Mètodes Numèrics I

Àlex Martorell LoCascio

Versió no definitiva

Grau de Matemàtiques 2n curs

Semestre de Primavera 2018

Assignatura impartida per Susana Romano

Índex

1 Problemes numèrics i errors

Notacions 1. És convenient començar donant una sèrie de notacions àmpliament usades:

- $a \ll b$ vol dir "a és molt més petit que b" (depèn del context)
- $a \simeq b$ vol dir "a aproximadament igual a b" i significa el mateix que $|a - b| \leq c$, on c s'escull adequat al context.
- $f(x) = O(g(x))$ amb $x \to a$ vol dir $\dfrac{f(x)}{g(x)}$ està acotat quan $x \to a$, és a dir $\lim_{x\to a} \dfrac{f(x)}{g(x)} \leq C < +\infty$
- $f(x) = o(g(x))$ amb $x \to a$ vol dir $\lim_{x\to a} \dfrac{f(x)}{g(x)} = 0$
- $f(x) \sim g(x)$ quan $x \to a$ si $\lim_{x\to a} \dfrac{f(x)}{g(x)} = 1$

Exemple 1. Tenim un problema de mida n i $T(n) = 4n^2 - 2n + 2$ és el nombre d'operacions. Si n és gran, aleshores $T(n) = O(n^2)$

Exemple 2. $e^x = 1 + x + \dfrac{x^2}{2} + O(x^2)$, d'altra banda $e^x - (1 + x + \dfrac{x^2}{2}) = o(x^2)$

Fonts de l'error

1. Error en les dades d'entrada: mesures incorrectes o finitud en la representació d'una dada.
2. Errors d'arrodoniment en els càlculs.
3. Errors de trucament del mètode utilitzat.
   - depèn del mètode.
   - és l'error produït en tallar un procés infinit (error de truncament, error de discretització)
4. Errors per la simplificació del model matemàtic i del model numèric.

1.1 Error absolut i error relatiu. Fites

Sigui x el valor exacte d'una certa magnitud i $\bar{x}$ un valor aproximat de x.

Definició 1. Es defineix **l'error absolut** en $\bar{x}$ com $e_a(x) = \bar{x} - x$. Escrivim també $e_a(\bar{x})$ i Δx

Definició 2. Es defineix **l'error relatiu** en $\bar{x}$ com $e_r(x) = \dfrac{e_a(x)}{x} \quad (x \neq 0)$
A la pràctica usarem: $e_r(x) \simeq \dfrac{e_a(x)}{x}$ Notem que $e_r(x)$ sovint s'expressa en tant per cent.

Exemples 3. Alguns exemples usant les definicions:

1. $x = \sqrt{3} = 1.7320508\ldots \quad \bar{x} = 1.732$
   $e_a(\sqrt{3}) = -5.08\ldots \cdot 10^{-5} \quad e_r(\sqrt{3}) \simeq 3 \cdot 10^{-5}$

2. $\pi = 3.1415926535 \quad \bar{\pi} = 3.1416$
   $e_a(\pi) = 3.1416 - 3.1415926535... \simeq 7.346 \cdot 10^{-6}$
   $e_r(\pi) \simeq 7.346 \cdot 10^{-6}/\pi \simeq 2.4 \cdot 10^{-6}$

3. Digues quina de les tres és la mesura més exacta si l'error absolut és de 1 cm $x_1 = 4, x_2 = 40, x_3 = 400$
   Clarament x_3 és la mesura més exacta ja que té l'error relatiu més petit.

Definició 3. $\varepsilon_a(x)$ és una fita de l'error absolut en $\bar{x}$ si $|e_a(x)| \leq \varepsilon_a(x)$

Definició 4. $\varepsilon_r(x)$ és una fita de l'error relatiu en $\bar{x}$ si $|e_r(x)| \leq \varepsilon_r(x)$

Exemple 4. Usant l'exemple 3.2, tenim $|e_a(\pi)| \leq 10^{-5}$ o $|e_a(x)| \leq 8 \cdot 10^{-6}$
Farem servir les següents notacions:

$$x = \bar{x} \pm \varepsilon_a(x)$$
$$\bar{x} = x \pm \varepsilon_a(x)$$

Interpretació:

$$|\bar{x} - x| \leq \varepsilon_a(x)$$
$$-\varepsilon_a(x) \leq \bar{x} - x \leq \varepsilon_a(x)$$
$$x - \varepsilon_a(x) \leq \bar{x} \leq x + \varepsilon_a(x)$$

i escrivim $\bar{x} = x \pm \varepsilon_a(x)$ fent un abús de notació.

Exemple 5. Si escrivim $x = 0.5432 \pm 0.0013$ això indica que $0.5419 \leq x \leq 0.5445$
Anàlogament per una **fita de l'error relatiu**:

$$x = \bar{x}(1 \pm \varepsilon_r(x))$$
$$\bar{x} = x(1 \pm \varepsilon_r(x))$$

Interpretació:

$$\left|\frac{\bar{x} - x}{x}\right| \leq \varepsilon_r(x)$$
$$|\bar{x} - x| \leq |x|\varepsilon_r(x)$$
$$-|x|\varepsilon_r(x) \leq \bar{x} - x \leq |x|\varepsilon_r(x)$$
$$x - |x|\varepsilon_r(x) \leq \bar{x} \leq x + |x|\varepsilon_r(x)$$

i obtenim $\bar{x} = x(1 \pm \varepsilon_r(x))$

1.2 Representació d'un nombre en punt flotant

Donada una base $b \geq 2$, una representació en punt flotant d'un nombre $x \neq 0$ ve donada per

$$x = (-1)^s \cdot m \cdot b^e$$

on e és un nombre enter, $s \in \{0, 1\}$ i m, anomenada mantissa, és $m = d_0.d_1d_2d_3\ldots$ amb $d_j \in \mathbb{N}$ i $0 \leq d_j \leq b, \forall j$. Si afegim $d_0 \neq 0$, és a dir $1 \leq m < b$ direm que la representació està **normalitzada**.

Nota. S'ha vist a l'assignatura de Programació Científica que per imprimir un nombre en pantalla fem servir , per exemple (en el cas d'un double) `printf("%23.15e", x);` és a dir`%p.q` tal que `p` $>$ `q` $+ 7$, on `p` és l'espai reservat i `q` el nombre de decimals. de la mantissa. Si només disposem de $t + 1$ dígits i la mantissa té més dígits, no podem guardar el nombre exactament. Codificarem:

$$fl(x) = (-1)^s \cdot \bar{m} \cdot b^e$$

La codificació aproximada es pot fer per:

1. tall, suprimint els dígits a partir de d_t

2. arrodoniment, codificant en funció del dígit d_{t+1}:

$$\begin{cases} \text{si} \quad d_{t+1} < \frac{1}{2}b \quad \text{prenem} \quad \bar{m} = d_0 \cdot d_1 d_2 d_3 \dots d_t \\ \text{si} \quad d_{t+1} \geq \frac{1}{2}b \quad \text{prenem} \quad \bar{m} = d_0 \cdot d_1 d_2 d_3 \dots d_t + b^{-t} \end{cases}$$

Exemple 6. Donat $\pi = 3.14159265358979323 84...$:

$fl(\pi) = 3.141592653 \cdot 10^0$ (si tallem)

$fl(\pi) = 3.141592654 \cdot 10^0$ (si arrodonim)

Fites en la representació en punt flotant Si $x = \pm m \cdot b^e$ i $fl(x) = \pm \bar{m} \cdot b^e$, on $\bar{m}$ té $t+1$ dígits:

- **tall**: $\varepsilon_a(m) = b^{-t}$ i $\varepsilon_r(m) = b^{-t}$
- **arrodoniment**: $\varepsilon_a(m) = \frac{1}{2}b^{-t} = \varepsilon_r(m) = \frac{1}{2}b^{-t}$

S'ometen demostracions.

Decimals correctes i dígits significatius

Sigui x el valor exacte i $\bar{x}$ el valor aproximat.

Definició 5. $\bar{x}$ aproxima x amb t **decimals correctes** si $t \in n$ és tal que $|e_a(x)| \leq \dfrac{1}{2} \cdot 10^{-t}$

Definició 6. Si $\bar{x}$ aproxima x amb t decimals correctes, llavors tots els dígits en posicions superiors a 10^{-t} són **dígits significatius** (excepte els zeros de davant)

Exemples 7. (de les definicions):

- $\bar{x}_1 = 0.001234 \pm 0.5 \cdot 10^{-5}$ $\quad \varepsilon_a(x_1) = 0.5 \cdot 10^{-5} \leq \frac{1}{2} \cdot 10^{-5}$ té 5 decimals correctes. S'observa que la fita de l'error està en el 3. En notació científica el nombre és $1.234e{-3}$. Té 3 dígits signficatius.
- $\bar{x}_2 = 56.789 \pm 0.5 \cdot 10^{-3}$ $\quad \varepsilon_a(x_2) = \frac{1}{2} \cdot 10^{-3}$. Aproxima x_2 amb 3 decimals correctes i 5 dígits signficatius.
- $x_3 = 1, \bar{x}_3 = 0.9997$ $\quad |e_a(x_3)| = 0.3 \cdot 10^{-3} \leq \frac{1}{2} \cdot 10^{-3}$ $\quad$ 3 decimals correctes.
  Si $\bar{x}'_3 = 0.9992$ $\quad |e_a(x_2)| = 0.8 \cdot 10^{-3} \leq \frac{1}{2} \cdot 10^{-2}$ $\quad$ 2 decimals correctes.

1.3 Propagació de l'error

Volem conèixer com es propaga l'error de les dades quan amb elles es realitza alguna de les operacions aritmètiques elementals.
Siguin x, y valors exactes i $\bar{x}, \bar{y}$ aproximacions.

Suma

$$e_a(x+y) = (\bar{x}+\bar{y}) - (x+y) = (\bar{x}-x) + (\bar{y}-y) = e_a(x) + e_a(y)$$
$$e_a(x-y) = (\bar{x}-\bar{y}) - (x-y) = (\bar{x}-x) - (\bar{y}-y) = e_a(x) - e_a(y)$$
$$|e_a(x+y)| = |e_a(x) + e_a(y)| \leq |e_a(x)| + |e_a(y)| \leq \varepsilon_a(x) + \varepsilon_a(y)$$
$$|e_a(x-y)| = |e_a(x) - e_a(y)| \leq |e_a(x)| + |e_a(y)| \leq \varepsilon_a(x) + \varepsilon_a(y)$$

Per tant, $\varepsilon(x \pm y) = \varepsilon_a(x) + \varepsilon_a(y)$. És important notar l'ús de la desigualtat triangular en la 3a i 4a fórmula. Anàlogament, per a l'error relatiu:

$$\varepsilon_r(x+y) = \left|\frac{x}{x+y}\right|\varepsilon_r(x) + \left|\frac{y}{x+y}\right|\varepsilon_r(y)$$
$$\varepsilon_r(x-y) = \left|\frac{x}{x-y}\right|\varepsilon_r(x) + \left|\frac{y}{x-y}\right|\varepsilon_r(y)$$

Deducció de la fórmula:

$$\varepsilon_r(x+y) = \frac{\varepsilon_a(x+y)}{|x+y|} = \frac{\varepsilon_a(x)+\varepsilon_a(y)}{|x+y|} = \frac{|x|\varepsilon_r(x)}{x+y} + \frac{|y|\varepsilon_r(y)}{x+y}$$

Producte

$$e_a(x \cdot y) = \bar{x}\bar{y} - x \cdot y = (x + e_a(x))(y + e_a(y)) - xy = ye_a(x) + xe_a(y) + e_a(x)e_a(y)$$

$$e_r(x \cdot y) = \frac{e_a(x \cdot y)}{x \cdot y} = \frac{e_a(x)}{x} + \frac{e_a(y)}{y} + \frac{e_a(x)}{x} \cdot \frac{e_a(y)}{y} = e_r(x) + e_r(y) + e_r(x)e_r(y)$$
$$\simeq e_r(x) + e_r(y)$$

Observem que a la segona fórmula es menyspreen els termes no lineals. Per trobar una fita de l'error relatiu, fem $|e_r(xy)| \simeq |e_r(x)+e_r(y)| \leq |e_r(x)|+|e_r(y)|$ d'on $\varepsilon_r(xy) = \varepsilon_r(x)+\varepsilon_r(y)$

Divisió

$$e_a\left(\frac{x}{y}\right) = \frac{\bar{x}}{\bar{y}} - \frac{x}{y} = \frac{x + e_a(x)}{(y + e_a(y)} - \frac{x}{y} = \ldots = \frac{1}{1 + e_r(y)} \cdot \frac{ye_a(x) - xe_a(y)}{y^2}$$

$$e_r\left(\frac{x}{y}\right) = e_r(x) - e_r(y) \Longrightarrow \left|e_r\left(\frac{x}{y}\right)\right| \leq |e_r(x)| + |e_r(y)|$$

Les deduccions omeses es deixen com a exercici.

Propagació de l'error en l'avaluació de funcions

Donat un x que coneixem amb un cert error, volem calcular $y = f(x)$ on $f : [a, b] \longrightarrow \mathbb{R}$ és una funció contínua i derivable en (a, b). Comencem recordant un resultat d'anàlisi:

Teorema 1. (del valor mitjà): Sigui $f : [a, b] \subset \mathbb{R} \longrightarrow R$ contínua en $[a, b]$ i derivable en (a, b). Llavors $\exists p \in [a, b]$ tal que $f(a) - f(b) = f'(p)(a - b)$.

En el nostre cas, $\exists \theta \in < x, \bar{x} >$ tal que $f(\bar{x}) - f(x) = f'(\theta)(\bar{x} - x)$. Això és $|e_a(y)| = |f'(\theta)||e_a(x)| \leq |f'(\theta)|\varepsilon_a(x) \simeq |f'(\bar{x})|\varepsilon_a(x)$

Per a funcions de diverses variables, sigui $x \in S \subset \mathbb{R}^n$ i $y = f(x) = f(x_1, \ldots, x_n)$ diferenciable en un entorn de x. La fórmula de la propagació de l'error és:

$$\bar{y} - y = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\bar{x}_k)\Delta x_i$$

La fita és:

$$\varepsilon_a(f(x_1, \ldots, x_n)) \simeq \sum_{k=1}^{n} \left|\frac{\partial f}{\partial x_k}(\bar{x})\right|\varepsilon_a(x_k)$$

Exemple 8. (1 variable): Sigui $f(x) = \log(x)$. Usant $\varepsilon_a(f(x)) \simeq |f'(\bar{x})|\varepsilon_a(x)$ obtenim:

$$\varepsilon_r(log(x)) = \frac{\varepsilon_a(x)}{|\log(x)|} = \frac{\varepsilon_r(x)}{|\log x|}$$

Per exemple: $x = 2.39, \bar{x} = 2.4$. Aleshores $\log(x) = 0.87129 \quad \log(\bar{x}) = 0.87546$. Obtenim $\varepsilon_r(x) = \frac{0.01}{2.39} \simeq 4.1841 \cdot 10^{-3}$ i $\varepsilon_a(\log(x)) = 4.17 \cdot 10^{-3}$

Exercici 1. Calcular l'efecte de l'error en f en l'avaluació de les funcions següents:
$y = \sqrt{x} \quad y = \sin(x)$

Exemples 9. (en diverses variables)

1. $f(x_1, x_2) = x_1 + x_2 \quad \dfrac{\partial f}{\partial x_1}(x_1, x_2) = 1 = \dfrac{\partial f}{\partial x_2}(x_1, x_2)$

$$\varepsilon_a(f(x_1, x_2)) \simeq \left|\frac{\partial f}{\partial x_1}(\bar{x})\right|\varepsilon_a(x_1) + \left|\frac{\partial f}{\partial x_2}(\bar{x})\right|\varepsilon_a(x_2)$$

2. $f(x_1, x_2) = x_1 \cdot x_2 \quad \dfrac{\partial f}{\partial x_1}(x_1, x_2) = x_2 \quad \dfrac{\partial f}{\partial x_2}(x_1, x_2) = x_1$

$$\varepsilon_a(f(x_1, x_2)) \simeq \varepsilon_a(x_1)|\bar{x}_2| + \varepsilon_a(x_2)|\bar{x}_1|$$

3. Considerem un cilindre de radi $r = 3 \pm 0'02$cm i $h = 5 \pm 0'01$cm. Recordem que la fórmula del volum d'un cilindre és $V = \pi r^2 h$ i escrivim $V(r, h) = \pi r^2 h$. Calculem les derivades parcials i surt $\dfrac{\partial V}{\partial r}(h, r) = 2\pi r h$ i $\dfrac{\partial V}{\partial r}(h, r) = \pi r^2$

$$\varepsilon_a(V) \simeq |2\pi \bar{r}\bar{h}|\varepsilon_a(r) + |\pi \bar{r}^2|\varepsilon_a(h) = 0'69 \cdot \pi$$

Per tant, $V = 45\pi \pm 0'69$ cm^3

Cancel·lació de termes

Aquest fenomen apareix en restar dues quantitats molt properes, es perd precisió. Siguin x, y tal que:

$$x = d_0 d_1 \dots d_s d_{s+1} \dots d_t \cdot 10^p$$
$$y = d_0 d_1 \dots \alpha_s \alpha_{s+1} \dots \alpha_t \cdot 10^p$$

Aleshores $x - y = 0'0 \dots 0\beta_{s+1} \dots \beta_t \cdot 10^p = \beta_{s+1}\beta_{s+2} \dots \beta_t \cdot 10^{p-s+1}$ Per tant, s'han perdut $t - (s+1)$ dígits. Usant la fórmula de l'error relatiu

$$\varepsilon_r(x) = \frac{\varepsilon_a(x - y)}{|x - y|} = \frac{\varepsilon_a(x) + \varepsilon_a(y)}{|x - y|}$$

Veiem que quan $x \sim y$ o $x - y \simeq 0$ o $|x - y| \ll 1$, aleshores cal buscar una fórmula millor numèricament.

Algunes tècniques per evitar cancel·lacions:

- Racionalització $\sqrt{x^2 + 1} - 1 = \dfrac{x^2}{\sqrt{x^2 + 1}} + 1$
- Sèrie de Taylor $x - \sin x$ quan $x \simeq 0$. Fem Taylor: $\sin x = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \dfrac{x^7}{7!}$

- Trigonomètriques $\cos(2\alpha) = \cos^2(\alpha) - \sin^2(\alpha) \quad \alpha = \frac{\pi}{4}$

Exemple 10. Resolució d'una equació de segon grau:

Recordem, tot i que no és necessari, que donada una equació $ax^2 + bx + c = 0$ aquesta té solucions $x_{1,2} = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. Observem que si $b^2 \gg 4ac$ aleshores $\sqrt{b^2 - 4ac} \simeq |b|$. Posant $q \simeq |b|$ tenim $\dfrac{-b \pm q}{2a}$ (1). És a dir, en un dels $x_{1,2}$ hi haurà problemes, perquè es produeix cancel·lació.

Dit això, posem un exemple numèric. Es considera l'equació $x^2 - 62.10x + 1 = 0$. Aquesta té solucions $x_1 = -1.6107237 \cdot 10^{-2}$ i $x_2 = -6.2083893 \cdot 10^1$. Anem a resoldre-la usant 4 dígits + arrodoniment:

$$\sqrt{(62.10)^2 - 4.000} = \sqrt{3856 - 4.000} = \sqrt{3852} \quad (62.10^2 = 3856.41)$$
$$\sqrt{3852} \simeq 62.0644825 \simeq 62.06$$
$$\sqrt{b^2 - 4ac} \simeq 62.06$$

Substituint a (1) obtenim $\bar{x}_1 = \dfrac{-62.10 + 62.06}{2.000} = \dfrac{-4.000 \cdot 10^{-2}}{2.000} = -2.000 \cdot 10^{-2}$

Decimals correctes: $|e_a(x_1)| = |\bar{x}_1 - x_1| \simeq 3.893 \cdot 10^{-2} \leq \frac{1}{2} \cdot 10^{-2} \Rightarrow$ 2 decimals correctes.

