CAPÍTULO III

REVISIÓN DE MÉTODOS NUMÉRICOS APLICABLES EN SIMULACIÓN DE PROCESOS EN ESTADO ESTACIONARIO

Por
H. J. Espinosa, P. Aguirre y G. A. Pérez

III.1 CONCEPTOS BÁSICOS

La solución de una ecuación no lineal o de una función trascendente, como método, supone la búsqueda de un valor tal que satisfaga la ecuación o función en cuestión.

Si *f (x)* es una función no lineal genérica y x^* es la solución o raíz, entonces se cumple que

$$f(x^*) = 0 \tag{1}$$

Otra forma alternativa de formular el problema es:

$$x^* = F(x^*) \tag{2}$$

donde *F(x)* es una función diferente de la original, y supone que puede explicitarse de la primitiva, cumpliéndose para ello el teorema de funciones implícitas.

Una *solución iterativa* significa comenzar con un valor inicial, x_0, y generar una sucesión (secuencia) $x_0, x_1, x_2, ..., x_n$, tal que:

$$\left\{ x_n \right\} / \lim_{n \to \infty} x_n = x^* \tag{3}$$

donde *n* representa el número de iteración (término de la correspondiente sucesión).

El *error exacto en la iteración n* es:

$$E_n = |x_n - x^*| \tag{4}$$

Las hipótesis usuales son que *f(x)*, *F(x)* y x^* satisfacen las siguientes restricciones:

i.- x^* se sitúa dentro de: $I = [a, b]/f(x) \wedge F(x) \varepsilon \mathbb{C}_1$.

Modelado, Simulación y Optimización de Procesos Químicos
Autor: Nicolás J. Scenna y col.
ISBN: 950-42-0022-2 - ©1999

ii.- Hay una sola raíz (única) en I y es real ($x^* \varepsilon \mathbb{R}$).

donde el símbolo $\wedge$ representa el operador lógico *y* y C_1: funciones continuas.

A pesar de que no se diga, todos los métodos usuales se basan en dichas hipótesis. De no ser así se requerirán métodos especiales.

III.2 MÉTODOS BÁSICOS. DISCUSIÓN DE LA CONVERGENCIA

En general, los métodos de mayor *orden de convergencia* llegan a la solución en un número menor de iteraciones. Las excepciones se producen al inicializarlos (o proponer las primeras estimaciones) con valores numéricos muy inapropiados. De todos modos, el menor número de iteraciones no necesariamente significa el menor tiempo computacional, dado que éste depende del esfuerzo de cálculo involucrado en cada iteración. El cálculo de *f(x)* es el mayor consumidor de tiempo junto con el cálculo de su derivada, *que para el caso de una ecuación o función única pueden suponerse equivalente*. El número de estas evaluaciones suele ser una medida más adecuada de la eficiencia del método de solución.

Los métodos iterativos hallan la solución exacta (si eso ocurre) *sólo en un número infinito de iteraciones* (Ecuación (3)). En la práctica, las iteraciones se detienen cuando el error es menor que una adecuada tolerancia, impuesta por el usuario. El valor exacto del error (Ecuación (4)) no puede usarse como criterio de terminación, porque normalmente no se conoce.

Por lo tanto, como criterio de finalización se usa una estimación del denominado error exacto, y es así como interviene la tolerancia de error adoptada, E_d.

Los criterios generalmente usados son:

$$|f(x_n)| < E_d \tag{5.1}$$

$$| x_n - x_{n-1} | < E_d \tag{5.2}$$

$$| x_n - x_{n-1} | < E_d \, |x_n| \tag{5.3}$$

Los dos primeros comparan errores absolutos, mientras que el tercero analiza errores relativos. Estos criterios pueden tener limitaciones, como veremos, y existen propuestas alternativas que resuelven el problema.

Una comparación de los métodos más usados puede verse en la tabla siguiente:

ISBN: 950-42-0022-2 - ©1999

Tabla III.1: Métodos más usuales.

MÉTODO	ORDEN DE CONVERGENCIA	INFORMACIÓN PARA CALCULAR x_{n+1}
BISECCIÓN	LINEAL (1)	$f(x_n)$, $f(x_{n-1})$
SECANTE	SUPERLINEAL (1.618)	$f(x_n)$, $f(x_{n-1})$
NEWTON-RAPHSON	CUADRÁTICO (2)	$f(x_n)$, $f'(x_n)$
SUSTITUCIÓN DIRECTA	LINEAL (1)	$F(x_n)$

Evidentemente, en nuestro campo de aplicaciones importan las raíces reales de $f(x) = 0$. El hecho de analizar una función trascendente primero (o una función o ecuación no lineal) es debido a la relación con los sistemas de ecuaciones, tanto lineales como no lineales, y también porque introduce y educa en la generación de *algoritmos*, y en los problemas asociados con la convergencia: factibilidad y velocidad.

Un concepto básico asociado con los métodos iterativos es el llamado *orden de convergencia* del método. Si definimos el error en la iteración *n*, como E_n, entonces si existe un número real $p \geq 1$ tal que:

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^p} \equiv \lim_{n\to\infty} \frac{|E_{n+1}|}{|E_n|^p} = K \neq 0 \qquad (6)$$

se dice que *el método es de orden p*, en x^*. La constante *K* se llama *constante de error asintótica*, y depende de *f(x)*. A mayor orden de convergencia, el método convergirá a mayor velocidad. Pero no implica, obviamente, garantía de convergencia.

Otra medida usual es aquella que se hace para saber cuánto se debe computar, para alcanzar una cierta precisión en la raíz buscada. En este caso, además del orden interviene el *costo computacional por iteración* que es importantísimo para definir la *eficiencia* de un método.

III.3 PRINCIPALES MÉTODOS

III.3.1 El Método de Bisección

Es interesante considerarlo porque naturalmente es lo primero que uno haría, si fuera capaz de graficar (o conocer la gráfica) de la función *f(x)* en cuestión. Además, su basamento se puede probar, y dentro de ciertas pautas da garantía de éxito.

Si las raíces de interés práctico son aisladas, esto es los valores de x han sido divididos en distintos intervalos, en cada uno de los cuales sospechamos que se encuentra alguna raíz de importancia para nuestro fines, posibilita garantía de convergencia a la misma. Existe un teorema, que dice: *Si f (x) es continua desde x=a hasta x=b, y si f (a) y f (b) tienen signos opuestos, luego hay como mínimo una raíz*

ISBN: 950-42-0022-2 - ©1999

real de f (x) = 0, entre a y b.

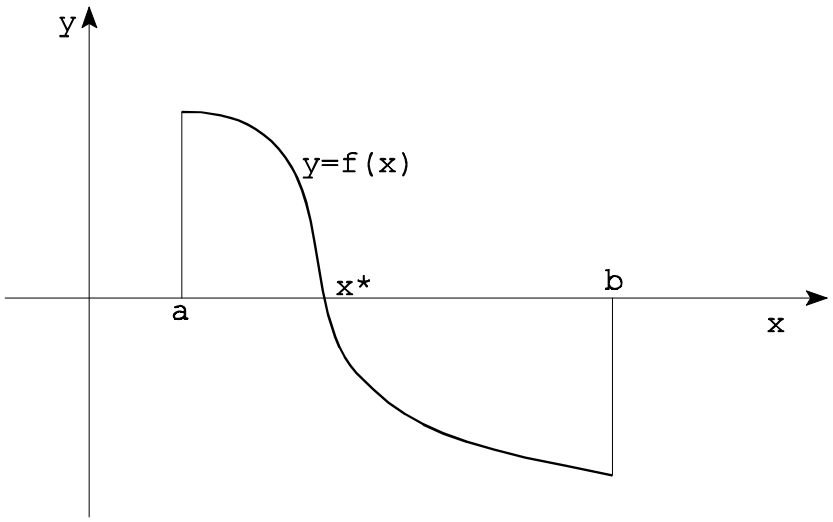


Figura III.1: Raíz aislada de una ecuación.

Basado en el teorema anterior, supongamos (tal como se aprecia en la Fig. (III.1)) que f (x) es continua y positiva en $x = a$ y negativa en $x = b$. Luego, habrá una raíz entre *a* y *b*. Si, entonces, dividimos el intervalo $I = [a, b]$ por la mitad, podemos calcular $f\,[(a+b)/2]$, teniendo tres posibilidades, a saber:

i.- Que sea cero (dentro de cierto error), en cuyo caso es la raíz.

ii.- Que sea negativo, en cuyo caso la raíz está entre $x = (a+b)/2$ y $x = a$.

iii.- Que sea positivo, entonces se encuentra entre $x = (a+b)/2$ y $x = b$.

De modo que el procedimiento continua bisectando el intervalo cada vez, hasta lograr la raíz dentro de la precisión deseada.

Cada repetición reduce el error máximo por un factor de *2* (dos), de manera que tres iteraciones producen una mejora aproximada de un orden de magnitud. Su gran virtud es que *asegura la convergencia*, que constituye una propiedad que no encontramos en los otros métodos. A cambio de ello, puede ser tremendamente lento lo que lo puede transformar en ineficiente, debido a que es lineal (orden 1).

