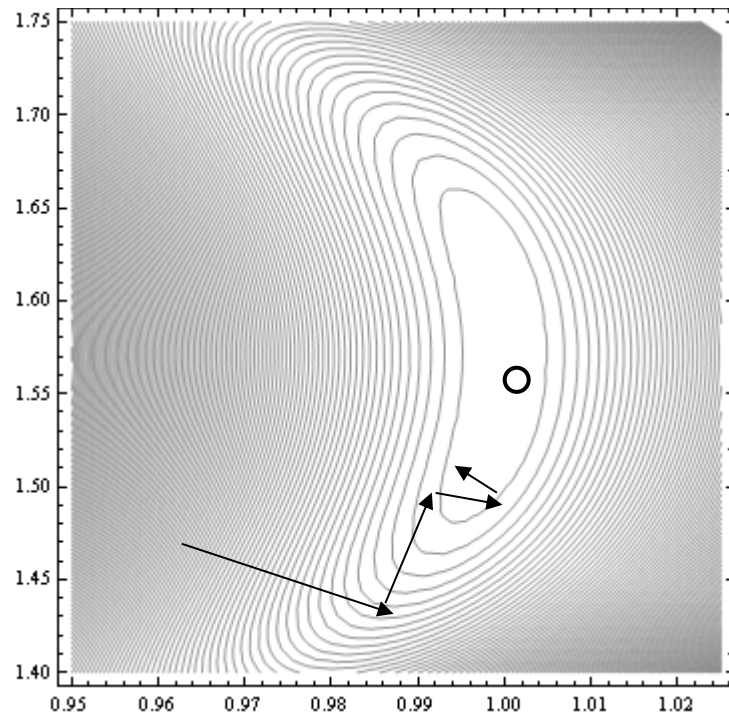


CAPITULO IX

SISTEMAS DE ECUACIONES NO LINEALES



EXORDIO

Los sistemas de ecuaciones no lineales son pesados y complejos, requieren un volumen importante de cálculo y el éxito depende tanto del método elegido como de los problemas numéricos involucrados y la habilidad del analista.

Descartados para este tema los denominados métodos exactos, que luego de un determinado número de pasos llevan a la solución, salvo alguna afortunada inspiración y a la aplicación oportuna de una triquiñuela apropiada; sólo quedan disponibles métodos aproximados, iterativos, que aproximan la solución hasta que ciertas condiciones quedan satisfechas.

Con ellos hay que operar.

Algunos son similares a los aplicados para la búsqueda de raíces de ecuaciones con una incógnita. La similitud es conceptual puesto que, en estos sistemas, el cálculo debe hacerse en espacios de n dimensiones, con todo lo que ello implica.

Se han consignado las similitudes correspondientes.

Se presentan en este trabajo los métodos de iteración simple o del punto fijo, el de Raphson Newton y el del gradiente o del descenso más rápido.

Este último, conceptualmente simple, es a veces operativamente inabordable, motivo por el cual no debe extrañar la cantidad de trabajo de investigación y comentarios que existen sobre el mismo en la bibliografía especializada.

CASOS

ASTRONOMÍA

Sistema Local

La esfera celeste es la superficie esférica en la que se proyectan todos los astros del firmamento, a la vista del observador situado en un determinado punto de la superficie terrestre.

Para ubicar uno de esos astros, naturalmente se utilizan coordenadas. Para ello se definen planos que permiten medir, con alta precisión, ángulos que determinan la posición del astro en cuestión en la esfera celeste.

Uno de esos planos es el plano horizontal que pasa por la vertical del lugar en que está situado el observador. Este plano es denominado el **horizonte astronómico local**.

El otro es un plano que contiene a la recta que une el zenit con el nadir, pasando por el centro de la esfera celeste y por el astro cuya ubicación se necesita.

El azimut A se mide sobre el horizonte astronómico local, desde el Sur, hasta el plano que pasa por el astro. Su rango es, naturalmente $0^\circ - 360^\circ$.

La altura h es medida desde el plano del horizonte astronómico hasta el astro en cuestión. Su rango es $0^\circ - 90^\circ$, positivo si se mide hacia el zenit y negativo si se mide hacia el nadir.

Obviamente este sistema es local, depende de la posición del observador razón por la cual los catálogos internacionales utilizan otro sistema de coordenadas, denominado **sistema ecuatorial celeste**.

Sistema Ecuatorial Celeste

El plano fundamental es el **ecuador celeste** que es la extensión del ecuador terrestre hasta su intersección con la esfera celeste.

El otro plano es el determinado por el eje polar celeste, norte - sur y el astro cuyas coordenadas se buscan.

Sobre el ecuador celeste se mide la denominada ascensión recta H medida de 0° a 360° desde el llamado punto vernal - intersección del ecuador celeste con la eclíptica- en el sentido NOSE, sigla de Norte, Oeste, Sur, Este.

Para recordar, la eclíptica es la órbita aparente que hace el sol en torno a la tierra, tiene una inclinación de $\varphi = 23^\circ 27'$ con respecto al ecuador terrestre.

La declinación δ es el ángulo medido en el plano determinado por el eje polar celeste y el astro. Se lo mide desde el ecuador celeste hasta el astro, con rango $0^\circ - 90^\circ$ positivo hacia el norte y negativo hacia el sur

Transformación de Coordenadas

La transformación de las coordenadas obtenidas en el sistema local al sistema ecuatorial celeste o viceversa es un problema que requiere la aplicación de conceptos de trigonometría plana y esférica, llegándose luego de una larga deducción al siguiente sistema de ecuaciones no lineales.

$$\begin{cases} \operatorname{sen} H \cos \delta = \operatorname{sen} A \cos h \\ \operatorname{sen} \delta = \operatorname{sen} h \operatorname{sen} \varphi - \cos \varphi \cos h \cos A \\ \cos \delta \cos H = \operatorname{sen} h \cos \varphi + \cos h \operatorname{sen} \varphi \cos A \end{cases}$$

Conste, el agradecimiento del autor a las autoras del trabajo *Sistemas de Ecuaciones no Lineales en Astronomía*, María Inés Ciancio, Elisa Silvia Oliva, Sol Molina. Universidad Nacional de San Juan- Facultad de Ciencias Exactas, Físicas y Naturales que inspiró esta aplicación.

POSICIONAMIENTO

Triangulación de señales GPS

Un aparato, como un teléfono celular, recibe señales de un satélite de la constelación GPS cuyas posiciones orbitales alrededor de la tierra son conocidas. La señal de un satélite permite al teléfono conocer que el mismo está en algún lugar sobre una línea circular, la de la órbita que recorre el satélite captado.

Otras señales provenientes de otros satélites del sistema identifican otra órbita circular cuyas intersecciones con la primera ajustan la posición del teléfono que las recibe.

Obviamente se trata de ecuaciones no lineales.

Tres o más señales reducen la solución del sistema no lineal a un punto, el de la ubicación del teléfono.

CONSERVACION DE LA ENERGÍA

La conservación de la energía mecánica puede producir un sistema de ecuaciones no lineales cuando existe un choque elástico. La energía cinética de los objetos depende de sus respectivas velocidades al cuadrado mientras que el impulso depende directamente de dichas velocidades. Las velocidades resultantes de los cuerpos después del choque, sin variación de masa, se resuelve mediante un sistema de ecuaciones no lineales.

DISEÑO

El diseño y la manufactura por computadora a través de herramientas específicas requiere la utilización de un software que permita crear formas tridimensionales mediante la intersección de curvas y superficies alabeadas.

Para visualizar y, por supuesto formular sus respectivos planes de construcción a través de equipos electrónicamente controlados, requiere que el antes mencionado software cuente con la capacidad de resolver complejos sistemas de ecuaciones no lineales.

Es posible que el usuario de este tipo de sistema de diseño no tenga acceso a los algoritmos que resuelven ese tipo de sistemas, pero definitivamente ellos están en el software que utiliza.

¿CUANDO NO! EULER

Se llaman ecuaciones diofánticas aquellas cuya solución se busca en números enteros.

Fermat no pudo resolver, en el siglo XVII, en enteros, la ecuación $61x^2 + 1 = y^2$. Euler menos de 100 años después lo hizo, obteniendo como solución $x = 226.153.980$ e $y = 1.766.319.049$. (3.119.882.982.860.264.401)

¿Cómo lo hizo? Es una buena pregunta.

¿Hace falta decir que esta es una ecuación no lineal?

I INTRODUCCION

1 Se considera la búsqueda de la solución de un sistema de ecuaciones no lineales del tipo

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ \dots\dots\dots \\ f_n(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

donde la f_i son funciones reales. En todo lo siguiente se supondrá, como hipótesis, que dichas funciones son diferenciables en la medida de lo requerido por los métodos que se exponen.

2 Desde el punto de vista vectorial, el sistema anterior puede escribirse

$$\vec{f} = \begin{vmatrix} f_1 \\ f_2 \\ \dots \\ f_n \end{vmatrix} \quad \vec{x} = \begin{vmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{vmatrix}$$

con lo cual el sistema anterior se escribe

$$\vec{f}(\vec{x}) = 0$$

3 También se supone que ha sido encontrada una aproximación inicial a la raíz buscada, que se denominará aproximación de orden 0

$$\vec{x}^{(0)} = \begin{vmatrix} x_1^{(0)} \\ x_2^{(0)} \\ \dots \\ x_n^{(0)} \end{vmatrix}$$

y que se tomará como punto de partida para los métodos que a continuación se presentan.