Substituint de nou a (1) obtenim $\bar{x}_2 = \dfrac{-62.10 - 62.06}{2.000} = \dfrac{-124.2}{2.000} = -62.10$

Alternatives per evitar cancel·lació en aquest problema:

1. Racionalitzar: $\dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a} \cdot \dfrac{-b - \sqrt{b^2 - 4ac}}{-b - \sqrt{b^2 - 4ac}} = \dfrac{-2a}{b + \sqrt{b^2 - 4ac}}$. Per calcular x_1, no x_2

2. Usar la relació de les arrels: $ax^2 + bx + c = 0 \Rightarrow a(x - x_1)(x - x_2)$ amb x_1, x_2 solucions. Se sap $x_1 x_2 = \frac{c}{a} x_1 + x_2 = \frac{-b}{a}$. Podem calcular primer la que no té cancel·lacions i l'altra fent servir una d'aquestes fórmules.

Usant les dades del cas anterior. Sabem $x_1 = 0.0161$. Aleshores $x_1 x_2 = \frac{1}{1}$. D'on $x_2 = \frac{1}{\bar{x}_1} = -1.610 \cdot 10^{-2}$ (aproximació més bona)

Propagació de l'error en punt flotant

Donats x, y, volem calcular $x \circ y$ (on $\circ$ denota qualsevol operació aritmètica) Sigui $fl(x \circ y)$ el valor aproximat. Estudiarem els dos casos següents:

1.No hi ha error en la representació de x i y:

$$e_a(x\circ y) = fl(x\circ y) - x\circ y$$
$$fl(x\circ y) = x\circ y + e_a(x\circ y) = (x\circ y)(1+\frac{e_a(x\circ y)}{x\circ y}) = (x\circ y)(1+e_r(x\circ y))$$

En termes de fites tenim:

$$fl(x\circ y) = (x\circ y)(1\pm\varepsilon_r(x\circ y)) = x\circ y \pm |x\circ y|\varepsilon_r(x\circ y)$$

Exemple 11. Volem calcular $x_1+x_2+x_3$, la suma es fa amb un error absolut fitat per ε
Es fa $(x_1+x_2)+x_3$
$x_1,x_2 \longrightarrow y_1 = fl(x_1+x_2) = (x_1+x_2)(1+\delta_1) \quad |\delta_1|\le\varepsilon$
$y_1,x_3 \longrightarrow y_2 = fl(y_1+x_3) = (y_1+x_3)(1+\delta_2) \quad |\delta_2|\le\varepsilon$

$$fl((x_1+x_2)+x_3)) = ((x_1+x_2)(1+\delta_1)+x_3)(1+\delta_2) = x_1+x_2+x_3+(x_1+x_2)\delta_1$$
$$+(x_1+x_2+x_3)\delta_2+(x_1+x_2)\delta_1\delta_2$$

Fitem l'error:

$$|fl(x_1+x_2+x_3)-(x_1+x_2+x_3)| \le |x_1+x_2|\delta_1+|x_1+x_2+x_3|\delta_2+|x_1+x_2||\delta_1||\delta_2|$$
$$\lesssim |x_1+x_2|\varepsilon+|x_1+x_2+x_3|\varepsilon$$

Una altra fita és $\varepsilon_a(x_1+x_2+x_3)\simeq(2|x_1|+2|x_2|+2|x_3|)\varepsilon$
2. Si hi ha error en la representació de x i y: Es considera $fl(x), fl(y)$

$$fl(x\circ y) = (fl(x)\circ fl(y))(1+e_r(x\circ y)) = fl(fl(x)\circ fl(y)) =$$
$$[x(1\pm\varepsilon_r(x)\circ(y(1\pm\varepsilon_r(y)](1\pm\varepsilon_r(x\circ y)]$$

Exemple 12. Volem calcular $(x_1+x_2)+x_3$, tenint en compte que hi ha error en la representació i també en la suma, tots fitats per ε.
Primerament, $x_i \longrightarrow y_i$ amb $fl(x_i) = x_i(1+\delta_i) \quad i=1,2,3,\ |\delta_i|\le\varepsilon$
$y_1+y_2 = z_1 = fl(y_1+y_2) = (y_1+y_2)(1+\delta_4) \quad |\delta_4|\le\varepsilon$
$z_1+y_3 = z_2 = fl(z_1+y_3) = (z_1+y_3)(1+\delta_5) \quad |\delta_5|\le\varepsilon$
Veiem que $z_1 = f(y_1,y_2,\delta_4)$
Recordem que $e_a(f(x_1,\dots,x_n))\simeq\sum\frac{\partial f}{\partial x_i}e_a(x_i)$
$|e_a(f(x_1,\dots,x_n))|\simeq\sum|\frac{\partial f}{\partial x_i}\varepsilon_a(x_i)$

Aleshores:

$$\varepsilon_a(z_1) = \varepsilon_a(f(y_1, y_2, \delta_4)) \simeq \left|\frac{\partial f}{\partial x_i}\right| \epsilon_a(y_1) + \left|\frac{\partial f}{\partial y_2}\varepsilon_a(y_2)\right| + \left|\frac{\partial f}{\partial \delta_4}\right|$$

$$\simeq |1+\delta_4|\varepsilon(y_1) + |1+\delta_4|\varepsilon_a(y_2) + |y_1+y_2|\varepsilon(\delta_4)$$

$$\simeq \varepsilon_a(y_1) + \varepsilon_a(y_2) + |y_1+y_2|\varepsilon_a(\delta_4) = |x_1|\varepsilon + |x_2|\varepsilon + |y_1+y_2|\varepsilon$$

$$\varepsilon_a(z_2) = \varepsilon_a(f(z_1, y_3, \delta_5)) \simeq \left|\frac{\partial f}{\partial z_1}\right| \varepsilon_a(z_1) + \left|\frac{\partial f}{\partial y_3}\varepsilon_a(y_3)\right| + \left|\frac{\partial f}{\partial \delta_5}\right|$$

$$= |1+\delta_5|\varepsilon_a(z_1) + |1+\delta_5|\varepsilon_a(y_3) + |z_1+y_3|\varepsilon_a(\delta_5) \simeq \varepsilon_a(z_1) + \varepsilon_a(y_3) + |x_1+x_2+x_3|\varepsilon_a(\delta_5)$$

$$\simeq (2|x_1| + 2|x_2|)\varepsilon + |x_3|\varepsilon + |x_1+x_2+x_3|\varepsilon \leq (3|x_1| + 3|x_2| + 2|x_3|)\varepsilon$$

1.4 Inestabilitat i estabilitat numèriques

Es considera la següent integral:

$$\int_0^1 x^n e^{x-1} dx \quad n = 1, 2 \ldots$$

integrant per parts obtenim $E_n = 1 - nE_{n-1}$. Fent els càlculs amb 6 xifres s'obté que per a $n = 9$, $E_9 < 0$!!. És per aquest motiu que parlem de **inestabilitat numèrica**. Cal buscar un algorisme que no sigui inestable.

Considerem $E_{n-1} = \dfrac{1-E_n}{E_n}$. En aquest cas notem que cal tenir un E_n per aplicar aquest algorisme (iterem cap enrere). Prenem un n prou gran ($E_n \simeq 0$). Tenim $E_{20} = 0$. Calculem $|e_a(E_{21})| \leq \frac{1}{21} \simeq 4'76 \cdot 10^{-2}$. Calculem $|e_a(E_{19})| \leq \frac{1}{21 \cdot 20} \simeq 0'0024$. Per a fer això, hem observat que:

$$E_n = \int_0^1 x^n e^{x-1} \leq \int_0^1 x^n dx = \frac{1}{n+1}$$

1.5 Problemes ben o mal condicionats

Si el resultat final d'un cert algorisme és molt sensible a petites variacions en les dades inicials, es diu que l'algorisme està mal condicionat o que és numèricament inestable.

Si aquesta dependència sensible del resultat final respecte les dades inicials és intrísneca al problema que volíem resoldre es diu que el *problema està mal condicionat* o que és (matemàticament) inestable.

1. Es consideren els següents sistemes d'equacions lineals

a) $\begin{cases} 2'000x + 0'6667y = 2'6667 \\ 1'000x + 0'333y = 1'3333 \end{cases}$ b) $\begin{cases} 2'000x + 0'6665y = 2'6667 \\ 1'000x + 0'333y = 1'3333 \end{cases}$

c) $\begin{cases} 2'000x + 0'6666y = 2'6667 \\ 1'000x + 0'333y = 1'3333 \end{cases}$ d) $\begin{cases} 2'000x + 0'6667y = 2'6666 \\ 1'000x + 0'333y = 1'3333 \end{cases}$

Solucions:

a) $x = 1 \; y = 1$

b) $x = 1'6667 \quad y = -1'000$

c) no té solució

d) infinites solucions

2. Polinomi de Wilkinson (1963)

Sigui $p(x) = (x-1)(x-2)\cdot\ldots\cdot(x-20) = x^{20} - 210x^{19} + \ldots$ que té arrels $1, 2\ldots, 20$ Es considera $p(x) + 2^{-23}x^{19}$ Notem $2^{-23} \simeq 1'19 \cdot 10^{-7}$. Aquest nou polinomi passa a tenir 10 solucions reals i 10 solucions complexes.

2 Àlgebra lineal numèrica

2.1 Sistemes d'equacions lineals

Recordem que un sistema de n equacions lineals amb n incògnites es pot escriure com:

$$\begin{cases} a_{11} + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21} + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1} + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n \end{cases}$$

Assumim que tots els coeficients són reals. Escrivim el sistema en forma matricial com $Ax = b$

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{22} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} \qquad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

on A és una matriu $n \times n$ i B un vector $n \times 1$.

Definició 7. Una matriu A s'anomena **regular** o **no singular** si $\det A \neq 0$.

El problema de la resolució de sistemes d'equacions és equivalent a represntar b com una combinació lineal dels vectors columna de A $a_1, \ldots a_n$ És a dir, el sistema té solució si i només si els vectors columna són una base de $\mathbb{R}^n$ o equivalentment que els vectors columna són linealment independents. És a dir, la matriu A és regular. Per tant, donada A regular, l'existència i la unicitat de la solució x està garantida.

Els mètodes per a resoldre sistemes lineals es divideixen en dos grans grups:

- Mètodes **directes**: Es troba la solució en un nombre finit de passos
- Mètodes **iteratius**: Es construeix una successió de vectors $x^{(0)}, x^{(1)}, x^{(2)}, \ldots$ que convergeix a la solució.

Comentari. En aquesta assignatura s'estudien els mètodes directes més importants. Els mètodes iteratius es veuran a l'assignatura de Mètodes Numèrics II.

Ens interessrà comptar el nombre d'operacions elementals $(+, -, *, /)$ que requereix fer un procés determinat. Per exemple, si volem comptar el nombre d'operacions que cal fer quan fem un producte escalar de dos vectors $x, y \in \mathbb{R}^n$ veiem que calen n productes i n sumes.

Exercici 2. Sigui A una matriu de dimensió 10×20, B de dimensió 20×30 i C de dimensió 30×40. Calcular el nombre de productes en fer $(AB)C$.

2.2 Sistemes triangulars

(a) Sistemes triangulars inferiors $Ly = b$ (L és lower, inferior en anglès)

$$\begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & & \ddots & \\ l_{n1} & \dots & \dots & l_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Tenim $\det L = \prod_{i=1}^{n} l_{ii} \neq 0 \Longrightarrow l_{ii} \neq 0 \quad \forall i$

Per resoldre aquest sistema fem servir l'algoritme de **substitució endavant**:

$$y_1 = \frac{b_1}{l_{11}}$$
$$y_i = \frac{b_i - \sum_{k=1}^{i-1} l_{ik} y_k}{l_{ii}} \qquad i = 2, \dots, n$$

- Sistemes triangulars superiors $Ux = y$ (U, de Upper)

$$\begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ & u_{22} & & \vdots \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Tenim $\det U = \prod_{i=1}^{n} u_{ii} \neq 0 \Longrightarrow u_{ii} \neq 0 \quad \forall i$

Apliquem l'algoritme de substiució endarrera:

$$x_n = \frac{y_n}{u_{nn}}$$
$$y_i = \frac{y_i - \sum_{k=1}^{i-1} u_{ik} x_k}{u_{ii}} \qquad i = n-1, \dots, 1$$

Nombre d'operacions en l'algoritme de substitució endavant:

operació	pas i	total
$+,-$	$i-1$	$\sum_{i=1}^{n}(i-1)=\frac{n(n-1)}{2}$
$*$	$i-1$	$\sum_{i=1}^{n}(i-1)$
$/$	1	$\sum_{i=1}^{n}1=n$

Notem que el cost de resoldre un sistema triangular d'ordre n és aproximadament proporcional a n^2

2.3 Eliminació gaussiana

L'objectiu és transformar el sistema lineal $Ax=b$ en un sistema $Ux=c$, triangular superior. Aquest darrer sistema es resoldrà mitjançant l'algoritme de substitució endarrera.
El mètode d'eliminació gaussiana consta de $n-1$ passos i, en cada pas $k=1,\ldots,n-1$, s'intenta eliminar x_k de les últimes $n-k$ equacions.

- Pas k-è. En iniciar la k-èna etapa del procés, el sistema és de la forma:

$$\left\{\begin{array}{cccccccc}
a_{11}^{(1)}x_1+ & a_{12}^{(1)}x_2+ & \ldots & +a_{1k}^{(1)}x_1+ & \ldots & +a_{1n}^{(1)}x_n & = & b_1^{(1)} \\
 & a_{22}^{(2)}x_2+ & \ldots & +a_{2k}^{(2)}x_k+ & \ldots & +a_{2n}^{(2)}x_n & = & b_2^{(2)} \\
 & & \ddots & \vdots & & \vdots & & \vdots \\
 & & & a_{kk}^{(k)}x_k+ & \ldots & +a_{kn}^{(k)}x_n & = & b_k^{(k)} \\
 & & & a_{k+1,k}^{(k)}x_k+ & \ldots & +a_{k+1,n}^{(k)}x_n & = & b_{k+1}^{(k)} \\
 & & & \vdots & & \vdots & & \vdots \\
 & & & a_{n,k}^{(k)}x_k+ & \ldots & +a_{nn}^{(k)}x_n & = & b_n^{(k)}
\end{array}\right.$$

Si el pivot $a_{kk}^{(k)}\neq 0$, podem eliminar x_k calculant primer els multiplicadors (m_{ik}) i després restant les equacions de manera adeuqada

$$m_{ik}=\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} \qquad i=k+1,\ldots,n$$

$$a_{ij}^{(k+1)}=a_{ij}^{(k)}-m_{ik}a_{kj}^{(k)} \qquad i=k+1,\ldots,n \quad j=k+1,\ldots,n$$

$$b_i^{(k+1)}=b_i^{(k)}-m_{ik}b_k^{(k)} \qquad i=k+1,\ldots,n$$

Comentaris.

1. A l'etapa k, només canviem les files de la $k+1$ a la n

2. Per a cada k i cada i, calculem cada m_{ik} adecuat de manera que posem un 0 a la posició (i,k). Notem que el càlcul per obtenir el 0 no es fa (seria el cas $j=k$).

3. Per a poder fer l'etapa k cal $a_{kk}^{(k)} \neq 0 \quad k = 1, \dots n-1$

4. ( ) indica a partir de quina etapa el terme en qüestió no es modifica

La pregunta que es planteja és si existeixen condicions per a poder assegurar (3).

Proposició 1. Per a cada k, sigui A_k la matriu $k \times k$, interssecció de k primeres files i k primeres columnes de A. Si $\det A_k \neq 0 \quad \forall k = 1, \dots, n-1$ aleshoers els sucessius pivots $a_{kk} \neq 0$.

Nota. A la pràctica no s'intenta comprovar les condicions sinó que es va fent eliminació gaussiana i abans de cadascuna comprovem $a_{kk}^{(k)} \neq 0$.

Recompte d'operacions:

operació	pas k	total
$/$	$\sum_{i=k+1}^{n} 1$	$\sum_{k=1}^{n-1}(\sum_{i=k+1}^{n} 1) = \frac{n(n-1)}{2}$
$*, -$		$\frac{n(n-1)(2n-1)}{6} + \frac{n(n-1)}{2}$

Recordatori. $1^2 + 2^2 + \ldots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$

Notes.

- S'ha posat a la mateixa fila tant els productes com les restes perquè se'n fa la mateixa quantitat.

- Per calcular el nombre de productes o restes fem:

$$\sum_{k=1}^{n-1}\sum_{i=k+1}^{n}\left[\left(\sum_{j=k+1}^{n} 1\right)+1\right] = \sum_{k=1}^{n-1}\sum_{i=k+1}^{n}\Big[(n+1)-k\Big] = \sum_{k=1}^{n-1}(n-k+1)\left(\sum_{i=k+1}^{n} 1\right) =$$

$$= \sum_{l=1}^{n-1} l(l+1) = \sum_{l=1}^{n-1} l^2 + \sum_{l=1}^{n-1} l = \frac{n(n-1)(2n-1)}{6} + \frac{n(n-1)}{2} = \frac{n^3-n}{3}$$

  on $l = n-k$.

Exemple 13. Es vol resoldre mitjançant el mètode d'eliminació gaussiana el sistema de matriu ampliada (treballem amb 4 xifres de mantissa i arrodoniment):

$$\left(\begin{array}{ccc|c} 0.15 & 2.11 & 30.75 & -26.38 \\ 0.64 & 1.21 & 2.05 & 1.01 \\ 3.21 & 1.53 & 1.04 & 5.23 \end{array}\right)$$

Pas 1. Veiem que $a_{11}^{(1)} = 0.15 \neq 0$. Calculem multplicadors: $m_{21} = \frac{a_{21}^{(1)}}{a_{11}^{(1)}} = \frac{0.64}{0.15} = 4.267$ $m_{31} = \frac{a_{31}^{(1)}}{a_{11}^{(1)}} = \frac{3.21}{0.15} = 21.4$. Fem el pas d'eliminació gaussiana i queda:

$$\left(\begin{array}{ccc|c} 0.15 & 2.11 & 30.75 & -26.38 \\ 0 & -7.793 & -129.2 & 113.6 \\ 0 & -43.62 & -657.0 & 569.8 \end{array}\right)$$

Pas 2. (hi ha n-1 passos) Veiem que $a_{22}^{(2)} = -7.793 \neq 0$ (puc fer eliminació gaussiana). Calculem multplicadors: $m_{32} = \frac{a_{32}^{(2)}}{a_{22}^{(2)}} = \frac{-43.62}{-7.793} = 5.597$. Fem el pas d'eliminació gaussiana i queda:

$$\left(\begin{array}{ccc|c} 0.15 & 2.11 & 30.75 & -26.38 \\ 0 & -7.793 & -129.2 & 113.6 \\ 0 & 0 & 66.13 & -66.02 \end{array}\right)$$

Com s'observa, obtenim un sistema de la forma $Ux = c$. Apliquem l'algoritme de **substitució endarrera**:

$$x_3 = \frac{-66.02}{66.13} = -0.9983$$

$$x_2 = \frac{113.6 - 129.2 \cdot -0.9983}{-7.793} = 1.974$$

$$x_1 = \frac{-26.38 - (2.11 \cdot 1.974 + 30.75 \cdot (-0.9983))}{0.15} = 1.017$$

Nota. Existeix una variant de d'eliminació gaussiana anomenada el **mètode de Jordan** (en algunes assignatures, mètode de Gauss-Jordan). S'obté un sistema diagonal(en comptes d'un triangular superior). La seva importància és relativa, ja que no aporta un millor cost pel que fa al nombre d'operacions.

