

# Mètodes Numèrics

## Pràctica 2: Zeros de funcions

Marc Seguí Coll: Conclusions individuals  
*Universitat Autònoma de Barcelona*

Març 2019

### 1 Conclusions personals

#### 1.1 Tria del mètode per a la determinació dels zeros simples d'una funció

Gràcies a la potència del teorema de Bolzano, es pot determinar de forma no gaire complicada l'interval on està continguda una arrel simple d'una certa funció. A partir d'aquí els diversos mètodes per al càlcul de zeros simples presenten certs avantatges uns en front dels altres. Al càlcul numèric, probablement el que més ens interessa és trobar mètodes amb ordre de convergència gran. Tot i això, com s'ha vist amb el mètode de Newton, una mala elecció de la llavor inicial pot portar a casos on la convergència sigui molt lenta o directament casos on no hiagi convergència directament.

El mètode de la bisecció no presenta aquest problema ja que un cop comprovada la condició dels extrems de l'interval per Bolzano, amb certes iteracions del mètode s'assoleix el valor de l'arrel amb precisió desitjada. Malauradament l'ordre de convergència és prou lent comparat amb la majoria dels restants.

Finalment, la secant presenta un mètode intermig entre els dos (reflectit també en el seu ordre de convergència) i per tant presenta una millora respecte la bisecció i sobre Newton inclús, si la llavor d'aquest darrer no s'acosta prou a l'arrel.

Altres mètodes com el de *Halley* permeten accelerar la convergència, però amb una combinació dels tres mètodes comentats anteriorment els resultats són prou bons. De totes formes, la recerca de mètodes iteratius amb ordre de convergència elevat és important per als càlculs quotidians, bastant més tediosos que els duts a terme en aquesta pràctica.

#### 1.2 Precisió limitada en els iterats

La representació en punt flotant limita la precisió amb la que s'efectuen els càlculs, com hem vist repetides vegades. En mètodes iteratius, el problema s'incrementa ja que hi intervenen diversos termes. En el cas observat, el càlcul successiu de mitjanes aritmètiques i geomètriques pot donar lloc a termes "iguals" per a la màquina, provocant que alguns dels termes de cada iterat no variïn o inclús cancel·lacions en altres termes. Això provoca resultats prou greus, com el fet que una successió convergent sigui divergent o que aquesta acabi convergint a un nombre prou diferent al límit analític.

Cal tenir cura doncs de tractar cada terme de la iteració per separat, trobar i eliminar qualsevol problema o directament descartar el mètode si aquest no pot satisfer els requeriments.

## 2 Problema opcional

Volem trobar una aproximació de l'arrel quadrada d'un cert nombre emprant l'expressió

$$\sqrt{1+x} = f(x)\sqrt{1+g(x)},$$

on  $g$  és un infinitèssim d'ordre més petit que  $x$  per a  $x$  tendint a 0. si triem  $f(x)$  com una aproximació de  $\sqrt{1+x}$  llavors podem calcular  $g(x)$  com

$$g(x) = \frac{1+x}{f(x)^2} - 1$$

### 2.1 Tria d' $f(x)$ com quocient de polinomis

Volem que  $f(x)$  sigui una aproximació de  $\sqrt{1+x}$ . Començarem considerant dos polinomis  $p(x)$  i  $q(x)$  del mateix grau i construint  $f(x) = p(x)/q(x)$  de forma que el seu desenvolupament de MacLaurin coincideixi fins a cert grau. Primer, suposarem que els polinomis són de grau 1, és a dir,  $p(x) = p_a(x) = ax+b$  i  $q(x) = q_a(x) = cx+d$ . Imposant que  $f(x) = f_a(x)$  i  $\sqrt{1+x}$  tinguin els primers tres coeficients iguals, s'ha de complir que el desenvolupament de MacLaurin de  $\omega(x) = p_a(x) - q_a(x)\sqrt{1+x}$  tingui els tres primers termes nuls ( $f_a$  és l'aproximant de Padé de  $\sqrt{1+x}$ ). Ha de ser doncs que  $\omega''(0) = \omega'(0) = \omega(0) = 0$ . Obtenim així que

$$\begin{aligned}\omega(0) = 0 &\Rightarrow b = d & \omega'(0) = 0 &\Rightarrow 2a - 2c = d & \omega''(0) = 0 &\Rightarrow 4c = d \\ f_a(x) &= \frac{ax+b}{cx+d} = \frac{3cx+4c}{cx+4c} = \frac{3x+4}{x+4}\end{aligned}$$