4 El lector debe estar al tanto de las denominadas **normas** de vectores y matrices puesto que el problema que en una variable se soluciona con el **valor absoluto** debe ser tratado mediante normas en espacios vectoriales de n dimensiones para vectores y matrices.

5 Estas normas, que usualmente se representan mediante $\|\cdot\|_*$ donde las dobles barras indican norma, el símbolo \circ indica el nombre del vector o la matriz y el subíndice $*$ la norma específica que se utiliza en el momento.

6 Se recuerda que se denomina norma sobre \mathbb{R}^n a toda aplicación definida en \mathbb{R}^n que toma valores reales no negativos y verifica que:

$$\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = 0$$

$$\|\lambda\vec{x}\| = |\lambda|\|\vec{x}\|, \forall \lambda \in \mathbb{R}, \forall \vec{x} \in \mathbb{R}^n$$

$$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|, \forall \vec{x}, \vec{y} \in \mathbb{R}^n$$

7 Así en el espacio vectorial \mathbb{R}^n son normas vectoriales posibles para el vector \mathbf{x}

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{i=n} |x_i|$$

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{i=n} |x_i|^2}$$

$$\|\mathbf{x}\|_\infty = \text{Máx}\{x_i\}_{1 \leq i \leq n}$$

8 Si \mathbf{x}^k y \mathbf{x}^m son vectores de \mathbb{R}^n , se llama **distancia** asociada a la norma $*$, al número

$$d(\mathbf{x}^k, \mathbf{x}^m) = \|\mathbf{x}^k - \mathbf{x}^m\|_*$$

9 Si se trata de matrices A_n definidas en \mathbb{R}^n se definen como normas los reales no negativos que verifican

$$\|A\| = 0 \Leftrightarrow A = 0$$

$$\|\lambda A\| = |\lambda| \|A\|, \forall \lambda \in \mathbb{R}, \forall A \in \mathbb{R}^n$$

$$\|A + B\| \leq \|A\| + \|B\|, \forall A, B \in \mathbb{R}^n$$

$$\|A \cdot B\| \leq \|A\| \|B\|, \forall A, B \in \mathbb{R}^n$$

10 Son posibles las siguientes

$$\|A\|_F = \sqrt{\sum_{j=1}^n \sum_{i=1}^n |a_{i,j}|^2}$$

$$\|A\|_1 = \text{Máx} \sum_{\substack{i=1 \\ 1 \leq j \leq n}}^n |a_{i,j}|$$

$$\|A\|_2 = \text{Máx} \sum_{\substack{j=1 \\ 1 \leq i \leq n}}^n |a_{i,j}|$$

11 También puede utilizarse como norma para matrices reales la siguiente

$$\|A\| = \sqrt{\rho(A^t \cdot A)}$$

donde A^t es la transpuesta de la matriz A y ρ indica el radio espectral, es decir el módulo del valor propio que tenga mayor módulo del producto $A^t A$

II ITERACIÓN PUNTO FIJO

12 La iteración de punto fijo para sistemas de ecuaciones no lineales tiene muchas similitudes conceptuales con el método de iteración simple visto en el capítulo RAICES DE ECUACIONES. La dificultad a salvar es la correspondiente al trabajo en espacios de n dimensiones donde los módulos antes utilizadas se valoran ahora en términos de normas de vectores y matrices. La convergencia será la convergencia de sucesiones de vectores, como a continuación se presenta.

13 Sea entonces el sistema no lineal

$$\vec{f}(\vec{x}) = 0$$

o su equivalente

$$\begin{cases} f_1(x_1, x_2, x_3, \dots, x_n) = 0 \\ f_2(x_1, x_2, x_3, \dots, x_n) = 0 \\ f_3(x_1, x_2, x_3, \dots, x_n) = 0 \\ \dots\dots\dots \\ f_n(x_1, x_2, x_3, \dots, x_n) = 0 \end{cases}$$

y una aproximación inicial a la raíz del sistema

$$\vec{x}^{(0)} = \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ \dots\dots \\ x_n^{(0)} \end{pmatrix}$$

14 El primer paso a dar es transformar el sistema dado en otro de la forma

$$\vec{x} = \vec{\varphi}(\vec{x})$$

o su equivalente

$$\begin{cases} x_1 = \varphi_1(x_1, x_2, x_3, \dots, x_n) \\ x_2 = \varphi_2(x_1, x_2, x_3, \dots, x_n) \\ x_3 = \varphi_3(x_1, x_2, x_3, \dots, x_n) \\ \dots\dots\dots \\ x_n = \varphi_n(x_1, x_2, x_3, \dots, x_n) \end{cases}$$

15 La elección de las funciones φ no es para nada trivial pues de ellas depende nada menos que la convergencia o no convergencia del método en estudio.

Algunas posibles son elementales, por ejemplo

$$\vec{x} = \vec{x} + \vec{f}(\vec{x}) \Rightarrow \vec{\varphi}(\vec{x}) = \vec{x} + \vec{f}(\vec{x})$$

$$\vec{x} = \vec{x} - \vec{f}(\vec{x}) \Rightarrow \vec{\varphi}(\vec{x}) = \vec{x} - \vec{f}(\vec{x})$$

otras sencillamente dependen de un despeje total o parcial de x_1 en la primera ecuación, de x_2 en la segunda y, por fin de x_n en la última ecuación dada.

16 A partir de la elección de las funciones φ_i se genera recursivamente la sucesión de vectores en el espacio de n dimensiones, a partir de la aproximación inicial x^0

$$\vec{x}^{k+1} = \vec{\varphi}(\vec{x}^k)$$

o su equivalente

$$\begin{cases} x_1^{k+1} = \varphi_1(x_1^k, x_2^k, x_3^k, \dots, x_n^k) \\ x_2^{k+1} = \varphi_2(x_1^k, x_2^k, x_3^k, \dots, x_n^k) \\ x_3^{k+1} = \varphi_3(x_1^k, x_2^k, x_3^k, \dots, x_n^k) \\ \dots\dots\dots \\ x_n^{k+1} = \varphi_n(x_1^k, x_2^k, x_3^k, \dots, x_n^k) \end{cases}$$

17 Se plantean, como siempre, las siguientes preguntas: ¿es convergente el proceso?, si lo es, ¿converge a la solución del problema, es decir, a las raíces del sistema?

18 Contestarlas requiere probar que

$$\lim_{k \rightarrow \infty} \vec{x}^{k+1} = \lim_{k \rightarrow \infty} \vec{\varphi}(\vec{x}^k)$$

si estos límites existen, deberá ser

$$\lim_{k \rightarrow \infty} \vec{x}^{k+1} = \vec{\Xi}$$

$$\lim_{k \rightarrow \infty} \vec{\varphi}(\vec{x}^k) = \vec{\Xi}$$

$$\vec{\Xi} = \vec{\varphi}(\vec{\Xi})$$

siendo entonces Ξ un punto fijo. Naturalmente, para resolver sistemas no lineales es necesario que existan puntos fijos y para ello, a su vez es necesario que la aplicación sea una contracción.

19 Para ello es nuevamente necesario que la aplicación de $D \subset \mathbb{R}^n$ en $D \subset \mathbb{R}^n$ definida por $\mathbf{x} = \varphi(\mathbf{x})$ sea Lipschitziana, es decir que se verifique que existe una constante real $L > 0$ tal que

$$d[\vec{\varphi}(\vec{x}), \vec{\varphi}(\vec{y})] \leq Ld(\vec{x}, \vec{y}), \forall \vec{x}, \vec{y} \in D$$

y que la constante L sea menor que 1

20 Si eso se cumple, existe un punto fijo. En efecto, siendo

$$\vec{x}^{k+1} = \vec{\varphi}(\vec{x}^k)$$

será

$$d(\vec{x}^1, \vec{x}^2) = d[\vec{\varphi}(\vec{x}^0), \vec{\varphi}(\vec{x}^1)] \leq Ld(\vec{x}^0, \vec{x}^1)$$

$$d(\vec{x}^2, \vec{x}^3) = d[\vec{\varphi}(\vec{x}^1), \vec{\varphi}(\vec{x}^2)] \leq Ld(\vec{x}^1, \vec{x}^2) \leq L^2d(\vec{x}^0, \vec{x}^1)$$

$$\dots\dots\dots$$

$$d(\vec{x}^n, \vec{x}^{n+1}) = d[\vec{\varphi}(\vec{x}^{n-1}), \vec{\varphi}(\vec{x}^n)] \leq Ld(\vec{x}^{n-1}, \vec{x}^n) \leq \dots \leq L^n d(\vec{x}^0, \vec{x}^1)$$

en particular, tomando un p natural cualquiera y aplicando la desigualdad triangular se tendrá:

$$d(\bar{x}^n, \bar{x}^{n+p}) \leq d(\bar{x}^n, \bar{x}^{n+1}) + d(\bar{x}^{n+1}, \bar{x}^{n+2}) + d(\bar{x}^{n+2}, \bar{x}^{n+3}) + \dots + d(\bar{x}^{n+p-1}, \bar{x}^{n+p}) \leq$$

