3^*$, con propiedad de que la gráfica de $\phi(\xi; \alpha_1^*, \alpha_2^*, \alpha_3^*)$ se ajuste lo mejor posible a los datos dados; es decir, se quiere determinar $a^* = [\alpha_1^*, \alpha_2^*, \alpha_3^*]^t$ tal que minimice

$$\begin{aligned} r(a) &= \frac{1}{2} \sum_{i=1}^m (\eta_i - (\alpha_1 + \alpha_2 \xi_i + \alpha_3 \xi_i^2))^2 \\ &= \frac{1}{2} \sum_{i=1}^m (\eta_i - \sum_{j=1}^3 \alpha_j \xi_i^{j-1})^2 \end{aligned}$$

Ahora, dado que $\sum_{j=1}^3 \alpha_j \xi_i^{j-1}$ se puede ver como el i -ésimo elemento del producto Ta , donde T es la matriz de elementos $(\tau_{ij}) = (\xi_i^{j-1})$, resulta que nuestro problema lo podemos plantear como: resolver

$$\min_a r(a) = \frac{1}{2} \|y - Ta\|^2$$

donde

$$y = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix}, T = \begin{bmatrix} 1 & \xi_1 & \xi_1^2 \\ \vdots & \vdots & \vdots \\ 1 & \xi_m & \xi_m^2 \end{bmatrix}, a = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

Sabemos que una condición necesaria para tener un punto estacionario de $r(a)$ es que

$$\frac{\partial r(a)}{\partial \alpha_i} = 0, \quad i = 1, 2, 3.$$

Usando este hecho y dado que

$$\frac{\partial r(a)}{\partial \alpha_1} = \sum_{i=1}^m (\eta_i - (\alpha_1 + \alpha_2 \xi_i + \alpha_3 \xi_i^2)) (-1)$$

$$\frac{\partial r(a)}{\partial \alpha_2} = \sum_{i=1}^m (\eta_i - (\alpha_1 + \alpha_2 \xi_i + \alpha_3 \xi_i^2)) (-\xi_i)$$

$$\frac{\partial r(a)}{\partial \alpha_3} = \sum_{i=1}^m (\eta_i - (\alpha_1 + \alpha_2 \xi_i + \alpha_3 \xi_i^2)) (-\xi_i^2)$$

se llega al sistema

$$\begin{aligned} \sum_{i=1}^m \alpha_1 + \alpha_2 \sum_{i=1}^m \xi_i + \alpha_3 \sum_{i=1}^m \xi_i^2 &= \sum_{i=1}^m \eta_i \\ \alpha_1 \sum_{i=1}^m \xi_i + \alpha_2 \sum_{i=1}^m \xi_i^2 + \alpha_3 \sum_{i=1}^m \xi_i^3 &= \sum_{i=1}^m \eta_i \xi_i \\ \alpha_1 \sum_{i=1}^m \xi_i^2 + \alpha_2 \sum_{i=1}^m \xi_i^3 + \alpha_3 \sum_{i=1}^m \xi_i^4 &= \sum_{i=1}^m \eta_i \xi_i^2 \end{aligned}$$

que puede escribirse como

$$\begin{bmatrix} m & & \sum_{i=1}^m \xi_i & \sum_{i=1}^m \xi_i^2 \\ \sum_{i=1}^m \xi_i & \sum_{i=1}^m \xi_i^2 & \sum_{i=1}^m \xi_i^3 & \\ \sum_{i=1}^m \xi_i^2 & \sum_{i=1}^m \xi_i^3 & \sum_{i=1}^m \xi_i^4 & \\ \sum_{i=1}^m \xi_i^3 & \sum_{i=1}^m \xi_i^4 & \sum_{i=1}^m \xi_i^5 & \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m \eta_i \\ \sum_{i=1}^m \eta_i \xi_i \\ \sum_{i=1}^m \eta_i \xi_i^2 \end{bmatrix}$$

o bien

$$(3) \quad (T^t T) a = T^t y$$

que es un sistema de ecuaciones algebraicas lineales, donde T es la matriz que vimos antes. Se puede checar fácilmente (bajo el supuesto de que $\xi_i \neq \xi_j$ si $i \neq j$) que $T^t T$ es no-singular, de donde resulta que la solución de (3) resuelve también nuestro problema original

2) Tomemos de nuevo la colección de datos

$$(\xi_i, \eta_i), \quad i = 1, 2, \dots, m; \quad \xi_i \neq \xi_j \quad \text{si} \quad i \neq j$$

y sea en este caso

$$\phi(\xi; a) = \alpha_1 e^{3\xi} + \alpha_2 \operatorname{sen} \xi + \alpha_3 \xi$$

El problema ahora es

$$\begin{aligned} \min_a r(a) &= \min_a \left[\frac{1}{2} \sum_{i=1}^m (\eta_i - (\alpha_1 e^{3\xi_i} + \alpha_2 \operatorname{sen} \xi_i + \alpha_3 \xi_i))^2 \right] \\ &= \min_a \|y - T_1 a\|^2 \end{aligned}$$

donde

$$y = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix} \quad T = \begin{bmatrix} e^{3\xi_1} & \text{sen } \xi_1 & \xi_1 \\ \vdots & \vdots & \vdots \\ e^{3\xi_m} & \text{sen } \xi_m & \xi_m \end{bmatrix} \quad , \quad a = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

Siguiendo un procedimiento análogo al del ejemplo anterior se llega al sistema

$$\begin{bmatrix} \sum_{i=1}^m e^{6\xi_i} & \sum_{i=1}^m e^{3\xi_i} \text{sen } \xi_i & \sum_{i=1}^m \xi_i e^{3\xi_i} \\ \sum_{i=1}^m e^{3\xi_i} \text{sen } \xi_i & \sum_{i=1}^m \text{sen}^2 \xi_i & \sum_{i=1}^m \xi_i \text{sen} \xi_i \\ \sum_{i=1}^m \xi_i e^{3\xi_i} & \sum_{i=1}^m \xi_i \text{sen} \xi_i & \sum_{i=1}^m \xi_i^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m \eta_i e^{3\xi_i} \\ \sum_{i=1}^m \eta_i \text{sen} \xi_i \\ \sum_{i=1}^m \eta_i \xi_i \end{bmatrix}$$

que tiene nuevamente la forma

$$(4) \quad (T_1^t T_1) a = T_1^t y$$

y cuya solución nos resuelve nuevamente el problema original

A los sistemas del tipo (3) y (4) se les conoce como las *ecuaciones normales* de los problemas respectivos

3) Determinar el vector

$$a^* = [\alpha_1^*, \quad \alpha_5^*]$$

tal que el modelo

$$\phi(\xi; a^*) = \alpha_1^* + \alpha_2^* e^{-\alpha_4^* \xi} + \alpha_3^* e^{-\alpha_5^* \xi}$$

resulte lo mejor ajustado a los datos

$$(\xi_i, \eta_i), \quad i = 1, 2, \dots, m; \quad \xi_i \neq \xi_j \text{ si } i \neq j$$

Problema a resolver:

$$\min_a r(a) = \min_a \left[\frac{1}{2} \sum_{i=1}^m (\eta_i - (\alpha_1 + \alpha_2 e^{-\alpha_4 \xi_i} + \alpha_3 e^{-\alpha_5 \xi_i}))^2 \right]$$

El tratar ahora de obtener a^* siguiendo el camino de igualar a cero el gradiente de r nos lleva al problema de calcular la solución de un sistema

$$(5) \quad F(a) = 0 \quad (\text{ver Apéndice})$$

de cinco ecuaciones no lineales. La tercera de ellas es por ejemplo,

$$\begin{aligned} \alpha_1 \sum_{i=1}^m e^{-\alpha_5 \xi_i} + \alpha_2 \sum_{i=1}^m e^{-(\alpha_4 \xi_i + \alpha_5 \xi_i)} + \alpha_3 \sum_{i=1}^m e^{-2\alpha_5 \xi_i} &= \\ &= \sum_{i=1}^m \eta_i e^{-\alpha_5 \xi_i} \end{aligned}$$

4) Sean

$$(\xi_i, \eta_i) \quad i = 1, 2, \dots, m; \quad \xi_i \neq \xi_j \text{ si } i \neq j$$

$$\phi(\xi; a) = \frac{\alpha_1}{1 + \alpha_2(\xi - \alpha_3)} + \frac{\alpha_4}{1 + \alpha_5(\xi - \alpha_6)}$$

Problema:

$$\min_a r(a) = \min_a \left[\frac{1}{2} \sum_{i=1}^m (\eta_i \left(\frac{\alpha_1}{1 + \alpha_2(\xi_i - \alpha_3)} + \frac{\alpha_4}{1 + \alpha_5(\xi_i - \alpha_6)} \right))^2 \right]$$

Al igual que en el caso anterior, el seguir el camino indicado antes para calcular a^* da como resultado el llegar a un sistema de ecuaciones no-lineales de la forma

$$(6) \quad F_1(a) = 0 \quad (\text{ver Apéndice})$$

cuya solución nos da la solución del problema original.