Otro método que converge para toda función continua es el denominado *Método de la Falsa Posición o Regula Falsi* (ver Figura (III.2)).

Mecánica (algoritmo)

x_1 y x_2/$f(x_1).f(x_2) < 0$

ISBN: 950-42-0022-2 - ©1999

luego → x_3

si: $f(x_3).f(x_1) < 0 \; ; \; i = 1, 2$

(en este caso $i = 1$)

se repite x_4 y continúa.

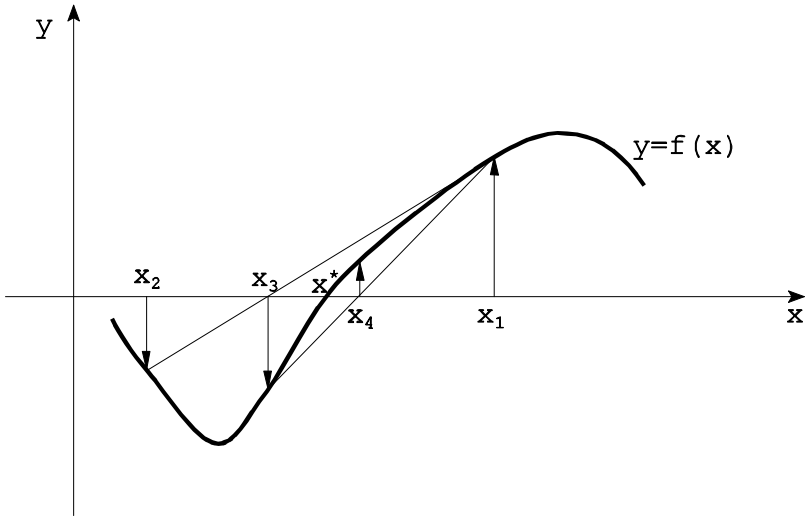


Figura III.2: Método Regula Falsi.

Se ve que:

$$x_3 = \frac{y_2}{y_2 - y_1} x_1 + \frac{y_1}{y_1 - y_2} x_2 \qquad (7.1)$$

$$x_4 = \frac{y_3}{y_3 - y_1} x_1 + \frac{y_1}{y_1 - y_3} x_3 \qquad (7.2)$$

Si *f(x)* es *cóncava* entre x_1 y x_2 se dice que el método es *estacionario*; esto es, el punto x_1 es siempre uno de los dos puntos usados para la siguiente iteración (ver Figura (III.3)). Lo mismo ocurriría si fuese convexa en las inmediaciones de la raíz, provocando una convergencia lineal en estos casos.

Una mejora que lo hace más eficiente consiste en aplicar la fórmula hallada (Ecuación (7.2)) a los puntos x_{i-1} y x_{i+1}, pero reemplazando

$$y_{i-1} \; por \; \alpha \; y_{i-1} = \bar{y}_{i-1}, \; tal \; que \; 0 < \alpha < 1 \qquad (8)$$

ISBN: 950-42-0022-2 - ©1999

Esto es:

$$x_{i+2} = \frac{\alpha\, y_{i-1}}{\alpha\, y_{i-1} - y_{i+1}}\, x_{i+1} + \frac{y_{i+1}}{y_{i+1} - \alpha\, y_{i-1}}\, x_{i-1} \qquad (9)$$

Si cambia el signo debe retornarse el método original, de lo contrario se continúa con la mejora presentada. Las elecciones más simples para α son:

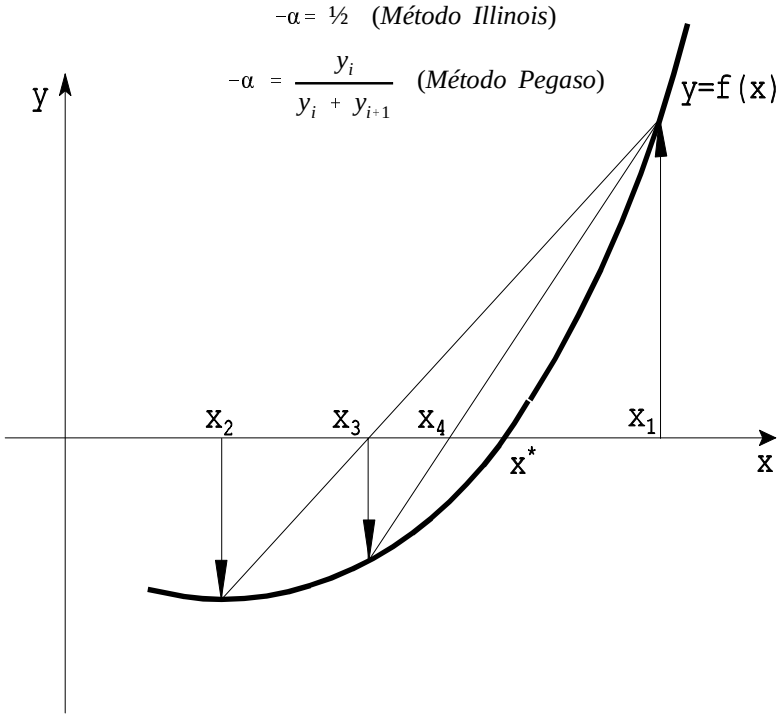


Figura III.3: Método Regula Falsi para funciones cóncavas.

Estas mejoras incrementan el orden, y los hace superlineales (Illinois: *1.44* y Pegaso: *1.64*). Son algoritmos muy buenos cuando tenemos dos puntos en los cuales *f(x)* tiene signo opuesto. En caso de no querer hacer un esfuerzo computacional en hallarlos, se pueden usar *interpolación lineal inversa* con los dos últimos puntos computados para generar el siguiente. En este caso resulta:

ISBN: 950-42-0022-2 - ©1999

$$x_{i+1} = \frac{y_i}{y_i - y_{i-1}} x_{i-1} + \frac{y_{i-1}}{y_{i-1} - y_i} x_i \qquad (10)$$

A este método estacionario se lo denomina *método de la secante*. Requiere de dos aproximaciones iniciales (recordar Tabla (III.1)), y sólo si son cercanas a la raíz el método es convergente. Pero tiene la ventaja de ser super lineal (orden de convergencia = *1.618*). Su nombre se debe, según la Ecuación (10), a que se puede escribir:

$$x_{i+1} = x_i - \frac{x_i - x_{i-1}}{y_i - y_{i-1}} y_i \qquad (11)$$

expresión que será de utilidad cuando discutamos métodos para sistemas de ecuaciones.

En la Ecuación (11) se pone de manifiesto un problema típico de errores en cálculo numérico. Se ve que cerca de la solución y_i e y_{i-1} son cantidades parecidas con lo que aparece el *problema de cancelación sustractiva* en ese segundo término; que es un término de corrección que estará aportando muy pocos dígitos significativos, lo que obliga a usar doble precisión (término computacional que implica la cantidad de dígitos retenidos) si queremos acercarnos a x^* con mucha precisión.

III.3.2 Métodos de Newton-Raphson. Usos de la Derivada de la Función

Si basados en la Ecuación (11), expresión de la secante generalizada, modificamos el término corrector de modo que aparezca la pendiente en el punto, obtendremos el conocido método de Newton-Raphson. Es decir, si:

$$x_{i+1} = x_i - \frac{y_i}{\dfrac{y_i - y_{i-1}}{x_i - x_{i-1}}} = x_i - \frac{f(x_i)}{\dfrac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}} \qquad (12)$$

se puede hacer que:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \qquad (13)$$

expresión del algoritmo que nos interesa.

Otra forma de entender la propuesta es suponer una expansión en serie de Taylor, truncada en el primer término (linealización de la función). Esto es:

ISBN: 950-42-0022-2 - ©1999

$$f(x_{i+1}) - f(x_i) + f'(x_i)(x_{i+1} - x_i) + \ldots \qquad (14)$$

$y = f(x_{i+1})$ es tal que es un valor muy cercano a x^*, tal que si: $y_{i+1} / f(x_{i+1}) = 0$, tenemos:

$$f(x_i) - f'(x_i)(x_{i+1} - x_i) \qquad (15)$$

que conduce, reordenando, a la Ecuación (13) básica del método.

Su ventaja es el orden de convergencia ($p = 2$, cuadrática) y su desventaja principal es la evaluación de la derivada de la función, ya sea analítica o numérica, como se verá en los casos de interés práctico.

***Ejemplo*:**

Sea la siguiente función, que por lo sencilla pueden conocerse sus raíces, de tal forma de ilustrar la mecánica del método: $f(x) = x^2 - 3x = 0$, con raíces 0 y 3. Encontrar una de sus raíces por N-R. Para ello como vimos necesitamos un valor inicial supuesto, o semilla. Sea éste $x_0 = 1$, y el criterio de tolerancia: $2f(x)2 \# 10^{-3}$.