$$L^n d(\bar{x}^0, \bar{x}^1) + L^{n+1} d(\bar{x}^0, \bar{x}^1) + L^{n+2} d(\bar{x}^0, \bar{x}^1) + \dots + L^{n+p-1} d(\bar{x}^0, \bar{x}^1) =$$

$$L^n d(\bar{x}^0, \bar{x}^1) (1 + L + L^2 + \dots + L^{p-1}) \leq \frac{L^n d(\bar{x}^0, \bar{x}^1)}{1 - L}$$

finalmente puede escribirse

$$d(\bar{x}^n, \bar{x}^{n+p}) \leq \frac{L^n}{1 - L} d(\bar{x}^0, \bar{x}^1)$$

como φ es una contracción L es menor que uno, entonces para cualquier $\varepsilon > 0$ basta con tomar

$$N \geq \frac{\log\left(\frac{\varepsilon(1-L)}{d(\bar{x}^0, \bar{x}^1)}\right)}{\log(L)}$$

para que se verifique que la distancia $d(x^n, x^{m=n+p}) < \varepsilon$ para todo par de índices n y m mayores que N . En consecuencia la sucesión

$$\bar{x}^{k+1} = \bar{\varphi}(\bar{x}^k)$$

es una sucesión de Cauchy (criterio general de convergencia de sucesiones) y, en consecuencia, tiene límite y ese límite es un punto fijo de la aplicación dado que:

$$\bar{x}^{k+1} \xrightarrow{k \rightarrow \infty} \Xi \wedge \bar{\varphi}(\bar{x}^k) \xrightarrow{k \rightarrow \infty} \bar{\varphi}(\Xi) \Rightarrow \Xi = \bar{\varphi}(\Xi)$$

donde Ξ es un punto fijo.

21 Además ese punto fijo es único. Supóngase que existen dos puntos fijos Ξ_0 y Ξ_1 entonces

$$d(\Xi_0, \Xi_1) = d[\varphi(\Xi_0), \varphi(\Xi_1)] \leq Ld(\Xi_0, \Xi_1) < d(\Xi_0, \Xi_1)$$

lo que evidentemente es un absurdo dado que ningún número real es estrictamente menor a si mismo. Salvo que Ξ_0 y Ξ_1 sean coincidentes, en cuyo caso la distancia es nula y entonces

$$0 = d(\Xi_0, \Xi_1) = d[\varphi(\Xi_0), \varphi(\Xi_1)] \leq Ld(\Xi_0, \Xi_1) = 0$$

lo que asegura la unicidad del punto fijo.

22 Asimismo, si en la última expresión del párrafo 20 precedente se pasa al límite cuando $p \rightarrow \infty$ se tiene:

$$\lim_{p \rightarrow \infty} d(\bar{x}^n, \bar{x}^{n+p}) = d(\bar{x}^n, \Xi) = \frac{L^n}{1-L} d(\bar{x}^1, \bar{x}^0)$$

que indica que la "distancia" (en términos de la norma utilizada) entre el vector de aproximación correspondiente al enésimo paso y el vector raíz del sistema depende, por un lado del valor de la constante L y, por otro de la bondad (cercanía) de la aproximación inicial a la raíz Ξ con el vector \bar{x}^0 .

23 Siendo necesariamente L positiva y $L < 1$ por hipótesis, el conjunto de números reales entre los que puede variar L define de cierta forma la velocidad de la convergencia. Si L es positiva y muy próxima a 0, L^n será un real extremadamente pequeño, con lo cual, en cada paso de cálculo la distancia entre el vector vigente y el vector solución decrece rápidamente. Obsérvese que el denominador $1-L$ es, en este caso muy próximo a 1.

24 Por el contrario si L es menor que 1 pero muy próxima a dicho valor, L^n no será insignificante y su divisor $1-L$ será bastante pequeño como para que $L^n / (1-L)$ signifique una contracción pero lenta o muy lenta (o desesperante).

25 Esto es claro si se determina

$$d(\bar{x}^k, \bar{x}^{k+1}) = d[\bar{\varphi}(\bar{x}^{k-1}), \bar{\varphi}(\bar{x}^k)] < Ld(\bar{x}^{k-1}, \bar{x}^k)$$

lo que indica que la contracción, en cada paso de cálculo depende de L . Si L , siendo siempre positiva y menor que 1 es grande (cercana a 1), la nueva distancia será más chica pero bastante "parecida" a la anterior, si es muy chica (cercana a cero) la nueva distancia será mucho menor que la correspondiente al paso anterior.

26 Se debe recordar al leer estos párrafos el método de iteración simple de una ecuación donde se demostró que el error en un paso es proporcional, con constante de proporcionalidad L , al error en el paso anterior. Las similitudes son evidentes. Allá el trabajo era con funciones de una variable independiente, en intervalos y valores absolutos, aquí es con funciones de n variables (vectores), con normas y distancias, pero el fondo del tema es el mismo. En cada paso de cálculo se está más cerca de la raíz, con un ritmo "lineal".

27 Otro punto a tomar en consideración es la bondad del ajuste inicial, es decir que tan próximo al vector solución Ξ está el vector de partida $x^{(0)}$ seleccionado. Dado que x^1 surge como primer paso de cálculo partiendo de x^0 , la distancia entre ambos será "grande" si x^0 es lejano a la solución y será "chica" si no lo es. Como esta distancia $d(x^1, x^0)$ es parte invariable de la fórmula del párrafo 25 precedente, el método será mejor en el sentido de velocidad de convergencia, cuanto menor sea esta distancia.

28 En resumen, cuando el sistema de ecuaciones no lineales $f_k(x_i) = 0$ k, i $1, n$ se transforma en el sistema

$$\begin{cases} x_1 = \varphi_1(x_1, x_2, x_3, \dots, x_n) \\ x_2 = \varphi_2(x_1, x_2, x_3, \dots, x_n) \\ x_3 = \varphi_3(x_1, x_2, x_3, \dots, x_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_n = \varphi_n(x_1, x_2, x_3, \dots, x_n) \end{cases}$$

hay que determinar si dicho sistema es una contracción en el dominio D en el cual se trabaja. Para ello es necesario determinar la constante L y verificar que, siendo positiva, es menor que uno. Esto puede ser muy laborioso razón por la cual se hace necesario analizar variantes más fácilmente aplicables en la práctica.

29 Una de ellas es utilizar la matriz Jacobiana de las funciones φ_k , asumiendo la derivabilidad de las mismas con respecto a todas las variables. Esto es usando la matriz

$$J[\vec{\varphi}(\vec{x})] = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_2} & \cdots & \frac{\partial \varphi_1}{\partial x_n} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} & \cdots & \frac{\partial \varphi_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \varphi_n}{\partial x_1} & \frac{\partial \varphi_n}{\partial x_2} & \cdots & \frac{\partial \varphi_n}{\partial x_n} \end{vmatrix}$$

y determinando que alguna de sus normas (matriciales) $\|J[\varphi(\mathbf{x})]\|$ cumpla

$$\|J[\vec{\varphi}(\vec{x})]\| \leq L < 1$$

que

esto es así porque, aplicando el teorema del valor medio puede escribirse

$$\vec{\varphi}(\vec{x}^p) - \vec{\varphi}(\vec{x}^q) = J[\vec{\varphi}(\vec{x}^r)](\vec{x}^p - \vec{x}^q), \vec{x}^r \in D$$

y, trabajando según las normas elegidas resulta

$$\|\vec{\varphi}(\vec{x}^p) - \vec{\varphi}(\vec{x}^q)\| = \|J[\vec{\varphi}(\vec{x}^r)](\vec{x}^p - \vec{x}^q)\| \leq \|J[\vec{\varphi}(\vec{x}^r)]\| \|\vec{x}^p - \vec{x}^q\| \leq L \|\vec{x}^p - \vec{x}^q\| < \|\vec{x}^p - \vec{x}^q\|$$

una contracción en el espacio de n dimensiones donde está D

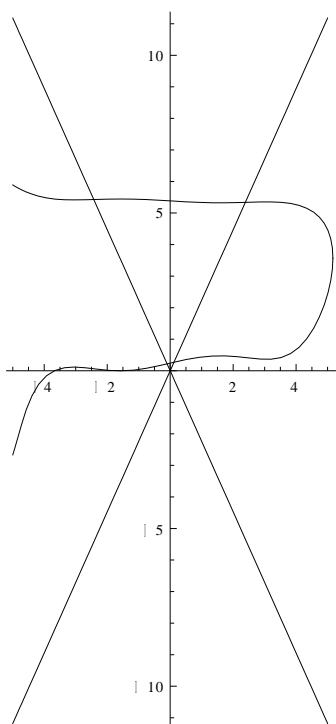
30 Como ejemplo se buscan raíces del siguiente sistema.

$$\begin{cases} 5x^2 - y^2 = 0 \\ y - 0.25(\text{Sen}(x) + \text{Cos}(y)) = 0 \end{cases}$$

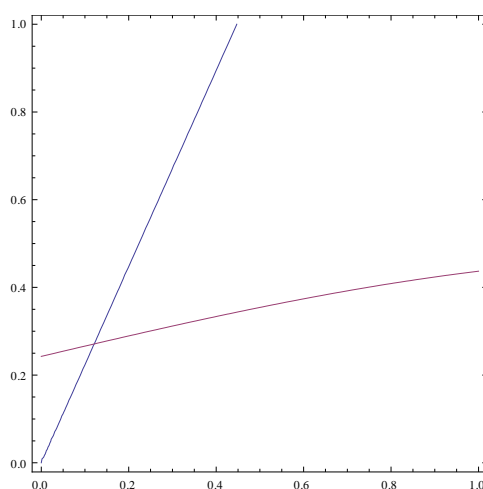
31 Para ello, en una primera etapa se buscan aproximaciones a dichas raíces mediante métodos gráficos. Si bien para la primera ecuación no hay problema para su representación, la segunda resulta dificultosa aún para el comando ImplicitPlot de MATHEMATICA, que "no se lleva bien" con las funciones trigonométricas para el cálculo de pares (x,y) de la función implícita.

