

EINES BÀSIQUES DE CÀLCUL NUMÈRIC

A. Aubanell

A. Benseny

A. Delshams

INTRODUCCIÓ

En els molt diversos camps de la ciència, la tecnologia, la medicina, l'economia, les ciències socials, etc, es descriuen tot sovint fenòmens reals mitjançant models matemàtics. L'estudi d'aquests models fa possible un coneixement més profund del fenomen i àdhuc una previsió de la seva evolució futura. Buscar i aplicar les eines més adients per trobar solucions a problemes basats en aquests models constitueix l'objectiu principal de la matemàtica aplicada. Es tracta d'un art apassionant que, dissortadament, no sempre pot recórrer als mètodes analítics clàssics per diverses raons: no s'adeqüen al model concret, llur aplicació resulta excessivament enrevessada, la solució formal resultant és tan complexa que fa impossible qualsevol interpretació posterior, etc. En tals casos són útils les tècniques numèriques que, mitjançant una labor de càlcul més intensa o menys, arriben a solucions aproximades. La notable tasca calculística que comporta l'aplicació de la majoria d'aquests mètodes fa que el seu ús estigui fortament lligat a la utilització d'aparells de càlcul automàtic. Així, sense el desenvolupament que s'ha produït en el camp de la informàtica, resultaria difícilment imaginable l'actual nivell d'utilització de tècniques numèriques en àmbits cada cop més diversos.

Aquest llibre és fruit d'una àmplia tasca docent en la Facultat de Matemàtiques de la Universitat de Barcelona. El seu contingut representa un primer curs de càlcul numèric i va dirigit a estudiants de carreres científiques, tècniques o socials que vulguin conèixer, de manera tan pràctica com sigui possible, eines bàsiques que els permetin de fer front a qüestions numèriques amb comoditat i rigor.

L'aplicació pràctica de molts dels mètodes que es presenten en aquest text requereix l'emprament de calculadores i d'ordinadors. Molts dels problemes numèrics d'aquest llibre poden ser abordats utilitzant calculadores de butxaca. La complexitat d'alguns mètodes o llur escala d'aplicació fa imprescindible l'ús d'ordinadors. Per aquest fet és altament aconsellable que el lector conegui algun llenguatge de programació que li permeti d'implementar mètodes numèrics en programes.

En aquest llibre es desenvolupen, agrupats en cinc capítols, els temes més bàsics del càlcul numèric:

- El concepte d'error és consubstancial amb el càlcul numèric. En cada problema, és d'una gran importància fer un seguiment dels errors comesos a fi de poder estimar el grau d'aproximació de la solució que se n'obté. En aplicar tècniques numèriques a un problema determinat, cal estudiar els diversos tipus d'error que afecten les seves solucions: els errors propis del mètode numèric emprat complementats amb els errors provinents de les dades i de les operacions. En el primer capítol només es poden donar tècniques per tal d'estimar o fitar els errors dels dos darrers tipus, tenint també en

compte les limitacions de l'eina feta servir (calculadora, ordinador ...). En els capítols següents, s'ha procurat acompanyar la presentació de cada mètode numèric amb la informació necessària per poder analitzar el seu error.

- Els models matemàtics més bàsics són els models lineals, que sovint provenen d'una reducció d'altres de més complexos. L'anàlisi dels problemes associats als models lineals comporta el càlcul de la solució de sistemes lineals o de llurs valors i vectors propis associats. El capítol II contempla una gran varietat de mètodes generals per a la realització d'aquests càlculs. En cada situació, l'elecció concreta d'un mètode depèn de diversos factors, com ara l'estructura del sistema (simetria, escassetat de coeficients no nuls ...), l'eficiència, l'estabilitat numèrica, etc. És així que es poden triar mètodes directes, mètodes iteratius, mètodes més estables a costa de més operacions, mètodes específics per a sistemes amb estructura especial, etc.
- Les tècniques d'aproximació i interpolació ofereixen procediments per extreure un model senzill d'una realitat més complexa, expressada bé per un conjunt de dades experimentals, bé per una funció d'avaluació complicada. Aquestes reduccions, explicades en el capítol III, permeten donar formes alternatives més simples de càlcul de funcions i són la base de la majoria dels mètodes numèrics per a l'obtenció de derivades, integrals i zeros de funcions, presentats en els capítols posteriors. Les funcions interpoladores i d'aproximació que s'hi fan servir són primordialment polinòmiques i trigonomètriques. Els mètodes d'aproximació tenen com a objectiu minimitzar la funció error segons algun criteri. Les aproximacions per mínims quadrats i minimax que es presenten corresponen a dos criteris diferents de mesura de la dita funció error.
- Els càlculs amb funcions sovint no es poden portar a terme analíticament o no és eficient de fer-ho. En el capítol IV es descriuen diferents fórmules de derivació, integració i sumació numèrica de funcions, acompanyades del seu error. L'existència d'expressions asimptòtiques d'aquests errors permet dissenyar estratègies d'extrapolació per tal d'assolir una utilització més eficaç de les dites fórmules. Es concedeix molta importància al càlcul amb operadors per la possibilitat que ofereix d'una reformulació més àgil d'algun dels mètodes presentats en aquest capítol i en l'anterior.
- Una de les tasques més genuïnes del càlcul numèric és la resolució d'equacions no lineals. Els mètodes de resolució es basen en la construcció d'una successió obtinguda iterativament que convergeixi a la solució. La seva elecció depèn de la funció i de la solució cercada. En cada cas, els diversos mètodes es diferencien en la velocitat de convergència de les successions prop de la solució i en l'esforç de càlcul que requereixen. En el capítol V s'exposa un recull dels procediments més usuals, tant per a equacions generals com per a equacions polinòmiques específicament. Es clou el capítol amb una breu introducció a la resolució de sistemes d'equacions no lineals.

En tot moment s'ha procurat defugir d'un text reduït a un catàleg de mètodes, receptes i trucs. S'ha insistit en les idees generals associades als temes presentats: plantejament del problema, conceptes que s'hi mouen, estratègies de resolució, estudi d'errors, qüestions de convergència ... En l'exposició de mètodes concrets s'ha buscat sempre un marc globalitzador (trets comuns, característiques generals, etc) que eviti una casuística excessivament dispersa.

Cada capítol consta, a banda de les seccions dedicades a l'exposició teòrica, d'una secció de comentaris bibliogràfics, d'una secció de problemes resolts i d'una de problemes proposats.

- En les seccions teòriques, s'ha cercat un equilibri entre la claredat del desenvolupament i la profunditat conceptual. El càlcul numèric és, però, més pragmàtic que fonamentalista, és més eina que matèria primera. Aquesta idea ha estat present en aquestes seccions, de manera que ens ha preocupat més la presentació del marc general i l'exposició concreta de les eines de càlcul que els detalls de fonamentació d'alguns conceptes.
- En els comentaris bibliogràfics se subministren referències complementàries al material teòric que, en alguns casos, contempnen demostracions no presentades explícitament. S'aporten també cites que contenen programes implementant els mètodes presentats en diferents llenguatges de programació per tal que el lector pugui utilitzar-los directament o bé comparar-los amb els elaborats per ell mateix.
- Les seccions de problemes resolts constitueixen el nucli de cada capítol. Tant la concepció inicial com l'objectiu final de la redacció d'aquest llibre han estat basats en el principi que la millor forma de familiaritzar-se en l'ús de qualsevol eina és fer-la servir degudament. Per això, s'ha tingut molta cura de l'exposició del procés de resolució dels problemes. La inclusió d'aquest material té dos objectius:
 - Simplificar les exposicions teòriques, remetent a la part de problemes aquells detalls que les farien massa feixugues.
 - Oferir exemples concrets d'aplicació dels mètodes exposats. En aquest sentit, s'ha intentat que els problemes resolts abastin essencialment totes les eines teòriques presentades i, a través de la pràctica, en permetin una millor comprensió.
- Els problemes proposats, de gran diversitat pel que fa a la seva dificultat, són arguments de reforç que permeten l'exercitació del lector.

No volem pas acabar aquesta introducció sense donar el nostre agraïment als companys de docència de l'assignatura de Càlcul Numèric a la Facultat de Matemàtiques. De tots ells hi ha alguna cosa en aquest text, però ens complau fer esment d'en Joan Llopart, col·laborador en uns apunts previs, d'en Joaquim Font, pacient calculador de zeros de polinomis i d'en Gerard Gómez, conseller en les darreres etapes.

Desitgem fer una menció especial al nostre mestre comú, en Carles Simó. Ell va introduir-nos en l'art del càlcul numèric i, tot fent camí al seu costat, va descobrir-nos paisatges d'una insospitada bellesa. A ell es deuen els encerts que aquest llibre pugui aportar, nosaltres ens fem responsables dels seus errors. Siguin aquestes paraules un testimoni de gratitud envers el seu amical mestratge.

La Maribel, la Fanny i la Rosalia han demostrat tenir una paciència que tendeix a l'infinit, la seva feina també és present en aquestes pàgines.

Els autors, juliol de 1991.

ÍNDEX DE CONTINGUTS

INTRODUCCIÓ	v
ÍNDEX DE CONTINGUTS	viii
1 ERRORS	1
1.1 CONCEPTES GENERALS	1
1.1.1 Definicions	1
1.1.2 Fonts d'error	1
1.2 ESTIMACIÓ I FITACIÓ D'ERRORS	4
1.2.1 Propagació dels errors de les dades	4
1.2.2 Propagació dels errors en els càlculs	5
1.2.3 Errors de truncament	5
1.3 TRACTAMENT INTERVALAR I ESTADÍSTIC	6
1.3.1 Anàlisi d'interval	6
1.3.2 Anàlisi estadística	7
COMENTARIS BIBLIOGRÀFICS	10
PROBLEMES RESOLTS	11
PROBLEMES PROPOSATS	46
2 SISTEMES LINEALS	49
2.1 CONCEPTES BÀSICS	49
2.1.1 Tipus de matrius	49
2.1.2 Definicions	50
2.1.3 Propietats importants	53
2.2 SISTEMES D'EQUACIONS	55
2.2.1 Introducció	55
2.2.2 Resolució de sistemes triangulars	56
2.2.3 Mètodes gaussians	56
2.2.4 Mètodes d'ortogonalització	60
2.2.5 Càlcul de determinants i inverses de matrius	67
2.2.6 Anàlisi de l'error	68
2.2.7 Mètodes iteratius	70
2.2.8 Mètodes iteratius de sobrerelaxació	72
2.3 VALORS I VECTORS PROPIS	73
2.3.1 Introducció	73
2.3.2 Deflació de matrius	74

2.3.3	Mètodes de la potència	76
2.3.4	Mètodes de Jacobi	79
2.3.5	Mètodes de reducció	81
2.3.6	Valors i vectors propis de matrius reduïdes	84
2.3.7	Mètodes de factorització	87
	COMENTARIS BIBLIOGRÀFICS	88
	PROBLEMES RESOLTS	90
	PROBLEMES PROPOSATS	126
3	INTERPOLACIÓ I APROXIMACIÓ DE FUNCIONS	145
3.1	INTERPOLACIÓ	145
3.1.1	Concepte d'interpolació	145
3.1.2	Interpolació polinòmica	146
3.1.3	Mètodes de càlcul del polinomi interpolador	149
3.1.4	Interpolació de Taylor	152
3.1.5	Interpolació d'Hermite	157
3.2	APROXIMACIÓ DE FUNCIONS	159
3.2.1	Introducció al problema general d'aproximació	159
3.2.2	Aproximació per mínims quadrats	163
3.2.3	Resolució de les equacions normals	168
3.2.4	Aproximació minimax	182
	COMENTARIS BIBLIOGRÀFICS	189
	PROBLEMES RESOLTS	190
	PROBLEMES PROPOSATS	251
4	DERIVACIÓ, INTEGRACIÓ I SUMACIÓ	262
4.1	DERIVACIÓ NUMÈRICA	262
4.1.1	Introducció	262
4.1.2	Derivades primeres	262
4.1.3	Derivades d'ordre superior	263
4.2	INTEGRACIÓ NUMÈRICA	265
4.2.1	Introducció	265
4.2.2	Integració amb abscisses donades	265
4.2.3	Regles (compostes) d'integració numèrica	270
4.2.4	Integració gaussiana	274
4.3	SUMACIÓ NUMÈRICA	280
4.3.1	Introducció	280
4.3.2	Fites dels residus de les sèries	282
4.3.3	Mètodes de sumació numèrica	284
4.4	EXTRAPOLACIÓ	287
4.4.1	Introducció	287
4.4.2	Mètode de Richardson d'extrapolació repetida	288
4.5	CÀLCUL AMB OPERADORS	289
4.5.1	Introducció	289
4.5.2	Propietats dels operadors	290
4.5.3	Aplicacions del càlcul amb operadors	293

COMENTARIS BIBLIOGRÀFICS	296
PROBLEMES RESOLTS	297
PROBLEMES PROPOSATS	333
5 EQUACIONS NO LINEALS	349
5.1 EQUACIONS EN UNA VARIABLE	349
5.1.1 Introducció	349
5.1.2 Mètodes iteratius d'aproximació de solucions	350
5.1.3 Ordre de convergència i constant asimptòtica de l'error	355
5.1.4 Acceleració de la convergència	356
5.1.5 Classificació de mètodes iteratius	357
5.2 EQUACIONS POLINOMIALS	360
5.2.1 Introducció	360
5.2.2 Avaluació i deflació de polinomis	361
5.2.3 Fitació de zeros de polinomis	362
5.2.4 Separació de zeros reals de polinomis. Mètode de Sturm	362
5.2.5 Mètodes numèrics per al càlcul de zeros de polinomis	363
5.3 SISTEMES NO LINEALS	369
5.3.1 Introducció	369
5.3.2 Mètode d'iteració simple en diverses variables	370
5.3.3 Mètode de Newton en diverses variables	370
COMENTARIS BIBLIOGRÀFICS	372
PROBLEMES RESOLTS	373
PROBLEMES PROPOSATS	396
BIBLIOGRAFIA	407

CAPÍTOL 1

ERRORS

1.1 CONCEPTES GENERALS

1.1.1 Definicions

Tot sovint no coneixem el *valor exacte* $\bar{x}$ d'una certa magnitud i hem de conformar-nos amb un *valor aproximat* x . Aleshores, definim:

- *Error absolut de x* : $e_a(x) = x - \bar{x}$.
- *Error relatiu de x* :

$$e_r(x) = \frac{e_a(x)}{\bar{x}} ,$$

si $\bar{x} \neq 0$.

A la pràctica, si $x \neq 0$, se sol prendre

$$e_r(x) \simeq \frac{e_a(x)}{x} .$$

En general, no coneixem aquests errors exactament, sinó només una *fitxa* d'aquests. Això és, un nombre $\epsilon_a(x)$ tal que $|e_a(x)| \leq \epsilon_a(x)$, per a l'error absolut; o bé, $\epsilon_r(x)$ tal que $|e_r(x)| \leq \epsilon_r(x)$, per a l'error relatiu. Usarem també la notació: $\bar{x} = x \pm \epsilon_a(x)$ o $x = \bar{x} \pm \epsilon_a(x)$, per a l'error absolut, i $\bar{x} = x(1 \pm \epsilon_r(x))$ o $x = \bar{x}(1 \pm \epsilon_r(x))$, per al relatiu.

1.1.2 Fonts d'error

Des d'un punt de vista numèric, ens interessen tres *fonts d'error*: errors en les dades d'entrada, errors d'arrodoniment durant el càlcul i error de truncament del mètode emprat.

Error en les dades d'entrada

Els errors en les dades poder ser deguts a dues causes:

MESURAMENTS INCORRECTES

Els aparells de mesura subministrats per la tecnologia no presenten una precisió indefinidament fina. Això fa que els valors mesurats estiguin afectats d'errors que depenen bàsicament de la precisió dels aparells.

FINITUD DE LA REPRESENTACIÓ DIGITAL D'UNA DADA

Elegida una base natural $b \geq 2$, qualsevol nombre real x no negatiu pot ser representat en la forma

$$x = a_n b^n + a_{n-1} b^{n-1} + \cdots + a_1 b + a_0 + a_{-1} b^{-1} + a_{-2} b^{-2} + \cdots ,$$

amb $a_j \in \mathbb{Z}$, $0 \leq a_j < b$ ($j \leq n$), que habitualment s'escriu

$$x = a_n a_{n-1} \dots a_1 a_0 . a_{-1} a_{-2} \dots b ,$$

i s'anomena *representació digital de x en la base b* . Els coeficients a_j ($j \leq n$) reben el nom de *dígits* o *xifres*. Aquesta representació és única, excepte per als nombres racionals x de la forma

$$x = \frac{k}{b^n} \quad (k, n \in \mathbb{Z}) ,$$

que en tenen dues. Així, per exemple,

$$\frac{132}{100} = 1.319999 \dots_{10} = 1.32_{10} ,$$

on 1.32_{10} és la representació finita de $\frac{132}{100}$, formada només per 3 xifres significatives: $a_0 = 1$, $a_{-1} = 3$, $a_{-2} = 2$ (per a més exemples, vegeu els problemes I.1-I.4).

Suposem ara que fem els nostres càlculs amb una calculadora que pot representar nombres amb t dígits en base b . Llavors la representació d'un nombre x amb més de t dígits no nuls no serà exactament igual a x , i l'anomenarem $\text{fl}(x)$, *representació en punt flotant de x* . En general,

$$\text{fl}(x) = m \cdot b^q ,$$

on $q \in \mathbb{Z}$ i $m = \pm 0.a_1 a_2 \dots a_t b$, amb $a_j \in \mathbb{Z}$, $0 \leq a_j < b$ ($j = 1 \div t$) i $a_1 \neq 0$; q s'anomena *exponent* i m , *mantissa*.

El pas de x a $\text{fl}(x)$ es pot fer per *tall*, simplement suprimint els dígits de x a partir de a_t , o per *arrodoniment*, escollint $\text{fl}(x)$ de manera que l'error $\text{fl}(x) - x$ sigui mínim. Quan aquesta condició dóna dos possibles arrodoniments, comunament s'escollix el que té valor absolut més gran (si aquests casos apareguessin dins d'una cadena llarga de càlculs, per tal d'evitar qualsevol tipus de desviació de l'error d'arrodoniment cap a alguna direcció fixa, seria preferible arrodonir de manera que l'última xifra de la mantissa sigui, per exemple, sempre parella).

Les fites de l'error relatiu produït, en cada cas, són:

$$\epsilon_T = b^{1-t} , \quad \epsilon_A = \frac{1}{2} b^{1-t} .$$

EXEMPLES

1. -99.962 (o $-0.99962 \cdot 10^2$) tallat a 3 xifres queda -99.9 (o també, $-0.999 \cdot 10^2$) i, arrodonit a 3 xifres, -100 . (o $-0.100 \cdot 10^3$).
2. 35.47846 arrodonit a 6 xifres queda 35.4785 , i 35.4785 arrodonit a 5 xifres esdevé 35.479 . Notem, però, que arrodonint (directament) 35.47846 a 5 xifres queda 35.478 .

3. El nombre π és 3.141592653589793238462643... Si π entra com a dada en una màquina que sols admet 10 xifres en base 10, serà representat per

$$\begin{aligned}\text{fl}_T(\pi) &= 0.3141592653 \cdot 10^1 \text{ (si talla) ,} \\ \text{fl}_A(\pi) &= 0.3141592654 \cdot 10^1 \text{ (si arrodoneix) ;}\end{aligned}$$

on fl_T indica la *representació en punt flotant de x per tall* i fl_A , la *representació en punt flotant de x per arrodoniment*. Observeu que l'error de la primera representació és més gran que el de la segona.

Error d'arrodoniment durant el càlcul

L'error en un resultat no solament pot provenir dels errors de les dades d'entrada, sinó també dels errors d'arrodoniment en els resultats dels càlculs intermedis.

EXEMPLE

Considerem una màquina que treballa amb 4 xifres decimals i talla o arrodoneix. Hi tenim dues dades:

$$a = 0.3425 \cdot 10^5, \quad b = 0.2517 \cdot 10^{-2}.$$

Aleshores,

$$\begin{aligned}ab &= 0.8620725 \cdot 10^2, \\ \text{fl}_T(ab) &= 0.8620 \cdot 10^2 \text{ (s'ha comès un error de } 0.725 \cdot 10^2), \\ \text{fl}_A(ab) &= 0.8621 \cdot 10^2 \text{ (s'ha comès un error de } 0.275 \cdot 10^2).\end{aligned}$$

Si no es diu el contrari, suposarem que, en cada *operació aritmètica* (+, −, ·, /), les fites d'error relatiu seran les mateixes que en la representació de les dades de partida: ϵ_T o ϵ_A .

Error de truncament del mètode emprat

Quan resollem un problema matemàtic per mètodes numèrics, tot i fent les operacions exactament, obtenim només una *aproximació numèrica* del resultat exacte (per exemple, quan aproximem una integral per una suma finita o una derivada per un quocient incremental, etc).

L'error produït depèn del mètode numèric emprat i rep el nom d'*error de truncament*. Per a alguns mètodes, disposem d'expressions d'aquest error, com veurem en els capítols següents.

En decidir sobre la conveniència de la utilització d'un mètode determinat, s'ha de tenir en compte no solament el seu error de truncament, sinó també els errors d'arrodoniment produïts per les operacions que el mètode comporta.

Per exemple, cal sempre defugir els mètodes o algorismes que comportin *cancel·lacions* de xifres en restar dues quantitats pròximes. Atès que en aquestes cancel·lacions es produeixen errors relatius considerablement grans, convé usar fórmules matemàticament equivalents que les evitin. Per copsar la importància d'aquest fet, vegeu-ne diversos exemples en els problemes I.8-I.11.

1.2 ESTIMACIÓ I FITACIÓ D'ERRORS

L'objectiu de qualsevol estudi d'errors és tractar de conèixer l'efecte que, sobre el resultat final d'un problema numèric, produeix cada un dels diferents tipus d'error que hi poden tenir lloc.

Distingirem tres tipus bàsics d'error: els de les dades, els dels càlculs intermedis i els errors de truncament del mètode numèric emprat. L'error total sobre el resultat final serà la suma de les contribucions dels tres tipus d'error.

1.2.1 Propagació dels errors de les dades

Per als problemes numèrics de fer operacions aritmètiques (+, −, ·, /) amb dues dades x_1 i x_2 afectades d'error, tenim les següents fites de l'error propagat:

$$\begin{aligned}\epsilon_a(x_1 + x_2) &= \epsilon_a(x_1) + \epsilon_a(x_2) , \\ \epsilon_a(x_1 - x_2) &= \epsilon_a(x_1) + \epsilon_a(x_2) ;\end{aligned}$$

i les següents fites aproximades, si els errors de x_1 i x_2 són petits:

$$\begin{aligned}\epsilon_r(x_1 x_2) &\simeq \epsilon_r(x_1) + \epsilon_r(x_2) , \\ \epsilon_r(x_1/x_2) &\simeq \epsilon_r(x_1) + \epsilon_r(x_2) .\end{aligned}$$

Per a un problema numèric consistent a calcular el resultat $y = f(x)$ a partir de només una dada x , obtenim la següent *fórmula aproximada de propagació de l'error*

$$e_a(y) \simeq f'(x) \cdot e_a(x) , \quad (1.1)$$

com a conseqüència directa del teorema del valor mitjà, per a funcions f d'una variable, derivables amb continuïtat. D'aquesta es dedueix una fita aproximada per a l'error absolut de y , en funció d'una fita de l'error absolut de x , donant lloc a la *fórmula aproximada de propagació de l'error maximal*

$$\epsilon_a(y) \simeq |f'(x)| \epsilon_a(x) . \quad (1.2)$$

Finalment, per al problema numèric més general, que consisteix a calcular un resultat $y = f(x_1, \dots, x_n)$ a partir d'unes dades $x_1, \dots, x_n$, disposem de la *fórmula aproximada de propagació de l'error*

$$e_a(y) \simeq \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) e_a(x_i) ; \quad (1.3)$$

d'on, conegudes fites de $e_a(x_i)$, podem fitar $e_a(y)$, obtenint així la *fórmula aproximada de propagació de l'error maximal*

$$\epsilon_a(y) \simeq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(x_1, \dots, x_n) \right| \epsilon_a(x_i) ; \quad (1.4)$$

especialment adequada quan n , el nombre de dades afectades d'error, no és gran.

EXAMPLE

Si calculem la suma $y = x_1 + \dots + x_n$ de n dades $x_1, \dots, x_n$, aleshores

$$e_a(y) \simeq \sum_{i=1}^n e_a(x_i) .$$

Si cada variable té una fita d'error ϵ , la fórmula de l'error maximal ens dóna $\epsilon_a(y) \simeq n\epsilon$, fita que només pot ser assolida quan tots els errors $e_a(x_i)$ tinguin el mateix signe i la màxima magnitud, cosa altament improbable quan n és gran.

1.2.2 Propagació dels errors en els càlculs

La propagació dels errors en els càlculs és estudiada en dues fases (per a diversos exemples, consulteu els problemes resolts I.14-I.18).

Anàlisi de l'error cap al darrera

Partint de dades inicials exactes, per culpa de l'acumulació dels errors en les operacions, obtenim un resultat afectat d'error. La idea bàsica de l'*anàlisi de l'error cap al darrera* consisteix a estudiar les modificacions que hauríem de fer sobre les dades d'entrada, de forma que, suposant que no hi haguessin errors en les operacions, s'obtingués el mateix error en el resultat.

Aquest estudi es fa emprant successivament la fórmula

$$\text{fl}(a * b) = (a * b)(1 + \delta_*) ,$$

amb $|\delta_*| \leq \epsilon_*$, a cada una de les operacions aritmètiques $*$ ($+$, $-$, $\cdot$, $/$) que componen el procés de càlcul, on ϵ_* indica una fita coneguda de l'error relatiu en l'operació $*$; a més, per a totes les funcions g que intervenen en els càlculs, s'escriu

$$\text{fl}(g(x)) = g(x)(1 + \delta_g) ,$$

amb $|\delta_g| \leq \epsilon_g$, on ϵ_g indica una fita coneguda de l'error relatiu en l'avaluació de g .

A continuació, s'escriu una expressió del resultat final que permeti imputar els errors dels càlculs a les dades. Amb aquest procediment es redueix l'anàlisi de l'error en els càlculs a una anàlisi de propagació d'errors en les dades sense errors en els càlculs.

Propagació dels errors imputats a les dades

Un cop feta la reducció anterior, s'aplica la fórmula de propagació de l'error maximal a les fites dels errors imputats a les dades, considerant que els càlculs ja es fan sense error. En el quadre de la figura 1.1 es mostra gràficament tot el procés.

1.2.3 Errors de truncament

L'error de truncament depèn de cada mètode numèric i el seu estudi es farà de manera específica per als diferents mètodes que es vagin presentant en els capítols que segueixen. De fet, l'estimació de l'error de truncament associat a cada mètode conforma una part fonamental de la seva exposició.

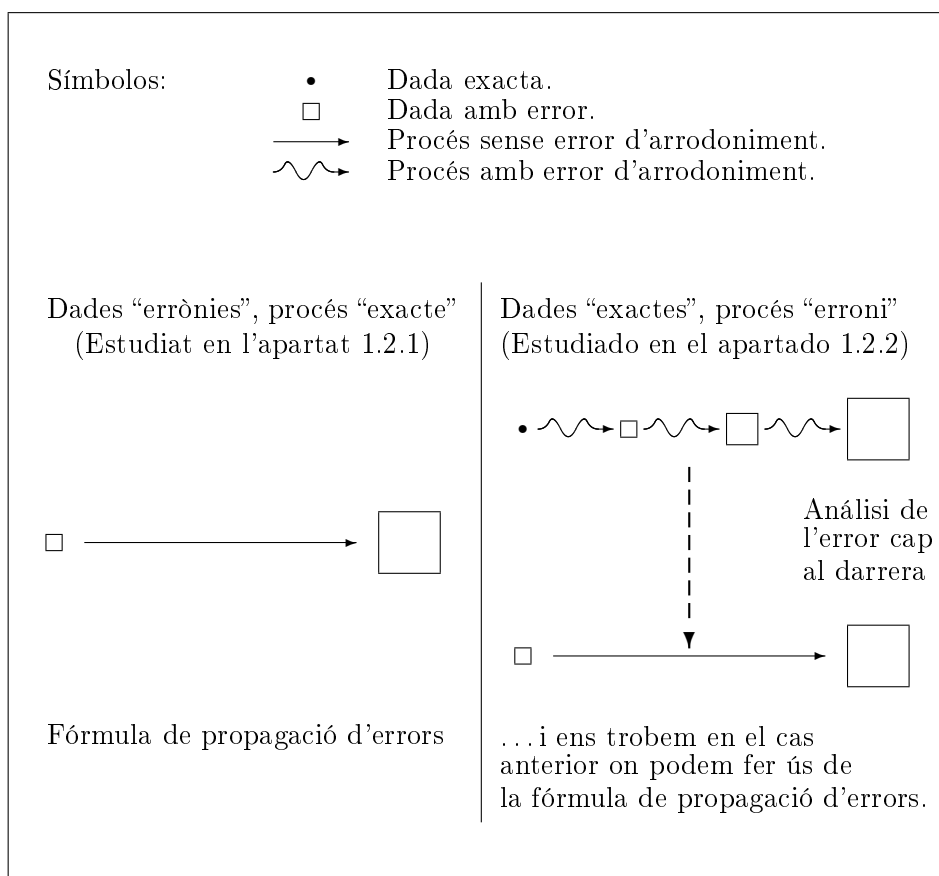


Figura 1.1: Quadre resum del tractament d'errors.

1.3 TRACTAMENT INTERVALAR I ESTADÍSTIC

La fórmula de propagació d'errors (1.3) és només una fórmula aproximada (de primer ordre) per a la fitació de l'error. El càlcul de l'error usant intervals permet d'obtenir-ne fites rigoroses; malauradament, aquestes fites molt sovint no són realistes quan el nombre de càlculs és gran; això és, són fites desproporcionades i poc probables dels errors reals. Per a estimacions més apropiades a la realitat convé molts cops analitzar l'error des d'un punt de vista estadístic. A continuació es presenten aquests dos vessants.

1.3.1 Anàlisi d'interval

L'anàlisi d'interval fa ús de les operacions entre intervals a fi de determinar un interval en què, amb tota seguretat, es troba el resultat final.

Un *interval (tancat i fitat)* I en $\mathbb{R}$ és un conjunt I de la forma $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$, amb $a, b \in \mathbb{R}$. Anomenarem $I(\mathbb{R})$ al conjunt dels intervals en $\mathbb{R}$.

Es defineixen a continuació les operacions amb intervals més senzilles. Definim primer les *operacions aritmètiques entre intervals*: si $I_1, I_2 \in I(\mathbb{R})$,

$$I_1 * I_2 = \{y \in \mathbb{R} : y = x_1 * x_2, x_1 \in I_1, x_2 \in I_2\},$$

per a qualsevol operació aritmètica $*$ entre nombres reals.

Per a un gran nombre de funcions entre nombres reals (per exemple, les contínues) poden definir les *funcions entre intervals*; així, donada la funció $f : \mathbb{R} \rightarrow \mathbb{R}$, definim

$$f : I(\mathbb{R}) \rightarrow I(\mathbb{R}) , \quad f(I) = \{y : y = f(x), x \in I\} .$$

Emprant operacions i funcions entre intervals, en comptes d'operacions i funcions entre nombres, l'anàlisi d'interval permet de fer un seguiment, en el decurs de la cadena de càlculs, de la regió de la recta real en què es troben els successius resultats parcials i, en darrer terme, el resultat final del procés de càlcul.

1.3.2 Anàlisi estadística

Per tal de donar unes estimacions més realistes de l'error del resultat, quan aquest s'obté a través d'un procés numèric que comporta moltes operacions, és convenient introduir un *tractament estadístic de l'error*.

Per això, se suposa que els errors en les dades d'entrada són *variables aleatòries independents* amb una certa *funció distribució* donada, com es mostra en els exemples següents.

EXEMPLES

a) Sigui e l'error en arrodonir un nombre a d xifres decimals. Aleshores e pren valors no nuls només sobre $[-\epsilon, \epsilon]$, on $\epsilon = \frac{1}{2}10^{-d}$; si suposem que cada valor sobre aquest interval és igualment probable, la *funció densitat* ρ d' e correspon a una *distribució uniforme* (és a dir, una funció que val $\frac{1}{2\epsilon}$ sobre $[-\epsilon, \epsilon]$ i és nul·la fora d'aquest interval); la *funció distribució*

$$F(x) = \int_{-\infty}^x \rho(t) dt$$

que ens dona la probabilitat que e prengui valors més petits o iguals que x , és una recta de pendent $\frac{1}{2\epsilon}$ en aquell interval. En la figura 1.2 es veuen representades ambdues funcions.

Donem a continuació diferents magnituds estadístiques de la variable aleatòria e :

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x\rho(x)dx = 0 \quad (\text{mitjana d}'e) , \\ \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 \rho(x)dx = \epsilon^2/3 \quad (\text{variància d}'e) , \\ \sigma &= \frac{\epsilon}{\sqrt{3}} = \frac{10^{-d}}{\sqrt{12}} \simeq 0.2887 \cdot 10^{-d} \quad (\text{desviació estàndard d}'e) . \end{aligned}$$

b) La majoria de les variables aleatòries e que hom troba a la pràctica tenen aproximadament una *distribució normal*; això és, tenen una funció densitat de probabilitat aproximadament igual a

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} ,$$

on μ és la mitjana i σ^2 la variància. En la figura 1.3 se'n troba una representació gràfica.

La probabilitat $P(\delta)$ que e prengui valors entre $\mu - \delta$ i $\mu + \delta$ ve donada per

$$P(\delta) = \int_{\mu-\delta}^{\mu+\delta} \rho(t)dt = \frac{2}{\sqrt{\pi}} \int_0^{\frac{\delta}{\sqrt{2}\sigma}} e^{-t^2} dt = \operatorname{erf}\left(\frac{\delta}{\sqrt{2}\sigma}\right) ,$$

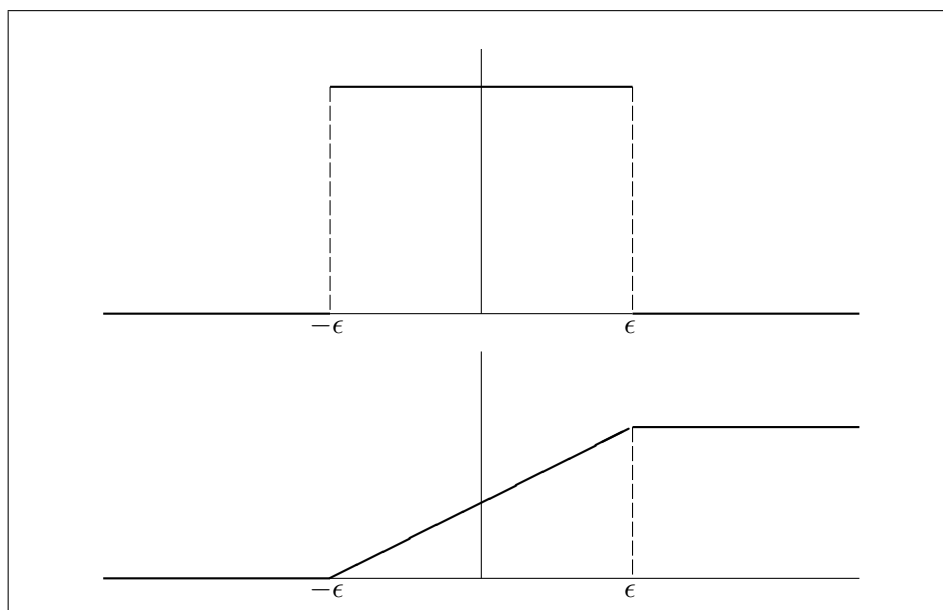


Figura 1.2: Funcions densitat i distribució de la distribució uniforme.

on

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

rep el nom de *funció d'error*.

Com a exemples, $Q(\delta) = 1 - P(\delta)$ val 0.317, 0.045 i 0.0027 quan δ val σ , 2σ i 3σ , respectivament. Això vol dir que, si e representa un error d'una dada amb mitjana $\mu = 0$ (per exemple) i desviació típica σ , l'afirmació que la magnitud de l'error sigui menor que σ , 2σ i 3σ és falsa tan sols en aproximadament el 32%, 5% i 0.3% dels casos, respectivament.

Retornant a la teoria general, el tractament estadístic dels errors es basa en els resultats següents:

Siguin $e_1, \dots, e_n$ variables aleatòries amb mitjanes $\mu_1, \dots, \mu_n$ i desviacions típiques $\sigma_1, \dots, \sigma_n$, i $d_1, \dots, d_n$ constants arbitràries; llavors, es compleixen les propietats següents:

1.

$$e = \sum_{i=1}^n d_i e_i$$

és una variable aleatòria amb mitjana

$$\mu = \sum_{i=1}^n d_i \mu_i .$$

2. Si $e_1, \dots, e_n$ són *independents* (és a dir, el valor que pren una variable no condiciona el valor de cap altra), e té per desviació típica σ , amb

$$\sigma^2 = \sum_{i=1}^n d_i^2 \sigma_i^2 .$$

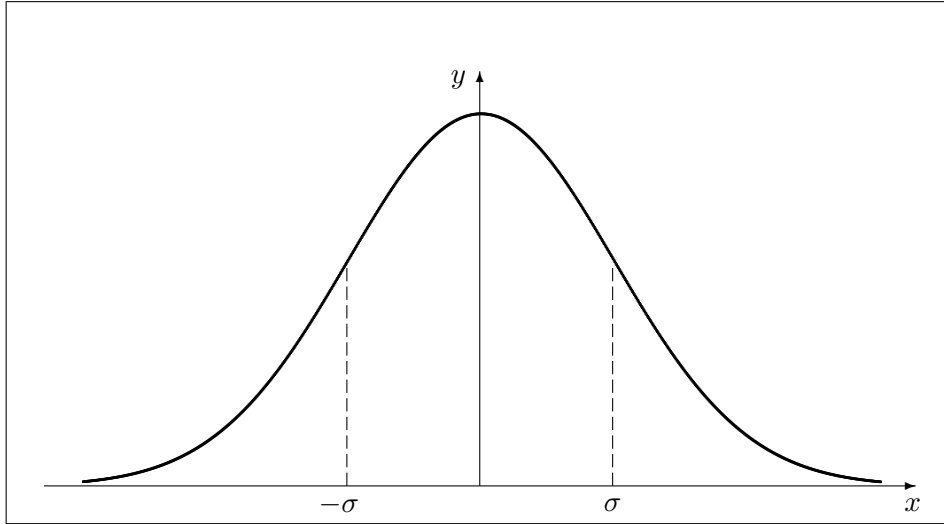


Figura 1.3: Funció densitat de la distribució normal.

Si, a més, $e_1, \dots, e_n$ són normals; també ho és e .

3. Si $e_1, \dots, e_n$ són independents i tenen la mateixa funció de distribució, la funció distribució de $\frac{e}{n}$ s'aproxima a una funció de distribució normal (quan $n \rightarrow \infty$) (*Teorema central del límit*).

Si volem calcular $y = f(x_1, \dots, x_n)$ i suposem que $e_a(x_i) = x_i - \bar{x}_i$ són variables aleatòries independents amb mitjana nul·la i desviació típica $\sigma(e_a(x_i))$, que escriurem simplement $\sigma(x_i)$; $e_a(y) = y - \bar{y}$ ve donada aleshores per la fórmula aproximada de propagació de l'error (1.3). Per tant, tenint en compte les propietats anteriors, és una variable aleatòria amb mitjana nul·la i desviació estàndard $\sigma(e_a(y))$, que també escriurem simplement $\sigma(y)$, donada per la *fórmula aproximada de propagació de l'error estàndard*

$$\sigma(y) \simeq \left(\sum_{i=1}^n \left[\frac{\partial f}{\partial x_i}(x_1, \dots, x_n) \right]^2 \sigma(x_i)^2 \right)^{1/2}. \quad (1.5)$$

Notem que l'aproximació de $e_a(y)$ donada en la fórmula (1.3) té una distribució normal, si les variables $e_a(x_i)$ la tenen, i aproximadament normal si aquestes tenen la mateixa funció distribució i n és gran.

En l'exemple de la suma $y = x_1 + x_2 + \dots + x_n$, tractant els errors d'arrodoniment a t xifres decimals $e_a(x_i)$ com a variables aleatòries independents amb distribució uniforme de mitjana nul·la i desviació típica $\epsilon = \frac{1}{2}10^{-t}$, la fórmula anterior dóna la desviació típica següent per a y

$$\sigma(y) \simeq \sqrt{n}\sigma = \sqrt{\frac{n}{3}}\epsilon.$$

Pel teorema central del límit, $e_a(y)$ té distribució aproximadament normal amb mitjana nul·la, si n és gran. Així, es pot dir que, per exemple, la magnitud de $e_a(y)$ és menor que

$$2\sqrt{\frac{n}{3}}\epsilon,$$

en aproximadament el 95% dels casos en què es facin sumes de n variables independents.

COMENTARIS BIBLIOGRÀFICS

La majoria de llibres de càlcul numèric porten algun tipus d'introducció a l'estudi d'errors. Per a un punt de vista similar al d'aquest capítol, remetem al lector a [DB74], [RR78], [SB80], [YG72].

Unes bones introduccions a la representació dels nombres es troben en [Hen64], [Ham73], [Knu69], i principalment en [Wil64]. El teorema central del límit es un dels més importants de la teoria de probabilitats, teoria que ultrapassa el contingut d'aquest llibre i es troba en els tractats clàssics [Cra46], [Fel50]. La referència bàsica sobre l'anàlisi d'interval és [Moo66].

PROBLEMES RESOLTS

Problema 1.1 *Convertiu*

- a) 26.32_{10} a base 2,
 b) 1314.96_{10} a base 16,
 c) $AF.3C_{16}$ a base 10.

SOLUCIÓ:

Estudiem primer com canvien les xifres d'un nombre r en canviar-lo de la base b a la base B .

Si $r_b = a_n \dots a_0.a_{-1} \dots$ i $r_B = A_N \dots A_0.A_{-1} \dots$, tenim

$$\sum_{i=-\infty}^n a_i b^i = \sum_{j=-\infty}^N A_j B^j.$$

Igualant la part entera d'ambdós membres

$$\sum_{i=0}^n a_i b^i = \sum_{j=0}^N A_j B^j$$

i dividint-los per B , tenim que A_0 és el residu de la divisió de la part entera de r per B . Prenent ara el quocient d'aquesta divisió i repetint la divisió per B , obtenim A_1 com a nou residu; la repetició d'aquest procés permet de conèixer totes les xifres de la part entera de r en la base B : $A_0, A_1, \dots, A_N$.