$f(x_0) = -2$, distinto de cero, luego:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

$f'(x) = 2x - 3$, luego:

$x_1 = 1 - (-2) / (-1) = -1$.

$f(x_1) = 1 + 3 = 4 \ldots 0$, luego:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = (-1) - 4/(-5) = -0.2$$

$f(-0.2) = 0.64, \ldots 0$, luego:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = (-0.2) - (0.64)/(-3.4) = -0.0118$$

$f(x_3) = -0.0.0354$

$$x_4 = x_3 - \frac{f(x_3)}{f'(x_3)} = (-0.0118) - (0.0354)/(-3.0235) = -4.577\ 10^{-5}$$

ISBN: 950-42-0022-2 - ©1999

ello implica que x_4 es raíz de la función dada, dentro del margen de tolerancia especificado.

III.3.3 Sustitución Directa o Aproximaciones Sucesivas

Dada una función o una ecuación, si puede resolverse en forma explícita para una variable, es decir si:

Dado: $f(x) = 0$

Se propone: $x = F(x)$

esto es, explicitando la variable independiente, se puede, entonces, establecer la fórmula para un *algoritmo de un solo punto y estacionario,* es decir:

$$x_{i\%1} \text{ ' } F(x_i) \qquad (16)$$

Su éxito dependerá, evidentemente, del arreglo logrado para la ecuación (esto es:

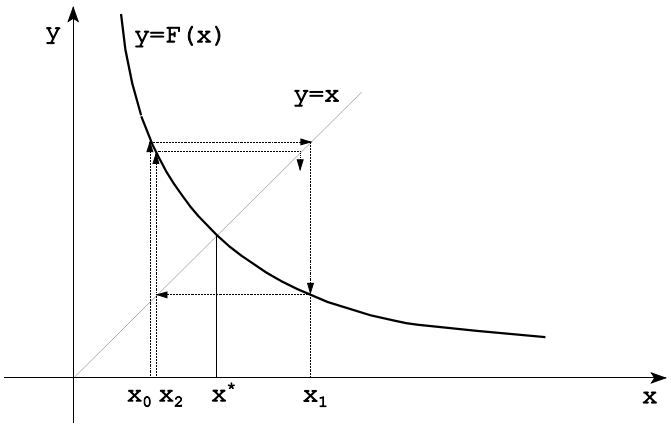


Figura III.4: Método de aproximaciones sucesivas.

F(x)), cuyo análisis permite ver la factibilidad de solución. Esto es, de acuerdo a la Figura (III.4), cuando es convergente (estabilidad del método). Se puede probar -y ver gráficamente- que si:

ISBN: 950-42-0022-2 - ©1999

$$|F'(x)| < 1 \quad ; \quad x \in I \subset \mathbb{R}$$

es convergente.

***Ejemplo*:**

Sea la misma función del ejemplo anterior. Encontrar una raíz por el método de sustitución directa. Para explicitar *x* operamos algebraicamente, y obtenemos:

$$x = \sqrt{3x} = F(x), \quad con\ x \geq 0$$

Hacemos $x_0 = 1$, criterio de error (5.10^{-3})

$x_{i+1} = F(x_i)$, entonces:

$x_1 = F(x_0) = 3^{½} = 1.732051$

$x_2 = F(x_1) = 2.279510$

$x_3 = F(x_2) = 2.615060$

$x_4 = F(x_3) = 2.800923$

$x_5 = F(x_4) = 2.898753$

$x_6 = F(x_5) = 2.948942$

$x_7 = F(x_6) = 2.974361$

$x_8 = F(x_7) = 2.987153$

$x_9 = F(x_8) = 2.993570$

$x_{10} = F(x_9) = 2.996783$

$|x_{10} - x_9| = 3.2\ 10^{-3}$, luego, x_{10} es raíz de $f(x)$ dentro del margen de error especificado.

$f(x_{10}) = 0.0096 \neq 0$

Aceleradores de Convergencia: Casos p=1

Son técnicas basadas en extrapolaciones. Si la iteración es convergente, será:

$$x' - x_{i+1} = K_i\ (x' - x_i);\ |K_i| < 1 \qquad (17)$$

donde: $|K_i| \cong K$: constante de error asintótico.

Cerca de la convergencia K_i es prácticamente constante (y parecido al valor final de *K*). De modo que:

$$(x' - x_{i+1}) = K\ (x' - x_i) \qquad (18)$$

y eliminando *K'* entre dos iteraciones

ISBN: 950-42-0022-2 - ©1999

$$\frac{(x^* - x_{i+2})}{(x^* - x_{i+1})} \cong \frac{(x^* - x_{i+1})}{(x^* - x_i)} \tag{19}$$

resolviendo para x^* se obtiene:

$$x^* \cong \frac{x_i\, x_{i+2} - x_{i+1}^2}{x_{i+2} - 2\, x_{i+1} + x_i} \tag{20}$$

Este resultado es el *procedimiento de Aitken*, de aceleración de convergencia.

De manera que: $x_{i+3} \cong x^*$ es una mejor aproximación (valor extrapolado); y así se continúa con el proceso iterativo, que logra mejorar la performance.

III.3.4 Procedimiento de Wegstein

Es el algoritmo más utilizado para acelerar el método de aproximaciones sucesivas. Incluso, es de gran importancia su implementación en problemas de sistemas de ecuaciones no lineales, como veremos en el próximo capítulo.

La base del mismo es proponer a la clásica iteración de aproximaciones sucesivas un valor *mejorado*, según la siguiente ecuación:

$$\overline{x}_{i+1} = q\, \overline{x}_i + (1 - q)\, x_{i+1} \tag{21}$$

de modo que:

$$x_{i+2} = F\,(\overline{x}_{i+1})$$

se corrige x_{i+2} y continúa. De la Ecuación (21) se ve que es necesario generar dos valores según el esquema tradicional, y conociendo q comenzar con esta propuesta.

Para analizar el cálculo de *q* puede ser útil la Figura (III.5), donde se aprecia la aproximación propuesta.

ISBN: 950-42-0022-2 - ©1999

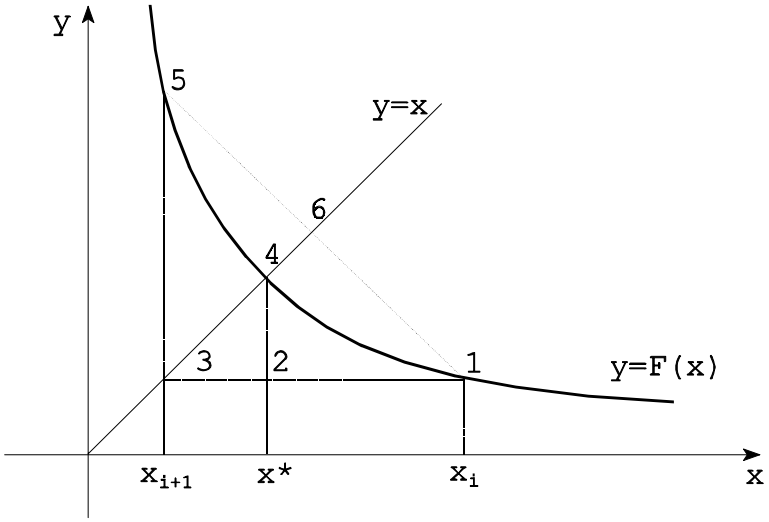


Figura III.5: Método de Wegstein.

Teniendo como idea que:

$$\bar{x}_{i+1} \approx x^*$$

es la definición de *q*, según:

$$x^* = q\, x_i + (1 - q)\, x_{i+1}$$

o sea

$$x^* - x_{i+1} = q\, (x_i - x_{i+1}) \qquad (22)$$

y, de igual modo,

$$x^* - x_i = (1 - q)\, (x_{i+1} - x_i) \qquad (23)$$

resultando

$$\frac{x^* - x_{i+1}}{x^* - x_i} \equiv - \frac{q}{(1 - q)} \qquad (24)$$

Como x^* no se conoce, se debe aproximar. Esto es:

ISBN: 950-42-0022-2 - ©1999

$$\frac{q}{(1-q)} = \frac{\overline{23}}{\overline{12}} = \frac{\overline{24}}{\overline{12}} \cong \frac{\overline{35}}{\overline{13}}$$

donde los segmentos indicados en los cocientes están determinados en Figura (III.5). Luego:

$$\frac{q}{(1-q)} \cong \frac{x_{i+2} - x_{i+1}}{x_i - x_{i+1}} \tag{25}$$

de donde obtenemos la expresión para el cálculo de q:

$$q \cong \frac{x_{i+2} - x_{i+1}}{x_i - 2\,x_{i+1} + x_{i+2}} \tag{26}$$

En el capítulo siguiente veremos ejemplos de aplicación tanto para una sola ecuación como para sistemas de ecuaciones no lineales, resueltas por este método.

III.3.5 Uso de Fracciones Continuadas

Para problemas de una variable la representación a través de fracciones continuadas ha demostrado tener ciertas ventajas sobre otros métodos tradicionales. La principal es el reducido número de operaciones que requiere su ejecución, con la correspondiente disminución del tiempo de computación.