32 Una solución a este nuevo problema consiste en aproximar a las funciones sen(x) y cos(y) por su aproximación mediante polinomios de Mc Laurin. Nótese que la convergencia de las series que se truncarán es muy buena y que los resultados que se buscan son "groseros".

33 Se obtiene con este método el siguiente gráfico. Pueden visualizarse cuatro raíces.



34 Para precisar un poco mejor la aproximación inicial, se utiliza otro recurso de MATHEMATICA, el comando *ContourPlot*, que permite obtener el siguiente gráfico para la raíz más próxima al origen en el primer cuadrante.



35 Se elige la que se encuentra en las cercanías de $(x,y) = (0.1,0.2)$ y se toman estos valores como valores iniciales.

36 Se transforma el sistema dado a la forma

$$\begin{cases} x_1 = \varphi_1(x_1, x_2) \\ x_2 = \varphi_2(x_1, x_2) \end{cases}$$

escribiendo

$$\begin{cases} x = \sqrt{(y^2)/5} \\ y = 0.25(\text{Sen}(x) + \text{Cos}(y)) \end{cases}$$

Donde se ha hecho $x_2 = y$

37 Corresponde ahora estudiar la convergencia. En el desarrollo teórico se ha visto que

$$\lim_{p \rightarrow \infty} d(\bar{x}^n, \bar{x}^{n+p}) = d(\bar{x}^n, \Xi) = \frac{L^n}{1-L} d(\bar{x}^1, \bar{x}^0)$$

Siendo necesario efectuar una estimación de L que, se recuerda, debe ser $L < 1$

38 En párrafo 29 precedente se expresó que puede tomarse como aproximación al número L alguna norma de la matriz jacobiana del sistema. Se elige la norma F , entonces

$$J(x, y) = \begin{bmatrix} \frac{\partial \varphi_1}{\partial x} & \frac{\partial \varphi_1}{\partial y} \\ \frac{\partial \varphi_2}{\partial x} & \frac{\partial \varphi_2}{\partial y} \end{bmatrix}_{(x,y)=(0.1,0.2)} = \begin{bmatrix} 0 & \frac{y}{\sqrt{5}} \\ 0.25\text{Cos}(x) & -0.25\text{Sen}(y) \end{bmatrix}_{(x,y)=(0.1,0.2)} = \begin{bmatrix} 0 & 0.089442 \\ 0.2487510 & 0.049967 \end{bmatrix}$$

Se calcula ahora

$$\left\| \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right\|_F = \sqrt{\sum_{i=1}^2 \sum_{j=1}^2 (a_{ij})^2} = \sqrt{0^2 + 0.089442^2 + 0.2487510^2 + 0.049967^2} = 0.269023 = L < 1$$

Este valor indica que el Método del Punto Fijo, para este sistema, es convergente

39 Se calcula, iterativamente

(x ₀ ,y ₀)	0.10000	0.20000
(x ₁ ,y ₁)	0.089442	0.34485
(x ₂ ,y ₂)	0.154222	0.257612
(x ₃ ,y ₃)	0.115208	0.280153
(x ₄ ,y ₄)	0.125288	0.268992
(x ₅ ,y ₅)	0.120297	0.27225
(x ₆ ,y ₆)	0.121754	0.270794
(x ₇ ,y ₇)	0.121103	0.271253
(x ₈ ,y ₈)	0.121308	0.271061

Obteniéndose los resultados ($x = 0.121308$; $y = 0.271061$), en ocho pasos de cálculo, con un "error de cierre" de $O(10^{-5})$

III RAPHSON NEWTON

40 Se recuerda que en el trabajo RAICES DE ECUACIONES se desarrolló el método de Newton para ecuaciones del tipo $f(x) = 0$. En esa oportunidad se demostró que la sucesión recurrentemente formada a partir de un valor x_0 tomado donde el signo de la función y su derivada segunda son coincidentes mediante la expresión

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Es convergente a la raíz de la ecuación dada en el sentido que

$$\lim_{k \rightarrow \infty} x_k = \xi$$

$$f(\xi) = 0$$

Y que la convergencia es cuadrática queriendo significar con ello que

$$|x_{k+1} - \xi| < Q|x_k - \xi|^2$$

41 Ahora, en el caso de sistemas de ecuaciones no lineales el problema a enfrentar es el siguiente:

Primero, un sistema de ecuaciones no lineales que vectorialmente puede escribirse

$$\vec{f}(\vec{x}) = 0$$

y escalarmente de la forma

$$\begin{cases} f_1(x_1, x_2, x_3, \dots, x_n) = 0 \\ f_2(x_1, x_2, x_3, \dots, x_n) = 0 \\ f_3(x_1, x_2, x_3, \dots, x_n) = 0 \\ \dots\dots\dots \\ f_n(x_1, x_2, x_3, \dots, x_n) = 0 \end{cases}$$

Segundo, un vector en el espacio de n dimensiones como primera aproximación a la raíz Ξ del sistema

$$\vec{x}^{(0)} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \\ \dots\dots\dots \\ x_n^{(0)} \end{bmatrix}$$

42 Como puede apreciarse, el planteo es similar. Pero ahora interviene la complicación emergente del trabajo en un espacio de n dimensiones, lo que modifica algo lo hecho anteriormente, aunque las semejanzas conceptuales son muy marcadas.

43 Se supone un incremento

$$\vec{h} = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ \dots \\ h_n \end{pmatrix}$$

de norma tal que el nuevo punto

$$\vec{x} = \vec{x}^{(0)} + \vec{h}$$

esté en un entorno de $\vec{x}^{(0)}$ y se recuerda que, siendo el vector

$$\vec{h} = \vec{x} - \vec{x}^{(0)}$$

el desarrollo en serie de Taylor para la función vectorial f puede escribirse

$$\begin{aligned} \vec{f}(\vec{x}^{(0)} + \vec{h}) &= \vec{f}(\vec{x}^{(0)}) + (\vec{h} \cdot \nabla) \vec{f}(\vec{x}^{(0)}) + \frac{1}{2!} (\vec{h} \cdot \nabla)^2 \vec{f}(\vec{x}^{(0)}) + \frac{1}{3!} (\vec{h} \cdot \nabla)^3 \vec{f}(\vec{x}^{(0)}) + \dots \\ &+ \frac{1}{(n-1)!} (\vec{h} \cdot \nabla)^{n-1} \vec{f}(\vec{x}^{(0)}) + \frac{1}{n!} (\vec{h} \cdot \nabla)^n \vec{f}(\vec{x}^{(0)}) \end{aligned}$$

(Cálculo Avanzado, Watson Fulks, Limusa Wiley, Cap 10, pto 5)

donde

$$\nabla = \frac{\partial}{\partial x_1} e_1 + \frac{\partial}{\partial x_2} e_2 + \frac{\partial}{\partial x_3} e_3 + \dots + \frac{\partial}{\partial x_n} e_n$$

Es el operador "nabla" y las "potencias" colocadas entre paréntesis son simbólicas. Por ejemplo

$$\left(\frac{\partial}{\partial x_1} h_1 + \frac{\partial}{\partial x_2} h_2 \right)^{(2)} f(x_1, x_2) = \frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} h_1^2 + 2 \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} h_1 h_2 + \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} h_2^2$$

44 Limitando el desarrollo a los términos lineales (ver similitud con lo expresado en párrafo N° precedente) se tiene

$$\vec{f}(\vec{x}^{(0)} + \vec{h}) = \vec{f}(\vec{x}^{(0)}) + (\vec{h} \cdot \nabla) \vec{f}(\vec{x}^{(0)})$$

Donde obviamente se comete un error igual al valor de los términos del desarrollo despreciados. La misma expresión vectorial anterior puede escribirse

$$\left\{ \begin{array}{l} f_1(x_1^{(0)} + h_1, x_2^{(0)} + h_2, \dots, x_n^{(0)} + h_n) = f_1(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}) + \sum_{i=1}^n \frac{\partial f_1(x_1, x_2, \dots, x_n)}{\partial x_i} \Big|_0 h_i \\ f_2(x_1^{(0)} + h_1, x_2^{(0)} + h_2, \dots, x_n^{(0)} + h_n) = f_2(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}) + \sum_{i=1}^n \frac{\partial f_2(x_1, x_2, \dots, x_n)}{\partial x_i} \Big|_0 h_i \\ f_3(x_1^{(0)} + h_1, x_2^{(0)} + h_2, \dots, x_n^{(0)} + h_n) = f_3(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}) + \sum_{i=1}^n \frac{\partial f_3(x_1, x_2, \dots, x_n)}{\partial x_i} \Big|_0 h_i \\ \dots \\ f_n(x_1^{(0)} + h_1, x_2^{(0)} + h_2, \dots, x_n^{(0)} + h_n) = f_n(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}) + \sum_{i=1}^n \frac{\partial f_n(x_1, x_2, \dots, x_n)}{\partial x_i} \Big|_0 h_i \end{array} \right.$$