Per al cas de polinomis de 3r grau, amb **SageMath** s'obté l'aproximat de Padé

$$f_c(x) = \frac{7x^3 + 56x^2 + 112x + 64}{x^3 + 24x^2 + 80x + 64} \Rightarrow g_c(x) = \frac{x^7}{(7x^3 + 56x^2 + 112x + 64)^2}$$

### 2.2 Identitats i desigualtats amb $g(x)$ i $f(x)$

Veurem per inducció sobre  $k$  que si  $a_0 = x$ ,  $a_{n+1} = g(a_n)$  i  $b_n = f(a_n)$  llavors

$$\sqrt{1+x} = \left( \prod_{j=0}^k b_j \right) \sqrt{1+a_{k+1}} \quad (1)$$

Suposem  $k = 0$ . D'aquesta forma

$$b_0\sqrt{1+a_1} = f(a_0)\sqrt{1+g(a_0)} = f(x)\sqrt{1+\frac{1+x}{f(x)^2}-1} = \sqrt{f(x)^2\frac{1+x}{f(x)^2}} = \sqrt{1+x}$$

Fixem  $k \geq 0$  i suposem la igualtat certa fins a  $k$ . Vegem-la per a  $k+1$ :

$$\begin{aligned}\left( \prod_{j=0}^{k+1} b_j \right) \sqrt{1+a_{k+2}} &= \left( \prod_{j=0}^k b_j \right) (b_{k+1})\sqrt{1+g(a_{k+1})} = \left( \prod_{j=0}^k b_j \right) f(a_{k+1})\sqrt{1+\frac{1+a_{k+1}}{f(a_{k+1})^2}-1} \\ &= \left( \prod_{j=0}^k b_j \right) \sqrt{f(a_{k+1})^2\frac{1+a_{k+1}}{f(a_{k+1})^2}} = \left( \prod_{j=0}^k b_j \right) \sqrt{1+a_{k+1}} = \sqrt{1+x}\end{aligned}$$

La darrera igualtat és deguda la hipòtesi d'inducció. Per tant, la igualtat és certa per a tot  $k \in \mathbb{N} \cup \{0\}$ .

Determinarem ara quin és l' $n > 0$  tal que  $g_a(x)$  (la obtinguda a partir de la  $f_a(x)$  com a quocient entre  $p_a(x)$  i  $q_a(x)$  lineals) satisfà que  $g_a(x) = O(x^n)$ . Calculem doncs:

$$g_a(x) = \frac{(x+4)^2(x+1) - (3x+4)^2}{(3x+4)^2} = \frac{x^3}{(3x+4)^2}$$

$$\lim_{x \rightarrow 0} \frac{g_a(x)}{x^3} = \lim_{x \rightarrow 0} \frac{x^3}{x^3(3x+4)^2} = \lim_{x \rightarrow 0} \frac{1}{(3x+4)^2} = \frac{1}{16}$$

Per tant, l' $n$  corresponent és 3, tenim que  $g_a(x) = O(x^3)$ . Veurem ara que  $g_a(x)$  és contractiva per a  $x > 0$ . Siguin  $x, y \in \mathbb{R}$  tals que  $x > 0$  i  $y > 0$ . Llavors:

$$|g_a(x) - g_a(y)| = \left| \frac{x^3}{(3x+4)^2} - \frac{y^3}{(3y+4)^2} \right| = \frac{1}{9} \left| \frac{x^3}{(x+\frac{4}{3})^2} - \frac{y^3}{(y+\frac{4}{3})^2} \right| \leq \frac{1}{9} \left| \frac{x^3}{x^2} - \frac{y^3}{y^2} \right| = \frac{1}{9} |x - y|$$

Com  $0 \leq 1/9 < 1$ ,  $g_a(x)$  és contractiva. La desigualtat anterior es deu al fet que com  $x > 0$  i  $y > 0$ , eliminant el sumand  $4/3$  del denominador cada un dels termes augmenta en valor absolut i per tant la diferència s'incrementa, raó del signe menor o igual.

