TEMA 3.

RESOLUCION DE ECUACIONES NO-LINEALES

1. Introducción

Se trata de encontrar un valor de x, que llamaremos x*, que satisfaga las ecuaciones

(1) $f(x^*)=0$

ó (2) $x^*=F(x^*)$ (ENL.1)

Cuando las ecuaciones f(x) y F(x) no son lineales el problema se debe resolver (salvo muy contadas excepciones) utilizando métodos iterativos. Una *solución iterativa* genera una secuencia x_0, x_1, x_2, x_3,...x_i donde x_0 es un valor estimado inicial y además:

$$\lim_{i \to \infty} x_i = x^* \qquad \text{(ENL.2)}$$

Es necesario conseguir una función iterativa para generar una secuencia de convergencia. La iteración se detiene cuando se cumple un determinado criterio, basado en una comparación entre la tolerancia en el error deseado ε_d y el error real ε_y.

El error exacto en cada iteración viene dado por $\left|x_i - x^*\right| = \varepsilon_i$

La relación de convergencia se define como $r_i = \dfrac{\varepsilon_{i+1}}{\varepsilon_i}$

Han sido propuestos distintos métodos para generar una secuencia de convergencia, todos ellos intentando mejorar la velocidad con que se alcanza la solución. En los métodos que vamos a ver a continuación la búsqueda de la solución estará sometida a las siguientes restricciones:

a) La raíz buscada se localizará en el intervalo cerrado [a,b] de tal manera que F(x) y f(x) serán funciones continuas en dicho intervalo.

b) Hay una única raiz en el intervalo

c) x* (la raíz) es un número real.

En el caso de sistemas de ecuaciones, el problema que se presenta trata de resolver un sistema de la forma **f**(**x**)=**0** donde **f** es un vector de funciones y **x** es un vector de variables desconocidas. Existen muchos métodos para resolver el problema entre ellos el de sustitución sucesiva, Newton, Broyden etc... Vamos a estudiar a continuación los métodos más extendidos.

2. Nomenclatura

f vector de funciones (que se han de igualar a cero)
F vector de funciones (que han de ser iguales a x)
x vector de variables desconocidas
k (subíndice) número de la iteración
i (superíndice) ecuación particular del sistema
j (subíndice) variable particular respecto con la que se trabaja
ε error permitido o tolerancia
x^* punto (vector) de partida
$\lambda(A)$ valor propio de la matriz A
w_k parámetro que determina la aceleración de algunos métodos
x^s punto donde se desarrolla la serie de Taylor en el método de Newton
J matriz jacobiana
H estimación de la matriz jacobiana

3. Resolución de una única ecuación de la forma x=F(x)

3.1. Método de Sustitución Sucesiva o Iteración Directa

El método de iteración directa está basado en generar una función de iteración de la forma:

$$x_{i+1} = F(x_i) \quad \text{i=1,2,3...} \qquad \text{(ENL.3)}$$

Como estimación del error se puede tomar:

$$\left|x_i - x_{i+1}\right| = \tilde{\varepsilon} \qquad \text{(ENL.4)}$$

En el método de iteración directa la condición necesaria y suficiente para la convergencia es que el valor estimado esté suficientemente cerca de la solución de tal manera que:

$$\left|F'(x^*)\right| < 1 \qquad \text{(ENL.5)}$$

Una posible demostración es la siguiente:

De acuerdo con la definición $x_{i+1} = F(x_i)$

En la solución $x^* = F(x^*)$

Restando ambas ecuaciones $x_{i+1} - x^* = F(x_i) - F(x^*)$

De acuerdo con el teorema del valor medio: $F(x_i) - F(x^*) = F'(\xi)(x_i - x^*)$

Recordando que $\varepsilon_{i+1} = x_{i+1} - x^*$ y que $\varepsilon_i = x_i - x^*$

De las ecuaciones anteriores $\dfrac{\varepsilon_{i+1}}{\varepsilon_i} = F'(\xi)$

Si x_i esta suficientemente cerca de la solución entonces: $F'(\xi) \approx F'(x^*)$ por lo tanto la velocidad limite de convergencia vendrá dada por:

$$\lim_{i\to\infty} r_i = \frac{\varepsilon_{i+1}}{\varepsilon_i} = \left|F'(x^*)\right| \qquad \text{(ENL.6)}$$

Solamente si $\left|F'(x^*)\right| < 1$ el error se reducirá en cada iteración.

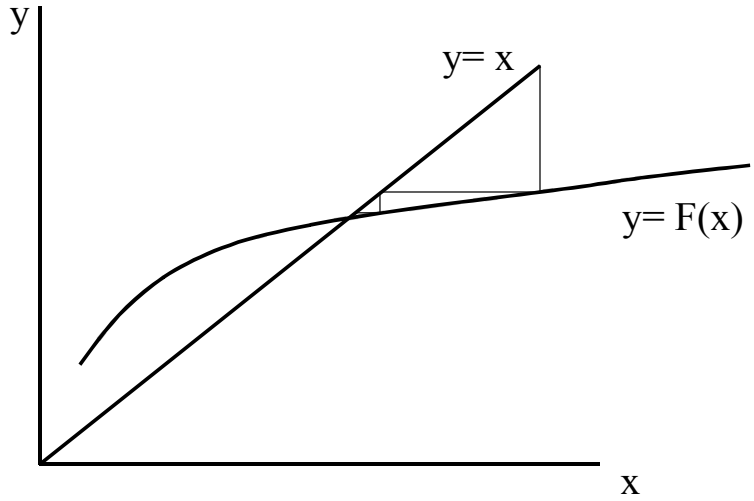

Figura 1a. El método de iteración directa converge.

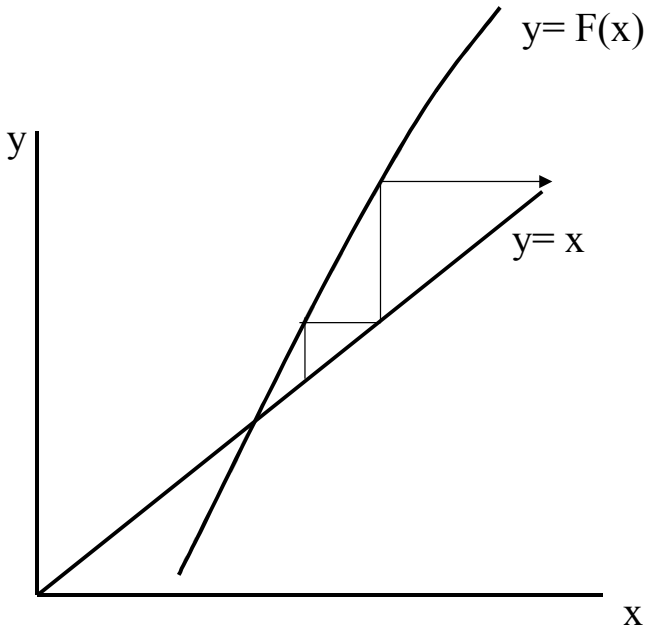

Figura 1b. El método de iteración directa diverge.

Si la pendiente es cercana a la unidad la velocidad de convergencia es lenta, y la estimación del error poco precisa (téngase en cuenta que la estimación del error se realiza entre dos iteraciones consecutivas y no con el verdadero valor). Si la pendiente es próxima a cero la velocidad de convergencia será grande y la estimación del error más precisa.

Cuando el método de sustitución sucesiva converge el límite de la relación de convergencia viene dado por la ecuación comentada anteriormente. Este tipo de convergencia donde el límite de la relación de convergencia se aproxima a un valor constante se llama *convergencia lineal.* Para un método de convergencia lineal, la representación del $log(\varepsilon_i)$ vs. el numero de iteraciones

se aproxima asintóticamente a una línea recta La pendiente de la asintota será $log\left(\mathrm{F}'\left(\mathrm{x}^{*}\right)\right)$. Ver Figura 1.

Si se asume que el valor de F'(x) no cambia mucho en las cercanías de la solución se puede obtener información adicional de las características de convergencia dentro del intervalo:

1.- Estimación de F'(x*):

$$\mathrm{r_i} \cong \frac{\widetilde{\varepsilon}_{\mathrm{i}+1}}{\widetilde{\varepsilon}_{\mathrm{i}}} \cong \left|\mathrm{F}'\left(\mathrm{x}^{*}\right)\right| \qquad \text{(ENL.7)}$$

2.- Estimación del error exacto:

$$\varepsilon_{\mathrm{i}} \cong \frac{\widetilde{\varepsilon}_{\mathrm{i}}}{\left[1-\mathrm{F}'\left(\mathrm{x}^{*}\right)\right]} \cong \frac{\widetilde{\varepsilon}_{\mathrm{i}}}{\left[1-\mathrm{r_i}\right]} \qquad \text{(ENL.8)}$$

3.- Estimación del número adicional de iteraciones para conseguir la solución dentro de un error de tolerancia deseado:

$$\mathrm{J} = \frac{log\frac{\varepsilon_{\mathrm{d}}}{\varepsilon_{\mathrm{i}}}}{log\,\mathrm{r_i}} \qquad \text{(ENL.9)}$$

3.2. Método de aceleración de convergencia de Wegstein

El uso de este método se basa en la suposición de que la derivada de la función F(x) no cambia mucho en las proximidades de la solución.

Se toman dos puntos iniciales: $[x_{i-1}, F(x_{i-1})]$ y $[x_i, F(x_i)]$ Se traza entonces la línea recta que une estos dos puntos, y la intersección de esta línea con la línea x=y se selecciona como la mejor estimación de la función. Así pues, la ecuación de la recta que pasa por los puntos $[x_{i-1}, F(x_{i-1})]$ y $[x_i, F(x_i)]$ viene dada por la expresión:

$$y - F(x_i) = \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}}(x - x_i) \qquad \text{(ENL.10)}$$

Para calcular el punto x_{i+1} se calcula la intersección de dicha recta con la recta $y = x$ de tal manera que:

$$x_{i+1} - F(x_i) = \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}}(x_{i+1} - x_i) \qquad \text{(ENL.11)}$$

Llamando $s_i = \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}}$, se obtiene

$$x_{i+1} - F(x_i) = s_i\ (x_{i+1} - x_i) \qquad \text{(ENL.12)}$$

despejando x_{i+1} se llega a:

$$x_{i+1} = \frac{1}{1-s_i} F(x_i) - \frac{s_i}{1-s_i} x_i = w_i\, F(x_i) + (1 - w_i)\, x_i \qquad \text{(ENL.13)}$$

donde $w_i = \frac{1}{1-s_i}$ (ENL.14)

Normalmente en lugar de utilizar 2 puntos al azar para la primera iteración se realiza una iteración directa.