45 Suponiendo ahora que los valores de h_i calculados a partir de este sistema son aquellos que anulan las funciones f_i , es decir los valores $x_i^{(0)} + h_i$ son exactamente las raíces del sistema no lineal que se está resolviendo, se tendrá

$$\begin{array}{c} \frac{\partial f_1}{\partial x_1} \quad \frac{\partial f_1}{\partial x_2} \quad \dots \quad \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} \quad \frac{\partial f_2}{\partial x_2} \quad \dots \quad \frac{\partial f_2}{\partial x_n} \\ \dots \\ \frac{\partial f_n}{\partial x_1} \quad \frac{\partial f_n}{\partial x_2} \quad \dots \quad \frac{\partial f_n}{\partial x_n} \end{array} \Big|_0 \begin{array}{c} h_1 \\ h_2 \\ \dots \\ h_n \end{array} = - \begin{array}{c} f_1(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \\ f_2(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \\ \dots \\ f_n(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \end{array}$$

un sistema lineal en las incógnitas h_i cuya solución puede hallarse por cualquiera de los métodos estudiados para este tipo de problemas.

46 Debe observarse que la matriz

$$\begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}$$

no es otra cosa que el Jacobiano de las funciones f_i con respecto a las variables x_i , determinante que en los cursos de Análisis Matemático II se simboliza

$$J\left(\begin{matrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{matrix}\right) = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}$$

47 Entonces, despejando queda

$$\begin{vmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{vmatrix} = -J^{-1}\left(\begin{matrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{matrix}\right)_0 \begin{vmatrix} f_1(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \\ f_2(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \\ \vdots \\ f_n(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \end{vmatrix}$$

Ajustando con estos valores o correcciones a las aproximaciones iniciales se tendrá.

$$\begin{vmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{vmatrix} = \begin{vmatrix} x_1^{(0)} \\ x_1^{(0)} \\ \vdots \\ x_1^{(0)} \end{vmatrix} - J^{-1}\left(\begin{matrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{matrix}\right)_0 \begin{vmatrix} f_1(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \\ f_2(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \\ \vdots \\ f_n(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}) \end{vmatrix}$$

$$\bar{x}^{(1)} = \bar{x}^{(0)} - J^{-1} \left[\frac{\bar{f}(\bar{x}^{(0)})}{\bar{x}^{(0)}} \right] \bar{f}(\bar{x}^{(0)})$$

48 Naturalmente estos NO son los valores de las raíces porque se han despreciado los términos no lineales del correspondiente desarrollo de Taylor pero puede demostrarse que se trata de una mejora en sus respectivas estimaciones, conformando los sucesivos vectores calculados de esta forma una sucesión convergente a las mismas. La expresión general es

$$\begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ \dots \\ x_n^{(k+1)} \end{pmatrix} = \begin{pmatrix} x_1^{(k)} \\ x_1^{(k)} \\ \dots \\ x_1^{(k)} \end{pmatrix} - J^{-1} \left(\frac{f_1, f_2, f_3, \dots, f_n}{x_1, x_2, x_3, \dots, x_n} \right)_k \begin{pmatrix} f_1(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_n^{(k)}) \\ f_2(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_n^{(k)}) \\ \dots \\ f_n(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_n^{(k)}) \end{pmatrix}$$

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} - J^{-1} \left[\frac{\bar{f}(\bar{x}^{(k)})}{\bar{x}^{(k)}} \right] \bar{f}(\bar{x}^{(k)})$$

vale la pena observar la similitud formal de esta última con aquella con la que se comenzó este tema

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

49 Obsérvese que, en cada paso de cálculo es necesario invertir una matriz de $n \times n$, la matriz Jacobiana de las funciones que componen el sistema o, en su defecto, resolver por algún método apto un sistema de ecuaciones lineales de $n \times n$.

50 El estudio de la convergencia y del error en este método requiere recursos que están fuera del alcance de estas notas.

51 Se resuelve a continuación el mismo sistema con que se ejemplificó el Método del Punto Fijo.

$$\begin{cases} f_1 = 5x^2 - y^2 \\ f_2 = y - 0.25 * (\text{Sen}(x) + \text{Cos}(y)) \end{cases}$$

Tomando como punto de partida el par $(x_0, y_0) = (0.1, 0.2)$

52 Se aplica la expresión adecuada a dos variables

$$\begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ \dots \\ x_n^{(k+1)} \end{pmatrix} = \begin{pmatrix} x_1^{(k)} \\ x_1^{(k)} \\ \dots \\ x_1^{(k)} \end{pmatrix} - J^{-1} \left(\frac{f_1, f_2, f_3, \dots, f_n}{x_1, x_2, x_3, \dots, x_n} \right)_k \begin{pmatrix} f_1(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_n^{(k)}) \\ f_2(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_n^{(k)}) \\ \dots \\ f_n(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_n^{(k)}) \end{pmatrix}$$

$$\begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} = \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix} - J^{-1} \left(\frac{f_1, f_2}{x, y} \right)_k \begin{pmatrix} f_1(x^{(k)}, y^{(k)}) \\ f_2(x^{(k)}, y^{(k)}) \end{pmatrix}$$

53 Se calcula

$$J^{-1} \left(\frac{f_1, f_2}{x, y} \right)_k = \begin{bmatrix} 10x & -2y \\ -0.25\text{Cos}(x) & 1 + 0.25\text{Sen}(y) \end{bmatrix}^{-1}$$

resultando

$$J^{-1} \left(\frac{f_1, f_2}{x, y} \right) = \begin{bmatrix} \frac{1 + 0.25\text{Sen}(y)}{10x - 0.5y\text{Cos}(x) + 2.5x\text{Sen}(y)} & \frac{2y}{10x - 0.5y\text{Cos}(x) + 2.5x\text{Sen}(y)} \\ \frac{0.25\text{Cos}(x)}{10x - 0.5y\text{Cos}(x) + 2.5x\text{Sen}(y)} & \frac{10x}{10x - 0.5y\text{Cos}(x) + 2.5x\text{Sen}(y)} \end{bmatrix}$$

Y se hace

$$\begin{bmatrix} x \\ y \end{bmatrix}^{(k+1)} = \begin{bmatrix} x \\ y \end{bmatrix}^{(k)} -$$

$$\begin{bmatrix} \frac{1 + 0.25 \operatorname{Sen}(y)}{10x - 0.5y \operatorname{Cos}(x) + 2.5x \operatorname{Sen}(y)} & \frac{2y}{10x - 0.5y \operatorname{Cos}(x) + 2.5x \operatorname{Sen}(y)} \\ \frac{0.25 \operatorname{Cos}(x)}{10x - 0.5y \operatorname{Cos}(x) + 2.5x \operatorname{Sen}(y)} & \frac{10x}{10x - 0.5y \operatorname{Cos}(x) + 2.5x \operatorname{Sen}(y)} \end{bmatrix}^{(k)} \begin{bmatrix} 5x^2 - y^2 \\ y - 0.25[\operatorname{Sin}(x) + \operatorname{Cos}(y)] \end{bmatrix}^{(k)}$$

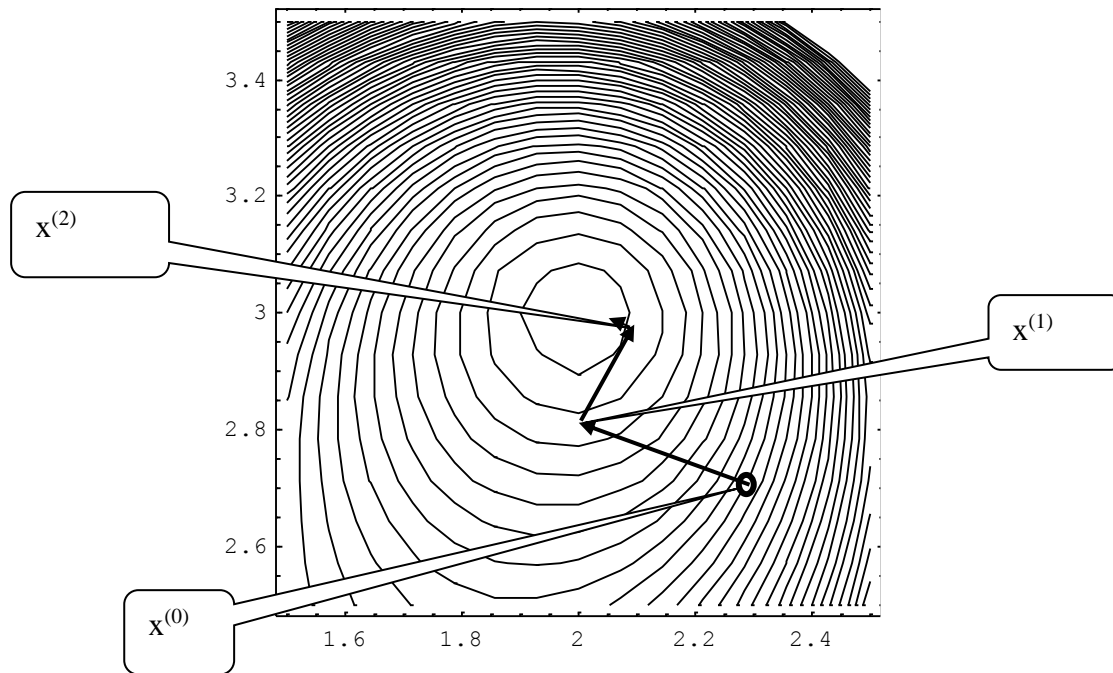
54 Partiendo, como se ha dicho, del punto $(x_0, y_0) = (0.1, 0.2)$

	X	Y
0	0,100000	0,200000
1	0,118411	0,271027
2	0,121280	0,271114
3	0,121242	0,271105

Se obtienen, en tres pasos, las raíces buscadas.