Estos dos últimos ejemplos, como se ha observado, difieren de los anteriores en que el sistema de ecuaciones a que da lugar

$$\frac{\partial r(a)}{\partial \alpha_i} = 0, \quad i = 1, 2, \dots, n$$

es en ambos casos un sistema no-lineal, aún a pesar de que algunos α_i aparezcan en forma lineal. El haber llegado en estos casos a este tipo de sistemas se debe a que los modelos propuestos son no lineales, y ésta es la razón para que a los problemas correspondientes se le conozca como problemas de mínimos cuadrados no-lineales

Los sistemas del tipo (5) y (6) son en general difíciles de resolver aún para el caso de una sola ecuación y es por ello que cuando se tiene un problema de ajuste no-lineal se trata en general de evitar resolverlos y se intenta en cambio atacar el problema original por otro camino. Esto será justamente lo que haremos nosotros aquí, y como un primer avance en este sentido en la sección siguiente discutiremos los métodos del descenso más rápido y el de Gauss Newton a que ya nos hemos referido antes.

Para finalizar esta sección presentamos en seguida una interpretación geométrica de nuestro problema

$$\min_a r(a), \quad a \in \Omega \subset \mathbb{R}^n$$

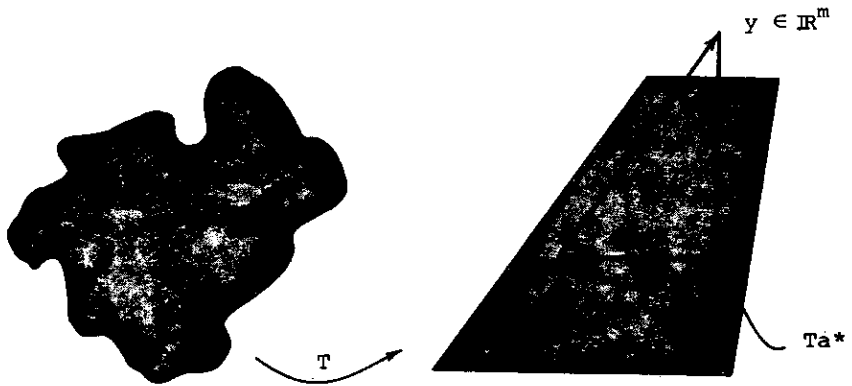
para los casos lineal y no-lineal.

1) Caso lineal

Hemos visto que cuando el modelo $\phi(\xi; a)$ propuesto es lineal, el problema $\min_a r(a)$ se puede escribir como

$$\min_a r(a) = \min_a \frac{1}{2} \|y - Ta\|^2$$

donde T es una matriz de dimensiones $m \times n$. La interpretación geométrica correspondiente a este caso es



Lo que nos dice esta figura es que, siendo T una transformación lineal ($T: \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$), su imagen resulta ser un hiperplano de dimensión a lo más n , contenido en \mathbb{R}^m . Además, dado $y \in \mathbb{R}^m$, la solución a^* del problema de mínimos cuadrados lineales es tal que su imagen bajo T es el punto $Ta^* \in \text{Im}(T) \subset \mathbb{R}^m$ más próximo a y . De esto resulta fácil probar que

$$Ta^* = \text{Proy}_{\text{Im}(T)} y.$$

2) Caso no-lineal.

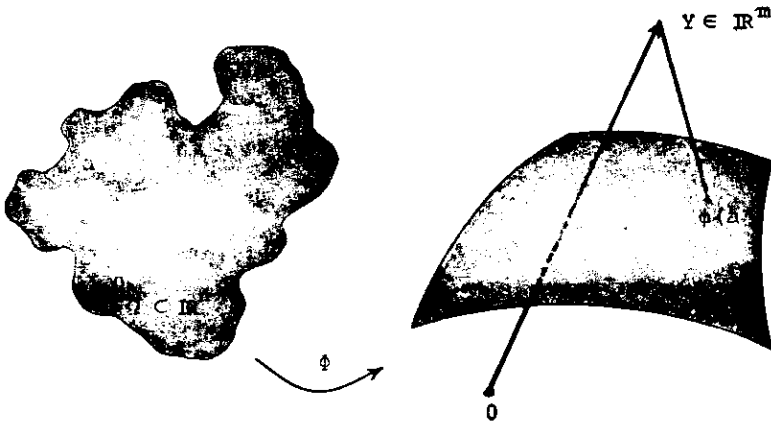
Si definimos la función

$$\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m; \phi(a) = \begin{bmatrix} \phi_1 = \phi(\xi_1; a) \\ \vdots \\ \phi_m = \phi(\xi_m; a) \end{bmatrix}$$

se tiene entonces

$$\min_a r(a) = \min_a \frac{1}{2} \|y - \phi(a)\|^2$$

que puede interpretarse geométricamente como sigue



En este caso la imagen de ϕ es una superficie no plana y nuevamente $\phi(a^*) \in \mathbb{R}^m$ resulta ser el punto en la imagen de ϕ más cercano a y .

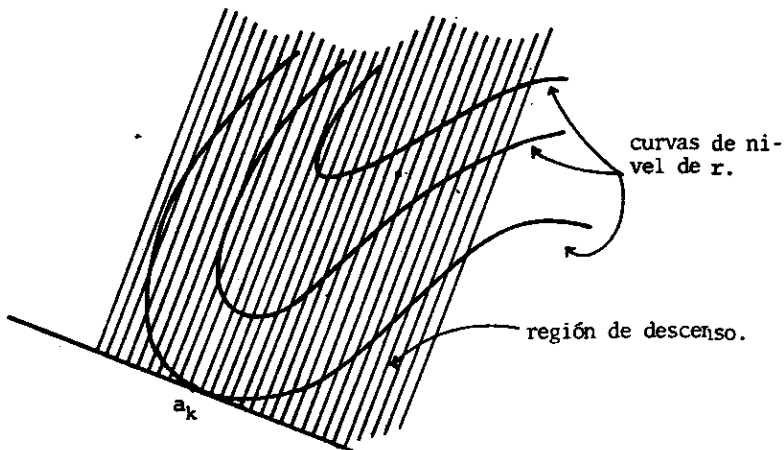
III.- DOS METODOS CLASICOS PARA LA SOLUCION DE PROBLEMAS
NO-LINEALES.

1) Método del descenso más rápido.