2.4 Pivotatge

Sigui la matriu $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 4 & 5 & 6 \end{pmatrix}$. Per poder aplicar eliminació gaussina, cal $a_{kk}^{(k)} \neq 0$. Veiem que aquesta condició ja no se satisfà pel primer pas. Aquest cas motiva la necessitat de pivotatge amb aritmètica exacta. Així doncs, si $a_{kk}^{(k)} = 0$, una estratègia natural és intercanviar files, la fila k amb una fila $i > k$ tal que $a_{ik} \neq 0$. Es planteja si existeixen sempre pivots $\neq 0$ per poder aplicar pivotatge. Necessitem la següent proposició.

Proposició 2. Si A és una matriu no singular, $\exists i$ tal que $a_{ik}^{(k)} \neq 0$.

Demostració. Comencem notant que el pas d'eliminació gaussiana es pot escriure matricialment, per exemple $A^{(1)} \longrightarrow A^{(2)}$. Tenim:

$$A^{(2)} = M_1 A^{(1)} \quad \text{on} \quad M_1 = \begin{pmatrix} 1 & & & \\ -m_{21} & 1 & & \\ \vdots & & \ddots & \\ -m_{n1} & & & 1 \end{pmatrix} \quad \text{,la resta 0}$$

Així doncs, escrivim

$$A^{(k)} = M_{k-1} A^{(k-1)} \Rightarrow$$

$$\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -m_{k,k-1} & \ddots & & \\ & & \vdots & & \ddots & \\ & & -m_{n,k-1} & & & 1 \end{pmatrix} \left(\begin{array}{ccccc|c} a_{11}^{(1)} & \dots & \dots & & a_{1n}^{(1)} & a_{1,n+1}^{(1)} \\ & \ddots & & & \vdots & \vdots \\ & & a_{k-1,k-1}^{(k-1)} & \dots & \dots \; a_{k-1,n}^{(k-1)} & a_{k-1,n+1}^{(k-1)} \\ & & \vdots & & \vdots & \vdots \\ & & \vdots & & \vdots & \vdots \\ & & a_{n,k-1}^{(k-1)} & \dots & \dots \; a_{n,n}^{(k-1)} & a_{n,n+1}^{(k-1)} \end{array}\right) =$$

$$\left(\begin{array}{ccccc|c} a_{11}^{(1)} & \dots & \dots & & a_{1n}^{(1)} & a_{1,n+1}^{(1)} \\ & \ddots & & & \vdots & \vdots \\ & & a_{k,k}^{(k)} & \dots & \dots \; a_{k-1,n}^{(k)} & a_{k,n+1}^{(k)} \\ & & \vdots & & \vdots & \vdots \\ & & \vdots & & \vdots & \vdots \\ & & a_{n,k}^{(k)} & \dots & \dots \; a_{n,n}^{(k)} & a_{n,n+1}^{(k)} \end{array}\right)$$

Anomenem $\widetilde{A}$ a la submatriu formada per files $k \div n$ i columnes $k \div n$.

Hem vist que $A^k = M_{k-1}A^{(k-1)} = M_{k-1}M_{k-2} \cdot \ldots \cdot M_1 A^{(1)}$. Anomenem $M = M_{k-1}M_{k-2} \cdot \ldots \cdot M_1$. Prenent determinants tenim $\det M \det A = a_{11}^{(1)} \cdot \ldots \cdot a_{k-1k-1}(k-1) \det \widetilde{A} = \det A$ (Notem que $\det M = 1$)

Aleshores, si per a tota $i \geq k$ tenim $\widetilde{A} = \begin{pmatrix} 0 & & \\ \vdots & & \\ 0 & & \end{pmatrix}$ això implica $\det \widetilde{A} = 0 \Rightarrow \det A = 0$, una contradicció.

És a dir, si A és no singular $\exists i \geq k$ tal que $a_{ik}^{(k)} \neq 0 \quad \square$

Exemple 14. Volem resoldre el sistema següent amb matriu ampliada (4 xifres):

$$\left(\begin{array}{cc|c} 0.0001 & 0.5 & 0.5 \\ 0.4 & -0.3 & 0.1 \end{array}\right)$$

Pas 1. $a_{11}^{(1)} = 0.0001 \neq 0$ Multiplicador: $m_{21} = \frac{0.4}{0.0001} = 4000$. Calculem $a_{22}^{(2)} = fl(a_{22}^{(1)} - m_{21}a_{12}^{(1)}) = fl(-0.3 - 4000(0.5)) = fl(-2000.3) = -2000$ (4 xifres)

$$\left(\begin{array}{cc|c} 0.0001 & 0.5 & 0.5 \\ 0.4 & -2000 & -2000 \end{array}\right)$$

Solució usant substitució endarrera: $x_1 = 0.000, \quad x_2 = 1.000$, una solució **incorrecta.**

Solució **exacta** (amb 4 xifres) : $x_1 = 0.9999, \quad x_2 = 0.9998$

Aquest exemple mostra que no usar pivotatge pot donar solucions incorrectes.

Hem vist que cal intentar que els pivots no siguin nombres massa petits. En el pas k, agafem $a_{kk}^{(k)} \neq 0$.

Definim el **pivotatge maximal per columnes** com el procés de buscar a la columna k l'element més gran en magnitud. És a dir, busquem $\max\limits_{i=k\div n} |a_{ik}^{(k)}| = |a_{rk}|$. Aleshores:

- Si $r = k$ fem el pas d'eliminació gaussiana
- Si $r \neq k$ intercanviem les files k i r i fem el pas d'eliminació gaussiana. Notem $|m_{ik}| = \left|\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}\right| \leq 1$

Exemple 15. Resoldrem el mateix sistema de l'exemple 13 usant pivotage maximal:

$$\left(\begin{array}{ccc|c} 0.15 & 2.11 & 30.75 & -26.38 \\ 0.64 & 1.21 & 2.05 & 1.01 \\ 3.21 & 1.53 & 1.04 & 5.23 \end{array}\right)$$

Pas 1. Intercanviem la 1a fila i la 3a. Fem el pas d'eliminació gaussiana i queda:

$$\left(\begin{array}{ccc|c} 3.21 & 1.53 & 1.04 & 5.23 \\ 0 & 0.9049 & 1.834 & -0.03826 \\ 0 & 2.039 & -3.070 & -26.62 \end{array}\right)$$

Pas 2. (hi ha $n-1$ passos) Intercanviem la 2a fila i la 3a. Fem el pas d'eliminació gaussiana i queda:

$$\left(\begin{array}{ccc|c} 3.21 & 1.53 & 1.04 & 5.23 \\ 0 & 2.039 & 3.070 & -26.62 \\ 0 & 0 & -11.78 & 11.78 \end{array}\right)$$

Aplicant l'algoritme de substitució endarrera obtenim $x_1 = 0.9995$, $x_2 = 2.001$, $x_3 = -1.000$

2.5 Factorizació LU (Descomposició)

Sigui A una matriu $n \times n$. Suposem que es pot fer la següent descomposició $A = LU$ on L és una matriu triangular inferior i U és una matriu triangular superior. Detallem algunes utilitats d'aquesta descomposició a continuació:

1. El sistema $Ax = b$ és equivalent a resoldre el següents sistemes en aquest ordre $Ly = b$ i $Ux = y$.
2. Càlcul del determinant de A: $\det A = \det L \cdot \det U = (\prod_{i=1}^{n} l_{ii})(\prod_{i=1}^{n} u_{ii})$
3. Si A és regular el càlcul de la inversa és equivalent a resoldre n sistemes tal que $AX = Id$. Però podem usar $A^{-1} = (LU)^{-1} = U^{-1}L^{-1}$

Cal preguntar-se també per l'existència i la unicitat de $A = LU$. Si comptem incògnites de L i U obtenim que cada matriu presenta $\frac{n(n+1)}{2}$ incògnites, que sumades donen $n(n+1)$. Eliminem n incògnites imposant que L tingui 1 a la diagonal.

Definició 8. Sigui A una matriu $n \times n$. Si $A = LU$ on L és una matriu triangular inferior amb 1's a la diagonal i U és triangular superior, es diu que A té **factorització LU.**

Teorema 2. (Existència i unicitat de la descomposició LU)
Sigui A una matriu $n \times n \quad \forall k = 1 \div n-1$, sigui A_k la matriu $k \times k$ formada per la intersecció de les k primeres files i columnes de A. Si es verifica $\det A_k \neq 0 \quad \forall k = 1 \div n-1$ aleshores A admet una única factorització LU.

Demostració. Es fa per inducció respecte n i treballant per blocs

Per a $n = 1$ és trivial ja que $(a_{11}) = (1) = (a_{11})$.

Per a $n = 2$ tenim $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} \\ l_{21}u_{12} & l_{21}u_{12} + u_{22} \end{pmatrix}$

Observem $a_{11} = u_{11} \neq 0$, $\frac{a_{21}}{u_{11}} = l_{21}$ i $u_{22} = a_{22} - l_{21}u_{12}$.

Suposem cert fins a $n - 1$.

$$\left(\begin{array}{c|c} A_{n-1} & b \\ \hline c^T & a_{nn} \end{array}\right) = \left(\begin{array}{c|c} L_{n-1} & 0 \\ \hline l^T & 1 \end{array}\right) \left(\begin{array}{c|c} U_{n-1} & u \\ \hline 0^T & u_{nn} \end{array}\right) = \left(\begin{array}{c|c} L_{n-1}U_{n-1} & L_{n-1}u \\ \hline l^T u_{n-1} & l^T u + u_{nn} \end{array}\right)$$

Per hipòtesis inductiva, existeix una única $L_{n-1}U_{n-1}$

Observem $b = L_{n-1}U$. Per tant existeix un única u. De la mateixa manera, $c^T = l^T u_{n-1}$ i per tant existeix una única l. Finalment, trobem u_{nn} restant el que ja és conegut. c.v.d.

Notes.

1. Si $\det A_k = 0$ per algun k no és possible trobar la descomposició LU.

2. Si eliminem 1's de la matriu L, la unicitat falla.

Descomposició LU per a matrius banda

Sigui una matriu A $n \times n$, que és banda (p, q). Si fem eliminació gaussiana, es conserva l'estructura banda.

2.6 Matrius simètriques i definides positives. Factorització de Cholesky

Definició 9. Una matriu simètrica A $n \times n$ és definida positiva si $\forall x \neq 0$ amb $x \in \mathbb{R}^n$ es té $x^T Ax > 0$

Caracteritzacions:

1. A és simètrica i definida positiva si i només si tots els valors propis de A són positius.

2. A és simètrica i definida positiva si i només si $\det A_k > 0 \quad k = 1 \div n$ (menors principals) on A_k és la matriu formada per les k primeres files i les k primeres columnes d'A.

Proposició 3. Si A és simètrica i es pot fer l'eliminació gaussiana sense pivotatge, llavors totes les matrius $\bar{A}^{(k)} = \left(a_{ij}^{(k)}\right)_{k \leq ij \leq n}$ són simètriques $k = 1 \div n$.

Demostració. Exercici.

Teorema 3. (existència i unicitat de la descomposició de Cholesky) Si A $n \times n$ simètrica i definida posititiva, llavors $\exists!$ una matriu R triangular inferior amb r_{ii} reals positius tals que $A = RR^T$. Anàlogament $\exists!$ U triangular superior tal que $A = U^T U$.

Demostració. (Idea):

Sabem que $\det A_k > 0$ $k = 1 \div n$, per tant $\exists!$ L triangular inferior amb $l_{ii} = 1$ i U triangular superior tal que $A = LU$. Es demostra $u_{kk} > 0$. Definim $D =$diag(u_{ii}) (matriu diagonal amb elements u_{ii}. Aleshores $A = LU = LDD^{-1}U = LDL^T$ $(D^{-1}U = L^T)$. Definim $D^{\frac{1}{2}}$ que consisteix en la matriu diagonal D els elements de la qual han estat elevats a $\frac{1}{2}$. Finalment, $A = LDL^T = LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T$ i anomeno $R = LD^{\frac{1}{2}}$, triangular inferior (per ser producte de triangulars inferiors) i els elements són reals i positius. Resta demostrar la unicitat. □

Càlcul de determinants i d'inverses

Ens interessa calcular el determinant d'una matriu. Ara, tenim les següents opcions:

- Descomposició LU. Tenim que $A = LU \Rightarrow \det A = \det L \cdot \det U = 1 \cdot u_{11} \cdot \ldots \cdot u_{nn}$
- $A = LDL^T$
- $A = RR^T$

Per calcular la inversa:

- Resolució de n sistemes lineals. $A^{-1} = X = (X_{(1)}, \ldots, X_{(n)})$ Fem $AX_{(i)} = e_i$, $i = 1 \div n$, on e_i és el vector i-èsim de la base canònica.
- Càlcul de LU, L^{-1}, U^{-1} i tenim $A^{-1} = U^{-1}L^{-1}$

2.7 Errors en àlgebra lineal numèrica

Sigui V un espai vectorial sobre $\mathbb{R}$.

Definició 10. La norma vectorial sobre V és una aplicació $\|\cdot\|: V \longrightarrow \mathbb{R}$ tal que:

1. $\|x\| > 0 \quad \forall x \in V$
2. $\|x\| = 0 \Longleftrightarrow x = 0$
3. $\|\alpha x\| = |\alpha| \|x\| \quad \forall x \in V \quad \forall \alpha \in \mathbb{R}$
4. Desigualtat triangular: $\|x + y\| \leq \|x\| + \|y\|$

Alguns exemples de normes són:

- $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ (suma de mòduls)
- $\|x\|_2 = \left(\sum_{i=1}^{n} |x_i| \right)^{\frac{1}{2}}$ (norma euclidiana)
- $\|x\|_\infty = \max_{1\le i\le n} |x_i|$ (norma del màxim)

Definició 11. Donada una norma vectorial $\|\cdot\|$ a $\mathbb{R}^n$ es defineix la norma matricial associada com:

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$$

que compleix les condicions de norma i dues propietats més:

(i) $\|AB\| \le \|A\|\|B\|$

(ii) $\|Ax\| \le \|A\|\|x\|$

Exemples de normes matricials:

- $\|A\|_1 = \max_{1\le j\le n} \left(\sum_{i=1}^{n} |a_{ij}| \right)$ (màxim de la suma per columnes)
- $\|A\|_2 = \sqrt{\rho(A^T A)}$
- $\|A\|_\infty = \max_{1\le i\le n} \left(\sum_{j=1}^{n} |a_{ij}| \right)$ (màxim de la suma per files)

Observació. Donada una matriu B, $\rho(B)$ és el radi espectral, definit com el valor propi amb mòdul màxim.

Teoria de perturbacions

La idea és la següent: Donat un sistema $Ax = b$, volem estudiar com afecta a la solució del sistema la modificació de b. I si modifiquem A?
Abans d'enunciar els resultats teòrics, veiem el següent exemple.

Exemple 16. Siguin els següents sistemes $Ax = b$ amb solucions ja calculades:

$$A = \begin{pmatrix} 1 & -11 \\ -9 & 100 \end{pmatrix} \qquad b_1 = \begin{pmatrix} 10 \\ -1 \end{pmatrix} \qquad x_1 = \begin{pmatrix} 989 \\ 89 \end{pmatrix}$$

Amb la perturbació següent $\bar{b}_1 = \begin{pmatrix} 9 \\ -1 \end{pmatrix}$ obtenim $\bar{x}_1 = \begin{pmatrix} 889 \\ 80 \end{pmatrix}$. Un error al voltant del 10% (fet que quadra) Si considerem ara:

$$b_2 = \begin{pmatrix} -1 \\ 10 \end{pmatrix} \qquad x_2 = \begin{pmatrix} 10 \\ 1 \end{pmatrix}$$

Amb la perturbació $\bar{b}_2 = \begin{pmatrix} 9 \\ -1 \end{pmatrix}$ obtenim $\bar{x}_2 = \begin{pmatrix} 889 \\ 80 \end{pmatrix}$. Fem notar que en aquest cas un 10% d'error produeix molt d'error !!

Per estudiar l'error, recuperem les definicions del Tema 1. Com que ara $x \in \mathbb{R}^n$ (solució del sistema) definirem l'error absolut com $\|e_a(x)\| = \|x - \bar{x}\| = \|\Delta x\|$ L'error relatiu el definim com $\dfrac{\|\Delta x\|}{\|x\|} = \dfrac{\|x - \bar{x}\|}{\|x\|}$

Volem resoldre $Ax = b$ amb A no singular i $b \neq 0$. Estudiarem els tres casos possibles:

1. Pertorbem $b \to b + \delta b$

   Obtenim un sistema de la forma següent: $A(x + \delta x) = b + \delta b$. Volem estudiar $\dfrac{\|\Delta x\|}{\|x\|}$:

$$Ax + A\delta x = b + \delta b$$
$$\delta x = A^{-1}\delta b$$
$$\|\delta x\| = \|A^{-1}\delta b\| \leq \|A^{-1}\|\|\delta b\|$$

   Per altra banda, $\|b\| = \|Ax\| \leq \|A\| \cdot \|x\|$. D'on $\dfrac{1}{\|x\|} \leq \dfrac{\|A\|}{\|b\|}$. Usant els dos fets:

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\|\|\delta b\|}{\|x\|} \leq \|A\|\|A^{-1}\|\frac{\|\delta b\|}{\|b\|}$$

   $\kappa(A) = \|A\|\|A^{-1}\|$ és el *nombre de condició* de A. Donada aquesta definició, podem escriure:

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\delta b\|}{\|b\|}$$

   Es diu que A està ben condicionada si $\kappa(A)$ és petit. Sempre és $\kappa(A) \geq 1$. Notem $\kappa(I) = \|I\|\|I\| = 1$. A està mal condicionada si $K(A) >> 1$

2. Pertorbem A: Obtenim un sistema de la forma següent: $(A + \Delta A)(x + \Delta x) = b$. Prenent normes:

$$\|\Delta x\| = \|A^{-1}\| \cdot \|\Delta A\| \cdot \|x + \Delta x\|$$
$$\frac{\|\Delta x\|}{\|x + \Delta x\|} \leq \|A\|\|A^{-1}\| \cdot \frac{\|\Delta A\|}{\|A\|}$$

3. Pertorbem A i b: Obtenim un sistema de la forma següent: $(A+\Delta A)(x+\Delta x) = b+\Delta b$

   És possible demostrar que l'error relatiu en x queda fitat per:

$$\frac{\|\Delta x\|}{\|x+\Delta x\|} \leq \kappa(A)\left(\frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|} \cdot \frac{\|x+\Delta x\|}{\|x\|}\right)$$

3 Interpolació polinomial i aplicacions

Donats $(x_k, y_k)_{k=0\div n}$ punts del pla $x_i \neq x_j$ si $i \neq j$ volem trobar una funció p tal que $p(x_k) = y_k \quad k = 0 \div n$. Es diu que p **interpola** les dades.

Si són els punts de la gràfica de la funció, volem saber el valor de la funció en $x \neq x_i \quad \forall i$.

En el cas $x \in \left[\min\limits_{i=0\div n} x_i, \max\limits_{i=0\div n} x_i\right]$ s'anomena **interpolació**. El cas contrari, en què $x \notin \left[\min\limits_{i=0\div n} x_i, \max\limits_{i=0\div n} x_i\right]$ s'anomena **extrapolació**.

Qüestions

1. La funció p de quin tipus és?

    - Si la funció original sembla periòdica, buscarem p en les funcions periòdiques
    - Si té asímptotes, en les funcions racionals

    En aquesta assignatura només considerarem polinomis. És per aquest motiu que el tema s'anomena interpolació polinomial.

2. Un cop tenim el tipus, existeix p ? És única?

3. És una bona aproximació de la funció fora dels $(x_k, y_k)_{k=0\div n}$?

3.1 Interpolació de Lagrange

Donats $(x_k, f_k) \quad k = 0 \div n$ amb $x_i \neq x_j$ si $i \neq j$ busquem un polinomi $p_n(x)$ de grau $\leq n$ tal que $p_n(x_k) = f_k$ $k = 0 \div n$. $p_n(x)$ s'anomena el polinomi interpolador de Lagrange.