IV METODO DEL GRADIENTE

55 El método del gradiente es conceptualmente simple. Para mejor apreciar esa simplicidad supóngase una superficie modular como la mencionada **en párrafo nnn precedente** correspondiente a la determinación de las raíces, reales o complejas, de un polinomio. La aproximación inicial $x^{(0)}$ deberá corresponder a un punto próximo a la raíz buscada, como se indica en el gráfico siguiente



56 Dado que se está buscando el punto donde la función (o su módulo) se anula es obvio que habrá que descender por la superficie modular hasta llegar hasta el fondo del valle en el que se encuentra la raíz. Varios son los caminos posibles, pero uno solo de ellos es el más rápido. Del análisis matemático se sabe que el gradiente es vector orientado según la más rápida variación de la función, motivo por el cual el camino más rápido es aquel que sigue la línea de gradientes (con dirección opuesta al mismo)

57 Cabe una pregunta ¿hasta dónde habrá que deslizarse por la línea del gradiente, con punto de partida en $x^{(0)}$? La respuesta es fácil, muy fácil para expresarla con palabras pero algo más compleja matemáticamente.

58 En palabras basta con decir que el deslizamiento hacia abajo sobre la superficie debe hacerse según la dirección del gradiente hasta el punto en que deje de descenderse. Pero, ¿cuando deja de descenderse? De nuevo, fácil: en el preciso instante en que se comienza a subir. Ese punto donde ya no se baja y se empieza a subir es sencillamente un mínimo. Se lo ha llamado $x^{(1)}$ en el gráfico anterior. Cabe ahora una nueva pregunta ¿mínimo de qué función?

59 La respuesta está dada en los párrafos anteriores. ¿Acaso no se expresó que el deslizamiento se hacía sobre la superficie modular?, entonces, la función cuyo extremo (mínimo) se busca sobre la línea del gradiente en $x^{(0)}$ es la función superficie modular u otra con ella íntimamente relacionada, cuya expresión se escribe $S(x)$.

60 En el punto (vector) $x^{(1)}$ se repite el mismo procedimiento, deslizándose hacia abajo según la línea de gradiente en ese punto ($x^{(1)}$) hasta que se presente nuevamente un mínimo, obteniéndose así el punto (vector) $x^{(2)}$ y, desde este, según su línea de gradiente hasta otro punto $x^{(3)}$ y así sucesivamente hasta estar en presencia de un punto (vector) $x^{(k)}$ que satisface al sistema dado con la tolerancia asumida como buena.

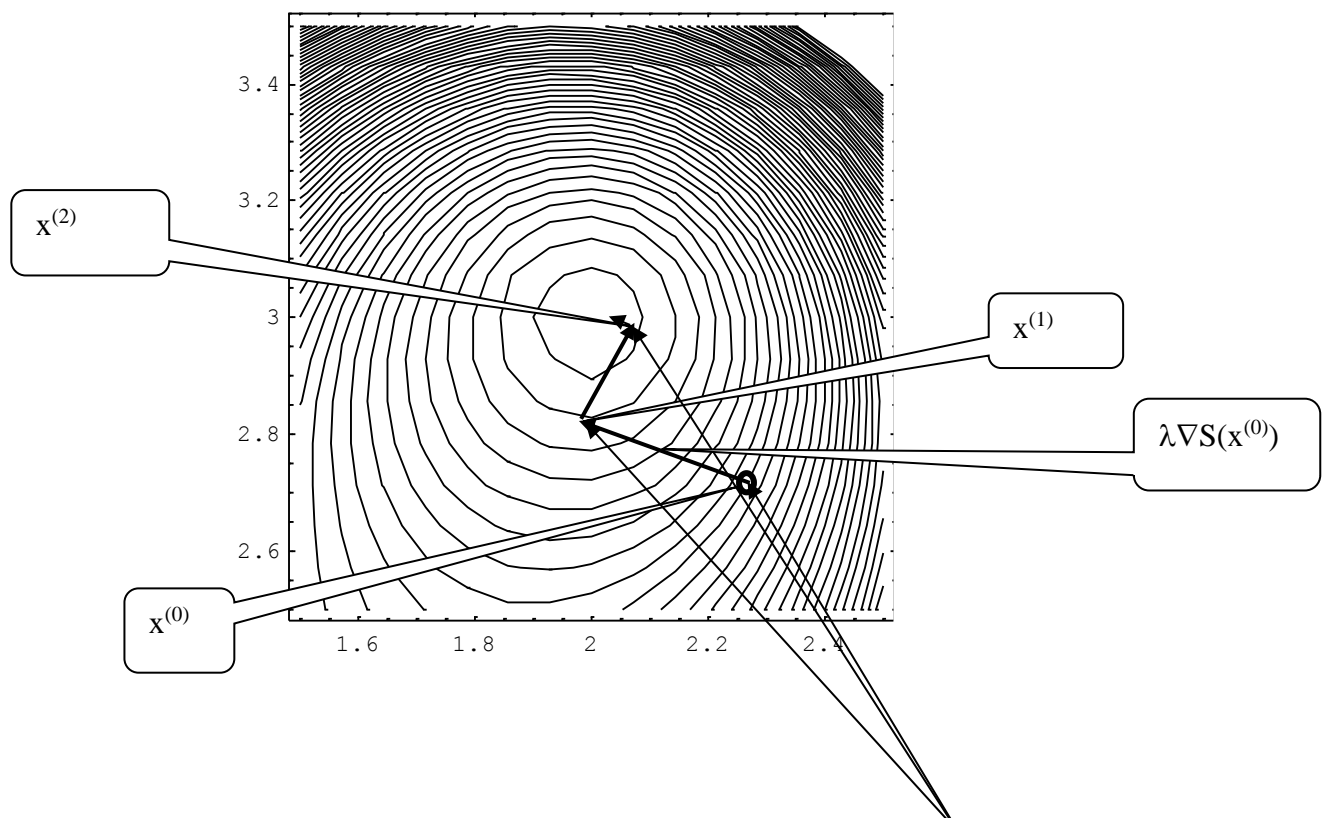
61 Naturalmente, el descenso se produce según la poligonal cuyos vértices son

$$\vec{x}^{(0)}, \vec{x}^{(1)}, \vec{x}^{(2)}, \vec{x}^{(3)}, \dots, \vec{x}^{(k)}, \dots$$

esperándose una buena tasa de convergencia. Esto es así cuando los "valles" están netamente definidos como lo es el del gráfico anterior. En cambio si el "fondo" del "valle" es estrecho y alargado la poligonal puede entrar en un casi interminable zig zag hasta llegar, si llega, al mínimo buscado.

62 Hasta aquí los conceptos insitos en el método. Si ellos están claros no debería ser dificultoso seguir el desarrollo matemático. No es elemental.

63 En primer lugar, del gráfico anterior pueden considerarse los siguientes esquemas vectoriales





de ellos resulta que

$$\vec{x}^{(k)} = \vec{x}^{(k-1)} - \lambda_{k-1} \nabla S(\vec{x}^{(k-1)})$$

donde λ_{k-1} es un parámetro a determinar para hacer mínima la función $S(X)$

64 Se toma como función $S(x)$ el cuadrado del módulo de la función vectorial f . Esto, por supuesto, no altera los mínimos y en consecuencia, las raíces. Es válido escribir

$$S(\vec{x}) = S(x_1, x_2, x_3, \dots, x_n) = \sum_{i=1}^n (f_i(x_1, x_2, x_3, \dots, x_n))^2 = (\vec{f}(\vec{x}) \bullet \vec{f}(\vec{x}))$$

dado que el vector raíz del sistema anula a esta función, haciendo cero a todos y cada uno de sus componentes f_i

65 Al reemplazar las variables x_i del párrafo 63 en la función $S(x)$ se obtiene una función del parámetro λ . Se la denomina $\Phi(\lambda)$. Se recomienda echar una mirada al ejemplo del párrafo 69.