El procedimiento gráfico se ilustra en la Figura 3.

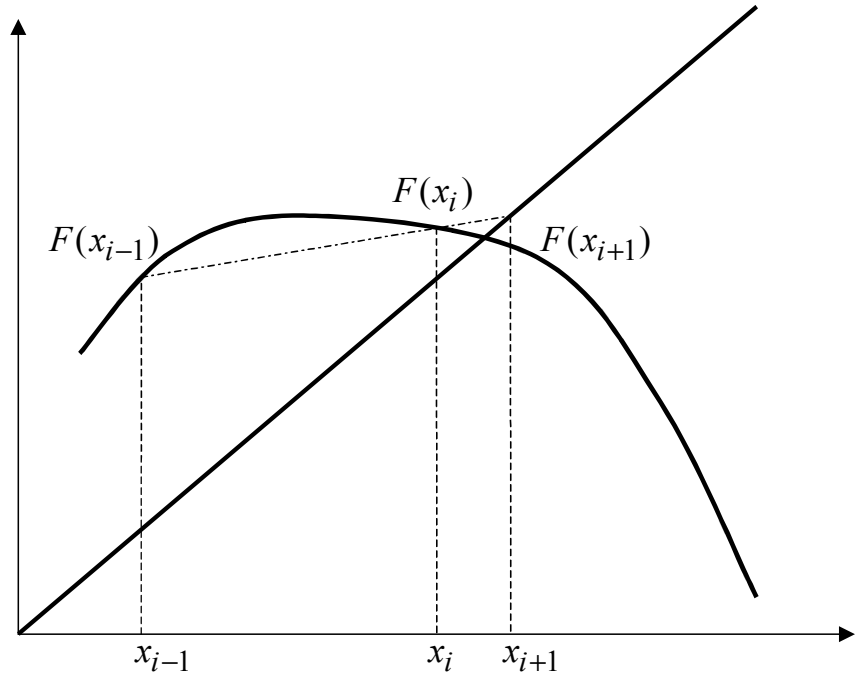

Figura 3. Método de Wegstein.

Se ha demostrado que si el método de iteración directa converge, el metodo de Wegstein lo hace más rápidamente. Con la única excepción del caso en que F'(x*)=0.

El método de Wegstein converge incluso en ocasiones en que el método de iteración directa no lo hace. Sin embargo pueden aparecer dificultades cuando $|F'(x)| = 1$ en algún punto dentro del intervalo de la solución.

4. Resolución de una única ecuación de la forma f(x)=0

En la primera parte vimos métodos numéricos para resolver una ecuación no lineal de la forma F(x)=x. Prácticamente todas las ecuaciones no lineales pueden ser expresadas de esta forma, pero no es necesario, existen otras técnicas numéricas para resolver ecuaciones no lineales de la forma f(x)=0.

4.1. Conceptos básicos

Como en el caso anterior se trata de encontrar la forma de generar una serie de secuencia x_0, x_1, x_2, ...donde x_0 es un valor supuesto inicial de tal manera que se cumpla la ecuación

$$\lim_{i \to \infty} x_i = x^*$$

donde i es el numero de iteraciones.

Como en el caso anterior es necesaria una *función de iteración* que permita generar la secuencia de valores de x. En la práctica es imposible generar una secuencia infinita de valores de x, por lo tanto no se podrá obtener la solución exacta, no obstante nos podremos aproximar a ella tanto como queramos. El criterio que nos dice para que valor de x_i nos debemos detener es el llamado *criterio de detención* y está basado en la comparación entre el error de tolerancia deseado que llamabamos ε_d y una estimación del error ε_i. La relación del error en dos iteraciones sucesivas da una idea de la velocidad de convergencia. Si definimos la *relación de convergencia* como:

$$r_i = \frac{|\varepsilon_{i+1}|}{|\varepsilon_i|^{\nu}}$$

donde ν es una constante. Se puede demostrar que si para algunos valores de $\nu \geq 1, r_i < 1 \;\; i = 0,1,2,3\ldots$ entonces la secuencia converge a la solución. Si para un valor especificado de ν el límite:

$$\lim_{i \to \infty} \frac{|\varepsilon_{i+1}|}{|\varepsilon_i|^{\nu}}$$

existe entonces, ν es el orden de convergencia del método iterativo. Como se mencionó anteriormente para el método de sustitución sucesiva el límite para el valor ν=1 existe presentando dicho método convergencia lineal o de primer orden.

Como en el caso anterior limitaremos nuestra búsqueda de la solución a un intervalo cerrado I=[a,b] tal que la función f(x) sea continua en dicho intervalo, presente en dicho intervalo una única raíz, y que ésta sea un número real.

4.2. El método de Bisección

Cuando se usa este método las iteraciones tienen que comenzar desde dos puntos iniciales x_0 y x_1 de tal manera que $f(x_0)$ y $f(x_1)$ tengan signos opuestos $(f(x_0).f(x_1) < 0)$. Se toma entonces el punto medio del intervalo $[x_{i-1}, x_i]$. La función de iteración de este método es:

$$x_{i+1} = (x_i + x_{i-1})/2 \qquad \text{(ENL.15)}$$

entre tres valores sucesivos de x: x_{i-1}, x_i, x_{i+1} debemos guardar para la siguiente iteración solamente dos de ellos. El valor de x_{i+1} nos lo guardamos siempre, de los otros dos nos quedaremos con aquel que cuyo valor de función sea de diferente signo que el valor de la función para x_{i+1}. De esta forma la raíz de la ecuación se encontrará siempre acotada por x_{i+1} y x_i. Dado que el intervalo se hace cada vez más pequeño la longitud del intervalo se puede usar como una estimación para calcular el error. Es fácil mostrar que después de "i" iteraciones el tamaño del intervalo se habrá reducido en un valor de:

$$L_i = 2^{-i} L_0 \qquad \text{(ENL.16)}$$

donde L_0 es el tamaño del intervalo inicial.

El método de la bisección es muy simple, incluso en ocasiones demasiado simple y no puede ser usado para muchas aplicaciones de análisis numérico. Algunas de sus propiedades lo hacen un método excelente para usos en ingeniería. Admitiendo que la función cumple los requerimientos básicos citados al principio del capítulo, si somos capaces de encontrar dos valores para la función con signos opuestos, la convergencia está asegurada. La principal desventaja del método de bisección es su lenta convergencia. Para varios tipos de problemas nosotros podemos desear usar un método iterativo de convergencia más rápida, el uso de estos métodos es esencial en casos tales como:

1. Se necesita una solución con una alta precisión

2. La función es muy complicada y su cálculo puede llevar bastante tiempo

3. La misma ecuación no lineal ha de ser resulta muchas veces (cientos o incluso miles).

4.3. El método de Newton-Raphson

Este método está basado en la expansión en una serie de Taylor de la función f(x) en las proximidades del punto x_i:

$$f(x_{i+1}) = f(x_i) + (x_{i+1} - x_i) f'(x_i) + \frac{(x_{i+1} - x_i)^2}{2} f''(x_i) + \ldots. \qquad \text{(ENL.17)}$$

dado que estamos buscando el valor de $f(x_{i+1})=0$, sustituyendo en la ecuación 2.6 y despreciando los términos de segundo orden, obtenemos la función de iteración de Newton-Raphson:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad i = 0,1,2,3..... \qquad \text{(ENL.18)}$$

Para que el método converja la función no debe presentar, en el intervalo de búsqueda máximos, mínimos o puntos de inflexión, en caso de no darse esta situación el método puede no conducir a la solución correcta. Aún cumpliéndose las condiciones anteriores la convergencia desde un punto inicial depende de la forma de la función.

Sokolnikov propone que aunque se cumplan las condiciones generales y la condición citada anteriormente, si la función presenta signos opuestos en los dos extremos del intervalo de búsqueda, debe elegirse como primer valor del intervalo de búsqueda para el tanteo el extremo en el que f(x) y f''(x) tengan el mismo signo.

El método de Newton-Raphson posee convergencia cuadrática o de segundo orden.

Una observación general del método de Newton dice que :"El método de Newton siempre converge si se parte de un valor x_0 suficientemente cerca del valor x*". Dado que no podemos conocer el valor de x* de antemano esta observación parece de poco uso, pero se puede desarrollar una estrategia adecuada basada en esta observación. Si no se dispone de un buen valor inicial se puede utilizar el método de la bisección para conseguir un valor suficientemente cerca de la solución y posteriormente terminar de resolver el problema por el método de Newton.

Cuando se usa el método de Newton se suele usar el valor de la función para estimar el error:

$$\tilde{\varepsilon} = \left| f(x_i) \right|$$

La estimación del error a partir del cálculo de $f(x_i)/f'(x_i)$ da mejores valores que solamente la función, pero la ecuación anterior no necesita del calculo de las derivadas que en ocasiones pueden ser costosas (en tiempo de cálculo).
Las iteraciones se suelen concluir cuando se cumple el siguiente test:

$$f(x_i + \varepsilon_d).f(x_i - \varepsilon_d) < 0$$

el que se cumpla esta última ecuación asegura que la diferencia (x^*-x_i) es menor que el error de tolerancia.

La Figura 4 muestra gráficamente tres iteraciones del método de Newton.