Teorema 4. (d'existència i unicitat) Donats (x_k, f_k) $k = 0 \div n$ $(x_i \neq x_j$ si $i \neq j)$ existeix un polinomi de grau $\leq n$, $p_n(x)$, que resol el problema.

Demostració. El polinomi $p_n(x) = a_0 + a_1 x + \ldots + a_n x^n$ té grau $\leq n$.

Imposem $p_n(x_k) = f_k \quad k = 0 \div n$. Obtenim:

$$\left\{\begin{array}{lllll} f_0 & = & p_n(x_0) & = & a_0 + a_1 x_0 + \ldots + a_n x_0^n \\ \vdots & & & & \\ f_n & = & p_n(x_n) & = & a_0 + a_1 x_n + \ldots + a_n x_n^n \end{array}\right.$$

És a dir:

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{pmatrix}$$

Calculem el determinant de la matriu:

$$\det M = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = V(x_0, \dots, x_n) = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & x_1 - x_0 & x_1(x_1 - x_0) & \dots & x_1^{n-1}(x_1 - x_0) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n - x_0 & x_n(x_n - x_0) & \dots & x_n^{n-1}(x_n - x_0) \end{vmatrix}$$

$$= \begin{vmatrix} x_1 - x_0 & \dots & x_1^{n-1}(x_1 - x_0) \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x_n - x_0 & \dots & x_n^{n-1}(x_1 - x_0) \end{vmatrix} = (x_1 - x_0) \cdots (x_n - x_0) \begin{vmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^{n-1} \end{vmatrix} =$$

$$\prod_{j \neq 0} (x_j - x_0) V(x_1 \dots, x_n) = \prod_{j<i} (x_i - x_j) = \det M$$

On $V(x_0, \dots, x_n)$ és el determinant de Vandermonde. El sistema d'equacions plantejat té solució i és única. cvd.

Mètodes de càlcul del polinomi interpolador de Lagrange

1. Resolució del sistema lineal que apareix en imposar les condicions (teoria sabuda)

2. Mètode de Lagrange

   Considerem els polinomis de Lagrange

   $$L_j(x) = \frac{(x - x_0) \cdot \ldots \cdot (x - x_{j-1})(x - x_{j+1}) \cdot \ldots \cdot (x - x_n)}{(x_j - x_0) \cdot \ldots \cdot (x_j - x_{j-1})(x_j - x_{j+1}) \cdot \ldots \cdot (x_j - x_n)} = \prod_{i \neq j} \frac{x - x_i}{x_j - x_i} \quad j = 0 \div n$$

   té grau n. Es té $L_j(x_k) = \delta_{jk} = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$ El polinomi interpolador serà $p_n(x) = \sum_{j=0}^{n} f_j L_j(x)$

3. Mètode de les diferències dividides de Newton

   Escriurem el polinomi com $p_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \ldots + a_n(x - x_0) \cdot$

$\ldots \cdot (x - x_{n-1})$ Observem que té grau $\leq n$. Abans de definir el polinomi interpolador, ens cal el concepte de **diferència dividida**. Notem $f[x_i] = f_i \quad i = 0 \div n$.

$$f[x_i, x_{i+1}] = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i} \quad \text{diferència dividida entre dos punts}$$

$$f[x_i, x_{i+1}, x_{i+2}] = \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i}$$

La generalització d'aquest concepte és:

$$f[x_i, x_{i+1}, \ldots, x_{i+k}] = \frac{f[x_{i+1}, \ldots, x_{i+k}] - f[x_i, \ldots, x_{i+k-1}]}{x_{i+k} - x_i}$$

Per calcular diferències dividides es fa servir el següent esquema triangular, que ho fa més senzill

x_i	f_i		
x_0	$f_0 = f[x_0]$		
		$f[x_0, x_1]$	
x_1	$f_1 = f[x_1]$		$f[x_0, x_1, x_2]$
		$f[x_1, x_2]$	
x_2	$f_2 = f[x_2]$		

Teorema 5. Donats $(x_k, f_k) \quad k = 0 \div n$ (amb $x_i \neq x_j$ si $i \neq j$). Sigui $p_n(x) = a_0 + a_1(x - x_0) + \ldots + a_n(x - x_0) \cdot \ldots \cdot (x - x_{n-1})$ el polinomi interpolador en $(x_k, f_k) \quad k = 0 \div n$, aleshores $a_i = f[x_0, \ldots, x_i] \quad i = 0 \div n$.

Demostració. Donarem només una idea de la demostració. Per fer-ho, necessitem el següent lema que si que demostrarem.

Lema 1. Siguin, per a $x_i \quad i = 0 \div n$ ($x_j \neq x_k$ si $j \neq k$)

$$P^{(1)}(x) \text{ polinomi interpolador de } f \text{ en } (x_0, f_0), \ldots, (x_{n-1}, f_{n-1})$$

$$P^{(2)}(x) \text{ polinomi interpolador de } f \text{ en } (x_1, f_1), \ldots, (x_n, f_n)$$

llavors $P^{(n)}(x) = P^{(1)}(x) + \frac{x - x_0}{x_n - x_0}(P^{(2)}(x) - P^{(1}(x))$ és el polinomi interpolador de f en $(x_0, f_0), \ldots, (x_n, f_n)$.

Demostració.

$$P_n(x_0) = P^{(1)}(x_0) = f_0$$

$$P_n(x_i) = P^{(1)}(x_i) + \frac{x_i - x_0}{x_n - x_0}\left[P^{(2)}(x_i) - P^{(1)}(x_i)\right] = P^{(1)}(x_i) = f_i$$

$$P_n(x_n) = P^{(1)}(x_n) + \frac{x_n - x_0}{x_n - x_0}\left[P^{(2)}(x_n) - P^{(1)}(x_n)\right] = P^{(2)}(x_n) = f_n \quad \square$$

Ara, per demostrar el teorema (4), fem inducció sobre n.

$$n = 0$$

$$p_0(x) = a_0 \quad f_0 = p_0(x_n) = a_0 \Rightarrow a_0 = f[x_0]$$

$$n = 1$$

$p_1(x) = a_0 + a_1(x - x_0)$ i $f_0 = a_0, f_1 = a_0 + a_1(x - x_0)$ obtenim

$$a_0 = f[x_0] \quad a_1 = \frac{f[x_1] - f[x_0]}{x_1 - x_0} = f[x_0, x_1]$$

Ho suposem cert fins a grau $n - 1$:

$P^{(1)}(x)$ polinomi interpolador de f en $(x_0, f_0), \ldots, (x_{n-1}, f_{n-1})$

$P^{(2)}(x)$ polinomi interpolador de f en $(x_1, f_1), \ldots, (x_n, f_n)$

Per tant:

$$P^{(1)}(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, \ldots, x_{n-1}](x - x_0) \cdot \ldots \cdot (x - x_{n-2})$$
$$P^{(2)}(x) = f[x_1] + f[x_1, x_2](x - x_1) + \ldots + f[x_1, \ldots, x_n](x - x_0) \cdot \ldots \cdot (x - x_{n-1})$$

i apliquem el lema anterior $\square$

Notació 2. $< x_0, \ldots, x_k >$ designa el mínim interval que conté tots $x_0, \ldots, x_k$.

Recordatori. El *Teorema de Rolle* diu que donada $f : [a, b] \to \mathbb{R}$ contínua en $[a, b]$, derivable en (a, b) i que compleix $f(a) = f(b)$ aleshores $\exists c \in (a, b)$ tal que $f'(c) = 0$.

Lema 2. Si $f \in \mathcal{C}^k([a, b])$, amb (x_i, f_i), $i = 0 \div n$ abscisses diferents; totes pertanyent a l'interval $[a, b]$ aleshores $\exists \xi \in < x_0, \ldots x_k >$ tal que $f[x_0, \ldots, x_k] = \dfrac{f^{(k)}(\xi)}{k!}$

Demostració. Sigui $p(x)$ el polinomi interpolador de f en $x_0, \ldots, x_k$. És a dir $p(x) = f[x_0] + f[x_0, x_1](x - x_0) + \ldots + f[x_0, \ldots, x_k](x - x_0) \cdot \ldots \cdot (x - x_{k-1})$

Considerem $d(x) = f(x) - p(x)$. Clarament $d \in \mathcal{C}^k([a, b])$ i $d(x_i) = 0$, $i = 0 \div n$.

Sabent això últim, separem en k intervals tals que $x_0 < x_1 < \ldots < x_k$. Apliquem el teorema de Rolle i obtenim que $d'(x)$ s'anul·la en almenys k punts. Aleshores $d''(x)$ s'anul·la en almenys $k-1$ punts. Reiterant, veiem que $d^{(k)}(x)$ s'anul·la en almenys 1 punt, ξ. Finalment:

$$0 = d^{(k)}(\xi) = f^{(k)}(\xi) - p^{(k)}(\xi) = f^{(k)}(\xi) - k! f[x_0, \ldots, x_k] \Rightarrow f[x_0, \ldots, x_k] = \frac{f^{(k)}(\xi)}{k!}$$

c.v.d.

Teorema 6. (fórmula de l'error en la interpolació de Lagrange)

Siguin $f \in \mathcal{C}^{n+1}([a,b])$, $x_0, \ldots, x_n$ abscisses diferents $\in [a,b]$ i $p_n(x)$ el polinomi interpolador de grau $\leq n$ de la funció f en $x_0, \ldots, x_n$, llavors $\forall x \in [a,b]$ existeix $\xi(x)$ tal que:

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}(x - x_0) \cdot \ldots \cdot (x - x_n)$$

Demostració. Sigui $\bar{x} \in [a,b]$. El cas en què existeix una i tal que $\bar{x} = x_i$ és trivial.

Suposem $\forall i \quad \bar{x} \neq x_i$, $\bar{p}(x)$ el polinomi interpolador de grau $\leq n+1$ de f en $x_0, \ldots, x_n, \bar{x}$.

És a dir: $\bar{p}(x) = p_n(x) + f[x_0, \ldots, x_n, \bar{x}](x - x_0) \cdot \ldots \cdot (x - x_n)$

Aplicant el lema anterior, existeix $\xi(\bar{x})$ tal que $f[x_0, \ldots, x_n, \bar{x}] = \dfrac{f^{(n+1)}(\xi(\bar{x}))}{(n+1)!}$

$$0 = f(\bar{x}) - p(\bar{x}) = f(\bar{x}) - \left[p_n(\bar{x}) + \frac{f^{(n+1)}(\xi(\bar{x}))}{(n+1)!}(\bar{x} - x_0) \cdot \ldots \cdot (\bar{x} - x_n)\right]$$

c.v.d.

Exemple 17. Per $n = 6$, considerem la funció $f(x) = \sin x$ i punts equiespaiats en $[a,b] = [0,9]$ Els punts són de la forma $x_i = a + ih \quad i = 0 \div n$ i $h = \frac{9}{6} = \frac{3}{2}$.

Volem calcular $f(4) - p(4)$. Aplicant el teorema que ens dona la fórmula de l'error, això és:

$$\begin{aligned} f(4) - p(4) &= \frac{f^{(7)}(\xi(4))}{7!}(4 - x_0) \cdot \ldots \cdot (4 - x_6) \\ |f(4) - p(4)| &= \left|\frac{f^{(7)}(\xi(4))}{7!}\right| |(4 - x_0) \cdot \ldots \cdot (4 - x_6)| \leq \frac{1}{7!} 175 = 3.47 \cdot 10^{-2} \end{aligned}$$

on hem usat $f^{(7)}(x) = -\cos x$ i $|f^{(7)}(x)| \leq 1$. Notem $|(4 - x_0) \cdot \ldots \cdot (4 - x_6)| = 175$.

3.2 Interpolació d'Hermite

Problema. Donats (x_i, f_i), $i = 0 \div n$ punts de la gràfica d'una funció i les pendents en aquests punts f_i, $i = 0 \div n$, busquem $p(x)$ polinomi de grau $\leq 2n+1$ tal que

$$p(x_i) = f_i$$

$$p'(x_i) = f_i'$$

per $i = 0, \ldots n$

Teorema 7. (d'existència i unicitat de la interpolació d'Hermite)

Donats (x_k, f_k), $k = 0 \div n$, punts de la gràfica de $f(x)$ amb $x_i \neq x_k$ si $i \neq k$ i les pendents en aquests punts $f'_k = f'(x_k) \quad k = 0 \div n$, llavors existeix un únic polinomi $p(x)$, de grau menor o igual a $2n+1$ tal que:

$$p(x_k) = f_k \quad k = 0, 1, 2, \ldots, n$$
$$p'(x_k) = f'_k \quad k = 0, 1, 2, \ldots, n$$

Demostració. Unicitat: Suposem $p(x), q(x)$ de grau $\leq 2n+1$ Imposem, per a $i = 0 \div n$

$$p(x_i) = q(x_i) = f_i$$
$$p'(x_i) = q'(x_i) = f'_i$$

Definim $R(x) = p(x) - q(x)$ de grau $\leq 2n+1$. Clarament $R(x_i) = 0$ i $R'(x) = p'(x) - q'(x) \Rightarrow R'(x_i) = 0 \quad i = 0 \div n$. Notem que R té $2(n+1) = 2n+2$ zeros, és a dir $R(x) = 0$. Per tant $p(x) = q(x)$ i la unicitat queda demostrada.

Existència: Mitjançant el mètode de Lagrange generalitzat. Recordem que els polinomis de Lagrange són de la forma

$$L_j(x) = \prod_{i \neq j} \frac{x - x_i}{x_j - x_i} \quad j = 0 \div n$$
$$L_j(x_k) = \delta_{jk}, \qquad \text{amb } L_j \text{ de grau n}$$

Definim

$$\phi_i(x) = (1 - 2(L'_i(x - x_i))L_i^2(x) \quad i = 0 \div n$$
$$\psi_i(x) = (x - x_i)L_i^2(x)$$

Es comprova el següent: $\phi_i(x_i) = 1$ $\phi_i(x_j) = 0$ $\psi_i(x_i) = 0$ $(\psi_i(x_j) = 0$ $\phi'_i(x_j) = 0$ i $\psi'_i(x_j) = \delta_{ij}$. El polinomi interpolador és:

$$p(x) = \sum_{j=0}^{n} f_j \phi_j(x) + \sum_{j=0}^{n} f'_j \psi_j(x)$$

on $p(x)$ té grau $\leq 2n+1$ A més a més:

$$p(x_k) = \sum_{j=0}^{n} f_j \phi_j(x_k) + \sum_{j=0}^{n} f'_j \psi_j(x_k) = f_k \phi_j(x_k) + 0) = f_k \quad k = 0 \div n$$

Per la derivada de $p(x)$:

$$p'(x) = \sum_{j=0}^{n} f_j \phi'_j(x) + \sum_{j=0}^{n} f'_j \psi'_j(x_k)$$
$$p'(x_k) = 0 + f'_k$$

c.v.d.

Mètodes de càlcul

1. Resolució del sistema lineal que surt de plantejar les condicions

2. Mètode de Lagrange generalitzat (molt semblant al que ja s'ha explicat)

3. Mètode de les diferències dividides generalitzades

   Volem calcular la diferència dividida $f[x_i, x_i]$. Notem el següent:

   $$f[x_i, x_i] = \lim_{x \to x_i} f[x_i, x] = \lim_{x \to x_i} \frac{f[x] - f[x_i]}{x - x_i} = \lim_{x \to x_i} \frac{f(x) - f(x_i)}{x - x_i} = f'(x_i) = f'_i$$

   Un cop feta aquesta observació, per calcular el polinomi interpolador farem servir el següent esquema triangular:

x_i	f_i					
x_0	$f[x_0]$					
		$f[x_0, x_0]$				
x_0	$f[x_0]$		$f[x_0, x_0, x_1]$			
		$f[x_0, x_1]$		$f[x_0, x_0, x_0, x_1]$		
x_1	$f[x_1]$		$f[x_0, x_1, x_1]$		$f[x_0, x_0, x_1, x_1, x_2]$	
		$f[x_1, x_1]$		$f[x_0, x_1, x_1, x_2]$		$f[x_0, x_0, x_1, x_1, x_2, x_2]$
x_1	$f[x_1]$		$f[x_1, x_1, x_2]$		$f[x_0, x_1, x_1, x_2, x_2]$	
		$f[x_1, x_2]$		$f[x_1, x_1, x_2, x_2]$		
x_2	$f[x_2]$		$f[x_1, x_2, x_2]$			
		$f[x_2, x_2]$				
x_2	$f[x_2]$					

Admetem, sense demostració, que el polinomi interpolador és:

$p(x) = f[x_0] + f[x_0, x_0](x - x_0) + f[x_0, x_0, x_1](x - x_0)^2 + f[x_0, x_0, x_1, x_1](x - x_0)^2(x - x_1) +$

$+ \ldots + f[x_0, x_0, \ldots, x_n, x_n](x - x_0)^2 \cdot \ldots \cdot (x - x_{n-1})^2(x - x_n)$

Prenem el *costat superior* del triangle per construir el polinomi interpolador

Exemple 18. Considerem la següent taula de valors:

x	f	f'
1	2	-1
-1	-2	-1

Per trobar el polinomi interpolador, és suficient construir l'esquema triangular detallat anteriorment.

i	x	f			
0	-1	-2			
			$f[x_0, x_0] = f'(x_0) = -1$		
0	-1	-2		$f[x_0, x_0, x_1] = \frac{3}{2}$	
			$f[x_0, x_1] = 2$		$f[x_0, x_0, x_1, x_1] = -\frac{3}{2}$
1	1	2		$f[x_0, x_1, x_1] = -\frac{3}{2}$	
			$f[x_1, x_1] = f'(x_1) = -1$		
1	1	2			

S'ometen els càlculs de diferències dividides per manca d'espai.

El polinomi interpolador és:

$p_3(x) = -2 + (-1)(x - (-1)) + \frac{3}{2}(x+1)^2 + (\frac{-3}{2})(x+1)^2(x-1) = -\frac{3}{2}x^3 + \frac{7}{2}x.$

En particular, s'observa que només té termes de grau senar.

Teorema 8. (error en la interpolació d'Hermite)

Sigui $f \in \mathcal{C}^{2n+2}([a,b])$ i $p_{2n+1}(x)$ el polinomi interpolador d'Hermite de f en els punts $(x_k, f_k) \quad k = 0 \div n$ ($x_i \neq x_j$ si $i \neq j$) i $x_k \in [a,b]$. Llavors per a tot $x \in [a,b]$:

$$f(x) - p_{2n+1}(x) = \frac{f^{(2n+2)}(\xi(x))}{(2n+2)!}\omega_n^2(x)$$

on $\xi(x) \in < x_0, \ldots, x_n >$ i $\omega_n(x) = \prod_{i=0}^{n}(x - x_i)$

Demostració. Fixem $\bar{x} \in [a,b]$. Si $\exists i \quad x_i = \bar{x}$ ja hem acabat.

Ara, per a $\bar{x} \neq x_i$ considerem la funció:

$$F(x) = f(x) - p_{2n+1}(x) - kw_n^2(x)$$

Volem determinar k tal que $F(\bar{x}) = 0$. Tenim que $F \in \mathcal{C}^{2n+1}([a,b])$. Aleshores $F(x)$ s'anul·la en $x_0, \ldots, x_n$, ja que:

$$F(x_i) = (f(x_i) - p_{2n+1}(x_i)) - k\omega_n^2(x_i) = 0 - 0 = 0$$

Fem la derivada, $F'(x) = f'(x) - p'_{2n+1}(x_i) - k2\omega_n(x)\omega'_n(x)$. Notem que F' s'anul·la en $x_0, \ldots, x_n$:

$$F'(x_i) = (f'(x_i) - p'_{2n+1}(x_i)) - k\omega_n^2(x_i)\omega'_n(x_i) = 0 - 0 = 0$$

I també, $F(\bar{x}) = 0$. Per tant, F té $(2n+1)+1 = 2n+3$ zeros.