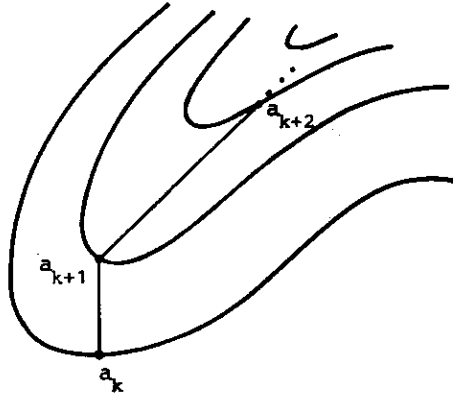
Pertenece éste a una familia de métodos de tipo general para el cálculo del mínimo de una función de varias variables, cuya idea central consiste en que a partir de una propuesta inicial a , de aproximación del mínimo, se genera, vía un proceso iterativo, una sucesión $\{a_n\}$ bajo el criterio de que (en nuestro caso)

$$r(a_{k+1}) < r(a_k)$$

Para lograr esto, una vez obtenida a_k lo que se hace es elegir una dirección de descenso entre todas las posibles -- que existen. (ver figura)



Así, una posibilidad para a_{k+1} , a_{k+2} , sería



En el caso del método del descenso más rápido lo que se hace es elegir la dirección de descenso local más veloz; esto es, se elige como dirección

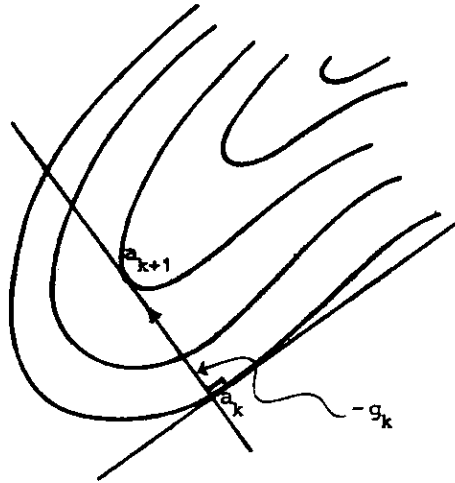
$$-\nabla r(a_k) = -g_k$$

una vez hecha esta elección, lo que se hace es resolver

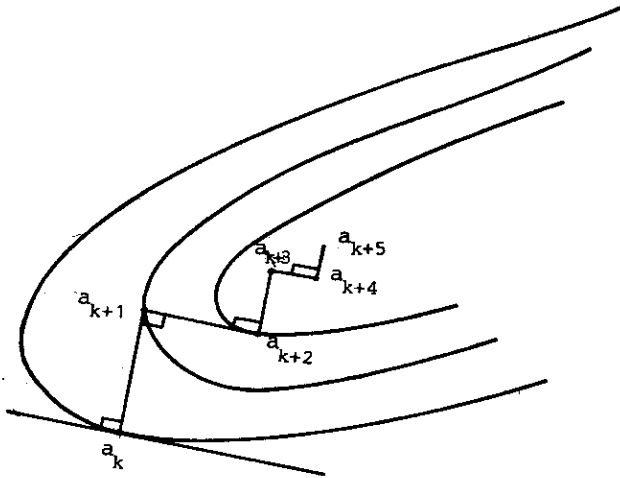
$$(7) \quad \min_{\lambda} r(a_k - \lambda g_k), \quad \lambda \in \mathbb{R}$$

Si llamamos λ_k a la solución de (7) entonces a_{k+1} se obtiene como

$$a_{k+1} = a_k - \lambda_k g_k$$



Como se advierte de la figura anterior, el punto a_{k+1} es aquél donde la recta dada por a_k y g_k es tangente a una de las curvas de nivel de r , de donde resulta que g_k y g_{k+1} son ortogonales para toda k . En base a esto la trayectoria definida por este método es de la forma



En la práctica el mayor problema con este método resulta ser el de una pobre velocidad de convergencia. Las razones que explican esto trataremos de ilustrarlas para el caso en que γ es una forma cuadrática positiva definida.

En efecto, si se tiene que $r(a)$ es del tipo

$$r(a) = a^t A a, \quad A > 0$$

puede mostrarse que cuando $k \rightarrow \infty$ se tiene

$$\|a_{k+1}\| < \rho \|a_k\|$$

donde

$$\rho = \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \quad [2]$$

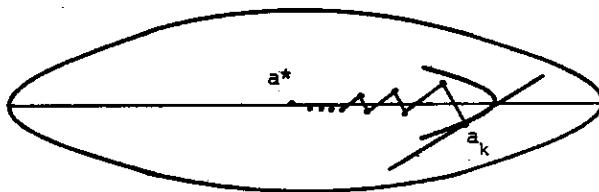
siendo λ_n el mayor eigenvalor de A y λ_1 , el mínimo.

Si escribimos ahora ρ como

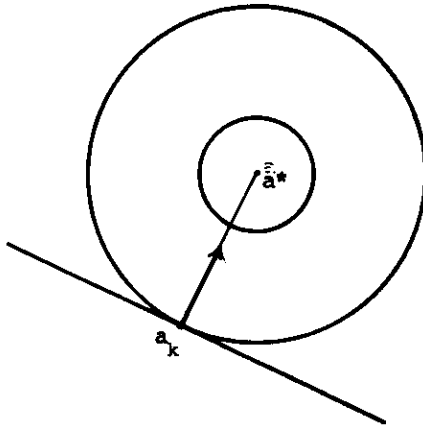
$$\rho = \frac{\frac{\lambda_n}{\lambda_1} - 1}{\frac{\lambda_n}{\lambda_1} + 1} = \frac{k(A) - 1}{k(A) + 1}, \quad k(A) = \frac{\lambda_n}{\lambda_1}$$

esto nos permite analizar la velocidad de convergencia en los siguientes términos: supongamos que $\lambda_n \gg \lambda_1$ ($k(A) \gg 1$); entonces debe suceder que $\rho \approx 1$ lo que implica que la disminución en $r(a_{k+1})$ con respecto a $r(a_k^*)$ será muy pequeña en cada paso y esto significa una pobre velocidad de convergencia. Si por el contrario se tiene $\lambda_n \approx \lambda_1$ ($k(A) \approx 1$) entonces resulta que $\rho \approx 0$ y en este caso se puede garantizar una convergencia veloz.

Geoméricamente, para el caso de dos variables, las curvas de nivel de $r(a)$ cuando $\rho \approx 1$ corresponden a elipses bastante "alargadas", y la trayectoria de las a_k es algo de la siguiente forma:



mientras que si $\rho \approx 0$, se tiende a una situación geométrica como la que se muestra en seguida:



Desgraciadamente, en la práctica el caso $\rho \approx 0$ ($k(A) \approx 1$) no es el más común y esto ayuda en parte a explicar el por qué de la lentitud en la convergencia del método del descenso más rápido.

2) Método de Gauss Newton.

Se trata de un método que, al igual que el anterior, pertenece a la familia que antes mencionamos, y en este caso la estrategia para la construcción de los términos de la sucesión $\{a_n\}$ es la siguiente: supóngase que conocemos a_k . Entonces para obtener a_{k+1} , lo que se hace en primer lugar es aproximar localmente a $\Phi(a)$ como

$$\phi(a) \approx \phi(a_k) + \phi'(a_k)(a-a_k)$$

Usando esta aproximación se tiene para $\gamma(a)$

$$r(a) \approx \frac{1}{2} //Y - \phi(a_k) - \phi'(a_k)(a-a_k) //^2$$

Si hacemos ahora

$$\phi(a_k) = \phi_k$$

$$\phi'(a_k) = \phi'_k$$

$$a-a_k = s$$

y

$$m(s) = \frac{1}{2} //Y - \phi_k - \phi'_k s //^2$$

tendremos

$$r(a) \approx m(s)$$

A continuación la idea consiste en resolver

$$(8) \quad \min_s m(s) = \min_s \left[\frac{1}{2} //Y - \phi_k - \phi'_k s //^2 \right]$$

que es un problema de mínimos cuadrados lineales, cuya solución (llamémosla s_k) es la misma que la solución del sistema

$$(9) \quad [\phi'_k{}^t \phi'_k] s_k = -\phi'_k{}^t (\phi_k - Y)$$

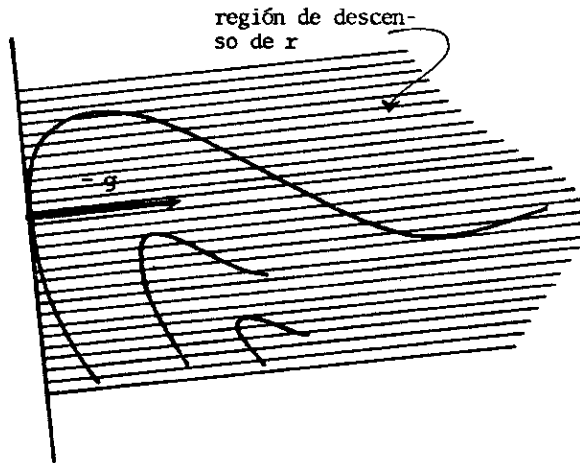
puesto que (9) no es otra cosa que las ecuaciones normales del problema (8). Una vez resuelto (8) a_{k+1} se obtiene como

$$a_{k+1} = a_k + s_k$$

Con respecto a este método, se sabe que no necesariamente converge; sin embargo puede mostrarse que bajo adecuadas -- condiciones (que el residual $\gamma(a)$ sea pequeño) se obtiene convergencia lineal o superlineal. [5]