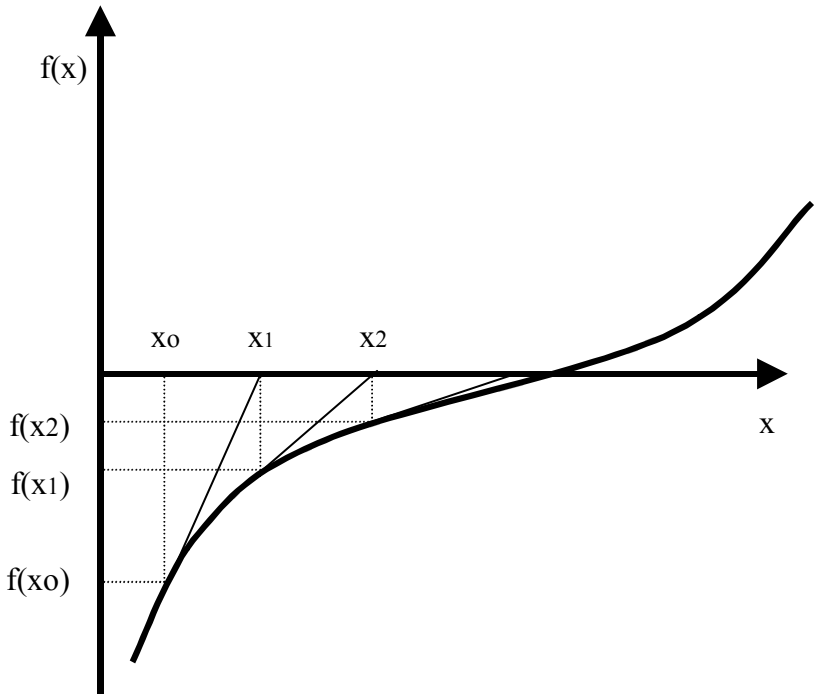

Figura 4. Representación gráfica del método de Newton

4.4. El método de la secante y la secante mejorado

Si nosotros no deseamos calcular la derivada podríamos aproximar la derivada por una línea recta que pasase a través de los puntos [x_{i-1}, $f(x_{i-1})$] y [x_i, $f(x_i)$] así :

$$f'(x_i) \approx \frac{\left[f(x_i) - f(x_{i-1})\right]}{x_i - x_{i-1}}$$

la derivada es pues aproximada por una secante entre las dos últimas iteraciones. La función de iteración para el método de la secante es pues:

$$x_{i+1} = x_i - f(x_i)\frac{(x_i - x_{i-1})}{\left[f(x_i) - f(x_{i-1})\right]} \qquad i=1,2,3,4....$$

Se puede demostrar que el orden de convergencia para el método de la secante es $\left[\left(1+\sqrt{5}\right)/2 = 1.618\right]$. Al igual que el método de Newton se puede decir que siempre converge desde un valor inicial suficientemente cerca a la solución. Se pueden utilizar, por lo tanto técnicas mixtas, como comenzar con el método de la bisección, y terminar con el método de la secante.

La velocidad de convergencia del método de la secante puede ser fácilmente acelerada. En lugar de aproximar f'(x) por una función lineal, parece natural obtener una convergencia más rápida aproximando f'(x) por un polinomio que pasa a través de todos los puntos previamente calculados, así podremos obtener la siguiente función de iteración:

$$x_{i+1} = \sum_{k=0}^{i} x_k \prod_{m=0}^{i} \frac{f(x_m)}{f(x_m) - f(x_k)} \qquad m \neq k$$

El orden de convergencia de este método se aproxima a 2 para valores elevados de i. Para i=2 la ecuación anterior se reduce al método de la secante.

Los métodos de la secante y la secante mejorado necesitan solamente una evaluación de la función por iteración y su convergencia es casi tan rápida como la del método de Newton. Su uso está recomendado en casos donde sea costosa la evaluación de la derivada. El método de la secante mejorado es el más costoso, y su uso está solamente recomendado cuando la evaluación de la función consume mucho tiempo y es importante la velocidad de convergencia.

4.5. El método de regula falsi

El método de regula falsi, o de la falsa posición, es realmente un método de secante que se desarrolló para intentar aumentar la velocidad del método de la bisección.

Como en el método de la bisección se parte de dos puntos *f(a) y f(b)* con signos opuestos. El método de la bisección utiliza el punto intermdedio del intervalo [a,b] para la siguiente iteración. Se puede encontrar una mejor aproximación si tomamos como punto para la siguiente iteración el punto de corte de la recta que pasa por (a,f(a)) y (b,f(b)) con el eje do abscisas: punto (c,0):

$$m = \frac{f(b) - f(a)}{b - a} = \frac{0 - f(b)}{c - b} \qquad \text{(ENL.19)}$$

Lo que nos da:

$$c = b - \frac{f(b)(b - a)}{f(b) - f(a)} \qquad \text{(ENL.20)}$$

La figura 5 muestra una representación gráfica del método de la falsa posición.

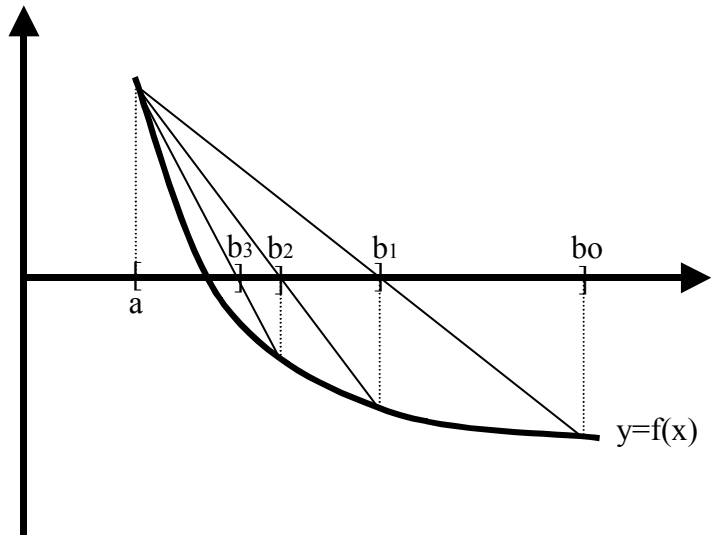

Figura 5. Método de Regula Falsi.

5. Métodos de resolución de sistemas de ecuaciones no lineales.

5.1. Método de sustitución sucesiva

El método es totalmente análogo al estudiado para una única variable. Se trata de reescribir el sistema de la forma **x**=**F**(**x**). Llamaremos 'k' a una iteración y 'k+1' a la siguiente. El esquema iterativo de este método es pues:

$$\boxed{\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k)} \qquad \text{(ENL.21)}$$

lo que implica que para cada ecuación:

$$\mathbf{x}^{i}{}_{k+1} = \mathbf{F}^{i}(\mathbf{x}_k) \qquad \text{(ENL.22)}$$

donde el superíndice 'i' hace referencia a las distintas ecuaciones del sistema. El criterio para detener las iteraciones suele ser

$$\left\| \mathbf{x}_{k+1} - \mathbf{x}_k \right\| < \varepsilon_d \qquad \text{(ENL.23)}$$

Se puede desarrollar un criterio de convergencia suficiente para el método de sustitución sucesiva cuando se parte de unos valores iniciales suficientemente cerca de la solución. Si consideramos la expansión en series de Taylor del sistema de ecuaciones F(x):

$$F(x_k) = F(x_{k-1}) + \left(\frac{\partial F}{\partial x}\right)^T_{x_{k-1}} (x_k - x_{k-1}) + \ldots \qquad \text{(ENL.24)}$$

Suponiendo que en las cercanías de la solución $\left(\frac{\partial F}{\partial x}\right) \approx cte$

$$x_{k+1} - x_k = F(x_k) - F(x_{k+1}) = \left(\frac{\partial F}{\partial x}\right)^T (x_k - x_{k-1}) \qquad \text{(ENL.25)}$$

o bien

$$x_{k+1} - x_k = \Delta x_{k+1} = \left(\frac{\partial F}{\partial x}\right)^T \Delta x_k \qquad \text{(ENL.26)}$$

Recordando que para las normas se cumple $\|AB\| \le \|A\|\|B\|$ tenemos:

$$\|\Delta x_{k+1}\| \le \left\|\left(\frac{\partial F}{\partial x}\right)\right\| \|\Delta x_k\| \qquad \text{(ENL.27)}$$

Utilizando la norma dos $\|A\|_2 = \sqrt{\lambda_{\max}\left(A^T A\right)}$ y recordando que partimos de un punto suficientemente cercano a la solución (es decir la derivada de F(x) es prácticamente constante en todas las iteraciones)

$$\|\Delta x_{k+1}\| \le \left(|\lambda|_{\max}\right)^k \|\Delta x_0\| \qquad \text{(ENL.28)}$$

Es decir el sistema converge cuando el valor absoluto del máximo valor propio de la matriz Jacobiana de F que contiene todas las derivadas parciales es menor que la unidad.

La ecuación anterior permite estimar el numero de iteraciones para alcanzar una tolerancia (ε) dada:

$$n_{iter} \geq \frac{Log\left(\frac{\varepsilon}{\|\Delta x_0\|}\right)}{Log\left(|\lambda|_{\max}\right)} \qquad \text{(ENL.29)}$$

Por ejemplo:

$\|\Delta x_0\| = 1$	ε=0.0001
λ_{max}	n
0.1	4
0.5	14
0.99	916

Un criterio equivalente al anterior y relativamente cómodo para comprobar la convergencia sin calcular los valores propios es comprobar que los sumatorios de las derivadas parciales respecto a cada variable en el punto de partida x* son menores que la unidad:

$\sum_{j=1}^{n}\left|\frac{\partial F_i(x^*)}{\partial x_j}\right| < 1$, para todas las funciones F_i que constituyen el sistema.

En el tema anterior vimos que cuando solo tenemos una ecuación, el criterio de convergencia es que la derivada primera sea menor que la unidad en x*. En el caso en que tengamos dos ecuaciones $g_1(x,y)$ y $g_2(x,y)$ debe cumplirse:

$$\left.\begin{aligned}\left|\frac{\partial g_1}{\partial x}(x^*, y^*)\right| + \left|\frac{\partial g_1}{\partial y}(x^*, y^*)\right| < 1 \\ \left|\frac{\partial g_2}{\partial x}(x^*, y^*)\right| + \left|\frac{\partial g_2}{\partial y}(x^*, y^*)\right| < 1\end{aligned}\right\} \text{ ambos} \qquad \text{(ENL.30)}$$

Las principales ventajas de este método son su facilidad para programarlo, y que para ciertos tipos de problemas que aparecen en Ingeniería Química este método es muy adecuado (recirculación). La principal desventaja está en que no converge en muchos casos y en otros la convergencia es muy lenta.