$F(x)$ s'anul·la en $x_0, \ldots, x_n, \bar{x}$, per tant podem aplicar el teorema de Rolle i tenim y_i $i = 1 \div n+1$ amb $F'(y_i) = 0$, diferents de tots els punts d'interpolació (per Rolle). És a dir, que F' s'anul·la en $x_0, \ldots, x_n, y_1, \ldots, y_{n+1}$ en almenys $2n+2$ punts. Argumentant com a la demostració del **Lema 2** obtenim que $F^{2n+2}(x)$ s'anul·la en almenys un punt $\xi(\bar{x})$. La derivada d'ordre $2n+2$ de la funció F és:

$$F^{(2n+2)}(x) = f^{(2n+2)}(x) - 0 - k(2n+2)! \cdot 1$$

Fem notar que surt un 0 perquè el polinomi $p_{2n+1}(x)$ és de grau $2n+1$ i l'he derivat $2n+2$ vegades. Finalment, només resta aïllar la k substituint $\xi(\bar{x})$ a la funció:

$$0 = F^{(2n+2)}(\xi(\bar{x})) = f^{(2n+2)}(\xi(\bar{x})) - k(2n+2)! \Rightarrow k = \frac{f^{(2n+2)}(\xi(\bar{x}))}{(2n+2)!}$$

c.v.d.

3.3 Interpolació d'Hermite generalitzada

En aquesta secció no es donaran demostracions dels resultats que s'enunciïn.

Donades $x_0, \ldots, x_n$ abscisses diferents 2 a 2 , donats $m_0, \ldots, m_n$ valors enters ≥ 0 i $f \in \mathcal{C}^r$ amb $r \geq \max\limits_{k=0 \div n} \{m_k\}$ volem construir un polinomi $p(x)$ tal que $p^{(j)}(x_k) = f^{(j)}(x_k)$ $k = 0 \div n$ $j = 0 \div m_k$. És a dir:

$$\begin{aligned} j = 0 \quad p(x_k) &= f(x_k) \\ j = 1 \quad p'(x_k) &= f'(x_k) \\ j = 2 \quad p''(x_k) &= f''(x_k) \end{aligned}$$

amb j tantes condicions com dicti m_k. Posant un exemple una mica més clar, suposem punts x_0, x_1, x_2 amb respectius valors m_0, m_1, m_2:

$$\begin{aligned} &x_0 \longrightarrow m_0 = 1 \quad p(x_0) = f(x_0) \quad p'(x_0) = f'(x_0) \\ &x_1 \longrightarrow m_1 = 0 \quad p(x_1) = f(x_1) \\ &x_2 \longrightarrow m_2 = 2 \quad p(x_2) = f(x_2) \quad p'(x_2) = f'(x_2) \quad p''(x_2) = f''(x_2) \end{aligned}$$

Passem a enunciar el teorema d'existència i unicitat del polinomi d'Hermite generalitzat i el teorema que determina l'error en la interpolació.

Teorema 9. (d'existència i unicitat)

El problema d'Hermite generalitzat té solució única si el grau de $p(x) \leq N = \sum_{k=0}^{n} m_k + n$.

Nombre de condicions: $\sum_{k=0}^{n}(m_k + 1) = \sum_{k=0}^{n} m_k + n + 1$

Teorema 10. (error en la interpolació d'Hermite generalitzada) Siguin $x_0, \ldots, x_n$ amb $x_i \neq x_j$ si $i \neq j$ i $m_0, \ldots, m_n$ valors enters amb $m_i \geq 0$, $i = 0, \ldots, n$ i sigui $f \in \mathcal{C}^{N+1}$ en l'interval $< x, x_0, x_1, \ldots, x_n >$ i p_N el polinomi d'Hermite generelitzat. Llavors, per a qualsevol $x \in [a, b]$, es verifica:

$$f(x) - p_N(x) = \frac{f^{(N+1)}(\xi(x)}{(N+1)!}(x - x_0)^{m_0+1}(x - x_1)^{m_1+1} \cdots (x - x_n)^{m_n+1}$$

on $\xi(x)$ depèn de x i $\xi(x) \in< x, x_0, x_1, \ldots, x_n >$.

Mètodes de càlcul

1. Resolució del sistema

2. Mètode de ls diferències dividides generalitzades

   Recordem $f[x_i, x_i] = f'(x_i)$. Es vol calcular $f[x_i, x_i, x_i]$.

$$f[x_i, x_i, x_i] = \lim_{x \to x_i} f[x_i, x_i, x] = \lim_{x \to x_i} \frac{f[x_i, x] - f[x_i, x_i]}{x - x_i} =$$
$$\lim_{x \to x_i} \frac{\frac{f[x]-f[x_i]}{x-x_i} - f'(x_i)}{x - x_i} = \lim_{x \to x_i} \frac{f(x) - f(x_i) - f'(x_i)(x - x_i)}{(x - x_i)^2} =$$
$$\lim_{x \to x_i} \frac{f(x_i) + f'(x_i)(x - x_i) + \frac{f'(\xi)}{2!}(x - x_i)^2 - f(x_i) - f'(x_i)(x - x_i)}{(x - x_i)^2} =$$
$$\lim_{x \to x_i} \frac{f''(\xi)}{2} = \frac{f''(x_i)}{2}$$

   on $\xi \in< x, x_i >$ De manera generalitzada, $f[x_i, \ldots, x_i] = \dfrac{f^{(j)}(x_i)}{j!}$ (per a $j + 1$ arguments)

Es farà també un esquema triangular, però de la manera següent:

x	f				
x_0	f_0				
		$f'(x_0)$			
x_0	f_0		$\frac{f''(x_0)}{2}$		
		$f'(x_0)$		$\dots$	$\frac{f^{(m_0)}(x_0)}{m_0!}$
x_0	f_0		$\frac{f''(x_0)}{2}$		
$\vdots$		$f'(x_0)$			
x_0	f_0		$f[x_0, x_0, x_1]$		
		$f[x_0, x_1]$			
x_1	f_1				
		$f'(x_1)$			
x_1	f_1		$\frac{f''(x_1)}{2}$		
$\vdots$		$f'(x_1)$			
x_1	f_1				

Per a x_0 tenim $m_0 + 1$ condicions i per a x_1 en tenim $m_1 + 1$. El polinomi interpolador és:

$$
\begin{aligned}
p(x) &= f[x_0] + f[x_0, x_0](x - x_0) + \dots + f[x_0, \dots, x_0](x - x_0)^{m_0} + \\
&+ f[x_0, \dots, x_0, x_1](x - x_0)^{m_0+1} + f[x_0, \dots, x_0, x_1, x_1](x - x_0)^{m_0+1}(x - x_1) \\
&+ \dots + f[x_0, \dots, x_0, x_n, \dots, x_n](x - x_0)^{m_0+1} \cdot \dots \cdot (x - x_{n-1})^{m_{n-1}+1}(x - x_n)^{m_n}
\end{aligned}
$$

Exemple 19. Es considera la taula següent:

x	f	f'	f''
-1	2	0	0
1	-2	0	-

Tenim 5 condicions (3 per a x_0, 2 per a x_1). Fem l'esquema triangular:

x	f				
-1	-2				
		0			
-1	-2		0		
		0		$\frac{1}{2}$	
-1	-2		1		$\frac{-3}{4}$
		2		-1	
1	2		-1		
		0			
1	2				

El polinomi interpolador és $p(x) = -2+0(x-(-1))+0(x+1)^2+\frac{1}{2}(x+1)^3+(\frac{-3}{4})(x+1)^3(x-1)$

Observacions.

1. $x_0, \ldots, x_n$ amb $m_k = 0$ $k = 0 \div n$ és la interpolació de Lagrange.

2. $x_0, \ldots, x_n$ amb $m_k = 1$ $k = 0 \div n$ és la interpolació d'Hermite.

3. $n = 0$ amb $m > 0$ és la interpolació de Taylor.

Finalment, es vol fer notar que conegut $f(z)$ amb z pertanyent a una taula $x_0, \ldots x_n$, si augmetem n això no necessàriament millora l'aproximació (encara que la intuïció digui el contrari). Per veure-ho, considerem els següents exemples.

1. Sigui $f(x) = \begin{cases} x \sin \frac{\pi}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$. Prenem $x_k = \frac{1}{k+1}$ $\quad k = 0, 1, 2, \ldots$ S'observa de seguida que $f(x_k) = 0 \quad \forall k$. Això voldrà dir que el polinomi interpolador és $p_n(x) \equiv 0$.

2. **Fenomen de Runge (1901)** Es considera la següent funció $f(x) = \dfrac{1}{1 + 25x^2}$ a l'interval $[-1, 1]$. Es prenen abscisses $x_k^n = -1 + \frac{2k}{n}$ $k = 0 \div n$. És un resultat comprovat que el polinomi interpolador $p_n(x) \longrightarrow f(x)$ quan $n \to \infty$ si i només si $|x| < 0.726...$. Això és perquè les abscisses són equiespaiades, aleshores el polinomi ha d'oscil·lar per poder passar per tots els punts.

3.4 Derivació numèrica

Problema. Se sap avaluar $f(c)$, sense tenir una fórmula de f, i es vol calcular $f^{(j)}(c)$.

Idea. Bàsicament el que farem és:

1. Construir el polinomi interpolador de f en $x_0, \ldots, x_n$ (pròxims a c)
2. Agafarem com a aproximació $f^{(j)}(c) \simeq p_n^{(j)}(c)$ $(j \leq n)$.
3. Caldrà estudiar l'error de l'aproximació, que es dividirà en error de discretització i error d'arrodoniment. L'error total serà la suma d'aquests dos.

Exemple I

Siguin $x_0 = c$, $x_1 = c + h$ i $f \in \mathcal{C}^3([c, c+h])$ Tenim la taula següent:

x	f
c	$f(c)$
$c+h$	$f(c+h)$

Pas 1. La diferència dividida és $\dfrac{f(c+h) - f(c)}{c+h-c} = \dfrac{f(c+h) - f(c)}{h}$. El polinomi interpolador és $p_1(x) = f(c) + \dfrac{f(c+h) - f(c)}{h}(x-c)$

Pas 2. $f'(c) \simeq p_1'(c) = \dfrac{f(c+h) - f(c)}{h}$

Pas 3. Sabem (usant **Teorema 6**):

$$f(x) - p_1(x) = \frac{f^2(\xi(x))}{2!}(x-c)(x-c-h) \quad \xi(x) \in < c, c+h >$$

Derivem:

$$f'(x) - p_1'(x) = \left(\frac{f^2(\xi(x))}{2!}\right)'(x-c)(x-c+h) + \frac{f^2(\xi(x))}{2!}[x-c-h+x-c]$$

Si x és un punt d'interpolació $(x-c)(x-c-h)$ val 0. En aquest cas, $x = c$ és

$$f'(c) - p_1'(c) = \left(\frac{f^2(\xi(c))}{2!}\right)[c-c-h+c-c] = -\frac{f^2(\xi(c))}{2!}h$$

Veiem que depèn de h.

La fórmula de l'exemple I rep el nom de *fórmula descentrada* (per la manera com prenem els punts) A l'exemple següent veurem la *fórmula centrada.*

Exemple II

Sigui $n = 2$ amb $x_0 = c - h$ $x_1 = c$ i $x_2 = c + h$

Pas 1. El càlcul del polinomi d'interpolació és immediat:

$p_2(x) = f(c-h) + f[c-h, c](x-c+h) + f[c-h, c, c+h](x-c+h)(x-c)$

Pas 2. Ometem el càlcul de $p_2'(x)$. Tenim que $p_2'(c) = \dfrac{f(c+h) - f(c-h)}{2h}$

Pas 3. Error d'interpolació:

$$f(x) - p_2(x) = \frac{f^3(\eta_x)}{3!}(x-c+h)(x-c)(x-c-h) \quad \eta_x \in < c-h, c+h >$$

Diem $\omega_2(x) = (x-c+h)(x-c)(x-c-h)$. Derivem:

$$f'(x) - p_2'(x) = \left(\frac{f^3(\eta_x)}{3!}\right)'\omega_2(x) + \frac{f^3(\eta_x)}{3!}\omega_2'(x)$$

Per $x = c$:

$$f'(c) - p_2'(c) = \frac{f^3(\eta_c)}{3!}\omega_2'(c)$$

Observem $\omega_2'(c) = -h^2$. Finalment:

$$f'(c) - p_2'(c) = -\frac{f^3(\eta_c)}{3!}h^2$$

Observacions a l'exemple II.

1. h dividida per 2 implica que l'error es divideix per 4.

2. La fórmula centrada dóna una millor aproximació (abans l'error es dividia per 2)

3. A vegades no podem fer servir la fórmula centrada (per això cal haver parlat de **I**)

Ara, volem calcular $f''(c)$. Vegem els següents exemples:

Exemple III

Sigui $n = 2$ amb $x_0 = c$ $x_1 = c+h$ $x_2 = c+2h$. Trobem que:

$$f''(c) \simeq \frac{f(c) - 2f(c+h) + f(c+2h)}{h^2} + \text{error}$$

Exemple IV (fórmula centrada)

Sigui $n = 2$ amb $x_0 = c-h$, $x_1 = c$, $x_2 = c+h$. El polinomi interpolador és:

$$p_2(x) = f(c-h) + \frac{f(c) - f(c-h)}{h}(x-c+h) + \frac{f(c+h) - 2f(c) + f(c-h)}{2h^2}(x-c+h)(x-c)$$

La segona derivada dona:

$$p_2''(c) = 2! \cdot \frac{f(c+h) - 2f(c) + f(c+h)}{2h^2}$$

Volem calcular l'error. Fem notar que el càlcul $f''(x) - p_2''(x)$ és llarg i pesat. Una altra manera de calcular l'error és fer Taylor. Per fer-ho, suposem $f \in \mathcal{C}^4$ prou derivable:

$$f(c+h) = f(c) + f'(c)h + \frac{f''(c)}{2!}h^2 + \frac{f^{(3)}(c)}{3!}h^3 + \frac{f^{(4)}(\alpha)}{4!}h^4 \quad \alpha \in (c, c+h)$$

$f(c)$

$$f(c-h) = f(c) - f'(c)h + \frac{f''(c)}{2!}h^2 - \frac{f^{(3)}(c)}{3!}h^3 + \frac{f^{(4)}(\beta)}{4!}h^4 \quad \beta \in (c-h, c)$$

Sumem $f(c+h) + f(c-h)$:

$$f(c+h) + f(c-h) = 2f(c) + 2\frac{f''(c)}{2!}h^2 + \left[\frac{f^{(4)}(\alpha)}{4!} + \frac{f^{(4)}(\beta)}{4!}\right]h^4$$

$$f(c+h) + f(c-h) - 2f(c) = f''(c)h^2 + \left[\frac{f^{(4)}(\alpha)}{4!} + \frac{f^{(4)}(\beta)}{4!}\right]h^4$$

Obtenim:

$$\frac{f(c+h) - 2f(c) + f(c+h)}{h^2} = f''(c) + \frac{1}{4!}\left[f^{(4)}(\alpha) + f^{(4)}(\beta)\right]h^2$$

Finalment, aplicarem un resultat conegut com el *Teorema del valor mig per a sumes*. L'enunciat és el següent:

Teorema 11. (Teorema del valor mig per a sumes)
Sigui $F \in \mathcal{C}(I)$ amb $x_i \in I$ i $a_i \in I$ amb $a_i \geq 0$. Llavors $\exists y \in I$ tal que $\displaystyle\sum_{i=1}^{m} a_i F(x_i) = F(y) \sum_{i=1}^{m} a_i$

La hipòtesi $f^{(4)}$ contínua és vertadera en el nostre cas. Per tant, per a la suma $f^{(4)}(\alpha) + f^{(4)}(\beta)$ $\exists \eta$ tal que $f^{(4)}(\alpha) + f^{(4)}(\beta) = f^{(4)}(\eta)[1+1]$. Obtenim:

$$F(h) = f''(c) + \frac{1}{4!}2f^{(4)}(\eta)h^2 = f''(c) + \frac{1}{12}f^{(4)}(\eta)h^2$$

una aproximació que depèn de h.
Sigui ara $f \in \mathcal{C}^8([c-h, c+h])$ Usant la fórmula centrada per a $f''(c)$, trobem:

$$F(h) = f''(c) + 2\frac{f^{(4)}(c)}{4!}h^2 + 2\frac{f^{(6)}(c)}{6!}h^4 + \frac{f^{(8)}(\alpha) + f^{(8)}(\beta)}{8!}h^6 \quad \alpha, \beta \in (c, c+h)$$

Aquest darrer exemple és per mostrar que $F(h) - f''(c) = a_1 h^2 + a_2 h^4$ amb a_1, a_2 constants. D'aquí deduim que com que la funció depèn de h, per a h molt propers a zero això comporta problemes (es produeix cancel·lació i propagació de l'error)
Vegem aquest fenomen amb un exemple.

Exemple 20. Sigui $f(x) = e^x$ amb $c = 1$ i $f'(1) = e = 2.718282$. Farem servir la fórmula centrada, que és:

$$f'(c) = \frac{f(c+h) - f(c-h)}{2h} - \frac{f^{(3)}(\xi)}{3!}h^2 \quad \xi \in (c-h, c+h)$$

Fem una taula per a diversos valors de h:

h	$F(h)$	error
0.2	2.736440	$1.81 \cdot 10^{-2}$
0.02	2.718475	$1.81 \cdot 10^{-4}$
0.002	2.718250	$1.81 \cdot 10^{-6}$
0.0002	2.720000	$1.81 \cdot 10^{-8}$

S'observa que si bé l'error va disminuint, perdem xifres. Per a $h = 10^{-6}$ es té $f'(1) = 3$. Per a trobar la derivada amb més exactitud tenim dues possibilitats:

1. Calcular el pas òptim

2. Aplicar l'extrapolació de Richardson

Recordem que mesurem l'error total com $E_{total} = E_{discr} + E_{arrod}$ (suma de l'error de discretització i de l'error d'arrodoniment). És pertinent notar que l'error de discretització tendeix a 0 amb h properes a 0 i en canvi l'error d'arrodoniment creix amb h properes a 0. Això es veurà clarament en el proper exemple.

Exemple 21. Considrem la mateixa funció de l'exemple anterior. Volem fitar l'error total. D'entrada, tenim:

$$|E_{\text{disc}}| = \left| \frac{f^{(3)}(\xi)}{6} h^2 \right| \quad \xi \in <1-h, 1+h>$$

Fitem $|f^{(3)}(x)| \leq M \ \forall x \in (1-0.2, 1+0.2)$. Aleshores per ser f creixent és $M = e^{1.2} \simeq 3.32017$. Per calcular l'error d'arrodoniment:

$$E_{\text{arrod}} = \varepsilon_a \left(\frac{f(c+h) - f(c-h)}{2h} \right) = \frac{\varepsilon_a(f(c+h)) - \varepsilon_a(f(c-h))}{2h} = \frac{2\varepsilon}{2h} = \frac{\varepsilon}{h}$$

on hem considerat $\varepsilon_a(f(x)) = \varepsilon$ (fita en l'avaluació de la funció). Aquí es veu que l'error d'arrodoniment creix quan $h \longrightarrow 0$. Així doncs, l'error total és:

$$|E_{\text{total}}| \leq \frac{M}{6}h^2 + \frac{\varepsilon}{h} = g(h)$$

$g(h)$ és una funció de h. Trobar el pas òptim és buscar el mínim d'aquesta funció. Derivant i imposant $g'(h) = 0$ s'obté que $h = \sqrt[3]{\frac{3\varepsilon}{h}}$. És un mínim perquè $g''(h) > 0 \quad \forall h$.
En el nostre cas, a tall d'exemple amb $\varepsilon = \frac{1}{2} \cdot 10^{-6}$ s'obté $h_{\text{òptim}} \simeq 0.0077$.