IV. LOS METODOS LEVENBERG-MARQUARDT-MORRISON.

Una primera cuestión que nos será de gran importancia en esta parte y que se desprende de lo expuesto en la sección anterior es que la región de descenso de γ en a_k se encuentra entre los 0° y $\pm 90^\circ$ con respecto a la dirección $-g$



Ahora bien, en relación a los métodos de Gauss Newton y del descenso más rápido, se ha observado que en una gran cantidad de casos las direcciones $-g$ y s generadas por ellos tienden a ser ortogonales, dándose con frecuencia el caso

$$80^\circ \leq \theta \leq 90^\circ, \quad \theta = \angle(s, -g)$$

lo que geométricamente puede verse como sigue



Y es justamente este hecho el que sugiere en forma natural el tratar de proponer métodos basados en los dos anteriores, buscando en cada paso obtener el mayor descenso posible. Por lo que toca a los métodos Levenberg-Marquardt-Morrison (que en lo sucesivo abreviaremos como L-M-M), de lo que tratan éstos es de explotar al máximo estas ideas basándose en ciertos resultados que serán expuestos más adelante. Por ahora, recordemos que en el caso del método de Gauss-Newton teníamos que en cada paso debía resolverse el problema lineal

$$[\phi_k'^t \phi_k'] s_k = - \phi_k'^t (\phi_k - y)$$

Con respecto a esto, resulta que en la práctica comenzó a observarse que la matriz $\phi_k'^t \phi_k'$ presentaba problemas en el

sentido de ser casi singular debido a la aparición de eigenvalores muy pequeños, por lo que se intentó resolver este problema sumando un cierto número $\lambda > 0$ a la diagonal. Con esto, el sistema lineal a resolver sería entonces

$$(10) \quad [\phi_k'{}^t \phi_k' + \lambda I] s_k = - \phi_k'{}^t (\phi_k - y)$$

Aunque esta idea funcionó en un buen número de ejemplos, la elección de un valor de λ adecuado estaba aún presente y existía una gran confusión con respecto al papel que dicho parámetro jugaba con relación al problema; y no es sino hasta principios de la década de los sesentas que se establece un resultado mediante el cual se demuestra que resolver (10) implica encontrar la solución de $\min_{\underline{s}} m(s)$ en una cierta vecindad de a_k . El resultado en cuestión es el siguiente:

Teorema. Dada $\lambda > 0$ y s_0 solución de

$$[\phi_k'{}^t \phi_k' + \lambda I] s_0 = - \phi_k'{}^t (\phi_k - y)$$

entonces s_0 ($s_0(\lambda)$) minimiza

$$m(s) = \| \phi_k' s + (\phi_k - y) \|^2$$

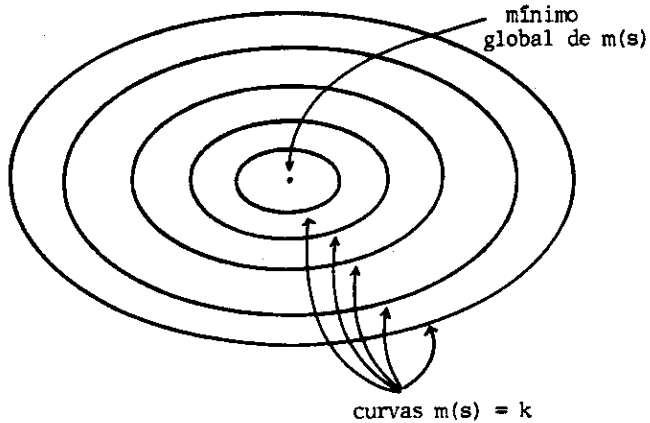
sobre la esfera cuyo radio $\|s\|$ cumple con

$$\|s\|^2 = \|s_0\|^2 \quad (6)$$

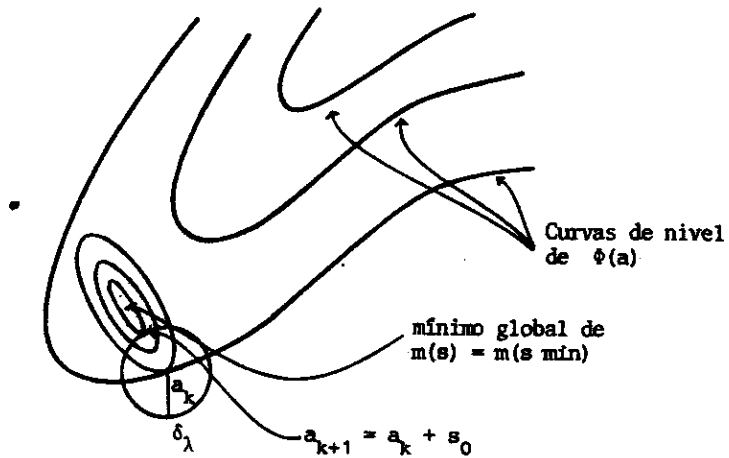
Lo que este teorema nos dice es que si de alguna forma sin llevar a cabo el cálculo de $s_0(\lambda)$ pudiéramos conocer $\|s_0(\lambda)\| = \delta\lambda$, entonces $s_0(\lambda)$ sería la solución de

$$\min_{\|s\| \leq \delta\lambda} m(s).$$

Fijémonos ahora un poco en la función $m(s)$. Dado que ésta es una cuadrática, sus curvas de nivel son hiperelipsoides que para el caso de dos dimensiones adoptan la forma de elipses concéntricas, como se muestra en la siguiente gráfica.

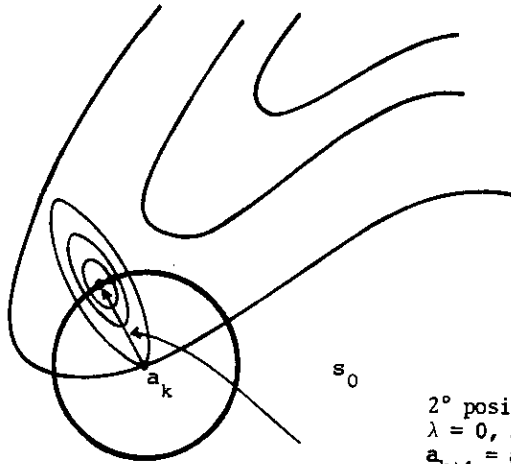


Así, dado el problema $\min_{\|s\| \leq \delta\lambda} m(s)$ se tienen tres posibilidades. La primera de ellas -que ilustramos a continuación- corresponde al caso en que el mínimo global de $m(s)$ queda fuera de la vecindad de radio $\delta\lambda$ de a_k .

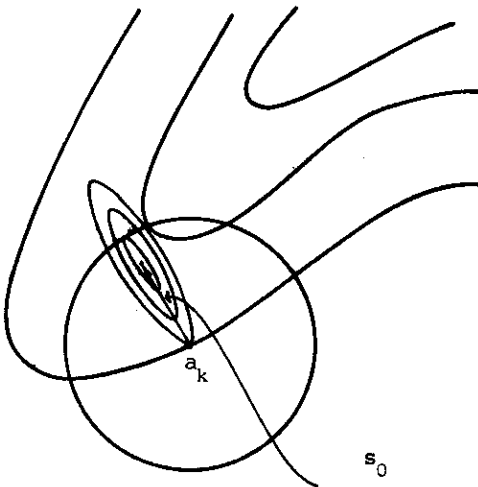


En este caso se tiene: $\lambda > 0$ y $m(s_{\min}) < m(s_0)$.