Finalment, demostrarem la següent desigualtat:

$$\left| \sqrt{1+x} - \prod_{j=0}^k b_j \right| \leq \frac{a_{k+1}}{2} \sqrt{1+x} \quad (2)$$

Com  $a_0 = x > 0$  i  $a_k \rightarrow 0$  monòtonament,  $\sqrt{1+a_k} > 1$  i té límit 1. Notem que podem aïllar el productori de la fórmula (1), obtenint així

$$\left| \sqrt{1+x} - \prod_{j=0}^k b_j \right| = \sqrt{1+x} \left| 1 - \frac{1}{\sqrt{1+a_{k+1}}} \right|$$

Per tant, podem treure valors absoluts ja que  $\sqrt{1+a_k} > 1$ . Emprant el desenvolupament de MacLaurin de  $1/\sqrt{1+x}$ :

$$1 - \frac{1}{\sqrt{1+x}} = 1 - 1 + \frac{x}{2} + O(x^2) = \frac{x}{2} - \frac{3}{8(1+\xi)^{5/2}} x^2 \leq \frac{x}{2}$$

La darrera desigualtat prové del fet que  $0 < \xi < x$ . Ajuntant aquests resultats:

$$\left| \sqrt{1+x} - \prod_{j=0}^k b_j \right| = \sqrt{1+x} \left( 1 - \frac{1}{\sqrt{1+a_{k+1}}} \right) \leq \frac{a_{k+1}}{2} \sqrt{1+x}$$

com volíem.

## 2.3 Programes, càlculs i comentaris

Mitjançant el codi `pop.c` es calcula el valor de  $\sqrt{1+x}$  de forma que aproximem  $\sqrt{1+x} \approx \prod_{j=0}^k b_j$ . D'aquesta manera es demana a l'usuari una entrada per pantalla del nombre del qual es vol calcular l'arrel i el nombre  $k$  que definirà el tamany del productori. Torna per pantalla el valor del productori corresponent i el valor de cada iterat  $a_k$  i  $b_k$ . El càlcul dels termes es porta a terme amb les funcions  $f_c(x)$  i  $g_c(x)$ .

Amb el vist anteriorment, si  $x > 0$  podem afirmar que

$$\left| \sqrt{1+x} - \prod_{j=0}^k b_j \right| \leq \frac{a_{k+1}}{2} \sqrt{1+x} \leq \frac{a_{k+1}}{2} (1+x)$$

Per demostrar el fet que  $|\sqrt{2} - b_0 b_1 b_2| < 5 \cdot 10^{-255}$  aplicarem la fórmula anterior, donant així

$$\left| \sqrt{1+1} - \prod_{j=0}^2 b_j \right| \leq \frac{a_{2+1}}{2} (1+1) \Leftrightarrow |\sqrt{2} - b_0 b_1 b_2| \leq a_3$$

Al programa `poperr.c` es determina el valor d' $a_3$  en *double* i s'obté que aquest és igual a  $1.042173076841195 \cdot 10^{-262}$ , per tant

$$|\sqrt{2} - b_0 b_1 b_2| \leq 1.042173076841195 \cdot 10^{-262} < 5 \cdot 10^{-255}$$

Com es pot observar aquest mètode per al càlcul d'arrels és prou efectiu, no només permet la determinació de l'arrel d'un nombre amb extremada precisió, sino que el procés no requereix gaires iterats per arribar a tenir unes 30 xifres correctes. Fixem-nos que en el cas de  $\sqrt{2}$  és suficient amb 3 iterats per tenir uns 250 decimals correctes!

El funcionament del mètode és degut que  $g$  és contractiva i que  $f$  és una aproximació de  $\sqrt{1+x}$ . La contractivitat de  $g$  i el fet de ser infinitèssim més petit que  $x$  quan  $x \rightarrow 0$  permet que els termes  $a_k$  vagin al 0 excessivament ràpid fent que el productori de  $b_j$  s'aproximi de manera quasi perfecta a l'arrel per a valors de  $k$  petits.

Aquest fet queda palpable si notem que  $g_c(x) = O(x^7)$  quan  $x \rightarrow 0$ . Llavors l'ús de polinomis de major grau en l'aproximació de Padé permet que la  $g$  sigui un infinitèssim cada vegada més petit i per tant, més efectiu. Efectivament, com es pot veure a `pop.c` un cop els  $a_k$  s'acosten a 0 l'ordre de convergència és de 7!