Extrapolació de Richardson

Referència: ELDEN, L. ; WITTMEYER-KOCH, L. *Numerical Analysis*

Sigui una funció $F(h)$ que podem avaluar per a valors $h \neq 0$ tot i que presenta dificultats per a $h \to 0$. Suposem que volem calcular $f'(c)$. Usarem la fórmula centrada:

$$F(h) = \frac{f(c+h) - f(c-h)}{2h}$$

S'ha vist en aquesta secció que $F(h) = f'(c) + a_1 h^2 + a_2 h^4 + \ldots$. Recordem que a_1, a_2 no són coneguts però si que sabem que no depenen del pas h. Ara, si avaluem f per $2h$:

$$F(2h) = f'(c) + a_1(2h)^2 + a_2(2h)^4 + \ldots$$

Negligim els termes amb grau superior a 2. Fent la següent resta obtenim:

$$F(2h) - F(h) = 3a_1 h^2$$

o de manera equivalent:

$$\frac{F(2h) - F(h)}{3} = a_1 h^2 + \ldots$$

Si tornem a l'expressió coneguda de $F(h)$ s'obté:

$$F(h) = f'(c) + \frac{F(2h) - F(h)}{3} + \ldots$$

Conclusió: Tenim una aproximació de $f'(c)$ amb un error de discretització proporcional a h^4 en comptes de h^2.

La generalització d'aquest procés es fa introduint les notacions següents: $F_1(h) = F(h)$ i $F_2(h) = F_2(h) = F_1(h) + \dfrac{F_1(h) - F_1(2h)}{3}$

De manera més general:

$$F(h) = Q + a_1 h^{p_1} + a_2 h^{p_2} + \ldots \qquad p_1 < p_2 < p_3 \ldots$$

, on Q és la quantitat que volem calcular ($f'(c)$ en el nostre cas).

Reiterarem el procés partint de $h = h_0$ i fent $\frac{h_0}{q}, \frac{h_0}{q^2}, \ldots$ amb $q > 1$. Per tant, amb $F_1(h) =$

$F(h)$ obtenim:

$$F_{k+1}(h) = F_k(h) + \frac{F_k(h) - F_k(2h)}{q^{p_k} - 1} \quad k = 1, 2, \ldots$$

llavors

$$F_n(h) = Q + a_n^{(n)} h^{p_n} + \ldots$$

Exemple 22. Seguim amb $f(x) = e^x$. Usant l'extrapolació de Richardson obtenim la taula següent:

h	$F(h)$	$F_2(h)$	$F_3(h)$
0.4	2.791352		
0.2	2.736440	2.718136	
0.1	2.722815	2.718273	2.718282
0.05	2.719414	2.718280	2.718280

Observació. S'ha escollit $q = 2$. $F_3(h) = F_2(h) + \dfrac{F_2(h) - F_2(2h)}{2^4 - 1}$

3.5 Integració numèrica

Problema. Es vol calcular $I(f) = \displaystyle\int_a^b f(x)dx$

Solució analítica: Trobar $F(x)$ tal que $F'(x) = f(x)$ i aplicar la regla de Barrow, $I(f) = F(b) - F(a)$. A vegades, però, passa el següent:

- No existeix $F(x)$ o és difícil de calcular.
- $F(x)$ és cara d'avaluar
- No coneixem $f(x)$, sinó només en un conjunt de punts

Solució numèrica:

1. Aproximar $f(x)$ per un polinomi $p(x)$.
2. Calcular $I(f) \simeq \displaystyle\int_a^b p(x)dx$ (un polinomi és fàcil d'integrar).
3. Estudiar l'error.

A continuació es veuran diverses fórmules interpolatòries. Començarem de forma general, després es veuran les fórmules de Newton-Cotes tancades, les fórmules de Newton-Cotes compostes i finalment el mètode de Romberg.

Fórmules interpolatòries

Nota. Suposarem sempre $[a, b]$ fitat.

Comencem prenent $(n+1)$ punts d'interpolació $a \leq x_0 \leq x_1 \leq \ldots \leq x_n \leq b$. Considerem $p(x)$ polinomi interpolador de grau $\leq n$ de $f(x)$ en $x_0, \ldots, x_n$ tal que $p(x_k) = f(x_k) = f_k$ per $k = 0 \div n$. Usant els polinomis de Lagrange, $\prod_{\substack{k=0 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}$ per $j = 0 \div n$. S'ha vist que $p(x)$ es pot escriure com $p(x) = \sum_{j=0}^{n} f_j L_j(x)$. Com que el nostre objectiu és integrar, tenim:

$$\int_a^b p(x)dx = \int_a^b \sum_{j=0}^{n} f_j L_j(x)dx = \sum_{j=0}^{n} f_j \int_a^b L_j(x)dx$$

Notació 3. $A_j = \int_a^b L_j(x)dx$, que s'anomenen pesos.

En resum, hem aproximat $\int_a^b f(x)dx \simeq \sum_{j=0}^{n} A_j f_j$. Fem notar que n és un valor variable i que els $\{x_i\}_{i=0 \div n}$ són variables, és dir que aquesta fórmula són moltes fórmules contigudes en una.

Càlcul dels pesos:

1. Resolent la integral $A_j = \int_a^b L_j(x)dx$ (això és el que es farà quan s'expliqui les diferents fórmules interpolatòries)

2. Imposant exactitud als polinomis, és a dir:

$$\begin{aligned}
f(x) = 1 \quad & \int_a^b 1dx = A_0 + A_1 + \ldots + A_n \\
f(x) = x \quad & \int_a^b xdx = A_0 x + A_1 x + \ldots + A_n x \\
\vdots \quad & \\
f(x) = x^n \quad & \int_a^b x^n dx = A_0 x_0^n + \ldots + A_n x_n^n
\end{aligned}$$

Això és un sistema amb $n+1$ A_i i $n+1$ equacions. Si veiem que el det $\neq 0$ es podran determinar de manera única. La matriu del sistema és:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_0 & x_1 & \dots & x_n \\ x_0^2 & x_1^2 & \dots & x_n^2 \\ \vdots & & \vdots & \\ x_0^n & x_1^n & \dots x_n^n & \end{pmatrix} \begin{pmatrix} A_0 \\ A_1 \\ \vdots \\ \vdots \\ A_n \end{pmatrix} = \begin{pmatrix} b-a \\ \frac{b^2-a^2}{2} \\ \vdots \\ \vdots \\ \frac{b^{n+1}-a^{n+1}}{n+1} \end{pmatrix}$$

És una matriu de Vandermonde. Com s'ha vist en la demostració d'un teorema previ, el determinant de Vandermonde és diferent de 0.

Error en les fórmules interpolatòries:

Si $f \in \mathcal{C}^{n+1}([a,b])$ i $p(x)$ és polinomi interpolador de f de grau $\leq n$ en $x_0, \dots, x_n$ llavors, com ja sabem: $f(x) - p(x) = \dfrac{f^{(n+1)}(\xi(x))}{(n+1)!}(x-x_0)\cdots(x-x_n)$. Ara, integrem:

$$\int_a^b f(x)dx - \int_a^b p(x) = \int_a^b f(x)dx - \sum_{j=0}^{n} A_j f_j = \int_a^b \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \prod_{j=0}^{n}(x-x_j)dx$$

1. Fórmules de Newton-Cotes (tancades)

Newton-Cotes (abreviat, NC) vol dir que interpolem en abscisses equiespaiades.

Tancades vol dir que a i b són punts d'interpolació.

Dividim l'interval $[a,b]$ en n subintervals de pas $h = \frac{b-a}{n}$. Obtenim $x_i = a + ih$ per $i = 0 \div n$ $(x_0 = a, x_n = b)$:

$$A_j = \int_a^b L_j(x)dx = \int_a^b \prod_{\substack{k=0 \\ k\neq j}}^{n} \frac{x-x_k}{x_j-x_k}dx = \int_0^n \left(\prod_{\substack{k=0 \\ k\neq j}}^{n} \frac{s-k}{j-k} \right) h \cdot ds = h \int_0^n \prod_{\substack{k=0 \\ k\neq j}}^{n} \frac{s-k}{j-k} ds$$

La integral que queda conforma els pesos de Newton-Cotes. Notem que s'ha fet un canvi de variable, que a continuació detallem:

$$x = a + hs \quad x - x_k = x_0 + sh - (x_0 + kh) = (s-k)h \qquad x = a \longrightarrow s = 0$$

$$dx = hds \quad x_j - x_k = (j-k)h \qquad x = b \longrightarrow s = n$$

1.1. Fórmula del trapezi

Consisteix en agafar 2 punts, per tant $n = 1$. Aleshores $h = \frac{b-a}{1} = b - a$. $x_0 = a$ i $x_1 = b$.

Calculem els pesos (usant el canvi de variable):

$$\alpha_0 = \int_0^1 \frac{s-1}{0-1}ds = \frac{1}{2}$$

$$\alpha_1 = \int_0^1 \frac{s-0}{1-0}ds = \frac{1}{2}$$

Aproximem:

$$\int_a^b f(x)dx \simeq A_0 f_0 + A_1 f_1 = h\alpha_0 + h\alpha_1 = (b-a)\left[\frac{1}{2}f(a) + \frac{1}{2}f(b)\right] = h\left[\frac{f(a)+f(b)}{2}\right]$$

Calculem l'error de la fórmula: Sabem que si $f \in \mathcal{C}^2([a,b])$ i $p(x)$ és el polinomi interpolador de f de grau ≤ 1 (en els punts a i b) l'error és:

$$f(x) - p(x) = \frac{f^{(2)}(\xi(x))}{2!}(x-a)(x-b)$$

Integrem:

$$\int_a^b f(x)dx - h\left[\frac{1}{2}f(a) + \frac{1}{2}f(b)\right] = \int_a^b \frac{f^{(2)}}{2!}(x-a)(x-b)dx =$$

Ara aplicarem un resultat conegut com *Teorema del valor mig per a integrals.*

Teorema 12. (Teorema del valor mitjà per a integrals) Sigui f contínua en $[a,b]$, g integrable en $[a,b]$ i que g no canvia de signe en $[a,b]$. Aleshores $\exists c$ amb $a < c < b$ tal que $\int_a^b f(x)g(x)dx = f(c)\int_a^b g(x)dx$

Reprenent el càlcul de l'error i usant aquest resultat, tenim que $\exists \eta \quad a < \eta < b$:

$$\frac{1}{2!}f^{(2)}(\eta)\int_a^b (x-a)(x-b)dx = \frac{1}{2!}f^{(2)}(\eta)\int_0^1 sh(s-1)hh \cdot hds = -\frac{f^{(2)}(\eta)}{12}h^3$$

Resumint, l'error de la fórmula del trapezi és:

$$\int_a^b f(x)dx - h\left[\frac{1}{2}f(a) + \frac{1}{2}f(b)\right] = -\frac{f^{(2)}(\eta)}{12}h^3$$

1.2. Fórmula de Simpson

Es pren $n = 2$. Els punts d'interpolació són $x_0 = a$, $x_1 = \frac{a+b}{2}$, $x_2 = b$. Per definició és $h = \frac{b-a}{2}$. Calculem els pesos:

$$\alpha_0 = \int_0^2 \frac{(s-1)(s-2)}{(0-1)(0-2)}ds = \frac{1}{3} = \alpha_2$$
$$\alpha_1 = \int_0^2 \frac{(s-0)(s-2)}{(1-0)(1-2)}ds = \frac{4}{3}$$

Aproximem i obtenim:

$$\int_a^b f(x)dx \simeq \frac{h}{3}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right]$$

Calculem l'error de la fórmula, procedint com abans:

$$f(x) - p(x) = \frac{f^{(2)}(\xi(x))}{2!}(x - x_0)(x - x_1)(x - x_2)$$
$$\int_a^b f(x)dx - \int_a^b p(x)dx = \int_a^b \frac{f^{(2)}(\xi(x))}{2!}(x - x_0)(x - x_1)(x - x_2)dx$$

Denotem $g(x) = (x - x_0)(x - x_1)(x - x_2)$ $g(x)$ canvia de signe, per tant no es pot aplicar el teorema del valor mitjà per a integrals. Observem, però, que la fórmula de Simpson és exacta per a polinomis de grau ≤ 3 (en particular ho és per a polinomis de grau 2). Vegem-ho:

Donada $f(x) = x^3$, calculem la integral $\displaystyle\int_a^b x^3 = \frac{b^4 - a^4}{4}$

Ara apliquem la Fórmula de Simpson:

$$S(x^3) = \frac{b-a}{6}\left[a^3 + 4\left(\frac{a+b}{2}\right) + b^3\right] = \frac{b^4 - a^4}{4}$$

Considerem el polinomi que verifica $p_3(x_0) = f(x_0)$, $p_3(x_1) = f(x_1)$, $p_3'(x_1) = f'(x_1)$, $p_3(x_2) = f(x_2)$. Sabem que p_3 existeix i és únic. Sabem $S(f) = S(p_3) = I(p_3)$ Calculem doncs l'error, que en aquest cas és l'error en la interpolació d'Hermite (per les condicions del polinomi)

$$f(x) - p(x) = \frac{f^{(4)}(\alpha)}{4!}(x - x_0)(x - x_1)^2(x - x_2)$$

Apliquem integrals:

$$\int_a^b f(x)dx - \int_a^b p_3(x)dx = \int_a^b \frac{f^{(4)}(\alpha)}{4!}(x - x_0)(x - x_1)^2(x - x_2)dx$$

Notem que ara, $g(x) = (x - x_0)(x - x_1)^2(x - x_2)$ no canvia de signe a $[a, b]$. Per tant podem aplicar el T.V.M per a integrals; aleshores $\exists \xi \in (a, b)$ tal que

$$\int_a^b f(x)dx - \int_a^b p_3(x)dx = \frac{f^{(4)}(\xi)}{4!}\int_a^b (x - x_0)(x - x_1)^2(x - x_2)dx$$

Apliquem el canvi de variable d'abans ($x = a + hs$):

$$\frac{f^{(4)}(\xi)}{4!}\int_0^2 sh(s-1)^2h^2(s-2)h \cdot h \cdot ds = -\frac{1}{90}h^5 f^{(4)}(\xi)$$

Observació. L'error depèn de la derivada 4a. És a dir, que s'anul·la per un polinomi de grau 3.

Les fórmules de Newton-Cotes d'ordre n són exactes:

- Si n és senar per a polinomis de grau $\leq n$

- Si n és parell per a polinomis de grau $\leq n+1$

Quins problemes generen l'ús d'aquestes fórmules? Si $[a,b]$ té longitud gran, com que es prenen pocs punts la fórmula no aproximarà bé. D'altra banda, molts punts poden no ser solució (com s'ha vist amb el fenomen de Runge) Per resoldre aquest problema, farem servir una aproximació amb subintervals que fa servir les fórmules de Newton-Cotes d'ordre petit.

2. Fórmules de Newton-Cotes compostes

Comencem dividint l'interval $[a,b]$ en m subintervals. A cadascun d'aquests aplicarem una fórmula d'ordre n.

Notació pel problema. Diem $l = \dfrac{b-a}{m}$ és la longitud d'un subinterval i el pas $h = \dfrac{l}{n} = \dfrac{b-a}{m\cdot n}$. Els extrems del subintervals els denotarem $x^0, x^1, \ldots, x^m$ o alternativament $x_0, x_n, x_{2n}, \ldots, x_{mn}$

Per calcular la integral aproximarem de la següent manera:

$$\int_a^b f(x)dx = \sum_{i=0}^{m-1} \int_{x^i=x_{in}}^{x^{i+1}=x_{(i+1)n}} f(x)dx \simeq \sum_{i=0}^{m-1}\sum_{k=0}^{n} A_i f_{k+in}$$

2.1. Fórmula dels trapezis

Prenem $n=1$ i $l=\frac{b-a}{m}$ i $h=\frac{l}{1}=l$. La fórmula és:

$$\int_a^b f(x)dx \simeq \sum_{i=0}^{m-1}\int_{x_i}^{x_{i+1}} f(x)dx \simeq \sum_{i=0}^{m-1}\left[h\left(\frac{f(x_i)}{2}+\frac{f(x_{i+1})}{2}\right)\right] =$$
$$h\left[\frac{f(x_0)}{2}+\frac{f(x_1)}{2}+\frac{f(x_1)}{2}+\frac{f(x_2)}{2}+\ldots+\frac{f(x_{m-1})}{2}+\frac{f(x_m)}{2}\right]$$

Finalment queda:

$$T(h) = h\left[\frac{f(x_0)}{2}+f(x_1)+\ldots+f(x_{m-1})+\frac{f(x_n)}{2}\right]$$

Notem $x_i = x^i$ (només en aquest cas).

Error en un interval:

$$\int_{x_i}^{x_{i+1}} f(x)dx = \frac{h}{2}[f(x_i)+f(x_{i+1})] = -\frac{h^3}{12}f^{(2)}(\eta_i) \quad \eta_i \in (x_i, x_{i+1})$$

Error total:

$$\int_a^b f(x)dx = \sum_{i=0}^{m-1}\int_{x_i}^{x_{i+1}} f(x)dx = \sum_{i=0}^{m-1}\left[\frac{h}{2}(f(x_i)+f(x_{i+1}) - \frac{h^3}{12}f^{(2)}(\eta_i)\right] =$$
$$= T(h) - \sum_{i=0}^{m-1}\frac{h^3}{12}f^{(2)}(\eta_i) = T(h) - \frac{b-a}{h}\cdot\frac{h^2}{12}\sum_{i=0}^{m-1} f^{(2)}(\eta_i) = T(h) - \frac{b-a}{12m}h^2 f^{(2)}(\beta)\left(\sum_{i=0}^{m-1} 1\right)$$

On s'ha aplicat a la darrera igualtat el T.V.M per a sumes. (**Teorema 11**) Així doncs:

$$\int_a^b f(x)dx = T(h) - \frac{b-a}{12}h^2 f^{(2)}(\beta)$$

2.2. Fórmula de Simpson composta

Prenem $n = 2$. Considerem, de forma general, l'interval $[x^i, x^{i+1}]$ i 3 punts, dos dels quals són els extrems, $x^i = x_{2i}$, $x^{i+1} = x_{2i+2}$ i x_{2i+1}, el punt mig dels altres dos. Tenim $h = \frac{l}{2} = \frac{b \cdot a}{2m}$. Necessitem f prou derivable, en particular $f \in \mathcal{C}^4([a,b])$. Per tant, la integral en aquest interval serà:

$$\int_{x_{2i}}^{x_{2i+2}} = \frac{h}{3}\big[f(x_{2i}) + 4f(x_{2i+1}) + f(x_{2i+2})\big] - \frac{h^5}{90}f^{(4)}(\eta_i)$$

En un interval $[a,b]$ tindrem:

$$\int_a^b f(x)dx = \sum_{i=0}^{m-1}\int_{x_{2i}}^{x_{2i+2}} f(x)dx = \frac{h}{3}[f_{2i} + 4f_{2i+1} + f_{2i+2}] - \frac{h^5}{90}f^{(4)}(\eta_i)$$

Notem $f_j = f(x_j)$. Diem $S(h)$ a la fórmula de Simpson sense tenir en compte l'error:

$$S(h) = \frac{h}{3}\big[f_0 + 4f_1 + f_2 + f_2 + 4f_3 + f_4 + \ldots + f_{2m-2} + 4f_{2m-1} + f_{2m}\big]$$

Aleshores,

$$\int_a^b f(x)dx = S(h) - \sum_{i=0}^{m-1}\frac{h^5}{90}f^{(4)}(\eta_i) = S(h) - \frac{b-a}{2m}\frac{h^4}{90}\sum_{i=0}^{m-1} f^{(4)}(\eta_i)$$

I ara apliquem el teorema del valor mig per a sumes: La funció en qüestió és continua a l'interval $[a,b]$ per tant $\exists \eta \in (a,b)$

$$\int_a^b f(x)dx = S(h) - \frac{b-a}{180}h^4\frac{1}{m}f^{(4)}(\eta)m = S(h) - \frac{b-a}{180}h^4 f^{(4)}(\eta)$$

3. Mètode de Romberg

Referència: Elden, L. ; Wittmeyer-Koch, L. *Numerical Analysis*

Recordatori. Direm $T(h)$ a la fórmula dels trapezis, on $T(h) = h[\frac{f(x_0)}{2} + f(x_1) + \ldots + f(x_{n-1}) + \frac{f(x_n)}{2}]$

S'ha vist que l'error en la fórmula del trapezis és $O(h^2)$ amb f'' contínua. Però existeix un resultat més potent, que s'anomena la fórmula d'Euler-Maclaurin.