5.2. Métodos de relajación (aceleración)

Para problemas donde $|\lambda|_{\max}$ está cercano a la unidad, el método de sustitución sucesiva converge lentamente. En su lugar podemos alterar la función de punto fijo F(x) con el objeto de acelerar la velocidad de convergencia:

$$x_{k+1} = w\, g(x_k) + (1-w)\, x_k \qquad \text{(ENL.31)}$$

donde el valor de *w* se adapta a los cambios en x y en F(x). Los dos métodos más importantes son el método del valor propio dominante y el método de Wegstein.

5.3. Método del valor propio dominante (DEM dominant eigenvalue method)

Este método hace una estimación de $|\lambda|_{\max}$ calculando la relación:

$$|\lambda|_{\max} \approx \frac{\|\Delta x_k\|}{\|\Delta x_{k-1}\|} \qquad \text{(ENL.33)}$$

Se puede demostrar que el valor óptimo del parámetro *w* viene dado por:

$$w = \frac{1}{1-|\lambda|_{\max}} \qquad \text{(ENL.34)}$$

(Nota: En la obtención de la ecuación anterior se admite que todos los valores propios son reales y que el valor propio máximo y mínimo no son muy diferentes. Si estas condiciones no se cumplen el método podría fallar).

5.4. Método de Wegstein

Este método obtiene el factor de relajación *w* aplicando el método de la secante (Wegstein unidimensional) a cada una de las variables x_i independientemente; es una extensión directa del método unidimensional:

$$x_{k+1}^i = w_k^i\, F(x_k^i) + (1-w_k^i)\, x_k^i \qquad \text{(ENL.35)}$$

donde: $w_k^i = \frac{1}{1 - s_k^i}$ (ENL.36)

$$s_k^i = \frac{F(x_k^i) - F(x_{k-1}^i)}{x_k^i - x_{k-1}^i} \quad \text{(ENL.37)}$$

donde 'i' representa una variable y 'k' una iteración.

El método de Wegstein funciona bien cuando no hay mucha interacción entre los componentes, por otra parte iteraciones y ciclos enlazados causan dificultades con este método.

Normalmente se suelen realizar de 2 a 5 iteraciones directas y sólo entonces se cambia a los métodos de valor propio dominante o Wegstein.

5.5. Método de Newton

Sin duda, el método de Newton es el método más extendido para resolver sistemas de ecuaciones no lineales y, sin ninguna duda, es el método que debería utilizarse –salvo casos especiales- para resolver un sistema de ecuaciones no lineal.

Consideremos un sistema de ecuaciones de la forma $\mathbf{f}(\mathbf{x}) = \mathbf{0}$, donde x es un vector de n variables reales y **f**() es un vector de n funciones reales. Si se supone un valor para las variables en un punto dado, digamos x^s es posible desarrollar en serie de Taylor alrededor del punto x^s para extrapolar la solución al punto x*. Escribiendo cada elemento del vector de funciones **f**:

$$f_i(x^*) \equiv 0 = f_i(x^*) + \left(\frac{\partial f_i}{\partial x}\right)^T \left(x^* - x^s\right) + \frac{1}{2}\left(x^* - x^s\right)^T \left(\frac{\partial^2 f_i}{\partial x^2}\right)\left(x^* - x^s\right) + \ldots \; i = 1 \ldots n$$

(ENL.38)

O bien

$$f_i(x^*) \equiv 0 = f_i(x^*) + \nabla f_i(x^s)^T \left(x^* - x^s\right) + \frac{1}{2}\left(x^* - x^s\right)^T \nabla^2 f_i(x^s)\left(x^* - x^s\right) + \ldots \; i = 1 \ldots n$$

(ENL.39)

Donde $\nabla f_i(x^s)^T$ y $\nabla^2 f_i(x^s)$ son el vector gradiente y la matriz Hessiana de la función $f_i(x)$ respectivamente. Si truncamos la serie de forma que se consideren sólo los términos de primer orden, se tiene que:

$$\mathbf{f}(\mathbf{x}^s + \mathbf{p}) \equiv 0 = \mathbf{f}(\mathbf{x}^s) + \mathbf{J}(\mathbf{x}^s)\mathbf{p} \qquad \text{(ENL.40)}$$

Donde se ha definido el vector $\mathbf{p} = (\mathbf{x}^* - \mathbf{x}^s)$ como una dirección de búsqueda hacia la solución del sistema de ecuaciones y los elementos de la matriz **J** corresponden a:

$\{J\}_{i,j} = \dfrac{\partial f_i}{\partial x_j}$ para la fila i y la columna j de dicha matriz. (ENL.41)

La matriz **J** se llama matriz *Jacobiana (o simplemente Jacobiano).* Si la matriz jacobiana no es singular, se puede resolver directamente el sistema lineal de ecuaciones:

$$\mathbf{J}(\mathbf{x}^s)\,\mathbf{p} = -\,\mathbf{f}(\mathbf{x}^s) \qquad \text{(ENL.42)}$$

O bien calcular explícitamente la inversa de la matriz jacobiana, aunque, como se comentó en le capítulo relacionado con la reslución de sistemas de ecuaciones lineales, es mucho más eficiente resolver el sistema que calcular explícitamente la inversa de la matriz Jacobiana:

$$\left[\mathbf{x}^* = \mathbf{x}^s - \mathbf{J}^{-1}(\mathbf{x}^s)\mathbf{f}(\mathbf{x}^s)\right] \qquad \text{(ENL.43)}$$

La ecuación anterior permite desarrollar una estrategia para encontrar el vector solución **x***. Comenzando con un valor supuesto inicial para el vector de variables **x** ($\mathbf{x}^0$), se puede establecer la siguiente fórmula de recursión:

$$\mathbf{J}(\mathbf{x}^k)\,\mathbf{p}^k = -\,\mathbf{f}(\mathbf{x}^k) \text{ donde } \mathbf{p}^k = \mathbf{x}^{k+1} - \mathbf{x}^k \qquad \text{(ENL.44)}$$

La fórmula de recursión anterior se puede formalizar en el siguiente algoritmo básico del método de Newton:

Algoritmo

1. Suponer un valor para el vector **x**. p.e. $\mathbf{x}^0$. Hacer k = 0
2. Calcular $\mathbf{f}(\mathbf{x}^k), \mathbf{J}\left(\mathbf{x}^k\right)$
3. Resolver el sistema de ecuaciones lineales $\mathbf{J}(\mathbf{x}^k)\,\mathbf{p}^k = -\mathbf{f}(\mathbf{x}^k)$ y calcular $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{p}^k$
4. Comprobar la convergencia: Si $\mathbf{f}^T(\mathbf{x}^k)\mathbf{f}(\mathbf{x}^k) \le \varepsilon_1$ y $(\mathbf{p}^k)^T(\mathbf{p}^k) \le \varepsilon_2$ parar. Donde $\varepsilon_1, \varepsilon_2$ son tolerancias para la terminación próximas a cero.
5. En otro caso, hacer k = k + 1 e ir al punto 1.

El método de Newton tiene unas propiedades de convergencia muy interesantes. En particular el método converge muy rápidamente en puntos cercanos a la solución. De forma más precisa se puede decir que el método de Newton tiene convergencia cuadrática, dada por la relación:

$$\frac{\left\|x^k - x^*\right\|}{\left\|x^{k-1} - x^*\right\|^2} \le K \qquad \text{(ENL.45)}$$

Donde $\|x\| = \left(x^T x\right)^{1/2}$ es la norma Euclídea, que es una medida de la longitud del vector x. Una forma de interpretar esta relación es pensar en el caso en el que K=1. Si en un momento dado se tiene una precisión para x^{k-1} de un dígito. Esto es $\left\|x^{k-1} - x^*\right\| = 0.1$. Entonces en la siguiente iteración se tendrán dos dígitos de precisión, en la siguiente 4 y en sucesivas iteraciones ocho, dieciséis etc.

Por otra parte, esta rápida velocidad de convergencia sólo ocurre si el método funciona de forma correcta. El método de Newton podría fallar por diferentes razones, sin embargo las condiciones suficientes para que el método de Newton converja son las siguientes:

1. Las funciones $\mathbf{f}(\mathbf{x})$ y $\mathbf{J}(\mathbf{x})$ existen y están acotadas para todos los valores de x.
2. El punto inicial $\mathbf{x}^0$, está suficientemente cerca de la solución.
3. La matriz **J**(**x**) debe ser no singular para todos los valores de x.

A continuación se considerará en detalle cada una de las condiciones anteriores así como las medidas a adoptar si no se cumple alguna de las condiciones.

5.6. Funciones y derivadas acotadas

Por simple inspección es posible re-escribir las ecuaciones para evitar divisiones por cero o dominios donde las funciones no están definidas. Además, es posible añadir nuevas variables y ecuaciones con el objetivo de eliminar problemas de tipo numérico. Los siguientes dos ejemplos sirven para ilustrar estos casos:

a) Resolver la ecuación $f(t)=10-e^{3/t}=0$. Para valores de t próximos a cero el valor tanto de la exponencial como de su derivada se hace excesivamente grande (tiende a infinito a medida que *t* tiende a cero). En su lugar, es posible definir una nueva variable $x=\frac{3}{t}$ y añadir la ecuación $xt-3=0$. Esto lleva a un sistema de ecuaciones mayor, pero en el que las ecuaciones están formadas por funciones acotadas. Así se resolvería el sistema:

$$\begin{aligned} f_1(x) &= 10-e^{x}=0 \\ f_2(x) &= x\,t-3\;=0 \end{aligned} \qquad \text{(ENL.46)}$$

Donde la matriz Jacobiana será: $J(x)=\begin{bmatrix} -e^{x} & 0 \\ t & x \end{bmatrix}$ (ENL.47)

Señalar que ambas funciones así como el Jacobiano permanecen acotados para valores finitos de x. Sin embargo, la matriz J todavía podría hacerse singular para ciertos valores de x y t.