Anàlogament, igualem ambdues parts fraccionàries

$$\sum_{i=-\infty}^{-1} a_i b^i = \sum_{j=-\infty}^{-1} A_j B^j,$$

i, si ara les multipliquem per B , obtenim com a part entera A_{-1} . Repetint-ho de nou, amb la part fraccionària del producte, tenim com a nova part entera A_{-2} i, així successivament, trobaríem totes les xifres de la part fraccionària de r en la base B : $A_{-1}, A_{-2}, \dots$

Per als casos demanats:

a)

$$26 = 13 \cdot 2 + \underline{0} \quad 13 = 6 \cdot 2 + \underline{1} \quad 6 = 3 \cdot 2 + \underline{0} \quad 3 = \underline{1} \cdot 2 + \underline{1} \quad .$$

Així, $26_{10} = 11010_2$.

$$\begin{array}{llll} 0.32 \cdot 2 = \underline{0.64} & 0.64 \cdot 2 = \underline{1.28} & 0.28 \cdot 2 = \underline{0.56} & 0.56 \cdot 2 = \underline{1.12} \\ 0.12 \cdot 2 = \underline{0.24} & 0.24 \cdot 2 = \underline{0.48} & 0.48 \cdot 2 = \underline{0.96} & 0.96 \cdot 2 = \underline{1.92} \\ 0.92 \cdot 2 = \underline{1.84} & 0.84 \cdot 2 = \underline{1.68} & 0.68 \cdot 2 = \underline{1.36} & 0.36 \cdot 2 = \underline{0.72} \\ 0.72 \cdot 2 = \underline{1.44} & 0.44 \cdot 2 = \underline{0.88} & 0.88 \cdot 2 = \underline{1.76} & 0.76 \cdot 2 = \underline{1.52} \\ 0.52 \cdot 2 = \underline{1.04} & 0.04 \cdot 2 = \underline{0.08} & 0.08 \cdot 2 = \underline{0.16} & 0.16 \cdot 2 = \underline{0.32} \\ \dots & & & \end{array}$$

Per tant, l'expressió demanada és

$$\boxed{26.32_{10} = 11010.\overline{01010001111010111000}_2} ,$$

on les línies per sobre indicaran a partir d'ara el període del nombre.

b)

$$1314 = 82 \cdot 16 + \underline{2} \quad 82 = \underline{5} \cdot 16 + \underline{2} \quad .$$

Tenim doncs, $1314_{10} = 522_{16}$.

$$\begin{array}{lll} 0.96 \cdot 16 = \underline{15}.36 & 0.36 \cdot 16 = \underline{5}.76 & 0.76 \cdot 16 = \underline{12}.16 \\ 0.16 \cdot 16 = \underline{2}.56 & 0.56 \cdot 16 = \underline{8}.96 & \dots \end{array} .$$

Resulta doncs,

$$\boxed{1314.96_{10} = 522.\overline{F5C28}_{16}} ,$$

emprant com a dígit en base 16: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F. Aquesta notació és molt comuna i la usarem d'ara endavant.

c) Simplement cal fer

$$AF.3C_{16} = 10 \cdot 16 + 15 + 3 \cdot 16^{-1} + 12 \cdot 16^{-2} .$$

Finalment,

$$\boxed{AF.3C_{16} = 175.234375_{10}} .$$

Problema 1.2 a) *Convertiu els nombres següents a base 10:*

$$100010001.100010001_2 , \quad 42356.66_7 , \quad 1F2D.124_{16} .$$

b) *Convertiu els nombres següents, en base 10, a les bases 2, 7 i 16:*

$$4235.66 , \quad \pi , \quad e .$$

c) *Feu les operacions següents:*

$$10001.1001_2 + 1001.1111_2 , \quad 11.1101_2 \cdot 11_2 , \quad 45.67_9 \cdot 44_9 ,$$

$$11.11_2 / 0.1111_2 , \quad 1221.2112_3 / 22_3 , \quad A1_{16} / F_{16} .$$

SOLUCIÓ:

a) De la mateixa definició de la representació digital d'un nombre en una base qualsevol, tenim

$$\begin{aligned}
 100010001.100010001_2 &= 1 \cdot 2^8 + 0 \cdot 2^7 + 0 \cdot 2^6 + 0 \cdot 2^5 \\
 &\quad + 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 \\
 &\quad + 1 \cdot 2^0 \\
 &\quad + 1 \cdot 2^{-1} + 0 \cdot 2^{-2} + 0 \cdot 2^{-3} + 0 \cdot 2^{-4} \\
 &\quad + 1 \cdot 2^{-5} + 0 \cdot 2^{-6} + 0 \cdot 2^{-7} + 0 \cdot 2^{-8} \\
 &\quad + 1 \cdot 2^{-9} \\
 &= 2^8 + 2^4 + 2^0 + 2^{-1} + 2^{-5} + 2^{-9} \\
 &= 256 + 16 + 1 + 0.5 + 0.03125 + 0.001953125 \\
 &= \underline{273.533203125} ,
 \end{aligned}$$

$$\boxed{100010001.100010001_2 = 273.533203125} .$$

La part entera de 42356.66_7 és

$$\begin{aligned}
 42356_7 &= 4 \cdot 7^4 + 2 \cdot 7^3 + 3 \cdot 7^2 + 5 \cdot 7^1 + 6 \\
 &= (((4 \cdot 7 + 2) \cdot 7 + 3) \cdot 7 + 5) \cdot 7 + 6 \\
 &= 10478 ;
 \end{aligned}$$

i la part fraccionària,

$$\begin{aligned}
 0.66_7 &= 6 \cdot 7^{-1} + 6 \cdot 7^{-2} = \frac{6 \cdot 7 + 6}{7^2} = \frac{48}{49} \\
 &= 0.\overline{979591836734693877551020408163265306122448} .
 \end{aligned}$$

Finalment,

$$\boxed{42356.66_7 = 10478.\overline{979591836734693877551020408163265306122448}} .$$

Recordem que les xifres en base 16 són: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A(=10), B(=11), C(=12), D(=13), E(=14) i F(=15). Així, de forma anàloga,

$$\boxed{1F2D.124_{16} = 7981.0712890625} .$$

b) Anàlogament al problema anterior,

$$\begin{array}{lll}
 4235 = 2117 \cdot 2 + \underline{1} & 2117 = 1058 \cdot 2 + \underline{1} & 1058 = 529 \cdot 2 + \underline{0} \\
 529 = 264 \cdot 2 + \underline{1} & 264 = 132 \cdot 2 + \underline{0} & 132 = 66 \cdot 2 + \underline{0} \\
 66 = 33 \cdot 2 + \underline{0} & 33 = 16 \cdot 2 + \underline{1} & 16 = 8 \cdot 2 + \underline{0} \\
 8 = 4 \cdot 2 + \underline{0} & 4 = 2 \cdot 2 + \underline{0} & 2 = \underline{1} \cdot 2 + \underline{0}
 \end{array}$$

$$\begin{array}{lll}
0.66 \cdot 2 = \underline{1}.32 & 0.32 \cdot 2 = \underline{0}.64 & 0.64 \cdot 2 = \underline{1}.28 \\
0.28 \cdot 2 = \underline{0}.56 & 0.56 \cdot 2 = \underline{1}.12 & 1.12 \cdot 2 = \underline{0}.24 \\
0.24 \cdot 2 = \underline{0}.48 & 0.48 \cdot 2 = \underline{0}.96 & 0.96 \cdot 2 = \underline{1}.92 \\
0.92 \cdot 2 = \underline{1}.84 & 0.84 \cdot 2 = \underline{1}.68 & 0.68 \cdot 2 = \underline{1}.36 \\
0.36 \cdot 2 = \underline{0}.72 & 0.72 \cdot 2 = \underline{1}.44 & 0.44 \cdot 2 = \underline{0}.88 \\
0.88 \cdot 2 = \underline{1}.76 & 0.76 \cdot 2 = \underline{1}.52 & 0.52 \cdot 2 = \underline{1}.04 \\
0.04 \cdot 2 = \underline{0}.08 & 0.08 \cdot 2 = \underline{0}.16 & 0.16 \cdot 2 = \underline{0}.32 \\
& \dots &
\end{array}$$

Resulta doncs,

$$\boxed{4235.66 = 1000010001011.101010001111010111000_2} \quad .$$

$$\begin{array}{lll}
4235 = 605 \cdot 7 + \underline{0} & 605 = 86 \cdot 7 + \underline{3} & 86 = 12 \cdot 7 + \underline{2} \\
12 = \underline{1} \cdot 7 + \underline{5} & &
\end{array} \quad ,$$

$$\begin{array}{lll}
0.66 \cdot 7 = \underline{4}.62 & 0.62 \cdot 7 = \underline{4}.34 & 0.34 \cdot 7 = \underline{2}.38 \\
0.38 \cdot 7 = \underline{2}.66 & \dots &
\end{array} \quad .$$

Per tant,

$$\boxed{4235.66 = 15230.\overline{4422}_7} \quad .$$

Agrupant de quatre en quatre les xifres binàries, s'obté

$$\boxed{4235.66 = 108B.A8\overline{F5C2}_{16}} \quad .$$

Les 25 primeres xifres significatives dels nombres π i e són:

$$\begin{array}{ll}
\pi & = 3.141592653589793238462643 \dots , \\
e & = 2.718281828459045235360287 \dots ,
\end{array}$$

respectivament.

Considerem el pas del nombre π a base 7. Anomenant $\pi^{(j)}$ al nombre π arrodonit a j xifres decimals trobem

$$\begin{aligned}
\pi^{(4)} &= 3.1416 = 3.0664_7 \pm \frac{1}{2}7^{-4} , \\
\pi^{(5)} &= 3.14159 = 3.06637_7 \pm \frac{1}{2}7^{-5} ;
\end{aligned}$$

i, per tant,

$$\begin{aligned}
\pi &= 3.0664_7 \pm \left(\frac{1}{2}7^{-4} + \frac{1}{2}10^{-4} \right) , \\
\pi &= 3.06637_7 \pm \left(\frac{1}{2}7^{-5} + \frac{1}{2}10^{-5} \right) .
\end{aligned}$$

Tindrem també, per exemple, que

$$\boxed{\pi = 3.141593 \pm \frac{1}{2}10^{-6} = 3.0663652_7 \pm 7^{-7}} ,$$

$$\boxed{\pi = 3.141592654 \pm \frac{1}{2}10^{-9} = 3.0663651432_7 \pm 7^{-10}} .$$

La resta de l'apartat b) es deixa per als lectors més pacients.

c) Realitzem les operacions aritmètiques demanades de forma anàloga a com es fan en base 10:

$$\begin{array}{r} 10001.1001_2 \\ 1001.1111_2 \\ \hline 11011.1000_2 \end{array}$$

$$\boxed{10001.1001_2 + 1001.1111_2 = 11011.1000_2} .$$

$$\begin{array}{r} 111101_2 \\ 11_2 \\ \hline 111101 \\ 111101 \\ \hline 10110111_2 \end{array}$$

$$\boxed{11.1101_2 \cdot 11_2 = 1011.0111_2} .$$

$$\begin{array}{r} 4567_9 \\ 44_9 \\ \hline 20501 \\ 20501 \\ \hline 225511_9 \end{array}$$

$$\boxed{45.67_9 \cdot 44_9 = 2255.11_9} .$$

$$\boxed{11.11_2 / 0.1111_2 = 100_2} .$$

$$\begin{array}{r} 12212112_3 \quad | \quad 22_3 \\ 00112 \quad \quad 20.1211_3 \\ 0201 \\ 0101 \\ 0022 \\ 00 \end{array}$$

$$\boxed{1221.2112_3 / 22_3 = 20.1211_3} .$$

$$\begin{array}{r} A1_{16} \quad | \quad F_{16} \\ 0B0 \quad \quad A.B \dots_{16} \\ 0B0 \end{array}$$

...

$$\boxed{A.1_{16} / F_{16} = A.\overline{B}_{16}} .$$

Problema 1.3 En un ordinador IBM, cada cel·la de memòria està formada per 32 posicions binàries (bits). Sabem que tot nombre p s'emmagatzema en punt flotant hexadecimal (en base 16) de la forma següent: un cop escrit p en la forma $\pm p' \cdot 16^{(p''-64)}$ amb $p'' \geq 0$, $\frac{1}{16} \leq p' < 1$, s'emmagatzemen en els 32 bits i successivament: el signe (0 si +, 1 si -) (1 bit), l'exponent modificat p'' en base 2 (7 bits) i les primeres xifres no enteres de la mantissa p' en base 2 (24 bits, en precisió simple, i 56 bits, en precisió doble). Notem que un nombre emmagatzemat en precisió doble ocupa 2 cel·les de memòria.

a) Indiqueu com s'emmagatzemen en precisió simple 1.0, 1.1 i -17.0 .

b) Expressen, en forma decimal, els nombres positius més gran i més petit que poden ser emmagatzemats.

c) Calculeu l'error relatiu màxim, en notació decimal, amb què pot ser emmagatzemat un nombre qualsevol en precisió simple i doble.

SOLUCIÓ:

a) El nombre $p = 1$ en la representació de punt flotant hexadecimal queda

$$1_{10} = 1_2 = \frac{1}{16} \cdot 16^1 = 2^{-4} \cdot 16^{65-64} ;$$

així,

$$p' = \frac{1}{16} = 2^{-4} = 0.0001_2, \quad p'' = 65 = 1000001_2,$$

i s'emmagatzema

$$\boxed{0 \mid 1000001 \mid 000100000000000000000000} .$$

Per a $p = 1.1$, tenim

$$1.1_{10} = 1.0001\overline{1}_2, \quad 1.1 = \frac{1.1}{16} \cdot 16^1 = 0.00010001\overline{1}_2 \cdot 16^{65-64},$$

i s'emmagatzema

$$\boxed{0 \mid 1000001 \mid 000100011001100110011010} ,$$

arrodonint adequadament.

Per a $p = -17$:

$$-17_{10} = -10001_2, \quad -17 = \frac{-17}{16^2} \cdot 16^2 = -0.00010001_2 \cdot 16^{66-64} .$$

Així, -17 es representa per

$$\boxed{1 \mid 1000010 \mid 000100010000000000000000} .$$

b) La representació del nombre positiu més gran correspon a

$$\boxed{0 \mid 1111111 \mid 111111111111111111111111} ;$$

es tracta, doncs, de

$$(1 - 2^{-24}) \cdot 16^{127-64} \simeq 16^{63} \simeq 7 \cdot 10^{75} .$$

La representació del més petit és

$$\boxed{0 \mid 0000000 \mid 000100000000000000000000} ,$$

atès que $p' \geq \frac{1}{16}$. Per tant, aquest nombre és

$$\frac{1}{16} \cdot 16^{-64} = 16^{-65} \simeq 6 \cdot 10^{-79} .$$

c) Suposant que s'arrodoneix en fer la representació finita de la mantissa de p , una fita de l'error absolut en aquesta representació és $\frac{1}{2}2^{-24} \cdot 16^{p''-64}$, en precisió simple, i $\frac{1}{2}2^{-56} \cdot 16^{p''-64}$, en precisió doble.

Com que $p'' \geq \frac{1}{16}$, tenim les fites d'error relatiu següents:

$$\epsilon_r(p) = \frac{2^{-25} \cdot 16^{p''-64}}{16^{-1} \cdot 16^{p''-64}} = 2^{-21} \simeq 0.5 \cdot 10^{-6} ,$$

en precisió simple, i

$$\epsilon_r(p) = \frac{2^{-57} \cdot 16^{p''-64}}{16^{-1} \cdot 16^{p''-64}} = 2^{-53} \simeq 1.1 \cdot 10^{-16} ,$$

en precisió doble.

Problema 1.4 En l'ordinador CDC 3200, els nombres s'emmagatzemen en punt flotant, emprant 48 posicions binàries de la manera següent:

$$\boxed{\begin{array}{c|cc} 1 & 11 & 36 \\ \hline s & \text{exponent} & \text{mantissa} \end{array}} .$$

El nombre p s'escriu en la forma $p = \pm p' \cdot 2^{p''}$ amb $\frac{1}{2} \leq p' < 1$. La mantissa p' s'emmagatzema aproximada per les seves 36 primeres xifres binàries no enteres, arrodonides adequadament. L'exponent p'' ha de complir la fita $|p''| < 2^{10}$ i es representa per

$$p'' + 2^{10} \text{ (si } p'' \geq 0) , \quad p'' + 2^{10} - 1 \text{ (si } p'' < 0) .$$

El bit corresponent al signe (s) és 0 quan $p \geq 0$; si $p < 0$, es pren el complement a 1 de tota la representació anterior (això és, es canvien el uns per zeros i els zeros per uns; en particular, el bit del signe serà 1).

a) Indiqueu com es representen els nombres següents:

$$1.0 , \quad 0.5 , \quad -0.0625 , \quad 0.1 .$$

- Doneu (en notació decimal) els nombres positius més gran i més petit que es poden emmagatzemar.
- Quants nombres reals es poden emmagatzemar exactament?
- Quin és l'error relatiu maximal en l'emmagatzemament dels nombres?

SOLUCIÓ:

- a) El nombre $p = 1$ en la representació de punt flotant binària queda

$$1_{10} = 1_2 = \frac{1}{2} \cdot 2^1 ;$$

així,

$$p' = \frac{1}{2} = 2^{-1} = 0.1_2 \text{ , } p'' = 1 = 1_2 \text{ ,}$$

i s'emmagatzema

$$\boxed{0 \mid 10000000001 \mid 10000000000000000000000000000000}.$$

Per a $p = 0.5$, tenim

$$0.5 = \frac{1}{2} \cdot 2^0$$

i s'emmagatzema

$$0 \mid 10000000000 \mid 1000000000000000000000000000000000$$

Per a $p = -0.0625$, considerem primer el seu valor absolut

$$0.0625_{10} = \frac{1}{16} = 0.1_2 \cdot 2^{-3} .$$

Així, 0.0625 es representa per

$$\boxed{0 \mid 0111111100 \mid 10000000000000000000000000000000} ;$$

per tant, $p = -0.0625$ tindrà la representació complementària a 1

$$1 \mid 10000000011 \mid 01111111111111111111111111111111 .$$

Finalment, per a $p = 0.1$, tenim

$$0.1 = 0.8 \cdot 2^{-3} = 0.\overline{1100}_2 \cdot 2^{-3}$$

i s'emmagatzema

$$\boxed{0 \mid 0111111100 \mid 11001100110011001100110011001101},$$

arrodonint convenientment.

- b) La representació del nombre positiu més gran correspon a

b) Calculeu l'error estàndard amb les mateixes dades d'a), suposant que les fites per als errors de x_1, x_2 són, de fet, desviacions estàndard.

SOLUCIÓ:

a) i) $x_1 = 2.0 \pm 0.1$ i $x_2 = 3.0 \pm 0.2$ equivalen a

$$\begin{aligned}\bar{x}_1 - \epsilon_1 &\leq x_1 \leq \bar{x}_1 + \epsilon_1 & (\bar{x}_1 = 2.0, \epsilon_1 = 0.1) , \\ \bar{x}_2 - \epsilon_2 &\leq x_2 \leq \bar{x}_2 + \epsilon_2 & (\bar{x}_2 = 3.0, \epsilon_2 = 0.2) .\end{aligned}$$

Com que $\bar{x}_1 \pm \epsilon_1$ i $\bar{x}_2 \pm \epsilon_2$ són positius,

$$(\bar{x}_1 - \epsilon_1)(\bar{x}_2 - \epsilon_2)^2 \leq x_1 x_2^2 \leq (\bar{x}_1 + \epsilon_1)(\bar{x}_2 + \epsilon_2)^2 . \quad (*)$$

Substituint,

$$14.896 = 1.8 \cdot 2.8^2 \leq x_1 x_2^2 \leq 2.1 \cdot 3.2^2 = 21.504 ;$$

per tant,

$$-3.104 \leq x_1 x_2^2 - \bar{x}_1 \bar{x}_2^2 \leq 3.504 .$$

Prenent la fita més gran,

$$\boxed{x_1 x_2^2 = 18 \pm 3.504} .$$

ii) Usarem $\epsilon_a(y_1 + y_2) = \epsilon_a(y_1) + \epsilon_a(y_2)$ i $\epsilon_r(y_1 \cdot y_2) \simeq \epsilon_r(y_1) + \epsilon_r(y_2)$.

En el nostre cas,

$$\epsilon_r(x_1 x_2^2) = \epsilon_r(x_1 x_2 x_2) \simeq \epsilon_r(x_1) + 2\epsilon_r(x_2) \quad (**) ,$$

$$\epsilon_r(x_1) = \frac{\epsilon_a(x_1)}{x_1} \simeq \frac{\epsilon_a(x_1)}{\bar{x}_1} = \frac{1}{20} , \quad \epsilon_r(x_2) = \frac{\epsilon_a(x_2)}{x_2} \simeq \frac{\epsilon_a(x_2)}{\bar{x}_2} = \frac{2}{30} .$$

Així,

$$\epsilon_r(x_1 x_2^2) \simeq \frac{1}{20} + \frac{4}{30} = \frac{11}{60} , \quad \epsilon_a(x_1 x_2^2) \simeq 18 \frac{11}{60} = 3.3$$

i, finalment,

$$\boxed{x_1 x_2^2 = 18 \pm 3.3} .$$

Una altra manera d'arribar a aquest resultat és operant en (*)

$$\begin{aligned}\bar{x}_1 \bar{x}_2^2 - \bar{x}_2^2 \epsilon_1 - 2\bar{x}_1 \bar{x}_2 \epsilon_2 + 2\bar{x}_2 \epsilon_1 \epsilon_2 + \bar{x}_1 \epsilon_2^2 - \epsilon_1 \epsilon_2^2 &\leq x_1 x_2^2 \leq \\ \bar{x}_1 \bar{x}_2^2 + \bar{x}_2^2 \epsilon_1 + 2\bar{x}_1 \bar{x}_2 \epsilon_2 + 2\bar{x}_2 \epsilon_1 \epsilon_2 + \bar{x}_1 \epsilon_2^2 + \epsilon_1 \epsilon_2^2 &\end{aligned}$$

i menyspreant els termes $\bar{x}_2 \epsilon_1 \epsilon_2$, $\bar{x}_1 \epsilon_2^2$ i $\epsilon_1 \epsilon_2^2$ en els quals apareixen productes d'errors de les dades (és a dir, que contenen més d'un ϵ),

$$\boxed{x_1 x_2^2 = \bar{x}_1 \bar{x}_2^2 \pm (\bar{x}_2^2 \epsilon_1 + 2\bar{x}_1 \bar{x}_2 \epsilon_2) = 18 \pm 3.3} .$$

iii) Usant que $\epsilon_r(y) \simeq \epsilon_a(\ln y)$, tenim

$$\epsilon_r(x_1 x_2^2) \simeq \epsilon_a(\ln x_1 + 2 \ln x_2) \simeq \epsilon_r(x_1) + 2\epsilon_r(x_2) ,$$

com en (**).

b) Usant la fórmula de propagació de l'error estàndard, amb $\sigma_1 = 0.1$, $\sigma_2 = 0.2$, tenim

$$\begin{aligned} \sigma(y) &\simeq [(x_2^2)^2 \sigma_1^2 + (2x_1 x_2)^2 \sigma_2^2]^{\frac{1}{2}} \\ &= [81\sigma_1^2 + 144\sigma_2^2]^{\frac{1}{2}} \simeq 2.6 . \end{aligned}$$

Problema 1.6 Donat el sistema d'equacions lineals

$$\left. \begin{array}{lcl} 3x & + & ay = 10 \\ 5x & + & by = 20 \end{array} \right\} ,$$

on $a = 2.100 \pm 5 \cdot 10^{-4}$ i $b = 3.300 \pm 5 \cdot 10^{-4}$; amb quina exactitud pot ser determinat $x + y$?

SOLUCIÓ:

Les components de la solució del sistema seran

$$x = \frac{10b - 20a}{3b - 5a} , \quad y = \frac{10}{3b - 5a} ;$$

i la suma

$$x + y = 10 \frac{b - 2a + 1}{3b - 5a} .$$

Donem a continuació els valors aproximats de diversos valors intermedis en el càlcul de $x + y$, així com les fites (aproximades) dels errors respectius:

$$\begin{array}{ll} a = 2.1 & \epsilon_a(a) = 5 \cdot 10^{-4} \equiv \epsilon , \\ b = 3.3 & \epsilon_a(b) = 5 \cdot 10^{-4} \equiv \epsilon , \\ N = b - 2a + 1 = 0.1 & \epsilon_a(N) = \epsilon_a(b) + 2\epsilon_a(a) = 3\epsilon , \\ D = 3b - 5a = -0.6 & \epsilon_a(D) = 3\epsilon_a(b) + 5\epsilon_a(a) = 8\epsilon , \\ x + y = 10 \frac{N}{D} = -\frac{10}{6} & \epsilon_a(x + y) = 10 \frac{N\epsilon_a(D) + |D| \epsilon_a(N)}{D^2} \\ & = \frac{650}{9} \epsilon \simeq 4 \cdot 10^{-2} . \end{array}$$

En resulta, doncs,

$$\boxed{x + y = -\frac{5}{3} \pm 4 \cdot 10^{-2}} .$$

Problema 1.7 *Tenim el sistema d'equacions lineals*

$$\left. \begin{array}{rcl} x & + & ay = 5 \\ bx & + & 2y = d \end{array} \right\} ,$$

on $a = 1.00 \pm 5 \cdot 10^{-3}$, $b = 1/a$ i $d = b - a$. Amb quina exactitud podem determinar xy ?

SOLUCIÓ:

Si interpretem l'enunciat de forma que b es calcula a partir d' a i d a partir de b i a , caldrà expressar primer el producte xy de les components de la solució del sistema en funció d' a , b i d i trobar els errors d'aquell a partir dels errors d' a , b i d (entenent que els errors de b i d són els propagats de l'error d' a).

Les components de la solució del sistema seran

$$x = \frac{10 - ad}{2 - ab} , \quad y = \frac{d - 5b}{2 - ab} ;$$

i el producte,

$$xy = \frac{(10 - ad)(d - 5b)}{(2 - ab)^2} .$$

Donem a continuació els valors aproximats de diversos valors intermedis en el càlcul de xy , així com les fites (aproximades) dels errors respectius:

$$\begin{array}{ll} a = 1 & \epsilon_a(a) = \epsilon \equiv 5 \cdot 10^{-3} , \\ b = 1 & \epsilon_a(b) = \frac{1}{a} = \frac{1}{a^2} \epsilon_a(a) = \epsilon , \\ d = 0 & \epsilon_a(d = b - a) = \epsilon_a(a) + \epsilon_a(b) = 2\epsilon , \\ 2 - ab = 1 & \epsilon(2 - ab) = a\epsilon_a(b) + b\epsilon_a(a) = 2\epsilon , \\ D = (2 - ab)^2 = 1 & \epsilon_a(D) = 2(2 - ab)\epsilon_a(2 - ab) = 4\epsilon , \\ N_1 = 10 - ad = 10 & \epsilon_a(N_1) = a\epsilon_a(d) + d\epsilon_a(a) = 2\epsilon , \\ N_2 = d - 5b = -5 & \epsilon_a(N_2) = \epsilon_a(d) + 5\epsilon_a(b) = 7\epsilon , \\ N = N_1 N_2 = -50 & \epsilon_a(N) = |N_2|\epsilon_a(N_1) + N_1\epsilon_a(N_2) = 80\epsilon , \\ xy = \frac{N}{D} = -50 & \epsilon_a(xy) = \frac{D\epsilon_a(N) + |N|\epsilon_a(D)}{D^2} = 280\epsilon . \end{array}$$

En resulta, doncs,

$$\boxed{xy = -50 \pm 1.4} .$$

Problema 1.8 Volem calcular $a = (7 - 4\sqrt{3})^4$, emprant el valor aproximat 1.73205 per a $\sqrt{3}$. Trieu, de les fórmules equivalents següents, la més adequada des del punt de vista numèric:

$$\frac{1}{(7 + 4\sqrt{3})}, \quad (97 - 56\sqrt{3})^2, \quad \frac{1}{(97 + 56\sqrt{3})^2},$$

$$18817 - 10864\sqrt{3}, \quad \frac{1}{18817 + 10864\sqrt{3}}.$$

SOLUCIÓ:

Considerem les funcions següents:

$$\begin{aligned} f_1(x) &= (7 - 4x)^4, \\ f_2(x) &= (7 + 4x)^{-4}, \\ f_3(x) &= (97 - 56x)^2, \\ f_4(x) &= (97 + 56x)^{-2}, \\ f_5(x) &= 18817 - 10864x, \\ f_6(x) &= (18817 + 10864x)^{-1}. \end{aligned}$$

És trivial comprovar que totes les funcions valen a en $x = \sqrt{3}$,

$$a = f_i(\sqrt{3}) \quad (i = 1 \div 6).$$

Quan procedim a avaluar $f_i(\sqrt{3})$ no disposem del valor exacte $\sqrt{3}$, sinó tan sols de l'aproximació $x = 1.73205$; hem de trobar aquella f_i que produeixi el mínim error en el càlcul d' a . Per això, utilitzarem la fórmula aproximada de propagació de l'error maximal en cada cas; així tenim:

$$\begin{aligned} \epsilon_a(f_1(x)) &\simeq 16(7 - 4x)^3 \epsilon_a(x) \simeq 0.005922 \epsilon_a(x), \\ \epsilon_a(f_2(x)) &\simeq 16(7 + 4x)^{-5} \epsilon_a(x) \simeq 0.00003052 \epsilon_a(x), \\ \epsilon_a(f_3(x)) &\simeq 112(97 - 56x) \epsilon_a(x) \simeq 0.5824 \epsilon_a(x), \\ \epsilon_a(f_4(x)) &\simeq 112(97 + 56x)^{-3} \epsilon_a(x) \simeq 0.00001534 \epsilon_a(x), \\ \epsilon_a(f_5(x)) &\simeq 10864 \epsilon_a(x), \\ \epsilon_a(f_6(x)) &\simeq 10864(18817 + 10864x)^{-2} \epsilon_a(x) \simeq 0.00000767 \epsilon_a(x). \end{aligned}$$

Observem que la millor fórmula des del punt de vista numèric és l'última, i que la pitjor és la donada per f_5 . Calculant a amb totes dues, tenim:

$$f_6(\sqrt{3}) \simeq 0.0000265717, \quad f_5(\sqrt{3}) \simeq 0.000976562(!).$$

L'estudi anterior assegura que les 5 primeres xifres significatives del primer valor són correctes; en canvi, no assegura cap xifra significativa correcta del segon valor obtingut, com és en realitat.

Cal notar que, en operacions similars a l'efectuada (en què es resten nombres molt propers), l'error relatiu pot ser molt gran i caldrà evitar de fer-les, substituint la fórmula que s'avalua per una altra d'equivalent però més bona des del punt de vista numèric.

Problema 1.9 Si, per calcular $\ln(x - \sqrt{x^2 - 1})$ per a $x = 30$, l'arrel quadrada es treu d'una taula que dóna 6 xifres decimals correctes amb arrodoniment, quin serà l'error absolut en el resultat? Doneu-ne una expressió equivalent que sigui millor numèricament, i fiteu l'error absolut en emprar-la en les mateixes condicions.

SOLUCIÓ:

Anomenem z la quantitat afectada d'error en l'expressió; tenim així que cal calcular l'error en $f(z) = \ln(30 - z)$, on $z = \sqrt{899}$ té una fita d'error de $\epsilon_a(z) = \frac{1}{2}10^{-6}$.

La fórmula de propagació de l'error dóna la fita aproximada següent:

$$\epsilon_a(f(z)) \simeq |f'(z)| \epsilon_a(z) = \frac{1}{30 - \sqrt{899}} \epsilon_a(z) \simeq \frac{\frac{1}{2}10^{-6}}{0.016671} \simeq 3 \cdot 10^{-5}.$$

Una manera d'obtenir una expressió equivalent, en què no es produeixi el problema de cancel·lació observat, consisteix a tornar a escriure l'argument del logaritme en la forma

$$(x - \sqrt{x^2 - 1}) \frac{x + \sqrt{x^2 - 1}}{x + \sqrt{x^2 - 1}} = \frac{1}{x + \sqrt{x^2 - 1}};$$

llavors, l'expressió equivalent serà

$$\ln(x - \sqrt{x^2 - 1}) = \ln\left(\frac{1}{x + \sqrt{x^2 - 1}}\right) = -\ln(x + \sqrt{x^2 - 1}).$$

Per a $x = 30$, ara caldrà calcular $g(z) = -\ln(30 + z)$ i l'error estarà fitat per

$$\epsilon_a(g(z)) \simeq |g'(z)| \epsilon_a(z) = \frac{1}{30 + \sqrt{899}} \epsilon_a(z) \simeq \frac{\frac{1}{2}10^{-6}}{59.983329} \simeq 8.4 \cdot 10^{-9}.$$

Destaquem que el factor d'expansió de l'error de z es redueix en un factor proper a 3600 en usar la segona fórmula i no la primera, i que aquesta reducció serà més gran com més gran sigui x .

Problema 1.10 Treballant amb 5 xifres decimals, calculeu

$$\sqrt[k]{2.15283} - \sqrt[k]{2.15263} \quad (k = 2, 3, 4).$$

- Directament.
- Usant fórmules equivalents, millors des del punt de vista numèric. (Indicació: Feu la divisió de polinomis $(a^k - b^k)/(a - b)$).
- Compareu els resultats i comenteu-los.

SOLUCIÓ:

	$k = 2$	$k = 3$	$k = 4$
$\sqrt[k]{2.15283}$	1.46725	1.29123	1.21130
$\sqrt[k]{2.15263}$	1.46718	1.29119	1.21127
$\sqrt[k]{2.15283} - \sqrt[k]{2.15263}$	0.00007	0.00004	0.00003

a) En la taula següent es donen els resultats dels càlculs fets directament, emprant tan sols 6 decimals:

Cal destacar que només s'obté una xifra significativa en tots els casos; hi ha, per tant, un problema de cancel·lació de termes que caldrà evitar.

b) Per tal de trobar fórmules equivalents, considerem la divisió del polinomi $a^k - b^k$ entre $a - b$, feta segons la regla de Ruffini,

$$\begin{array}{r|rrrrr} 1 & 0 & 0 & \cdots & 0 & -b^k \\ b & b & b^2 & \cdots & b^{k-1} & b^k \\ \hline 1 & b & b^2 & \cdots & b^{k-1} & 0 \end{array}$$

així,

$$a^k - b^k = (a^{k-1} + a^{k-2}b + a^{k-3}b^2 + \cdots + ab^{k-2} + b^{k-1})(a - b) .$$

Per tant, fent en cada cas $a = \sqrt[k]{A}$ i $b = \sqrt[k]{B}$, tenim

$$\begin{aligned} \sqrt{A} - \sqrt{B} &= \frac{A - B}{\sqrt{A} + \sqrt{B}} , \\ \sqrt[3]{A} - \sqrt[3]{B} &= \frac{A - B}{(\sqrt[3]{A})^2 + \sqrt[3]{A}\sqrt[3]{B} + (\sqrt[3]{B})^2} , \\ \sqrt[4]{A} - \sqrt[4]{B} &= \frac{A - B}{(\sqrt[4]{A})^3 + (\sqrt[4]{A})^2\sqrt[4]{B} + \sqrt[4]{A}(\sqrt[4]{B})^2 + (\sqrt[4]{B})^3} ; \end{aligned}$$

usant aquestes fórmules amb el numerador exacte $A - B = 2 \cdot 10^{-4}$, es troba

$$\begin{aligned} \sqrt[k]{2.15283} - \sqrt[k]{2.15263} &= 6.81563 \cdot 10^{-5} \quad (k = 2) , \\ &3.99866 \cdot 10^{-5} \quad (k = 3) , \\ &2.81340 \cdot 10^{-5} \quad (k = 4) . \end{aligned}$$

c) $\sqrt[k]{2.15283}$ i $\sqrt[k]{2.15263}$ estan calculats amb 5 xifres decimals; per tant, una fita de l'error en fer la resta en a) és de $\frac{1}{2}10^{-5} + \frac{1}{2}10^{-5} = 10^{-5}$. Així, els resultats d'a) amb les fites d'error corresponents són:

$$(7 \pm 1)10^{-5} , (4 \pm 1)10^{-5} , (3 \pm 1)10^{-5} .$$

Si usem, en canvi, les fórmules trobades a b), per exemple amb $k = 2$, i anomenem $D = A - B$, $a = \sqrt{A}$ i $b = \sqrt{B}$, obtenim la fita

$$\epsilon_a\left(\frac{D}{a+b}\right) \simeq \frac{D}{(a+b)^2}(\epsilon_a(a) + \epsilon_a(b)) = \frac{2 \cdot 10^{-4}}{2.93^2}10^{-5} \simeq 0.2 \cdot 10^{-9} .$$

Aquesta fita assegura que podem considerar correctes les 9 primeres xifres decimals del resultat del càlcul demanat

$$\boxed{\sqrt{2.15283} - \sqrt{2.15263} = 6.8156 \cdot 10^{-5}} .$$

Fent càlculs anàlegs, es veu que això també és cert per als altres valors de k :

$$\boxed{\sqrt[3]{2.15283} - \sqrt[3]{2.15263} = 3.9987 \cdot 10^{-5}} ,$$

$$\boxed{\sqrt[4]{2.15283} - \sqrt[4]{2.15263} = 2.8134 \cdot 10^{-5}} .$$

Observeu el considerable guany de 4 xifres significatives que s'obté amb l'ús de les fórmules de b) en comptes de les d'a).

Problema 1.11 Doneu expressions equivalents per a les fórmules següents que siguin més bones des del punt de vista numèric, quan ϵ és molt més petit que x :

- a) $\sqrt[k]{x + \epsilon} - \sqrt[k]{x} ,$
- b) $\frac{1}{\sqrt{x - \epsilon}} - \frac{1}{\sqrt{x}} ,$
- c) $\frac{1}{x + \epsilon} + \frac{1}{x - \epsilon} - \frac{2}{x} ,$
- d) $\sin(x + \epsilon) - \sin(x) ,$
- e) $\cos(x + \epsilon) - \cos(x) ,$
- f) $\tan(x + \epsilon) - \tan(x) ,$
- g) $\int_x^{x+\epsilon} \frac{dt}{t} .$

SOLUCIÓ:

a) Emprant la tècnica de l'exercici anterior, tenim

$$\begin{aligned} \sqrt[k]{x + \epsilon} - \sqrt[k]{x} &= \\ &= \frac{\epsilon}{(\sqrt[k]{x + \epsilon})^{k-1} + (\sqrt[k]{x + \epsilon})^{k-2} \sqrt[k]{x} + \dots + \sqrt[k]{x + \epsilon} (\sqrt[k]{x})^{k-2} + (\sqrt[k]{x})^{k-1}} . \end{aligned}$$

b) Posant denominador comú, tenim

$$\begin{aligned} \frac{1}{\sqrt{x - \epsilon}} - \frac{1}{\sqrt{x}} &= \frac{\sqrt{x} - \sqrt{x - \epsilon}}{\sqrt{x - \epsilon} \sqrt{x}} = \frac{(\sqrt{x} - \sqrt{x - \epsilon})(\sqrt{x - \epsilon} + \sqrt{x})}{(\sqrt{x} \sqrt{x - \epsilon})(\sqrt{x - \epsilon} + \sqrt{x})} \\ &= \frac{\epsilon}{(\sqrt{x} \sqrt{x - \epsilon})(\sqrt{x - \epsilon} + \sqrt{x})} . \end{aligned}$$

c) Posant també denominador comú, resulta

$$\begin{aligned}\frac{1}{x+\epsilon} + \frac{1}{x-\epsilon} - \frac{2}{x} &= \frac{(x-\epsilon)x + (x+\epsilon)x - 2(x+\epsilon)(x-\epsilon)}{(x+\epsilon)(x-\epsilon)x} \\ &= \frac{2\epsilon^2}{(x^2 - \epsilon^2)x}.\end{aligned}$$

d)

$$\begin{aligned}\sin(x+\epsilon) - \sin(x) &= 2\sin\left(\frac{x+\epsilon-x}{2}\right)\cos\left(\frac{x+\epsilon+x}{2}\right) \\ &= 2\sin\left(\frac{\epsilon}{2}\right)\cos\left(x+\frac{\epsilon}{2}\right).\end{aligned}$$

e)

$$\begin{aligned}\cos(x+\epsilon) - \cos(x) &= -2\sin\left(\frac{x+\epsilon-x}{2}\right)\sin\left(\frac{x+\epsilon+x}{2}\right) \\ &= -2\sin\left(\frac{\epsilon}{2}\right)\sin\left(x+\frac{\epsilon}{2}\right).\end{aligned}$$

f)

$$\begin{aligned}\tan(x+\epsilon) - \tan(x) &= \frac{\sin(x+\epsilon)}{\cos(x+\epsilon)} - \frac{\sin(x)}{\cos(x)} \\ &= \frac{\sin(x+\epsilon)\cos(x) - \sin(x)\cos(x+\epsilon)}{\cos(x+\epsilon)\cos(x)} = \frac{\sin(\epsilon)}{\cos(x+\epsilon)\cos(x)}.\end{aligned}$$

g)

$$\int_x^{x+\epsilon} \frac{dt}{t} = \ln(x+\epsilon) - \ln(x) = \ln\left(\frac{x+\epsilon}{x}\right).$$

Problema 1.12 Calculeu, fitant l'error propagat, l'acceleració deguda a la gravetat en la superfície del Sol (g_S), donada per la llei de la gravitació universal de Newton i partint de les següents dades aproximades, que suposarem donades per arrodoniment: massa del Sol, $m_S = 1.90 \cdot 10^{30} \text{ Kg}$; massa de la Terra, $m_T = 5.98 \cdot 10^{24} \text{ Kg}$; radi solar (r_S) igual a 112 radis terrestres (r_T), i la gravetat en la superfície terrestre, $g_T = 9.81 \text{ m/s}^2$.

SOLUCIÓ:

Per la llei de la gravitació universal de Newton, tenim

$$g_S = K \frac{m_S}{r_S^2}, \quad g_T = K \frac{m_T}{r_T^2};$$

on K és la constant de la gravitació universal, que no es té com a dada del problema; així que, aïllant-la de la segona igualtat i substituint-la en la primera, tenim la fórmula per al càlcul de g_S

$$g_S = \frac{m_S \cdot r_T^2}{m_T \cdot r_S^2} g_T.$$

Considerant les variables x_i ($i = 1 \div 4$) tals que

$$m_S = 10^{30} x_1, \quad m_T = 10^{24} x_2, \quad r_S = x_3 r_T, \quad g_T = x_4;$$

llavors

$$g_S = \frac{x_1 x_4}{x_2 x_3^2} 10^6.$$

Emprant els valors aproximats

$$x_1 = 1.90, \quad x_2 = 5.98, \quad x_3 = 112, \quad x_4 = 9.81 \text{ m/s}^2,$$

s'obté l'aproximació

$$g_S \simeq 248.47 \text{ m/s}^2.$$

La fórmula de propagació de l'error maximal ens dóna de forma aproximada una fita de l'error absolut de g_S

$$\begin{aligned} \epsilon_a(g_S) &\simeq \sum_{i=1}^4 \left| \frac{\partial g_S}{\partial x_i}(x_1, x_2, x_3, x_4) \right| \epsilon_a(x_i) \\ &= 10^6 \left(\frac{x_4}{x_2 x_3^2} \epsilon_a(x_1) + \frac{x_4 x_1}{x_2^2 x_3^2} \epsilon_a(x_2) + 2 \frac{x_4 x_1}{x_2 x_3^3} \epsilon_a(x_3) + \frac{x_1}{x_2 x_3^2} \epsilon_a(x_4) \right) 10^6. \end{aligned}$$

Les fites $\epsilon_a(x_i)$ són les fites en l'arrodoniment a les xifres donades dels valors corresponents de x_i : $\epsilon_a(x_1) = \epsilon_a(x_2) = \epsilon_a(x_4) = \frac{1}{2} 10^{-2}$, $\epsilon_a(x_3) = \frac{1}{2}$. Substituint-los en la fórmula anterior, resulta

$$\epsilon_a(g_S) \simeq 3.2 \text{ m/s}^2.$$

El treball de càlcul és molt més senzill si es fa emprant els errors relatius i no els absoluts. Els errors relatius de les dades estan fitats per:

$$\begin{aligned} \epsilon_r(m_S) &= \frac{1}{2} \frac{10^{-2}}{1.9}, \quad \epsilon_r(m_T) = \frac{1}{2} \frac{10^{-2}}{5.8}, \\ \epsilon_r(g_T) &= \frac{1}{2} \frac{10^{-2}}{9.81}, \quad \epsilon_r\left(\frac{r_S}{r_T}\right) = \frac{1}{2} \frac{1}{112}. \end{aligned}$$

Així, directament,

$$\begin{aligned} \epsilon_r(g_S) &\simeq \epsilon_r(m_S) + \epsilon_r(g_T) + \epsilon_r(m_T) + 2\epsilon_r\left(\frac{r_S}{r_T}\right) \\ &= \frac{1}{2} 10^{-2} \left(\frac{1}{1.9} + \frac{1}{9.81} + \frac{1}{5.98} \right) + \frac{1}{112} \simeq 1.3 \cdot 10^{-2}. \end{aligned}$$

L'error absolut està, doncs, fitat per

$$\epsilon_a(g_S) = g_S \epsilon_r(g_S) \simeq 3.2 ,$$

com abans.

Finalment,

$$\boxed{g_S = (248.47 \pm 3.2) \text{ m/s}^2} .$$

Problema 1.13 Considerem una barra de longitud ℓ i secció rectangular, d'amplada a i alçada b , encastada per un dels seus extrems. Si en l'extrem lliure s'aplica una força F perpendicular a la barra, la flexió s experimentada ve donada per l'expressió

$$s = \frac{4}{E} \frac{\ell^3}{ab^3} F ,$$

on E és una constant que depèn només del material, anomenada mòdul de Young.