Las posibilidades segunda y tercera corresponden al valor $\lambda = 0$ y para ambas se tiene que $m(s_{\min}) = m(s_0)$. Geométricamente estos casos equivalen a que el mínimo global de $m(s)$ quede en la frontera o en el interior de la vecindad de a_k de radio δ_λ , según se muestra en seguida.



2° posibilidad:
 $\lambda = 0, s \text{ min} = s_0$
 $a_{k+1} = a_k + s_0$

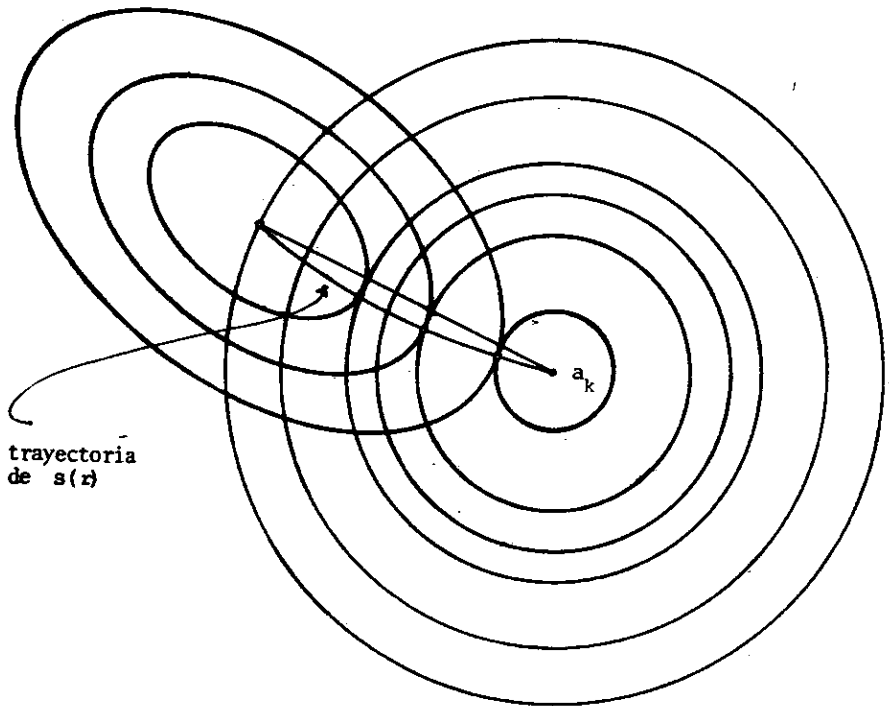


3° posibilidad:
 $\lambda = 0, s \text{ min} = s_0$
 $a_{k+1} = a_k + s_0$

Otra cosa que se tiene es que si dado el problema

$$\min_{\|s\| \leq r} m(s)$$

nos interesamos en pensar a s como función de r ; resulta que la trayectoria de $s(r)$ tiene la forma siguiente



Regresando ahora al teorema anterior, resulta que éste sigue aún sin decirnos nada acerca de cómo elegir un valor de λ en la práctica. Ahora bien, éste es un problema de interés debido a que en ocasiones se tiene algún conocimiento acerca de la magnitud del radio r de la vecindad del punto a_k en la cual la aproximación $m(s)$ se sabe que es buena. En otras palabras, resulta que en un buen número de casos el problema consistiría en poder, a partir de un radio r dado, determinar la λ correspondiente a dicho valor. Sin embargo, éste es en general un problema difícil desde el punto de vista práctico debido a que si dado r se quiere resolver

$$(11) \quad \min_{\|s\|^2 = r^2} m(s)$$

(que es el caso que en realidad nos interesa), se tiene que una condición necesaria para tener un punto estacionario de (11) está dada por

$$0 = - [\phi_k'^t (y - \phi_k) - \phi_k'^t \phi_k' s] + \lambda s$$

$$0 = \|s\|^2 - r^2$$

que es un sistema lineal en s y no-lineal en λ . Dadas las dificultades que implica el tratar de resolverlo, lo que se ha hecho es sugerir distintas estrategias para la elección de un valor de λ en la práctica. Estas estrategias tienen como sustento ciertos hechos que respecto a la forma como varía

s en términos de dicho parámetro, los cuales serán vistos a continuación.

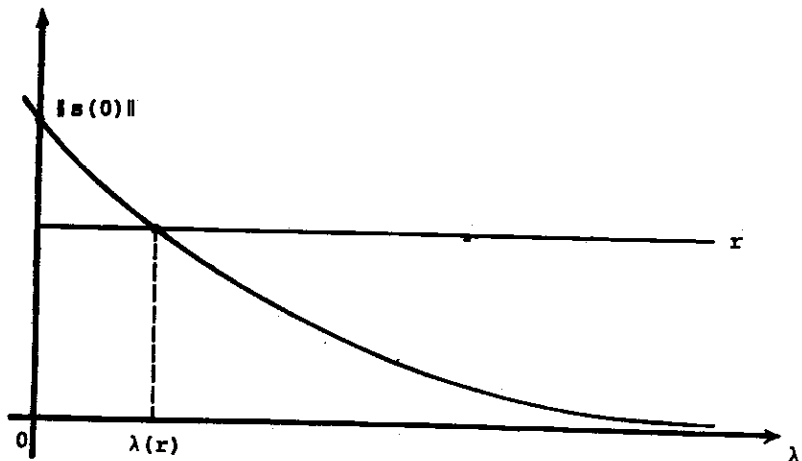
Teorema. Sea $s(\lambda)$ solución de

$$(\phi^t \phi^{t-1} + \lambda I)s = \phi^t(\phi - y)$$

para un valor dado de λ . Entonces, $\|s(\lambda)\|^2$ es una función continua y decreciente de λ ; esto es,

$$\begin{aligned} \|s(\lambda)\|^2 &\rightarrow 0 \\ \lambda &\rightarrow \infty. \end{aligned} \quad [6]$$

En este contexto, el problema de, dado r determinar λ equivaldría en términos geométricos a encontrar la intersección que se muestra en seguida.



Teorema. Sea θ el ángulo entre s y $-g$. Entonces θ es una función continua, monótonamente decreciente de λ ; esto es,

$$\theta \rightarrow 0$$

$$\lambda \rightarrow \infty$$

La idea de la demostración de este resultado consiste en hacer ver que $\cos \theta$ se puede escribir como

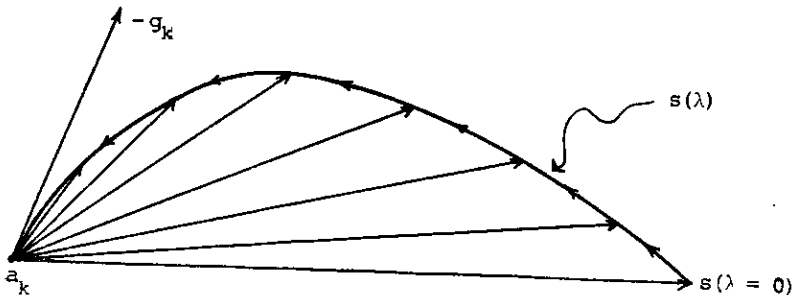
$$\cos \theta = \frac{\sum_{j=1}^m \frac{\xi_j^2}{\delta_j + \lambda}}{\left(\sum_{j=1}^m \frac{\xi_j^2}{(\delta_j + \lambda)^2} \right)^{1/2} \left([\phi^t(\phi - y)]^t [\phi^t(\phi - y)] \right)^{1/2}}$$

donde los δ_j son los eigenvalores de la matriz $\phi^t \phi$. Una vez establecido esto, todo se reduce a hacer ver que

$$\cos \theta \rightarrow 1$$

$$\lambda \rightarrow \infty.$$

Estos resultados se pueden interpretar en forma geométrica de manera conjunta como sigue