Resolver la ecuación $f(x)=Ln(x)-5=0$. En este caso el logaritmo no está definido para valores no positivos de la variable x. Este problema se puede reescribir utilizando una nueva variable y una nueva ecuación. Así

pues definimos la variable $x_2 = Ln(x_1)$ y por lo tanto el nuevo sistema de ecuaciones queda como:

$$\begin{aligned} f_1 &= x_1 - e^{x_2} = 0 \\ f_2 &= x_2 - 5 \;\; = 0 \end{aligned} \qquad \text{(ENL.48)}$$

Donde la matriz Jacobiana viene dada por: $J(x) = \begin{bmatrix} 1 & -e^{x_2} \\ 0 & 1 \end{bmatrix}$

De nuevo, todas las funciones están definidas y acotadas para valores finitos de la variable x.

5.7. Cercanía a la solución

En general, asegurar que un punto está "cerca" de la solución no es práctico y además el concepto de cercanía a la solución depende de cada problema en particular. Por lo tanto, si el punto de partida es "malo", se necesita controlar de alguna manera que el método avance hacia la solución. En el método de Newton esto se puede conseguir controlando la longitud de paso en cada iteración, de tal forma que el avance hacia la solución puede venir dado por un paso completo del método de Newton o una fracción de este. Matemáticamente se puede expresar de la siguiente manera:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{p}^k \qquad \text{(ENL.49)}$$

Donde alfa es un parámetro que puede tomar valores entre cero y uno, y $\mathbf{p}^k$ es la dirección predicha por el método de Newton. Por supuesto, si α=1 se recupera el paso completo de Newton. Es necesario desarrollar ahora una estrategia que permita elegir automáticamente el valor de α de forma que asegure que el método de Newton converge.

Comenzaremos definiendo una función objetivo $\phi(\mathbf{x}) = \frac{1}{2}\mathbf{f}(x)^T\mathbf{f}(\mathbf{x})$ y buscaremos el mínimo de $\phi(\mathbf{x})$ en la dirección dada por el método de Newton en función del parámetro alfa.

Teniendo que en cuenta que $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{p}^k$ si desarrollamos en serie de Taylor:

$$\phi(\mathbf{x}^{k+1}) = \phi(\mathbf{x}^k) + \alpha \frac{d\phi}{d\alpha} + \frac{\alpha^2}{2} \frac{d^2\phi}{d\alpha^2} + ... = \\ \phi(\mathbf{x}^k) + \nabla\phi(\mathbf{x}^k)^T(\alpha \mathbf{p}^k) + \frac{\alpha^2}{2}(\mathbf{p}^k)^T \nabla^2\phi(\mathbf{x}^k)(\mathbf{p}^k) + ... \quad \text{(ENL.50)}$$

Si para simplificar la notación hacemos que: $\mathbf{J}^k = \mathbf{J}(\mathbf{x}^k)$; $\left\{J(x^k)\right\}_{ij} = \frac{\partial f_i}{\partial x_j}$ y teniendo en cuenta que:

$$\nabla\phi(\mathbf{x}^k)^T = \nabla\left(\frac{1}{2}\mathbf{f}(\mathbf{x}^k)^T\mathbf{f}(\mathbf{x}^k)\right)^T = \mathbf{f}(\mathbf{x}^k)^T\nabla\mathbf{f}(\mathbf{x}^k) \quad \text{(ENL.51)}$$

Y que un paso de Newton viene dado por:

$$\mathbf{p}^k = -\mathbf{J}(\mathbf{x}^k)^{-1}\,\mathbf{f}(\mathbf{x}^k) \quad \text{(ENL.52)}$$

Si multiplicamos por detrás la derivada de $\phi(\mathbf{x})$, ecuación (ENL.51), por $\mathbf{p}^k$ y sustituimos en el paso de Newton ecuación (ENL.52):

$$\nabla\phi(\mathbf{x}^k)\mathbf{p}^k = -\left(\mathbf{f}(\mathbf{x}^k)^T\mathbf{J}^k\left(\mathbf{J}^k\right)^{-1}\mathbf{f}(\mathbf{x}^k)\right) = -\mathbf{f}(\mathbf{x}^k)^T\mathbf{f}(\mathbf{x}^k) = -2\phi(\mathbf{x}^k) < 0 \quad \text{(ENL.53)}$$

Si sustituimos ahora en la serie de Taylor, ecuación (ENL.50), suponiendo que $\alpha \to 0$ con lo que nos podemos quedar con el término de primer orden:

$$\phi(\mathbf{x}^{k+1}) - \phi(\mathbf{x}^k) \approx -2\alpha\phi(\mathbf{x}^k) < 0 \quad \text{(ENL.54)}$$

Lo que dice la ecuación (ENL.54) es que para valores suficientemente pequeños del parámetro alfa un paso del método de Newton siempre reduce el valor de $\phi(\mathbf{x})$. Esta importante propiedad se conoce como *"propiedad de descenso del método de Newton"* y asegura que el método modificado con una longitud de paso variable siempre producirá una mejora.

El siguiente paso sería minimizar el valor de $\phi(\mathbf{x}^k + \alpha\mathbf{p}^k)$ para encontrar el valor óptimo del parámetro α que produce la máxima mejora en la dirección dada por $\mathbf{p}^k$ Sin embargo, esto es muy costoso en términos de número de

evaluaciones de las ecuaciones (o lo que es lo mismo de la función $\phi(\mathbf{x})$). En su lugar se elegirá un valor de alfa que produzca una reducción suficiente en $\phi(\mathbf{x})$. Este procedimiento se conoce *como búsqueda unidireccional de Armijo.* La Figura 6 muestra gráficamente

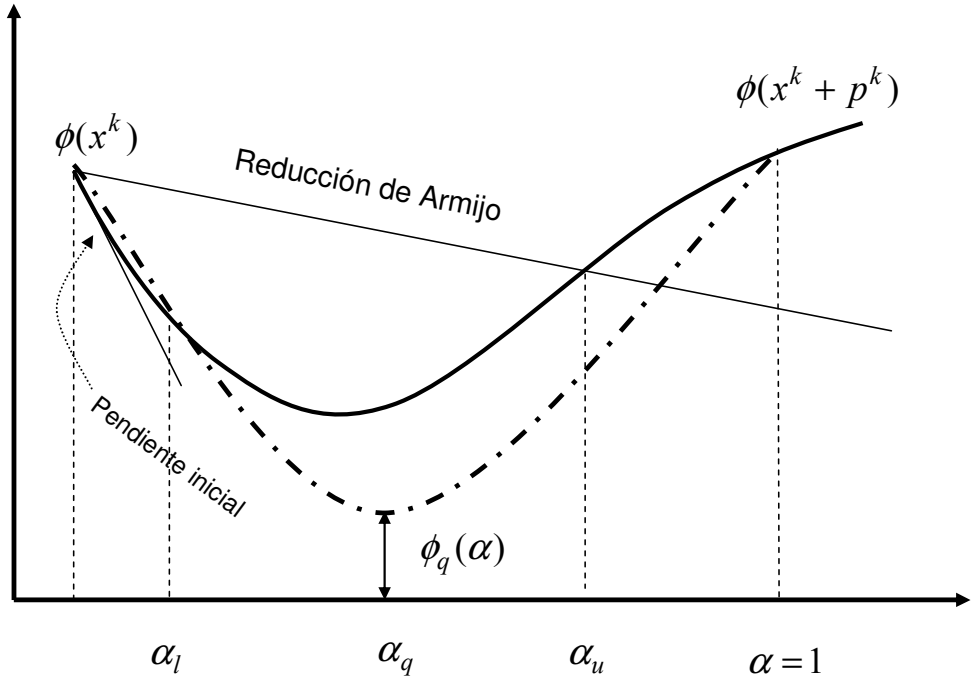

Figura 6. Esquema de la búsqueda unidireccional de Armijo.

En lugar de minimizar el valor de alfa lo que se busca es un valor de alfa que produzca una reducción suficiente, bajo el criterio de Armijo lo que se busca es un valor de alfa que haga que $\phi(\mathbf{x}^k + \alpha\mathbf{p}^k)$ caiga por debajo de la "cuerda" de Armijo, definida como:

$$\phi(\mathbf{x}^k + \alpha\mathbf{p}^k) - \phi(\mathbf{x}^k) \leq -2\,\delta\,\alpha\,\phi(\mathbf{x}^k) \qquad \text{(ENL.55)}$$

donde δ es una fracción de la pendiente (con valores típicos entre cero y 0.5) que define la pendiente de la cuerda. De esta forma se asegura una disminución satisfactoria de $\phi(\mathbf{x})$ que será al menos una fracción δ de la velocidad de reducción del error en el último punto $\mathbf{x}^k$. Si esta relación se satisface para un valor suficientemente grande de alfa, entonces se toma este paso.

Por otra parte, considérese el caso de la Figura 6, donde $\alpha = 1$. En este caso el valor de $\phi(\mathbf{x}^k + \mathbf{p}^k)$ está por encima la cuerda de Armijo, lo que significa que o bien no hay reducción de $\phi(x)$ o esta reducción no se considera

suficiente, y por lo tanto se necesita un tamaño de paso más pequeño. También necesitamos asegurarnos de que ese tamaño de paso no es excesivamente pequeño (por ejemplo en el intervalo $[\alpha_l, \alpha_u]$) para evitar que el algoritmo se quede 'estancado' con muchos movimientos excesivamente pequeños. Una forma de conseguir este objetivo es llevar a cabo una interpolación cuadrática de la función ϕ en términos del parámetro α. Dicha función de interpolación (a la que llamaremos $\phi_q(\alpha)$) se puede calcular utilizando el valor de la función en el punto $\mathbf{x}^k$, en el punto $\mathbf{x}^k + \alpha\mathbf{p}^k$ y el valor de la derivada en el punto base ($\alpha = 0$) $\frac{\partial \phi_q(0)}{\partial \alpha} = -2\phi(\mathbf{x}^k)$. Así:

$$\phi(\mathbf{x}^k + \alpha\mathbf{p}^k) = a + b\,\alpha + c\,\alpha^2 = \phi_q(\alpha) \qquad \text{(ENL.56)}$$

$$\phi_q(0) = a = \phi(\mathbf{x}^k) \qquad \text{(ENL.57)}$$

$$\left(\frac{\partial \phi_q(\alpha)}{\partial \alpha}\right)_{\alpha=0} = b = \left(\frac{\partial \phi(\mathbf{x}^k + \alpha\mathbf{p}^k)}{\partial \alpha}\right)_{\alpha=0} = -2\phi(\mathbf{x}^k) \qquad \text{(ENL.58)}$$