Sabent que una força de 140 Kp, aplicada sobre una barra de ferro de 125 cm de longitud i secció quadrada de 2.5 cm de costat, li produeix una flexió d'1.71 mm, calculeu el mòdul de Young del ferro i fiteu l'error propagat dels errors de les dades (suposant que aquestes tenen només l'error d'aproximació per tall a les xifres donades).

SOLUCIÓ:

El valor aproximat E del mòdul de Young es troba aïllant primer E de l'expressió donada per a la flexió i substituint després els valors aproximats de les dades

$$E = \frac{4}{s} \frac{\ell^3}{ab^3} F = 16374.27 .$$

L'error en E es troba còmodament treballant amb els errors relatius; així,

$$\epsilon_r(E) = \epsilon_r(s) + 3\epsilon_r(\ell) + \epsilon_r(a) + 3\epsilon_r(b) + \epsilon_r(F) ;$$

$$\text{on } \epsilon_r(s) = \frac{1}{171}, \epsilon_r(\ell) = \frac{1}{125}, \epsilon_r(a) = \epsilon_r(b) = \frac{1}{25}, \epsilon_r(F) = \frac{1}{140} .$$

D'on,

$$\epsilon_r(E) = \frac{1}{171} + \frac{3}{125} + \frac{1}{25} + \frac{3}{25} + \frac{1}{140} = 1.97 \cdot 10^{-1};$$

per tant,

$$\epsilon_a(E) = E \epsilon_r(E) = 3226 .$$

El mòdul de Young del ferro amb la fita d'error trobada ve donat per

$$\boxed{E = (1.6374 \pm 0.3226) 10^4} .$$

Volem fer notar que tan sols la primera xifra significativa es pot considerar correcta.

Problema 1.14 Si usem un ordinador que comet errors relatius fitats per ϵ , en la representació dels nombres i en les operacions aritmètiques, i per 5ϵ en el logaritme, fiteu l'error comès en els càlculs de

$$\begin{aligned} a) \quad & \sum_{i=1}^n a_i, \\ b) \quad & \prod_{i=1}^n a_i, \\ c) \quad & \sum_{i=1}^n x_i y_i, \\ d) \quad & \sum_{i=1}^n x_i \ln x_i, \\ e) \quad & \sum_{i=0}^n a_i x^i, \end{aligned}$$

emprant diferents algorismes i tractant de decidir (si es pot) quina seria la millor manera de realitzar-los.

SOLUCIÓ:

L'enunciat diu que un nombre x s'emmagatzema com a

$$\text{fl}(x) = x(1 + \delta),$$

amb $|\delta| \leq \epsilon$; anàlogament, $x + y$, xy , $\ln x$ s'emmagatzemen respectivament com a

$$\text{fl}(x + y) = (x + y)(1 + \delta_+), \quad \text{fl}(xy) = xy(1 + \delta.), \quad \text{fl}(\ln x) = (\ln x)(1 + \delta_{\ln}),$$

amb $|\delta_+| \leq \epsilon$, $|\delta.| \leq \epsilon$ i $|\delta_{\ln}| \leq 5\epsilon$.

a) Per tal de calcular $a_1 + a_2 + a_3$, prenem $A_i = \text{fl}(a_i) = a_i(1 + \delta_i)$ ($i = 1, 2, 3$) i trobem

$$S_2 = \text{fl}(A_1 + A_2) = (A_1 + A_2)(1 + \delta_{+2}), \quad S_3 = \text{fl}(S_2 + A_3) = (S_2 + A_3)(1 + \delta_{+3}).$$

El valor calculat de la suma serà

$$\text{fl}(a_1 + a_2 + a_3) = \{[a_1(1 + \delta_1) + a_2(1 + \delta_2)](1 + \delta_{+2}) + a_3(1 + \delta_3)\}(1 + \delta_{+3}),$$

on els valors absoluts de δ amb qualsevol subíndex són menors que ϵ .

Menyspreant els termes que donarien una contribució no lineal en els diferents valors de δ , l'error en el càlcul es pot fitar per

$$\begin{aligned} & |S_3 - (a_1 + a_2 + a_3)| \\ & \leq |a_1\delta_1 + a_2\delta_2 + a_3\delta_3 + (a_1 + a_2)\delta_{+2} + (a_1 + a_2 + a_3)\delta_{+3}| \\ & \leq (|a_1| + |a_2| + |a_3| + |a_1 + a_2| + |a_1 + a_2 + a_3|)\epsilon, \end{aligned}$$

que també es pot fitar per $(3|a_1| + 3|a_2| + 2|a_3|)\epsilon$.

En el cas general, fitant els valors absoluts de δ_{+i} per ϵ , queda

$$\begin{aligned} & |S_n - \sum_{i=1}^n a_i| \leq |a_1\delta_1 + a_2\delta_2 + \dots + a_n\delta_n \\ & \quad + (a_1 + a_2)\delta_{+2} + (a_1 + a_2 + a_3)\delta_{+3} + \dots + (a_1 + a_2 + \dots + a_n)\delta_{+n}| \\ & \leq \sum_{i=1}^n |a_i| |\delta_i| \\ & \quad + [(n-1)|a_1| + (n-1)|a_2| + (n-2)|a_3| + \dots + |a_n|]\epsilon; \quad (*) \end{aligned}$$

fitant com abans $|\delta_i| \leq \epsilon$, tenim la fita de l'error absolut

$$\epsilon_a \left(\sum_{i=1}^n a_i \right) \leq [n|a_1| + n|a_2| + (n-1)|a_3| + \dots + 2|a_n|]\epsilon.$$

S'aconsella, doncs, de fer les sumes dels nombres sumant primer els termes de menys valor absolut i acabant amb els de més gran valor absolut per tal de minimitzar l'expressió trobada per a la fita de l'error absolut.

b) Per al producte $\prod_{i=1}^n a_i$ s'actua de forma similar i es troba per a $n = 3$ el valor calculat P_3

$$\{[(a_1(1 + \delta_1)a_2(1 + \delta_2))(1 + \delta_2)a_3(1 + \delta_3)](1 + \delta_3)\}.$$

Menyspreant els termes no lineals, s'escriu com

$$a_1a_2a_3(1 + \delta_1 + \delta_2 + \delta_3 + \delta_{.2} + \delta_{.3})$$

i l'error relatiu està fitat per 5ϵ .

En general,

$$P_n = a_1a_2 \dots a_n(1 \pm (2n-1)\epsilon)$$

i l'error relatiu està fitat per

$$\epsilon_r \left(\prod_{i=1}^n a_i \right) = (2n-1)\epsilon,$$

independentment de l'ordenació dels a_i ($i = 1 \div n$).

c) Refent els procediments anteriors, trobarem fàcilment una fita de l'error absolut en el càlcul de

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

els productes $x_i y_i$ tindran errors absoluts fitats per $3|x_i y_i|\epsilon$ i errors relatius fitats per $\epsilon_i = 3\epsilon$, segons la fórmula anterior per a l'error relatiu en el producte; la suma de tots ells tindrà un error que, prenent $a_i = x_i y_i$, i fitant $|\delta_i|$ per ϵ_i en (*) dóna

$$\epsilon_a \left(\sum_{i=1}^n x_i y_i \right) \leq [(n+2) |x_1 y_1| + (n+2) |x_2 y_2| + (n+1) |x_3 y_3| + \cdots + 4 |x_n y_n|] \epsilon .$$

Es recomana, com en el primer cas, de començar a fer els càlculs amb els sumands més petits, en valor absolut, i acabar amb els més grans.

d) La fita de l'error absolut en

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

es troba de forma similar al cas anterior un cop es conegui una fita de l'error absolut dels $\ln x_i$. Observeu que calen x_i positius a fi de poder calcular llur logaritme.

El valor calculat de $\ln x_i$ serà

$$\begin{aligned} \text{fl}(\ln x_i) &= \ln[x_i(1 + \delta_i)](1 + \delta_{\ln}) \\ &\simeq \ln x_i + \delta_i + \ln x_i \delta_{\ln} , \end{aligned}$$

on s'ha aplicat la fórmula de propagació de l'error a la funció $\ln$ i s'han menyspreat els termes no lineals en els valors de δ . Així, l'error absolut en els productes $a_i = x_i \ln x_i$ està fitat per $(1 + 5 |\ln x_i|)\epsilon$, i l'error relatiu per

$$\epsilon_i = \left(5 + \frac{1}{|\ln x_i|} \right) \epsilon .$$

Fitant de nou $|\delta_i|$ per ϵ_i en (*), tenim la fita de l'error absolut

$$\begin{aligned} \epsilon_a \left(\sum_{i=1}^n x_i \ln(x_i) \right) &= [x_1 + x_2 + x_3 + \cdots + x_n \\ &\quad + (n+4) |x_1 \ln x_1| + (n+4) |x_2 \ln x_2| \\ &\quad + (n+3) |x_3 \ln x_3| + \cdots + 6 |x_n \ln x_n|] \epsilon . \end{aligned}$$

e) L'avaluació del polinomi

$$p(x) = \sum_{i=0}^n a_i x^i$$

s'analitza aquí quant als errors; es tria la regla d'avaluació de Horner perquè és la que realitza menys operacions. Notem, però, que l'algorisme d'avaluació consistent a trobar primer tots els termes $a_i x^i$ i sumar-los després pot analitzar-se emprant la tècnica que s'ha fet servir fins al moment en aquest exercici: trobar l'error en cada sumand i usar (*) per tal de trobar l'error absolut en la suma.

La regla de Horner troba $p(x)$ mitjançant la recurrència

$$b_{n-1} = a_n, \quad b_{j-1} = b_j x + a_j \quad (j = n-1 \div 0); \quad p(x) = b_{-1}$$

(vegeu l'apartat 3.1.2).

Establím ara una recurrència entre fites ϵ_j dels errors absoluts de les b_j ($j = n-1 \div 0$)

$$\begin{aligned} \text{fl}(b_{j-1}) &= \{[(b_j \pm \epsilon_j)x(1 \pm \epsilon)](1 \pm \epsilon) + a_j(1 \pm \epsilon)\}(1 \pm \epsilon) \\ &\simeq b_{j-1} \pm [|x| \epsilon_j + 2 |b_j x| \epsilon + |a_j| \epsilon + |b_{j-1}| \epsilon], \\ \text{fl}(b_{j-1}) - b_{j-1} &\leq |x| \epsilon_j + [2 |b_j x| + |a_j| + |b_{j-1}|] \epsilon \\ &\leq |x| \epsilon_j + [|a_j| + 2 |b_j x| + |b_{j-1}|] \epsilon. \end{aligned}$$

Considerarem, doncs, la següent recurrència per a les fites:

$$\begin{aligned} \epsilon_n &= |a_n| \epsilon, \\ \epsilon_{j-1} &= |x| \epsilon_j + [|a_j| + 2 |b_j x| + |b_{j-1}|] \epsilon \quad (j = n \div 0). \end{aligned}$$

Introduint

$$d_j = \frac{\epsilon_j}{\epsilon} + 2 |b_j| \quad (j = 0 \div n),$$

queda

$$\begin{aligned} d_{n-1} &= |a_n| + 2 |b_{n-1}| = 3 |a_n|, \\ d_{j-1} &= d_j |x| + |a_j| + 3 |b_{j-1}| \quad (j = n-1 \div 0); \end{aligned}$$

per tant, com que $\epsilon_{-1} = \epsilon(d_{-1} - 2 |b_{-1}|)$, una fita de l'error absolut en $p(x)$ es pot trobar, doncs, avaluant en $|x|$ el polinomi de coeficients positius

$$c_n = 3 |a_n|, \quad c_j = |a_j| + 3 |b_{j-1}| \quad (j = n-1 \div 1), \quad c_0 = |a_0| + |b_{-1}|.$$

Multiplicant finalment per ϵ , trobem la fita demanada

$$\epsilon_a(p(x)) = \left(\sum_{j=0}^n c_j |x|^j \right) \epsilon.$$

Problema 1.15 Un ordinador comet errors relatius fitats per ϵ , ϵ , 2ϵ , 4ϵ i 5ϵ en l'emmagatzematge de dades, operacions aritmètiques, arrel quadrada, logaritme i exponencial, respectivament. Per a quins valors d' a i b l'error relatiu en el càlcul de a^b com $\exp[b \ln(a)]$ és menor que 12ϵ ?

Aplicació: Estudieu el cas $\sqrt[3]{x}$.

SOLUCIÓ:

Volem escriure l'aproximació $\text{fl}(a^b)$ en la forma $a^b(1 + \delta)$ i després fitar δ . Per això, farem una anàlisi en cadena de les fites dels errors produïts en l'emmagatzematge i en cada operació, tenint en compte que

$$\text{fl}(x) = x(1 + \delta_x); \quad \text{fl}(x * y) = \text{fl}(x) * \text{fl}(y)(1 + \delta_*);$$

amb $|\delta_x| \leq \epsilon$ i $|\delta_*| \leq \epsilon$, per a les operacions aritmètiques; i

$$\text{fl}(f(x)) = f(\text{fl}(x))(1 + \delta_f),$$

amb $|\delta_f| \leq 2\epsilon$, per a l'arrel quadrada, $|\delta_f| \leq 4\epsilon$, per al logaritme, i $|\delta_f| \leq 5\epsilon$, per a l'exponenciació.

Així,

$$\text{fl}(a) = a(1 + \delta_a) \quad (\text{representació inicial d'a}),$$

$$\text{fl}(\ln(a)) = \ln[a(1 + \delta_a)](1 + \delta_{\ln}) \quad (\text{logaritme}),$$

$$\text{fl}(b) = b(1 + \delta_b) \quad (\text{representació inicial de b}),$$

$$\text{fl}(b \ln(a)) = b(1 + \delta_b) \ln[a(1 + \delta_a)](1 + \delta_{\ln})(1 + \delta_{\cdot}) \quad (\text{producte}),$$

$$\text{fl}(a^b) = \exp\{b \ln[a(1 + \delta_a)](1 + \delta_b)(1 + \delta_{\ln})(1 + \delta_{\cdot})\}(1 + \delta_{\exp}) \quad (\text{exponencial}).$$

Usant la fórmula de propagació de l'error en una variable, tenim

$$\ln[a(1 + \delta_a)] \simeq \ln(a) + \delta_a, \quad \exp[b \ln(a) + \gamma] \simeq a^b(1 + \gamma).$$

Fent-ho servir en l'expressió anterior, amb

$$\gamma = b\delta_a + b \ln(a)(\delta_{\ln} + \delta_b + \delta_{\cdot}),$$

s'arriba a

$$\text{fl}(a^b) \simeq a^b[1 + \delta_{\exp} + b\delta_a + b \ln(a)(\delta_{\ln} + \delta_b + \delta_{\cdot})],$$

que ja dóna una expressió aproximada per a δ en la qual s'han tingut en compte només els errors lineals en ϵ .

Fitant, l'aproximació de δ en aquesta expressió i usant la informació donada de les fites dels errors comesos en l'emmagatzematge i en els càlculs, s'obté

$$\text{fl}(a^b) = a^b[1 \pm (5 + |b|(1 + 6|\ln(a)|))\epsilon];$$

d'on, la condició sobre a i b demanada en l'enunciat serà

$$|b|(1 + 6|\ln(a)|) \leq 7.$$

Tal cosa fa que $|b| \leq 7$ i, en aquest cas,

$$e^{-c} \leq a \leq e^c, \quad c = \frac{7}{6|b|} - \frac{1}{6}.$$

L'aplicació correspon a prendre $b = \frac{1}{7}$; així, l'error en el l'arrel setena, calculada com a $\exp[\frac{1}{7} \ln(a)]$, serà menor que 12ϵ per als valors d' a en l'interval $[e^{-8}, e^8]$.

Problema 1.16 Trobeu una expressió de la fita aproximada de l'error absolut en avaluar la funció

$$f(x) = \sqrt{3 + \ln^2(x)} .$$

Hom suposa que els errors relatius en la representació de nombres, operacions aritmètiques, arrels quadrades i logaritmes són, en valor absolut, menors que ϵ , 2ϵ , 3ϵ i 5ϵ , respectivament; a més, hom comet un error relatiu menor, en valor absolut, que 4ϵ a l'hora de mesurar x .

SOLUCIÓ:

Per tal de calcular $f(x) = \sqrt{3 + \ln^2(x)}$ fem els passos següents:

1. Emmagatzematge de x : $x \rightarrow x_1 = \text{fl}(x) = x(1 \pm 4\epsilon)(1 \pm \epsilon)$.
2. Càlcul de $\ln x$: $x_1 \rightarrow x_2 = \text{fl}(\ln x_1) = \ln x_1(1 \pm 5\epsilon)$.
3. Càlcul de $\ln^2 x$: $x_2 \rightarrow x_3 = \text{fl}(x_2^2) = x_2^2(1 \pm 2\epsilon)$.
4. Càlcul de $3 + \ln^2 x$: $x_3 \rightarrow x_4 = \text{fl}(3 + x_3) = (3 + x_3)(1 \pm 2\epsilon)$.
5. Càlcul de $\sqrt{3 + \ln^2 x}$: $x_4 \rightarrow x_5 = \text{fl}(\sqrt{x_4}) = \sqrt{x_4}(1 \pm 3\epsilon)$.

Notem que

$$x_1 \simeq x, \quad x_2 \simeq \ln x, \quad x_3 \simeq \ln^2 x, \quad x_4 \simeq 3 + \ln^2 x, \quad x_5 \simeq \sqrt{3 + \ln^2 x} .$$

Volem trobar una fita per a l'error absolut. Emprarem repetidament la fórmula per a l'error en un producte i la fórmula aproximada de propagació dels errors:

1. $\epsilon_a(x_1) = 5x\epsilon$.
2. $\epsilon_a(x_2) = \epsilon_a(\ln x_1) + 5|x_2|\epsilon \simeq \frac{1}{x_1}\epsilon_a(x_1) + 5|\ln x_1|\epsilon$
 $\simeq 5(1 + |\ln x_1|)\epsilon$.
3. $\epsilon_a(x_3) = \epsilon_a(x_2^2) + 2x_2^2\epsilon \simeq 2x_2\epsilon_a(x_2) + 2x_2^2\epsilon$
 $\simeq [10|\ln x|(1 + |\ln x|) + 2\ln^2 x]\epsilon = (12\ln^2 x + 10|\ln x|)\epsilon$.
4. $\epsilon_a(x_4) = \epsilon_a(x_3) + 2(3 + x_3)\epsilon \simeq (14\ln^2 x + 10|\ln x| + 6)\epsilon$.
5. $\epsilon_a(x_5) = \frac{1}{2\sqrt{x_4}}\epsilon_a(x_4) + 3\sqrt{x_4}\epsilon$
 $\simeq \left[\frac{1}{2\sqrt{3 + \ln^2 x}}(14\ln^2 x + 10|\ln x| + 6) + 3\sqrt{3 + \ln^2 x} \right] \epsilon$.

L'expressió final per a la fita és

$$\epsilon_a(\sqrt{3 + \ln^2 x}) = \frac{1}{\sqrt{3 + \ln^2 x}}(10\ln^2 x + 5|\ln x| + 12)\epsilon .$$

Problema 1.17 Hem de calcular $\cos(\cos \sqrt{x})$ i tenim representat x amb un error relatiu menor que ϵ i els algorismes de càlcul de l'arrel quadrada i el cosinus tenen errors relatius fitats per 3ϵ i 5ϵ , respectivament. Fiteu aproximadament l'error relatiu total en funció de x , suposant ϵ prou petit, i demostreu que, per a tot $x \in (0, 1)$, la fita és menor que $(5 + \sqrt{37.25} \tan 1)\epsilon$.

SOLUCIÓ:

Per tal de calcular $\cos(\cos \sqrt{x})$ fem els passos següents:

1. Emmagatzematge de x : $x \rightarrow x_1 = \text{fl}(x) = x(1 \pm \epsilon)$.
2. Càlcul de $\sqrt{x}$: $x_1 \rightarrow x_2 = \text{fl}(\sqrt{x_1}) = \sqrt{x_1}(1 \pm 3\epsilon)$.
3. Càlcul de $\cos \sqrt{x}$: $x_2 \rightarrow x_3 = \text{fl}(\cos x_2) = \cos x_2(1 + 5\epsilon)$.
4. Càlcul de $\cos(\cos \sqrt{x})$: $x_3 \rightarrow x_4 = \text{fl}(\cos x_3) = \cos x_3(1 + 5\epsilon)$.

Notem que

$$x_1 \simeq x, \quad x_2 \simeq \sqrt{x}, \quad x_3 \simeq \cos \sqrt{x}, \quad x_4 \simeq \cos(\cos \sqrt{x}).$$

Volem trobar una fita per a l'error relatiu

$$\epsilon_r(x_4) \simeq \frac{\epsilon_a(x_4)}{\cos(\cos \sqrt{x})};$$

per això, emprarem repetidament la fórmula per a l'error en un producte i la fórmula aproximada de propagació de l'error:

1. $\epsilon_a(x_1) = x\epsilon$.
2. $\epsilon_a(x_2) = \epsilon_a(\sqrt{x_1}) + 3\sqrt{x_1}\epsilon \simeq \frac{1}{2\sqrt{x_1}}\epsilon_a(x_1) + 3\sqrt{x_1}\epsilon$
 $\simeq \frac{7}{2}\sqrt{x}\epsilon$.
3. $\epsilon_a(x_3) = \epsilon_a(\cos x_2) + 5\cos x_2\epsilon \simeq \sin x_2\epsilon_a(x_2) + 5\cos x_2\epsilon$
 $\simeq \left(\frac{7}{2}\sqrt{x}\sin \sqrt{x} + 5\cos \sqrt{x}\right)\epsilon$.
4. $\epsilon_a(x_4) = \epsilon_a(\cos x_3) + 5\cos x_3\epsilon \simeq \sin x_3\epsilon_a(x_3) + 5\cos x_3\epsilon$
 $\simeq \left[\left(\frac{7}{2}\sqrt{x}\sin \sqrt{x} + 5\cos \sqrt{x}\right)\sin(\cos \sqrt{x}) + 5\cos(\cos \sqrt{x})\right]\epsilon$.

Tenim ja, doncs, la fita per a l'error relatiu per a cada $x \in (0, 1)$

$$\epsilon_r(x_4) = \left[\frac{7}{2}\sqrt{x}\sin \sqrt{x} + 5\cos \sqrt{x}\right]\tan(\cos \sqrt{x}) + 5\epsilon;$$

per demostrar que està fitat per $(5 + \sqrt{37.25} \tan 1)\epsilon$, caldrà veure tan sols que la funció

$$f(t) = \frac{7}{2}t \sin t + 5 \cos t$$

està fitada per $\sqrt{37.25}$ per a tota $t \in (0, 1)$, atès que la funció $\tan \cos t$ és una funció decreixent en l'interval $(0, 1)$ i pot fitar-se per $\tan \cos 0 = \tan 1$.

Fitant $f(t)$ per la funció $g(t) = \frac{7}{2} \cos t + 5 \sin t$; aquesta funció pren el seu valor màxim per a $t = t^*$ tal que $g'(t^*) = 0$ (això és, $\tan t^* = \frac{7}{10}$); calculant el seu cosinus trobem

$$\cos t^* = \frac{1}{\sqrt{1 + \tan^2 t^*}} = \frac{10}{\sqrt{149}}$$

i llavors, com es volia demostrar

$$g(t) \leq g(t^*) = \left(\frac{7}{2} \tan t^* + 5\right) \cos t^* = \left(\frac{7}{2} \frac{7}{10} + 5\right) \frac{10}{\sqrt{149}} = \sqrt{\frac{149}{4}} = \sqrt{37.25}.$$

Problema 1.18 Els errors relatius en la representació de nombres, operacions aritmètiques (+, −, ·, /), arrel quadrada i càlcul de l'arcsin són, en valor absolut, menors que ϵ , ϵ , 2ϵ i 5ϵ , respectivament. Per a quins valors de x es pot garantir que el valor absolut de l'error relatiu d' $\arctan(x)$ és menor que 20ϵ , si el calculem mitjançant la fórmula

$$\arctan x = \arcsin \frac{x}{\sqrt{1+x^2}} ?$$

SOLUCIÓ:

Per al càlcul de $f(x) = \arctan x = \arcsin \frac{x}{\sqrt{1+x^2}}$ cal fer els passos intermedis següents:

1. Emmagatzematge de x : $x \rightarrow x_1 = \text{fl}(x) = x(1 \pm \epsilon)$.
2. Càlcul de x^2 : $x_1 \rightarrow x_2 = \text{fl}(x_1^2) = x_1^2(1 \pm \epsilon)$.
3. Càlcul de $1 + x^2$: $x_2 \rightarrow x_3 = \text{fl}(1 + x_2) = (1 + x_2)(1 \pm \epsilon)$.
4. Càlcul de $\sqrt{1 + x^2}$: $x_3 \rightarrow x_4 = \text{fl}(\sqrt{x_3}) = \sqrt{x_3}(1 \pm 2\epsilon)$.
5. Càlcul de $\frac{x}{\sqrt{1+x^2}}$: $x_4 \rightarrow x_5 = \text{fl}\left(\frac{x_1}{x_4}\right) = \frac{x_1}{x_4}(1 \pm \epsilon)$.
6. Càlcul d' $\arctan x$: $x_5 \rightarrow x_6 = \text{fl}(\arcsin x_5) = \arcsin x_5(1 \pm 5\epsilon)$.

Notem que $x_1 \simeq x$, $x_2 \simeq x^2$, $x_3 \simeq 1 + x^2$, $x_4 \simeq \sqrt{1 + x^2}$, $x_5 \simeq \frac{x}{\sqrt{1+x^2}}$, $x_6 \simeq \arctan x$.

Volem trobar una fita per a l'error relatiu de x_6 per això, emprarem repetidament la fórmula per a l'error en un producte i la fórmula aproximada de propagació de l'error:

1. $\epsilon_a(x_1) = |x| \epsilon$.
2. $\epsilon_a(x_2) = \epsilon_a(x_1^2) + x_2 \epsilon \simeq 2 |x_1| \epsilon_a(x_1) + x_1^2 \epsilon \simeq 3x^2 \epsilon$.
3. $\epsilon_a(x_3) = \epsilon_a(1 + x_2) + (1 + x_2) \epsilon \simeq \epsilon_a(x_2) + (1 + x_2) \epsilon \simeq (1 + 4x^2) \epsilon$.
4. $\epsilon_a(x_4) = \epsilon_a(\sqrt{x_3}) + 2\sqrt{x_3} \epsilon \simeq \frac{1}{2\sqrt{x_3}} \epsilon_a(x_3) + 2\sqrt{x_3} \epsilon$

$$\begin{aligned}
& \simeq \left(\frac{1+4x^2}{2\sqrt{1+x^2}} + 2\sqrt{1+x^2} \right) \epsilon = \frac{5+8x^2}{2\sqrt{1+x^2}} \epsilon . \\
5. \quad \epsilon_a(x_5) &= \epsilon_a\left(\frac{x_1}{x_4}\right) + \frac{|x_1|}{x_4} \epsilon \simeq \frac{|x_1 \epsilon_a(x_4) + x_4 \epsilon_a(x_1)|}{x_4^2} + \frac{|x_1|}{x_4} \epsilon \simeq \\
& \simeq \frac{3|x|(3+4x^2)}{2\sqrt{1+x^2}(1+x^2)} \epsilon . \\
6. \quad \epsilon_a(x_6) &= \epsilon_a(\arcsin x_5) + 5 |\arcsin x_5| \epsilon \simeq \frac{\epsilon_a(x_5)}{\sqrt{1-x_5^2}} + 5 \arcsin x_5 \epsilon \\
& \simeq \left[\frac{3|x|(3+4x^2)}{2(1+x^2)} + 5 |\arctan x| \right] \epsilon .
\end{aligned}$$

L'expressió obtinguda per a la fita de l'error relatiu és

$$\epsilon_r(\arctan x) \simeq \left[\frac{3|x|(3+4x^2)}{2|\arctan x|(1+x^2)} + 5 \right] \epsilon ;$$

la condició d'error relatiu menor que 20ϵ pot assegurar-se per als x complint

$$\frac{3|x|(3+4x^2)}{2|\arctan x|(1+x^2)} < 15 .$$

Aquesta es compleix per als valors de x tals que

$$|x| < 3.25 .$$

Problema 1.19 Usant un mètode recurrent, calculeu el valor de les integrals

$$I_j = \int_0^1 x^j \sin(\pi x) \, dx \quad (j = 2, 4, \dots, 20) .$$

Estudieu l'estabilitat del mètode trobat.

SOLUCIÓ:

La recurrència es troba prenent l'expressió d' I_j i integrant-la per parts dues vegades:

$$\begin{aligned}
 I_j &= \int_0^1 x^j \sin(\pi x) dx = -\frac{1}{\pi} x^j \cos(\pi x) \Big|_0^1 - \frac{j}{\pi} \int_0^1 x^{j-1} \cos(\pi x) dx \\
 &= \frac{1}{\pi} - \frac{j}{\pi} \left[x^{j-1} \sin(\pi x) \Big|_0^1 - \frac{j-1}{\pi} \int_0^1 x^{j-2} \sin(\pi x) dx \right] \\
 &= \frac{1}{\pi} - \frac{j(j-1)}{\pi^2} I_{j-2} .
 \end{aligned}$$

La recurrència que compleixen les integrals I_j ($j = 2, 4, 6, \dots$) és, doncs,

$$I_0 = \frac{2}{\pi} = 0.636619772\dots, \quad I_j = \frac{1}{\pi} - \frac{j(j-1)}{\pi^2} I_{j-2} \quad (j \geq 2) .$$

j	I_j
0	0.636620
2	0.189304
4	0.088144
6	0.050384
8	0.032433
10	0.022560
12	0.016588
14	0.012423
16	0.016211
18	-0.018429
20	7.413890

Taula 1.1: Recurrència cap al davant.

j	I_j
0	1.000
2	0.202
4	0.246
6	0.748
8	4.249
10	38.749
12	518.249
14	9556.798
16	232392.245
18	7205154.731
20	277413226.170

Taula 1.2: Factors d'amplificació d'errors.

j	I_j
26	0.000000
24	0.004833
22	0.005605
20	0.006680
18	0.008094
16	0.010006
14	0.012679
12	0.016574
10	0.022561
8	0.032433
6	0.050384
4	0.088144
2	0.189304
0	0.636620

Taula 1.3: Recurrència cap al darrera.

En la taula 1.1 es mostren els resultats obtinguts en emprar la recurrència amb simple precisió (que és equivalent a treballar amb 6 xifres decimals).

La recurrència es mostra decididament *inestable*: notem que les integrals haurien de decreïxer en augmentar el valor de j , i això no s'observa; a més, cap integral I_j pot ser negativa, i I_{18} ho és. Això s'explica pel fet que els errors s'amplifiquen a cada pas de la recurrència per un factor $\frac{j(j-1)}{\pi^2}$. Aquest factor augmenta a mesura que augmenta j , i el factor d'amplificació total per al càlcul d' I_j és $\frac{j!}{\pi^j}$ que pren els valors aproximats donats a la taula 1.2.

Els fets anteriors expliquen la inestabilitat i deixen entreveure una forma d'usar la recurrència anterior de forma estable: si s'aïlla I_{j-2} en la recurrència anterior es troba

$$I_{j-2} = \frac{\pi^2}{j(j-1)} \left(\frac{1}{\pi} - I_j \right) .$$

Per tal d'utilitzar aquesta nova recurrència cal disposar d'algun valor d' I_j ; ara bé, usant el fet que la successió d'integrals tendeix a 0, es pot prendre l'aproximació $I_n = 0$, per a un valor prou gran de n . Els errors comesos en aquesta aproximació s'aniran reduint a mesura que s'apliqui la recurrència i podrem fer el càlcul demanat. Si es pren $n = 26$ s'obté la taula 1.3 que dona el valor de les integrals demanades amb 6 xifres decimals correctes, encara que els tres primers valors no tinguin totes aquestes xifres correctes.

Problema 1.20 Feu el mateix que en l'exercici anterior amb les integrals

$$I_j = \int_0^1 \frac{x^j}{x+10} dx \quad (j = 1 \div 20) .$$

SOLUCIÓ:

La recurrència es troba simplement escrivint el numerador de l'integrand com $x^j + 10x^{j-1} - 10x^{j-1}$, llavors

$$I_j = \int_0^1 x^{j-1} dx - 10 \int_0^1 \frac{x^{j-1}}{x+10} dx = \frac{1}{j} - 10I_{j-1} .$$

La recurrència pot començar-se amb $I_0 = \ln(1.1) = 0.0953101798\dots$ i en cal esperar, com en l'exercici anterior, una desestabilització numèrica del procés.

j	I_j	j	I_j
0	0.095310	11	0.007622
1	0.046898	12	0.007114
2	0.031018	13	0.005785
3	0.023154	14	0.013577
4	0.018465	15	0.069105
5	0.015353	16	0.753549
6	0.013138	17	-7.476665
7	0.011481	18	74.822208
8	0.010194	19	-748.169449
9	0.009167	20	7481.744486
10	0.008329		

Taula 1.4: Recurrència cap al davant.

j	I_j	j	I_j
24	0.000000	11	0.007629
23	0.004167	10	0.008328
22	0.003931	9	0.009167
21	0.004152	8	0.010194
20	0.004347	7	0.011481
19	0.004565	6	0.013138
18	0.004807	5	0.015353
17	0.005075	4	0.018465
16	0.005375	3	0.023154
15	0.005713	2	0.031018
14	0.006095	1	0.046898
13	0.006533	0	0.095310
12	0.007039		

Taula 1.5: Recurrència cap al darrera.

En la taula 1.4 es mostren els resultats obtinguts en emprar la recurrència amb doble precisió (que és equivalent a treballar amb 16 xifres decimals), donant-ne les 6 primeres xifres decimals.

Les integrals haurien de decreixer en augmentar el valor de j , i això no s'observa; a més, cap integral I_j pot ser negativa, i I_{17} ja ho és. L'explicació del fet es basa en què els errors s'amplifiquen a cada pas de la recurrència per un factor 10. A mesura que augmenta j , el factor d'amplificació total per al càlcul d' I_j és 10^j .

Emprant la recurrència a l'inrevés

$$I_{j-1} = \frac{1}{10} \left(\frac{1}{j} - I_j \right) ,$$

el mètode de càlcul d'integrals esdevé estable; si partim d' $I_n = 0$ per a n prou gran, a cada pas d'utilització de la recurrència l'error es divideix per 10, es guanya aproximadament una xifra decimal correcta. Si es pren $n = 24$, ja s'obtenen els valors demanats amb 6 xifres decimals correctes; les darreres xifres d' I_j ($j = 24, 23, 22, 21$) estan encara afectades d'error. Els resultats dels càlculs es mostren en la taula 1.5.

Problema 1.21 *Volem avaluar en $x = 1$ funcions de Bessel d'ordre enter*

$$J_n(x) = \sum_{j=0}^{\infty} (-1)^j \frac{\left(\frac{x}{2}\right)^{2j+n}}{j! (j+n)!} ,$$

que compleixen la recurrència

$$J_{n+1}(x) = \frac{2n}{x} J_n(x) - J_{n-1}(x) \quad (n \geq 1) .$$

a) Sabent que $J_0(1) = 0.765198$ i $J_1(1) = 0.440051$, calculeu $J_7(1)$ emprant la recurrència.

b) Compareu amb el resultat exacte $J_7(1) = 0.1502326 \cdot 10^{-6}$ i expliqueu el fenomen.

c) Emprant la relació

$$J_0(x) + 2 \sum_{r=1}^{\infty} J_{2r}(x) = 1 ,$$

calculeu $J_n(1)$ ($n = 0 \div 20$) , amb 6 xifres significatives correctes.

SOLUCIÓ:

n	$J_n(1)$
1	0.440051
2	0.114904
3	0.019565
4	0.002486
5	0.000321
6	0.000725
7	0.008377

Taula 1.6: Recurrència cap al davant.

n	$\bar{J}_n(1)$	n	$J_n(1)$
24	1	24	$9.507130 \cdot 10^{-32}$
23	48	23	$4.563423 \cdot 10^{-30}$
22	2207	22	$2.098224 \cdot 10^{-28}$
21	97060	21	$9.227621 \cdot 10^{-27}$
20	4074313	20	$3.873502 \cdot 10^{-25}$
19	162875456	19	$1.548478 \cdot 10^{-23}$
18	6185192960	18	$5.880344 \cdot 10^{-22}$
17	222504075264	17	$2.115375 \cdot 10^{-20}$
16	7558953172992	16	$7.186395 \cdot 10^{-19}$
15	241664002097152	15	$2.297531 \cdot 10^{-17}$
14	7242361222463488	14	$6.885407 \cdot 10^{-16}$
13	202544455746584576	13	$1.925617 \cdot 10^{-14}$
12	$5.25891353811773030 \cdot 10^{18}$	12	$4.999718 \cdot 10^{-13}$
11	$1.26011377280803144 \cdot 10^{20}$	11	$1.198007 \cdot 10^{-11}$
10	$2.76699134485810958 \cdot 10^{21}$	10	$2.630615 \cdot 10^{-10}$
9	$5.52138161795883650 \cdot 10^{22}$	9	$5.249249 \cdot 10^{-9}$
8	$9.91081733664729666 \cdot 10^{23}$	8	$9.422343 \cdot 10^{-8}$
7	$1.58020943277800528 \cdot 10^{25}$	7	$1.502326 \cdot 10^{-6}$
6	$2.20238241593444582 \cdot 10^{26}$	6	$2.093834 \cdot 10^{-5}$
5	$2.62705690163896132 \cdot 10^{27}$	5	$2.497577 \cdot 10^{-4}$
4	$2.60503302398405905 \cdot 10^{28}$	4	$2.476639 \cdot 10^{-3}$
3	$2.05775589444304340 \cdot 10^{29}$	3	$1.956335 \cdot 10^{-2}$
2	$1.20860318753651952 \cdot 10^{30}$	2	$1.149035 \cdot 10^{-1}$
1	$4.62863725514910340 \cdot 10^{30}$	1	$4.400506 \cdot 10^{-1}$
0	$8.04867147387741473 \cdot 10^{30}$	0	$7.651977 \cdot 10^{-1}$

Taula 1.7: Recurrència cap al darrera.

a) En la taula 1.6 es mostren els resultats obtinguts en emprar la recurrència

$$J_0(1) = 0.765198, \quad J_1(1) = 0.440051, \quad J_{n+1}(1) = 2nJ_n(1) - J_{n-1}(1),$$

amb precisió doble (que és equivalent a treballar amb 16 xifres decimals), donant-ne les 6 primeres xifres decimals. b) No hi ha cap xifra significativa de coincidència amb el valor

exacte, els errors relatius comesos esdevenen tràgicament grans. D'una banda, els errors absoluts poden amplificar-se a cada pas per un factor $2n$ (si no tenim en compte els errors de $J_{n-1}(1)$). La seva contribució en el factor d'amplificació total de l'error absolut per al càlcul de $J_n(1)$ esdevé de $(2n-1)!$, molt gran quan n és gran; d'altra banda, els valors de $J_n(1)$ són cada cop més petits. El resultat combinat d'aquests dos efectes és el d'un error relatiu enorme per a $J_7(1)$.

c) Emprant la recurrència a l'inrevés

$$J_{n-1}(1) = 2nJ_n(1) - J_{n+1}(1) ,$$

els errors absoluts van també en augment de la mateixa forma que en a); sembla en principi que, si procedim a emprar aquesta fórmula no obtindrem cap resultat acurat; ara però, a mesura que anem trobant valors, aquests són cada cop més grans i es té l'esperança d'un bon funcionament. A més, la relació

$$J_0(1) + 2 \sum_{r=1}^{\infty} J_{2r}(1) = 1$$

ens procura una ajuda molt valuosa.

El procediment triat es basa en l'establiment de la recurrència per a un valor d' a qualsevol

$$\begin{aligned} \bar{J}_{N+1}(1) &= 0 , \quad \bar{J}_N(1) = a , \\ \bar{J}_{n-1}(1) &= 2n\bar{J}_n(1) - \bar{J}_{n+1}(1) \quad (n = N \div 1) , \end{aligned}$$

on $\bar{J}_{N+1}(1) = 0$ és una bona aproximació si N és prou gran. El valor d' a permet controlar la grandària dels valors de la successió, atès que són proporcionals a aquesta a . La successió buscada es dona per a un cert valor d' a que es pot determinar aproximadament emprant la relació anterior. Així, triant N parell ($N = 2s$) i algun valor d' a , usem la recurrència anterior per trobar $\bar{J}_n(1)$ ($n = N - 1 \div 0$). Calculem

$$S_{2s} = \bar{J}_0(1) + 2 \sum_{r=1}^s \bar{J}_{2r}(1)$$

ja que és una aproximació de $\frac{a}{\bar{J}_N(1)}$. La successió buscada es trobarà llavors dividint per S_{2s} els valors trobats per als $\bar{J}_n(1)$.

En la taula 1.7 es mostren els valors trobats per a ambdues successions, fent servir $a = 1$ i $N = 24$. Les 6 primeres xifres significatives dels valors demanats han estat verificades com a correctes.

El valor trobat per a la suma és de $S_{24} = 1.05184210896169095 \cdot 10^{31}$; els valors de $J_n(1)$ s'han trobat dividint els de $\bar{J}_n(1)$ per aquest valor.

Problema 1.22 *Siguin*

$$I_j = \int_0^1 \frac{x^j}{x^2 + x + 6} dx \quad (j \geq 0) .$$

a) Trobeu una recurrència per als I_j i indiqueu com s'ha d'emprar per calcular-los de forma numèricament estable.

b) Si es coneixen I_{n-1} i I_n amb errors absoluts menors que ϵ i 2ϵ , respectivament; amb quantes xifres decimals correctes es podrà obtenir I_j mitjançant el mètode trobat a a) (suposant aritmètica exacta)?

c) Expliqueu com es poden obtenir tots el I_j amb t xifres decimals correctes sense calcular cap integral.

SOLUCIÓ:

a) La recurrència es troba escrivint el numerador de l'integrand com $x^j + x^{j-1} + 6x^{j-2} - x^{j-1} - 6x^{j-2}$, llavors

$$\begin{aligned} I_j &= \int_0^1 x^{j-2} dx - \int_0^1 \frac{x^{j-1}}{x^2 + x + 6} dx - 6 \int_0^1 \frac{x^{j-2}}{x^2 + x + 6} dx \\ &= \frac{1}{j-1} - I_{j-1} - 6I_{j-2} . \end{aligned}$$

La recurrència s'ha d'emprar en la forma

$$I_{j-2} = \frac{1}{6} \left(\frac{1}{j-1} - I_j - I_{j+1} \right)$$

per tal que sigui numèricament estable; així, la suma dels errors d' I_j i I_{j-1} es divideix per 6 a cada pas d'utilització de la recurrència, la qual cosa fa que els errors decreixin.

b) Considerem I_n i I_{n-1} amb fites d'errors absoluts ϵ i 2ϵ . L'error absolut en I_{n-2} pot fitar-se per

$$\epsilon_a(I_{n-2}) = \frac{\epsilon_a(I_{n-1}) + \epsilon_a(I_n)}{6} = \frac{\epsilon}{2} ,$$

que és la meitat del d' I_{n-1} ; repetint-ho de nou $n-j$ cops, trobem

$$\epsilon_a(I_j) = \frac{\epsilon}{2^{n-j-1}} .$$

L'error trobat no afectarà la xifra decimal t -ena si

$$\frac{\epsilon}{2^{n-j}} < \frac{1}{2} 10^{-t} ; \quad (*)$$

així doncs, podem assegurar que el nombre t de xifres decimals correctes per a I_j és la part entera de

$$\log \frac{2^{n-j-2}}{\epsilon} .$$

c) Si prenem el valor $j = n$ i aproximem I_n per zero es comet un error fitat per

$$\epsilon_a(I_n) = \int_0^1 \frac{x^n}{6} dx = \frac{1}{6(n+1)} \equiv \epsilon ;$$

si prenem també l'aproximació per zero per a I_{n-1} , tenim que l'error està fitat per

$$\int_0^1 \frac{x^{n-1}}{6} dx = \frac{1}{6n} \leq 2\epsilon \equiv \epsilon_a(I_{n-1}) .$$

Així, aplicant el resultat trobat a b), tenim assegurades t xifres decimals correctes si es compleix la relació (*); cal prendre, per a cada valor de j , un valor de n complint

$$3(n+1)2^{n-j-1} > 10^t .$$

Notem que, a partir de $j = 166666$, totes les xifres decimals d' I_j són nul·les.