La importancia de los teoremas anteriores radica en que nos permiten construir algoritmos cuya iteración general sea de la forma

$$a_{k+1} = a_k + s_k(\lambda_k)$$

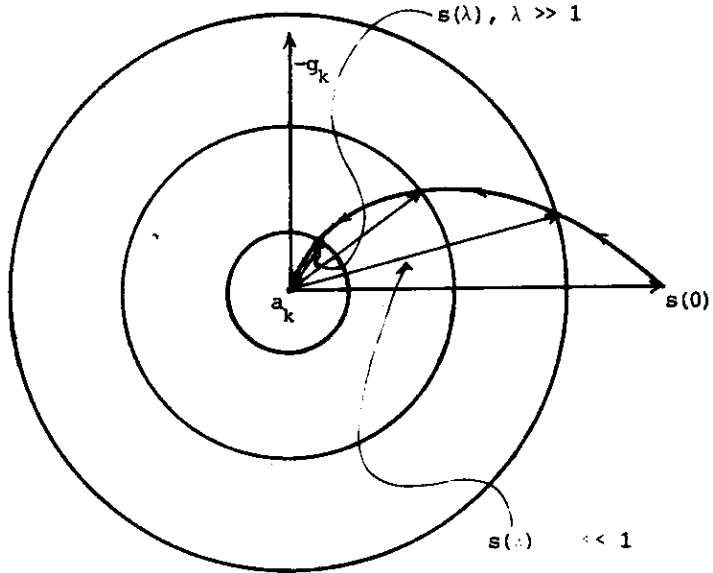
con la propiedad de que

$$(12) \quad r(a_{k+1}) \leq r(a_k)$$

para λ_k adecuada. La razón de esto es que dichos resultados nos aseguran que debe existir alguna λ_k (tal vez bastante -- grande) con la propiedad de que (12) se satisfaga, ya que al crecer λ nos acercamos cada vez más a $-g$ que es una dirección de descenso.

En la práctica resulta interesante el que la elección de λ en cada paso puede hacerse sin necesidad de usar el valor de r . Una estrategia en este sentido es la propuesta por D.W. Marquardt, la cual consiste en que en cada paso se trata de tener la mayor vecindad posible en la que la aproximación $m(s)$ sea buena. Para ello lo que propone Marquardt es comenzar con un valor de λ pequeño, lo que no sólo garantiza una vecindad lo más grande posible sino que además fija una dirección cercana a la de Gauss-Newton, conocida como de descenso veloz (cuando funciona). Si resultara que con la λ propuesta no se consiguiera que $r(a_{k+1}) \leq r(a_k)$ lo que debe hacerse entonces es usar valores de λ cada vez mayores hasta lograr una

disminución de $r(a)$ Es claro que el tomar valores grandes de λ implica que la venciad de a_k será cada vez menor, lo que debe traducirse en un avance más lento cada vez (ver figura)



V SOBRE LA CONVERGENCIA DE LOS METODOS
LEVENBERG - MARQUARDT MORRISON.

Para empezar permitasenos introducir algo de notación.
En primer lugar definiremos la función

$$f: \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$f(a) = \begin{bmatrix} \eta_1 - \phi_1 \\ \vdots \\ \eta_m - \phi_m \end{bmatrix}$$

Con esto nuestro problema $\min_a \gamma(a)$ lo podemos escribir como

$$\min_a \frac{1}{2} \|f(a)\|^2$$

Además, el sistema lineal a resolver en cada iteración para el caso de los métodos L-M-M resulta ser

$$(13) \quad [f'_k{}^t f'_k + \lambda_k I] s_k = -f'_k{}^t f_k; \quad f'_k = f'_k(a_k)$$

cuya solución s_k es la misma que la del problema de mínimos cuadrados lineales

$$(14) \quad \min_s \left\| \begin{bmatrix} f'_k \\ \sqrt{\lambda_k} \end{bmatrix} s + \begin{bmatrix} f_k \\ 0 \end{bmatrix} \right\|^2$$

dado que (13) es justamente el sistema de ecuaciones normales de (14).

Por último definiremos la función

$$R(a, \lambda) = \left(\frac{f'_k}{\sqrt{\lambda} I} \right) s + \left(\frac{f_k}{0} \right), \quad s = a - a_k$$

En base a estos elementos, pasaremos primeramente a ocuparnos del problema de la convergencia para una cierta clase de algoritmos en los que la aproximación a_{k+1} se obtiene como

$$(15) \quad a_{k+1} = a_k + \gamma_k s_k$$

donde γ_k es un parámetro cuya elección se hace de tal manera que minimice

$$\| f(a_k + \gamma_k s_k) \|$$

El esquema general de estos algoritmos es:

i) Resolver

$$[f'_k{}^t f'_k + \lambda_k I] s_k = - f'_k{}^t f_k$$

ii) Determinar λ_k con la propiedad de que:

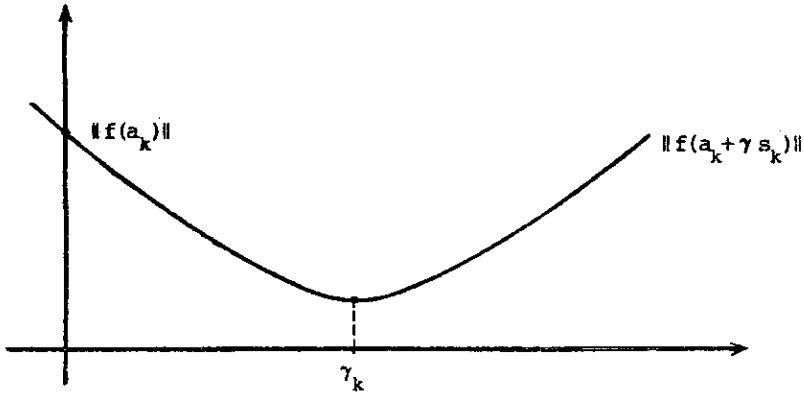
$$\min_{\lambda} \| f(a_k + \gamma s_k) \| = \| f(a_k + \gamma_k s_k) \| .$$

iii) Hacer

$$a_{k+1} = a_k + \gamma_k s_k$$

iv) Si a_{k+1} es ya una buena aproximación a a^* , terminar
Si no, repetir nuevamente el proceso.

Como se observa, el segundo paso del algoritmo lo que nos dice es que deberá buscarse el mínimo de $\|f(a_k + \gamma s_k)\|$ en la dirección del s_k obtenido en el paso uno. Este proceso es conocido como búsqueda lineal y equivale gráficamente a lo siguiente



Con respecto a este punto resulta importante hacer notar que este tipo de algoritmos pueden resultar costosos debido a un posible gran número de evaluaciones de f durante el proceso de búsqueda lineal, y es por esta razón que resulta de interés poder determinar los casos en los que se da $\gamma = 1$; es to es, los casos en que

$$a_{k+1} = a_k + s_k$$

Este tipo de iteración da lugar al llamado método de paso completo.