Por lo tanto

$$\phi(\mathbf{x}^k + \alpha\mathbf{p}^k) = \phi(\mathbf{x}^k) - 2\phi(\mathbf{x}^k)\,\alpha + c\,\alpha^2 \qquad \text{(ENL.59)}$$

De donde el valor de c vendrá dado por:

$$c = \frac{\phi(\mathbf{x}^k + \mathbf{p}^k) - \phi(\mathbf{x}^k) + 2\,\alpha\,\phi(\mathbf{x}^k)}{\alpha^2} \qquad \text{(ENL.60)}$$

El mínimo de la función cuadrática se puede obtener derivando con respecto de alfa e igualando a cero:

$$\min: \phi_q(\alpha) \rightarrow \frac{\partial \phi_q(\alpha)}{\partial \alpha} = 0 = b + 2\,c\,\alpha \rightarrow \alpha_q = -\frac{b}{2c} \qquad \text{(ENL.61)}$$

Sustituyendo las ecuaciones (ENL.58 y ENL.60) en la ecuación (ENL.61) se llega a que:

$$\alpha_q = \frac{\phi(\mathbf{x}^k)\,\alpha^2}{\phi(\mathbf{x}^k + \alpha\mathbf{p}^k) - \phi(\mathbf{x}^k) + 2\,\phi(\mathbf{x}^k)\,\alpha} \qquad \text{(ENL.62)}$$

Basado en las propiedades de salvaguarda que tiene el criterio de Armijo es posible modificar el punto 3 del método de Newton incluyendo la búsqueda unidireccional de Armijo de la siguiente manera:

Algoritmo para la búsqueda unidireccional de Armijo (modificación del punto 3 del método de Newton)

a. Hacer $\alpha = 1$
b. Evaluar $\phi(\mathbf{x}^k + \alpha\mathbf{p}^k)$
c. Si $\phi(\mathbf{x}^k + \alpha\mathbf{p}^k) - \phi(\mathbf{x}^k) \leq 2\,\delta\,\alpha\,\phi(\mathbf{x}^k)$ se ha encontrado una longitud de paso que cumple el criterio de Armijo. Hacer $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha\mathbf{p}^k$ y continuar con el punto 4 del método de Newton. En otro caso continuar con el punto d.
d. Hacer $\lambda = \max\{\eta, \alpha_q\}$ donde el valor de α_q viene dado por la ecuación (ENL.42) Hacer $\alpha = \lambda\,\alpha$ y volver la punto b.

Un valores típico de los parámetros δ y η es 0.1 (para ambos). Este procedimiento añade robustez y confiabilidad al método de Newton, especialmente si el punto inicial no es muy bueno. Sin embargo, si no se encuentra un tamaño de paso en 4 o 5 vueltas del algoritmo anterior, la dirección $\mathbf{p}^k$ podría ser una mala dirección de búsqueda debido a problemas de condicionamiento (por ejemplo un Jacobiano muy próximo a ser singular). En el caso extremo, si $\mathbf{J}(\mathbf{x}^k)$ es singular entonces la dirección de Newton no existe y el algoritmo falla. A continuación se presentan alternativas para evitar este problema

5.8. Singularidad en el Jacobiano. Modificación de la dirección de Newton

Si el Jacobiano es singular, o cercano a la singularidad (y por lo tanto mal condicionado) la dirección dada por el método de Newton es ortogonal (o

casi ortogonal) a la dirección de máximo descenso de $\phi(\mathbf{x})$. La dirección de máximo descenso viene dada por la dirección contra gradiente $(-\nabla\phi(\mathbf{x}))$ que tiene da la máxima reducción en la función $\phi(\mathbf{x})$ para longitudes de paso suficientemente pequeñas. Como consecuencia se podría considerar utilizar la dirección de máximo descenso en lugar de la dirección del método de Newton:

$$\mathbf{p}^{md} = -\nabla\phi(\mathbf{x}^k) = -\mathbf{J}(\mathbf{x}^k)^T\mathbf{f}(\mathbf{x}^k) \qquad \text{(ENL.63)}$$

Esta nueva dirección posee la propiedad de descenso, pero sólo tiene velocidad de convergencia lineal, definida como:

$$\left\|\mathbf{x}^k - \mathbf{x}^*\right\| < \left\|\mathbf{x}^{k-1} - \mathbf{x}^*\right\| \qquad \text{(ENL.64)}$$

Una ventaja del método de máximo descenso es que en tanto en cuanto se cumpla que $\mathbf{p}^{md} - \mathbf{J}(\mathbf{x}^k)^T\mathbf{f}(\mathbf{x}^k) \neq 0$ siempre se encontrará un punto mejor, incluso si el Jacobiano es singular. Sin embargo, la convergencia podría ser muy lenta.

Una situación de compromiso es combinar las direcciones de máximo descenso y de Newton. Dos de estas estrategias son el método de Levenverg-Marquardt y el método “dogleg” de Powell. En el primero de estos métodos se combinan ambas direcciones y se resuelve el siguiente sistema lineal para conseguir la dirección de búsqueda:

$$\left(\mathbf{J}(\mathbf{x}^k)^T\mathbf{J}(\mathbf{x}^k) + \lambda\mathbf{I}\right)\mathbf{p}^k = -\mathbf{J}(\mathbf{x}^k)\mathbf{f}(\mathbf{x}^k) \qquad \text{(ENL.65)}$$

Donde λ es un parámetro escalar, no negativo, que ajusta la longitud y dirección de cada paso. Para valores de $\lambda = 0$ se obtiene el método de Newton:

$$\mathbf{p}^k = -\left(\mathbf{J}(\mathbf{x}^k)^T\mathbf{J}(\mathbf{x}^k)\right)^{-1}\mathbf{J}(\mathbf{x}^k)^T\mathbf{f}(\mathbf{x}^k) = -\mathbf{J}(\mathbf{x}^k)^{-1}\mathbf{f}(\mathbf{x}^k) \qquad \text{(ENL.66)}$$

Por otra parte, si λ se hace grande y domina al valor $\mathbf{J}(\mathbf{x}^k)^T\mathbf{J}(\mathbf{x}^k)$ el sistema de ecuaciones se aproxima a:

$$\mathbf{p}^k = -\left(\lambda \mathbf{I}\right)^{-1} \mathbf{J}(\mathbf{x}^k)^T \mathbf{f}(\mathbf{x}^k) = -\frac{\mathbf{J}(\mathbf{x}^k)^T \mathbf{f}(\mathbf{x}^k)}{\lambda} \qquad \text{(ENL.67)}$$

Que es la dirección de máximo descenso con una longitud de paso muy pequeña. Con un valor intermedio de λ se obtiene una dirección que cae en un punto intermedio entre la dirección dada por el método de Newton y la de máximo descenso.

Una desventaja del método de Levenverg y Marquardt es que cada vez que se cambia el valor de λ es necesario resolver un nuevo sistema de ecuaciones lineales, lo que puede ser numéricamente costoso, sobre todo si se tiene en cuenta que es posible que sea necesario probar varios valores de λ antes de conseguir una longitud de paso adecuada. En su lugar consideraremos un algoritmo que utiliza una combinación entre los métodos de Newton y de máximo descenso y elige la dirección entre ellos de forma automática. Este método conocido como "dogleg" (a algunos autores la forma de generar la dirección le recuerda la curva de la pata de un perro) fue desarrollado por Powell. Y gráficamente se ilustra en la Figura 7:

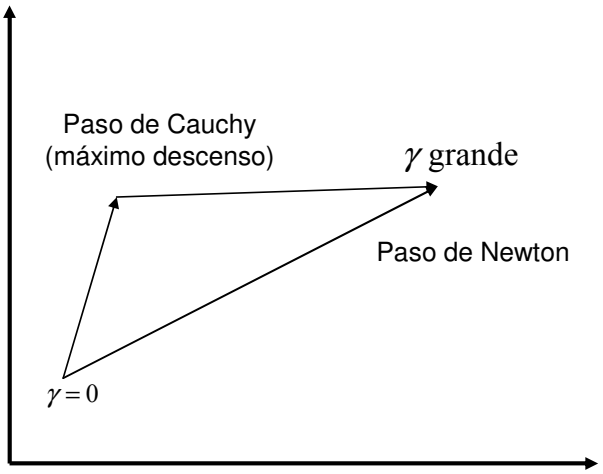

Figura 7 Ilustración del Método Dogled de Powell

Aquí el paso más largo es debido al método de Newton, y el más pequeño a una combinación entre el método de Newton y el de máximo descenso. Para pasos más pequeños que el dado por la dirección de máximo descenso, se conserva la dirección de máximo descenso.