Teorema 13. (Fórmula d'Euler-Maclaurin) Si $f \in \mathcal{C}^{2r+2}([a,b])$ llavors

$$T(h) = \int_a^b f(x)dx + a_1h^2 + a_2h^4 + \ldots + a_rh^r + O(h^{2r+2})$$

on els coeficients a_i no depenen d'h.

Observem que aquest resultat ens diu que podem aproximar $T(h)$ per a diferents valors de h i usar l'extrapolació de Richardson (vegeu secció 3.4) prenent $q = 2$ per reduir l'error. Aquest procediment s'anomena el mètode de Romberg.
Per tant, procedim de la següent manera:

$$
\begin{aligned}
&T_1(h) = T(h) \\
&T_2(h) = T_1(h) + \frac{T_1(h) + T_1(2h)}{2^2 - 1} = \frac{4T_1(h) - T_1(2h)}{3} \qquad \text{1r pas d'extrapolació} \\
&T_3(h) = T_2(h) + \frac{T_2(h) - T_2(2h)}{2^4 - 1} \\
&\dots
\end{aligned}
$$

En el primer pas d'extrapolació el terme h^2 en l'error de truncament s'elimina, i obtenim aproximacions amb error de truncament $O(h^4)$.

Exercici 3. Demostrar que un pas d'extrapolació de trapezis correspon a la fórmula de Simpson composta.

Exemple 23. Calcular $\displaystyle\int_1^{2.2} \ln x dx$. Usarem 7 dígits
La primitiva de $\ln x$ es pot calcular integrant per parts:

$$\int_1^{2.2} \ln x dx = x(\ln x - 1)\Big|_1^{2.2} = 0.5346061928$$

La calcularem, però usant el mètode de Romberg. Com que apliquem la fórmula dels trapezis, a $T_1(h)$ $h = \frac{2.2-1}{2} = 0.6$. S'ometen alguns càlculs, però simplement es tracta de fer els 2 passos d'extrapolació aplicant les fórmules donades

$$
\begin{aligned}
T(0.6) &= 0.6\left(\frac{\ln 1}{2} + \ln 1.6 + \frac{\ln 2.2}{2}\right) \\
T(0.3) &= \frac{1}{2}\big[T(0.6) + 0.6[\ln 1.3 + \ln 1.9]\big]
\end{aligned}
$$

h	$T(h)$ $(O(h^2))$	$T_2(h) = S(h)$ $(O(h^4))$	$T_3(h)$ $(O(h^6))$
0.6	0.5185394		
0.3	0.5305351	0.5345337	
0.15	0.5335847	0.5346013	0.5346058

4 Zeros de funcions

Problema. Se sap avaluar $f : \mathbb{R} \longrightarrow \mathbb{R}$. Es vol buscar $z \in \mathbb{R}$ tal que $f(z) = 0$, anomenats zeros o arrels de la funció.

Idea. Bàsicament, el procés de càlcul d'arrels consta de tres etapes:

- Localització: Es busquen regions que continguin els zeros
- Separació: Es vol una regió que només contingui un zero
- Aproximació numèrica: Sempre s'usen mètodes iteratius, que generen sucessions $(x_n)_{n\geq 0}$ que convergeixen al zero z.

Un cas típic és el càlcul dels màxims i mínims d'una funció $f : \mathbb{R} \longrightarrow \mathbb{R}$ derivable. Per exemple, si tenim la funció $p(x) = 5x^6 + 6x^2 - 18x + 5$, calculem la derivada $p'(x) = 30x^5 + 12x - 18$ i la igualem a 0, $p'(x) = 0$.

Observació. Existeixen fórmules per resoldre equacions de quart grau. Però per a grau superior, només podem aproximar la solució mitjançant mètodes numèrics.

Exemple 24. Es considera $f(x) = x^2 + 10\cos x$. Notem que trobar els zeros és equivalent a trobar els punts de tall entre $g(x) = x^2$ i $h(x) = -10\cos x$

x	$g(x)$	$h(x)$	$f(x)$
$\frac{\pi}{4}$	0.61685	-7.07	7.68
$\frac{\pi}{2}$	2.4674	0	2.46
$\frac{3\pi}{4}$	5.5516	7.071	-1.5

Aplicant el teorema de Bolzano veiem $\exists c_1 \in [\frac{\pi}{2}, \frac{3\pi}{4}]$. Si seguissim fent la taula, observaríem que $\exists c_2 \in [\pi, \frac{5\pi}{4}]$. Argumentant, veiem que no hi ha cap altre zero ja que g és creixent.

Aquest exemple ha consistit en els primers passos del procés de càlcul dels zeros: S'han localitzat i s'han separat. Ara, tocaria aproximar-los numèricament. Com s'ha dit, els mètodes generen una successió $(x_n)_{n\geq 0}$ que $x_n \longrightarrow 0$ quan $n \to \infty$. La idea és que donats uns punts $x_n, x_{n-1}, \ldots, x_{n-k}$ el càlcul de $x_{n+1} = G(x_n, x_{n-1}, \ldots, x_{n-k})$ depèn dels anteriors, on G és una funció concreta. Finalment, tallarem el procés quan $|x_{n+1} - x_n| < \text{tol}_1$ o $|f(x_n)| < \text{tol}_2$.

4.1 Mètodes iteratius

S'estudiaran els següents mètodes iteratius: mètode de bisecció, mètode de Newton-Raphson i mètode de la secant.

Mètode de bisecció

Recordem el *Teorema de Bolzano*: Donada f contínua en $[a, b]$, si $f(a)f(b) < 0$ aleshores existeix $c \in (a, b)$ amb $f(c) = 0$.

El mètode de bisecció es basa en el teorema de Bolzano. La idea és la següent: Es construeix una successió d'intervals $[a, b] = [a_0, b_0] \supset [a_1, b_1] \supset \ldots \supset [a_n, b_n] \supset \ldots$ sempre amb $f(a_i)f(b_i) < 0$

Comencem considerant $[a, b] = [a_0, b_0]$ amb $f(a_0)f(b_0) < 0$. Calculem el punt mig $c_0 = \dfrac{a_0 + b_0}{2}$.

Ara, l'interval $[a_1, b_1]$ ve definit de la següent manera:

$$[a_1, b_1] = \begin{cases} [a_0, c_0] & \text{si } f(a_0)f(c_0) < 0 \\ [c_0, b_0] & \text{si } f(c_0)f(b_0) < 0 \end{cases}$$

Per a cada i, farem $c_i = \frac{a_i+b_i}{2}$. Si $f(c_i) = 0$, aleshores c_i és el zero. En general:

$$[a_{i+1}, b_{i+1}] = \begin{cases} [a_i, c_i] & \text{si } f(a_i)f(c_i) < 0 \\ [c_i, b_i] & \text{si no} \end{cases}$$

La longitud de l'interval la trobem fent $[a_{i+1}, b_{i+1}] = \dfrac{b_0 - a_0}{2^{i+1}}$

Si volem aturar el procés donada una certa tolerància ε, farem:

$$\frac{b-a}{2^{i+1}} < \varepsilon \Longrightarrow i \geq \frac{\log\left(\frac{b-a}{\varepsilon}\right)}{\log 2} - 1$$

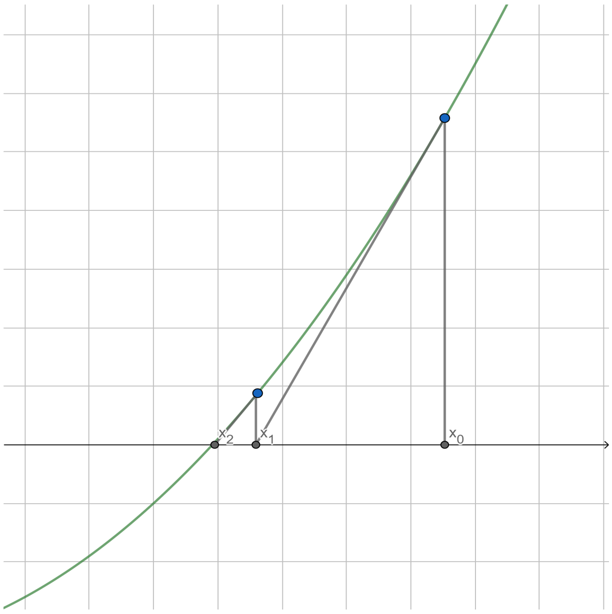
Exemple 25. Tenim l'interval $[0.5, 0.6]$, $f(x) = x - e^{-x}$ i $\varepsilon = 10^{-6}$. Aleshores fem $\frac{0.1}{2^{i+1}} \leq 10^{-6}$. Obtenim $i > 15.6$. Per tant caldran 16 iteracions.

Exemple 26. Seguim amb la funció $f(x) = x^2 + 10\cos x$. Hem vist que a $[\frac{\pi}{2}, \frac{3\pi}{4}] = [1.57, 2.36]$ la funció té una arrel. Diem $[a_0, b_0] = [1.57, 2.36]$. A continuació es detallen algunes iteracions del mètode de bisecció:

$$\begin{array}{lll} [a_0, b_0] = [1.57, 2.36] & c_1 = 1.965 & f(c_0) = 2 \cdot 10^{-2} > 0 \\ [a_1, b_1] = [1.965, 2.36] & c_1 = 2.1625 & f(c_1) = -9 \cdot 10^{-1} < 0 \\ [a_2, b_2] = [1.965, 2.1625] & c_2 = 2.06375 & f(c_2) = -2 \cdot 10^{-1} < 0 \\ [a_3, b_3] = [1.965, 2.06375] & ... & \end{array}$$

Mètode de Newton-Raphson (o de la tangent)

Donada $y = f(x)$ i x_0, es traça la recta tangent a la corba en el punt $(x_0, f(x_0))$. Es pren x_1 com la intersecció de la recta tangent amb l'eix d'abscisses (veure figura).



L'equació de la tangent és:

$$y - f(x_0) = f'(x_0)(x - x_0)$$

La intersecció amb l'eix d'abscisses consisteix en prendre $y = 0$ i solucionar per x:

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

La fórmula general és:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad n = 0, 1, 2 \ldots$$

El problema del mètode de Newton-Raphson és que no està assegurada la convergència. Però si convergeix, ho farà rapidament, tal com es pot observar a la figura. Un segon problema és quan per algun x_n, $f'(x_n) = 0$. Geomètricament, es tracta d'una tangent horitzontal. Analíticament, és que no podem dividir per 0. Per tant, fixarem $|f'(x_n)| < \text{tol}$.

Exemple 27. Es vol trobar una arrel $f(x) = x^2 + 10 \cos x$ amb $x_0 = 1.57$ mitjançant el mètode de Newton-Raphson. Construïm la taula tot aplicant el mètode:

$$
\begin{array}{ll}
x_0 = 1.57 & f(x_0) = 2.47286 \\
x_1 = 1.93047586 & f(x_1) = 2.07 \cdot 10^{-1} \\
x_2 = 1.9681170087 & f(x_2) = 4 \cdot 10^{-3} \\
x_3 = 1.96887262037 & f(x_3) = 1.6 \cdot 10^{-6} \\
x_4 = 1.9688729378 &
\end{array}
$$

S'observa la ràpida convergència del mètode. A la tercera iteració ja s'està molt proper al zero.

Mètode de la secant

Donats x_n, x_{n-1} es traça la recta entre els punts $(x_n, f(x_n))$ i $(x_{n-1}, f(x_{n-1}))$. Hom pot saber que l'equació d'aquesta recta és:

$$y = f(x_n) + \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}(x - x_n)$$

Ara, si tallem la recta amb l'eix de les x imposant $y = 0$ i resolent per x, obtenim:

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \qquad n = 1, 2, 3 \ldots$$

que és la fórmula general del mètode de la secant.

Comentaris.

- No convergeix sempre.
- En general, Newton-Raphson convergeix més ràpid que secant.
- Si calcular $f'(x)$ comporta moltes operacions o directament no es pot calcular, aplicarem secant.

Exemple 28. Es vol trobar una arrel $f(x) = x^2 + 10\cos x$ amb $x_0 = 1.57$ i $x_1 = 2.36$ mitjançant el mètode de la secant. Construïm la taula tot aplicant el mètode:

$$
\begin{array}{r|l}
x_2 = 2.058245 & f(x_2) = -4.5 \cdot 10^{-1} \\
x_3 = 1.933363 & f(x_3) = 1.9 \cdot 10^{-1} \\
x_4 = 1.970747 & f(x_4) = -9.8 \cdot 10^{-3} \\
x_5 = 1.968909 & f(x_5) = -1.9 \cdot 10^{-4}
\end{array}
$$

Fem notar que perquè convergeixi calen 7 iteracions. El zero és $\alpha = 1.968873$.

Mètode de regula falsi

Aquest mètode és una variació de bisecció i secant i té poca rellevància.

Donat un interval $[x_0, x_1]$ amb $f(x_0)f(x_1) < 0$ es considera $[a_1, b_1] = [x_0, x_1]$ i es calcula x_2 usant el mètode de la secant. Aleshores definim l'interval $[a_2, b_2]$:

$$[a_2, b_2] = \begin{cases} [a_1, x_2] & \text{si } f(a_1)f(x_2) < 0 \\ [x_2, b_1] & \text{si no} \end{cases}$$

i es reitera el procés. Aquest mètode convergeix sempre però ho fa lentament.

4.2 Teoria general de la iteració simple

Seguim amb el problema de buscar z tal que $f(z) = 0$. Ara, reescriurem l'equació $f(x) = 0$ com $x = g(x)$. Per tant, busquem z tal que $z = g(z)$, que s'anomena **punt fix**.

La idea és ben senzilla: Partint de x_0 es construeix un mètode iteratiu amb $x_1 = g(x_0)$, $x_{n+1} = g(x_n)$ $n \geq 0$. Aquest mètode iteratiu s'anomena *procés d'iteració simple*. La convergència depèn de g i de x_0. De fet, n'hi ha prou observant que si g és contínua i $x_n \longrightarrow z$ llavors:

$$z = \lim_{n\to\infty} x_n = \lim_{n\to\infty} g(x_{n-1}) = g(\lim_{n\to\infty} x_{n-1}) = g(z)$$

Abans d'enunciar els resultats teòrics, vegem alguns exemples de canvis (reescriptura de $f(x) = 0$).

Exemple 29. Sigui $f(x) = x^2 + 10\cos x = 0$. Tres possibles opcions de canvis són:

(a) $8x = 8x + x^2 + 10\cos x \Longrightarrow x = \dfrac{8x + x^2 + 10\cos x}{8} = g_1(x)$

(b) $3x^2 = 3x^2 - (x^2 + 10\cos x) = 2x^2 - 10\cos x \Longrightarrow x = \dfrac{2x^2 - 10\cos x}{3x} = g_2(x)$

(c) $x^2 + 10\cos x = 0 \Longrightarrow \cos x = \dfrac{-x^2}{10} \Longrightarrow x = \arccos\left(\dfrac{-x^2}{10}\right) = g_3(x)$

Teorema 14. (d'existència i unicitat de punts fixos)

(a) Sigui $g : [a, b] \to \mathbb{R}$ una funció contínua en l'interval $[a, b]$ tal que $g(x) \in [a, b]$ $\forall x \in [a, b]$. Llavors la funció g té un punt fix en $[a, b]$

(b) Si, a més, g és derivable en l'interval (a, b) i $|g'(x)| \leq k < 1$ $\forall x \in (a, b)$ llavors en $[a, b]$ la funció té un únic punt fix.

Demostració.

(a) Si $g(a) = a$ o $g(b) = b$ aleshores existeix un punt fix i ja estem.

Suposem $g(a) \neq a$ i $g(b) \neq b$. Aleshores $g(a) > a$ i $g(b) < b$. Definim $h(x) = x - g(x)$ que és clarament contínua. Notem

$$h(a) = a - g(a) < 0$$

$$h(b) = b - g(b) > 0$$

Aleshores pel teorema de Bolzano $\exists c \in (a, b)$ tal que $h(c) = 0$. Observem que $h(c) = 0$ és $c - g(c) = 0$, és a dir $c = g(c)$; per tant c és un punt fix de g.

(b) Suposem que hi ha 2 punts fixos $p \neq q$:

$$|p - q| = |g(p) - g(q)| = |g'(r)||p - q| \leq k|p - q| < |p - q| \quad !!$$

Tenim una contradicció. A la segona igualtat (en virtut de la primera) hem aplicat el teorema del valor mitjà: $\exists r \in (p, q) \subset [a, b]$. Per tant el punt fix és únic c.v.d.

Exemple 30.

- $g(x) = x^2$ té un punt fix a $[0.5, 1.2]$ Però observem que $g(0.5) = 0.25 \notin [0.5, 1.2]$ i que $g(1.2) = 1.44 \notin [0.5, 1.2]$. Per tant, no es verifica el teorema
- $g(x) = 3^{-x}$ $g(x) \in [0, 1]$ $\forall x \in [0, 1]$. Observem que verifica (a) del teorema. Però $|g'(x)| \not\leq 1$

En resum, les condicions del teorema són suficients però podrien ser no necessàries.

Teorema 15. (determinació de punts fixos)
Suposem $g \in \mathcal{C}([a, b])$ i $g(x) \in [a, b]$ $\forall x \in [a, b]$ i que g' existeix en (a, b) i verifica $|g'(x)| \leq k < 1$ $\forall x \in (a, b)$. Agafem un punt qualsevol $p_0 \in [a, b]$ i definim la successió $p_{n+1} = g(p_n)$ $n \geq 0$, llavors $(p_n)_{n \geq 0}$ convergeix a l'únic punt fix de g.

Demostració. Per les hipòtesis del teorema anterior, sabem que $\exists!$ punt fix p de g en $[a, b]$. Suposem $p_0 \in [a, b]$. La successió $p_{n+1} = g(p_n) \in [a, b]$ està ben definida. Aleshores, usant $|g'(x)| < k$ pel teorema del valor mitjà existeix un r tal que:

$$|p_n - p| = |g(p_{n-1}) - g(p)| = |g'(r)||p_{n-1} - p| \leq k|p_{n-1} - p|$$

Si apliquem aquesta desigualtat de manera inductiva obtenim:

$$|p_n - p| \leq k|p_{n-1} - p| \leq k^2|p_{n-2} - p| \leq \ldots \leq k^n|p_0 - p|$$

Com que $k < 1$:

$$\lim_{n\to\infty} |p_n - p| \leq \lim_{n\to\infty} k^n |p_0 - p| = 0$$

És a dir, $p_n \longrightarrow p$ c.v.d.

Corol·lari 1. Sota les hipòtesis del teorema anterior:

$$|p_n - p| \leq k^n \max\{p_0 - a, b - p_0\}$$

Corol·lari 2. Sota les hipòtesis del teorema anterior:

$$|p_n - p| \leq \frac{k^n}{1-k}|p_1 - p_0| \quad n \geq 1$$

Demostració.