Este tipo de representación ha sido usado para problemas de interpolación de datos; y se puede extender al caso de búsqueda de raíces que nos interesa.

Si se conocen valores tabulados de una función, es decir: (x_o , f_0), (x_1 , f_1),..., (x_n , f_n), para aproximar $f(x)$ se puede proponer la siguiente expresión:

$$f(x) \cong \varnothing_0(x) = a_0 + \cfrac{x - x_0}{a_1 + \cfrac{x - x_1}{a_2 + \cfrac{x - x_2}{a_3 + \cfrac{x - x_3}{\cdots\cdots \\ a_{n-1} + \cfrac{x - x_{n-1}}{a_n}}}}} \tag{27}$$

Que, en forma compacta, puede ser escrita como:

ISBN: 950-42-0022-2 - ©1999

$$f(x) \cong \varnothing_0(x) = a_0 + \frac{x - x_0}{\varnothing_1(x)} \tag{28}$$

donde:

$$\begin{aligned}
\varnothing_1(x) &= a_1 + \frac{x - x_1}{\varnothing_2(x)} \\
\varnothing_2(x) &= a_2 + \frac{x - x_2}{\varnothing_3(x)} \\
&\cdots\cdots\cdots\cdots \\
\varnothing_{i-1}(x) &= a_{i-1} + \frac{x - x_{i-1}}{\varnothing_i(x)} \\
&\cdots\cdots\cdots\cdots \\
\varnothing_n(x) &= a_n
\end{aligned} \tag{29}$$

Los coeficientes a_i pueden evaluarse con la Ecuación (29). Esto es, si $\varnothing_0(x_i)$ se iguala a los valores f_i, luego se resuelven las ecuaciones que resultan. Es decir: en $x = x_o$

$$\varnothing_0(x_0) = f(x_0) \equiv f_0 = a_0 \tag{30}$$

en $x = x_1$

$$\varnothing_0(x_1) = f_1 = f_0 + \frac{x_1 - x_0}{a_1}$$

luego,

$$a_1 = \frac{x_1 - x_0}{f_1 - f_0} \tag{31}$$

De igual forma se pueden obtener todos los coeficientes.

En el caso de búsqueda de raíces, es decir resolver para los valores de x tal que: $f(x) = 0$, se plantea

$$x = \psi(f) \tag{32}$$

ecuación de la función inversa. Si ésta se aproxima por fracciones continuadas,

ISBN: 950-42-0022-2 - ©1999

$$x = \psi_0(f) = a_0 + \cfrac{f - f_0}{a_1 + \cfrac{f - f_1}{a_2 + \cfrac{f - f_2}{a_3 + \cfrac{f - f_3}{\cdots\cdots\cdots + \cfrac{f - f_{n-1}}{a_{n-1} + \cfrac{f - f_{n-1}}{a_n}}}}}}} \qquad (33)$$

Luego, sustituimos $f = 0$ en Ecuación (33), de modo que el lado izquierdo sea una aproximación a la raíz buscada,

$$\bar{x} = a_0 - \cfrac{f_0}{a_1 - \cfrac{f_1}{a_2 - \cfrac{f_2}{a_3 - \cfrac{f_3}{\cdots\cdots\cdots - \cfrac{f_{n-1}}{a_{n-1} - \cfrac{f_{n-1}}{a_n}}}}}}} \qquad (34)$$

La determinación de los coeficientes a_i se lleva a cabo en una forma similar a la ya vista. Esto es, si:

$$x = \Omega_0(f) = a_0 + \frac{f - f_0}{\Omega_1(f)} \qquad (35)$$

donde:

$$\begin{aligned} \Omega_1(f) &= a_1 + \frac{f - f_1}{\Omega_2(f)} \\ &\cdots\cdots\cdots \\ \Omega_{i-1}(f) &= a_{i-1} + \frac{f - f_{i-1}}{\Omega_i(f)} \\ &\cdots\cdots\cdots \\ \Omega_n(f) &= a_n \end{aligned} \qquad (36)$$

la evaluación se hace con la Ecuación (36), poniendo $\Omega_0(f_i)$ igual a x_i y resolviendo

ISBN: 950-42-0022-2 - ©1999

sucesivamente. En $f = f_0$, $\psi(f_0) = a_0$ (Ecuación 33).

Un algoritmo muy interesante, basado en este método, es el de Shacham (1989). Se denomina de memoria mejorada y se aplica a cualquier ecuación no lineal.

El reemplazo de los polinomios de interpolación de Lagrange por fracciones continuadas mejora los métodos de interpolación inversa, sobre todo en el uso de memoria que requieren. El mencionado algoritmo, para el caso de una ecuación del tipo $f(x) = 0$, se puede describir de la siguiente manera:

1.- Elija dos valores iniciales, x_0 y x_1. Calcule $y_o = f(x_0)$, $y_1 = f(x_1)$.

2.- Compute x_2, con las siguientes ecuaciones:
$a_0 = x_0$
$a_1 = (y_1 - y_0)/(x_1 - x_0)$
$x_2 = a_0 - y_0/a_1$

3.- Haga $n = 2$ y calcule $y_2 = f(x_2)$.

4.- Con las ecuaciones recursivas se calcula x_{n+1}, de modo que:
$b_0 = x_n$
$b_i = (y_n - y_{i-1})/(b_{i-1} - a_{i-1})$; $i = 1,2,3, ..., n-1$
$\psi_n = a_n = (y_n - y_{n-1})/(b_{n-1} - a_{n-1})$
$\psi_{i-1} = a_{i-1} - y_{i-1}/\psi_i$; $i=n, n-1, ..., 1$
$x_{n+1} = \psi_0$

5.- Calcule $y_{n+1} = f(x_{n+1})$.

6.- Verifique la convergencia. Si no converge, haga $n = n+1$, y vuelva a *4*.

III.4 SOLUCIÓN DE SISTEMAS DE ECUACIONES LINEALES SIMULTÁNEAS

III.4.1 Planteo del Problema. Teoremas Básicos

Se trata de resolver n ecuaciones lineales simultáneas con n incógnitas:

$$\sum_{j=1}^{n} a_{ij}\, x_j = b_i \; ; \quad i = 1,\ 2,...,\ n \tag{37}$$

o, en su forma compacta matricial

$$\underline{\underline{A}} \cdot \underline{x} = \underline{b} \tag{38}$$

donde $\boldsymbol{A} = [a_{ij}]$ es la matriz de coeficientes (cuadrada, nxn) y $\boldsymbol{x}^t = (x_1, ..., x_n)$ es el vector de incógnitas, siendo $\boldsymbol{b}^t = (b_1, ..., b_n)$ el vector de términos independientes. Para el caso que nos interesa, tanto $\boldsymbol{A}$ como $\boldsymbol{b}$ son reales. En adelante se obviará la notación transpuesta de la matriz, suponiendo que en las operaciones entre matrices se

ISBN: 950-42-0022-2 - ©1999

disponen éstas de tal forma que sean compatibles para las mismas. Si definimos la denominada matriz aumentada como:

$$\underline{\underline{A}}_b = [\underline{\underline{A}} \quad \underline{b}] \; ; \; n \; . \; (n + 1) \tag{39}$$

dado que $\boldsymbol{b}$ es un vector columna; y recordando la definición de rango de una matriz: v(**A**), el teorema básico de existencia de solución establece:

1.- El sistema de ecuaciones (Ecuación(38)) tiene solución *sí y sólo sí:* $v(\boldsymbol{A}) = v(\boldsymbol{A}_b)$.

2.- Si $v(\boldsymbol{A}) = v(\boldsymbol{A}_b) = k < n$, luego las $x_{i1}, x_{i2},..., x_{ik}$ son variables cuyas columnas son linealmente independientes en $\boldsymbol{A}$, de modo que las restantes $(n-k)$ variables pueden asignarse arbitrariamente. O dicho de otra forma, hay una familia paramétrica de $(n-k)$ soluciones.

3.- Si $v(\boldsymbol{A}) = v(\boldsymbol{A}_b) = n$, hay una única solución.

Corolario: Para el caso homogéneo ($\boldsymbol{b} = \boldsymbol{0}$), o sea $\boldsymbol{A} \,.\, \boldsymbol{x} = 0$ habrá solución no trivial, si y sólo si $v(\boldsymbol{A}) < n$.

Para estos problemas, cosa que no ocurre en el caso no lineal, existe solución analítica (recordemos la denominada *Regla de Cramer*), pero la dificultad reside principalmente en computar esa solución. La evaluación de determinantes no hace práctico dicho procedimiento analítico; luego, el problema es desarrollar algoritmos computacionales más eficientes, es decir que sean más rápidos, sobre todo en el número de operaciones necesarias y que además sean robustos de modo que la solución calculada sea lo más precisa posible.