Para desarrollar este método se necesita primero encontrar la longitud correcta (dada por un escalar β) a lo largo de la dirección de máximo descenso:

$$\mathbf{p}^{md} = -\mathbf{J}^T \mathbf{f}(\mathbf{x}^k) \qquad \text{(ENL.68)}$$

Para calcular dicho valor consideramos la minimización de un modelo cuadrático formado por las ecuaciones linealizadas a lo largo de la dirección de máximo descenso:

$$\min_{\beta} : \frac{1}{2}\left(\mathbf{f}(\mathbf{x}^k) + \beta\, \mathbf{J}\, \mathbf{p}^{md}\right)^T \left(\mathbf{f}(\mathbf{x}^k) + \beta\, \mathbf{J}\, \mathbf{p}^{md}\right) \qquad \text{(ENL.69)}$$

Sustituyendo la dirección de máximo descenso ecuación (ENL.68) en la anterior se tiene que:

$$\beta = \frac{\left[\mathbf{f}(\mathbf{x}^k)^T \mathbf{J}\mathbf{J}^T \mathbf{f}(\mathbf{x}^k)\right]}{\left[\mathbf{f}(\mathbf{x}^k)^T \mathbf{J}(\mathbf{J}^T\mathbf{J})\mathbf{J}^T \mathbf{f}(\mathbf{x}^k)\right]} = \frac{\left\|\mathbf{p}^{md}\right\|^2}{\left\|\mathbf{J}\, \mathbf{p}^{md}\right\|^2} \qquad \text{(ENL.70)}$$

El paso $\beta\, \mathbf{p}^{md}$ es conocido como *paso de Cauchy*, y se puede demostrar que su longitud nunca es mayor que la longitud dada por un paso del método de Newton ($\mathbf{p}^N = -\mathbf{J}^{-1}\mathbf{f}(\mathbf{x}^k)$). Para una longitud de paso γ para el paso global, se puede calcular la dirección del método dogleg de Powell de la siguiente manera, donde la longitud de paso γ se puede ajustar de forma automática:

- Para $\gamma \le \beta \left\|\mathbf{p}^{md}\right\|$; $\mathbf{p} = \gamma \dfrac{\mathbf{p}^{md}}{\left\|\mathbf{p}^{md}\right\|}$
- Para $\gamma \ge \left\|\mathbf{p}^{N}\right\|$; $\mathbf{p} = \mathbf{p}^N$
- Para $\left\|\mathbf{p}^{N}\right\| > \gamma > \beta \left\|\mathbf{p}^{md}\right\|$; $\mathbf{p} = \eta \mathbf{p}^N + (1-\eta)\beta\, \mathbf{p}^{md}$

donde $\eta = \dfrac{\gamma - \beta \left\|\mathbf{p}^{md}\right\|}{\left\|\mathbf{p}^{N}\right\| - \beta \left\|\mathbf{p}^{md}\right\|}$

Finalmente señalar que los métodos de Levenverg y Marquardt y el el método dogleg caen dentro de una clase general de algoritmos denominados de "*región de confiabilidad*" para estos problemas, la longitud de paso γ corresponde al tamaño de la región alrededor del punto $\mathbf{x}^k$ para el que se confía que el modelo cuadrático es una aproximación precisa de $\phi(\mathbf{x})$. Una

minimización aproximada de este modelo cuadrático necesita ajustar λ o η (dependiendo del método, Levenverg y Marquardt, o dogleg) en cada iteración. Si bien los métodos basados en regiones de confiabilidad son más costosos en términos numéricos que la búsqueda unidireccional de Armijo, presentan características de convergencia mucho mejores, particularmente para problemas mal condicionados.

5.8. Métodos cuasi-newton. El método de Broyden

Los métodos cuasi-Newton tienen en común que tratan de evitar el cálculo de la matriz de derivadas parciales en cada iteración. La matriz se estima al principio del procedimiento y se actualiza en cada iteración. Existen varios métodos cuasi-Newton, el más utilizado de todos es el debido a Broyden. La función de iteración de este método es:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k \mathbf{f}(\mathbf{x}_k) \qquad \text{(ENL.71)}$$

donde $\mathbf{H}_k$ es la k-esima estimación de la inversa de la matriz jacobiana.

Si desarrollamos en serie de Taylor $f_i(x_{k+1})$ en los alrededores de x_k:

$$f_i(x_{k+1}) = f_i(x_k) + \frac{\partial f_i}{\partial x_1}(x_{1,k+1} - x_{1,k}) + \frac{\partial f_i}{\partial x_2}(x_{2,k+1} - x_{2,k}) + \ldots + \frac{\partial f_i}{\partial x_n}(x_{n,k+1} - x_{n,k})$$

Si suponemos conocidos los valores de $(\mathbf{x}_k, \mathbf{f}(\mathbf{x}_k))$ y $(\mathbf{x}_{k+1}, \mathbf{f}(\mathbf{x}_{k+1}))$ como un método de secante, entonces llamando

$$\mathbf{p}_k = \mathbf{x}_{k+1} - \mathbf{x}_k = \Delta x_k; \quad \mathbf{y}_k = \mathbf{f}(\mathbf{x}_{k+1}) - \mathbf{f}(\mathbf{x}_k) \qquad \text{(ENL.72)}$$

De esta forma la ecuación para $f_i(x_{k+1})$ se puede reescribir como:

$$\mathbf{y}_k = \mathbf{B}_{k+1}\mathbf{p}_k \qquad \text{(ENL.73)}$$

donde $\mathbf{B}_{k+1}$ es la estimación del jacobiano que estamos buscando. La matriz $\mathbf{B}_{k+1}$ la podemos descomponer en suma de dos, de tal forma que:

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mathbf{D}_k \qquad \text{(ENL.74)}$$

sustituyendo obtenemos:

$$\mathbf{y}_k = (\mathbf{B}_k + \mathbf{D}_k)\mathbf{p}_k \quad \text{(ENL.75)}$$

$$\mathbf{y}_k = \mathbf{B}_k \mathbf{p}_k + \mathbf{D}_k \mathbf{p}_k \quad \text{(ENL.76)}$$

$$\mathbf{D}_k\mathbf{p}_k = \mathbf{y}_k - \mathbf{B}_k\mathbf{p}_k \quad \text{(ENL.77)}$$

La parte derecha de esta última ecuación contiene solamente vectores conocidos, la ecuación tiene que ser resuelta para $\mathbf{D}_k$. Esta ecuación tiene infinito número de soluciones de la forma:

$$\mathbf{D}_k = \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{p}_k)\mathbf{z}^T}{\mathbf{z}^T\mathbf{p}_k} \quad \text{(ENL.78)}$$

donde $\mathbf{z}^T$ es un vector arbitrario.

Por lo tanto la matriz $\mathbf{B}_{k+1}$ que estábamos buscando será:

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{p}_k)\mathbf{z}^T}{\mathbf{z}^T\mathbf{p}_k} \quad \text{(ENL.79)}$$

Esta no es la forma final todavía, ya que la ecuación anterior proporciona una manera de estimar la matriz jacobiana, pero nos interesa la matriz inversa de la jacobiana, para ello utilizamos la fórmula debida a Sherman-Morrison, que dice:

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \frac{(\mathbf{H}_k\mathbf{y}_k - \mathbf{p}_k)\mathbf{z}_k^T\mathbf{H}_k}{\mathbf{z}_k^T\mathbf{H}_k\mathbf{y}_k} \quad \text{(ENL.80)}$$

Broyden encontró por experimentación numérica que el mejor valor para $\mathbf{z}^T$ es $\mathbf{p}_k^T$, por lo tanto la forma final de la ecuación para calcular la matriz $\mathbf{H}_{k+1}$ es:

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \frac{(\mathbf{H}_k\mathbf{y}_k - \mathbf{p}_k)\mathbf{p}_k^T\mathbf{H}_k}{\mathbf{p}_k^T\mathbf{H}_k\mathbf{y}_k} \quad \text{(ENL.81)}$$

como estimación inicial de $\mathbf{H}_0$ hay varias opciones, o bien se hace $\mathbf{H}_0=\mathbf{I}$ (equivalente a la sustitución sucesiva para la primera iteración), se calcula la inversa del verdadero Jacobiano en la primera iteración, se toma como

primera aproximación del Jacobiano una matriz diagonal con los valores correspondientes a la diagonal principal del Jacobiano.

El algoritmo se puede resumir en los siguientes pasos:

1.- Especificar los valores iniciales para $\mathbf{x}_0$ y $\mathbf{H}_0$

2.- Calcular $\mathbf{f}(\mathbf{x}_0)$

3.- Calcular $\mathbf{x}_{k+1}$ y $\mathbf{f}(\mathbf{x}_{k+1})$

4.- Calcular $\mathbf{y}_k$, $\mathbf{p}_k$, y $\mathbf{H}_{k+1}$

5.- Chequear el criterio de finalización (por ejemplo $\|\mathbf{x}_{k+1}\| < \varepsilon_d$), si se satisface terminar, en caso contrario incrementar el contador k=k+1 y volver a la etapa 3.

El método de Broyden se usa para problemas a gran escala. Es también muy adecuado para usar en sistemas de recirculación.

5.9. Métodos homotópicos o de continuación

Debido a las dificultades de convergencia global de los métodos tipo Newton o cuasi-Newton, se introdujeron a finales de los años 70 principios de los 80 lo que se conoce como métodos homotópicos o de continuación. Aunque dichos métodos se han venido refinando hasta prácticamente hoy en día.

La idea básica de los métodos de continuación consiste en introducir las ecuaciones del modelo f(x) en una combinación lineal de la forma:

$$H(x,\theta)=\theta f(x)+(1-\theta)g(x)=0 \qquad \text{(ENL.82)}$$

donde x es el vector de variables del problema. θ es el parámetro homotópico y g(x) una función de la que se conoce su solución (g(x)=0) o es fácilmente resoluble. Existen diferentes opciones para g(x), produciendo diferentes comportamientos del método. Algunas alternativas son:

$$H(x,\theta)=\theta f(x)+(1-\theta)(x-x_0) \qquad \textit{Homotopía de punto fijo}$$
$$H(x,\theta)=\theta f(x)+(1-\theta)\big(f(x)-f(x_0)\big) \qquad \textit{Homotopía de Newton}$$
$$H(x,\theta)=\theta f(x)+(1-\theta)A(x-x_0) \qquad \textit{Homotopía afin}$$

donde A es una matriz para prevenir problemas de condicionamiento por el escalado, típicamente se hace igual a $f'(x_0)$. Las diferentes funciones homotópicas presentadas anteriormente tienen la peculiaridad de que el error de los componentes de $f(x)$ decrece linealmente desde los valores dados por x_0.

La función homotópica más utilizada es la de punto fijo, que además de su simplicidad evita complicaciones adicionales causadas por la multiplicidad adicional que pueden presentarse al añadir funciones no lineales.

Una forma de resolver el problema es ir dando valores al parámetro θ desde 0 hasta 1 en intervalos pequeños utilizando el resultado de cada paso como valor inicial para el siguiente. Sin embargo, suele ser mejor reformular el sistema de ecuaciones no lineales como un sistema de ecuaciones diferenciales de valor inicial. Así, derivando la función homotópica con respecto al parámetro θ:

$$\frac{d\,H(x,\theta)}{d\,\theta} = \frac{\partial H}{\partial x}\frac{d\,x}{d\,\theta} + \frac{\partial H}{\partial \theta} \qquad \text{(ENL.83)}$$

dado que se conoce el valor de H para x_0, se puede integrar la ecuación anterior desde θ=0 hasta θ=1, punto donde se debe cumplir que f(x)=0.