66 Si se considera que el parámetro λ es pequeño en la expresión resultante

$$\Phi(\lambda) = \sum_{i=1}^n \left\{ f_i \left[\vec{x}^{(k-1)} - \lambda \nabla S(\vec{x}^{(k-1)}) \right] \right\}^2$$

se puede efectuar un desarrollo en serie de Taylor truncado en sus términos lineales obteniéndose

$$\Phi(\lambda) = \sum_{i=1}^n \left[f_i(\vec{x}^{(k-1)}) - \lambda \frac{\partial f_i(\vec{x}^{(k-1)})}{\partial \vec{x}} \nabla S(\vec{x}^{(k-1)}) \right]^2$$

de donde, derivando con respecto a λ se tiene

$$\Phi'(\lambda) = -2 \sum_{i=1}^n \left[f_i(\vec{x}^{(k-1)}) - \lambda \frac{\partial f_i(\vec{x}^{(k-1)})}{\partial \vec{x}} \nabla S(\vec{x}^{(k-1)}) \right] \bullet \frac{\partial f_i(\vec{x}^{(k-1)})}{\partial \vec{x}} \nabla S(\vec{x}^{(k-1)})$$

de donde, a su vez igualando a cero y despejando se obtiene para el parámetro λ

$$\lambda_{k-1} = \frac{\sum_{i=1}^n f_i(\bar{x}^{(k-1)}) \frac{\partial f_i(\bar{x}^{(k-1)})}{\partial \bar{x}} \nabla S(\bar{x}^{(k-1)})}{\sum_{i=1}^n \left[\frac{\partial f_i(\bar{x}^{(k-1)})}{\partial \bar{x}} \nabla S(\bar{x}^{(k-1)}) \right]^2}$$

67 Teniendo presente que

$$\frac{\partial f_i(\bar{x})}{\partial \bar{x}} = J \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

y que siendo

$$\nabla = \frac{\partial}{\partial x_1} e_1 + \frac{\partial}{\partial x_2} e_2 + \frac{\partial}{\partial x_3} e_3 + \dots + \frac{\partial}{\partial x_n} e_n$$

resulta

$$\nabla S(\bar{x}) = \left(\frac{\partial}{\partial x_1} e_1 + \frac{\partial}{\partial x_2} e_2 + \frac{\partial}{\partial x_3} e_3 + \dots + \frac{\partial}{\partial x_n} e_n \right) \sum_{i=1}^n [f_i(\bar{x})]^2 = 2 \sum_{i=1}^n f_i(\bar{x}) \frac{\partial f_i(\bar{x})}{\partial x_j}$$

pero en esta última la derivada vectorial que aparece es la transpuesta del Jacobiano anterior, razón por la cual puede escribirse

$$\nabla S(\vec{x}) = 2J^t \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} \vec{f}(\vec{x})$$

de donde

$$\lambda_{k-1} = \frac{\sum_{i=1}^n f_i(\vec{x}^{(k-1)}) \frac{\partial f_i(\vec{x}^{(k-1)})}{\partial \vec{x}} \nabla S(\vec{x}^{(k-1)})}{\sum_{i=1}^n \left[\frac{\partial f_i(\vec{x}^{(k-1)})}{\partial \vec{x}} \nabla S(\vec{x}^{(k-1)}) \right]^2} =$$

$$2 \frac{\left(\vec{f}(\vec{x}^{(k-1)}), J \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} J^t \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} \vec{f}(\vec{x}^{(k-1)}) \right)}{\left(J \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} J^t \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} \vec{f}(\vec{x}^{(k-1)}), J \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} J^t \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix} \vec{f}(\vec{x}^{(k-1)}) \right)}$$

y, por último con este valor del parámetro

$$\vec{x}^{(k)} = \vec{x}^{(k-1)} - 2\lambda_{k-1} J^t \begin{pmatrix} f_1, f_2, f_3, \dots, f_n \\ x_1, x_2, x_3, \dots, x_n \end{pmatrix}_{k-1} \vec{f}(\vec{x}^{(k-1)})$$

se obtiene (o se debería obtener) una nueva aproximación al vector solución.

68 El uso del verbo en potencial en el párrafo anterior tiene su razón de ser en las enormes dificultades de cálculo que se presentan. Aparte de los inevitables problemas numéricos, la forma del valle en el que se encuentra la raíz y la complejidad de las expresiones a tratar hacen a este método de muy laboriosa y pesada aplicación.

69 A título de ejemplo de lo mencionado, se intenta resolver por este método el problema trivial

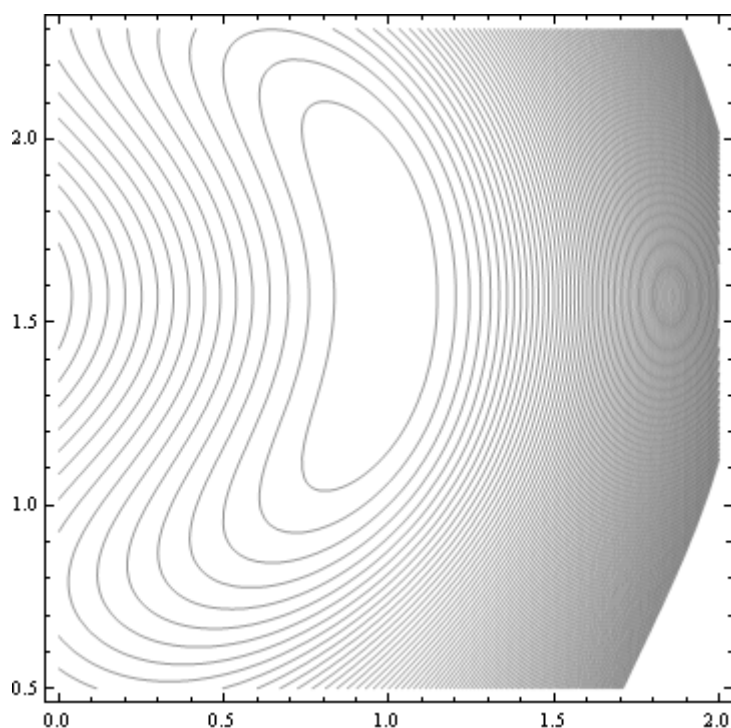
$$\begin{cases} f_1(x, y) = x \operatorname{sen}(y) - 1 = 0 \\ f_2(x, y) = x^2 + \cos(2y) = 0 \end{cases}$$

Cuya raíz es, evidentemente, $x = 1$ e $y = \pi/2$

La función $S(x)$ es, en este caso

$$S(x, y) = [f_1(x, y)]^2 + [f_2(x, y)]^2 = [x \operatorname{sen}(y) - 1]^2 + [x^2 + \cos(2y)]^2$$

70 Para tener una idea del "valle" donde se encuentra la raíz, el gráfico siguiente es explícito al respecto



Es fácil ver que la raíz buscada es el punto más profundo de la "lúnula" central (mejor sería decir la banana central), forma que puede ser causa de múltiples rebotes en la ruta de descenso siguiendo la línea de gradientes.

71 El gradiente, por su parte es

$$\nabla S(x, y) = \{2[x \operatorname{sen}(y) - 1] \operatorname{sen}(y) + 4x[x^2 + \cos(2y)]\}I + \{2 \cos(y)[x \operatorname{sen}(y) - 1] - 4 \operatorname{sen}(2y)[x^2 + \cos(2y)]\}J$$

haciendo

$$x_1 = x_0 - \lambda [2[x_0 \operatorname{sen}(y_0) - 1] \operatorname{sen}(y_0) + 4x_0 [x_0^2 + \cos(2y_0)]]$$

$$y_1 = y_0 - \lambda [2 \cos(y_0)[x_0 \operatorname{sen}(y_0) - 1] - 4 \operatorname{sen}(2y_0)[x_0^2 + \cos(2y_0)]]$$

se calcula la función

$$\Phi(\lambda) = S[x_1(\lambda), y_1(\lambda)]$$

$$\begin{aligned} S[x_1(\lambda), y_1(\lambda)] &= [x_1(\lambda)\text{sen}(y_1(\lambda)) - 1]^2 + [x_1(\lambda)^2 + \cos(2y_1(\lambda))] = \\ &= \left\{ \begin{aligned} &[x_0 - \lambda \ 2[x_0\text{sen}(y_0) - 1]\text{sen}(y_0) + 4x_0[x_0^2 + \cos(2y_0)]] \\ &\text{sen}[y_0 - \lambda 2\cos(y_0)[x_0\text{sen}(y_0) - 1] - 4\text{sen}(2y_0)[x_0^2 + \cos(2y_0)] \end{aligned} \right\}^2 + \\ &\left\{ \begin{aligned} &[x_0 - \lambda \ 2[x_0\text{sen}(y_0) - 1]\text{sen}(y_0) + 4x_0[x_0^2 + \cos(2y_0)]]^2 + \\ &\cos[2(y_0 - \lambda 2\cos(y_0)[x_0\text{sen}(y_0) - 1] - 4\text{sen}(2y_0)[x_0^2 + \cos(2y_0)])] \end{aligned} \right\}^2 \end{aligned}$$

cuyo mínimo hay que hallar. Como es usual, el procedimiento requiere derivar con respecto a λ , igualar esta derivada a cero y obtener el valor de λ que corresponde al mínimo.

72 Lo escrito constituye un fenomenal problema de cálculo numérico, hecho que justificaría, como se ha hecho en párrafo 66 anterior, efectuar un desarrollo de Taylor en λ hasta términos de primer orden para despejar λ de la expresión lineal resultante.