Un punto vital es discutir cómo se espera que sea la matriz de coeficientes. En general, puede encontrarse entre alguna de estas dos categorías:

i.- Llena pero no muy grande. Es decir con muy pocos ceros, y en donde n no sea mayor que 100, por ejemplo.

ii.- Dispersa y relativamente muy grande, denominadas también ralas. En estos casos, son muy pocos (en relación al orden) los elementos distintos de cero y n puede ser mayor a 1000. Estas matrices son típicas al resolver problemas con ecuaciones diferenciales parciales y también aparecen al plantear modelos de simulación de plantas completas, o bien de procesos con múltiple etapas en serie, como veremos en los próximos capítulos.

Naturalmente, los métodos desarrollados deben estar dirigidos a resolver alguna de estas dos categorías, y si es posible haciendo uso de sus características para incrementar su eficiencia.

ISBN: 950-42-0022-2 - ©1999

Un problema, cual es la condición del sistema, se puede analizar considerando un vector residual ***r***, cuando se tiene una solución calculada $\boldsymbol{x_c}$. Es decir:

$$\underline{r} = \underline{b} - \underline{\underline{A}} \cdot \underline{x}_c \tag{40}$$

se sabe que:

$$\underline{\underline{A}} \cdot \underline{x}^* = \underline{b} \text{ , } o \text{ } \underline{b} - \underline{\underline{A}} \cdot \underline{x}^* = 0 \tag{41}$$

entonces será:

$$\underline{r} = \underline{\underline{A}} \cdot (\underline{x}^* - \underline{x}_c)$$

luego,

$$(\underline{x}^* - \underline{x}_c) = \underline{\underline{A}}^{-1} \cdot \underline{r} \tag{42}$$

de donde se ve que aunque ***r*** tenga elementos muy chicos, si $\boldsymbol{A^{-1}}$ (matriz inversa) contiene muy grandes coeficientes, la diferencia entre $\boldsymbol{x^*}$ y $\boldsymbol{x_c}$ puede ser aún muy grande. Esto permite anticipar la importancia de un escalado en los coeficientes de la matriz ***A*** original, ya que aunque el vector residual ***r*** impuesto sea pequeño, el error encontrado para la solución puede ser muy grande.

III.4.2 Métodos Directos

Un método directo para hallar la solución es uno en el cual, si todos los cálculos (computaciones) fueran llevados a cabo sin error de redondeo conduciría a la solución exacta del sistema dado. Prácticamente todos están basados en la *técnica de eliminación*. El error de truncamiento para estos métodos es intrascendente.

Eliminación Gaussiana

Desarrollando la Ecuación (42), se obtiene el sistema en la siguiente forma:

$$\begin{aligned}
&a_{11}\,x_1 + a_{12}\,x_2 + ... + a_{1n}\,x_n = b_1 \equiv a_{1,\,n+1}\\
&a_{21}\,x_1 + a_{22}\,x_2 + ... + a_{2n}\,x_n = b_2 \equiv a_{2,\,n+1}\\
&\qquad\qquad .....\\
&\qquad\qquad .....\\
&a_{n1}\,x_1 + a_{n2}\,x_2 + ... + a_{nn}\,x_n = b_n \equiv a_{n,\,n+1}
\end{aligned} \tag{43}$$

Se supone que la matriz es no singular, y que $a_{11} \neq 0$ de manera de poder dividir la primer columna por a_{i1} y así restar para las ecuaciones, donde $i=2, ..., n$. Esto da como resultado:

ISBN: 950-42-0022-2 - ©1999

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + \ldots\ldots\ldots + a_{1n}\,x_n &= a_{1,\,n+1} \\ a_{22}^{(1)}\,x_2 + \ldots\ldots\ldots + a_{2n}^{(1)}\,x_n &= a_{2,\,n+1}^{(1)} \\ \ldots\ldots\ldots & \\ \ldots\ldots\ldots & \\ a_{n2}^{(1)}\,x_2 + \ldots + a_{nn}^{(1)}\,x_n &= a_{n,\,n+1}^{(1)} \end{aligned} \qquad (44)$$

siendo los $a^{(1)}_{ij}$, tal que:

$$a_{ij}^{(1)} = a_{ij} - \frac{a_{1j}}{a_{11}}\,a_{i1} \;;\; i=2,\ \ldots\ ,\ n \quad ;\quad j=2,\ \ldots\ ,\ n+1 \qquad (45)$$

si fuese $a_{11} = 0$ se intercambian columnas, y se opera en consecuencia. Por comodidad se define: $m_{i1} = a_{i1}/a_{11}$ con $i=2, \ldots, n$.

De igual forma se continúa con el procedimiento haciendo ahora (si $a^{(1)}_{22} \neq 0$) $m_{i2} = a^{(1)}_{i2}/a^{(1)}_{22}$ con $i=3, \ldots, n$ resultando, al restar, el siguiente sistema:

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + \ldots\ldots\ldots\ldots + a_{1n}\,x_n &= a_{1,\,n+1} \\ a_{22}^{(1)}\,x_2 + \ldots\ldots\ldots\ldots + a_{2n}^{(1)}\,x_n &= a_{2,\,n+1}^{(1)} \\ a_{33}^{(2)}\,x_3 + \ldots\ldots\ldots\ldots + a_{3n}^{(2)}\,x_n &= a_{3,\,n+1}^{(2)} \\ \ldots\ldots\ldots\ldots & \\ a_{n3}^{(2)}\,x_3 + \ldots\ldots\ldots\ldots + a_{nn}^{(2)}\,x_n &= a_{n,\,n+1}^{(2)} \end{aligned} \qquad (46)$$

donde:

$$a_{ij}^{(2)} = a_{ij}^{(1)} - m_{i2}\,a_{2j}^{(1)} \quad ;\quad i=3,\ \ldots\ ,\ n \qquad j=3,\ \ldots\ ,\ n+1 \qquad (47)$$

Y, continuando con el procedimiento hasta $(n-1)$ pasos llegamos al sistema final:

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + \ldots\ldots + a_{1n}\,x_n &= a_{1,\,n+1} \\ a_{22}^{(1)}\,x_2 + \ldots\ldots + a_{2n}^{(1)}\,x_n &= a_{2,\,n+1}^{(1)} \\ \ldots\ldots\ldots & \\ a_{nn}^{(n-1)}\,x_n &= a_{n,\,n+1}^{(n-1)} \end{aligned} \qquad (48)$$

con los elementos en la diagonal (distintos de cero), tal que:

ISBN: 950-42-0022-2 - ©1999

$$a_{ij}^{k} = a_{ij}^{k-1} - m_{ik}\, a_{kj}^{k-1} \;; \quad k = 1,....., n-1 \qquad j = k+1,..., n+1 \qquad i = k+1,..., n \qquad a_{ij}^{0} = a_{ij} \tag{49}$$

con $m_{ik} = a_{ik}^{(k-1)} / a_{kk}^{(k-1)}$

Luego, la solución es fácilmente calculada por sustitución hacia atrás, al terminar el procedimiento de eliminación. Esto es:

$$x_i = \frac{1}{a_{ii}^{i-1}} \left[a_{i,n+1}^{(i-1)} - \sum_{j=i+1}^{n} a_{ij}^{(i-1)}\, x_j\right] \;,\; i = n,...,1 \tag{50}$$

Una variante muy importante es el *procedimiento de reducción o eliminación* de Gauss-Jordan. En este caso se procede a eliminar con los elementos diagonales en toda la columna, dejando sólo ese elemento; y se deriva en un sistema que comparando con Ecuación (44), en una primera pasada tiene la forma:

$$\begin{aligned}
a_{11}\, x_1 + a_{13}^{(2)}\, x_3 + ... + a_{1n}^{(2)}\, x_n &= a_{1,\,n+1}^{(2)} \\
a_{22}^{(1)}\, x_2 + a_{23}^{(1)}\, x_3 + ... + a_{2n}^{(1)}\, x_n &= a_{2,\,n+1}^{(1)} \\
a_{33}^{(2)}\, x_3 + ... + a_{3,n}^{(2)}\, x_n &= a_{3,n+1}^{(2)} \\
&..... \\
a_{n3}^{(2)}\, x_3 + ... + a_{nn}^{(2)}\, x_n &= a_{n,\,n+1}^{(2)}
\end{aligned} \tag{51}$$

Notar que:

$$a_{2,\,n+1}^{(2)} = a_{2,\,n+1}^{(1)}$$

Siguiendo con el procedimiento se hacen cero todos los elementos, excepto los correspondientes a la diagonal, resultando:

ISBN: 950-42-0022-2 - ©1999

$$\begin{aligned} a_{11}\ x_1 \qquad\qquad &= a_{1,\ n+1}^{(n-1)} \\ a_{22}^{(1)}\ x_2 \qquad &= a_{2,\ n+1}^{(n-1)} \\ \ldots\ldots\ldots\ldots & \\ a_{nn}^{(n-1)}\ x_n \quad &= a_{n,\ n+1}^{(n-1)} \end{aligned} \tag{52}$$

de modo que la solución es simplemente:

$$x_i = \frac{a_{i,\ n+1}^{(n-1)}}{a_{ii}^{(i-1)}} \quad ; \quad i=1,\ \ldots\ ,\ n \tag{53}$$

A pesar de lo que resulta del procedimiento, la eliminación Gaussiana es la más eficiente de las dos, considerando sólo las multiplicaciones y divisiones (recordar acumulación de error, además), llegando, para grandes sistemas ($n >> 1$), el método Gauss-Jordan a requerir cerca de un *50%* más de operaciones que el de Gauss.