Las condiciones bajo las cuales (15) converge son:

a) Las a_k deben cumplir con

$$a_k \in \Omega \subset \mathbb{R}, \Omega \text{ acotado.}$$

b) $f(a)$ debe ser suficientemente lisa, de tal forma - que tenga sentido escribir

$$f(a+\gamma s) = f(a) + \gamma f'(a)s + \gamma^2 \|s\|^2 \omega(a, \gamma, s),$$

con $\|\omega\| \leq k$ para toda $a \in \Omega$. Esto es, f debe al menos cumplir $f \in C^2(\Omega)$ y tener segundas derivadas acotadas. [7]

Un primer resultado sobre convergencia es:

Teorema.- Si λ_k se escoge de tal forma que el eigenvalor δ_k más pequeño de la matriz

$$[f_k^t f_k + \lambda_k I]$$

satisfaga que $\delta_k > \delta > 0$, entonces la sucesión $\{\|f_k\|\}$ converge y además, los puntos límites de la sucesión $\{a_k\}$ resultan ser puntos estacionarios de $\|f(a)\|$ [7]

Las partes centrales de la demostración son:

a) Mostrar que a partir de la relación

$$f(a+\gamma s) = f(a) + \gamma f'(a)s + \gamma^2 \|s\| \omega(a, \gamma, s)$$

se puede probar la desigualdad

$$\|f(a_k + \gamma s_k)\| \leq Q_k(\gamma) = (1-\gamma) \|f(a_k)\| + \gamma \|R_k^{(1)}(a, \lambda)\| + \gamma^2 \|s_k\|^2 \omega$$

donde $R_k^{(1)}(a, \lambda)$ contiene las primeras m componentes de $R_k(a, \lambda)$

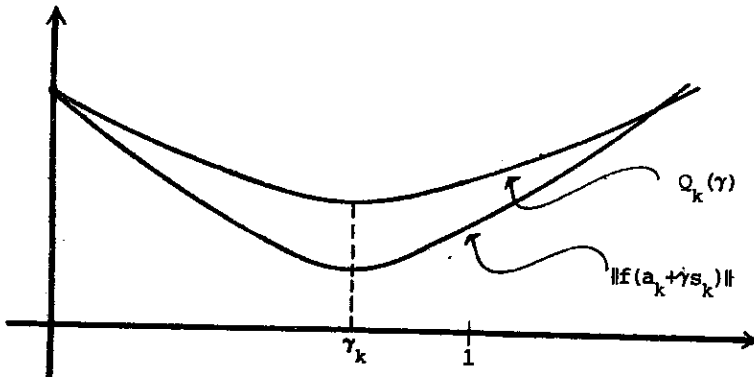
b) Hacer ver que si a_{k+1} se converge como

$$a_{k+1} = a_k + \gamma_k s_k$$

con γ_k tal que minimice $\|f(a_k + \gamma s_k)\|$, entonces se cumple

$$\| \gamma(a_{k+1}) \| \leq \min_{0 \leq \gamma \leq 1} Q_k(\gamma)$$

(ver figura)



c) A partir de lo anterior, es posible encontrar cotas para $\hat{\gamma}_k$ en términos del mínimo de $Q_k(\gamma)$ en $[0,1]$. Si llamamos γ_k^* a dicho mínimo entonces se tiene que

$$\gamma_k^* = \min \{1, \hat{\gamma}_k\}$$

donde $\hat{\gamma}_k$ es el mínimo global de $Q_k(\gamma)$. Se puede demostrar además que existe $\mu(\delta) > 0$ tal que

$$\gamma_k^* > \mu(\delta) > 0.$$

d) Finalmente se puede probar que

$$\|f(a_{k+1})\| \leq \|f(a_k)\| - \frac{\gamma_k^*}{2} (\|f(a_k)\| - \|R_k^{(1)}\|)$$

Este resultado implica que la sucesión $\{\|f(a_k)\|\}$ es monótona decreciente y por lo tanto convergente. En forma análoga se demuestra que los puntos límites de la sucesión $\{a_k\}$ son puntos estacionarios de $\|f(a)\|$.

En la práctica se ha observado que el caso $\gamma_k^* = 1$ es usual, con lo que se garantiza en general la convergencia de los métodos L-M-M. Más aún, se puede demostrar que en este caso la convergencia es lineal; es decir, existe ϵ tal que

$$\|a_{k+1} - a^*\| \leq \epsilon \|a_k - a^*\| ; 0 < \epsilon < 1.$$

En la actualidad la teoría de convergencia para los métodos L-M-M ha hecho progresos considerables y ha podido abor- darse sin necesidad de llevar a cabo la minimización de $\|f(a_k + \gamma s_k)\|$ sino trabajando directamente sobre la itera- ción

$$a_{k+1} = a_k + s_k$$

Por lo que toca al problema del ajuste de la longitud del paso en cada etapa, éste es realiza en función exclusiva- mente de λ , conocido como parámetro de Marquardt o de amorti- guamiento. Una de las ideas al respecto consiste en cons- truir una función mediante la cual se tenga un criterio para la selección de λ_k . Una propuesta en este sentido es tomar la función

$$g(a_k, \lambda_k) = \frac{\|f(a_k)\|^2 - \|f(a_{k+1})\|^2}{2(\|f(a_k)\|^2 - \|R(a_k, \lambda_k)\|^2)}$$

y dado que deseamos tener $\|f(a_{k+1})\|^2 < \|f(a_k)\|^2$ y como $\|R(a_k, \lambda_k)\|^2 \approx \|f(a_{k+1})\|^2$, esto sugiere pedir que g cumpla con

$$0 < g(a_k, \lambda_k) \leq 1$$

de donde, si podemos escoger los $\{\lambda_k\}$ tales que se tenga

$$0 < \sigma \leq g(a_k, \lambda_k)$$

entonces podremos asegurar la convergencia de nuestro algoritmo. Esto se establece formalmente en el siguiente

Teorema.- Si es posible escoger $\{\lambda_k\}$ acotada de tal forma que se cumpla

$$0 < \sigma \leq g(a_k, \lambda_k)$$

donde σ es una constante independiente de k , y si

$$a_{k+1} = a_k + s_k, \quad g(a_k, \lambda_k) = \frac{\|f(a_k)\|^2 - \|f(a_{k+1})\|^2}{2(\|f(a_k)\|^2 - \|R(a_k, \lambda_k)\|^2)}$$

entonces la sucesión $\{\|f(a_k)\|^2\}$ es convergente y los puntos límites de la sucesión $\{a_k\}$ son puntos estacionarios de $\|f(a)\|^2$ [8]

Nótese que aquí el papel de la desigualdad

$$0 < \sigma \leq g(a_k, \lambda_k)$$

es de una gran importancia ya que de hecho nos proporciona un criterio para la elección de un λ_k adecuado en términos de cantidades conocidas. Además, la elección de un valor de σ pequeño puede ayudar en el sentido de que el valor de λ_k sea también pequeño, lo que, como ya se ha visto, nos sitúa en una dirección cercana a la de Gauss-Newton. Que esto es al menos probable puede verse como consecuencia del resultado que sigue:

En la actualidad la teoría de convergencia para los métodos L-M-M ha hecho progresos considerables y ha podido abor- darse sin necesidad de llevar a cabo la minimización de $\|f(a_k + \gamma s_k)\|$ sino trabajando directamente sobre la iteración.

$$a_{k+1} = a_k + s_k$$

Por lo que toca al problema del ajuste de la longitud del paso en cada etapa, éste se realiza en función exclusivamente de λ , conocido como parámetro de Marquardt o de amortiguamiento. Una de las ideas al respecto consiste en construir una función mediante la cual se tenga un criterio para la selección de λ_k . Una propuesta en este sentido es tomar la función

$$g(a_k, \lambda_k) = \frac{\|f(a_k)\|^2 - \|f(a_{k+1})\|^2}{2(\|f(a_k)\|^2 - \|R(a_k, \lambda_k)\|^2)}$$

y dado que deseamos tener $\|f(a_{k+1})\|^2 < \|f(a_k)\|^2$ y como $\|R(a_k, \lambda_k)\|^2 \approx \|f(a_{k+1})\|^2$, esto sugiere pedir que g cumpla con

$$0 < g(a_k, \lambda_k) \leq 1$$

de donde, si podemos escoger los $\{\lambda_k\}$ tales que se tenga

$$0 < \sigma \leq g(a_k, \lambda_k)$$

Teorema.- Para cada $\sigma < 1$ y cada s_k , existe una λ_k tal que

$$(16) \quad 0 < \sigma \leq g(a_k, \lambda_k)$$

se satisface. [8]

La idea de la demostración en este caso consiste en probar que

$$\begin{aligned} g(a_k, \lambda) &\rightarrow 1 + 0 \quad (1/\gamma^2) \\ \lambda &\rightarrow \infty \end{aligned}$$

de donde resulta que una elección de σ pequeño favorece el que (16) se satisfaga y esto puede ayudar a que la λ correspondiente sea también pequeña.

VI.- COMENTARIO FINAL.