Si el sistema tuviese más soluciones extendiendo el intervalo de integración más allá de los valores $0 \le \theta \le 1$ se puede encontrar, en muchos casos, otras soluciones al problema.

Ejemplo de aplicación de un método homotópico

Se desea calcular el volumen molar de CO_2 a 298 K y 50 atm utilizando la ecuación de estado de Redlich-Kwong

$$P = \frac{RT}{V-b} - \frac{a}{V(V-b)\sqrt{T}} \qquad \begin{cases} a = 0.42747\left(\dfrac{R^2\,T_c^{5/2}}{Pc}\right) \\ b = 0.08664\left(\dfrac{R\,T_c}{P_c}\right) \end{cases}$$

R = 0.08206 atm l mol^{-1} K^{-1}
Pc = Presión crítica, atm
Tc = temperatura crítica, K
V = Volumen molar, l mol^{-1}

Para el CO_2 la presión crítica es de 72.9 atm y la temperatura crítica de 304.2 K.

Para aplicar el método de Newton reescribimos la ecuación igualada a cero:

$$f(V) = \frac{RT}{V-b} - \frac{a}{V(V-b)\sqrt{T}} - P = 0$$

Dependiendo de cual sea el punto inicial un algoritmo básico de Newton (sin búsqueda unidireccional ni región de confiabilidad) puede converger o no. Por ejemplo partiendo de los siguientes valores

V_0 = 1 litro/mol	no converge
V_0 = 0.1 litros/mol	no converge
V_0 = 0.2 litros/mol	converge a V = 0.33354 litros/mol

Si se aplica la homotopía de Newton partiendo de V0 = 1 litro/mol:

$$f(V_0) = -28.38382209795442$$

$$H(V,\theta) = f(V) - \theta\, f(V_0) = 0 = \frac{RT}{V-b} - \frac{a}{V(V-b)\sqrt{T}} - P - \theta\, f(V_0)$$

$$\frac{dH(V,\theta)}{d\theta} = \frac{\partial H}{\partial V}\frac{dV}{d\theta} + \frac{\partial H}{\partial \theta} = 0$$

$$\left.\begin{aligned} \frac{\partial H}{\partial V} &= \frac{-RT}{(V-b)^2} + \frac{a\sqrt{T}(2V+b)}{\left[V(V+b)\sqrt{T}\right]^2} = g(V) \\ \frac{\partial H}{\partial \theta} &= -f(V_0) \end{aligned}\right\} \; g(V)\frac{dV}{d\theta} - f(V_0) = 0$$

Despejando:

$$\frac{dV}{d\theta} = \frac{f(V_0)}{g(V)}$$

Integrando numéricamente entre 1 y 0 se obtiene:

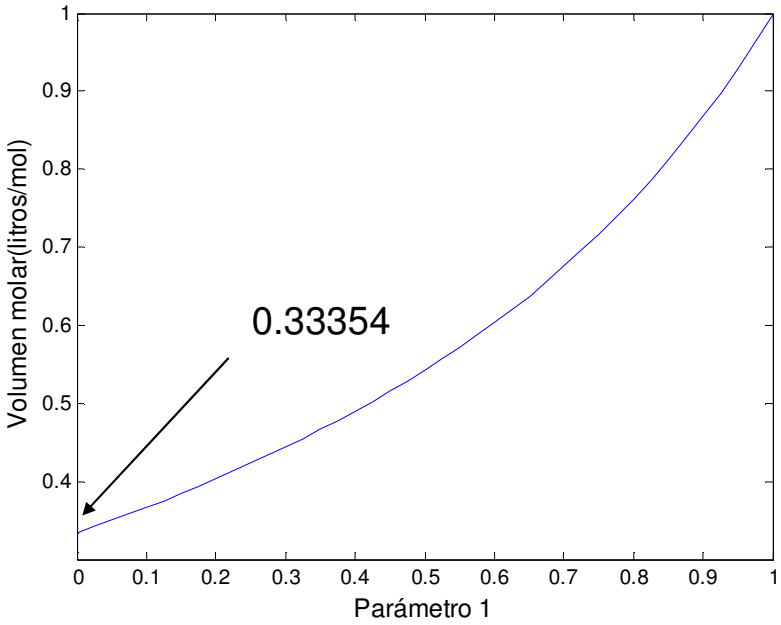

Figura 8 Ejemplo de homotopía de Newton. Ec de Redlich-Kwong para CO_2

6. RESUMEN

La discusión y resolución de sistemas de ecuaciones lineales, empleando distintos procedimientos, completa el estudio de la solución de ecuaciones y sistemas. Con esta unidad se pretende que el alumnado aplique lo estudiado en las unidades anteriores a la discusión y resolución de los sistemas de ecuaciones.

Comienza con la discusión de los métodos más comunes, como es el de sustitución sucesiva. Posteriormente, como paso previo a su resolución en los casos en que sea posible, se discuten lo métodos de relajación. Por último, se describen los métodos de Newton y Broyden, este último adecuado para evitar el complejo cálculo de la matriz jacobiana.

El dominio de los métodos para discutir y resolver un sistema de ecuaciones lineales y no lineales permitirá al alumnado afrontar el planteamiento y resolución de problemas diversos.

7. Programación en Matlab®

7.1. Métodos tipo F(x)=x: iteracion directa

$$x_{i+1} = F(x_i) \quad \text{i=1,2,3...}$$

```
function x=sustsuc(F,x0,tol)

x1=feval(F,x0);
error=norm(x1-x0);
contador=0;

% Comprobamos la convergencia:
J=jacobiano(F,x0);
Jp=abs(J);
sumas=sum(Jp,2);% Vector columna con las sumas de las filas de Jp
A=(sumas>=1);
if sum(A)>=1
   disp('A lo mejor no converge')
end


while error>=tol
   x2=feval(F,x1);
   error=norm(x2-x1);
   x1=x2;
   contador=contador+1;
end
x=x1;
contador
```

7.2. Métodos tipo F(X)=X: Wegstein

$$x_{i+1} = \frac{1}{1-s_i} F(x_i) - \frac{s_i}{1-s_i} x_i = w_i\, F(x_i) + (1-w_i)\, x_i$$

$$s_i = \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}}; \qquad w_i = \frac{1}{1-s_i}$$

```
function x=wegstein(F,x0,tol)

x1=feval(F,x0);
error=abs(x1-x0);
contador=0;

while error>=tol
   s=(feval(F,x1)-feval(F,x0))/(x1-x0);
   w=1/(1-s);
   x2=w*feval(F,x1)+(1-w)*x1;
   error=abs(x2-x1);
   x0=x1;
   x1=x2;
   contador=contador+1;
end
x=x1;
contador
```

7.3. Método de Wegstein adaptado a la solución de sistemas de ecuaciones

```
function x=wegsteins(F,x0,tol)

x1=feval(F,x0);
```

```
error=norm(x1-x0);
contador=0;

while error>=tol
   s=(feval(F,x1)-feval(F,x0))./(x1-x0);
   w=1./(1-s);
   x2=w.*feval(F,x1)+(1-w).*x1;
   error=norm(x2-x1);
   x0=x1;
   x1=x2;
   contador=contador+1;
end
x=x1;
contador
```

7.4. Ejemplo para Wegstein sistemas

```
function F=ejemplo18(X)
x=X(1);
y=X(2);
z=X(3);
F(1)=(x^2-2*x+y^2-z+3)^(1/3);
F(2)=(0.5*z^2-x+y+z+5)^(1/3);
F(3)=(0.5*x*y+0.1*z^2);
F=F'; %Para que salga como columna
```

7.5. Calculo del Jacobiano de una función

```
function J=jacobiano(F,x)
incx=x*(eps^0.5);
dincx=diag(incx);
Fx=feval(F,x);
J=[];

for i=1:length(x)
   xincx=x+dincx(:,i); %Solo incrementamos la variable 'i'
   Fxincx=feval(F,xincx);
   Deriva=(Fxincx-Fx)/incx(i);
   J=[J Deriva];
End
```

7.6. Métodos tipo f(x)=0: Newton

$$\begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}_{\substack{x=x_k \\ y=y_k}} \begin{bmatrix} \Delta x_k \\ \Delta y_k \end{bmatrix} = -\begin{bmatrix} f_1(x_k, y_k) \\ f_2(x_k, y_k) \end{bmatrix}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{J}_k^{-1}\,\mathbf{f}(\mathbf{x}_k)$$

```
function x=newton(F,x0,tol)
% La funcion F debe estar en la forma f(x)=0

J=jacobiano(F,x0);
x1=x0-inv(J)*feval(F,x0);
error=norm(x1-x0);
```

```
contador=0;

while error>=tol
   J=jacobiano(F,x1);
   x2=x1-inv(J)*feval(F,x1);
   error=norm(x2-x1);
   x1=x2;
   contador=contador+1;
end
x=x1;
contador
function F=ejemplonewton(X)
x=X(1);
y=X(2);
F(1)=2*x+y^2+x*y-1;
F(2)=x^3+x^2+3*y-2;
F=F';
```

7.7. Métodos tipo f(x)=0: Broyden

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k \mathbf{f}(\mathbf{x}_k)$$

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \frac{(\mathbf{H}_k \mathbf{y}_k - \mathbf{p}_k)\mathbf{z}_k^T \mathbf{H}_k}{\mathbf{z}_k^T \mathbf{H}_k \mathbf{y}_k}$$

$$\mathbf{p}_k = \mathbf{x}_{k+1} - \mathbf{x}_k = \Delta x_k; \quad \mathbf{y}_k = \mathbf{f}(\mathbf{x}_{k+1}) - \mathbf{f}(\mathbf{x}_k)$$

H0= Identidad o Jacobiano del punto inicial

```
function x=broyden(F,x0,tol)
% La funcion F debe estar en la forma f(x)=0
n=length(x0);% Numero de ecuaciones
H0=eye(n);

x1=x0-H0*feval(F,x0);
error=norm(x1-x0);
contador=0;

while error>=tol
   p0=x1-x0;
   y0=feval(F,x1)-feval(F,x0);
   H1=H0-((H0*y0-p0)*p0'*H0)/(p0'*H0*y0);
   x2=x1-H1*feval(F,x1);
   error=norm(x2-x1);
   x0=x1;
   x1=x2;
   contador=contador+1;
end
x=x1;
contador
```