Se ha trabajado mucho para lograr *formas compactas* del método de Gauss, no sólo para ahorrar espacio de almacenamiento (memoria) sino también para mejorar la precisión en los cálculos que más inciden en el resultado. Para ello se han definido matrices especiales en cuanto a la característica de su formulación, las cuales permiten ahorrar tiempo de cálculo una vez aplicadas. No trataremos este punto aquí, remitiendo al lector a la bibliografía recomendada al final del capítulo.

Análisis de errores

Debido a que generalmente no es posible obtener la solución exacta, se considerarán los posibles errores y sus cotas. Las fuentes de error son variaciones en los elementos de ***A*** y de ***b***, ya sea originales o debidas al redondeo. Estudiando primero el caso más simple, que considera cambios sólo en ***b***, será:

$$\underline{\underline{A}} \cdot \underline{x} = \underline{b} + \underline{\delta b} \tag{54}$$

si resulta a su vez:

$$\underline{x} = \underline{x}^* + \underline{\delta x}$$

se ve que:

$$\underline{\underline{A}} \cdot (\underline{x}^* + \underline{\delta x}) = \underline{b} + \underline{\delta b}$$

con lo que:

ISBN: 950-42-0022-2 - ©1999

$$\underline{\underline{A}} \cdot \underline{\delta x} = \underline{\delta b}$$

$$\underline{\delta x} = \underline{\underline{A}}^{-1} \cdot \underline{\delta b} \tag{55}$$

y/o, tomando normas:

$$\|\underline{\delta x}\| \leq \|\underline{\underline{A}}^{-1}\| \, \|\underline{\delta b}\| \tag{56}$$

de igual modo:

$$\|\underline{b}\| \leq \|\underline{\underline{A}}\| \cdot \|\underline{x}^*\|$$

tal que:

$$\|\underline{x}^*\| \geq \frac{\|\underline{b}\|}{\|\underline{\underline{A}}\|}$$

resultando el error relativo:

$$\frac{\|\underline{\delta x}\|}{\|\underline{x}^*\|} \leq \|\underline{\underline{A}}\| \, \|\underline{\underline{A}}^{-1}\| \, \frac{\|\underline{\delta b}\|}{\|\underline{b}\|} \tag{57}$$

La cantidad $\|\mathbf{A}\| \, \|\mathbf{A}^{-1}\|$ se denomina el *número de condición de* $\mathbf{A}$, $K(\mathbf{A}) \geq 1$ siempre. De modo que:

$$\frac{\|\underline{\delta x}\|}{\|\underline{x}^*\|} \leq K(\underline{\underline{A}}) \, \frac{\|\underline{\delta b}\|}{\|\underline{b}\|} \tag{58}$$

Así, si $K(\mathbf{A})$ es cercano a 1, se dice que $\mathbf{A}$ es *bien condicionada*; y si es muy grande nos encontramos frente a un caso *mal condicionado*.

Si la fuente de error fueran los elementos de la matriz de coeficientes, esto es:

$$(\underline{\underline{A}} + \underline{\underline{\delta A}}) \cdot (\underline{x}^* + \underline{\delta x}) = \underline{b} \tag{59}$$

por un procedimiento similar al anterior se llega a que:

ISBN: 950-42-0022-2 - ©1999

$$\frac{\|\delta \underline{x}\|}{\|\underline{x}^* + \delta \underline{x}\|} \leq K(\underline{\underline{A}}) \frac{\|\delta \underline{\underline{A}}\|}{\|\underline{\underline{A}}\|} \qquad (60)$$

con igual connotación que en Ecuación (57). Los efectos del valor $\|\delta \boldsymbol{A}\|$ pueden subsanarse, en parte, trabajando con mayor precisión (mayor retención de cifras significativas durante el cálculo).

Refinamiento iterativo

Hemos visto que:

$$\underline{r} = \underline{b} - \underline{\underline{A}} \cdot \underline{x}_c$$

En las situaciones que nos interesa una solución muy cercana a la verdadera $\boldsymbol{x}^* = \boldsymbol{A}^{-1} \cdot \boldsymbol{b}$, con un vector tal que **r** sea pequeño o, en caso en que $\boldsymbol{x}_c$ sea tal que:

$$\underline{e} = \underline{x}^* - \underline{x}_c \qquad (61)$$

y sea muy pequeño en términos relativos a $\boldsymbol{x}^*$. Si:

$$\underline{b} = \underline{\underline{A}} \cdot \underline{x}^*$$

Resulta:

$$\underline{r} = \underline{\underline{A}} \cdot \underline{x}^* - \underline{\underline{A}} \cdot \underline{x}_c = \underline{\underline{A}} \cdot (\underline{x}^* - \underline{x}_c) = \underline{\underline{A}} \cdot \underline{e} \qquad (62)$$

Tomando normas, y procediendo como ya se ha visto, se puede probar que:

$$\frac{1}{K(\underline{\underline{A}})} \frac{\|\underline{r}\|}{\|\underline{b}\|} \leq \frac{\|\underline{e}\|}{\|\underline{x}^*\|} \leq K(\underline{\underline{A}}) \frac{\|\underline{r}\|}{\|\underline{b}\|} \qquad (63)$$

Por lo que, si el número de condición $K(\boldsymbol{A})$ es cercano a la unidad, pequeños errores relativos de $\boldsymbol{r}$ y $\boldsymbol{e}$ siguen la misma tendencia (en coincidencia). Este hecho no ocurre para sistemas mal condicionados, como se había anticipado.

La respuesta al problema dependerá de la condición de $\boldsymbol{A}$ y de la precisión de la aritmética (redondeo). Esto lleva a que el cómputo del residuo $\boldsymbol{r}$ debe hacerse con doble precisión, por la razón de que $\boldsymbol{r}$ suele ser del mismo orden de magnitud que el error de redondeo.

Formalizar un procedimiento que contemple el problema es lo que se denomina *refinamiento iterativo*. En realidad se arma un esquema iterativo con una solución computada previamente, tal que el residuo para una etapa *m* de cálculo sea:

ISBN: 950-42-0022-2 - ©1999

$$\underline{r}^{(m)} = \underline{b} - \underline{\underline{A}} \cdot \underline{x}^{(m)} \; ; \; m = 1, 2, ..., \tag{64}$$

y de acuerdo a la Ecuación (61)

$$\underline{x}^{(m+1)} = \underline{x}^{(m)} + \underline{e}^{(m)} \tag{65}$$

con $\boldsymbol{e}^{(m)}$ calculada según la ecuación anterior.

Si $\boldsymbol{x}^{(m+1)}$ no es satisfactorio se procede a la etapa *(m+1)* de cálculo. En cambio, si $\|\boldsymbol{e}^{(m+1)}\| / \|\boldsymbol{e}^{(m)}\|$ es menor que una tolerancia establecida aceptamos la solución $\boldsymbol{x}^{(m)}$. Así, se conforma un criterio de terminación que funciona muy bien en la práctica.

Cada etapa de este proceso, además, es mucho más rápida que la solución del problema original. Se debe recordar que la computación de Ecuación (64) se realiza en doble precisión, lo que implica aumento en el requerimiento de memoria.

III.4.3 Métodos Iterativos

Su basamento es idéntico al método de aproximaciones sucesivas visto en el caso de funciones no lineales, con lo que empezando con un vector inicial se genera una sucesión de vectores, tal que:

$$\underline{x}_{i+1} = \underline{F}_i \, [\underline{x}_i, \, \underline{x}_{i-1}, ..., \, \underline{x}_{i-k}] \tag{66}$$

Si la F_i (función iteradora) no es dependiente sobre i (nivel de iteración) se llama, a la recurrencia, estacionaria.

Para la mayoría de las matrices estos métodos requieren más computación, para un deseado grado de convergencia, que los métodos directos; pero para la matrices ralas (de gran interés en aplicaciones) el esfuerzo computacional es comparable. Además, como hemos mencionado, para estas matrices es posible lograr una mejor utilización de la memoria. Luego para matrices ralas grandes, los métodos iterativos son los mas aconsejados por su eficiencia computacional.

Sólo nos interesarán *procesos iterativos lineales* por consideraciones de eficacia, una vez más. Un procedimiento de iteración *matricial lineal en un punto* tiene la forma general.

$$\underline{x}_{i+1} = \underline{\underline{B}}_i \cdot \underline{x}_i + \underline{c}_i \tag{67}$$

dónde $\boldsymbol{B}_i$ y $\boldsymbol{c}_i$ son independientes de *i* (iteración estacionaria).