PROBLEMES PROPOSATS

1. Una calculadora treballa en base 2 i mantissa de 20 bits i aproxima per arrodoniment, i un ordinador treballa en base 16 i mantissa de 24 bits i aproxima per tall. Quin dels dos es pot considerar més precís?
2. Amb quina exactitud s'ha de mesurar el radi d'una esfera i amb quants decimals s'ha de donar el número π perquè el seu volum es conegui amb un error relatiu menor que el 0.01%? Considereu ambdós efectes per separat.
3. Coneixem el sinus i la tangent d'un angle amb la mateixa precisió. Indiqueu si és millor usar la funció arcsin o la funció arctan per trobar el valor de l'angle.
4. Trobeu les arrels de l'equació de segon grau $x^2 - 200x + 1 = 0$ amb 9 xifres decimals, emprant una calculadora de mà. Per què no és correcte usar l'expressió $100 \pm \sqrt{100^2 - 1}$?
5. Coneixent les longituds a i b de dos costats d'un triangle i el valor C de l'angle que formen, la longitud c del costat restant es pot trobar emprant la *fórmula del cosinus*

$$c = \sqrt{a^2 + b^2 - 2ab \cos(C)} .$$

a) Si es coneixen exactament a i b , però de C tan sols sabem que és a l'interval I , trobeu l'interval d'incertesa de c ,

- exactament, operant amb intervals,
- aproximadament, emprant la fórmula de propagació de l'error maximal,

i en els casos següents:

i) $a = 10$, $b = 20$, $I = [\frac{\pi}{6}, \frac{\pi}{3}]$;

ii) $a = b = 20$, $I = [-\epsilon, \epsilon]$.

b) Si $a = b = 20$ i $C = 0.00001$, calculeu c amb el mateix nombre de xifres correctes que la vostra calculadora admeti.

6. Calculeu l'error permès en la inclinació d'un canó per assegurar que encerti un objectiu rectangular de 40 metres de longitud i 20 metres d'amplada, situat horitzontalment a la mateixa alçada del canó, i el centre del qual és a una distància de 3.000 metres d'aquest. La velocitat de sortida del projectil és de 600 metres per segon. Suposeu que totes les magnituds donades són exactes i que les dimensions de l'objectiu

són petites en relació a la distància a què es troba el canó. Considereu també que s'apunta al centre de l'objectiu i que és permès de tocar qualsevol lloc d'aquest.

7. Quina és la millor manera de sumar 9999, 999, 99 i 9 amb una calculadora que fa servir representació decimal en punt flotant de 4 dígits de mantissa?

8. Estudieu, en funció d' a , b i c , quina és la millor manera de calcular en punt flotant

$$a + b + c, \quad abc, \quad (a + b)^2, \quad a^2b^2, \quad (a + b)(a + c).$$

9. Hem de calcular x^n per a n natural. Suposem que x està emmagatzemat amb un error relatiu menor que ϵ , que les multiplicacions es fan amb error relatiu menor que ϵ i que les funcions $\ln$ i $\exp$ donen errors relatius fitats per 4ϵ i 6ϵ , respectivament. Compareu l'acumulació dels errors en els dos algorismes següents per al seu càlcul:

i) $x^n = \exp[n \ln(x)]$;

ii) es representa n en base 2, $n = a_k \cdots a_1 a_0$ ($a_k \neq 0$), es calculen les potències x^{2^j} ($j = 1 \div k$) i es multipliquen aquelles que es corresponen amb els uns de la representació en base 2 (exemple: x^{21} es calcula fent

$$x \rightarrow x^2 \rightarrow x^4 \rightarrow x^8 \rightarrow x^{16} \rightarrow x^{20} = x^{16} \cdot x^4 \rightarrow x^{21} = x^{20} \cdot x).$$

10. Volem calcular $\cos(20x)$ per a $x \in [0, \pi/2]$. Disposem d'un ordinador que treballa en base 2 i, per tant, fa exactament les multiplicacions per enters i les divisions per potències de 2. Les operacions aritmètiques i les representacions dels nombres es realitzen amb un error relatiu fitat per ϵ . La funció que calcula $\cos y$ per a $y \in [0, \pi/2]$ comet un error relatiu fitat per 3ϵ .

Fiteu aproximadament l'error relatiu en el càlcul de $\cos(20x)$ per a $x \in [0, \pi/2]$, usant els algorismes següents:

i) Reduir primer l'angle $20x$ al primer quadrant i aplicar-li després la funció de càlcul del cosinus.

ii) Aplicar la funció directament a x per trobar $\cos x$ i fer servir la recurrència: $c_0 = 1$, $c_1 = \cos x$, $c_{k+1} = 2c_1c_k - c_{k-1}$ ($k = 1 \div 19$), per obtenir $\cos(20x) = c_{20}$.

11. Per a l'avaluació del polinomi $p(x) = a_nx^n + \cdots + a_1x + a_0$ hom disposa dels algorismes següents:

i) $p \leftarrow a_0$	ii) $p \leftarrow a_n$
$z \leftarrow 1$	
$(k = 1 \div n)$	$(k = n - 1 \div 1)$
$z \leftarrow zx$	
$p \leftarrow p + a_k z$	$p \leftarrow px + a_k$
	(Algorisme de Horner) .

Sigui ϵ una fita de l'error relatiu en fer les operacions aritmètiques i no considerem error de representació.

- a) Expliciteu l'error que hom comet en el càlcul de p en cada pas, en funció de l'error en el pas anterior en cadascun dels algorismes.
- b) Doneu una fita de l'error absolut comès en calcular un valor de $p(x)$, emprant cadascun dels algorismes. (Indicació: feu-ho per recurrència).
- c) Suposant que $|a_i| \simeq 1$ ($i = 0 \div n$), digueu quin algorisme convé usar: i) quan x pren valors grans, ii) quan x pren valors petits.

12. Resoleu l'equació

$$\frac{a}{x} = 0.2 - \frac{x}{100},$$

on $a = 1.000 \pm 0.001$ i estudeu els efectes de la incertesa d' a sobre les solucions de l'equació.

13. Estudieu el comportament de les solucions del sistema

$$\left. \begin{aligned} 2x + 3y + z &= 6 \\ 2\epsilon x + 2y + 2\epsilon z &= 2 + 4\epsilon \\ 2\epsilon x + 2y - \epsilon z &= 1 + \epsilon \end{aligned} \right\},$$

per a ϵ proper a 0.

14. Les solucions de l'equació de tercer grau $z^3 + a_2 z^2 + a_1 z + a_0 = 0$ poden expressar-se en funció dels coeficients a_2 , a_1 i a_0 com s'indica a continuació.

Fent primerament

$$q = \frac{a_1}{3} - \frac{a_2^2}{3}, \quad r = \frac{a_1 a_2 - 3a_0}{6} - \frac{a_2^3}{27}, \quad s_{1,2} = \sqrt[3]{r \pm \sqrt{q^3 + r^2}},$$

les solucions z_0 , z_1 i z_2 s'expressen així:

$$z_0 = (s_1 + s_2) - \frac{a_2}{3}, \quad z_{1,2} = -\frac{s_1 + s_2}{2} - \frac{a_2}{3} \pm \frac{\sqrt{3}(s_1 - s_2)}{2} i.$$

Per simplificar, suposarem $a_2 = 0$.

- a) Estudieu per a quins valors d' a_0 i a_1 apareixen cancel·lacions en el problema quan se cerquen les solucions emprant les fórmules anteriors.
- b) Doneu mètodes numèricament més bons per al cas $|a_0| \gg |a_1|$.
- c) Aplicació: Per a $a_0 = 100$ i $a_1 = 0.01$ i suposant que ambdós tenen error absolut fitat per ϵ , estimeu el valor de ϵ que asseguri aproximadament una fita d'error absolut per a z_0 de 10^{-10} .

CAPÍTOL 2

SISTEMES LINEALS

2.1 CONCEPTES BÀSICS

2.1.1 Tipus de matrius

Denotarem per $\mathcal{M}_{m,n}$ l'espai vectorial de les matrius complexes $m \times n$ de m files i n columnes. Els *vectors de dimensió m* són elements de $\mathcal{M}_{m,1}$.

Els *elements d'una matriu* $M \in \mathcal{M}_{m,n}$ seran denotats per

$$m_{kj} \quad (k = 1 \div m) \quad (j = 1 \div n)$$

i els *elements d'un vector* y , per

$$y_k \quad (k = 1 \div m) .$$

En símbols, $M = (m_{kj})$ i $y = (y_k)$.

Per a una matriu $M \in \mathcal{M}_{m,n}$ complexa, denotarem la seva *transposada* per $M^\top$, la seva *conjugada* per $\overline{M}$ i la seva *adjunta* per M^* : $M^\top \in \mathcal{M}_{n,m}$ es forma canviant files per columnes, $\overline{M} \in \mathcal{M}_{m,n}$ es forma conjugant els elements de M i $M^* \in \mathcal{M}_{n,m}$ és la transposada de la matriu conjugada de M .

Les matrius tractades en aquest capítol seran considerades reals, llevat que s'especifiqui el contrari.

Les matrius A amb el mateix nombre de files que columnes ($m = n$) reben el nom de *matrius quadrades*. Denotarem la *matriu identitat* per I i la *inversa* de la matriu A per A^{-1} , si existeix. Els elements de la matriu identitat es poden denotar per δ_{ij} , fent servir la *delta de Kronecker*:

$$\delta_{ii} = 1 \quad , \quad \delta_{ij} = 0 \quad (i \neq j) .$$

Les matrius $k \times k$ formades per les k primeres fileres de les k primeres columnes d' A seran anomenades *submatrius principals d'ordre k d' A* i denotades per $(A)_k$.

Denotarem el *determinant d' A* per $\det A$. Anomenarem *determinants principals d' A* els determinants de les seves submatrius principals i els denotarem per $\Delta_k \equiv \det (A)_k$.

Vegem ara diferents tipus de matrius quadrades:

- *A singular* : $\det A = 0$.

- *A regular* : $\det A \neq 0$.
- *A hermitica* (*A simètrica*) : $A^* = A$ ($A^\top = A$).
- *A unitària* (*A ortogonal*) : $A^{-1} = A^*$ ($A^{-1} = A^\top$).
- *A definida positiva*: *A* hermitica i $x^*Ax > 0 \quad \forall x \neq 0$.
- *A estrictament diagonal dominant* :

$$|a_{ii}| > \sum_{j \neq i, j=1}^n |a_{ij}| \quad (i = 1 \div n) .$$

- *A banda* (p, q) : $a_{ij} = 0 \quad (i \geq j + p \text{ o } j \geq i + q)$.
 - *A Hessenberg superior* : $p = 2, \quad q = n$.
 - *A Hessenberg inferior* : $p = n, \quad q = 2$.
 - *A triangular superior* : $p = 1, \quad q = n$.
 - *A triangular inferior* : $p = n, \quad q = 1$.
 - *A diagonal, tridiagonal, pentadiagonal, ...* : $p = q = 1, 2, 3, \dots$

Les definicions de matrius unitàries i hermitiques corresponen a matrius complexes i les de matrius ortogonals i simètriques, a matrius reals. Noteu que les matrius unitàries i hermitiques reals són, respectivament, matrius ortogonals i simètriques.

Dues matrius A i B s'anomenen *semblants* si existeix una matriu invertible C , anomenada *transformació de semblança d'A a B*, tal que es compleix $B = C^{-1}AC$.

Una matriu A es diu *diagonalitzable* quan és semblant a una matriu diagonal.

2.1.2 Definicions

Sistemes d'equacions lineals

Tot sistema lineal de n equacions amb n incògnites pot ser escrit, en forma matricial, com

$$Ax = b \quad \text{o breument} \quad (A|b) ,$$

on A és una matriu $n \times n$, anomenada *matriu del sistema*, i b és un vector de dimensió n , anomenat *vector de termes independents*. Els elements de la matriu A són denotats per a_{ij} ($i, j = 1 \div n$) i, els del vector b , per b_i ($i = 1 \div n$).

Se suposa en aquest capítol que el sistema és *compatible* i *determinat*; és a dir, que el determinant de la matriu A és no nul. L'objectiu és llavors cercar el vector x de components x_j ($j = 1 \div n$) tal que $Ax = b$, anomenat *solució del sistema*. En el capítol següent es tractaran sistemes de la forma

$$Ma = y ,$$

on M és una matriu $m \times n$ amb $m > n$ (amb més equacions que incògnites).

Valors i vectors propis

Donada una matriu $n \times n$ $A = (a_{ij})$, direm que $v \neq 0$ és un *vector propi* d' A de *valor propi* λ quan

$$Av = \lambda v .$$

La condició anterior pot ser escrita en la forma $(A - \lambda I)v = 0$. A fi que aquest sistema d'equacions tingui solucions no nul·les és necessari i suficient que λ compleixi l'*equació característica*

$$p_A(\lambda) \equiv \det(A - \lambda I) = 0 ;$$

així doncs, els valors propis d' A són els zeros del polinomi p_A , anomenat *polinomi característic* d' A .

Els vectors propis d' A i $A^\top$ s'anomenen també *vectors propis per la dreta* i *vectors propis per l'esquerra* d' A , respectivament.

S'anomena *radi espectral* d' A al mòdul màxim de tots els valors propis d' A i es denota per $\rho(A)$.

A continuació es presenten dues consideracions que il·lustren les definicions fetes de valors i vectors propis:

INTERPRETACIÓ GEOMÈTRICA

La matriu A pot considerar-se com a representació, en una certa base, d'una transformació lineal de l'espai vectorial $\mathbb{R}^n$ en ell mateix. Un vector propi és aquell vector v que, sotmès a la transformació, conserva la seva direcció (si λ és complex, entenem que λv representa un vector en la mateixa direcció que v , respecte al cos dels nombres complexos). Si $\lambda > 0$ en conserva també el sentit i, si $\lambda < 0$, l'inverteix. Si $|\lambda| = 1$, en conserva el mòdul; si $|\lambda| > 1$, el mòdul augmenta i, si $|\lambda| < 1$, disminueix.

Si λ és un valor propi d' A , la solució del sistema $(A - \lambda I)v = 0$ no és única; existeix tot un subspai vectorial de vectors solució que s'anomena *subspai propi* del valor propi λ . Un subspai generat per un sol vector propi s'anomena també *direcció pròpia*.

EXEMPLE

La matriu

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

té per equació característica

$$p_A(\lambda) = (3 - \lambda)(3 - \lambda) - 1 = \lambda^2 - 6\lambda + 8 = 0 ,$$

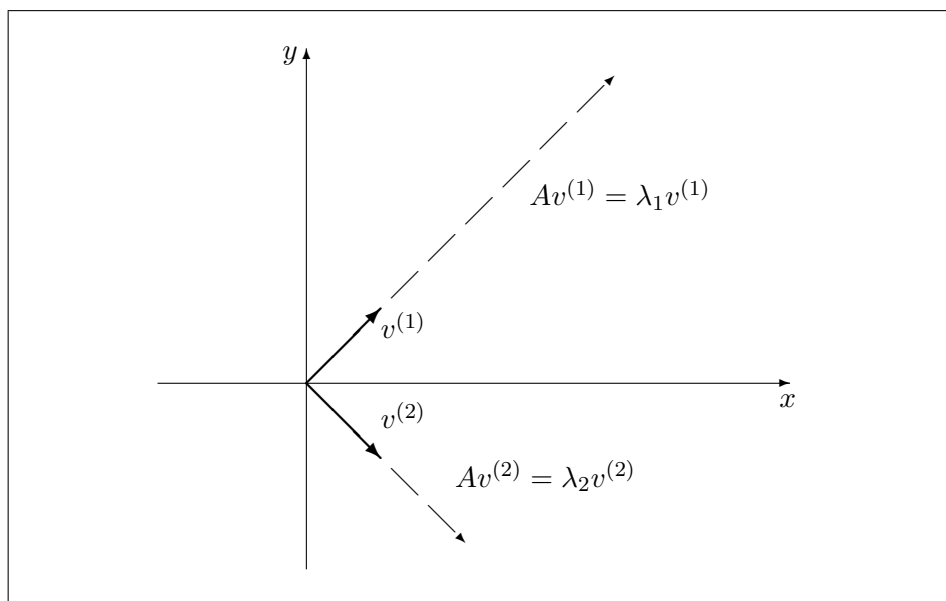
de solucions $\lambda_1 = 4$ i $\lambda_2 = 2$. Aquestes solucions són llavors els valors propis d' A .

Els subspais propis $\mathcal{V}_1$ i $\mathcal{V}_2$ dels valors propis $\lambda_1 = 4$ i $\lambda_2 = 2$ són les direccions pròpies generades, per exemple, pels vectors propis

$$v^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \quad v^{(2)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} ,$$

respectivament.

En la figura 2.1 es pot veure el comportament de la transformació feta per A sobre les direccions pròpies.

Figura 2.1: Efecte d' A sobre els seus vectors propis.

Normes vectorials i normes matricials

Sigui E un espai vectorial real o complex. Una *norma* a E és una aplicació

$$\begin{array}{ccc} \|\cdot\| : E & \longrightarrow & \mathbb{R}^+ \\ x & \longrightarrow & \|x\| \end{array}$$

complint

- $\|x\| = 0 \Leftrightarrow x = 0$;
- $\|cx\| = |c| \|x\| \quad \forall \text{ escalar } c \quad \forall x \in E$;
- $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in E$.

Les *normes vectorials* són aquelles que estan definides en espais vectorials de la forma $E = \mathbb{R}^n$ o $E = \mathbb{C}^n$. Les normes vectorials més emprades són les *normes de Hölder*

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (p \geq 1) .$$

Els casos particulars més importants pel seu ús són:

- *La norma suma de mòduls*

$$\|x\|_1 = \sum_{i=1}^n |x_i| \quad (p = 1) ,$$

- la norma euclidiana

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (p=2) .$$

Una altra norma molt emprada és el cas límit (quan p tendeix a infinit) de les normes de Hölder, la *norma del màxim*

$$\|x\|_\infty = \max_{i=1 \div n} |x_i| .$$

Una *norma matricial* és una norma en l'espai vectorial $\mathcal{M}_{n,n}$ que sigui *multiplicativa*; això és, que compleixi

$$\|AB\| \leq \|A\| \|B\| \quad \forall A, B \in \mathcal{M}_{n,n} .$$

Una norma matricial $\| \cdot \|$ és *consistent amb una norma vectorial* (que notarem igual) si i només si

$$\|Ax\| \leq \|A\| \|x\| \quad \forall A \in \mathcal{M}_{n,n} , \quad \forall x \in \mathbb{R}^n .$$

Donada una norma vectorial $\| \cdot \|$, sempre pot definir-se una norma matricial consistent amb ella fent

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

i és anomenada *norma matricial subordinada* a la norma vectorial donada.

2.1.3 Propietats importants

S'enuncien a continuació algunes propietats importants que poden ser útils en la resolució de sistemes lineals, en el càlcul de determinants i inverses de matrius i en la resolució de problemes de valors i de vectors propis de matrius.

Matrius

1. $(A+B)^\top = A^\top + B^\top$.
2. $\overline{A+B} = \overline{A} + \overline{B}$.
3. $(AB)^\top = B^\top A^\top$.
4. $\overline{AB} = \overline{A} \overline{B}$.
5. $\det A^\top = \det A$.
6. $\det AB = \det A \det B$.
7. $(AB)^{-1} = B^{-1}A^{-1}$, si A i B són regulars.
8. Una matriu quadrada té inversa si i només si és regular.
9. *Lema de Schur*. Tota matriu A és semblant a una matriu triangular superior, mitjançant una transformació de semblança amb una matriu unitària.
10. *Criteri de Sylvester*. Una matriu simètrica és definida positiva si i només si té tots els determinants principals positius.

Valors i vectors propis

1. Les matrius A i $A^\top$ tenen els mateixos valors propis. Els vectors propis per la dreta són ortogonals als vectors propis per l'esquerra de valors propis diferents.
2. Una matriu A és regular si i només si té tots els valors propis diferents de zero; en tal cas, els valors propis d' A^{-1} són els recíprocs dels valors propis d' A i v és un vector propi de valor propi λ d' A si i només si v és un vector propi de valor propi $\frac{1}{\lambda}$ d' A^{-1} .
3. *Teorema de Gerschgorin.* Els valors propis d' A estan localitzats, dins el pla complex, en la unió $\mathcal{F}$ dels discs

$$F_i = \{ \lambda \mid |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \} \quad (i = 1 \div n) ,$$

i també, en la unió $\mathcal{C}$ dels discs

$$C_j = \{ \lambda \mid |\lambda - a_{jj}| \leq \sum_{i \neq j} |a_{ij}| \} \quad (j = 1 \div n) .$$

A més, si una component connexa de $\mathcal{F}$ (o de $\mathcal{C}$) està formada per d discs F_i (o C_j), aquesta conté exactament d valors propis d' A comptats segons llur multiplicitat. Conseqüentment, tota matriu estrictament diagonal dominant és invertible.

4. Dues matrius semblants A i B tenen els mateixos valors propis, i v és un vector propi de valor propi λ d' A si i només si $C^{-1}v$ és un vector propi de valor propi λ de B , essent C la transformació de semblança d' A a B .
5. Si una matriu A és de la forma

$$A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array} \right) ,$$

amb les matrius A_{11} i A_{22} quadrades, el conjunt de valors propis d' A és la unió dels conjunts de valors propis d' A_{11} i d' A_{22} . Així, els valors propis d'una matriu triangular són els elements de la diagonal.

6. Si p és un polinomi no nul, λ és un valor propi d' A si i només si $p(\lambda)$ ho és de $p(A)$; v és un vector propi de valor propi λ d' A si i només si és vector propi de valor propi $p(\lambda)$ de $p(A)$.
7. Sigui $\lambda_1, \dots, \lambda_n$ els valors propis d' A , repetits segons llur multiplicitat; llavors,

$$\sum_{j=1}^n \lambda_j = \operatorname{tr} A \equiv \sum_{j=1}^n a_{jj} , \quad \prod_{j=1}^n \lambda_j = \det A .$$

8. Sigui A diagonalitzable per una transformació de semblança V . Aleshores $D = V^{-1}AV$ és diagonal, els elements de la diagonal de D són els valors propis d' A , les columnes de V formen una base de vectors propis (per la dreta) d' A , i les fileres de V^{-1} , una base de vectors propis per l'esquerra d' A .

9. Vectors propis corresponents a valors propis diferents són linealment independents. Per tant, si una matriu té tots els valors propis diferents és diagonalitzable.
10. Els valors propis d'una matriu hermítica són reals. Tota matriu hermítica A és diagonalitzable mitjançant una transformació de semblança amb una matriu unitària; si U^*AU és diagonal amb U unitària, les columnes d' U formen una base de vectors propis ortonormals d' A . En el cas real, si A és simètrica, és diagonalitzable mitjançant una transformació amb una matriu ortogonal, les columnes de la qual formen també una base de vectors propis ortonormals d' A .

Normes vectorials i matricials

1. Les normes matricials subordinades a les normes vectorials $\| \cdot \|_1$, $\| \cdot \|_2$ i $\| \cdot \|_\infty$ resulten ser

$$\|A\|_1 = \max_{j=1 \div n} \sum_{i=1}^n |a_{ij}|, \quad (2.1)$$

$$\|A\|_2 = \sqrt{\rho(A^*A)}, \quad (2.2)$$

$$\|A\|_\infty = \max_{i=1 \div n} \sum_{j=1}^n |a_{ij}| \quad (2.3)$$

(vegeu el problema II.4 per a les normes $\| \cdot \|_1$ i $\| \cdot \|_\infty$).

2. El radi espectral d'una matriu és menor o igual que qualsevol norma matricial de la dita matriu:

$$\rho(A) \leq \|A\|.$$

3. Donada una matriu A qualsevol i per a qualsevol $\epsilon > 0$, es pot trobar una norma matricial, subordinada a una certa norma vectorial, tal que

$$\|A\| \leq \rho(A) + \epsilon.$$

2.2 SISTEMES D'EQUACIONS

2.2.1 Introducció

Aquesta secció està dedicada a la resolució de sistemes d'equacions lineals i a problemes relacionats com són el càlcul de determinants i inverses de matrius.

Els mètodes de resolució de sistemes lineals es classifiquen en mètodes directes i mètodes iteratius.

Els *mètodes directes* permeten obtenir la solució del sistema després d'un nombre finit de passos. Hi ha tipus de matrius per a les quals és extraordinàriament fàcil la resolució directa dels sistemes corresponents. Alguns són: el de les matrius triangulars i, òbviament, el de les matrius unitàries (ortogonals, si són reals). Per a altres tipus de matrius més generals presentarem els mètodes gaussians i els mètodes d'ortogonalització. Les tècniques

Mètode de Gauss

El *mètode de Gauss* consta de dues fases.

En la fase anomenada d'eliminació gaussiana, es fa la reducció del sistema inicial a un sistema triangular superior.

Aquest objectiu s'aconsegueix mitjançant un procés d'eliminació que, partint de la matriu $A = A_1 = (a_{ij}^{(1)})$, va obtenint matrius $A_k = (a_{ij}^{(k)})$ ($k = 2 \div n$), amb $a_{ij}^{(k)} = 0$ ($i > j$, $j = 1 \div k - 1$),

$$A_k = \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \cdots & a_{1k}^{(k)} & \cdots & a_{1n}^{(k)} \\ & a_{22}^{(k)} & \cdots & a_{2k}^{(k)} & \cdots & a_{2n}^{(k)} \\ & & \ddots & \vdots & \cdots & \vdots \\ & & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ & & & \vdots & \cdots & \vdots \\ & & & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} \end{pmatrix} \quad (k = 2 \div n) ,$$

de forma que A_n resulti ser triangular superior. Paral·lelament a aquestes modificacions sobre la matriu del sistema, es realitzen modificacions sobre el vector b de termes independents, per tal que les transformacions del sistema no alterin les solucions; així obtenim els vectors $b^{(k)}$ de components $b_i^{(k)}$ ($i = 1 \div n$) ($k = 2 \div n$).

El pas k -è d'aquesta primera fase es realitza canviant només els valors dels elements de les fileres que van de la $(k+1)$ -ena fins a l'última i els dels elements corresponents del vector de termes independents, segons:

$$\begin{aligned} m_{ik} &= \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} , \quad b_i^{(k+1)} = b_i^{(k)} - m_{ik} b_k^{(k)} , \\ a_{ij}^{(k+1)} &= a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)} \quad (j = k \div n) \\ &\quad (i = k+1 \div n) \quad (k = 1 \div n-1) . \end{aligned} \quad (2.4)$$

En la segona fase es resol el sistema triangular superior $A_n x = b^{(n)}$ pel procediment de substitució cap al darrera detallat en l'apartat 2.2.2.

Per a n gran, el nombre d'operacions que es realitzen, al llarg de tot el procés, és de $\frac{n^3}{3} + \mathcal{O}(n^2)$. Per a matrius amb característiques especials el mètode pot optimitzar-se molt en el sentit d'evitar operacions innecessàries (vegeu el problema II.2).

Noteu que, en l'etapa k -ena del procés de reducció, hem de dividir per l'element diagonal $a_{kk}^{(k)}$, anomenat *pivot*. Convé fer tres observacions al respecte:

1. Els determinants principals coincideixen amb els productes de pivots

$$\Delta_k = a_{11}^{(1)} \cdots a_{kk}^{(k)} .$$

El mètode de Gauss pot, doncs, aplicar-se directament si i només si tots aquests determinants són no nuls (vegeu el problema II.1).

2. Si $a_{kk}^{(k)} = 0$, llavors podem cercar $a_{ik}^{(k)} \neq 0$ ($i > k$) (que existeix sempre, en ser $\det A \neq 0$) i permutar l'equació k -ena amb la i -ena; així es pot continuar el procés.
3. Si $a_{kk}^{(k)} \neq 0$ però és petit, el mètode de Gauss, encara que pugui aplicar-se, és molt inestable numèricament en el sentit que els errors de la solució, propagats a partir dels errors de les dades, poden ser augmentats notablement. Així doncs, convindrà modificar el mètode de Gauss també en aquest cas.

Podem triar entre les opcions següents:

- *Pivotatge maximal per columnes.* Es pren com a nou pivot $a_{kk}^{(k)}$ l'element de màxim valor absolut dels elements $a_{ik}^{(k)}$ ($i = k \div n$).
- *Pivotatge complet.* El nou pivot passa a ser l'element de màxim valor absolut dels elements $a_{ij}^{(k)}$ ($i, j = k \div n$).

En el primer cas, hem de canviar fileres i termes independents; en el segon, fileres i termes independents, així com columnes i les incògnites corresponents.

Mètodes LU i de Txoleski

El mètode LU produeix primer la *factorització LU de la matriu*

$$A = LU ,$$

on $L = (l_{ik})$ és triangular inferior amb uns a la diagonal i $U = (u_{kj})$, triangular superior.

Aquesta factorització es pot fer de manera única sempre que tots els determinants principals d' A siguin no nuls: condició equivalent a la de la possibilitat d'aplicació del mètode de Gauss sense pivotatge. Els elements de L i d' U es poden trobar llavors usant només els elements de la matriu A i aleshores resulta:

$$\begin{aligned} l_{kk} &= 1, \quad l_{ik} = m_{ik} \quad (i = k + 1 \div n) , \\ u_{kj} &= a_{kj}^{(k)} \quad (j = k \div n) \\ &\quad (k = 1 \div n) . \end{aligned} \tag{2.5}$$

En cas que no es compleixi la condició exigida i sempre que A sigui regular, serà possible de permutar les equacions de manera que la nova matriu admeti una tal factorització.

Un cop feta la factorització LU, la solució del sistema $Ax = b$ és també la solució del sistema $Ux = y$, essent y la solució de $Ly = b$. Així, caldrà resoldre successivament els sistemes triangulars $Ly = b$ i $Ux = y$.

Aquesta segona fase és comuna a tots els mètodes gaussians que segueixen i ja no hi serà explicitada.

Per a matrius simètriques, és millor fer la factorització

$$A = LDL^{\top} ,$$

on $L = (l_{ik})$ és una matriu triangular inferior amb uns a la diagonal i D és una matriu diagonal amb elements diagonals d_{kk} ($k = 1 \div n$):

$$\begin{aligned} d_{kk} &= a_{kk} - \sum_{r=1}^{k-1} l_{kr}^2 d_{rr} , \\ l_{ik} &= \frac{1}{d_{kk}} \left(a_{ik} - \sum_{r=1}^{k-1} l_{ir} d_{rr} l_{kr} \right) \quad (i = k+1 \div n) \\ &\quad (k = 1 \div n) . \end{aligned} \quad (2.6)$$

Noteu que el nombre d'operacions es redueix essencialment a la meitat.

Per a matrius definides positives, els elements diagonals d_{kk} són positius, segons el criteri de Sylvester (vegeu la propietat 10 de l'apartat 2.1.3). Si $\mathcal{L}$ és la matriu $LD^{1/2}$, amb $D^{1/2} = \text{diag}(d_{11}^{1/2}, \dots, d_{nn}^{1/2})$, llavors A pot expressar-se en la forma

$$A = \mathcal{L}\mathcal{L}^\top,$$

anomenada *factorització de Txoleski*. La dita factorització, amb elements diagonals de $\mathcal{L}$ positius, existeix i és única si i només si A és definida positiva.

En el problema II.1 s'estudia l'efecte de l'eliminació gaussiana sobre matrius definides positives i s'exposa un exemple de factorització de Txoleski.

Si la matriu A és complexa, existeixen factoritzacions anàlogues dels tipus $A = LDL^*$ i $A = \mathcal{L}\mathcal{L}^*$.

Mètode de Doolittle

El *mètode de Doolittle* persegueix el mateix objectiu inicial que el mètode LU, però ofereix, sota les mateixes condicions, un algorisme de càlcul dels elements de les matrius que no fa servir l'eliminació gaussiana:

$$\begin{aligned} u_{kj} &= a_{kj} - \sum_{r=1}^{k-1} l_{kr} u_{rj} \quad (j = k \div n) , \\ l_{kk} &= 1 , \quad l_{ik} = \frac{1}{u_{kk}} \left(a_{ik} - \sum_{r=1}^{k-1} l_{ir} u_{rk} \right) \quad (i = k+1 \div n) \\ &\quad (k = 1 \div n) . \end{aligned} \quad (2.7)$$

Els elements u_{kk} ($k = 1 \div n$) seran no nuls, sota les hipòtesis d'existència de la factorització LU: $\det(A)_k \neq 0$ ($k = 1 \div n$).

L'ordre de càlcul és important: u_{1j} ($j = 1 \div n$), l_{i1} ($i = 2 \div n$), u_{2j} ($j = 2 \div n$), l_{i2} ($i = 3 \div n$), ..., $u_{n-1,n-1}$, $u_{n-1,n}$, $l_{n,n-1}$ i u_{nn} .

Mètode de Crout

El *mètode de Crout* és semblant al de Doolittle amb la diferència que ara és la matriu U la que posseeix els elements diagonals iguals a 1:

$$l_{ik} = a_{ik} - \sum_{r=1}^{k-1} l_{ir} u_{rk} \quad (i = k \div n) ,$$

$$u_{kk} = 1, \quad u_{kj} = \frac{1}{l_{kk}} \left(a_{kj} - \sum_{r=1}^{k-1} l_{kr} u_{rj} \right) \quad (j = k+1 \div n) \\ (k = 1 \div n). \quad (2.8)$$

Els elements l_{kk} ($k = 1 \div n$) seran no nuls, sota les mateixes hipòtesis que abans.

L'ordre de càlcul és: l_{i1} ($i = 1 \div n$), u_{1j} ($j = 2 \div n$), l_{i2} ($i = 2 \div n$), u_{2j} ($j = 3 \div n$), $\dots$, $l_{n-1,n-1}$, $l_{n,n-1}$, $u_{n-1,n}$, l_{nn} .

2.2.4 Mètodes d'ortogonalització

Els *mètodes d'ortogonalització* assoleixen la reducció a forma triangular mitjançant l'ús de matrius ortogonals. Aquests mètodes estan basats en factoritzacions QR de matrius, que es defineixen tot seguit, i reben també el nom de *mètodes QR*.

Hi observem també dues fases:

- En la primera, es factoritza A en la forma:

$$A = QR,$$

on Q és ortogonal i R triangular superior. Aquesta factorització s'anomena *factorització QR* de la matriu A .

- En la segona, s'acaba la resolució del sistema $Ax = QRx = b$, resolent el sistema triangular superior $Rx = Q^T b$ pel mètode de substitució cap al darrera.

La dita factorització es fa emprant algun dels mètodes QR següents que poden ser generalitzats, de forma natural, a matrius no quadrades (vegeu l'apartat 3.2.3).

Mètode d'ortogonalització modificat de Gram-Schmidt

El *mètode d'ortogonalització modificat de Gram-Schmidt* és una millora, des d'un punt de vista numèric, del mètode d'ortogonalització clàssic de Gram-Schmidt (vegeu tots els detalls en l'apartat 3.2.3).

Començant per $A_1 = A$, un cop coneguda

$$A_k = (q_1 \dots q_{k-1} \ a_k^{(k)} \dots a_n^{(k)})$$

amb columnes q_j ($j = 1 \div k-1$) i $a_s^{(k)}$ ($s = k \div n$) complint les relacions d'ortogonalitat

$$q_j^T q_l = \delta_{jl}, \quad q_j^T a_s = 0 \quad (j, l = 1 \div k-1, \ s = k \div n),$$

es normalitza la k -ena columna i s'ortogonalitzen, respecte a ella, totes les que la segueixen:

$$r_{kk} = \|a_k^{(k)}\|_2, \quad q_k = \frac{a_k^{(k)}}{r_{kk}}; \\ r_{ks} = q_k^T a_s^{(k)}, \quad a_s^{(k+1)} = a_s^{(k)} - r_{ks} q_k \\ (s = k+1 \div n). \quad (2.9)$$

$$(2.10)$$

S'obté

$$A_{k+1} = (q_1 \ \dots \ q_k \ a_{k+1}^{(k+1)} \ \dots \ a_n^{(k+1)}) ,$$

verificant les mateixes condicions d'ortogonalitat anteriors, substituint k per $k+1$ ($k = 1 \div n$).

Al cap de n passos,

$$A_{n+1} = (q_1 \ q_2 \ \dots \ q_n)$$

és una matriu ortogonal, perquè $q_j^\top q_l = \delta_{jl}$ ($j, l = 1 \div n$).

Les matrius $Q = A_{n+1}$ i $R = (r_{ks})$ formen una factorització QR d' A . Com que A és regular, aquesta és única si imposem que els elements diagonals de R siguin positius.

Matrius de Householder

Abans de començar a descriure el mètode d'ortogonalització de Householder, presentarem les matrius de Householder, que seran útils també en altres apartats.

Per a cada vector no nul u de $\mathbb{R}^n$, definim la *matriu de Householder* $P(u)$ associada a u per

$$P(u) \equiv I - \frac{2}{u^\top u} uu^\top .$$

Aquestes matrius satisfan les propietats següents:

- $P(cu) = P(u)$, si $c \neq 0$. En altres paraules, les matrius de Householder estan realment associades a *direccions* o subspais unidimensionals més que a vectors no nuls.
- $P(u)$ és simètrica: $P(u)^\top = P(u)$.
- Si denotem per $u^\perp$ el subspai ortogonal a u :

$$u^\perp = \{v \in \mathbb{R}^n : u^\top v = 0\} ,$$

aleshores $P(u)v = v \ \forall v \in u^\perp$. Com que $P(u)u = -u$, resulta que $P(u)$ representa una simetria (o reflexió) respecte a l'hiperplà $u^\perp$. És així que tot $x \in \mathbb{R}^n$ es pot descompondre com

$$x = cu + v$$

amb

$$c = \frac{u^\top x}{u^\top u} , \quad v = x - cu \in u^\perp ,$$

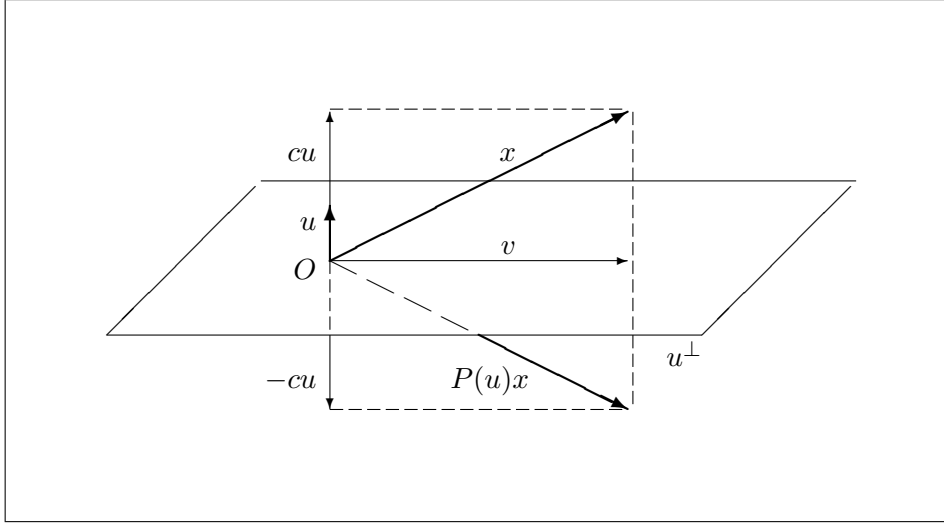
i

$$P(u)x = -cu + v$$

(vegeu la figura 2.2).

- $P(u)P(u) = I$ o $P(u)^{-1} = P(u) = P(u)^\top$. En particular, $P(u)$ és ortogonal i, per tant, preserva la norma euclidiana:

$$\|P(u)x\|_2^2 = x^\top P(u)^\top P(u)x = x^\top P(u)P(u)x = x^\top x = \|x\|_2^2 .$$

Figura 2.2: $P(u)x$ és la reflexió de x respecte a l'hiperplà $u^\perp$.

- $\det P(u) = -1$. Això és conseqüència del fet que $P(u)$ és una reflexió.
- Donat un vector $x \in \mathbb{R}^n$, el càlcul de $P(u)x$ es porta a terme molt fàcilment de la forma següent: si prèviament es calcula

$$\alpha = \frac{2}{u^\top u} ,$$

aleshores

$$P(u)x = x - (\alpha(u^\top x))u$$

només requereix $2n + 1$ operacions.

- Si $\|a\|_2 = \|b\|_2$, $a \neq b$, aleshores $P(a - b)a = b$. Això respon a la pregunta següent: donats $a, b \in \mathbb{R}^n$, existeix $u \neq 0$ tal que $P(u)a = b$? Com que $P(u)$ és ortogonal, cal suposar $\|a\|_2 = \|b\|_2$. Escrivint

$$a = cu + v$$

amb

$$c = \frac{u^\top a}{u^\top u} , \quad v = a - cu \in u^\perp ,$$

$P(u)a = b$ equival a $b = -cu + v$ i, per tant, $a - b = 2cu$; és a dir, u és paral·lel al vector $a - b$, que suposem no nul. Així doncs, podem escollir $u = a - b$. Noteu que u està determinat llevat d'una constant multiplicativa no nul·la.

- En particular, si $a \in \mathbb{R}^n$ no té les $n-1$ darreres components nul·les: $a_2^2 + a_3^2 + \dots + a_n^2 > 0$ o, en altres paraules, a no és de la forma $ce^{(1)}$, aleshores

$$P(a + \|a\|_2 e^{(1)})a = -\|a\|_2 e^{(1)} , \quad P(a - \|a\|_2 e^{(1)})a = \|a\|_2 e^{(1)} ;$$

o sigui, podem trobar dues matrius de Householder que transformen a en un vector molt simple, amb totes les components nul·les, llevat de la primera: un múltiple del primer element de la base canònica denotat per $e^{(1)} = (1 \ 0 \ \dots \ 0)^\top$.

Més breument: si $s = \|a\|_2$ o $s = -\|a\|_2$, $P(a - se^{(1)})a = se^{(1)}$.

Per tal d'evitar cancel·lacions, molt sovint s'escull s amb el signe canviat del d' a_1 .

El càlcul amb aquest tipus de matrius es fa com s'exposa tot seguit.

CÀLCUL DE $P(u) = P(a - se^{(1)})$:

Si $a = (a_1, a_2, \dots, a_n)^\top$ amb $a_2^2 + \dots + a_n^2 \neq 0$, es calcula:

- s tal que

$$s^2 = \|a\|_2^2 = \sum_{i=1}^n a_i^2.$$

Així, $s = \pm\|a\|_2$.

- $u_1 = a_1 - s$, $u_i = a_i$ ($i = 2 \div n$). Per tal d'evitar cancel·lacions en u_1 , se sol escollir $s = \|a\|_2$ si $a_1 \leq 0$ i $s = -\|a\|_2$ si $a_1 > 0$.

-

$$\alpha = \frac{2}{u^\top u} = \frac{1}{\|a\|_2(\|a\|_2 + |a_1|)} = \frac{-1}{su_1}.$$

Aleshores $P(u) = P(a - se^{(1)}) = I - \alpha uu^\top$ satisfà $P(u)a = se^{(1)}$, on el càlcul d' u , α requereix $n + 2$ operacions i una arrel quadrada. Una vegada tenim u i α podem afrontar diversos càlculs.

CÀLCUL DE $P(u)x$:

$P(u)x = x - (\alpha(u^\top x))u$ requereix $2n + 1$ operacions.

CÀLCUL DE $B = P(u)A$:

Es calcula

- $w = \alpha u$.
- $p \in \mathbb{R}^n$ tal que $p^\top = w^\top A$ (és a dir, $p = A^\top w$).
- $B = P(u)A = (I - \alpha uu^\top)A = A - up^\top$; en components,

$$b_{ij} = a_{ij} - u_i p_j \quad (i, j = 1 \div n).$$

El càlcul requereix $2n^2 + n$ operacions.