Resulta interesante hacer notar el hecho de que en buena parte, la teoría existente con respecto a los métodos L-M-M ha ido surgiendo con posterioridad a la experiencia y como una necesidad natural de dotar a ésta de un marco adecuado en el que por una parte puedan ser explicados los resultados empíricos y por otra, pueda contarse con elementos que permitan la elaboración de estrategias cada vez mejores en la práctica. Respecto de esto, existen actualmente diversos algoritmos e implementaciones en los que se busca optimizar tanto el espacio de trabajo como el tiempo de cómputo. En la bibliografía se mencionan algunos trabajos de interés en este sentido.

VII.-

BIBLIOGRAFIA

- [1] JENNINGS, L.S. Applications of Orthogonal Matrix Transformations to the Solution of Systems of Linear and Nonlinear Equations, Australian Nat. Univ., Computer Centre Tech. Rept. No. 37, Canberra, Australia, 1970.
- [2] KOWALIK, J & OSBORNE, M.R., Methods for Unconstrained Optimization Problems, American Elsevier Publishing Co. Inc., 1968.
- [3] LAWSON, CH.J. & HANSON R.J , Solving Least Squares Problems, Prentice Hall Inc., 1974.
- [4] LEVENBERG, K., A Method for the Solution of Certain Non-Linear Problems in Least Squares, Quart. Appl. Math., 2, 1944.
- [5] LOPEZ E. JOSE Y BARRERA S.PABLO Un Enfoque Geométrico al Problema Clásico de Estimación de Parámetros, Comunicación Interna No. 10, Depto. de Matemáticas de la Fac. de Ciencias de la U.N.A.M., 1977.

- [6] MARQUARDT, D.W An Algorithm for Least Squares Estimation of Nonlinear Parameters, J. Soc. Indust. Appl. Math., 11, 1963.
- [7] OSBORNE, M.R., Some Aspects of Nonlinear Least Squares Calculations, in Numerical Methods for Nonlinear Optimization, Academic Press, 1972.
- [8] -----, Nonlinear Least Squares — The Levenberg Algorithm Revisited, J Aust. Math. Soc., 19, 1976.

Mecanografiado: Mirna Ocampo Fernández

Dibujos: María Rosas Torres

APENDICE.

VIII.

a) En relacion al problema

$$\min_a \left\{ \frac{1}{2} \sum_{i=1}^m (\eta_i - (\alpha_1 + \alpha_2 e^{-\alpha_4 \xi_i} + \alpha_3 e^{-\alpha_5 \xi_i}))^2 \right\}$$

enunciado en la página 11, el sistema no-lineal a que da lugar es:

$$\sum_{i=1}^m \alpha_1 + \alpha_2 \sum_{i=1}^m e^{-\alpha_4 \xi_i} + \alpha_3 \sum_{i=1}^m e^{-\alpha_5 \xi_i} = \sum_{i=1}^n \eta_i$$

$$\alpha_1 \sum_{i=1}^m e^{-\alpha_4 \xi_i} + \alpha_2 \sum_{i=1}^m e^{-2\alpha_4 \xi_i} + \alpha_3 \sum_{i=1}^m e^{-(\alpha_5 \xi_i + \alpha_4 \xi_i)} = \sum_{i=1}^m \eta_i e^{-\alpha_4 \xi_i}$$

$$\alpha_1 \sum_{i=1}^m e^{-\alpha_5 \xi_i} + \alpha_2 \sum_{i=1}^m e^{-(\alpha_4 \xi_i + \alpha_5 \xi_i)} + \alpha_3 \sum_{i=1}^m e^{-2\alpha_5 \xi_i} = \sum_{i=1}^m \eta_i e^{-\alpha_5 \xi_i}$$

$$\alpha_1 \alpha_2 \sum_{i=1}^m \xi_i e^{-\alpha_4 \xi_i} + \alpha_2^2 \sum_{i=1}^m \xi_i e^{-2\alpha_4 \xi_i} + \alpha_2 \alpha_3 \sum_{i=1}^m \xi_i e^{-(\alpha_4 \xi_i + \alpha_5 \xi_i)} = \alpha_2 \sum_{i=1}^m \eta_i \xi_i e^{-\alpha_4 \xi_i}$$

$$\alpha_1 \alpha_3 \sum_{i=1}^m \xi_i e^{-\alpha_5 \xi_i} + \alpha_2 \alpha_3 \sum_{i=1}^m \xi_i e^{-(\alpha_4 \xi_i + \alpha_5 \xi_i)} + \alpha_3^2 \sum_{i=1}^m \xi_i e^{-2\alpha_5 \xi_i} = \alpha_3 \sum_{i=1}^m \eta_i \xi_i e^{-\alpha_5 \xi_i}$$

b) Por lo que toca al problema

$$\min_a \left[\frac{1}{2} \sum_{i=1}^m (\eta_i - \left(\frac{\alpha_1}{1 + \alpha_2 (\xi_i - \alpha_3)} + \frac{\alpha_4}{1 + \alpha_5 (\xi_i - \alpha_6)} \right))^2 \right]$$

el sistema no-lineal asociado en este caso es:

$$\alpha_1 \sum_{i=1}^m \frac{1}{(1+\alpha_2(\xi_1-\alpha_2))^2} + \alpha_4 \sum_{i=1}^m \frac{1}{(1+\alpha_2(\xi_1-\alpha_2))(1+\alpha_5(\xi_1-\alpha_6))} = \sum_{i=1}^m \frac{\eta_1}{1+\alpha_2(\xi_1-\alpha_2)}$$

$$\alpha_1^2 \sum_{i=1}^m \frac{(\xi_1 - \alpha_2)}{(1+\alpha_2(\xi_1-\alpha_2))^2} + \alpha_1 \alpha_4 \sum_{i=1}^m \frac{(\xi_1 - \alpha_2)}{(1+\alpha_2(\xi_1+\alpha_2))(1+\alpha_5(\xi_1-\alpha_6))} = \alpha_1 \sum_{i=1}^m \frac{\eta_1 (\xi_1 - \alpha_2)}{(1+\alpha_2(\xi_1-\alpha_2))^2}$$

$$\alpha_1^2 \alpha_2 \sum_{i=1}^m \frac{1}{(1+\alpha_2(\xi_1-\alpha_2))^2} + \alpha_1 \alpha_4 \sum_{i=1}^m \frac{1}{(1+\alpha_2(\xi_1-\alpha_2))^2 (1+\alpha_5(\xi_1-\alpha_6))} = \alpha_1 \alpha_2 \sum_{i=1}^m \frac{\eta_1}{(1+\alpha_2(\xi_1-\alpha_2))^2}$$

$$\alpha_1 \sum_{i=1}^m \frac{1}{(1+\alpha_2(\xi_1-\alpha_2))(1+\alpha_5(\xi_1-\alpha_6))} + \alpha_4 \sum_{i=1}^m \frac{1}{(1+\alpha_5(\xi_1-\alpha_6))^2} = \sum_{i=1}^m \frac{\eta_1}{1+\alpha_5(\xi_1-\alpha_6)}$$

$$\alpha_1 \alpha_4 \sum_{i=1}^m \frac{(\xi_1 - \alpha_6)}{(1+\alpha_2(\xi_1-\alpha_2))(1+\alpha_5(\xi_1-\alpha_6))} + \alpha_4^2 \sum_{i=1}^m \frac{(\xi_1 - \alpha_6)}{(1+\alpha_5(\xi_1-\alpha_6))^2} = \alpha_4 \sum_{i=1}^m \frac{\eta_1 (\xi_1 - \alpha_6)}{(1+\alpha_5(\xi_1-\alpha_6))^2}$$

$$\alpha_1 \alpha_4 \alpha_5 \sum_{i=1}^m \frac{1}{(1+\alpha_2(\xi_1-\alpha_2))(1+\alpha_5(\xi_1-\alpha_6))} + \alpha_4^2 \alpha_5 \sum_{i=1}^m \frac{1}{(1+\alpha_5(\xi_1-\alpha_6))^2} = \alpha_4 \alpha_5 \sum_{i=1}^m \frac{\eta_1}{(1+\alpha_5(\xi_1-\alpha_6))^2}$$

o