Recordar que en los métodos iterativos basados en substitución directa (o Wegstein) debe explicitarse la variable independiente. A modo de introducción se puede ver dicha forma haciendo:

ISBN: 950-42-0022-2 - ©1999

$$\underline{\underline{A}} \cdot \underline{x} = \underline{b}$$

$$\underline{\underline{A}} \cdot \underline{x} + \underline{x} = \underline{b} + \underline{x}$$

$$(\underline{\underline{A}} + \underline{\underline{I}}) \cdot \underline{x} = \underline{x} + \underline{b}$$

$$\underline{x} = (\underline{\underline{A}} + \underline{\underline{I}}) \cdot \underline{x} - \underline{b}$$

o, en la ley de recurrencia:

$$\underline{x}_{i+1} = (\underline{\underline{I}} + \underline{\underline{A}}) \cdot \underline{x}_i - \underline{b} \qquad (68)$$

que nos da una forma genérica simple de los métodos iterativos estacionarios que son los universalmente usados.

Método de Jacobi
Se escribe la matriz ***A*** como:

$$\underline{\underline{A}} = \underline{\underline{D}} + \underline{\underline{L}} + \underline{\underline{U}} \qquad (69)$$

siendo ***D*** la matriz diagonal (formada precisamente con esos elementos). ***L*** y ***U*** son triangulares inferior y superior, respectivamente, con el resto de los elementos; y ceros en su diagonal principal. Luego, si:

$$\underline{\underline{A}} \cdot \underline{x} = \underline{b}$$

$$(\underline{\underline{D}} + \underline{\underline{L}} + \underline{\underline{U}}) \cdot \underline{x} = \underline{b}$$

$$\underline{\underline{D}} \cdot \underline{x} = -(\underline{\underline{L}} + \underline{\underline{U}}) \cdot \underline{x} + \underline{b}$$

obtenemos:

$$\underline{x}_{i+1} = -\underline{\underline{D}}^{-1} \cdot (\underline{\underline{L}} + \underline{\underline{U}}) \cdot \underline{x}_i + \underline{\underline{D}}^{-1} \cdot \underline{b} \qquad (70)$$

por supuesto que supone que la diagonal principal no tiene todos ceros (en ***A***).

De no ser así, si ***A*** es no singular, se debe permutar filas y columnas para obtener una forma que permita definir ***D***. Es deseable que sea de la forma diagonal dominante, esto que los elementos de la diagonal sean lo más grandes posibles respecto a los demás. Para este método de Jacobi la matriz ***B*** puede entonces escribirse como:

ISBN: 950-42-0022-2 - ©1999

$$\underline{\underline{B}} = - \underline{\underline{D}}^{-1} \cdot (\underline{\underline{L}} + \underline{\underline{U}}) \qquad (71)$$

haciendo:

$$\underline{x}_{i+1} = \underline{\underline{B}} \cdot \underline{x}_i + \underline{\underline{D}}^{-1} \cdot \underline{b} \qquad (72)$$

Método de Gauss-Seidel

Este es el método ampliamente usado que siempre converge si converge Jacobi, (y aún en casos en que éste no converge); y en general lo hace más rápido, siendo su implementación más eficiente.

La diferencia fundamental es que a medida que es calculada cada componente de $\boldsymbol{x_{i+1}}$ es usada inmediatamente en la misma iteración. Al proceder de ese modo, la ecuación obtenida para Jacobi, resulta:

$$\underline{x}_{i+1} = - \underline{\underline{D}}^{-1} \cdot (\underline{\underline{L}} \cdot \underline{x}_{i+1}) + \underline{\underline{D}}^{-1} \cdot \underline{\underline{U}} \cdot \underline{x}_i + \underline{\underline{D}}^{-1} \cdot \underline{b} \qquad (73)$$

reordenando y sacando factor común $\boldsymbol{x_{i+1}}$ se tiene:

$$(\underline{\underline{I}} + \underline{\underline{D}}^{-1} \cdot \underline{\underline{L}}) \cdot \underline{x}_{i+1} = - \underline{\underline{D}}^{-1} \cdot \underline{\underline{U}} \cdot \underline{x}_i + \underline{\underline{D}}^{-1} \cdot \underline{b}$$

por lo que premultiplicando por $\boldsymbol{D}$, obtenemos:

$$(\underline{\underline{D}} + \underline{\underline{L}}) \cdot \underline{x}_{i+1} = - \underline{\underline{U}} \cdot \underline{x}_i + \underline{b}$$

luego:

$$\underline{x}_{i+1} = - (\underline{\underline{D}} + \underline{\underline{L}})^{-1} \cdot \underline{\underline{U}} \cdot \underline{x}_i + (\underline{\underline{D}} + \underline{\underline{L}})^{-1} \cdot \underline{b} \qquad (74)$$

entonces será, tomando $\boldsymbol{B}$ para la expresión generalizada:

$$\underline{x}_{i+1} = + \underline{\underline{B}} \cdot \underline{x}_i + (\underline{\underline{D}} + \underline{\underline{L}})^{-1} \cdot \underline{b} \qquad (75)$$

$$\underline{\underline{B}} = - (\underline{\underline{D}} + \underline{\underline{L}})^{-1} \cdot \underline{\underline{U}} \qquad (76)$$

Existe un teorema que prueba que: *Si una matriz es definida positiva, el procedimiento de Gauss-Seidel converge independientemente del vector inicial propuesto.*

Otro hecho importante es que los métodos matriciales iterativos convergen linealmente. Es natural pensar que el error de redondeo es mayor en un método

ISBN: 950-42-0022-2 - ©1999

iterativo que en uno directo; sin embargo como siempre se usa la matriz de coeficientes original, el error por redondeo en que se incurre en un método iterativo es sólo aquel producido en la última iteración. En consecuencia, ambos tipos de métodos tienen un error equivalente, tan serio para unos como para otros.

El error de truncamiento, en el caso de los métodos iterativos, suele ser de un orden mayor de magnitud que el mencionado de redondeo. No obstante, dado que en los métodos iterativos se tiene un criterio de control de error, este es acotado convenientemente.

PROBLEMAS PROPUESTOS

P1) Sea la siguiente ecuación definida para $x > 0$:

$$x^2 - \ln(3x) + 1/9 = 0$$

Encontrar las raíces de la misma por los siguientes métodos:

a) Wegstein
b) Newton-Raphson

P2) Para el flujo isoentrópico de un gas perfecto que fluye desde un reservorio a través de una boquilla convergente - divergente operando con velocidad sónica en la constricción, se puede demostrar que:

$$\frac{A_c^2}{A^2} = \left(\frac{\gamma + 1}{2}\right)^{(\gamma+1)/(\gamma-1)} \left(\frac{\gamma - 1}{2}\right) \left[\left(\frac{P}{P_r}\right)^{2/\gamma} - \left(\frac{P}{P_r}\right)^{(\gamma+1)/\gamma}\right]$$

donde P es la presión sobre el área transversal A de la boquilla, P_r es la presión en el reservorio, A_c es la sección transversal en la constricción y γ es la relación entre el calor específico a presión constante y el calor específico a volumen constante.

Si A_c, γ, P_r y A ($> A_c$) se conocen, idear un esquema para calcular las *dos posibles presiones P* que satisfacen la ecuación de arriba. Implemente su método en la computadora.

Datos sugeridos para prueba:
$A_c = 0.1$ *pie*^{2}, $\gamma = 1.41$, $P_r = 100$ *psia* y $A = 0.12$ *pie*^{2}.

P3) Una bolsa esférica de gas a alta presión, inicialmente de radio r_0 y presión p_0, se expande radialmente en una explosión submarina adiabática. Para el caso especial de un gas con $\gamma = 4/3$ (relación de calor específico a presión constante a calor específico a volumen constante), el radio r en cualquier instante de tiempo t posterior a la explosión se expresa como:

ISBN: 950-42-0022-2 - ©1999

$$\frac{t}{r_0}\sqrt{\frac{p_0}{\rho}} = \left(1 + \frac{2}{3} + \alpha + \frac{1}{5}\alpha^2\right)(2\ \alpha)^{1/2}$$

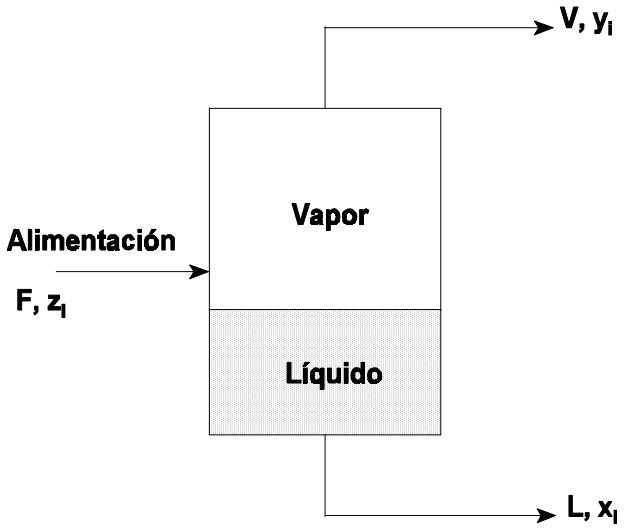
donde $\alpha = (r/r_0) - 1$ y ρ es la densidad del agua. Durante la expansión adiabática, la presión del gas se expresa como $(p/p_0) = (r_0/r)^{\gamma}$.