CÀLCUL D' $A' = P(u)AP(u)$:

$A' = P(u)AP(u) = BP(u) = B - (\alpha Bu)u^\top$. Es calcula:

- $w = \alpha u$, $p = A^\top w$ i $B = A - up^\top$ segons el càlcul anterior.
- $q = Bw$.
- $A' = B - qu^\top$; en components,

$$a'_{ij} = b_{ij} - q_i u_j \quad (i, j = 1 \div n).$$

El càlcul requereix, en total, $4n^2 + n$ operacions.

CÀLCUL D' $A' = P(u)AP(u)$, A SIMÈTRICA:

Només cal conèixer a'_{ij} per a $i \leq j$; per exemple, calculant $w = \alpha u$ i $p = A^\top w$ en primer lloc. Ara no cal calcular B :

$$q = Bw = P(u)Aw = P(u)A^\top w = P(u)p = p - (w^\top p)u$$

es calcula amb $2n$ operacions, i

$$A' = A - up^\top - qu^\top ;$$

en components,

$$a'_{ij} = a_{ij} - u_i p_j - q_i u_j \quad (1 \leq i \leq j \leq n) .$$

Així, el càlcul total requereix $2n^2 + 4n$ operacions.

De fet, l'expressió per a A' es pot escriure de manera simètrica introduint el vector

$$q^{(s)} = \frac{q + p}{2} = p - \frac{1}{2}(w^\top p)u ,$$

i procedint als càlculs:

- $w = \alpha u$.
- $p = Aw$.
- $q^{(s)} = p - \frac{1}{2}(w^\top p)u$.
- $A' = A - uq^{(s)\top} - q^{(s)}u^\top$; en components,

$$a'_{ij} = a_{ij} - u_i q_j^{(s)} - q_i^{(s)} u_j \quad (1 \leq i \leq j \leq n) .$$

Els procediments de càlcul que s'acaben de detallar seran molt útils a l'apartat que segueix, a l'apartat 2.3.2 en tractar la deflació de matrius i a l'apartat 2.3.5 en tractar els mètodes de reducció de matrius per al càlcul de llurs valors i vectors propis.

Mètode d'ortogonalització de Householder

El *mètode d'ortogonalització de Householder* per trobar una factorització QR d'una matriu A amb n files i n columnes consta de $n - 1$ passos. El mètode s'inicia amb $A_1 = A$ i, en el pas k -è ($k = 1 \div n - 1$), es parteix de la matriu trobada A_k de la forma

$$A_k = \left(\begin{array}{ccc|cc} \dots & \dots & \dots & \dots & \dots \\ & & \ddots & \dots & \dots \\ & & & \ddots & \dots \\ \hline & & & & \tilde{A}_k \end{array} \right)$$

$$= \left(\begin{array}{c|c} R_k & M_k \\ \hline 0 & \tilde{A}_k \end{array} \right),$$

on R_k és una matriu $(k-1) \times (k-1)$ triangular superior, i M_k , $\tilde{A}_k$ són matrius $(k-1) \times (n-k+1)$ i $(n-k+1) \times (n-k+1)$, respectivament, essent

$$\tilde{a}^{(k)} = \begin{pmatrix} a_{kk}^{(k)} \\ a_{k+1,k}^{(k)} \\ \vdots \\ a_{nk}^{(k)} \end{pmatrix} \in \mathbb{R}^{n-k+1}$$

la primera columna d' $\tilde{A}_k$.

Si les darreres $n-k$ components d' $\tilde{a}^{(k)}$ són nul·les

$$\sum_{i=k+1}^n (a_{ik}^{(k)})^2 = 0,$$

la matriu formada per les k primeres files de les k primeres columnes d' A_k és ja triangular superior, i prenem $P_k = I$, $A_{k+1} = P_k A_k = A_k$.

En cas contrari, escollim s_k tal que

$$s_k^2 = \|\tilde{a}^{(k)}\|_2^2 = \sum_{i=k}^n (a_{ik}^{(k)})^2;$$

per exemple, $s_k = \|\tilde{a}^{(k)}\|_2$, si $a_{kk}^{(k)} \leq 0$, i $s_k = -\|\tilde{a}^{(k)}\|$, si $a_{kk}^{(k)} > 0$. Prenem llavors la matriu $(n-k+1) \times (n-k+1)$ de Householder $\tilde{P}_k$ tal que

$$\tilde{P}_k \tilde{a}^{(k)} = \begin{pmatrix} s_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = s_k e^{(1)} \in \mathbb{R}^{n-k+1};$$

és a dir,

$$\tilde{P}_k = \tilde{P}(\tilde{u}^{(k)}) = I_{n-k+1} - \alpha_k \tilde{u}^{(k)} \tilde{u}^{(k)\top},$$

amb

$$\begin{aligned} \tilde{u}^{(k)} &= \begin{pmatrix} a_{kk}^{(k)} - s_k \\ a_{k+1,k}^{(k)} \\ \vdots \\ a_{nk}^{(k)} \end{pmatrix} = \tilde{a}^{(k)} - s_k e^{(1)} \in \mathbb{R}^{n-k+1}, \\ \alpha_k &= \frac{-1}{s_k(a_{kk}^{(k)} - s_k)} = \frac{1}{s_k(s_k - a_{kk}^{(k)})}. \end{aligned}$$

Definint aleshores

$$P_k = \left(\begin{array}{c|c} I_{k-1} & 0 \\ \hline 0 & \tilde{P}_k \end{array} \right),$$

o equivalentment $P_k = P(u^{(k)})$, amb

$$u^{(k)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \tilde{u}^{(k)} \end{pmatrix},$$

resulta que

$$\begin{aligned} A_{k+1} &\equiv P_k A_k = \left(\begin{array}{c|c} R_k & M_k \\ \hline 0 & \tilde{P}_k \tilde{A}_k \end{array} \right) \\ &= \left(\begin{array}{c|c|c} R_k & & \tilde{M}_k \\ \hline 0 & \cdots & 0 & s_k & \cdots & \cdots \\ \hline & & 0 & & & \\ 0 & & \vdots & & & \tilde{A}_{k+1} \\ & & 0 & & & \end{array} \right) \\ &= \left(\begin{array}{c|c} R_{k+1} & M_{k+1} \\ \hline 0 & \tilde{A}_{k+1} \end{array} \right); \end{aligned}$$

és a dir, la matriu R_{k+1} , formada per les k primeres files de les k primeres columnes d' A_{k+1} , és triangular superior i té la matriu $(n-k) \times k$ nul·la sota ella.

La matriu $R \equiv R_n = A_n$ és, doncs, triangular superior i es té

$$R = P_{n-1} P_{n-2} \cdots P_1 A$$

amb P_k ($k = 1 \div n-1$) matrius de Householder o matrius identitat i, per tant, complint $P_k = P_k^\top = P_k^{-1}$; d'aquí es dedueix la factorització

$$A = QR, \quad Q = P_1 P_2 \cdots P_{n-1}.$$

Els càlculs per a $\tilde{B}_k \equiv \tilde{P}_k \tilde{A}_k$ requereixen $2(n-k+1)^2 + \mathcal{O}(n-k)$ operacions i una arrel quadrada.

Sumant per a $k = 1 \div n-1$, cal efectuar $\frac{2}{3}n^3 + \mathcal{O}(n^2)$ operacions i $n-1$ arrels quadrades per tenir

$$R = P_{n-1}P_{n-2} \cdots P_1 A .$$

El sistema $Ax = b$ equival a

$$Rx = z \equiv P_{n-1}P_1P_2b .$$

El càlcul de z requereix aleshores $2n + 2(n-1) + \cdots + 2 \cdot 2 = n^2 + \mathcal{O}(n)$ operacions i la resolució del sistema $Rx = z$, $\frac{1}{2}n^2 + \mathcal{O}(n)$. Hem de fer el doble d'operacions que amb el mètode de Gauss; però, emprant un mètode QR, obtenim en general un sistema triangular superior més ben condicionat, com s'observarà en l'apartat II.2.6.

2.2.5 Càlcul de determinants i inverses de matrius

Determinants

Les transformacions que experimenta una matriu A al llarg del procés d'eliminació gaussiana amb pivotatge consisteixen a fer combinacions lineals de files de manera que, llevat del signe, que pot variar en permutar files i permutar columnes, es preserva el determinant. És així que

$$\det A = \sigma \det A_n = \sigma \prod_{k=1}^n a_{kk}^{(k)} ,$$

on σ val $+1$ o -1 i A_n és la matriu triangular final que té els elements $a_{kk}^{(k)}$ a la diagonal.

Cal fer dues observacions:

- El valor de σ serà $+1$ o -1 segons s'hagi fet un nombre parell o senar, respectivament, d'intercanvis entre files i entre columnes.
- Podria donar-se el cas que, en el k -è pas de l'eliminació gaussiana, $a_{ik}^{(k)} = 0$ ($k \leq i \leq n$). Aquesta circumstància clouria el càlcul del determinant: $\det A = 0$.

El mètode d'ortogonalització de Householder ofereix una altra tècnica de càlcul de determinants

$$\det A = \det P_1P_2 \cdots P_{n-1}R = \sigma \det R ,$$

on $\det R$ és el producte dels elements diagonals de R i σ és el producte dels determinants de les matrius P_k ($k = 1 \div n-1$). Com que la matriu identitat té determinant $+1$ i les matrius de Householder, -1 ; σ valdrà $+1$ o -1 segons que el nombre de matrius de Householder emprades sigui parell o senar.

Inverses

Donada una matriu $n \times n$ regular A , la matriu $X = A^{-1}$ compleix el sistema d'equacions matricial $AX = I$. Atenent a la columna k -ena d'aquesta equació, tindrem que

$$Ax^{(k)} = e^{(k)} \quad (k = 1 \div n) ,$$

on $e^{(k)}$ indica el k -è vector de la base canònica i $x^{(k)}$, la k -ena columna de la matriu X .

Així doncs, el càlcul d' A^{-1} quedarà reduït a la resolució de n sistemes amb una matriu comuna i amb termes independents molt senzills.

Aquests fets comporten un avantatge de càlcul: si s'utilitza el mètode de Gauss, es podrà realitzar la fase d'eliminació simultàniament en els n sistemes, situant la matriu identitat I en el lloc de la columna de termes independents. Una generalització del procediment que s'acaba de descriure dona lloc al *mètode de Gauss-Jordan*, que parteix dels sistemes d'equacions $Ax^{(k)} = e^{(k)}$ ($k = 1 \div n$) escrits en forma matricial

$$(A|I) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & \cdots & 1 \end{array} \right),$$

i va modificant les files de $(A|I)$, usant els elements de la diagonal com a pivots (com en el mètode de Gauss) i aconseguint que es vagin anul·lant els elements no diagonals (també els de sobre la diagonal!), columna a columna. S'obté, finalment, $(D|B)$ amb D diagonal, llavors $A^{-1} = D^{-1}B$.

Alternativament, l'ús de les factoritzacions LU i QR ens ofereix també la possibilitat de calcular la inversa segons:

$$A^{-1} = U^{-1}L^{-1}, \quad A^{-1} = R^{-1}Q^T.$$

Hi ha altres procediments per calcular la inversa d'una matriu A . En el problema II.9 se'n descriu un que es basa en la construcció d'una successió de matrius que, sota certes condicions, convergeix cap a la matriu inversa A^{-1} . En els problemes proposats se'n presenten altres mètodes.

2.2.6 Anàlisi de l'error

En resoldre un sistema $Ax = b$, A i b apareixen com a dades i x com a solució. Els errors de x poden provenir de:

- la propagació dels errors de les dades A i b ,
- l'acumulació dels errors d'arrodoniment en els càlculs produïts en el decurs del procés de resolució.

Propagació dels errors de les dades

Suposem que no disposem exactament de la matriu A i del vector b , sinó d'una matriu $A + \delta A$ i d'un vector $b + \delta b$. En aquestes condicions, no obtindrem la solució cercada x del sistema $Ax = b$, sinó una solució $x + \delta x$ del sistema pertorbat

$$(A + \delta A)(x + \delta x) = b + \delta b.$$

Disposem de la fórmula següent per tal de fitar, de forma relativa, una norma qualsevol del vector d'error δx

$$\frac{\|\delta x\|}{\|x\|} \leq \mu(A) \left[\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \frac{\|x + \delta x\|}{\|x\|} \right],$$

sobre la qual cal fer les observacions següents:

- En general i a efectes pràctics, $\|x + \delta x\| \simeq \|x\|$.
- $\mu(A) = \|A\| \|A^{-1}\|$ s'anomena *nombre de condició* de la matriu A i representa el factor màxim d'ampliació dels errors relatius d' A i b . Parlarem de matrius *mal condicionades* quan tinguin nombres de condició excessivament grans.

El problema II.5 presenta una fitació de l'error relatiu de x , sota unes condicions particulars, i posa de manifest, amb un exemple concret, les conseqüències del mal condicionament de la matriu sobre l'amplificació dels errors de les dades.

- Les normes matricials emprades en la fórmula han de ser consistentes amb les normes vectorials.
- Si U és una matriu ortogonal (i, per tant, unitària) tenim que, emprant el nombre de condició euclidià $\mu_2(A) = \|A\|_2 \|A^{-1}\|_2$,

$$\mu_2(UA) = \mu_2(A) ;$$

així, en el mètode QR, la matriu R del sistema reduït $Rx = Q^T b$ té el mateix nombre de condició que la matriu A de partida. Es té, doncs, la garantia que el condicionament del sistema triangular no empitjora, cosa que no pot assegurar-se en general per als mètodes gaussians.

Al problema II.3 es detalla un estudi de propagació d'errors en un procés de factorització de Txoleski.

Acumulació dels errors d'arrodoniment

En resoldre el sistema $Ax = b$, per culpa dels errors d'arrodoniment durant el procés, no obtenim la solució exacta x , sinó una solució aproximada $x + \delta x$. Per tal de fitar l'error δx (en alguna norma), es pot fer servir l'anàlisi de l'error cap al darrera, tractant de cercar primerament δA tal que

$$(A + \delta A)(x + \delta x) = b .$$

Es redueix, doncs, el problema a un d'equivalent en què es culpa els elements de la matriu A dels errors d'arrodoniment; després caldrà aplicar la fórmula de propagació dels errors de les dades acabada de trobar per tal de fitar l'error en la solució x , atès que no cal considerar errors en les operacions.

La matriu δA depèn, evidentment, del mètode de resolució que se segueixi. En concret, si el mètode fos el de factorització LU, es pot arribar a la fita següent per a l'error carregat a la matriu A després de fer l'anàlisi de l'error cap al darrera

$$\|\delta A\|_\infty \leq (n^3 + 3n^2) \bar{g}_n \epsilon \|A\|_\infty ,$$

on n és la dimensió del sistema, ϵ és la fita de l'error relatiu en les operacions aritmètiques (vegeu l'apartat 1.1.2) i

$$\bar{g}_n \equiv \frac{\max_{i,j,k} |\bar{a}_{ij}^{(k)}|}{\max_{i,j} |a_{ij}|}$$

indica el factor màxim de creixement del mòdul dels elements calculats $\bar{a}_{ij}^{(k)}$ pel mètode d'eliminació gaussiana respecte al màxim dels mòduls dels elements de la matriu A de partida.

Si definim g_n de forma anàloga, però considerant els coeficients $a_{ij}^{(k)}$ sense els errors acumulats, es poden obtenir les fites següents:

- $g_n \leq 2^{n-1}$, usant pivotatge maximal per columnes;
- $g_n \leq 3n^{\frac{2+\ln n}{4}}$, usant pivotatge complet.

Aquestes fites són molt pessimistes en la major part de les aplicacions pràctiques; això és, l'error real és molt més petit que la fita obtinguda.

Cal però observar que, si no s'usa pivotatge en la resolució dels sistemes lineals no es té generalment cap control sobre la grandària del factor $\bar{g}_n$; aquest pot arribar a ser molt gran si, durant el procés, ens trobem amb pivots propers a zero. Aquests faran créixer molt els elements calculats $\bar{a}_{ij}^{(k)}$ i consegüentment l'error en la solució.

2.2.7 Mètodes iteratius

Generalitats

Els mètodes iteratius estan basats en la idea següent: donat un sistema lineal $Ax = b$, li associem d'alguna manera determinada un sistema lineal equivalent $x = Bx + c$ que resollem de forma iterativa. Això és, comencem amb $x^{(0)}$ arbitrari i obtenim després $x^{(1)}$, $x^{(2)}$, ... mitjançant la recurrència

$$x^{(k+1)} = Bx^{(k)} + c \quad (k \geq 0) ;$$

si la successió de vectors $x^{(k)}$ ($k \geq 0$) és convergent a un vector x , aquest serà la solució del sistema $x = Bx + c$ i, per tant, del sistema de partida $Ax = b$. La matriu B s'acostuma a anomenar *matriu d'iteració*.

Analitzant el comportament dels errors $e^{(k)} \equiv x^{(k)} - x$ quan k tendeix a infinit, i tenint en compte les propietats de les normes vectorials i matricials donades en l'apartat II.1.3, s'arriba als resultats següents sobre la convergència dels mètodes iteratius:

- El mètode iteratiu $x^{(k+1)} = Bx^{(k)} + c$ ($k \geq 0$) és convergent, independentment de l'elecció de $x^{(0)}$, si i només si el radi espectral de la matriu d'iteració $\rho(B)$ és menor que 1. El teorema de Gershgorin, presentat a l'apartat 2.1.3, és una bona eina per a la fitació de $\rho(B)$ (vegeu el problema II.11).
- El mètode iteratiu anterior és, doncs, convergent si i només si $\|B\| < 1$ per a alguna norma matricial consistent amb alguna norma vectorial (per exemple, les normes matricials subordinades $\| \cdot \|_1$ i $\| \cdot \|_\infty$).

Descriurem a continuació els mètodes iteratius més coneguts.

Mètode iteratiu de Jacobi

Suposant que els elements diagonals d' A són no nuls, si dividim cada fila per l'element de la diagonal corresponent, queda una matriu amb uns a la diagonal. Podem imaginar aquesta matriu formada per la suma d'una matriu triangular inferior L amb zeros a la diagonal, de la matriu identitat I i d'una matriu triangular superior U amb zeros a la diagonal; així queda descomposta la matriu A com a producte de la seva diagonal D per la suma de L , I i U ,

$$A = D(L + I + U) .$$

El sistema $Ax = b$ és llavors equivalent al sistema

$$x = -(L + U)x + D^{-1}b .$$

Noteu que aquest sistema equivalent s'obté simplement aïllant en l'equació i -ena la variable x_i ($i = 1 \div n$).

Aquest sistema permet de definir el *mètode iteratiu de Jacobi*

$$x^{(k+1)} = -(L + U)x^{(k)} + D^{-1}b ,$$

que, en components, pren la forma

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j \neq i} a_{ij} x_j^{(k)} \right) \quad (i = 1 \div n) .$$

Mètode iteratiu de Gauss-Seidel

Igualment com el mètode iteratiu de Jacobi, només serà aplicable a sistemes $Ax = b$ en què els elements diagonals d' A siguin no nuls i està basat en l'equivalència amb el sistema

$$x = -(I + L)^{-1}Ux + (I + L)^{-1}D^{-1}b .$$

Així, es defineix el *mètode iteratiu de Gauss-Seidel* mitjançant

$$x^{(k+1)} = -(I + L)^{-1}Ux^{(k)} + (I + L)^{-1}D^{-1}b ,$$

que equival a

$$x^{(k+1)} = -Lx^{(k+1)} - Ux^{(k)} + D^{-1}b ,$$

i que, en components, s'escriu

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right) \quad (i = 1 \div n) .$$

Els dos mètodes que s'acaben de descriure són molt semblants, atès que la fórmula d'avaluació de la component i -ena del nou vector és pràcticament la mateixa: la diferència rau en el fet que, en el mètode iteratiu de Jacobi, la dita component es calcula a partir de les components del vector anterior de components $x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_{i+1}^{(k)}, \dots, x_n^{(k)}$, i en el mètode iteratiu de Gauss-Seidel es prenen les anteriors components del nou vector ja calculades: $x_1^{(k+1)}, \dots, x_{i-1}^{(k+1)}, x_{i+1}^{(k)}, \dots, x_n^{(k)}$.

2.2.8 Mètodes iteratius de sobrerelaxació

Es tracta d'una família de mètodes que generalitzen el mètode iteratiu de Gauss-Seidel. Estan basats en l'equivalència d' $Ax = b$ amb

$$x = x - \omega(Lx + (I + U)x - D^{-1}b) ,$$

on ω és un paràmetre real, anomenat *factor de relaxació*. Aquest sistema també és equivalent al sistema

$$x = B_{\omega}x + c_{\omega} ,$$

amb

$$B_{\omega} = (I + \omega L)^{-1}[(1 - \omega)I - \omega U] , \quad c_{\omega} = \omega(I + \omega L)^{-1}D^{-1}b .$$

Es defineix el *mètode iteratiu de sobrerelaxació amb factor ω* per la recurrència

$$x^{(k+1)} = B_{\omega}x^{(k)} + c_{\omega} ,$$

que també pot escriure's com

$$x^{(k+1)} = x^{(k)} - \omega(Lx^{(k+1)} + (I + U)x^{(k)} - D^{-1}b) .$$

En components, pren la forma

$$x_i^{(k+1)} = x_i^{(k)} + \frac{\omega}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i}^n a_{ij}x_j^{(k)} \right) \quad (i = 1 \div n) .$$

Cal observar que, per a $\omega = 1$, es retroba el mètode iteratiu de Gauss-Seidel.

Els mètodes de sobrerelaxació que s'acaben de descriure es basen en una transformació del sistema $Ax = b$ en un altre d'equivalent del tipus $x = B(\omega)x + c(\omega)$, sobre el qual es defineix la recurrència. La construcció del sistema $x = B(\omega)x + c(\omega)$ es pot fer de diverses maneres que donen lloc a nous mètodes iteratius. En el problema II.8 se'n defineix un i s'estudia la seva convergència.

Convergència dels mètodes iteratius de Jacobi, de Gauss-Seidel i de sobrerelaxació

D'acord amb el que s'ha dit abans, els mètodes iteratius de Jacobi, Gauss-Seidel i sobrerelaxació convergiran si i només si els radis espectrals de llurs matrius d'iteració ($B_J = -(L + U)$, $B_{GS} = -(I + L)^{-1}U$ i B_{ω}) són menors que 1. En el problema II.6 s'estudia la convergència dels mètodes de Jacobi i de Gauss-Seidel en aplicar-los a un sistema lineal de dues equacions i dues incògnites.

A continuació es donen alguns resultats relacionats amb la convergència d'aquests mètodes en casos particulars i que no requereixen el càlcul del radi spectral:

- Si A és estrictament diagonal dominant, els mètodes iteratius de Jacobi i de Gauss-Seidel convergeixen.

La rapidesa amb què es produeix aquesta convergència depèn de les característiques del sistema. En el problema II.7 es presenta una estimació del nombre d'iteracions del mètode iteratiu de Jacobi que fan falta per tal de poder garantir una fita d'error donada en resoldre un sistema on, entre altres condicions, la matriu se suposa estrictament diagonal dominant.

- Els mètodes iteratius de sobrerelaxació divergeixen per a factors ω que no pertanyen a l'interval $(0, 2)$.
- Si A és hermitica amb elements positius a la diagonal, els mètodes iteratius de sobrerelaxació amb factors $\omega \in (0, 2)$ convergeixen si i només si la matriu A és definida positiva.

El mètode iteratiu de sobrerelaxació té l'avantatge que permet de modificar el factor ω de forma que pugui assolir-se una convergència més ràpida per a certs tipus de matrius; per exemple, es disposa del resultat següent en aquest sentit:

- Per a una matriu A , definida positiva i tridiagonal per blocs i tal que els blocs de la diagonal siguin matrius diagonals, tenim que

$$\rho(B_{GS}) = \rho(B_J)^2 .$$

Si, a més, els valors propis de B_J són a $(-1, 1)$; llavors, per a

$$\omega = \tilde{\omega} = \frac{2}{1 + \sqrt{1 - \rho(B_J)^2}} ,$$

tenim que

$$\rho(B_{\tilde{\omega}}) = \tilde{\omega} - 1 = \left(\frac{\rho(B_J)}{1 + \sqrt{1 - \rho(B_J)^2}} \right)^2 \leq \rho(B_{\omega}) \quad \forall \omega .$$

$\tilde{\omega}$ és el valor de ω que minimitza el radi espectral de les matrius B_{ω} ; per tant, el mètode iteratiu de sobrerelaxació corresponent serà el més ràpidament convergent. Per aquesta raó, $\tilde{\omega}$ s'anomena *factor òptim de relaxació*.

2.3 VALORS I VECTORS PROPIS

2.3.1 Introducció

El problema de cerca de valors i vectors propis d'una matriu és, de fet, un problema no lineal, però té una forta component lineal. Així, el càlcul de valors propis es redueix a cercar els zeros d'un polinomi i pot ser portat a terme mitjançant els mètodes generals de cerca de zeros de polinomis (que seran presentats en el capítol V), o bé pels mètodes iteratius lineals explicats en aquest capítol. Un cop coneguts els valors propis, el càlcul dels vectors propis sí que és un problema lineal.

En general, però, és excessivament laboriós el càlcul directe del polinomi característic d'una matriu. És per aquesta raó que el càlcul efectiu de valors i vectors propis no passa per l'obtenció del dit polinomi, excepte quan l'especial naturalesa de la matriu ho permet: matrius de dimensions petites, tridiagonals, de Frobenius (vegeu els problemes II.10 i II.11), etc.

El concepte de deflació de matrius és semblant al de deflació de polinomis (vegeu l'apartat 5.2.2) i permet de simplificar el problema de trobar valors i vectors propis, un cop coneguts alguns d'ells. Es descriuen tot seguit els mètodes de Hotteling, Wieland i de Householder.

Els mètodes de càlcul de valors i vectors propis són de diversa naturalesa: uns, com són els mètodes iteratius de la potència, de Jacobi, LR i QR, calculen els valors propis mitjançant tècniques iteratives; d'altres, com els de Givens i Householder, redueixen la matriu inicial a una matriu semblant Hessenberg superior (tridiagonal simètrica, si la matriu és simètrica), en un nombre finit de passos. Per a aquests tipus de matrius reduïdes, es donen mètodes específics de càlcul dels valors i vectors propis que sí que passen per l'obtenció del polinomi característic. En el problema II.10 es descriu un mètode d'aquest tipus (el de Danilevski), que redueix la matriu inicial a la forma normal de Frobenius, el polinomi característic de la qual s'obté d'immediat.

2.3.2 Deflació de matrius

Partint del coneixement d'un valor propi λ i d'un vector propi associat v d'una matriu $n \times n$ A , el procés de *deflació* consisteix a trobar una nova matriu $\tilde{A}$ (de característiques més simples, sovint de dimensió més baixa) de manera que, trobant els valors i vectors propis d' $\tilde{A}$, es puguin trobar els d' A , amb facilitat.

Convé fer servir la deflació, alternant-la amb mètodes que permetin trobar els valors propis d'un a un, i així poder obtenir tots els valors propis treballant cada cop amb matrius més senzilles. A més, anàlogament al que passa amb la deflació polinomial, que s'estudiarà en el capítol V, és millor que els dits valors propis es vagin calculant des del més petit al més gran, a fi de guanyar estabilitat numèrica; els valors i vectors propis trobats al final del procés convé que siguin corregits tenint en compte la matriu de partida, per tal d'eliminar els errors acumulats en tot el procés de deflació.

Existeixen diversos tipus de deflació de matrius; esmentarem aquí els de Hotteling, Wielandt i Householder. Cal destacar els avantatges d'aquest últim, que emprava transformacions de semblança amb matrius ortogonals.

Deflació de Hotteling

Considerem coneguts un valor propi λ d' A i vectors propis associats v i u , per la dreta i per l'esquerra respectivament, satisfent $u^\top v \neq 0$; es construeix la matriu

$$A_{(H)} = A - \frac{\lambda}{u^\top v} v u^\top,$$

que té els mateixos valors propis que A , llevat que el valor propi λ ha estat substituït pel valor propi zero; els vectors propis són els mateixos.

Si A és simètrica, pot triar-se $u = v$.

Deflació de Wielandt

Conegut un vector propi v de valor propi λ d' A , amb $v_1 \neq 0$ (per exemple) i denotant per $a^{(1)\top}$ la primera filera d' A , la matriu

$$A_{(V)} = A - \frac{1}{v_1} v a^{(1)\top},$$

té la primera fila nul·la. Només cal ara cercar valors i vectors propis de la matriu $(n-1) \times (n-1)$, $\tilde{A}_{(V)}$, formada per les darreres $n-1$ fileres de les darreres $n-1$ columnes d' A .

Això es deu al fet que, si $\lambda' \neq \lambda$ és un valor propi d' $\tilde{A}_{(V)}$ amb vector propi (de $n - 1$ components) associat $\tilde{v}'$ i fem

$$w = \begin{pmatrix} 0 \\ \tilde{v}' \end{pmatrix}, \quad c = \frac{(\lambda' - \lambda)v_1}{a^{(1)\top}w},$$

aleshores $v' = v + cw$ és un vector propi d' A de valor propi λ' .

Deflació de Householder

És el tipus de deflació al qual prestem més atenció, atès que es porta a terme mitjançant matrius ortogonals; així s'aconsegueix un control de l'estabilitat numèrica i la conservació de la simetria de les matrius, quan s'escau.

Conegut un vector propi v de valor propi λ d' A , si v és de la forma

$$v = \begin{pmatrix} v_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

podem prendre $v = e^{(1)}$ i aleshores A és de la forma

$$A = \left(\begin{array}{c|c} \lambda & c^\top \\ \hline 0 & \tilde{A} \end{array} \right).$$

Si $\tilde{v}'$ és un vector propi d' $\tilde{A}$ de valor propi $\lambda' \neq \lambda$, aleshores

$$v' = \begin{pmatrix} b \\ \tilde{v}' \end{pmatrix} \quad \text{amb} \quad b = \frac{c^\top \tilde{v}}{\tilde{\lambda} - \lambda}$$

serà un vector propi de valor propi $\tilde{\lambda}'$ d' A .

Aquest resultat ens permet calcular els vectors i valors propis d' A , a partir dels d' $\tilde{A}$, de dimensió $n - 1$.

Si A és simètrica, $c = 0$ i $\tilde{A}$ és també simètrica i, si $\tilde{v}'$ és un vector propi d' $\tilde{A}$ de valor propi λ' , aleshores

$$v' = \begin{pmatrix} 0 \\ \tilde{v}' \end{pmatrix}$$

serà un vector propi de valor propi λ' d' A .

Retornant al cas general, si v és vector propi de valor propi λ d' A , amb $v_2^2 + \dots + v_n^2 \neq 0$, podem trobar matrius de Householder P tals que

$$Pv = se^{(1)}, \quad s = \pm \|v\|_2$$

(vegeu l'apartat 2.2.4 per al càlcul efectiu).

Com que $P = P^\top = P^{-1}$,

$$PAPe^{(1)} = \frac{1}{s}PAv = \frac{\lambda}{s}Pv = \lambda e^{(1)} ;$$

per tant, la matriu PAP és de la forma volguda:

$$A' = PAP = \left(\begin{array}{c|c} \lambda & c'^\top \\ \hline 0 & \tilde{A}' \end{array} \right) .$$

El vector v' és un vector propi de valor propi λ' d' A' si i només si Pv' és un vector propi de valor propi λ' d' A . Si A és simètrica, també A' ($c' = 0$) (vegeu el problema II.12).

2.3.3 Mètodes de la potència

Mètode de la potència

El mètode de la potència i les seves variants són molt útils quan només són requerits alguns valors propis i llurs vectors propis associats. També serveixen per calcular tots els valors propis de la matriu si es combinen amb algun dels tipus de deflació descrits anteriorment.

El mètode de la potència pròpiament dit permet calcular aproximacions successives del valor propi de mòdul màxim (si existeix), així com vectors propis associats a aquest.

Donada A , matriu $n \times n$, siguin λ_j ($j = 1 \div n$) els seus valors propis, repetits segons llur multiplicitat i ordenats segons llur mòdul

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n| .$$

Suposarem, per simplicitat, que A és diagonalitzable i anomenarem $v^{(j)}$ ($j = 1 \div n$) els vectors propis linealment independents associats als valors propis λ_j ($j = 1 \div n$), i també que existeix un únic valor propi de mòdul màxim λ_1 : $|\lambda_1| > |\lambda_2|$. Uns altres casos són tractats en els problemes proposats.

El *mètode de la potència* considera les iteracions

$$x^{(k+1)} = Ax^{(k)} \quad (k \geq 0)$$

per a algun vector

$$x^{(0)} = \sum_{j=1}^n \alpha_j v^{(j)} \quad \text{amb } \alpha_1 \neq 0 ;$$

així tenim

$$x^{(k)} = \lambda_1^k \left[\alpha_1 v^{(1)} + \alpha_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k v^{(2)} + \dots + \alpha_n \left(\frac{\lambda_n}{\lambda_1} \right)^k v^{(n)} \right] .$$

Si tenen sentit les successions de quocients

$$q_i^{(k)} = \frac{x_i^{(k+1)}}{x_i^{(k)}} \quad (i = 1 \div n) ;$$

és a dir, si $x_i^{(k)} \neq 0$, aquestes tenen per límit λ_1 ; a més, quan $k \rightarrow \infty$:

$$q_i^{(k)} = \lambda_1 \left[1 + \mathcal{O} \left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right) \right] \quad (i = 1 \div n) , \quad (2.11)$$

$$\frac{x^{(k)}}{\lambda_1^k} = \alpha_1 v^{(1)} \left[1 + \mathcal{O} \left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right) \right] . \quad (2.12)$$

Aquestes són expressions asimptòtiques per a λ_1 i per a algun vector propi associat a ell, tractades a l'apartat 4.3.1; la velocitat de convergència depèn essencialment de $|\frac{\lambda_2}{\lambda_1}|$ i pot ser molt petita, quan $|\lambda_2|$ sigui proper a $|\lambda_1|$.

Com que la successió $(\lambda_1^k)_{k \geq 0}$ generalment tendeix a zero o no està fitada, convé normalitzar d'alguna manera la successió dels vectors $x^{(k)}$. Un possible procediment és el següent: un cop tenim $x^{(k)}$, el normalitzem segons

$$y^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|}$$

(emprant una norma qualsevol) i fem $x^{(k+1)} = Ay^{(k)}$ ($k \geq 0$); llavors,

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{y_i^{(k)}} = \lambda_1 \quad (i = 1 \div n) , \quad \lim_{k \rightarrow \infty} y^{(k)} = \pm \frac{v^{(1)}}{\|v^{(1)}\|} .$$

A fi d'assegurar la convergència de les successions a λ_1 , s'ha hagut de suposar que la component α_1 de $x^{(0)}$, segons $v^{(1)}$, no era nul·la. En cas contrari i suposant, per exemple, que $\alpha_2 \neq 0$ i $|\lambda_2| > |\lambda_3|$, s'obtindria analíticament que

$$\lim_{k \rightarrow \infty} \frac{x_i^{(k+1)}}{x_i^{(k)}} = \lambda_2 .$$

Malgrat això, a la pràctica, els errors d'arrodoniment que es cometien durant el càlcul introdueixen en els $x^{(k)}$ una component no nul·la segons $v^{(1)}$, que implica la tendència de la successió a λ_1 , i no a λ_2 .

Per a matrius A simètriques, és més eficaç calcular la successió de *quocients de Rayleigh* de $x^{(k)}$

$$\sigma_k = \frac{x^{(k)\top} A x^{(k)}}{x^{(k)\top} x^{(k)}} \quad (k \geq 0)$$

emprant la recurrència

$$y^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|_2} , \quad x^{(k+1)} = Ay^{(k)} , \quad \sigma_k = y^{(k)\top} x^{(k+1)} \quad (k \geq 0) .$$

Aleshores obtenim la convergència de les successions anteriors:

$$\lim_{k \rightarrow \infty} \sigma_k = \lambda_1 , \quad \lim_{k \rightarrow \infty} y^{(k)} = \pm \frac{v^{(1)}}{\|v^{(1)}\|_2} ,$$

segons les expressions asimptòtiques, quan $k \rightarrow \infty$:

$$\begin{aligned}\sigma_k &= \lambda_1 \left[1 + \mathcal{O} \left(\left| \frac{\lambda_2}{\lambda_1} \right|^{2k} \right) \right], \\ y^{(k)} &= \pm \frac{v^{(1)}}{\|v^{(1)}\|_2} \left[1 + \mathcal{O} \left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right) \right].\end{aligned}\quad (2.13)$$

La utilització dels quocients de Rayleigh comporta un augment de la velocitat de convergència en el càlcul del valor propi, en ser

$$\left| \frac{\lambda_2}{\lambda_1} \right|^2 < \left| \frac{\lambda_2}{\lambda_1} \right|.$$

Algunes possibles variants del mètode de la potència són les indicades tot seguit.

Mètode de la potència desplaçada

Aplicant el mètode de la potència a la matriu $A - dI$, que té com a valors propis $\mu_j = \lambda_j - d$ amb vectors propis associats $v^{(j)}$ ($j = 1 \div n$), s'obté el valor propi μ_j de mòdul màxim d' $A - dI$ (sempre que sigui únic) i un vector pròpi associat $v^{(j)}$: permet, doncs, calcular el valor propi $\lambda_j = d + \mu_j$ d' A més allunyat de d .

Mètode de la potència inversa

Suposant ara que A és invertible, A^{-1} té els mateixos vectors propis que A amb valors propis μ_j que són els recíprocs dels d' A : $\mu_j = \frac{1}{\lambda_j}$ ($j = 1 \div n$). Si el valor propi de mòdul mínim és únic ($|\lambda_n| < |\lambda_{n-1}|$), l'aplicació del mètode de la potència a A^{-1} dóna el valor propi $\frac{1}{\lambda_n}$ d' A^{-1} i un vector propi associat: podem conèixer així el valor propi de mòdul mínim λ_n d' A .

Mètode de la potència inversa desplaçada

És una combinació de les dues variants anteriors. Consisteix a aplicar el mètode de la potència a $(A - dI)^{-1}$, que té per valors propis els $\mu_j = \frac{1}{\lambda_j - d}$ ($j = 1 \div n$). L'aplicació del mètode dóna el valor propi μ_j de mòdul màxim d'aquesta matriu, sempre que sigui únic, i un vector propi associat $v^{(j)}$: així podem trobar el valor propi $\lambda_j = d + \frac{1}{\mu_j}$ d' A més proper a d .

Aquesta variant és molt útil per refinar aproximacions d de valors propis, i per trobar vectors propis associats.

En el problema II.13 es presenta un exemple d'ús combinat del mètode de la potència amb el mètode de la potència inversa per al càlcul de tots els valors propis d'una matriu 4×4 .

2.3.4 Mètodes de Jacobi

Mètode de Jacobi

Sigui A una matriu $n \times n$ simètrica. Sabem, per la propietat 10 de valors i vectors propis en l'apartat 2.1.3, que existeix una matriu ortogonal O ($O^{-1} = O^T$) tal que $\Lambda = O^T A O$ és diagonal: $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$; la columna j -ena d' O és un vector propi d' A associat al valor propi λ_j i les dites columnes formen una base de vectors propis ortonormals d' A .

El *mètode de Jacobi* dóna precisament un procediment de cerca de successives aproximacions A_k i O_k de Λ i O , respectivament, on O_k és un producte de matrius ortogonals senzilles ($k \geq 1$). Es tracta, per tant, d'un mètode iteratiu, el mecanisme de construcció del qual es dóna a continuació.

Comencem el procés amb $A_1 = A$ i $O_1 = I$ i, coneguts $A_k \equiv B = (b_{ij})$ i $O_k \equiv \Omega = (\omega_{ij})$, triem en B un element no diagonal b_{pq} (amb $p < q$) complint $b_{pq} \neq 0$.

Els elements b_{pp} , b_{pq} , $b_{qp}(=b_{pq})$ i b_{qq} formen una matriu 2×2 simètrica. Cerquem una matriu de rotació

$$\begin{pmatrix} c & -s \\ s & c \end{pmatrix}, \quad c = \cos \varphi, \quad s = \sin \varphi,$$

de forma que la matriu simètrica

$$\begin{pmatrix} d_{pp} & d_{pq} \\ d_{qp} & d_{qq} \end{pmatrix} = \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix}$$

sigui diagonal.

Per això, l'angle φ ha de complir

$$\cot 2\varphi = \frac{b_{pp} - b_{qq}}{2b_{pq}}$$

i es pot triar φ complint $|\varphi| < \frac{\pi}{4}$.

Les fórmules que s'obtenen llavors per a c i s són:

$$\begin{aligned} \text{Si } b_{pp} = b_{qq} : \quad & \epsilon = \text{sgn}(b_{pq}), \quad c = \frac{\sqrt{2}}{2}, \quad s = \epsilon \frac{\sqrt{2}}{2}. \\ \text{Si } b_{pp} \neq b_{qq} : \quad & \epsilon = \text{sgn}[b_{pq}(b_{pp} - b_{qq})], \end{aligned} \tag{2.14}$$

$$\begin{aligned} c &= \sqrt{\frac{1}{2} \left(1 + \frac{|b_{pp} - b_{qq}|}{d} \right)}, \\ s &= \epsilon \sqrt{\frac{1}{2} \left(1 - \frac{|b_{pp} - b_{qq}|}{d} \right)}, \\ d &= \sqrt{(b_{pp} - b_{qq})^2 + 4b_{pq}^2}. \end{aligned} \tag{2.15}$$

Prenem ara la matriu

$$R_k \equiv R = \begin{pmatrix} 1 & & & & & & & & & \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & c & & & -s & & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & s & & & & & c & & \\ & & & & & & & & 1 & \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \end{pmatrix},$$

on els elements no escrits són nuls; aquesta matriu representa una rotació d'angle φ en el pla coordinat $p - q$.

Calculem ara $\bar{B} = R^\top B R = (\bar{b}_{ij})$ (canviant només les fileres i -ena i j -ena de B) i $\bar{\Omega} = \Omega R = (\bar{\omega}_{ij})$ (canviant només les columnes i -ena i j -ena de Ω) segons les fórmules:

$$\begin{aligned} \bar{b}_{pj} &= \bar{b}_{jp} = b_{pj}c + b_{qj}s \quad (j = 1 \div n, \quad j \neq p, q), \\ \bar{b}_{iq} &= \bar{b}_{qi} = -b_{ip}s + b_{iq}c \quad (i = 1 \div n, \quad i \neq p, q), \\ \bar{b}_{pp} &= b_{pp}c^2 + 2b_{pq}cs + b_{qq}s^2 = b_{pp} + b_{pq}t, \\ \bar{b}_{qq} &= b_{pp}s^2 - 2b_{pq}cs + b_{qq}c^2 = -b_{pq}t + b_{qq}, \\ \bar{b}_{pq} &= \bar{b}_{qp} = 0, \end{aligned} \tag{2.16}$$

$$\begin{aligned} \bar{\omega}_{ip} &= \omega_{ip}c + \omega_{iq}s, \\ \bar{\omega}_{iq} &= -\omega_{ip}s + \omega_{iq}c \quad (i = 1 \div n); \end{aligned} \tag{2.17}$$

on s'ha emprat $t = \frac{s}{c}$; les altres components no canvien.

De manera anàloga, en la k -ena iteració del procés iteratiu s'obté:

$$\begin{aligned} A_{k+1} &= R_k^\top A_k R_k = \cdots = R_k^\top \cdots R_1^\top A R_1 \cdots R_k, \\ O_{k+1} &= O_k R_k = \cdots = R_1 \cdots R_k \quad (k \geq 0). \end{aligned}$$

Les matriu A_k són, doncs, semblants a la matriu A , mitjançant les transformacions ortogonals O_k :

$$A_k = O_k^\top A O_k \quad (k \geq 0).$$

Variants del mètode de Jacobi

La tria dels elements no diagonals b_{pq} que volem anul·lar pot ser feta de diverses maneres que donen lloc a les diferents variants del mètode de Jacobi.

Així, si fem que

$$|b_{pq}| = \max_{i < j} |b_{ij}|$$

tenim el *mètode clàssic de Jacobi* (vegeu el problema II.14).