Comencem aplicant el procediment usat en la demostració del teorema anterior:

$$|p_{n+1} - p_n| = |g(p_n) - g(p_{n-1})| = |g'(\xi)||p_n - p_{n-1}| \leq k|p_n - p_{n-1}| \leq \ldots \leq k^n|p_1 - p_0|$$

Fixem n i prenem $m > n \geq 1$:

$$\begin{aligned}
|p_m - p_n| &= |p_m - p_{m-1} + p_{m-1} - p_{m-2} + \ldots + p_n| \\
&\leq |p_m - p_{m-1}| + |p_{m-1} - p_{m-2}| + \ldots + |p_{n+1} - p_n| \\
&\leq k^{m-1}|p_1 - p_0| + k^{m-2}|p_1 - p_0| + \ldots + k^n|p_1 - p_0| \\
&= k^n(k^{m-n-1} + \ldots + k + 1)|p_1 - p_0| \leq
\end{aligned}$$

Fem notar que l'expressió entre parèntesis és una progressió geomètrica de raó $k < 1$, que convergeix a $\frac{1}{1-k}$

$$\leq k^n \sum_{i=0}^{\infty} k^i |p_1 - p_0| = \frac{k^n}{1-k}|p_1 - p_0|$$

Recapitulant, hem obtingut:

$$|p_m - p_n| \leq \frac{k^n}{1-k}|p_1 - p_0|$$

Fent $m \longrightarrow \infty$ queda $p_m \longrightarrow p$, i per tant:

$$|p - p_n| \leq \frac{k^n}{1-k}|p_1 - p_0|$$

c.v.d.

Observació. Aquesta desigualtat ens dona la velocitat en què $(p_n)_{n\geq 0}$ convergeix a la cota k a la primera derivada. La velocitat de convergència depèn del factor k^n. Com més petit sigui el valor de k, més ràpida serà la convergència, que pot ser molt petita si k és molt proper a 1.

Exemple 31. Per a la funció $f(x) = x^2 + 10\cos x$ amb valors $x = 1,96$ i $x = 3.16$ s'han vist algunes possibles funcions d'iteració, de les quals ara calcularem la derivada.

$$g_1(x) = \frac{8x + x^2 + 10\cos x}{8} \Longrightarrow g_1'(x) = \frac{1}{8}(8 + 2x - 10\sin x)$$

Ara, substituim els valors de les arrels a la derivada. Obtenim que $g_1'(1.96) \simeq 0.3335$ és a dir $|g_1'(1.96)| < 1$ convergirà a prop de 1.96. D'altra banda, veiem $g_1'(3.16) \simeq 1.813$, per tant en aquest cas no convergirà.

4.3 Ordre de convergència

Definició 12. Una successió $(x_n)_{n\geq 0}$ convergent a α direm que té ordre de convergència almenys p si $\exists N \geq 0$ i $c \geq 0$ tals que $\forall n \geq N$ $|x_{n+1} - \alpha| \leq c \cdot |x_{n+1} - \alpha|^p$

Un mètode iteratiu de k ($k \geq 1$) punts $x_{n+1} = G(x_n, \ldots, x_{n-k+1})$ $n \geq k-1$ té ordre de convergència almenys p si la successió $(x_n)_{n\geq 0}$ té ordre de convergència almenys p. Fem notar que el mètode iteratiu és per trobar el zero d'una funció f. En particular, anomenem l'error $e_n = x_n - \alpha$ si existeix

$$\lim_{n\to\infty} \frac{|e_{n+1}|}{|e_n|^p} = c \neq 0$$

aleshores diem que el mètode té ordre de convergència p. Anomenem c el coeficient asimptòtica de l'error (o constant asimptòtica).

Observació. Per $p = 1$ es parla de convergència lineal i per $p = 2$ es parla de convergència quadràtica. Fem notar que p no és necessàriament enter, i que quan $1 < p < 2$ parlarem de convergència superlineal.

Observació. En el límit anterior, per n gran, prenem $e_{n+1} \simeq c \cdot e_n^p$

Exemple 32. Partim de $|e_{n+1}| \leq M|e_n|^p$ amb $M = 0.7$ i $|e_n| = 1$ i es vol veure l'evolució de l'error en funció de l'ordre. Obtenim la següent taula:

i	ordre 1	ordre 2
1	0.7	0.7
2	0.49	0.343
3	0.343	$8.2 \cdot 10^{-2}$
4	0.2401	$4.75 \cdot 10^{-3}$
5	0.16807	$1.58 \cdot 10^{-5}$
6	0.11765	$1.74 \cdot 10^{-10}$

Tal com s'observa, si tenim convergència quadràtica, en pocs iterats l'error ja es fa molt petit. Ara, per caractertizar l'ordre de convergència tenim la següent proposició.

Proposició 4. (Ordre de convergència dels mètodes d'iteració simple)
Suposem $g \in \mathcal{C}^p([a, b])$ i sigui $\alpha \in [a, b]$ un punt fix de g i $x_{n+1} = g(x_n)$ $n \geq 0$ que convergeix a α. Aleshores $x_{n+1} = g(x_n)$ té ordre $p \in \mathbb{N}$ si i només si $g^{(i)}(\alpha) = 0$ per $i = 1 \div p - 1$ i $g^{(p)}(\alpha) \neq 0$

Demostració. Cal veure les dues implicacions.
$\Leftarrow$) Partim de $g^{(i)}(\alpha) = 0$ $i = 1 \div p - 1$ i $g^{(p)}(\alpha) \neq 0$. Usant desenvolupament de Taylor:

$$g(x) = g(\alpha) + g'(\alpha)(x - \alpha) + \ldots + \frac{g^{(p-1)}(\alpha)}{(p-1)!}(x - \alpha)^{p-1} + \frac{g^p(\xi)}{p!}(x - \alpha)^p \quad \xi \in < x, \alpha >$$

Observem que els primers $p - 1$ termes són 0 per hipòtesi. Així doncs, tenim

$$x_{n+1} = g(x_n) = g(\alpha) + \frac{g^{(p)}(\xi_n)}{p!} \quad \xi_n \in < x, \alpha >$$

I usem $g(\alpha) = \alpha$ punt fix de g. Aleshores:

$$e_{n+1} = x_{n+1} - \alpha = \frac{g^{(p)}(\xi_n)}{p!} e_n^p$$

Passem el terme e_n^p dividint i fem $x_n \to \alpha$ $\xi_n \to \alpha$:

$$\frac{|e_{n+1}|}{|e_n|^p} = \left| \frac{g^{(p)}(\xi_n)}{p!} \right| \longrightarrow \left| \frac{g^{(p)}(\alpha)}{p!} \right| \neq 0$$

que és la constant asimptòtica del mètode d'ordre p. Hem usat $g^{(p)}(\alpha) \neq 0$
$\Rightarrow$) Suposem $\exists j \in \{1, \ldots, p - 1\}$ amb $g^j(\alpha) = 0$ amb $g^{(i)}(\alpha) \neq 0$ per $i < j$. Però això vol dir que el mètode té ordre $j < p$, contradicció. c.v.d.

Ordre de convergència del mètode de Newton-Raphson

Recordem que es vol solucionar $f(x) = 0$ i el mètode de Newton-Raphson construeix la successió

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

que és un mètode d'iteració simple amb $x_{n+1} = g(x_n)$ amb $g(x) = x - \dfrac{f(x)}{f'(x)}$.
Buscarem l'ordre de convergècia a partir de la proposició anterior. La idea és la següent: El zero és simple, i tenim $f(\alpha) = 0$ i $f'(\alpha) \neq 0$.

$$g'(x) = \frac{f(x)f''(x)}{(f'(x))^2}$$

Notem $g'(\alpha) = 0$. Calculem $g''(x)$:

$$g''(x) = \frac{(f'(x))^3 f''(x) + f(x)f^{(3)}(x)(f'(x))^2 - 2f(x)f'(x)(f''(x))^2}{(f'(x))^4}$$

Substituint per α:

$$g''(\alpha) = \frac{(f'(\alpha))^3 f''(\alpha)}{(f'(\alpha))^4} = \frac{f''(\alpha)}{f'(\alpha)}$$

Observem que si $f''(\alpha) = 0$ té ordre almenys 3. En cas contrari, si $f''(\alpha) \neq 0$ llavors $g''(\alpha) \neq 0$ té ordre 2. Per tant, el coeficient asimptòtic és $\left|\dfrac{g''(\alpha)}{2!}\right| = \dfrac{1}{2}\left|\dfrac{f''(\alpha)}{f'(\alpha)}\right|$

Cas 1. Suposem α un zero simple. Fent desenvolupament de Taylor a $f(x)$ i $f'(x)$ obtenim:

$$f(x) = f(\alpha) + f'(\alpha)(x-\alpha) + \frac{f''(\alpha)}{2!}(x-\alpha)^2 + \frac{f^{(3)}(\alpha)}{3!}(x-\alpha)^3 + \dots$$
$$f'(x) = f'(\alpha) + f''(\alpha)(x-\alpha) + \frac{f^{(3)}(\alpha)}{2!}(x-\alpha)^2 + \frac{f^{(4)}(\alpha)}{3!}(x-\alpha)^3 + \dots$$

Substituïm a la fórmula $g(x)$ del mètode iteratiu i li restem α per obtenir un resultat en termes de l'error:

$$g(x) - \alpha =$$

$$x - \alpha - \frac{f(\alpha) + f'(\alpha)(x-\alpha) + \dots}{f'(\alpha) + f''(\alpha)(x-\alpha) + \dots} = \frac{\frac{1}{2}f''(\alpha)(x-\alpha)^2 + O((x-\alpha)^3)}{f'(\alpha) + f''(\alpha)(x-\alpha) + \frac{f^{(3)}(\alpha)}{2!}(x-\alpha)^2 + O((x-\alpha)^3)}$$

$$g(x_n) - \alpha = e_{n+1} = \frac{\frac{1}{2}e_n^2 + O(e_n^3)}{f'(\alpha) + f''(\alpha)e_n + O(e_n^2)}$$

Dividim per e_n^2:

$$\frac{e_{n+1}}{e_n^2} = \frac{\frac{1}{2}f''(\alpha) + O(e_n)}{f'(\alpha) + f''(\alpha)e_n + O(e_n^2)} \longrightarrow \frac{1}{2}\frac{f''(\alpha)}{f'(\alpha)}$$

Cas 2. Suposem α zero doble. És anàleg al que acabem de fer però es té $f'(\alpha) = 0$. Així doncs:

$$e_{n+1} = \frac{\frac{1}{2}f''(\alpha)e_n^2 + O(e_n^3)}{f'(\alpha) + f''(\alpha) + O(e_n^2)} = \frac{\frac{1}{2}f''(\alpha)e_n + O(e_n^3)}{f''(\alpha) + O(e_n^2)}$$

Aleshores, si $f''(\alpha) \neq 0$:

$$\frac{e_{n+1}}{e_n} = \frac{\frac{1}{2}f''(\alpha) + O(e_n)}{f''(\alpha) + O(e_n)} \longrightarrow \frac{1}{2}$$

Per tant, per un zero doble es té convergència lineal i coeficient asimptòtic $\frac{1}{2}$.

Resta el cas amb zero de multiplicitat m. Però abans introduïm la següent proposició.

Proposició 5. Sigui $x_{n+1} = G(x_n, \ldots, x_{n-k+1})$ un mètode de k ($k \geq 1$) punts per trobar un zero de f que té ordre de convergència p (i.e. $|x_{n+1} - \alpha| \leq M|x_n - \alpha|^p$. Sigui $\varepsilon_n = |x_n - \alpha|$, aleshores:

(1) Per $p = 1$ tenim $\varepsilon_n \leq M^n \varepsilon_0$

(2) Per $p > 1$ tenim $\varepsilon_n \leq \dfrac{(\overline{M}\varepsilon_0)^{p^n}}{\overline{M}}$ amb $\overline{M} = M^{\frac{1}{p-1}}$

Demostració.

(1) És immediat: $\varepsilon_n \leq M\varepsilon_{n-1} \leq \ldots \leq M^n \varepsilon_0$

(2) Per $p > 1$ fem: $\varepsilon_n \leq \varepsilon_{n-1}^p \leq M(M\varepsilon_{n-2}^p)^p = M^{1+p}\varepsilon_{n-2}^{p^2} \leq M^{1+p}(M\varepsilon_{n-1}^p)^{p^2} = M^{1+p+p^2}\varepsilon_{n-1}^{p^3} \leq \ldots \leq M^{1+p+p^2+\ldots+p^{n-1}}\varepsilon_0^{p^n} = M^{\frac{p^n-1}{p-1}}\varepsilon_0^{p^n} = \dfrac{\overline{M}^{p^n}\varepsilon_0^{p^n}}{\overline{M}}$

c.v.d.

Cas 3. Suposem zero α de multplicitat $m > 1$. Per tant

$$0 = f(\alpha) = f'(\alpha) = \ldots = f^{(m-1)}(\alpha) \quad f^m(\alpha) \neq 0$$

Estudiarem també l'ordre de convergència pel mètode de Newton modificat. Això és:

$$x_{n+1} = x_n - m\frac{f^(x_n)}{f'(x_n)}$$

Comencem l'estudi fent els desenvolupaments de Taylor corresponents:

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}}{i!}(x-\alpha)^i + \frac{f^{(m+1)}(\xi)}{(m+1)!}(x-\alpha)^{m+1} = \frac{f^{(m)}}{m!}(x-\alpha)^m + \frac{f^{(m+1)}(\xi)}{(m+1)!}(x-\alpha)^{m+1}$$

amb $\xi \in < x, \alpha >$. Hem aplicat que les primeres $m-1$ derivades s'anul·len. Anàlogament per a f':

$$f'(x) = \sum_{i=1}^{n} \frac{f^{(i)}}{(i-1)!}(x-\alpha)^{i-1} + \frac{f^{(m+1)}(O)}{m!}(x-\alpha)^m \quad O \in < x, \alpha >$$

Notem $e_n = x_n - \alpha$. Obtenim:

$$f(x) = \frac{f^{(m)}}{m!}e_n^m + \frac{f^{(m+1)}(\xi_n)}{(m+1)!}e_n^{m+1} \quad \xi_n \in < x_n, \alpha >$$

$$f'(x_n) = \frac{f^{(m)}}{(m-1)!}e_n^{m-1} + \frac{f^{(m+1)}(O_n)}{m!}e_n^m \quad O_n \in < x_n, \alpha >$$

Calculem el quocient $\dfrac{f(x_n)}{f'(x_n)}$:

$$\frac{f(x_n)}{f'(x_n)} = \frac{\dfrac{f^{(m)}}{m}e_n + \dfrac{f^{(m+1)}(\xi_n)}{m(m+1)}e_n^2}{f^m(\alpha) + \dfrac{f^{(m+1)}(O_n)}{m}e_n} = \frac{\dfrac{1}{m}e_n + \dfrac{f^{(m+1)}(\xi_n)}{m(m+1)f^{(m)}(\alpha)}e_n^2}{1 + \dfrac{f^{(m+1)}(O_n)}{mf^{(m)}}e_n}$$

On a la primera igualtat hem dividit per $\dfrac{e_n^{m-1}}{(m-1)!}$ i a la segona igualtat hem dividit per $f^{(m)}(\alpha)$. Anomenem $\beta = \dfrac{f^{(m+1)}(O_n)}{mf^{(m)}}e_n$. Podem aplicar el següent resultat: Per β prou petit $(1+\beta)^{-1} = 1 - \beta + \beta^2$. Per tant:

$$\left(\frac{1}{m}e_n + \frac{f^{(m+1)}(\xi_n)}{m(m+1)f^{(m)}(\alpha)}e_n^2\right)\left(1 - \frac{f^{(m+1)}(O_n)}{mf^{(m)}}e_n + O(e_n^2)\right)$$

Finalment:

$$\frac{1}{m}e_n + \left(\frac{f^{(m+1)}(\xi_n)}{m(m+1)f^{(m)}(\alpha)} - \frac{f^{(m+1)}(O_n)}{m^2 f^{(m)}(\alpha)}\right)e_n^2 + O(e_n^3)$$

Tenim calculat el quocient $\dfrac{f(x_n)}{f'(x_n)}$. Ara cal substituir-ho al mètode de Newton-Raphson:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

$$e_{n+1} = x_{n+1} - \alpha = x_n - \alpha - \frac{f(x_n)}{f'(x_n)}$$

$$e_{n+1} = e_n - \left[\frac{1}{m}e_n + O(e_n^2)\right] = \left(1 - \frac{1}{m}\right)e_n + O(e_n^2)$$

$$\frac{e_{n+1}}{e_n} = 1 - \frac{1}{m} + O(e_n) \longrightarrow 1 - \frac{1}{m} \neq 0$$

És a dir: Tenim convergència lineal amb constant asimptòtica $1-\frac{1}{m}$.

Substituïm-ho ara al Newton modificat:

$$e_{n+1} = e_n - m\frac{f(x_n)}{f'(x_n)} = e_n - m\left[\frac{1}{m}e_n + \left(\frac{f^{(m+1)}(\xi_n)}{m(m+1)f^{(m)}(\alpha)} - \frac{f^{(m+1)}(O_n)}{m^2 f^{(m)}(\alpha)}\right)e_n^2 + O(e_n^3)\right]$$

$$\lim_{n\to\infty}\left|\frac{e_{n+1}}{e_n^2}\right| = \left|\frac{f^{(m+1)}(\alpha)}{f^{(m)}(\alpha)}\left(\frac{1}{m+1} - \frac{1}{m}\right)\right| = \frac{1}{m+1}\left|\frac{f^{(m+1)}(\alpha)}{f^{(m)}(\alpha)}\right|$$

on $\xi_n, O_n \longrightarrow \alpha$ si $n \to \infty$. La convergència serà quadràtica si $f^{(m+1)}(\alpha) \neq 0$. En cas contrari, serà almenys quadràtica.

Observacions. Si bé no ho demostrarem, els mètodes vistos en aquest tema tenen els següents ordres de convergència:

- Mètode de la bisecció: convergència lineal
- Mètode de la secant per a un zero simple: és $p = \frac{1+\sqrt{5}}{2} \simeq 1.618$ raó àuria !!
- Regula falsi: convergència almenys lineal si $f^{(2)}(\alpha) \neq 0$

Problema. Hem definit $\frac{|e_{n+1}|}{|e_n|^p} = c$, però $e_n = x_n - \alpha$ pressuposa conèixer l'error, i per tant el coneixement del zero α. Per solucionar-ho, definim l'operador $\Delta x_n = x_{n+1} - x_n$. Tenim que si $\dfrac{|e_{n+1}|}{|e_n|^p} \to c$ també $\dfrac{\Delta x_{n+1}}{\Delta x_n} \to c$

Demostració.

Observem $\Delta x_n = x_{n+1} - x_n = x_{n+1} - \alpha + \alpha - x_n = e_{n+1} - e_n = e_n\left[\dfrac{e_{n+1}}{e_n} - 1\right]$. Anàlogament $\Delta x_{n+1} = e_{n+1}\left[\dfrac{e_{n+2}}{e_{n+1}} - 1\right]$. Fem el quocient:

$$\frac{\Delta x_{n+1}}{\Delta x_n} = \frac{e_{n+1}\left(\frac{e_{n+2}}{e_{n+1}} - 1\right)}{e_n\left(\frac{e_{n+1}}{e_n} - 1\right)} = \frac{e_{n+1}}{e_n}\left(\frac{\frac{e_{n+2}}{e_{n+1}} - 1}{\frac{e_{n+1}}{e_n} - 1}\right) \longrightarrow c \cdot \frac{c-1}{c-1} = c \quad \square$$

Aquest resultat és anàleg per a convergència quadràtica:

$$\frac{e_{n+1}}{e_n^2} \longrightarrow c_Q \Longrightarrow \frac{\Delta x_{n+1}}{(\Delta x_n)^2} \longrightarrow c_Q$$

Per veure-ho observem que:

$$\frac{e_{n+1}}{e_n} = \frac{e_{n+1}}{e_n^2} \cdot e_n \longrightarrow c_Q \cdot 0 = 0$$

Així doncs:

$$\frac{\Delta x_{n+1}}{(\Delta x_n)^2} = \frac{e_{n+1}}{e_n^2} \cdot \frac{\left(\dfrac{e_{n+2}}{e_{n+1}} - 1\right)}{\left(\dfrac{e_{n+1}}{e_n} - 1\right)} \longrightarrow c_Q \cdot \frac{0-1}{0-1} = c_Q \quad \square$$