73 La observación de la expresión anterior basta para abandonar este camino para hallar la solución del sistema dado. Sin embargo, la claridad conceptual de método lleva a intentar otros caminos, tal vez menos rigurosos, pero efectivos.

74 Uno de ellos, consiste simplemente en estimar un valor para el parámetro λ y observar si, con ese valor adoptado, la función $S(x,y)$ decrece, lo que indica que la poligonal sobre sucesivos gradientes se dirige hacia el mínimo, raíz del sistema.

75 En el caso en estudio, tomando $x_0 = 0.9$ e $y_0 = 1.4$ se obtiene - utilizando MATHEMATICA como máquina de calcular - $S(x_0, y_0) = 0.0302733$.

76 Inmediatamente se hacen

$$x_1 = x_0 - \lambda \left. \frac{\partial S(x, y)}{\partial x} \right|_0 = 0.9 - \lambda \left\{ 2[x_0 \operatorname{sen}(y_0) - 1] \operatorname{sen}(y_0) + 4x_0 [x_0^2 + \cos(2y_0)] \right\}$$

$$y_1 = y_0 - \lambda \left. \frac{\partial S(x, y)}{\partial y} \right|_0 = 1.4 - \lambda \left\{ 2 \cos(y_0) [x_0 \operatorname{sen}(y_0) - 1] - 4 \operatorname{sen}(2y_0) [x_0^2 + \cos(2y_0)] \right\}$$

Se hace $\lambda = 0.2$ y se obtienen $x_1 = 1.03978$ e $y_1 = 1.42851$. Con estos valores, la función $S(x, y) = 0.0155853$ y se observa una disminución en su valor al pasar del punto P_0 al punto P_1

77 Se calculan ahora x_2 e y_2 en base al gradiente calculado en P_1 y se obtienen $x_2 = 0.983509$ e $y_2 = 1.45404$. Con estos valores, la función $S(x, y)$ vale 0.000568658 que indica una nueva disminución, o en forma equivalente, un mejor acercamiento a la raíz del sistema.

78 Un paso más arroja los siguientes resultados $x_3 = 0.997103$, $y_3 = 1.45407$ y la función $S(x, y)$ pasa a valer 0.000549178 , menor que el anterior pero lamentablemente muy poco menor. Esto indica que, para llegar a las raíces serán necesarios muchos más pasos de cálculo. Tal vez esto es debido a la "forma" del valle donde se encuentra la raíz buscada.

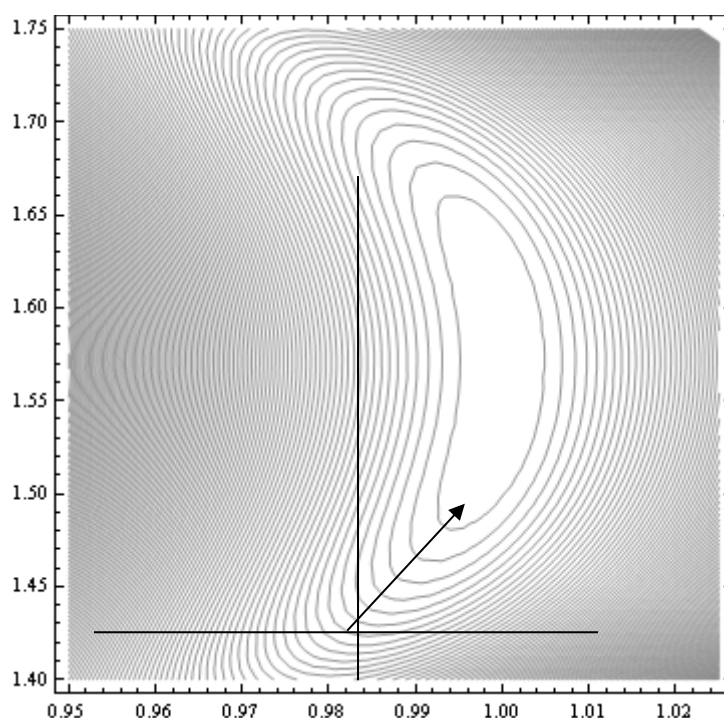
79 Un nuevo "experimento" ahora con $\lambda = 0.15$ arroja los siguientes resultados

n	X_n	Y_n	$S(x_n, y_n)$
0	0.90000	1.40000	0.0302733
1	1.00483	1.42139	0.00295744
2	0.97416	1.43121	0.00139841
3	0.991839	1.43061	0.000839455

4	0.983591	1.43514	0.000663708
5	0.988776	1.4368	0.000582263
6	0.986816	1.43973	0.000532164

En la cuarta columna puede apreciarse como se "desciende" por la superficie $S(x,y)$ pero también se aprecia que, al avanzar en el cálculo, el descenso existe pero es cada vez más lento.

80 El siguiente gráfico y la ubicación del punto (x_6, y_6) permite apreciar hasta dónde se ha llegado. Parecería apropiado dar ahora un buen salto en la dirección del gradiente en P_6 , $\nabla S(x_6, y_6)$ como marca la flecha



81 Quedan por realizar otros "experimentos", por ejemplo, cambiar el valor de λ en cada paso de cálculo, pero se carece de un criterio para hacerlo, salvo el de "prueba y error" que en esta materia puede traducirse en "¿baja más o baja menos?"

82 El tratamiento que se da a este tema en el excelente texto "Análisis Numérico" de Richard Burden y J. Douglas Faires consiste en asignar tres valores no negativos a λ : λ_1 , λ_2 y λ_3 , calcular Φ_1 , Φ_2 y Φ_3 y hallar el polinomio cuadrático en λ . Se toma como valor de cálculo para el descenso el valor de λ que minimiza este polinomio. Esto quizá porque en todo entorno de la raíz la función $S(x, y)$ puede ser aproximada suficientemente bien por una función cuadrática, mediante un desarrollo limitado de Taylor.

83 Los valores de λ se obtienen aproximando la expresión:

$$\Phi(\lambda) = S(X^{(0)} - \lambda \nabla S(X^{(0)}))$$

por medio de una parábola de segundo grado, para lo cual se construye el polinomio que interpola en (λ_1, Φ_1) ; (λ_2, Φ_2) y (λ_3, Φ_3)

84 Para iniciar el procedimiento se tomen $\lambda_1 = 0$ (valor que obviamente facilita el cálculo); λ_3 se elige arbitrariamente con la condición $\Phi(\lambda_3) < \Phi(\lambda_1)$ y se hace $\lambda_2 = \lambda_3 / 2$. Con los valores λ y Φ se determina la parábola de segundo grado que aproxima a la superficie S

85 A continuación se busca el mínimo de esta parábola de segundo grado en el intervalo comprendido entre λ_1 y λ_3 . El valor en el cual ocurre este mínimo λ_m corresponde al λ que debe ser utilizado en la expresión:

$$\bar{x}^{(k)} = \bar{x}^{(k-1)} - \lambda_{k-1} \nabla S(\bar{x}^{(k-1)})$$

para calcular el siguiente punto (x, y) .

86 Este proceso se repite hasta llegar a la aproximación deseada, tomando como parámetro de corte alguna convención, que puede ser el valor que toma la función $S(x, y)$. (En este caso este valor deber ser lo suficiente cercano a cero y dependerá de la exactitud con que se desea trabajar. Recuérdese que esta función es la suma del cuadrado del valor de cada una de las funciones que componen el sistema).

87 A modo de ejemplo se resuelve nuevamente el sistema utilizado en el punto 30, partiendo del punto $(x_0, y_0) = (1, 0)$

$$\begin{cases} 5x^2 - y^2 = 0 \\ y - 0.25(\text{Sen}(x) + \text{Cos}(y)) = 0 \end{cases}$$

$$\lambda_1 = 0$$

$$\lambda_3 = 0.5 \text{ (arbitrario)}$$

$$\lambda_2 = \lambda_3/2 = 0.25$$

Y se obtienen los siguientes valores

n	x	y	λ	S(x,y)
0	1.000000	0.000000	0.577133	25.2119
1	0.040855	0.424315	0.181025	0.0641780
2	0.097935	0.252524	0.026653	0.0004458
3	0.121060	0.265778	0.003860	0.0000387
4	0.119025	0.269059	0.002579	5.09127 10 ⁻⁶
5	0.121214	0.270424	0.0004896	6.30662 10 ⁻⁷
6	0.120957	0.270840	0.0003295	8.55129 10 ⁻⁸
7	0.121238	0.271013	0.0000662	1.15818 10 ⁻⁸
8	0.121203	0.271069	0.0000448	1.58500 10 ⁻⁹
9	0.121241	0.271092	9.0571 10 ⁻⁶²	2.16801 10 ⁻¹⁰
100	0.121237	0.27110	6.1304 10 ⁻⁶²	2.97003 10 ⁻¹¹
110	0.121242	0.271103	1.2405 10 ⁻⁶	4.069 10 ⁻¹²
120	0.121241	0.271104		5.574 10 ⁻¹³

Se toma como la aproximación a la solución buscada el punto (0.121241,0.271104) con un valor de $S(x,y) = 5.574 \cdot 10^{-13}$.



volver al Índice General