Els elements no diagonals $b_{pq} \neq 0$ poden ser triats també d'altres maneres: per exemple, prenent llurs índexs segons l'ordre: $(1, 2), (1, 3), \dots, (1, n), (2, 3), \dots, (2, n), \dots, (n-1, n)$. Repetint aquest procés les vegades que faci falta, el mètode corresponent s'anomena *mètode cíclic de Jacobi*. Una variant d'aquest consisteix a efectuar només aquelles iteracions anul·ladores dels elements més grans, en valor absolut, que unes certes quantitats prefixades, anomenats *llindars*, i que poden canviar en cada cicle; el mètode es diu llavors *mètode cíclic de Jacobi amb llindars*.

Notem que els elements no diagonals anul·lats en una iteració poden tornar a ser diferents de zero després d'alguna de les iteracions posteriors; malgrat això, es pot demostrar que, si definim

$$\epsilon^2(B) \equiv \sum_{i \neq j} b_{ij}^2,$$

resulta que, per al mètode clàssic de Jacobi,

$$\epsilon^2(\bar{B}) = \epsilon^2(B) + \sum_{j=1}^n b_{jj}^2 \leq - \sum_{j=1}^n \bar{b}_{jj}^2 \leq \epsilon^2(B) \left(1 - \frac{2}{n(n-1)}\right);$$

això implica que els elements no diagonals d' A_k tendeixen a 0 quan k tendeix a infinit i que, sota condicions molt generals, les successions $(A_k)_{k \geq 0}$ i $(O_k)_{k \geq 0}$ tenen límits respectius Λ i O tals que $\Lambda = O^\top A O$, amb Λ diagonal i O ortogonal.

2.3.5 Mètodes de reducció

L'objectiu d'aquests mètodes consisteix a reduir una matriu A real a una matriu Hessenberg superior, mitjançant transformacions de semblança ortogonals. En el cas que la matriu de partida sigui simètrica, quedarà reduïda a forma tridiagonal simètrica.

La matriu reduïda tindrà els mateixos valors propis que la matriu inicial, i els vectors propis de la matriu inicial es trobaran a partir dels de la matriu reduïda per aplicació de la transformació ortogonal trobada. En l'apartat següent es donaran mètodes de càlcul de valors i vectors propis de les matrius reduïdes.

Tractarem primer el cas simètric, el cas no simètric s'entendrà com una extensió d'aquest, on els procediments de transformació portaran a matrius Hessenberg superior, atès que en aquest cas no es produeix l'anul·lació dels elements simètrics durant el procés.

Un cop obtinguts els valors i vectors propis de les matrius tridiagonals simètriques (o Hessenberg superior), els valors propis de la matriu inicial coincidiran amb els d'aquests; els vectors propis de les matrius reduïdes hauran de ser transformats més endavant a vectors propis de la matriu de partida per les transformacions de semblança fetes.

Mètode de Givens

Com el mètode de Jacobi, el *mètode de Givens* emprà transformacions de semblança successives que són rotacions en plans coordenats; ara, però, es vol que, en cada transformació, es faci zero un element no tridiagonal, i també el simètric respecte a la diagonal, de forma que es mantinguin nuls tots els elements anul·lats en transformacions anteriors. Es tracta, per tant, d'un mètode directe que aconsegueix anul·lar els $m = \frac{(n-1)(n-2)}{2}$ elements en un màxim de m transformacions.

A cada pas se cerca una rotació R_k en el pla coordinat $p - q$ a fi que es faci zero l'element $(p - 1, q)$ i el seu simètric. Aquest procés es fa en el següent ordre dels plans coordinats $p - q$: $2 - 3, \dots, 2 - n, 3 - 4, \dots, 3 - n, \dots, (n - 1) - n$ de manera que es vagin anul·lant els elements $(1, 3), \dots, (1, n), (2, 4), \dots, (2, n), \dots, (n - 2, n)$, i llurs simètrics.

Al cap d'aquestes m transformacions de semblança tindrem la matriu

$$A_{m+1} = R_m^\top \cdots R_1^\top A R_1 \cdots R_m$$

que serà ja tridiagonal simètrica.

Emprant una notació anàloga a la feta servir en el mètode de Jacobi, l'angle φ de la rotació en el pla $p - q$ que fa $\bar{b}_{p-1,q} = 0$ compleix que

$$\tan \varphi = \frac{b_{p-1,q}}{b_{p-1,p}} ;$$

així doncs, cal triar els valors següents de c i s :

$$c = \frac{b_{p-1,p}}{\sqrt{b_{p-1,p}^2 + b_{p-1,q}^2}} , \quad s = \frac{b_{p-1,q}}{\sqrt{b_{p-1,p}^2 + b_{p-1,q}^2}} .$$

Es pot comprovar que els zeros absoluts són preservats en les successives transformacions. En aquest sentit, l'ordre indicat és important.

Mètode de Householder

Les transformacions de semblança ortogonals es realitzen ara mitjançant matrius de Householder (vegeu l'apartat 2.2.4). Es vol aconseguir, en cada transformació, que els elements no tridiagonals de la columna corresponent (i els de la filera simètrica) esdevinguin nuls.

Partint d' $A_2 = A$, cal considerar en el pas k -è una transformació ortogonal $A_{k+1} = P_k A_k P_k$ de la matriu simètrica

$$A_k = \left(\begin{array}{cc|cc} & & & 0 \\ & T_k & & \\ \hline & & \tilde{a}^{(k)\top} & \\ 0 & \tilde{a}^{(k)} & & \tilde{A}_k \end{array} \right) ,$$

on

$$T_k = \begin{pmatrix} r_1 & s_2 & & & \\ s_2 & r_2 & s_3 & & \\ & \ddots & \ddots & \ddots & \\ & & s_{k-2} & r_{k-2} & s_{k-1} \\ & & & s_{k-1} & r_{k-1} \end{pmatrix}$$

ja és tridiagonal simètrica i

$$\tilde{a}^{(k)} = \begin{pmatrix} a_{k,k-1}^{(k)} \\ a_{k+1,k-1}^{(k)} \\ \vdots \\ a_{n,k-1}^{(k)} \end{pmatrix} \in \mathbb{R}^{n-k+1} .$$

Si $(a_{k+1,k-1}^{(k)})^2 + \dots + (a_{n,k-1}^{(k)})^2 = 0$, la matriu $(A_k)_k$ (formada per les k primeres files de les k primeres columnes d' A_k) ja és de fet una matriu tridiagonal simètrica, amb $s_k = a_{k,k-1}^{(k)}$, $r_k = a_{k,k}^{(k)}$ i només cal escollir $P_k = I$, $A_{k+1} = P_k A_k P_k = A_k$.

En cas contrari, triant s_k tal que

$$s_k^2 = \|\tilde{a}^{(k)}\|_2^2 = \sum_{i=k}^n (a_{i,k-1}^{(k)})^2 ,$$

podem trobar una matriu $(n-k+1) \times (n-k+1)$ de Householder $\tilde{P}_k$ tal que

$$\tilde{P}_k \tilde{a}^{(k)} = \begin{pmatrix} s_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = s_k e^{(1)} \in \mathbb{R}^{n-k+1} .$$

De fet

$$\tilde{P}_k = I_{n-k+1} - \alpha_k \tilde{u}^{(k)} \tilde{u}^{(k)\top} ,$$

amb

$$\tilde{u}^{(k)} = \tilde{a}^{(k)} - s_k e^{(1)} \in \mathbb{R}^{n-k+1} , \quad \alpha_k = \frac{1}{s_k(s_k - a_{k,k-1}^{(k)})}$$

(vegeu l'apartat 2.2.4).

Definint aleshores

$$P_k = \left(\begin{array}{c|c} I_k & 0 \\ \hline 0 & \tilde{P}_k \end{array} \right) ,$$

o equivalentment $P_k = I_k - \alpha_k u^{(k)} u^{(k)\top}$, amb

$$u^{(k)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \tilde{u}^{(k)} \end{pmatrix} ,$$

resulta que

$$A_{k+1} = P_k A_k P_k = \left(\begin{array}{c|cccc} T_k & & & & 0 \\ & & & & \\ \hline & s_k & 0 & \cdots & 0 \\ & 0 & & & \\ 0 & \vdots & & & \tilde{P}_k \tilde{A}_k \tilde{P}_k \\ & 0 & & & \end{array} \right)$$

Suposarem que $b_j \neq 0$ ($j = 2 \div n$), ja que sinó la descompondríem en dues matrius del mateix tipus T_1 i T_2

$$T = \left(\begin{array}{c|c} T_1 & 0 \\ \hline 0 & T_2 \end{array} \right)$$

i, gràcies a la propietat 6 de valors i vectors propis en l'apartat 2.1.3, per trobar els valors propis de T només cal trobar per separat els de T_1 i T_2 .

Com que $\det(\lambda I - T) = (-1)^n \det(T - \lambda I)$, cercar els valors propis de T equival a resoldre l'equació polinomial $p_n(\lambda) = \det(\lambda I - T) = 0$.

Els valors $p_n(\lambda)$ es poden trobar per recurrència: començant amb

$$p_0(\lambda) = 1, \quad p_1(\lambda) = \lambda - a_1,$$

definim llavors el determinant $p_{j+1}(\lambda)$ de la matriu formada per les j primeres fileres de les j primeres columnes de la matriu $\lambda I - T$; aquest determinant es troba a partir dels dos determinants anàlegs anteriors fent

$$p_j(\lambda) = (\lambda - a_j)p_{j-1}(\lambda) - \gamma_j p_{j-2}(\lambda), \quad \gamma_j = b_j^2 \geq 0 \quad (j \geq 2).$$

Observem que $p_j(\lambda)$ és un polinomi de grau j . La recurrència que s'acaba d'establir és molt semblant a la que compleixen els polinomis de Txebixev. El problema II.16 fa ús d'aquesta similitud.

Notem que no es precisa conèixer explícitament els coeficients de $p_n(\lambda)$ a fi de calcular els valors propis de T , i que, donat un valor qualsevol de λ , $p_n(\lambda)$ es calcula amb aproximadament $2n$ operacions.

La successió de polinomis $\{p_n(\lambda), p_{n-1}(\lambda), \dots, p_1(\lambda), p_0(\lambda)\}$ conforma una successió d'Sturm per a $p_n(\lambda)$ sobre $\mathbb{R}$ (vegeu l'apartat 5.2.4 per a més detalls); aplicant el teorema d'Sturm, tenim que

- Els valors propis de T són tots reals i diferents (suposant $b_j \neq 0$ ($j = 2 \div n$)). A més, si $p_n(a)p_n(b) \neq 0$, el nombre de valors propis de T entre a i b és igual a $V(a) - V(b)$, on $V(\lambda)$ és el nombre de canvis de signe en la successió $\{p_0(\lambda), p_1(\lambda), \dots, p_n(\lambda)\}$.

Amb l'ajut d'aquest resultat, podem localitzar i separar els valors propis de T ; després podem aplicar qualsevol mètode de l'apartat 5.1.2 per trobar zeros d'una funció, com els mètodes de la secant o de Newton, a fi de refinar el seu valor. Observem que, si es vol fer servir el mètode de Newton, el càlcul de $p'_n(\lambda)$ es pot dur a terme també per recurrència

$$p'_0(\lambda) = 0, \quad p'_1(\lambda) = 1,$$

$$p'_j(\lambda) = p_{j-1}(\lambda) + (\lambda - a_j)p'_{j-1}(\lambda) - \gamma_j p'_{j-2}(\lambda) \quad (j = 2 \div n).$$

Aquesta recurrència s'ha de fer servir conjuntament amb l'anterior per tal de trobar $p_n(\lambda)$ i $p'_n(\lambda)$ al mateix temps.

Un cop trobats els valors propis λ_j ($j = 1 \div n$) de T , cal trobar vectors propis associats. Si volem trobar un vector propi v de T , associat a un valor propi λ , s'ha de resoldre el sistema

$$\begin{aligned}(\lambda - a_1)v_1 - b_2v_2 &= 0 \\ -b_iv_{i-1} + (\lambda - a_i)v_i - b_{i+1}v_{i+1} &= 0 \quad (i = 2 \div n-1), \\ -b_nv_{n-1} + (\lambda - a_n)v_n &= 0\end{aligned}$$

amb la condició $v_1 \neq 0$ per tal que v no sigui nul.

Prenent $v_1 = 1$, s'obté

$$v_i = \frac{p_{i-1}(\lambda)}{b_2 \cdots b_i} \quad (i = 2 \div n)$$

(vegeu el problema II.15).

Aquest procediment no sempre està ben condicionat, atès que només coneixem aproximacions de λ . En tal cas, s'obtindrien resultats més precisos aplicant unes quantes iteracions del mètode de la potència inversa a $T - \lambda I$, encara que s'hagin de fer més operacions.

Mètode per a matrius Hessenberg superior

En el cas de matrius Hessenberg superior

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & \cdots & h_{1,n-1} & h_{1n} \\ h_{21} & h_{22} & h_{23} & \ddots & \ddots & h_{2n} \\ & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & h_{n-1,n-2} & h_{n-1,n-1} & h_{n-1,n} \\ & & & & h_{n,n-1} & h_{n,n} \end{pmatrix},$$

s'obté també una relació de recurrència per als determinants de les matrius formades per les j primeres fileres de les j primeres columnes de $\lambda I - H$, $p_j(\lambda)$ ($j = 1 \div n$), que porten a l'equació característica $p_n(\lambda) = 0$ que compleixen els valors propis de H .

Suposem $h_{j+1,j} \neq 0$ ($j = 1 \div n-1$), com en el cas tridiagonal simètric; sinó passariem a trobar per separat els valors propis de les matrius Hessenberg superior H_1 i H_2 tals que

$$H = \left(\begin{array}{c|c} H_1 & N_1 \\ \hline 0 & H_2 \end{array} \right).$$

A fi d'escriure millor les relacions de recurrència de la successió de polinomis $p_j(\lambda)$ ($j = 0 \div n$), prenem

$$q_0(\lambda) = p_0(\lambda) = 1, \quad q_j(\lambda) = \frac{p_j(\lambda)}{h_{21}h_{32} \cdots h_{j+1,j}} \quad (j = 2 \div n);$$

la recurrència queda llavors en la forma

$$q_0(\lambda) = 1, \quad q_1(\lambda) = \frac{\lambda - h_{11}}{h_{21}},$$

$$q_j(\lambda) = \frac{(\lambda - h_{j,j})q_{j-1}(\lambda) - [h_{1,j}q_0(\lambda) + \cdots + h_{j-1,j}q_{j-2}(\lambda)]}{h_{j+1,j}}$$

$$(j = 2 \div n).$$

A diferència del que succeïa amb matrius simètriques, els valors propis de matrius qualssevol no són necessàriament reals. Això ens diu que els zeros de $q_n(\lambda)$ no sempre seran reals. S'hauran de cercar usant mètodes de càlcul de zeros de polinomis que permetin trobar zeros complexos (vegeu el capítol V); els vectors propis associats es poden cercar llavors amb l'ajut del mètode de la potència inversa.

2.3.7 Mètodes de factorització

Els mètodes que es tractaran estan basats en les factoritzacions LU (dita aquí LR) i QR, esmentades en els apartats 2.2.3 i 2.2.4, respectivament.

Mètode iteratiu LR

Donada una matriu $n \times n$ real, el mètode iteratiu LR parteix de la matriu $A_1 = A$ i, quan coneix la matriu A_k , calcula la seva factorització LU (si existeix)

$$A_k = L_k R_k,$$

amb L_k triangular inferior amb uns a la diagonal i R_k , triangular superior; a continuació, forma $A_{k+1} = R_k L_k$ ($k \geq 1$).

Com que

$$A_{k+1} = L_k^{-1} A_k L_k = \cdots = (L_1 \cdots L_k)^{-1} A (L_1 \cdots L_k),$$

les matrius A_k ($k \geq 1$) són semblants a la matriu A .

Mètode iteratiu QR

El mètode iteratiu QR usa la factorització QR en comptes de la LU: així,

$$A_k = Q_k R_k, \quad A_{k+1} = R_k Q_k,$$

i llavors

$$A_{k+1} = Q_k^\top A_k Q_k = \cdots = (Q_1 \cdots Q_k)^\top A (Q_1 \cdots Q_k).$$

Per tant, les matrius A_k ($k \geq 1$) són també semblants a la matriu A .

Consideracions generals

Les successions $(A_k)_{k \geq 0}$ convergeixen en una matriu triangular superior, sota condicions bastant generals; n'hi ha prou, per exemple, que es verifiquin les tres condicions següents:

- els valors propis d' A estan separats en mòdul

$$|\lambda_1| > \cdots > |\lambda_n|,$$

- A és diagonalitzable: existeix una matriu regular X tal que la matriu $D = X^{-1}AX$ és diagonal,
- existeix la factorització LU de les matrius X i X^{-1} .

El mètode iteratiu QR conserva la simetria de la matriu; això és, si la matriu A és simètrica, totes les matrius A_k ($k \geq 1$) són simètriques, i la convergència a una matriu triangular superior es tradueix en convergència a una matriu diagonal.

Els mètodes iteratius LR i QR conserven a més les formes Hessenberg superior de les matrius; el mètode iteratiu QR conserva, doncs, el caràcter de matriu tridiagonal simètrica.

El nombre d'operacions que requereixen aquests mètodes és elevat i són bastant costosos quan s'apliquen a matrius plenes qualssevol; és, però, molt eficient transformar prèviament la matriu A a una forma reduïda (Hessenberg superior o tridiagonal simètrica) per algun dels mètodes donats en l'apartat anterior (de Givens o de Householder) i aplicar llavors els mètodes descrits ací a les matrius reduïdes.

Notem, per exemple, que la factorització QR de matrius Hessenberg superior (o de matrius tridiagonals simètriques, en particular) es produeix molt eficaçment mitjançant un mètode que empru rotacions R_j en plans coordinats $(n-j+1) - (n-j)$ ($j = 1 \div n-1$) en l'ordre següent: $n - (n-1)$, $(n-1) - (n-2)$, ..., $2 - 1$, i de forma que es vagin anul·lant els elements $(n, n-1)$, $(n-1, n-2)$, ..., $(2, 1)$ de les matrius transformades. En cada transformació es modifiquen només dues fileres de la matriu i es preserven els elements anul·lats en els passos anteriors.

COMENTARIS BIBLIOGRÀFICS

S'han suposat coneguts els resultats teòrics elementals sobre resolució de sistemes lineals i problemes de valors i vectors propis, que poden ser trobats, per exemple, en [HS74], [Que71], ... Diversos llibres hi dediquen algun capítol introductori amb un punt de vista més algorímic, com [Cia82], [DM73], [Fro69], [Hou64], [IK66], on poden ésser trobats els resultats de la secció 2.1.3.

[Cia82] és una de les millors referències per a la resolució de sistemes lineals, juntament amb [Wil64] per a mètodes directes, i [You71] per a mètodes iteratius. Aquests tres llibres cobreixen essencialment tots els resultats d'aquest capítol. Unes altres referències generals bàsiques són [DB74], [Gas66], [Hil74], [Jac77], [Ral65]. En [CdB72] i, principalment, [WR71] es presenten programes amb la implementació efectiva dels diferents mètodes, tant de sistemes lineals com de valors i vectors propis. Les matrius escasses no han estat tractades, encara que apareixen sovint en la resolució de molts problemes reals; una referència molt útil pot ser [Ric81]. L'estudi de propagació dels errors es troba en [IK66] i sobretot en

[Wil64]. Per a factoritzacions QR, la referència estàndard és [LH74], que també incorpora rutines en llenguatge Fortran. Els càlculs sobre el nombre d'operacions dels diferents mètodes d'aquest capítol apareixen desenvolupats en moltes referències com [Cia82], [IK66] i [LH74].

Pel que fa al càlcul de valors i vectors propis, cal destacar [Cia82], [Ham70], [RR78], [Ste73] i [Wil65]. Per a diferents resultats sobre localització de valors propis, consulteu [SB80]. Els diversos mètodes presentats de deflació es troben a [DM73], [Fro69] i [RR78]. Una implementació més efectiva del mètode de la potència i variants utilitza altres tècniques, com l'acceleració d'Aitken i una elecció adequada dels desplaçaments d per a la matriu $A - dI$, vegeu [IK66], [RR78] i [Wil64]. El mètode de Jacobi i les seves diverses variants pot ser trobat d'una manera més detallada a [Cia82] i [IK66]. Aquest últim text també incorpora els càlculs sobre el nombre d'operacions necessàries per portar a terme els mètodes de Givens i de Householder. Un mètode alternatiu per reduir matrius no simètriques a matrius Hessenberg superior consisteix a usar eliminació gaussiana amb pivotatge per columnes, vegeu [RR78]. En aquest llibre i a [Cia82] hi ha una bona discussió sobre el teorema de Sturm. La convergència dels mètodes iteratius LR i QR, així com la seva implementació efectiva, usant matrius desplaçades, com en el mètode de la potència, es pot trobar a [Cia82], [LH74], [RR78] i, sobretot, en l'article de B.N. Parlett a [RW67] i a [Par80].

PROBLEMES RESOLTS

Problema 2.1 a) Si $A = (a_{ij})$ és una matriu definida positiva, demostreu que:

- i) $a_{ii} > 0$ ($i = 1 \div n$);
 - ii) $\max_{i,j} |a_{ij}| = \max_i a_{ii}$;
 - iii) les submatrius $A^{(k)} = (a_{ij}^{(k)})_{i,j=k \div n}$ ($k = 2 \div n$), trobades en aplicar el mètode de Gauss, són definides positives i, per tant, que es pot portar a terme el procés sense haver de recórrer als pivotatges;
 - iv) $a_{ii}^{(k+1)} \leq a_{ii}^{(k)}$ ($i = k + 1 \div n$, $k = 1 \div n - 1$).
- b) Aplicació: Demostreu que la matriu

$$\begin{pmatrix} 13 & 11 & 11 \\ 11 & 13 & 11 \\ 11 & 11 & 13 \end{pmatrix}$$

és definida positiva, trobeu-ne la factorització de Txoleski i el seu determinant.

SOLUCIÓ:

a) i) Es dedueix de la mateixa definició de matriu definida positiva, prenent $x = e^{(i)}$ (vector i -è de la base canònica) :

$$a_{ii} = e^{(i)\top} A e^{(i)} > 0 \quad (i = 1 \div n) .$$

ii) El criteri de Sylvester assegura que una matriu és definida positiva si i només si els determinants principals Δ_k ($k \geq 0$) són estrictament positius. Una matriu definida positiva continua sent-ho després de permutar dues de les seves fileres al mateix temps que es permuten les columnes corresponents, atès que això representa només un canvi de variable consistent a permutar les components corresponents dels vectors. Així, canviant les fileres i columnes i i j per les fileres i columnes 1 i 2 de la matriu A , s'obté una matriu definida positiva amb determinant principal d'ordre 2

$$\begin{vmatrix} a_{ii} & a_{ij} \\ a_{ij} & a_{jj} \end{vmatrix} = a_{ii}a_{jj} - a_{ij}^2 > 0 ;$$

per tant,

$$|a_{ij}| < a_{ii} \quad \text{o} \quad |a_{ij}| < a_{jj} \quad (i, j = 1 \div n, \quad i \neq j) ,$$

i obtenim la relació demanada

$$\max_{i,j} |a_{ij}| = \max_i a_{ii} .$$

iii) Ho demostrarem per inducció. L'enunciat del problema ens diu que $A^{(1)} = A$ és definida positiva. Suposarem que $A^{(k)}$ n'és i deduirem que $A^{(k+1)}$ també és definida positiva per a $k = 1 \div n - 1$.

Si $A^{(k)}$ és definida positiva, resulta que:

- $a_{kk}^{(k)} > 0$.

Es dedueix aplicant l'apartat i) a la matriu $A^{(k)}$. Per tant, el pas k -è del mètode de Gauss podrà realitzar-se sense pivotatge.

- $A^{(k+1)}$ és simètrica.

Aquesta afirmació es dedueix de la fórmula d'obtenció dels elements d' $A^{(k+1)}$ a partir dels d' $A^{(k)}$ i del fet que $A^{(k)}$ és simètrica:

$$\begin{aligned} a_{ij}^{(k+1)} &= a_{ij}^{(k)} - \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} a_{kj}^{(k)} \\ &= a_{ji}^{(k)} - \frac{a_{jk}^{(k)}}{a_{kk}^{(k)}} a_{ki}^{(k)} = a_{ji}^{(k+1)} \quad (i, j = k + 1 \div n) . \end{aligned}$$

- $A^{(k+1)}$ és definida positiva.

Com que és simètrica, aplicant el criteri de Sylvester, serà suficient demostrar que tots els seus determinants principals són positius. Atès que el procés d'eliminació gaussiana sense pivotatge preserva el determinant, tenim que

$$\det (A^{(k)})_j = a_{kk}^{(k)} \det (A^{(k+1)})_{j-1} \quad (j = 2 \div n - k + 1) .$$

Aplicant el criteri de Sylvester a la matriu $A^{(k)}$, $\det (A^{(k)})_j > 0$. Com que $a_{kk}^{(k)} > 0$, resulta que $\det (A^{(k+1)})_{j-1} > 0$ ($j = 2 \div n - k + 1$), que és el que volíem demostrar.

Observeu que, a més de demostrar que les submatrius $A^{(k)}$ ($k = 2 \div n$) són definides positives, hem deduït que es pot aplicar el mètode d'eliminació gaussiana sense necessitat de recórrer al pivotatge.

iv) Usant la simetria d' $A^{(k)}$ i el fet que $a_{kk}^{(k)} > 0$, resulta que

$$a_{ii}^{(k+1)} = a_{ii}^{(k)} - \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} a_{ki}^{(k)} = a_{ii}^{(k)} - \frac{a_{ik}^{(k)2}}{a_{kk}^{(k)}} \leq a_{ii}^{(k)} \quad (i = k + 1 \div n) .$$

b) A fi de trobar la factorització de Txoleski, cercarem primer una factorització de la forma $LDL^\top$ vàlida per a una matriu simètrica qualsevol; les fórmules recurrents de càlcul dels elements de les matrius L i D es dedueixen, per exemple, de les fórmules de mètode de Crout, entenent que la matriu U ara és $DL^\top$.

Tenim, doncs,

$$\begin{aligned} d_{11} &= a_{11} = 13, & l_{21} &= \frac{a_{21}}{d_{11}} = \frac{11}{13}, \\ d_{22} &= a_{22} - l_{21}^2 d_{11} = \frac{48}{13}, & l_{31} &= \frac{a_{31}}{d_{11}} = \frac{11}{13}, \\ l_{32} &= \frac{a_{32} - l_{31} l_{21} d_{11}}{d_{22}} = \frac{11}{24}, & d_{33} &= a_{33} - l_{31}^2 d_{11} - l_{32}^2 d_{22} = \frac{35}{12}. \end{aligned}$$

Així, en forma matricial,

$$A = LDL^T = \begin{pmatrix} 1 & 0 & 0 \\ \frac{11}{13} & 1 & 0 \\ \frac{11}{13} & \frac{11}{24} & 1 \end{pmatrix} \begin{pmatrix} 13 & 0 & 0 \\ 0 & \frac{48}{13} & 0 \\ 0 & 0 & \frac{35}{12} \end{pmatrix} \begin{pmatrix} 1 & \frac{11}{13} & \frac{11}{13} \\ 0 & 1 & \frac{11}{24} \\ 0 & 0 & 1 \end{pmatrix}.$$

La matriu A és definida positiva atès que tots els elements diagonals de D són positius i la seva factorització de Txoleski $A = \mathcal{L}\mathcal{L}^T$ s'obté amb

$$\mathcal{L} = \begin{pmatrix} \sqrt{13} & 0 & 0 \\ \frac{11}{\sqrt{13}} & \sqrt{\frac{48}{13}} & 0 \\ \frac{11}{\sqrt{13}} & \frac{11}{24}\sqrt{\frac{48}{13}} & \sqrt{\frac{35}{12}} \end{pmatrix} = \begin{pmatrix} 3.6056 & 0 & 0 \\ 3.0509 & 1.9215 & 0 \\ 3.0509 & 0.8807 & 1.7078 \end{pmatrix}.$$

Finalment,

$$\det A = d_{11}d_{22}d_{33} = 13 \frac{48}{13} \frac{35}{12} = 140.$$

Problema 2.2 Calculeu el nombre d'operacions necessàries per resoldre un sistema $Ax = b$, on A és una matriu pentadiagonal tal que $a_{ij} = 0$, si $|i - j| = 1$.

SOLUCIÓ:

La matriu A és de la forma

$$A = \begin{pmatrix} a_1 & 0 & c_1 & & & & \\ 0 & a_2 & 0 & c_2 & & & \\ d_3 & 0 & a_3 & 0 & c_3 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & d_{n-2} & 0 & a_{n-2} & 0 & c_{n-2} \\ & & & & d_{n-1} & 0 & a_{n-1} & 0 \\ & & & & & d_n & 0 & a_n \end{pmatrix}.$$

La resolució del sistema pel mètode de Gauss es compon de dues fases ben definides:

TRIANGULARITZACIÓ

En el pas k -è hem d'anul·lar únicament l'element d_{k+2} ; deixant a part aquesta anul·lació, els únics elements modificats són el a_{k+2} de la matriu A i el b_{k+2} del vector b , que se substitueixen per

$$\bar{a}_{k+2} = a_{k+2} - \frac{d_{k+2}}{\bar{a}_k} c_k, \quad \bar{b}_{k+2} = b_{k+2} - \frac{d_{k+2}}{\bar{a}_k} b_k \quad (k = 1 \div n-2),$$

on s'ha fet inicialment $\bar{a}_1 = a_1$, $\bar{a}_2 = a_2$, $\bar{b}_1 = b_1$ i $\bar{b}_2 = b_2$, atès que aquests elements no es modifiquen en el procés.

En cada pas s'han de fer, doncs, 2 restes, 2 multiplicacions i 1 divisió; en total: $2(n-2)$ restes i multiplicacions i $n-2$ divisions.

SUBSTITUCIÓ CAP AL DARRERA

Cal trobar les components del vector solució x recurrentment, segons

$$x_n = \frac{\bar{b}_n}{\bar{a}_n}, \quad x_{n-1} = \frac{\bar{b}_{n-1}}{\bar{a}_{n-1}};$$

$$x_k = \frac{\bar{b}_k - c_k x_{k+2}}{\bar{a}_k} \quad (k = n-2 \div 1).$$

En aquesta etapa calen, doncs, $n-2$ restes i multiplicacions i n divisions.

En la taula següent es resumeixen els nombres d'operacions trobats:

	—	*	/
Triangularització	$2(n-2)$	$2(n-2)$	$n-2$
Substitució	$n-2$	$n-2$	n
Total	$3(n-2)$	$3(n-2)$	$2(n-1)$

Comptant les operacions com a multiplicacions/divisions+sumes/restes, en total cal fer-ne $5n-8$.

Problema 2.3 Considerem la matriu $n \times n$

$$A = \begin{pmatrix} a & 1 & & & & & \\ 1 & a & 1 & & & & \\ & 1 & a & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & \ddots \\ & & & & & 1 & a & 1 \\ & & & & & & 1 & a \end{pmatrix}.$$

a) Demostreu que la matriu A és definida positiva per a $a \geq 2$.

b) Si $a \geq 2$, trobeu un mètode recurrent per fer la factorització de Txoleski de la matriu A , $A = \mathcal{L}\mathcal{L}^\top$ amb

$$\mathcal{L} = \begin{pmatrix} \alpha_1 & & & & & \\ \beta_2 & \alpha_2 & & & & \\ & \beta_3 & \alpha_3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \beta_{n-1} & \alpha_{n-1} \\ & & & & & \beta_n & \alpha_n \end{pmatrix}.$$

c) Si l'element a_{11} d' A té un error absolut δ i els altres elements són exactes, doneu una estimació de l'error relatiu de α_n en fer la factorització, suposant δ prou petit.

SOLUCIÓ:

a) Pel criteri de Sylvester, cal demostrar que els determinants principals Δ_j ($j = 1 \div n$) són positius.

És evident que $\Delta_1 = a$ i $\Delta_2 = a\Delta_1 - 1$ i, desenvolupant el determinant d' $(A)_{j+1}$ per l'última fila o columna,

$$\Delta_{j+1} = a\Delta_j - \Delta_{j-1} \quad (j = 2 \div n - 1).$$

Veurem per inducció que $\Delta_{j+1} > \Delta_j > 0$ ($j = 1 \div n - 1$).

La relació es compleix per a $j = 1$: $\Delta_2 = a^2 - 1 > a = \Delta_1 > 0$, ja que $a \geq 2$; suposada la validesa de la relació fins a $j - 1$, tenim

$$\Delta_{j+1} = a\Delta_j - \Delta_{j-1} \geq 2\Delta_j - \Delta_{j-1} = \Delta_j + (\Delta_j - \Delta_{j-1}) > \Delta_j > 0,$$

ja que s'ha suposat que $\Delta_j > \Delta_{j-1} > 0$.

b) Suposem realitzada la factorització $\mathcal{L}\mathcal{L}^\top$ de la matriu A amb

$$\mathcal{L} = \begin{pmatrix} \alpha_1 & & & & & \\ \beta_2 & \alpha_2 & & & & \\ & \ddots & \ddots & & & \\ & & \beta_{n-1} & \alpha_{n-1} & & \\ & & & \beta_n & \alpha_n \end{pmatrix}.$$

Cal cercar els valors de α_j ($j = 1 \div n$) i de β_j ($j = 2 \div n$) de manera que es compleixi la igualtat matricial $A = \mathcal{L}\mathcal{L}^\top$ que, component a component, s'escriu:

$$a = \alpha_1^2;$$

$$1 = \alpha_{j-1}\beta_j, \quad a = \alpha_j^2 + \beta_j^2 \quad (j = 2 \div n).$$

Aïllant les incògnites, tenim la recurrència següent:

$$\alpha_1 = \sqrt{a}; \quad \beta_j = \frac{1}{\alpha_{j-1}}, \quad \alpha_j = \sqrt{a - \beta_j^2} \quad (j = 2 \div n).$$

c) Aplicant la fórmula aproximada de propagació de l'error a les recurrències trobades en l'apartat b):

$$e_a(\alpha_1) \simeq \frac{1}{2\alpha_1} \delta ;$$

$$e_a(\beta_j) \simeq -\frac{1}{\alpha_{j-1}^2} e_a(\alpha_{j-1}) , \quad e_a(\alpha_j) \simeq -\frac{\beta_j}{\alpha_j} e_a(\beta_j) \quad (j = 2 \div n) .$$

Així,

$$e_a(\alpha_j) \simeq \frac{\beta_j}{\alpha_j \alpha_{j-1}^2} e_a(\alpha_{j-1}) = \frac{1}{\alpha_j \alpha_{j-1}^3} e_a(\alpha_{j-1})$$

i

$$e_a(\alpha_n) \simeq \frac{1}{\alpha_n \alpha_{n-1}^3} \cdots \frac{1}{\alpha_2 \alpha_1^3} e_a(\alpha_1) = \frac{\alpha_n^3 \alpha_1}{\prod_{j=1}^n \alpha_j^4} e_a(\alpha_1) \simeq \frac{\alpha_n^3}{2 \prod_{j=1}^n \alpha_j^4} \delta .$$

Tenint en compte que

$$\det A = \prod_{j=1}^n \alpha_j^2 ,$$

resulta

$$e_a(\alpha_n) \simeq \frac{\alpha_n^3}{2(\det A)^2} \delta .$$

D'on, l'error relatiu en α_n pot estimar-se per

$$e_r(\alpha_n) \simeq \frac{\alpha_n^2}{2(\det A)^2} \delta .$$

Problema 2.4 Demostreu que les normes matricials subordinades a les normes vectorials

$$\|x\|_1 = \sum_i |x_i| , \quad \|x\|_\infty = \max_i |x_i| ,$$

són respectivament:

$$\|A\|_1 = \max_j \sum_i |a_{ij}| , \quad \|A\|_\infty = \max_i \sum_j |a_{ij}| .$$

SOLUCIÓ:

La definició de norma matricial subordinada a la norma vectorial $\|\cdot\|_1$ és

$$\|A\|_1 = \max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1} .$$

Com que

$$\|Ax\|_1 = \sum_i \left| \sum_j a_{ij} x_j \right| \leq \sum_j \left(\sum_i |a_{ij}| \right) |x_j| \leq \max_j \sum_i |a_{ij}| \|x\|_1 ,$$

tenim que

$$\|A\|_1 \leq \max_j \sum_i |a_{ij}| .$$

Ara bé, si

$$\max_j \sum_i |a_{ij}| = \sum_i |a_{il}| ,$$

prenent $\bar{x}$ de components $\bar{x}_j = \delta_{jl}$ tenim que $\|\bar{x}\|_1 = 1$ i

$$\|A\bar{x}\|_1 = \sum_i \left| \sum_j a_{ij} \bar{x}_j \right| = \sum_i |a_{il}| = \max_j \sum_i |a_{ij}| .$$

D'on,

$$\|A\|_1 = \max_j \sum_i |a_{ij}| .$$

De manera anàloga, tenim per a $\|\cdot\|_\infty$:

$$\|Ax\|_\infty = \max_i \left| \sum_j a_{ij} x_j \right| \leq \max_i \sum_j |a_{ij}| |x_j| \leq \max_i \sum_j |a_{ij}| \|x\|_\infty ,$$

que implica

$$\|A\|_\infty \leq \max_i \sum_j |a_{ij}| .$$

Si ara

$$\max_i \sum_j |a_{ij}| = \sum_j |a_{kj}| ,$$

triaent $\bar{x}$ de components $\bar{x}_j = \text{sgn}(a_{kj})$, tenim que $\|\bar{x}\|_\infty = 1$ i

$$\begin{aligned} \|A\bar{x}\|_\infty &= \max_i \left| \sum_j a_{ij} \bar{x}_j \right| = \max_i \left| \sum_j a_{ij} \text{sgn}(a_{kj}) \right| \\ &= \sum_j |a_{kj}| = \max_i \sum_j |a_{ij}| . \end{aligned}$$

D'on,

$$\|A\|_\infty = \max_i \sum_j |a_{ij}| .$$

Problema 2.5 a) Partint de $(A + \delta A)(x + \delta x) = b + \delta b$, demostreu que, si

$$\|A^{-1}\| \|\delta A\| < 1 ,$$

llavors

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\mu(A)}{1 - \mu(A) \frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \right) ,$$

sempre que la norma matricial emprada sigui consistent amb la norma vectorial.

b) Trobeu una matriu 3x3 A tal que compleixi $\mu_\infty(A) > 10^6$, malgrat que $\|A\|_\infty < 10^{-6}$.

c) Resoleu els sistemes corresponents $Ax = b$ amb

$$i) b = b^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} , \quad ii) b = b^{(2)} = \begin{pmatrix} 0.9 \\ 1.1 \\ 1 \end{pmatrix} .$$

d) Comenteu els resultats trobats a l'apartat c).

SOLUCIÓ:

a) De $(A + \delta A)(x + \delta x) = b + \delta b$ i $Ax = b$, es dedueix

$$\delta x = A^{-1}\delta b - A^{-1}\delta A(x + \delta x) .$$

Prenent normes consistents i emprant les seves propietats,

$$\begin{aligned} \|\delta x\| &= \|A^{-1}\delta b - A^{-1}\delta A(x + \delta x)\| \\ &\leq \|A^{-1}\delta b\| + \|A^{-1}\delta A(x + \delta x)\| \\ &\leq \|A^{-1}\| \|\delta b\| + \|A^{-1}\| \|\delta A\| \|x + \delta x\| ; \end{aligned}$$

com que $b = Ax$, resulta que

$$\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|} , \quad \text{si } x, b \neq 0 .$$

Aleshores,

$$\begin{aligned} \frac{\|\delta x\|}{\|x\|} &\leq \mu(A) \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|x + \delta x\|}{\|x\|} \frac{\|\delta A\|}{\|A\|} \right) \\ &\leq \mu(A) \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \right) + \mu(A) \frac{\|\delta A\|}{\|A\|} \frac{\|\delta x\|}{\|x\|} ; \end{aligned}$$

i, fent servir que

$$\mu(A) \frac{\|\delta A\|}{\|A\|} = \|A^{-1}\| \|\delta A\| < 1 ,$$

tenim finalment

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\mu(A)}{1 - \mu(A) \frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|} \right) .$$

b) Considerem δ suficientment petit i ϵ molt més petit que δ ; llavors la matriu

$$A = \delta \begin{pmatrix} \epsilon & 1 & 0 \\ -\epsilon & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

compleix que $\|A\|_\infty = \delta(1 + \epsilon) \simeq \delta$ i la seva inversa

$$A^{-1} = \frac{1}{2\epsilon\delta} \begin{pmatrix} 1 & -1 & 0 \\ \epsilon & \epsilon & 0 \\ 0 & 0 & 2\epsilon \end{pmatrix}$$

compleix que $\|A^{-1}\|_\infty = \frac{1}{\epsilon\delta}$; d'on, el seu nombre de condició és molt gran

$$\mu_\infty(A) = \frac{1 + \epsilon}{\epsilon} \simeq \frac{1}{\epsilon}.$$

Prenent, per exemple, $\epsilon = 10^{-7}$ i $\delta = 10^{-7}$ tenim:

$$\|A\|_\infty = 10^{-7}(1 + 10^{-7}) < 10^{-6}, \quad \mu_\infty(A) = \frac{1 + 10^{-7}}{10^{-7}} > 10^6.$$

c) Resolem primerament el sistema amb $b = b^{(1)} = (1 \ 1 \ 1)^\top$, la solució exacta serà

$$x^{(1)} = A^{-1}b^{(1)} = (0 \ \frac{1}{\delta} \ \frac{1}{\delta})^\top.$$

Si fem el mateix per a $b^{(2)} = (0.9 \ 1.1 \ 1)^\top$, la solució serà $x^{(2)} = x^{(1)} + \delta x$ amb $\delta x = A^{-1}\delta b$, essent $\delta b = (-0.1 \ 0.1 \ 0)^\top$, resulta així

$$\delta x = A^{-1}\delta b = -\frac{1}{10\epsilon\delta} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Per als valors donats de ϵ i δ resulten:

$$\|x^{(1)}\|_\infty = 10^7, \quad \|\delta x\|_\infty = 10^{13}.$$

d) Els resultats de l'apartat c) ens fan observar que una pertorbació petita, en termes relatius, del vector $b^{(1)}$

$$\frac{\|\delta b\|_\infty}{\|b^{(1)}\|_\infty} = 10^{-1}$$

produeix una pertorbació important en la solució del sistema

$$\frac{\|\delta x\|_\infty}{\|x^{(1)}\|_\infty} = 10^6.$$

Aquesta ampliació és deguda al fet que la matriu A té un nombre de condició molt alt, que eleva notablement el valor de la fita de l'error relatiu de x deduïda a l'apartat a)

$$\frac{\|\delta x\|_{\infty}}{\|x^{(1)}\|_{\infty}} \leq \mu_{\infty}(A) \frac{\|\delta b\|_{\infty}}{\|b^{(1)}\|_{\infty}} = \frac{1+10^{-7}}{10^{-7}} \frac{1}{10} = (1+10^{-7})10^6 .$$

Adoneu-vos que, en aquest cas, la fita anterior és pràcticament assolida.

Problema 2.6 Demostreu que, per al sistema

$$\left. \begin{array}{lcl} ax & + & by = p \\ cx & + & dy = q \end{array} \right\} ,$$

una condició necessària i suficient de convergència dels mètodes iteratius de Jacobi i Gauss-Seidel és $|bc| < |ad|$.

SOLUCIÓ:

Els mètodes iteratius de resolució de sistemes lineals convergeixen si i només si la matriu d'iteració té tots els valors propis menors que 1 en mòdul. Per als mètodes iteratius de Jacobi i de Gauss-Seidel, un cop feta la descomposició d' $A = D(L+I+U)$ amb D diagonal, L triangular inferior amb zeros a la diagonal i U triangular superior amb zeros a la diagonal, les matrius d'iteració són $B_J = -(L+U)$ i $B_{GS} = -(I+L)^{-1}U$, respectivament.