Desarrollar un procedimiento para calcular la presión y el radio de la bolsa para cualquier instante de tiempo.

Datos sugeridos para prueba:

$p_0 = 10^4\ lb_f/pulgada^2$, $\rho = 64\ lb_m/pie^3$, $r_0 = 1$ *pie*, $t = 0.5, 1, 2, 3, 5$ *y 10 milisegundos.*

P4) *F moles/hr*. de una corriente de gas natural licuado de *n-componentes* se introduce como corriente de alimentación a un tanque de vaporización flash como se indica en la figura siguiente:



Las corrientes resultantes de vapor y de líquido se extraen a las velocidades de V y L moles/hr., respectivamente. Las fracciones molares de los componentes en la alimentación, en el vapor y en el líquido, se designan, z_i, y_i, e x_i, respectivamente ($i = 1, 2, ..., n$). Suponiendo equilibrio líquido/vapor y operación en estado estacionario (en el capítulo IX desarrollaremos con mayor extensión el análisis de este equipo),

ISBN: 950-42-0022-2 - ©1999

tenemos:

Balance de materia: $F = L + V$

Balance individual: $z_i F = x_i L + y_i V$

Relación de equilibrio: $K_i = y_i/x_i$

Aquí, K_i, es la constante de equilibrio para la i-ésima componente a la temperatura y presión prevalecientes en el tanque. A partir de estas ecuaciones y del hecho que:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 1$$

probar que:

$$\sum_{i=1}^{n} \frac{z_i \left(K_i - 1\right)}{\Theta \left(K_i - 1\right) + 1} = 0$$

donde $\Theta = V/F$ representa la fracción vaporizada.

Escriba un programa que lea los valores de F, z_i, y los K_i como datos y entonces utilice el método de Newton para resolver esta última ecuación en Θ. El programa debería también calcular la fracción de líquido L/F, los x_i e y_i utilizando las tres primeras ecuaciones dadas más arriba. Los datos de prueba, que se muestran en la tabla siguiente, están relacionados con el flasheo de una corriente de gas natural a 1600 psia y 120 ºF.

Componente	i	z_i	K_i
Dióxido de carbono	1	0.0046	1.650
Metano	2	0.8345	3.090
Etano	3	0.0381	0.720
Propano	4	0.0163	0.390
Isobutano	5	0.0050	0.210
n-Butano	6	0.0074	0.175
Pentanos	7	0.0287	0.093
Hexanos	8	0.0220	0.065
Heptanos	9	0.0434	0.036
	Total	1.0000	

Suponer que *F = 1000 moles/hr*. Además deberían ser leídos como datos la tolerancia

ISBN: 950-42-0022-2 - ©1999

ε y un número máximo de iteraciones. ¿Cuál sería un buen valor de arranque para Θ_i?.

P5) Para el flujo turbulento de un fluido a través de un tubo liso, es posible establecer la siguiente relación entre el factor de fricción c_f y el número de Reynolds *Re*:

$$\sqrt{\frac{1}{c_f}} = -0.4 + 1.74 \ln\left(Re \sqrt{c_f} \right)$$

Calcular c_f para Re = 10^4, 10^5 y 10^6.

P6) Para el flujo estacionario de un fluido incompresible a través de un tubo rugoso de longitud *L* y diámetro interior *D*, la caída de presión viene expresada por la siguiente relación:

$$\Delta p = \frac{f_M \rho u_M^2 L}{2 D}$$

donde ρ es la densidad del fluido, u_m es la velocidad media del fluido y f_M es el factor de fricción de Moody (adimensional). El factor de fricción de Moody es una función de la rugosidad ϵ y del número de Reynolds,

$$Re = \frac{D \rho u_M}{\mu}$$

donde μ es la viscosidad del fluido. Para $Re \leq 2000$,

$$f_M = \frac{64}{Re}$$

mientras que para $Re > 2000$, f_M viene expresada por la Ecuación de Colebrook,

$$\sqrt{\frac{1}{f_M}} = -2 * \log_{10}\left(\frac{\varepsilon}{3.7 D} + \frac{2.51}{Re \sqrt{f_M}} \right)$$

Un buen punto de arranque para la solución iterativa de esta ecuación puede encontrarse a partir de la Ecuación de Blasius,

$$f_M = 0.316 \, Re^{-0.25}$$

apropiada para flujo turbulento en tubos lisos.

Escribir una función (function), *PDELTA*, que podría ser llamada a través de

ISBN: 950-42-0022-2 - ©1999

la sentencia: *DELTAP = PDELTA (Q, D, L, RHO, MU, E)* donde el valor de *PDELTA* representa la caída de presión para el flujo volumétrico (caudal) *Q* de un fluido con densidad *RHO* y viscosidad *MU* a través de un tubo de longitud *L*, diámetro interior *D* y rugosidad *E*. Observe que *MU* y *L deben* definirse como variables reales.

Escriba un programa que lea los valores de *Q*, *D*, *L*, *RHO*, *MU* y *E*, llame a *PDELTA* para calcular la caída de presión, imprima los datos y resultados y vuelva a leer otro conjunto de datos.

Datos sugeridos para prueba:

Q	gal/min	170	4
D	pulgada	3.068	0.622
L	pie	10000	100
RHO	lb_m/pie^3	62.4	80.2
MU	lb_m/pie seg.	0.0007	0.05
E	pulgada	0.002	0.0005

P7) Sea el siguiente sistema de ecuaciones lineales:

$$\begin{aligned} 9x_1 + 3x_2 + 5x_3 &= -3 \\ 2x_2 - x_3 &= 0 \\ 4x_1 - 3x_2 + 7x_3 &= -1 \end{aligned}$$

Encontrar la solución por los siguientes métodos:

a) Gauss
b) Jacobi
c) Gauss-Seidel
d) Sustitución directa
e) Wegstein

P8) Sea el siguiente sistema:

$$\begin{aligned} 9x_1 + 2x_2 &= 3 \\ 2x_1 + 3x_2 + x_3 &= 4 \\ 5x_2 + 4x_3 + 8x_4 &= -8 \\ 3x_3 - x_4 &= 0 \end{aligned}$$

Encontrar la solución mediante el método de Gauss. ¿Puede aprovechar la estructura particular del sistema para facilitar el cálculo?.

ISBN: 950-42-0022-2 - ©1999

P9) Sea el siguiente sistema:

$$\begin{aligned} 2\,x + 3\,y - z &= 0 \\ 4\,x - 3\,y - 9\,z &= 2 \\ 3\,x - 3\,y - z &= 1/3 \end{aligned}$$

Aplicar el método de Gauss-Seidel para encontrar la solución.

P10) Sea el siguiente sistema expresado en forma matricial:

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 3 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}$$

Encontrar los valores para Δx_1, Δx_2 y Δx_3 por los siguientes métodos:

a) Gauss
b) Gauss-Seidel
c) Jacobi

P11) Confeccionar un diagrama de flujos para un programa que ingrese como datos la matriz de coeficientes y el término independiente de un sistema de ecuaciones lineales de dimensión *N*, y calcule la solución. Suponer que dispone de subrutinas de inversión de matrices.

P12) Utilizando software comercial (rutinas IMSL, Numerical Recipes, etc.) confeccionar el programa correspondiente al diagrama de flujos anterior.

BIBLIOGRAFÍA RECOMENDADA

- Carnahan, B., H. A. Luther y J. O. Wilkes, *Applied Numerical Methods*, John Wiley and Sons, Inc., New York (1969).
- Cohen, A. M., J. F. Cutts, R. Fielder, D. E. Jones, J. Ribbans y E. Stuart, *Análisis Numérico*, Editorial Reverté S. A., Barcelona (1977).
- Hamming, R. W., *Numerical Methods for Scientists and Engineers*, McGraw-Hill, New York (1962).
- Hildebrand, F. B., *Introduction to Numerical Analysis*, McGraw-Hill, New York (1956).
- Hornbeck, R. W., *Numerical Methods*, Quantum Publishers, Inc., New York.
- Lapidus, L., *Digital Computation for Chemical Engineers*, McGraw-Hill, New York (1962).

ISBN: 950-42-0022-2 - ©1999

- Luthe, R., A. Olivera y F. Schutz, *Métodos Numéricos*, Editorial Limusa, México (1978).
- Milne, W. E., *Numerical Solution of Differential Equations*, John Wiley and Sons, New York (1953).
- O'Neill, P. V., *Advanced Engineering Mathematics*, 3ra. Ed., Wadsworth, Inc., USA (1991).
- Perry, J. H., *Chemical Engineers Handbook*, McGraw-Hill, New York.
- Ralston, A. y P. Rabinowitz, *First Course in Numerical Analysis*, McGraw-Hill Inc., USA (1978).

ISBN: 950-42-0022-2 - ©1999