En el cas en qüestió:

$$\begin{aligned} A &= \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \left[\begin{pmatrix} 0 & 0 \\ \frac{c}{d} & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & \frac{b}{a} \\ 0 & 0 \end{pmatrix} \right] \\ &= D(L+I+U) , \\ B_J &= - \left[\begin{pmatrix} 0 & 0 \\ \frac{c}{d} & 0 \end{pmatrix} + \begin{pmatrix} 0 & \frac{b}{a} \\ 0 & 0 \end{pmatrix} \right] = \begin{pmatrix} 0 & -\frac{b}{a} \\ -\frac{c}{d} & 0 \end{pmatrix} , \\ B_{GS} &= - \left[\begin{pmatrix} 0 & 0 \\ \frac{c}{d} & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 0 & \frac{b}{a} \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -\frac{b}{a} \\ 0 & \frac{cb}{da} \end{pmatrix} . \end{aligned}$$

Els mòduls dels valors propis de les matrius d'iteració són

$$\begin{aligned} &\sqrt{\left| \frac{bc}{ad} \right|} , \quad \text{per a } B_J ; \\ &0 , \quad \left| \frac{bc}{ad} \right| , \quad \text{per a } B_{GS} . \end{aligned}$$

Per tant, els mètodes convergeixen si i només si

$$|bc| < |ad| \quad .$$

Notem finalment que, en aquest cas especial, la velocitat de convergència del mètode iteratiu de Gauss-Seidel és el doble de la del de Jacobi.

Problema 2.7 Volem resoldre pel mètode iteratiu de Jacobi el sistema $Ax = b$ partint de $x^{(0)} = 0$. Tenim la informació següent sobre les dades:

$$|a_{ii}| > \alpha \sum_{j \neq i} |a_{ij}| \quad (i = 1 \div n), \quad \alpha > 1,$$

$$0 < \gamma < |a_{ii}| \quad (i = 1 \div n), \quad \|b\|_{\infty} = \beta,$$

on α, β, γ són constants conegudes.

a) Quantes iteracions k del mètode calen per tal de garantir un error en la norma del màxim menor que ϵ , $\|x^{(k)} - x\|_{\infty} < \epsilon$?

b) Aplicació: Quan val k si $\alpha = 2$, $\beta = 10$, $\gamma = 1$ i $\epsilon = 10^{-6}$?

SOLUCIÓ:

a) El mètode iteratiu de Jacobi es pot aplicar sempre a matrius d'aquest tipus perquè els elements diagonals són no nuls. Si descomponem la matriu en la forma $A = D(I + N)$, on D és una matriu diagonal i N té els elements diagonals nuls, el mètode iteratiu de Jacobi es basa en l'equivalència del sistema $Ax = b$ de partida i el sistema $x = D^{-1}b - Nx$; el mètode iteratiu en qüestió serà, doncs,

$$x^{(0)} = 0; \quad x^{(k+1)} = D^{-1}b - Nx^{(k)} \quad (k \geq 0).$$

La convergència del mètode està assegurada perquè A és estrictament diagonal dominant. El problema consisteix ara en fitar $\|x^{(k)} - x\|_{\infty}$.

Noteu que el comportament dels errors vectorials $x^{(k)} - x$ és

$$x^{(k)} - x = -N(x^{(k-1)} - x)$$

i, per tant,

$$\|x^{(k)} - x\|_{\infty} \leq \|N\|_{\infty} \|x^{(k-1)} - x\|_{\infty}.$$

Calculem ara una fita de $\|N\|_{\infty}$.

Els elements de $N = (n_{ij})$ són

$$n_{ij} = \begin{cases} \frac{a_{ij}}{a_{ii}} & (\text{si } i \neq j) \\ 0 & (\text{si } i = j) \end{cases};$$

aleshores,

$$\|N\|_{\infty} = \max_i \sum_j |n_{ij}| = \max_i \sum_{j \neq i} \left| \frac{a_{ij}}{a_{ii}} \right| = \max_i \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| < \frac{1}{\alpha} .$$

D'on,

$$\|x^{(k)} - x\|_{\infty} < \frac{1}{\alpha} \|x^{(k-1)} - x\|_{\infty} < \dots < \frac{1}{\alpha^k} \|x^{(0)} - x\|_{\infty} = \frac{1}{\alpha^k} \|x\|_{\infty} .$$

Com que $x = D^{-1}b - Nx$, prenent normes, tenim

$$\|x\|_{\infty} \leq \|D^{-1}\|_{\infty} \|b\|_{\infty} + \|N\|_{\infty} \|x\|_{\infty} < \frac{1}{\gamma} \beta + \frac{1}{\alpha} \|x\|_{\infty} ;$$

per tant,

$$\|x\|_{\infty} < \frac{\alpha \beta}{(\alpha - 1)\gamma} , \quad \|x^{(k)} - x\|_{\infty} < \frac{\beta}{\alpha^{k-1}(\alpha - 1)\gamma} .$$

En conseqüència, si

$$\alpha^{k-1} > \frac{\beta}{\epsilon(\alpha - 1)\gamma} ,$$

llavors $\|x^{(k)} - x\|_{\infty} < \epsilon$; prenent logaritmes, obtenim la fita inferior següent del nombre d'iteracions necessàries, a fi d'assegurar que l'error, en la norma del màxim, sigui menor que ϵ :

$$k > 1 + \frac{\log \frac{\beta}{\epsilon(\alpha-1)\gamma}}{\log \alpha} .$$

b) En l'aplicació demanada, per a $\alpha = 2$, $\beta = 10$, $\gamma = 1$ i $\epsilon = 10^{-6}$, s'obté $k > 24.25$; així és que 25 iteracions seran suficients per tal de garantir aquella fita d'error.

Problema 2.8 Sigui A una matriu $n \times n$ amb valors propis λ_j ($j = 1 \div n$), reals entre $m > 0$ i M :

$$0 < m = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n = M.$$

Per tal de resoldre el sistema $Ax = b$, es considera la família de mètodes iteratius

$$x^{(k+1)} = B_{\omega} x^{(k)} + c_{\omega} ,$$

on $B_{\omega} = I - \omega A$ i $c_{\omega} = \omega b$.

a) Demostreu que són convergents per a tots els valors $\omega \in (0, \frac{2}{M})$.

b) Si $\rho(\omega)$ indica el radi espectral de B_{ω} , demostreu que

$$\rho(\omega^*) = \min_{\omega} \rho(\omega) = \frac{M - m}{M + m} ,$$

per al valor òptim de ω , $\omega^* = \frac{2}{M+m}$.

SOLUCIÓ:

a) El mètode iteratiu serà convergent quan tots els valors propis de B_ω tinguin mòdul més petit que 1.

Els valors propis de B_ω són de la forma $\mu_j = 1 - \omega\lambda_j$ ($j = 1 \div n$), complint-se la relació

$$1 > 1 - \omega m = 1 - \omega\lambda_1 \geq 1 - \omega\lambda_2 \geq \cdots \geq 1 - \omega\lambda_n = 1 - \omega M .$$

Es compleix, doncs, que

$$\rho(\omega) = \max(|1 - \omega M|, |1 - \omega m|) .$$

Representant les funcions, tal com mostra la figura 2.3, s'observa que:

1. Aquestes funcions coincideixen per a $\omega = 0$ i per a un altre valor $\omega = \omega^*$ i el màxim $\rho(\omega)$ d'ambdues funcions coincideix amb la segona d'elles només en l'interval $[0, \omega^*]$, i amb la primera, per als altres valors.
2. La funció ρ pren el valor 1 per a $\omega = 0$ i $\omega = \omega_c$ en què la funció $y = |1 - \omega M|$ val 1; això és, quan $\omega_c M - 1 = 1$, $\omega_c = \frac{2}{M}$. La dita funció es manté menor que 1 en l'interval $(0, \omega_c)$ i és més gran o igual que 1 per a la resta de valors reals.
3. La funció ρ és decreixent, fins al valor de $\omega = \omega^*$, i creixent, a partir d'aquest valor.

Figura 2.3: Gràfiques de les funcions $y = |1 - \omega M|$ i $y = |1 - \omega m|$.

Del punt 2 de l'estudi de la funció ρ es dedueix la convergència del mètode només per a valors de l'interval $(0, \frac{2}{M})$.

b) El valor òptim de ω es troba quan el radi espectral de la matriu d'iteració és mínim; és a dir, quan la funció ρ és mínima. Segons el punt 3, el dit mínim es troba en $\omega = \omega^*$ descrit en el punt 1 com el valor de ω , diferent de 0, en què coincideixen les dues funcions. Això és,

$$1 - \omega^* m = \omega^* M - 1, \quad \omega^* = \frac{2}{M + m}.$$

El valor mínim de la funció ρ serà, doncs,

$$\rho(\omega^*) = \omega^* M - 1 = \frac{M - m}{M + m}.$$

Problema 2.9 *Siguin A una matriu $n \times n$ regular i X_0 , una aproximació d' A^{-1} . Considerem la successió $(X_k)_{k \geq 0}$ de matrius donada per la recurrència:*

$$E_k = I - AX_k, \quad X_{k+1} = X_k(I + E_k + E_k^2) \quad (k \geq 0).$$

- a) *Vegeu que $E_{k+1} = E_k^3$ i que $X_k = A^{-1}(I - E_0^{3k})$ ($k \geq 0$).*
- b) *Doneu condicions necessàries i suficients de convergència de $(X_k)_{k \geq 0}$ cap a A^{-1} .*
- c) *Si partim d'una matriu X_0 tal que*

$$E_0 = \begin{pmatrix} 0.1 & 0.2 & 0.3 \\ 0.2 & 0.3 & 0.1 \\ 0.3 & 0.1 & 0.2 \end{pmatrix},$$

quantes iteracions s'han de fer a fi de poder garantir que $\|E_k\|_\infty < 10^{-12}$?

SOLUCIÓ:

- a) Tenim

$$\begin{aligned} E_{k+1} &= I - AX_{k+1} = I - AX_k(I + E_k + E_k^2) \\ &= I - (I - E_k)(I + E_k + E_k^2) \\ &= I - (I - E_k + E_k - E_k^2 + E_k^2 - E_k^3) = E_k^3 \end{aligned}$$

i, per inducció,

$$E_k = E_0^{3^k} \quad (k \geq 0).$$

De la relació $E_k = I - AX_k$, tenim

$$X_k = A^{-1}(I - E_k) = A^{-1}(I - E_0^{3^k}) \quad (k \geq 0).$$

b) La convergència de $(X_k)_{k \geq 0}$ cap a A^{-1} equival a la de $(E_0^{3^k})_{k \geq 0}$ cap a la matriu 0, que es dóna si i només si

$$\rho(E_0) = \rho(I - AX_0) < 1 .$$

c) Treballant amb la norma del màxim, tenim que $\|E_0\|_\infty = 0.6$ i que $\|E_k\|_\infty \leq \|E_0\|_\infty^{3^k} = 0.6^{3^k}$. La condició $0.6^{3^k} < 10^{-12}$ equival a

$$k > \frac{\log \frac{12}{- \log 0.6}}{\log 3} \simeq 3.63 .$$

Per tant, són suficients quatre iteracions.

Problema 2.10 Sigui B una matriu de Frobenius; això és, una matriu de la forma

$$B = \begin{pmatrix} b_1 & b_2 & b_3 & \cdots & \cdots & b_{n-1} & b_n \\ 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 1 & 0 \end{pmatrix} .$$

a) Trobeu el polinomi característic de B .

b) Si λ és un valor propi de B , trobeu un vector propi associat a λ .

c) Sigui $A = A_1 = (a_{ij}^{(1)})$ una matriu $n \times n$ qualsevol. Considereu les successives transformacions de semblança

$$A_{l+1} = M_{n-l}^{-1} A_l M_{n-l} \quad (l = 1 \div n-1) ,$$

essent

$$M_k = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ m_{k1} & \cdots & \cdots & m_{kk} & \cdots & \cdots & m_{kn} \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} ,$$

amb

$$m_{kj} = -\frac{a_{k+1,j}^{(n-k)}}{a_{k+1,k}^{(n-k)}} \quad (j \neq k) , \quad m_{kk} = \frac{1}{a_{k+1,k}^{(n-k)}} ,$$

suposant que $a_{k+1,k}^{(n-k)} \neq 0$ ($k = n-1 \div 1$).

Demostreu que $A^{(n)}$ és una matriu de Frobenius. Se l'anomena forma normal de Frobenius d' A . La combinació dels apartats a), b) i c) ofereix un mètode de càlcul de valors i vectors propis anomenat mètode de Danilevski.

d) Apliqueu aquest mètode de reducció a la forma normal de Frobenius a fi de trobar els valors i vectors propis de la matriu

$$A = \begin{pmatrix} 39 & -16 & -56 \\ 87 & -36 & -126 \\ -2 & 1 & 3 \end{pmatrix}.$$

SOLUCIÓ:

a) El polinomi característic es troba desenvolupant per la primera filera el determinant

$$\begin{aligned} \det(B - \lambda I) &= \begin{vmatrix} b_1 - \lambda & b_2 & b_3 & \cdots & \cdots & b_{n-1} & b_n \\ 1 & -\lambda & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 1 & -\lambda & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 1 & -\lambda \end{vmatrix} \\ &= (b_1 - \lambda) \begin{vmatrix} -\lambda & 0 & \cdots & \cdots & 0 & 0 \\ 1 & -\lambda & \cdots & \cdots & 0 & 0 \\ 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & -\lambda \end{vmatrix} \\ &\quad - b_2 \begin{vmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & -\lambda & \cdots & \cdots & 0 & 0 \\ 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & -\lambda \end{vmatrix} + \cdots \\ &\quad + (-1)^{n+1} b_n \begin{vmatrix} 1 & -\lambda & \cdots & \cdots & 0 & 0 \\ 0 & 1 & -\lambda & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{vmatrix} \\ &= (b_1 - \lambda)(-\lambda)^{n-1} - b_2(-\lambda)^{n-2} + \cdots + (-1)^{n+1} b_n \\ &= (-1)^n (\lambda^n - b_1 \lambda^{n-1} - b_2 \lambda^{n-2} - \cdots - b_n). \end{aligned}$$

Així doncs, el polinomi característic de B serà

$$p_B(\lambda) = (-1)^n (\lambda^n - b_1 \lambda^{n-1} - b_2 \lambda^{n-2} - \cdots - b_n).$$

L'expressió anterior del polinomi característic es pot obtenir també per inducció, de forma anàloga a la que es fa servir en el problema II.11.

b) Si λ és un valor propi de B , ha de complir l'equació característica

$$\lambda^n - b_1\lambda^{n-1} - b_2\lambda^{n-2} - \dots - b_n = 0 .$$

Els vectors propis $y = (y_1 \ y_2 \ \dots \ y_n)^\top$ de B , associats a λ , compleixen $By = \lambda y$; això és, el sistema d'equacions

$$\begin{aligned} b_1y_1 + b_2y_2 + b_3y_3 + \dots + b_{n-1}y_{n-1} + b_ny_n &= \lambda y_1 \\ y_{i-1} &= \lambda y_i \quad (i = 2 \div n) . \end{aligned}$$

Les $n-1$ darreres equacions permeten posar totes les components en funció de l'última y_n :

$$y_{n-i} = \lambda^i y_n \quad (i = 1 \div n-1) ;$$

la primera equació s'escriu llavors com

$$(\lambda^n - b_1\lambda^{n-1} - b_2\lambda^{n-2} - \dots - b_n)y_n = 0 ,$$

equació que és satisfeta per a qualsevol y_n , atès que λ és precisament una arrel de l'equació característica.

Així, els valors propis y de B associats a λ s'escriuen en la forma

$$y = y_n \begin{pmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \lambda \\ 1 \end{pmatrix} ;$$

prenent $y_n = 1$, donem com a vector propi associat a λ

$$y = \begin{pmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \lambda \\ 1 \end{pmatrix} .$$

c) Per portar a terme la primera iteració $A_2 = M_{n-1}^{-1}A_1M_{n-1}$, notem que la inversa de

$$M_{n-1} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ -\frac{a_{n1}^{(1)}}{a_{n,n-1}^{(1)}} & \dots & \dots & \frac{1}{a_{n,n-1}^{(1)}} & -\frac{a_{nn}^{(1)}}{a_{n,n-1}^{(1)}} \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix}$$

ve donada per

$$M_{n-1}^{-1} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ a_{n1}^{(1)} & \cdots & \cdots & a_{n,n-1}^{(1)} & a_{nn}^{(1)} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix};$$

aleshores,

$$\begin{aligned} A_2 &= M_{n-1}^{-1} A_1 M_{n-1} \\ &= \begin{pmatrix} a_{11}^{(2)} & a_{12}^{(2)} & \cdots & \cdots & a_{1,n-1}^{(2)} & a_{1n}^{(2)} \\ \vdots & \vdots & & & \vdots & \vdots \\ a_{n-1,1}^{(2)} & a_{n-1,2}^{(2)} & \cdots & \cdots & a_{n-1,n-1}^{(2)} & a_{n-1,n}^{(2)} \\ 0 & 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}. \end{aligned}$$

Usant que la matriu M_k té per inversa

$$M_k^{-1} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ a_{k+1,1}^{(n-k)} & \cdots & \cdots & a_{k+1,k}^{(n-k)} & \cdots & a_{k+1,n}^{(n-k)} \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix},$$

i sabent que A_l té les darreres $l-1$ darreres fileres de les $l-1$ darreres columnes en forma de matriu de Frobenius, es dedueix que

$$A_{l+1} = M_{n-l}^{-1} A_l M_{n-l}$$

té les l darreres fileres de les l darreres columnes en forma de matriu de Frobenius ($l = 1 \div n-1$).

Així, amb $n-1$ passos s'arriba a la forma normal de Frobenius de la matriu A

$$B \equiv A_n = \begin{pmatrix} a_{11}^{(n)} & a_{12}^{(n)} & a_{13}^{(n)} & \cdots & a_{1,n-1}^{(n)} & a_{1n}^{(n)} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

d) Emprant el mètode anterior sobre

$$A_1 \equiv A = \begin{pmatrix} 39 & -16 & -56 \\ 87 & -36 & -126 \\ -2 & 1 & 3 \end{pmatrix},$$

trobem recursivament:

$$M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}, \quad M_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix},$$

$$A_2 = M_2^{-1} A_1 M_2 = \begin{pmatrix} 7 & -16 & -8 \\ 1 & -1 & -2 \\ 0 & 1 & 0 \end{pmatrix};$$

$$M_1 = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad M_1^{-1} = \begin{pmatrix} 1 & -1 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$B = A_3 = M_1^{-1} A_2 M_1 = \begin{pmatrix} 6 & -11 & 6 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

El polinomi característic de B és, doncs,

$$p_B(\lambda) = -\lambda^3 + 6\lambda^2 - 11\lambda + 6.$$

Els valors propis de B , i també d' A , són així:

$$\lambda_1 = 1, \quad \lambda_2 = 2, \quad \lambda_3 = 3.$$

Usant l'apartat b), podem donar una base de vectors propis de B ,

$$y^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad y^{(2)} = \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}, \quad y^{(3)} = \begin{pmatrix} 9 \\ 3 \\ 1 \end{pmatrix}.$$

Com que B es troba a partir d' A aplicant dues transformacions de semblança amb les matrius M_2 i M_1 , trobarem una base de vectors propis d' A fent

$$v^{(j)} = M_2 M_1 y^{(j)} \quad (j = 1 \div 3).$$

Resulten

$$v^{(1)} = \begin{pmatrix} 4 \\ 6 \\ 1 \end{pmatrix}, \quad v^{(2)} = \begin{pmatrix} 8 \\ 15 \\ 1 \end{pmatrix}, \quad v^{(3)} = \begin{pmatrix} 14 \\ 28 \\ 1 \end{pmatrix}.$$

Problema 2.11 Sigui B_n una matriu de la forma

$$\begin{pmatrix} b_1 & b_2 & b_3 & \cdots & \cdots & b_{n-1} & b_n \\ a & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & a & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & a & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & a & 0 \end{pmatrix},$$

amb $b_j = \frac{K}{2^j}$ ($j = 1 \div n$).

a) Considerem el cas en què a i K són reals no negatius. Responen a les preguntes següents:

- i) si n és parell, poden ser tots els valors propis de B_n reals positius?
- ii) poden ser negatives totes les parts reals dels valors propis?

b) Trobeu l'equació característica i els valors propis de B_n per a $a = 2$ i $K = -2$.

c) Considerem un mètode iteratiu general del tipus

$$x^{(k+1)} = B_n x^{(k)} + c$$

per a c qualsevol. Estudieu la seva convergència per a $|a| < 1$ i $|K| \leq 1$.

SOLUCIÓ:

a) i) Si n és parell, no poden ser tots els valors propis de B_n reals positius; si fos així, el producte de tots ells seria positiu, però

$$\prod_{j=1}^n \lambda_j = \det B_n = (-1)^{n-1} b_n a^{n-1} = (-1)^{n-1} \frac{K}{2^n} a^{n-1} \leq 0,$$

si $a \geq 0$, $K \geq 0$ i n parell.

Les parts reals dels valors propis no poden ser totes negatives; en tal cas, la seva suma seria també negativa, però, d'altra banda,

$$\sum_{j=1}^n \Re(\lambda_j) = \sum_{j=1}^n \lambda_j = \operatorname{tr} B_n = b_1 = \frac{K}{2} \geq 0,$$

si $K \geq 0$.

La primera igualtat es basa en què els valors propis són els zeros d'un polinomi a coeficients reals; els zeros no reals d'aquest polinomi han de ser conjugats dos a dos de manera que la seva suma serà sempre real.

b) Hem de calcular $\Delta_n = \det (B_n - \lambda I)$ per a

$$b_j = -\frac{2}{2^j} = -\frac{1}{2^{j-1}} \quad \text{ i } \quad a = 2.$$

Fem-ho primerament per a $n = 2$ i per a $n = 3$:

$$\begin{aligned}\Delta_2 &= \begin{vmatrix} b_1 - \lambda & b_2 \\ a & -\lambda \end{vmatrix} = \lambda^2 + \lambda + 1, \\ \Delta_3 &= \begin{vmatrix} b_1 - \lambda & b_2 & b_3 \\ a & -\lambda & 0 \\ 0 & a & -\lambda \end{vmatrix} = -(\lambda^3 + \lambda^2 + \lambda + 1) .\end{aligned}$$

Això ens fa pensar que

$$\Delta_n = (-1)^n(\lambda^n + \lambda^{n-1} + \cdots + \lambda + 1) .$$

Ho demostrarem per inducció. Suposem que

$$\Delta_{n-1} = (-1)^{n-1}(\lambda^{n-1} + \lambda^{n-2} + \cdots + \lambda + 1) ;$$

desenvolupant B_n per l'última columna, tenim

$$\Delta_n = -\lambda\Delta_{n-1} + (-1)^{n-1}b_na^{n-1} .$$

Així doncs,

$$\begin{aligned}\Delta_n &= -\lambda(-1)^{n-1}(\lambda^{n-1} + \lambda^{n-2} + \cdots + \lambda + 1) \\ &\quad + (-1)^{n-1}\left(-\frac{1}{2^{n-1}}\right)2^{n-1} \\ &= (-1)^n(\lambda^n + \lambda^{n-1} + \cdots + \lambda^2 + \lambda) + (-1)^n \\ &= (-1)^n(\lambda^n + \lambda^{n-1} + \cdots + \lambda + 1) ,\end{aligned}$$

que és el que es volia demostrar.

El valors propis de B_n són aquells valors de λ que anul·len

$$\Delta_n(\lambda) = (-1)^n(\lambda^n + \lambda^{n-1} + \cdots + \lambda + 1) = (-1)^n \frac{\lambda^{n+1} - 1}{\lambda - 1} \quad (\lambda \neq 1) ;$$

això és,

$$\lambda^{n+1} = 1 , \quad \lambda \neq 1 .$$

Els valors propis són, doncs, les arrels $(n+1)$ -enes de la unitat, exceptuant $\lambda = 1$.

c) Una condició necessària i suficient de convergència és que el radi espectral de B_n sigui menor que 1.

Aplicant el teorema de Gerschgorin, sabem que els valors propis de B_n són a la unió dels discs:

$$\{\lambda : |\lambda - b_1| \leq \sum_{j=2}^n |b_j|\} , \quad \{\lambda : |\lambda - 0| \leq |a|\} .$$

Per als valors de λ al primer disc, tenim

$$|\lambda - \frac{K}{2}| \leq \sum_{j=2}^n \frac{|K|}{2^j} = |K| \left(\frac{1}{2} - \frac{1}{2^n} \right) < \frac{|K|}{2} ;$$

així,

$$|\lambda| \leq \left| \lambda - \frac{K}{2} \right| + \left| \frac{K}{2} \right| < |K| \leq 1,$$

ja que se suposa $|K| \leq 1$.

Per als valors de λ a l'altre disc:

$$|\lambda| \leq |a| < 1,$$

suposant $|a| < 1$.

Així doncs, ambdós discs estan continguts en el disc unitat, centrat a l'origen i, per tant, el radi espectral de B_n és més petit que 1; consegüentment, el mètode iteratiu donat és sempre convergent.

Problema 2.12 Sabem que el vector $(1 \ 1 \ 1 \ 1)^\top$ és un vector propi de valor propi 32 de la matriu

$$A = \begin{pmatrix} 19 & 11 & 1 & 1 \\ 3 & 3 & 13 & 13 \\ 5 & 5 & 15 & 7 \\ 9 & 9 & 7 & 7 \end{pmatrix}.$$

Utilitzeu la deflació de Householder per tal de trobar els altres valors propis i una base de vectors propis d' A .

SOLUCIÓ:

Dient $v = (1 \ 1 \ 1 \ 1)^\top$, hem d'escollir s tal que $s^2 = \|v\|_2^2 = 4$ per tal de formar el vector $u = v - se_1$ tal que $P(u)v = se_1$. Escollim, per exemple, $s = 2$ (normalment, s'escull s de signe contrari al de v_1 , ja que no tenim problemes greus de cancel·lació).

Així,

$$u = v - se_1 = (-1 \ 1 \ 1 \ 1)^\top, \quad \alpha = \frac{2}{v^\top v} = \frac{-1}{s_1 u_1} = \frac{1}{2}.$$

La matriu de Householder $P = P(u) = I - \alpha uu^\top$ satisfà $Pv = 2e_1$, i amb ella procedim al càlcul d' $A' = PAP$.

Primerament calculem

$$B = PA = (I - \alpha uu^\top)A = A - u(\alpha u)^\top A = A - uw^\top A = A - up^\top,$$

on

$$w = \alpha u = \frac{1}{2}(-1 \ 1 \ 1 \ 1)^\top, \quad p^\top = w^\top A = (-1 \ 3 \ 17 \ 13);$$

aleshores,

$$B = A - \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 3 & 17 & 13 \end{pmatrix} = \begin{pmatrix} 18 & 14 & 18 & 14 \\ 4 & 0 & -4 & 0 \\ 6 & 2 & -2 & -6 \\ 10 & 6 & -10 & -6 \end{pmatrix}.$$

Finalment, procedim al càlcul de la matriu transformada

$$A' = PAP = BP = B(I - \alpha uu^\top) = B - (Bw)u^\top = B - qu^\top,$$

on $q = Bw = (14 \quad -4 \quad -6 \quad -10)^\top$, resultant

$$A' = B - \begin{pmatrix} 14 \\ -4 \\ -6 \\ -10 \end{pmatrix} \begin{pmatrix} -1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 32 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 0 & 8 & 4 & 0 \\ 0 & 16 & 0 & 4 \end{pmatrix}.$$

La matriu A' té els mateixos valors propis que A ; un dels valors propis és $\lambda_1 = 32$ i els altres valors propis són els de la matriu reduïda

$$\tilde{A} = \begin{pmatrix} 4 & 0 & 4 \\ 8 & 4 & 0 \\ 16 & 0 & 4 \end{pmatrix},$$

que trobem calculant el seu polinomi característic

$$\det(\tilde{A} - \lambda I) = \begin{vmatrix} 4 - \lambda & 0 & 4 \\ 8 & 4 - \lambda & 0 \\ 16 & 0 & 4 - \lambda \end{vmatrix} = (4 - \lambda)(12 - \lambda)(-4 - \lambda).$$

Els seus zeros

$$\lambda_2 = 12, \quad \lambda_3 = 4, \quad \lambda_4 = -4,$$

són llavors els valors propis d' $\tilde{A}$.

Notem que, si $\tilde{v}$ és un vector propi d' $\tilde{A}$ de valor propi λ , aleshores

$$v = \begin{pmatrix} v_1 \\ \tilde{v} \end{pmatrix}, \quad \text{amb} \quad v_1 = \frac{c^\top \tilde{v}}{\lambda - \lambda_1},$$

és valor propi de

$$A' = \left(\begin{array}{c|c} \lambda_1 & c^\top \\ \hline 0 & \tilde{A} \end{array} \right).$$

Una base de vectors propis d' $\tilde{A}$, associats a λ_2, λ_3 i λ_4 , es troba directament:

$$\tilde{v}^{(2)} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad \tilde{v}^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{v}^{(4)} = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}.$$

Calculant primer

$$\tilde{v}_1^{(2)} = \frac{4 \cdot 1}{12 - 32} = -\frac{1}{5}, \quad \tilde{v}_1^{(3)} = \frac{4 \cdot 1}{4 - 32} = -\frac{1}{7}, \quad \tilde{v}_1^{(4)} = \frac{4 \cdot 1}{-4 - 32} = -\frac{1}{9};$$

tenim que

$$v'^{(2)} = \begin{pmatrix} -\frac{1}{5} \\ 1 \\ 1 \\ 2 \end{pmatrix}, \quad v'^{(3)} = \begin{pmatrix} -\frac{1}{7} \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad v'^{(4)} = \begin{pmatrix} -\frac{1}{9} \\ -1 \\ 1 \\ 2 \end{pmatrix},$$

són vectors propis d' $A' = PAP$ de valors propis associats λ_2 , λ_3 i λ_4 , respectivament.

Aplicant-los la matriu P trobarem els vectors propis d' A :

$$v^{(2)} = \frac{1}{10} \begin{pmatrix} 19 \\ -11 \\ -11 \\ -1 \end{pmatrix}, \quad v^{(3)} = \frac{1}{14} \begin{pmatrix} 6 \\ -8 \\ 6 \\ -8 \end{pmatrix}, \quad v^{(4)} = \frac{1}{18} \begin{pmatrix} 17 \\ -37 \\ -1 \\ 17 \end{pmatrix}.$$

Una base de vectors propis d' A serà, doncs, la formada pels vectors:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 19 \\ -11 \\ -11 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 3 \\ -4 \\ 3 \\ -4 \end{pmatrix}, \quad \begin{pmatrix} 17 \\ -37 \\ -1 \\ 17 \end{pmatrix}.$$

Problema 2.13 Trobeu tots els valors propis de la matriu

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 1 & 1 & 1 & 4 \end{pmatrix},$$

emprant els mètodes de la potència i de la potència inversa.

SOLUCIÓ:

Construïm primerament la successió de vectors definits per

$$x^{(0)} = (1 \ 1 \ 1 \ 1)^\top, \quad x^{(k+1)} = Ax^{(k)} \quad (k \geq 0),$$

i la de vectors de quocients $q^{(k)}$ de components

$$q_i^{(k)} = \frac{x_j^{(k)}}{x_j^{(k-1)}} \quad (i = 1 \div n) \quad (k \geq 1).$$

Els resultats s'exposen en la taula següent:

k	1	2	3	4	5	6	7	8	9
$x^{(k)}$	4	22	126	728	4218	24464	141946	823740	4780650
	5	27	153	881	5099	29563	171509	995249	5775899
	6	34	194	1116	6450	37364	216674	1257088	7294826
	7	43	255	1493	8697	50555	293611	1704573	9894369
$q^{(k)}$	4	5.5	5.73	5.778	5.7999	5.8022	5.80319	5.80359	5.80376
	5	5.4	5.67	5.758	5.7978	5.8014	5.80289	5.80347	5.80371
	6	5.7	5.70	5.753	5.7929	5.7999	5.80175	5.80296	5.80348
	7	6.2	5.93	5.855	5.8129	5.8077	5.80555	5.80460	5.80420

Observem la convergència lineal de les components dels vectors de quocients al valor propi de mòdul màxim $\lambda_1 \simeq 5.804$.

L'ús de quocients de Rayleigh porta a una successió més ràpidament convergent

$$\begin{aligned}\sigma_k &= \frac{x^{(k+1)\top} x^{(k)}}{x^{(k)\top} x^{(k)}} \\ &= 5.5, 5.77, 5.7999, 5.80319, 5.80376, 5.803863, 5.8038820, \\ &\quad 5.80388554, 5.803886331, \dots \quad (k \geq 0)\end{aligned}$$

al valor propi $\lambda_1 \simeq 5.8038863$, que podem donar amb 7 xifres decimals correctes.

La matriu inversa d' A és

$$A^{-1} = \frac{1}{6} \begin{pmatrix} 17 & -6 & -3 & -2 \\ -6 & 6 & 0 & 0 \\ -3 & 0 & 3 & 0 \\ -2 & 0 & 0 & 2 \end{pmatrix};$$

repetint el procés per a $B = 6A^{-1}$, per tal de portar a terme les iteracions amb enters, i partint ara de $x^{(0)} = (1 \ 0 \ 0 \ 0)^\top$, obtenim

k	1	2	3	4	5	6	7
$x^{(k)}$	17	338	6830	138332	2802884	56796776	1150932152
	-6	-138	-2856	-58116	-1178688	-23889432	-484117248
	3	-60	-1194	-24072	-487212	-9870288	-200001192
	2	-38	-752	-15164	-306992	-6219752	-126033056
$q^{(k)}$		19.	20.21	20.254	20.262	20.26404	20.264114
		23.	20.70	20.349	20.282	20.26491	20.264298
		19.	20.21	20.254	20.262	20.26404	20.263879
		19.	20.21	20.254	20.262	20.26404	20.263972

La convergència al valor propi dominant μ_1 de B és més ràpida, si usem els quocients de Rayleigh; notem que

$$\sigma_5 = 20.26413315, \quad \sigma_6 = 20.26413316.$$

El valor propi més petit en mòdul d' A serà, llavors,

$$\lambda_4 = \frac{6}{\mu_1} \simeq \frac{6}{\sigma_6} = 0.296090.$$

Els altres valors propis λ_2 i λ_3 es calculen ara fàcilment, usant que $\text{tr } A$ i $\det A$ són respectivament la suma i el producte dels valors propis:

$$\begin{aligned}\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = \text{tr } A &= 10 , \\ \lambda_1 \lambda_2 \lambda_3 \lambda_4 = \det A &= 6 .\end{aligned}$$

O sigui,

$$\begin{aligned}\lambda_2 + \lambda_3 &= 10 - \lambda_1 - \lambda_4 = 3.990024 , \\ \lambda_2 \lambda_3 &= \frac{6}{\lambda_1 \lambda_4} = 3.491473 ;\end{aligned}$$

d'on, els dos valors propis restants d' A són

$$\lambda_2 = 2.507752 , \quad \lambda_3 = 1.392272 .$$

Problema 2.14 *Apliqueu el mètode de Jacobi per tal de trobar els valors i vectors propis de la matriu simètrica*

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 6 \end{pmatrix} ,$$

fins que els elements no diagonals siguin tots menors que 10^{-6} .

SOLUCIÓ:

Emprarem el mètode clàssic de Jacobi. L'element maximal no diagonal de $B \equiv A_1 \equiv A$ és $b_{23} = 3$; cal aplicar primer una rotació R_1 , en el pla coordinat 2-3, que anul·li l'element $\bar{b}_{23}$ de $\bar{B} = R_1^\top B R_1$. El cosinus c_1 i el sinus s_1 de l'angle de la rotació es troben fent:

$$\begin{aligned}d_1 &= \sqrt{(b_{22} - b_{33})^2 + 4b_{23}^2} = \sqrt{37} , \\ c_1 &= \sqrt{\frac{1}{2}\left(1 + \frac{1}{d_1}\right)} = 0.7630199825 , \\ s_1 &= -\sqrt{\frac{1}{2}\left(1 - \frac{1}{d_1}\right)} = -0.6463748961 ;\end{aligned}$$

i els elements de B es transformen segons

$$\begin{aligned}\bar{b}_{12} &= b_{12}c_1 + b_{13}s_1 = 0.8796650689 , \\ \bar{b}_{13} &= -b_{12}s_1 + b_{13}c_1 = 2.055769775 , \\ \bar{b}_{22} &= b_{22}c_1^2 + 2b_{23}s_1c_1 + b_{33}s_1^2 = b_{22} + b_{23}t_1 = 2.458618735 , \\ \bar{b}_{33} &= b_{22}s_1^2 - 2b_{23}s_1c_1 + b_{33}c_1^2 = -b_{23}t_1 + b_{33} = 8.541381266 ,\end{aligned}$$

on s'ha emprat $t_1 = \frac{s_1}{c_1}$. Recordem que $\bar{b}_{11} = b_{11}$ i que $\bar{b}_{23} = 0$, per construcció.
Tenim així,

$$A_2 \equiv \bar{B} = \begin{pmatrix} 4 & 0.879665069 & 2.055769775 \\ 0.879665069 & 2.458618735 & 0 \\ 2.055769775 & 0 & 8.541381266 \end{pmatrix}.$$

Ara, l'element no diagonal maximal de $B \equiv A_2$ és b_{13} ; la rotació s'ha de fer, doncs, en el pla 1-3. Els sinus i cosinus es troben de forma anàloga, essent

$$c_2 = 0.9330913253, \quad s_2 = -0.3596395121;$$

la matriu resultant és

$$A_3 = R_2^\top A_2 R_2 = \begin{pmatrix} 3.207648792 & 0.820807845 & 0 \\ 0.820807845 & 2.458618735 & 0.316362316 \\ 0 & 0.316362316 & 9.333732474 \end{pmatrix}.$$

Repetint el procés, s'obtenen els sinus i cosinus dels angles de les rotacions i matrius transformades:

$$c_3 = 0.8411621222, \quad s_3 = 0.540783029;$$

$$A_4 = R_3^\top A_3 R_3 = \begin{pmatrix} 3.735346059 & 0 & 0.171083372 \\ 0 & 1.930921469 & 0.266111997 \\ 0.171083372 & 0.266111997 & 9.333732474 \end{pmatrix}.$$

$$c_4 = 0.999356178, \quad s_4 = 0.035877979;$$

$$A_5 = R_4^\top A_4 R_4 = \begin{pmatrix} 3.735346059 & -0.00613812 & 0.170973224 \\ -0.00613812 & 1.921367757 & 0 \\ 0.170973224 & 0 & 9.343286185 \end{pmatrix}.$$

$$c_5 = 0.999536434, \quad s_5 = -0.0304452993;$$

$$A_6 = R_5^\top A_5 R_5 = \begin{pmatrix} 3.730138314 & -0.00613528 & 0 \\ -0.00613528 & 1.921367757 & -0.000186877 \\ 0 & -0.00018688 & 9.343286185 \end{pmatrix}.$$

$$c_6 = 0.999994248, \quad s_6 = -0.003391903;$$

$$A_7 = R_6^\top A_6 R_6 = \begin{pmatrix} 3.730159124 & 0 & 0.000000634 \\ 0 & 1.921346947 & -0.000186877 \\ 0.000000634 & -0.00018688 & 9.348493935 \end{pmatrix}.$$

$$c_7 = 0.999999997, \quad s_7 = 0.0000251595;$$

$$A_8 = R_7^T A_7 R_7 = \begin{pmatrix} 3.730159124 & 0.000000000 & 0.000000634 \\ 0.000000000 & 1.921346942 & 0 \\ 0.000000634 & 0 & 9.348493935 \end{pmatrix}.$$

Observem que les 9 primeres xifres significatives dels elements de la diagonal no han variat en l'última iteració del mètode; donem així les aproximacions següents dels valors propis:

$$\lambda_1 \simeq 3.73015912, \quad \lambda_2 \simeq 1.92134694, \quad \lambda_3 \simeq 9.34849394.$$

Una aproximació dels vectors propis ortonormals associats a λ_1 , λ_2 i λ_3 és donada per les columnes de la matriu ortogonal O_8 que és producte de totes les rotacions fetes

$$O_8 = R_1 R_2 \cdots R_7,$$

atès que

$$O_8^T A O_8 = A_8 \simeq \text{diag}(\lambda_1, \lambda_2, \lambda_3).$$

Per tal de calcular O_3 a partir de $\Omega = O_2 = R_1$, s'ha de fer el producte de matrius

$$\bar{\Omega} = O_3 = R_2 O_2 = R_2 \Omega,$$

en què només es modifiquen les fileres 1 i 3 de la matriu Ω segons

$$\bar{\omega}_{1j} = \omega_{1j} c_2 + \omega_{3j} s_2, \quad \bar{\omega}_{3j} = -\omega_{1j} s_2 + \omega_{3j} c_2 \quad (j = 1, 2, 3).$$

Resulta així,

$$O_3 = \begin{pmatrix} 0.933091325 & 0 & 0.359639512 \\ -0.232461952 & 0.763019983 & 0.603126808 \\ -0.274412134 & 0.646374896 & 0.711967327 \end{pmatrix}.$$

Repetint el procés, es troba finalment la matriu formada pels vectors propis d' A

$$O_8 = \begin{pmatrix} 0.774686009 & -0.488736998 & 0.401245229 \\ 0.196800996 & 0.789366503 & 0.581523767 \\ -0.600600794 & -0.517791163 & 0.713020526 \end{pmatrix}.$$

Problema 2.15 Reduïu la matriu

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 1 & 1 & 1 & 4 \end{pmatrix}$$

a forma tridiagonal simètrica pel mètode de Householder i trobeu després els seus valors i vectors propis.

SOLUCIÓ:

Comencem amb el procés de reducció a forma tridiagonal simètrica. Això s'aconseguirà a través de dues transformacions de semblança per matrius de Householder sobre la matriu A .

Per això, partim d' $A_2 = A$ i triem una matriu de Householder de la forma

$$P_2 = \left(\begin{array}{c|cc} 1 & & 0 \\ \hline & & \\ 0 & & \tilde{P}_2 \end{array} \right),$$

amb $\tilde{P}_2 = P(\tilde{u}^{(2)}) = I_3 - \alpha_2 \tilde{u}^{(2)} \tilde{u}^{(2)\top}$, on I_3 indica la matriu identitat de dimensió 3 i el vector $\tilde{u}^{(2)}$, que té dimensió 3, és tal que la matriu $\tilde{P}_2$ transforma el vector $\tilde{a}^{(2)} = (1 \ 1 \ 1)^\top$ en un vector amb la segona i tercera components nul·les,

$$\tilde{P}_2 \tilde{a}^{(2)} = s_2 \tilde{e}^{(1)} \equiv \begin{pmatrix} s_2 \\ 0 \\ 0 \end{pmatrix},$$

on $s_2^2 = \|\tilde{a}^{(2)}\|_2^2$.

Els càlculs de s_2 , α_2 i $\tilde{u}^{(2)}$ es fan a continuació:

$$\begin{aligned} s_2 &= -\|\tilde{a}^{(2)}\|_2 = -\sqrt{3}, \\ \tilde{u}^{(2)} &= \begin{pmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{pmatrix}, \\ \alpha_2 &= \frac{1}{\sqrt{3}(\sqrt{3} + 1)} = \frac{3 - \sqrt{3}}{6}, \end{aligned}$$

La matriu transformada $A_3 = P_2 A_2 P_2$ és llavors de la forma

$$A_3 = \left(\begin{array}{c|ccc} 1 & -\sqrt{3} & 0 & 0 \\ \hline -\sqrt{3} & & & \\ 0 & & \tilde{P}_2 \tilde{A}_2 \tilde{P}_2 & \\ 0 & & & \end{array} \right).$$

Per al càlcul de $\tilde{P}_2 \tilde{A}_2 \tilde{P}_2$, es fan els càlculs previs següents:

$$\begin{aligned} \tilde{w}^{(2)} &= \alpha_2 \tilde{u}^{(2)} = \frac{1}{3 + \sqrt{3}} \begin{pmatrix} 1 + \sqrt{3} \\ 1 \\ 1 \end{pmatrix}, \\ \tilde{p}^{(2)} &= \alpha_2 \tilde{A}_2 \tilde{w}^{(2)} = \frac{1}{3 + \sqrt{3}} \begin{pmatrix} 4 + 2\sqrt{3} \\ 5 + \sqrt{3} \\ 6 + \sqrt{3} \end{pmatrix}, \\ \beta_2 &= \tilde{w}^{(2)\top} \tilde{p}^{(2)} = \frac{21 + 8\sqrt{3}}{(3 + \sqrt{3})^2}, \end{aligned}$$

$$\begin{aligned}
\tilde{q}^{s(2)} &= \tilde{p}^{(2)} - \frac{1}{2}\beta_2\tilde{u}^{(2)} = \frac{1}{12} \begin{pmatrix} 9 - 9\sqrt{3} \\ 6 + \sqrt{3} \\ 12 - \sqrt{3} \end{pmatrix}, \\
\tilde{N}_2 &= \tilde{u}^{(2)}\tilde{q}^{s(2)\top} = \frac{1}{12} \begin{pmatrix} -18 & 9 + 7\sqrt{3} & 9 + 11\sqrt{3} \\ 9 - 9\sqrt{3} & 6 + \sqrt{3} & 12 - \sqrt{3} \\ 9 - 9\sqrt{3} & 6 + \sqrt{3} & 12 - \sqrt{3} \end{pmatrix}, \\
\tilde{N}_2 + \tilde{N}_2^\top &= \begin{pmatrix} -3 & \frac{3}{2} - \frac{\sqrt{3}}{6} & \frac{3}{2} + \frac{\sqrt{3}}{6} \\ \frac{3}{2} - \frac{\sqrt{3}}{6} & 1 + \frac{\sqrt{3}}{6} & \frac{3}{2} \\ \frac{3}{2} + \frac{\sqrt{3}}{6} & \frac{3}{2} & 2 - \frac{\sqrt{3}}{6} \end{pmatrix}.
\end{aligned}$$

La matriu $\tilde{P}_2\tilde{A}_2\tilde{P}_2$ és llavors igual a $\tilde{A}_2 - (\tilde{N}_2 + \tilde{N}_2^\top)$ i la matriu transformada és

$$A_3 = \begin{pmatrix} 1 & -\sqrt{3} & 0 & 0 \\ -\sqrt{3} & 5 & -\frac{1}{2} + \frac{\sqrt{3}}{6} & -\frac{1}{2} - \frac{\sqrt{3}}{6} \\ 0 & -\frac{1}{2} + \frac{\sqrt{3}}{6} & 2 - \frac{\sqrt{3}}{6} & -\frac{1}{2} \\ 0 & -\frac{1}{2} - \frac{\sqrt{3}}{6} & -\frac{1}{2} & 2 + \frac{\sqrt{3}}{6} \end{pmatrix},$$

que notem en la forma

$$\left(\begin{array}{cc|cc} & & 0 & 0 \\ T_3 & & & \\ \hline & & \tilde{a}^{(3)\top} & \\ 0 & & & \\ & \tilde{a}^{(3)} & & \tilde{A}_3 \\ 0 & & & \end{array} \right).$$

Triem ara, d'una manera anàloga, una matriu de Householder de la forma

$$P_3 = \left(\begin{array}{c|c} I_2 & 0 \\ \hline 0 & \tilde{P}_3 \end{array} \right),$$

amb $\tilde{P}_3 = I_2 - \alpha_3\tilde{u}^{(3)}\tilde{u}^{(3)\top}$, on I_2 indica la matriu identitat de dimensió 2 i el vector $\tilde{u}^{(3)}$, que té dimensió 2, és tal que $\tilde{P}_3$ transforma el vector

$$\tilde{a}^{(3)} = \begin{pmatrix} -\frac{1}{2} + \frac{\sqrt{3}}{6} \\ -\frac{1}{2} - \frac{\sqrt{3}}{6} \end{pmatrix}$$

en un vector amb la segona component nul·la,

$$\tilde{P}_3\tilde{a}^{(3)} = s_3\tilde{e}^{(2)} \equiv \begin{pmatrix} s_3 \\ 0 \end{pmatrix},$$

on $s_3^2 = \|\tilde{a}^{(3)}\|_2^2$.

Els càlculs de s_3 , α_3 i $\tilde{u}^{(3)}$ es fan a continuació:

$$\begin{aligned} s_3 &= \|a^{(3)}\|_2 = \frac{\sqrt{6}}{3}, \\ \tilde{u}^{(3)} &= \begin{pmatrix} -\frac{1}{2} + \frac{\sqrt{3}}{6} - \frac{\sqrt{6}}{3} \\ -\frac{1}{2} - \frac{\sqrt{3}}{6} \\ 6 \end{pmatrix}, \\ \alpha_3 &= \frac{6}{4 + \sqrt{6} - \sqrt{2}}. \end{aligned}$$

La matriu transformada $A_4 = P_3 A_3 P_3$ és llavors de la forma

$$A_4 = \left(\begin{array}{cc|cc} 1 & -\sqrt{3} & 0 & 0 \\ -\sqrt{3} & 5 & \frac{\sqrt{6}}{3} & 0 \\ 0 & \frac{\sqrt{6}}{3} & \tilde{P}_3 \tilde{A}_3 \tilde{P}_3 & \\ 0 & 0 & & \end{array} \right).$$

El càlcul de $\tilde{P}_3 \tilde{A}_3 \tilde{P}_3$, es fa d'una manera similar i es troba finalment la matriu tridimensional simètrica

$$T \equiv A_4 = \begin{pmatrix} 1 & -\sqrt{3} & 0 & 0 \\ -\sqrt{3} & 5 & \frac{\sqrt{6}}{3} & 0 \\ 0 & \frac{\sqrt{6}}{3} & 2 & -\frac{\sqrt{3}}{3} \\ 0 & 0 & -\frac{\sqrt{3}}{3} & 2 \end{pmatrix}.$$

El polinomi característic $p_4(\lambda)$ es calcula per recurrència dels determinants $p_j(\lambda)$ ($j = 1 \div 4$) de les matrius principals de $\lambda I - T$:

$$\begin{aligned} p_0(\lambda) &\equiv 1, \\ p_1(\lambda) &= \lambda - 1, \\ p_2(\lambda) &= (\lambda - 5)p_1(\lambda) - 3p_0(\lambda), \\ p_3(\lambda) &= (\lambda - 2)p_2(\lambda) - \frac{2}{3}p_1(\lambda), \\ p_4(\lambda) &= (\lambda - 2)p_3(\lambda) - \frac{1}{3}p_1(\lambda). \end{aligned}$$

Aquests polinomis formen la successió de Sturm

$$\{p_4(\lambda), p_3(\lambda), p_2(\lambda), p_1(\lambda), p_0(\lambda)\}$$

a la qual aplicarem el teorema de Sturm per tal de trobar els zeros de $p_4(\lambda)$. Aquests són els valors propis de T i també els d' A .

Aquests zeros són tots reals, en ser la matriu simètrica, i pel teorema de Gerschgorin aplicat a la matriu original A , sabem que són a l'interval $[-2, 7]$.

Emprarem a continuació el teorema d'Sturm per tal de separar els valors propis. Caldrà calcular, per a diversos valors de $\lambda \in [-2, 7]$, el nombre de variacions de signe de la successió

$V(\lambda)$; recordem que $V(\lambda)$ indica, llavors, el nombre de valors propis de la matriu més grans que λ .

En la taula següent es representen els signes dels valors dels polinomis $p_j(\lambda)$ ($j = 0 \div 4$) i el nombre corresponent de variacions de signe de la successió, per a diversos valors de λ ,

λ	-2	-1	0	1	2	3	4	5	6
$p_0(\lambda)$	+	+	+	+	+	+	+	+	+
$p_1(\lambda)$	-	-	-	0	+	+	+	+	+
$p_2(\lambda)$	+	+	+	-	-	-	-	-	+
$p_3(\lambda)$	-	-	-	+	-	-	-	-	+
$p_4(\lambda)$	+	+	+	-	+	-	-	-	+
$V(\lambda)$	4	4	4		3	2	1	1	0

Així tenim una arrel en $(0, 1)$, una en $(1, 2)$, una en $(2, 3)$ i finalment una altra en $(5, 6)$.

Fent ara unes iteracions dels mètodes de bisecció i de la secant, descrits en el capítol V, es troben els valors propis següents:

$$\lambda_1 = 5.803886359... \quad , \quad \lambda_2 = 2.507748705... \quad ,$$

$$\lambda_3 = 1.392275291... \quad , \quad \lambda_4 = 0.296089645...$$

A fi de trobar els vectors propis $v^{(j)}$ ($j = 1 \div 4$) de

$$T = \begin{pmatrix} a_1 & b_2 & & \\ b_2 & a_2 & b_3 & \\ & b_3 & a_3 & b_4 \\ & & b_4 & a_4 \end{pmatrix} \quad ,$$

es resolen els sistemes tridiagonals simètrics $(\lambda I - T)v'' = 0$:

$$\left. \begin{aligned} (\lambda - a_1)v_1'' &+ b_2v_2'' & & & & & = 0 \\ b_2v_1'' &+ (\lambda - a_2)v_2'' & & + b_3v_3'' & & & = 0 \\ & b_3v_2'' &+ (\lambda - a_3)v_3'' & & + b_4v_4'' & & = 0 \\ & & + b_4v_3'' &+ (\lambda - a_4)v_4'' & & & = 0 \end{aligned} \right\} \quad ,$$

per a $\lambda = \lambda_j$ i $v'' = v''^{(j)}$ ($j = 1 \div 4$).

Prenent $v_1'' = 1$, resulta

$$v_2'' = \frac{p_1(\lambda)}{b_2} \quad , \quad v_3'' = \frac{p_2(\lambda)}{b_2b_3} \quad , \quad v_4'' = \frac{p_3(\lambda)}{b_2b_3b_4} \quad .$$

Aquestes fórmules donen els vectors propis de T associats als valors propis trobats:

$$v^{(1)} = \begin{pmatrix} 1.0000000000 \\ -2.7735250825 \\ -0.6093695727 \\ 0.0924895353 \end{pmatrix} \quad , \quad v^{(2)} = \begin{pmatrix} 1.0000000000 \\ -0.8704991207 \\ 4.7784074785 \\ -5.4334256556 \end{pmatrix} \quad ,$$

$$v''^{(3)} = \begin{pmatrix} 1.0000000000 \\ -0.2264802449 \\ 3.1220328934 \\ 2.9659918380 \end{pmatrix}, \quad v''^{(4)} = \begin{pmatrix} 1.0000000000 \\ 0.4064028329 \\ -0.2200029869 \\ -0.0745454645 \end{pmatrix}.$$

Per trobar els vectors propis $v = v^{(j)}$ ($j = 1 \div 4$) d' A , cal aplicar les transformacions de semblança trobades a aquests vectors propis:

$$v'^{(j)} = P_3 v''^{(j)}, \quad v^{(j)} = P_2 v'^{(j)} \quad (j = 1 \div 4).$$

Notem que la transformació amb P_3 afecta només les dues darreres components i la transformació amb P_2 , les tres darreres. L'aplicació de les matrius de Householder $P_i = I - 2\alpha_i u^{(i)} u^{(i)\top}$ ($i = 2, 3$) a un vector v'' es troba simplement fent

$$\begin{aligned} v' &= P_3 v'' = v'' - \alpha_3 (u^{(3)\top} v'') u^{(3)}, \\ v &= P_2 v' = v' - \alpha_2 (u^{(2)\top} v') u^{(2)}. \end{aligned}$$

Els vectors propis de $T = A_4$ són així transformats per P_3 en els vectors propis d' A_3 següents:

$$\begin{aligned} v'^{(1)} &= \begin{pmatrix} 1.0000000000 \\ -2.7735250825 \\ 0.0683784201 \\ 0.6125438612 \end{pmatrix}, \quad v'^{(2)} = \begin{pmatrix} 1.0000000000 \\ -0.8704991207 \\ 4.0115433053 \\ -6.0218612318 \end{pmatrix}, \\ v'^{(3)} &= \begin{pmatrix} 1.0000000000 \\ -0.2264802449 \\ -3.6729696891 \\ -2.2479970270 \end{pmatrix}, \quad v'^{(4)} = \begin{pmatrix} 1.0000000000 \\ 0.4064028329 \\ 0.1289463524 \\ 0.1932127810 \end{pmatrix}. \end{aligned}$$

Finalment, aquests són transformats per P_2 en els vectors propis d' $A_2 = A$ següents:

$$\begin{aligned} v^{(1)} &= \begin{pmatrix} 1.0000000000 \\ 1.2081647906 \\ 1.5257780636 \\ 2.0699435048 \end{pmatrix}, \quad v^{(2)} = \begin{pmatrix} 1.0000000000 \\ 1.6632404977 \\ 4.9389563722 \\ -5.0944481649 \end{pmatrix}, \\ v^{(3)} &= \begin{pmatrix} 1.0000000000 \\ 3.5492301578 \\ -2.2909637644 \\ -0.8659911023 \end{pmatrix}, \quad v^{(4)} = \begin{pmatrix} 1.0000000000 \\ -0.4206354474 \\ -0.1737706681 \\ -0.1095042395 \end{pmatrix}. \end{aligned}$$

Problema 2.16 *Escriuiu una matriu tridiagonal simètrica (no diagonal) $n \times n$ que tingui per valors propis els zeros del polinomi de Tchebixev T_n a $[-1, 1]$.*

SOLUCIÓ:

El polinomi característic d'una matriu tridiagonal

$$\begin{pmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & b_3 & a_3 & b_4 & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots & \ddots \\ & & & & b_{n-1} & a_{n-1} & b_n \\ & & & & & b_n & a_n \end{pmatrix}$$

es pot calcular recurrentment. Prenent $p_j(x) = \det(xI - A_j)$, on A_j és la matriu formada per les j primeres fileres de les j primeres columnes d' A , es té la recurrència

$$p_0(x) = 1, \quad p_1(x) = x - a_1;$$

$$p_j(x) = (x - a_j)p_{j-1}(x) - b_j^2 p_{j-2}(x) \quad (j \geq 1).$$

El polinomi característic d' A s'escriu llavors com

$$p_A(\lambda) = \det(A - \lambda I) = (-1)^n \det(\lambda I - A) = (-1)^n p_n(\lambda).$$

No farem cap distinció entre els polinomis $p_A(\lambda)$ i $p_n(\lambda)$ a l'hora de cercar valors propis, i entendrem que $p_n(\lambda)$ és el polinomi característic d' A .

Fixem-nos primer que els polinomis obtinguts per la recurrència anterior són mònics, ja que ho són p_0 i p_1 . Si volem que p_n tingui els mateixos zeros que T_n , p_n ha de ser un múltiple de T_n ; en realitat, ha de ser el polinomi mònic de Txebixev (vegeu el capítol III).

Els polinomis de Txebixev compleixen la recurrència

$$T_0(x) = 1, \quad T_1(x) = x; \quad T_j(x) = 2xT_{j-1}(x) - T_{j-2}(x) \quad (j \geq 2).$$

Com que el coeficient principal del polinomi de grau j és 2^{j-1} , per a $j \geq 1$, els polinomis mònics de Txebixev seran

$$\tilde{T}_0(x) = T_0(x), \quad \tilde{T}_j(x) = \frac{T_j(x)}{2^{j-1}} \quad (j \geq 1);$$

aïllant-ne T_j i substituint-lo en la relació de recurrència anterior, obtenim la relació de recurrència següent d'aquests polinomis mònics:

$$\tilde{T}_0(x) = 1, \quad \tilde{T}_1(x) = x, \quad \tilde{T}_2(x) = x\tilde{T}_1(x) - \frac{1}{2}\tilde{T}_0(x);$$

$$\tilde{T}_j(x) = x\tilde{T}_{j-1}(x) - \frac{1}{4}\tilde{T}_{j-2}(x) \quad (j \geq 2).$$

Imposant la igualtat entre els p_j ($j = 0 \div n$) i els $\tilde{T}_j$ ($j = 0 \div n$), sobre les seves relacions de recurrència, tenim que els coeficients de la matriu tridiagonal han de complir

$$a_j = 0 \quad (j = 1 \div n); \quad b_2^2 = \frac{1}{2}, \quad b_j^2 = \frac{1}{4}.$$

Triant, per exemple, els coeficients positius, trobem la matriu cercada

$$\begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & & & & & \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{2} & & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \ddots & \ddots & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & & & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Problema 2.17 a) Resoleu el problema de valors propis en equacions diferencials lineals següent: trobeu els valors de λ per als quals existeixen funcions $y(x)$ no idènticament nul·les tals que compleixen l'equació diferencial

$$y''(x) + \lambda y(x) = 0 \quad \forall x \in (0, 1),$$

i les condicions de frontera $y(0) = y(1) = 0$.

Discretitzem el problema imposant la validesa de l'equació només en els $m+1$ abscisses equidistants $x_k = kh$ ($k = 0 \div m$), $h = \frac{1}{m}$, i aproximant després la derivada segona emprant diferències centrades; en resulta un problema de valors propis en dimensió finita en les aproximacions y_k de y en x_k :

$$y_0 = 0, \quad \frac{1}{h^2}(y_{k+1} - 2y_k + y_{k-1}) + \lambda y_k = 0 \quad (k = 1 \div m-1), \quad y_m = 0.$$

b) Resoleu exactament el problema de valors propis resultant.

c) Compareu les solucions del problema discretitzat amb les solucions exactes.

SOLUCIÓ:

a) Les solucions analítiques (exactes) de l'equació diferencial són de la forma

$$y(x) = A \cos(\sqrt{\lambda}x) + B \sin(\sqrt{\lambda}x).$$

Imposant la condició de frontera $y(0) = 0$, resulta $A = 0$, i imposant l'altra condició $y(1) = 0$, resulta $B = 0$ a menys que $\lambda = \lambda_j = j^2\pi^2$ ($j \geq 0$).

Aquests valors de λ són els valors propis del problema; les funcions pròpies del problema associades a un valor propi $\lambda_j = j^2\pi^2$ són múltiples de la funció $\sin j\pi x$.

b) El *problema discretitzat* consisteix a trobar els valors i vectors propis de la matriu tridiagonal simètrica

$$\frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots & \ddots \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 2 \end{pmatrix}.$$

Es comprova fàcilment que els valors propis de la matriu són

$$\tilde{\lambda}_j = \frac{2}{h^2} (1 - \cos \frac{j\pi}{m}) \quad (j = 1 \div m-1),$$

i que les components d'un dels vectors propis associats $y^{(j)}$ són

$$y_k^{(j)} = \sin \frac{jk\pi}{m} = \sin jx_k \quad (k = 1 \div m-1);$$

els altres vectors propis associats a $\tilde{\lambda}_j$ són múltiples d'aquest vector.

c) Notem que els vectors propis del problema discretitzat són les taules de les funcions pròpies del problema de partida en les abscisses en què l'hem discretitzat.

En el problema discretitzat es tenen $m-1$ valors propis que tendeixen asimptòticament, quan h tendeix a zero o m tendeix a infinit, als infinits valors propis del problema de partida:

$$\begin{aligned} \tilde{\lambda}_j &= \frac{2}{h^2} (1 - \cos j\pi h) = \frac{2}{h^2} \left(\frac{j^2 \pi^2 h^2}{2} - \frac{j^4 \pi^4 h^4}{24} + \dots \right) \\ &= \lambda_j - \frac{j^2 \pi^4}{12} h^2 + \dots \quad (h = \frac{1}{m} \rightarrow 0) \end{aligned}$$

(consulteu la taula 3.5 per al desenvolupament de Taylor de la funció $\cos x$ prop de zero).

Tenim, a més, la fita d'error

$$\frac{|\tilde{\lambda}_j - \lambda_j|}{|\lambda_j|} \leq \frac{\pi^4}{12} \frac{j^2}{m^2} \quad (j = 1 \div m-1),$$

que ens mostra la poca precisió dels valors aproximats $\tilde{\lambda}_j$ enfront dels exactes λ_j , per als últims valors de l'índex j .

PROBLEMES PROPOSATS

1. Considerem el sistema triangular superior $U_n(a)x = e^{(n)}$, on $e^{(n)}$ és el vector n -è de la base canònica i

$$U_n(a) = (u_{ij}) = \begin{pmatrix} 1 & a & a & \cdots & a \\ & 1 & a & \cdots & a \\ & & 1 & \cdots & a \\ & & & \ddots & a \\ & & & & 1 \end{pmatrix}.$$

- a) Determineu la solució x .
 b) Fiteu les components x_i de la solució del sistema, suposant que $|a| \leq 1$.
2. Resoleu el sistema lineal següent per eliminació gaussiana sense pivotatge:

$$\left. \begin{array}{rrrrr} 20x_1 & + & 1.0x_2 & - & 0.1x_3 & + & 1.0x_4 & = & 2.7 \\ 0.4x_1 & + & 0.5x_2 & + & 4.0x_3 & - & 8.5x_4 & = & 21.9 \\ 0.3x_1 & - & 1.0x_2 & + & 1.0x_3 & + & 5.2x_4 & = & -3.9 \\ 1.0x_1 & + & 0.2x_2 & + & 2.5x_3 & - & 1.0x_4 & = & 9.9 \end{array} \right\}.$$

3. Resoleu el sistema

$$(A|b) = \left(\begin{array}{ccc|c} 1 & 1 & 0 & 4 \\ 2 & 0 & 2 & 4 \\ 0 & 3 & 3 & 4 \end{array} \right),$$

pel mètode de Gauss

- a) sense pivotatge,
 b) amb pivotatge maximal per columnes, i
 c) amb pivotatge complet.

4. Resoleu el sistema

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

pel mètode de Gauss

- a) sense pivotatge,
 b) amb pivotatge maximal per columnes, i
 c) amb pivotatge complet.

5. Resoleu els sistemes

$$\left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 1 \\ 1 & 4 & 9 & 16 & 1 \\ 1 & 8 & 27 & 64 & 1 \\ 1 & 16 & 81 & 216 & 1 \end{array} \right), \quad \left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 1 \\ 1 & 4 & 9 & 16 & 0 \\ 1 & 8 & 27 & 64 & 0 \\ 1 & 16 & 81 & 216 & 1 \end{array} \right);$$

usant

- a) el mètode LU,
- b) el mètode de Doolittle, i
- c) el mètode de Crout.

6. Trobeu la factorització LU de la matriu

$$\left(\begin{array}{ccccc} 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & 5 \end{array} \right).$$

7. Considereu les matrius A i B

$$\left(\begin{array}{cccc} 1 & 2 & 1 & -1 \\ 2 & -3 & -2 & 0 \\ 0 & 5 & -3 & 2 \\ -3 & 1 & 4 & 0 \end{array} \right), \quad \left(\begin{array}{ccccc} 2 & -5 & 6 & 4 & -1 \\ 3 & 0 & -2 & -1 & 3 \\ 0 & 4 & -5 & 0 & 1 \\ 1 & -1 & 1 & -3 & 0 \\ 5 & 1 & 6 & 1 & -1 \end{array} \right),$$

i els vectors

$$b^{(1)} = \begin{pmatrix} 7 \\ 7 \\ -17 \\ -5 \end{pmatrix}, \quad b^{(2)} = \begin{pmatrix} 4 \\ -6 \\ 3 \\ 9 \end{pmatrix}, \quad b^{(3)} = b^{(1)} - b^{(2)}.$$

Resoleu els sistemes

$$Ax^{(1)} = b^{(1)}, \quad Ax^{(2)} = b^{(2)}, \quad Ax^{(3)} = b^{(3)}, \quad Bx^{(4)} = b^{(4)}.$$

8. a) Resoleu el sistema

$$(A|b) = \left(\begin{array}{cccc|c} 10 & 7 & 8 & 7 & 4 \\ 7 & 5 & 6 & 5 & 3 \\ 8 & 6 & 10 & 9 & 3 \\ 7 & 5 & 9 & 10 & 1 \end{array} \right),$$

trobant primerament la factorització $A = LDL^T$ amb L triangular inferior amb uns a la diagonal i D diagonal.

b) Trobeu la factorització de Txoleski d' A .

9. Resoleu el sistema següent pel mètode de Txoleski:

$$\left. \begin{array}{cccccc} 1.00x_1 & + & 0.42x_2 & + & 0.54x_3 & + & 0.66x_4 & = & 0.3 \\ 0.42x_1 & + & 1.00x_2 & + & 0.32x_3 & + & 0.44x_4 & = & 0.5 \\ 0.54x_1 & + & 0.32x_2 & + & 1.00x_3 & + & 0.22x_4 & = & 0.7 \\ 0.66x_1 & + & 0.44x_2 & + & 0.22x_3 & + & 1.00x_4 & = & 9.9 \end{array} \right\}.$$

10. a) Expliciteu el mètode LU per resoldre sistemes tridiagonals, pentadiagonals i heptadiagonals.

b) Compteu el nombre d'operacions requerides en cada cas.

c) Repetiu els apartats anteriors, suposant que la matriu del sistema és simètrica.

1

11. a) Demostreu que el mètode LU conserva el caràcter banda de les matrius; això és, si A és una matriu banda (p, q) , les matrius L i U corresponents són banda $(p, 1)$ i banda $(1, q)$, respectivament.

b) Compteu el nombre d'operacions en resoldre un sistema banda (p, q) .

c) Expliciteu la resolució d'un sistema banda $(4, 3)$ de dimensió $n \geq 4$ pel mètode LU.

12. a) Expliciteu un mètode per resoldre directament el sistema amb *matriu tridiagonal per blocs*

$$\left(\begin{array}{cccc|c} A_1 & C_1 & & & b^{(1)} \\ B_2 & A_2 & C_2 & & b^{(2)} \\ & B_3 & A_3 & C_3 & b^{(3)} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & B_{n-1} & A_{n-1} & C_{n-1} & b^{(n-1)} \\ & & & & B_n & A_n & & b^{(n)} \end{array} \right),$$

on els blocs A_i ($i = 1 \div n$) són tridiagonals i els B_i ($i = 2 \div n$) i C_i ($i = 1 \div n - 1$) diagonals, tots $n \times n$.

b) Aplicació: Resoleu el sistema anterior en el cas $n = 4$ per a

$$b^{(i)} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad A_i = \begin{pmatrix} -4 & 1 & 0 & 0 \\ 1 & -4 & 1 & 0 \\ 0 & 1 & -4 & 1 \\ 0 & 0 & 1 & -4 \end{pmatrix} \quad (i = 1 \div 4),$$

$$B_{i+1} = C_i = I \quad (i = 1 \div 3) .$$

13. a) Doneu un mètode directe de resolució de sistemes de la forma següent, dits *sistemes quasi-tridiagonals*:

$$\left(\begin{array}{cccccccc|c} a_1 & c_1 & & & & & & e & z_1 \\ b_2 & a_2 & c_2 & & & & & & z_2 \\ & b_3 & a_3 & c_3 & & & & & z_3 \\ & & \ddots & \ddots & \ddots & & & & \vdots \\ & & & \ddots & \ddots & \ddots & & & \vdots \\ & & & & \ddots & \ddots & \ddots & & \vdots \\ & & & & & b_{n-1} & a_{n-1} & c_{n-1} & z_{n-1} \\ d & & & & & & b_n & a_n & z_n \end{array} \right) .$$

- b) Si la matriu fos a més simètrica, com s'hauria de variar el mètode a fi de minimitzar la memòria emprada i el nombre d'operacions?

- c) Aplicació: Resoleu el sistema

$$\left(\begin{array}{ccccc|c} 4 & -1 & 0 & 0 & 1 & 1 \\ -1 & 4 & -1 & 0 & 0 & 2 \\ 0 & -1 & 4 & -1 & 0 & 3 \\ 0 & 0 & -1 & 4 & -1 & 4 \\ 1 & 0 & 0 & -1 & 4 & 5 \end{array} \right) .$$

14. Sigui A una matriu regular donada. Per tal de resoldre el sistema lineal $A^2x = b$, és millor calcular A^2 i resoldre el sistema, o bé, resoldre els dos sistemes $Ay = b$, $Ax = y$, sempre usant la factorització LU?

15. Sigui A una matriu definida positiva. Fem una factorització de la forma $A = LDL^\top$, amb L triangular inferior amb uns a la diagonal i D diagonal amb elements diagonals positius.

- a) Quin és el nombre total d'operacions aritmètiques (+, −, *, /) que cal fer?

Un *ordinador vectorial* és un ordinador que permet fer moltes operacions amb una sola instrucció (anomenada *instrucció vectorial*); així, si x i y són dos vectors i $\#$ és una operació aritmètica qualsevol, $x\#y$ calcula el vector de components $x_i\#y_i$. Òbviament, en el càlcul de la factorització esmentada hi ha moltes operacions que es poden fer vectorialment.

- b) Calculeu el nombre total d'operacions escalars i operacions vectorials que es requereixen per portar a terme la dita factorització en un ordinador vectorial.

16. Es vol factoritzar la matriu A com a producte d'una matriu ortogonal Q i una matriu triangular superior R usant
- i) el mètode d'ortogonalització de Gram-Schmidt,
 - ii) el mètode d'ortogonalització de Householder.
- a) Realitzeu les factoritzacions QR de la matriu

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \\ 1 & 16 & 81 & 216 \end{pmatrix},$$

pels mètodes esmentats.

- b) Resoleu el sistema $Ax = b$ amb $b^T = (1 \ 1 \ 1 \ 1)$.

17. Trobeu la inversa de la matriu

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \\ 1 & 16 & 81 & 216 \end{pmatrix},$$

- a) fent la factorització LU, invertint L i U , i després fent el producte $U^{-1}L^{-1}$;
 - b) fent una factorització QR, invertint la matriu R i després fent el producte $R^{-1}Q^T$;
 - c) compareu els mètodes a) i b) quant al nombre d'operacions.
18. El *mètode de Gauss-Jordan* d'inversió de matrius parteix dels sistemes d'equacions $Ax^{(j)} = e^{(j)}$ ($j = 1 \div n$) escrits en forma matricial

$$(A|I) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & 1 & 0 & 0 \\ \vdots & & \vdots & 0 & \ddots & 0 \\ a_{n1} & \cdots & a_{nn} & 0 & 0 & 1 \end{array} \right),$$

i va modificant les fileres de $(A|I)$, usant els elements de la diagonal com a pivots (com en el mètode de Gauss), aconseguint que s'anul·lin els elements no diagonals d' A (també els de sobre la diagonal!), columna a columna. S'obté finalment $(D|B)$ amb D diagonal; llavors, tenim $A^{-1} = D^{-1}B$.

Sigui A una matriu regular $n \times n$.

- a) Expliciteu el mètode de Gauss-Jordan per al càlcul de la seva inversa.
- b) Compteu el nombre d'operacions a realitzar, en general.
- c) Aplicació: Invertiu la matriu

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \\ 1 & 16 & 81 & 216 \end{pmatrix}$$

pel mètode esmentat.

19. Sigui A una matriu regular $n \times n$ que considerem partida en quatre blocs

$$A = \begin{pmatrix} P & Q \\ R & S \end{pmatrix},$$

amb P i S matrius $p \times p$ i $s \times s$, respectivament ($p + s = n$).

Suposant que P i $S - RP^{-1}Q$ tenen inversa:

- Doneu un mètode per trobar la inversa d' A .
- En general, quantes operacions calen per trobar A^{-1} usant aquest mètode?
- Quan serà útil el mètode esmentat?
- Expliciteu un procediment per invertir una matriu A , emprant $n - 1$ vegades el procediment anterior.
- Sota quines condicions funcionarà el procediment de l'apartat d)?
- Demostreu que és segur que funciona si A és definida positiva.
- Aplicació: Trobeu

$$\left(\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 2 & -1 & 0 \end{array} \right)^{-1}, \text{ usant l'apartat a),}$$

$$i \left(\begin{array}{ccc} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{array} \right)^{-1}, \text{ usant l'apartat d).}$$

20. Calculeu el determinant i la inversa de les matrius següents:

$$\begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & 3 \\ 4 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 4 & 2 & 1 & 1 \\ 1 & -3 & -2 & 0 \\ 7 & 10 & 1 & 3 \\ 2 & 4 & 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 3 & 6 & 9 & 5 \\ -2 & -7 & -5 & -17 & -12 \\ -1 & -4 & -5 & 1 & -4 \\ 3 & 11 & 2 & 32 & 23 \\ 1 & 4 & 5 & 3 & 4 \end{pmatrix}.$$

21. Volem calcular la inversa d'una *matriu bidiagonal (inferior)*

$$A = \begin{pmatrix} a_1 & & & & & \\ b_2 & a_2 & & & & \\ & b_3 & a_3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & b_{n-1} & a_{n-1} \\ & & & & & b_n & a_n \end{pmatrix}.$$

- Doneu un mètode per al càlcul de cadascuna de les seves columnes.
- Calculeu el nombre d'operacions necessàries per tal d'invertir-la, en el cas més general.
- Trobeu la inversa quan $a_i = a$ ($i = 1 \div n$), $b_i = b$ ($i = 2 \div n$).
- Si les multiplicacions i les divisions es fan en precisió doble, fiteu l'error relatiu en l'element $(A^{-1})_{n1}$, degut només a les dites operacions.

22. a) Calculeu el determinant

$$\begin{vmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 3 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{vmatrix},$$

- usant el mètode de Gauss,
 - usant el mètode d'ortogonalització de Householder.
- b) Compareu ambdós mètodes quant al nombre d'operacions, en el cas general.
23. a) Trobeu una recurrència per al càlcul de determinants de matrius tridiagonals.
- b) Aplicació: Trobeu la dimensió màxima de les matrius

$$A_n = \begin{pmatrix} 2 & 1.02 & & & & \\ 0.99 & 2 & 1.02 & & & \\ & 0.99 & 2 & 1.02 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \ddots \\ & & & & 0.99 & 2 & 1.02 \\ & & & & & 0.99 & 2 \end{pmatrix},$$

per tal que siguin definides positives.

- Trobeu també una recurrència per al càlcul de determinants de matrius pentadiagonals.

24. Siguin $e^{(i)}$ ($i = 1 \div n$) els vectors (columna) de la base canònica de $\mathbb{R}^n$; donats els nombres reals a_i ($i = 1 \div n$), considerem la matriu

$$S = \sum_{i=1}^{n-1} a_{i+1} e^{(i+1)} (e^{(i)})^\top .$$

a) Calculeu S^k per a tot k enter positiu.

(Indicació: Trobeu S^2 i generalitzeu-ho, no fa falta usar inducció).

b) Trobeu $(I - S)^{-1}$.

c) Trobeu $(I - S)(I - S)^\top$.

d) Aplicació: Quina és la factorització de Txoleski de la matriu B que té com a únics elements no nuls els següents:

$$\begin{aligned} b_{11} &= 1, & b_{ii} &= i^2 + 1 \quad (i = 2 \div n), \\ b_{i,i-1} &= i \quad (i = 2 \div n), & b_{i,i+1} &= i \quad (i = 1 \div n-1). \end{aligned}$$

25. a) Donada una matriu $n \times n$ regular A i 2 vectors u i v de dimensió n , demostreu que $A - uv^\top$ és regular si i només si

$$\alpha = 1/(1 - v^\top A^{-1}u) \neq -1,$$

i llavors

$$(A - uv^\top)^{-1} = A^{-1} + \alpha A^{-1}uv^\top A^{-1}$$

(fórmula de Sherman-Morrison).

b) Sigui B una matriu que difereix d' A només en una filera i en una columna; suposant coneguda A^{-1} , doneu un procediment de càlcul de B^{-1} basat en l'apartat a).

c) Aplicació: Tenim que

$$\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}^{-1} = \frac{1}{4} \begin{pmatrix} 3 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 3 \end{pmatrix};$$

calculeu, emprant l'apartat b),

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}^{-1}.$$

d) Generalitzeu la fórmula de Sherman-Morrison a una altra que permeti fer el càlcul de $(A - UV^\top)^{-1}$ a partir d' A^{-1} , essent ara U i V matrius $n \times r$.

26. Considerem els vectors u i v de n components reals i generem la matriu $A = I + uv^\top$.

a) Proveu que existeix un polinomi de segon grau $P(x) = x^2 + ax + b$ amb a, b reals tal que $P(A) = A^2 + aA + bI = 0$.

b) Demostreu que A és invertible si i només si $v^\top u \neq -1$ i trobeu, en aquest cas, la inversa d' A . Estudieu el cas particular en què u i v són ortogonals.

Considerem el sistema lineal $Cx = c$, essent $C = (c_{ij})$ la matriu de terme general

$$c_{ij} = \begin{cases} \alpha & (\text{si } i = j) \\ \beta & (\text{si } i \neq j) \end{cases},$$

on α i β són no nuls i diferents.

c) Demostreu que el sistema anterior és compatible i determinat si i només si $\alpha + (n-1)\beta \neq 0$.

(Indicació: Trobeu λ per tal que $C - \lambda I$ tingui rang 1).

e) Expliqueu com es poden aplicar els resultats anteriors a la resolució de l'esmentat sistema.

27. a) Demostreu que la norma de matrius (no multiplicativa)

$$\|A\|_E = \left(\sum_{i,j} |a_{ij}|^2 \right)^{\frac{1}{2}},$$

no està subordinada a cap norma vectorial.

(Indicació: Considereu la norma de la matriu identitat).

b) Demostreu que la norma subordinada a la norma euclidiana

$$\|x\|_2 = \left(\sum_i |x_i|^2 \right)^{\frac{1}{2}}$$

és

$$(\rho(A^*A))^{\frac{1}{2}}.$$

(Indicació: Useu el fet que, si N és hermitiana, existeix una matriu U unitària tal que U^*NU és diagonal).

c) Demostreu les relacions següents:

i) $\|A^*A\|_2 = \|A\|_2^2 = \|A^*\|_2^2$;

ii) $\|U^*AU\|_2 = \|AU\|_2 = \|UA\|_2$, si U és unitària;

iii) $\|A\|_2 \leq \|A\|_E \leq \sqrt{n}\|A\|_2$ i $\frac{1}{\sqrt{n}}\|A\|_E \leq \|A\|_2 \leq \|A\|_E$;

iv)

$$\|A\|_2 = \max_{x,y \neq 0} \frac{|y^*Ax|}{\|x\|_2\|y\|_2}.$$

28. a) Considereu una matriu tal que $\max_{i,j} |a_{ij}| \leq a$, fiteu en funció d' a : $\|A\|_\infty$, $\|A\|_1$, $\|A\|_2$ i $\|A\|_E$.

b) Comproveu el resultat obtingut a l'apartat a) amb la matriu

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \\ 0 & 1 & 1 \end{pmatrix}.$$

29. Considerem una norma matricial $\|\cdot\|$ subordinada a una norma vectorial qualsevol.

a) Demostreu que, si $\|R\| \leq 1$, $I + R$ és regular, i llavors

$$\frac{1}{1 + \|R\|} \leq \|(I + R)^{-1}\| \leq \frac{1}{1 - \|R\|}.$$

b) Proveu que, si A i $A + E$ són regulars, llavors

$$\|(A + E)^{-1} - A^{-1}\| \leq \|A^{-1}\| \frac{\alpha}{1 - \alpha},$$

amb $\alpha = \|A^{-1}E\|$.

30. Sigui $\|\cdot\|_2$ la norma matricial subordinada a la norma vectorial euclidiana.

a) Demostreu que, per a qualsevol matriu $A = (a_{ij})$, es verifica

$$\|A\|_2 \geq |a_{ii}| \quad (i = 1 \div n).$$

b) Demostreu que, si A és triangular amb elements diagonals diferents de zero, llavors

$$\|A^{-1}\|_2 \geq \frac{1}{|a_{ii}|} \quad (i = 1 \div n).$$

Trobeu, en aquest cas, una fita inferior per al nombre de condició d' A en aquesta norma.

31. a) Demostreu que les sèries geomètriques matricials

$$A + AB + \cdots + AB^j + \cdots, \quad A + BA + \cdots + B^j A + \cdots$$

convergeixen si $\|B\| < 1$ per a alguna norma matricial subordinada a una norma vectorial qualsevol i, en tal cas,

$$\sum_{j=0}^{\infty} AB^j = A(I - B)^{-1}, \quad \sum_{j=0}^{\infty} B^j A = (I - B)^{-1}A.$$

b) Fiteu la norma del residu n -è en funció de $\|B\|$ i $\|A\|$.

c) Aplicació: Calculeu

$$\begin{pmatrix} 0.99 & 1.01 \\ 0.98 & 1.02 \end{pmatrix}^{-1}$$

amb un error menor que 10^{-6} en la norma $\| \cdot \|_{\infty}$.

32. Per a una matriu $n \times n$ A , definim

$$\begin{aligned} e^A &= \sum_{j=0}^{\infty} \frac{A^j}{j!}, \\ \sin A &= \sum_{r=0}^{\infty} (-1)^r \frac{A^{2r+1}}{(2r+1)!}, \\ \cos A &= \sum_{r=0}^{\infty} (-1)^r \frac{A^{2r}}{(2r)!}. \end{aligned}$$

a) Trobeu expressions per a les fites de les restes d'aquestes sèries matricials en una norma donada.

b) Aplicació: Donada

$$A = \begin{pmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{pmatrix},$$

calculeu e^A , $\sin A$ i $\cos A$ amb un error menor que 10^{-6} en la norma $\| \cdot \|_{\infty}$.

33. Sigui R una matriu regular. Donada una norma vectorial qualsevol $\| \cdot \|$ definim $\|x\|_R \equiv \|Rx\|$.

a) Demostreu que $\| \cdot \|_R$ és una norma vectorial i que

$$\frac{\|x\|}{\|R^{-1}\|} \leq \|x\|_R \leq \|R\| \|x\|,$$

on la norma matricial $\| \cdot \|$ indica la norma subordinada a la norma vectorial donada.

b) Expressen la norma matricial $\| \cdot \|_R$ subordinada a la norma vectorial definida (i denotada de la mateixa manera) en termes de la norma matricial subordinada a $\| \cdot \|$.

c) Vegeu que els nombres de condició

$$\mu_R(A) \equiv \|A\|_R \|A^{-1}\|_R$$

poden ser arbitràriament grans per a una matriu A fixada, triant R adequadament.

34. Considerem el sistema

$$(A|b) = \left(\begin{array}{cccc|c} 10 & 7 & 8 & 7 & 32 \\ 7 & 5 & 6 & 5 & 23 \\ 8 & 6 & 10 & 9 & 33 \\ 7 & 5 & 9 & 10 & 31 \end{array} \right).$$

- a) Resoleu-lo exactament.
 b) Comproveu que els vectors

$$x = x^{(1)} = \begin{pmatrix} 6 \\ -7.2 \\ -2.9 \\ -0.1 \end{pmatrix} \text{ i } x = x^{(2)} = \begin{pmatrix} 1.50 \\ 0.18 \\ 1.19 \\ 0.89 \end{pmatrix}$$

tenen *residus vectorials* $r_x = b - Ax$ molt petits en $\| \cdot \|_\infty$.

- c) Trobeu els nombres de condició $\mu_\infty(A)$ i $\mu_1(A)$.
 d) Expliqueu el fenomen descobert als apartats a) i b).

35. a) Demostreu que la raó

$$g_n = \frac{\max_{i,j,k} |a_{ij}^{(k)}|}{\max_{i,j} |a_{ij}|},$$

que apareix en l'estudi dels errors en els càlculs en resoldre sistemes lineals pel mètode de Gauss amb pivotatge maximal per columnes, és menor que 2^{n-1} .

- b) Proveu que g_n assoleix aquest valor màxim 2^{n-1} per a les matrius

$$A_n = \begin{pmatrix} 1 & 0 & 0 & \ddots & \ddots & 0 & 1 \\ -1 & 1 & 0 & \ddots & \ddots & 0 & 1 \\ -1 & -1 & 1 & \ddots & \ddots & 0 & 1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \ddots & \ddots & 0 & 1 \\ -1 & -1 & -1 & \ddots & \ddots & 1 & 1 \\ -1 & -1 & -1 & \ddots & \ddots & -1 & 1 \end{pmatrix}.$$

- c) Fiteu l'error en fer la factorització LU d'aquestes matrius (en funció de n), a causa dels errors en els càlculs, si empren un ordinador treballant amb doble precisió.

36. Fiteu el valor de g_n , quan usem pivotatge maximal per columnes per resoldre el sistema $Ax = b$, en els casos particulars següents:

- i) A és una matriu Hessenberg superior;
 ii) A és una matriu tridiagonal.

37. Hem de resoldre el sistema $Cx = b$, on $C = A^m$, A és una matriu $n \times n$ i $\|A\|_\infty = a$; sabem també que $\|A^{-1}\|_\infty = \|A\|_\infty$. Disposem de la fórmula següent per a l'error relatiu en x , usant eliminació gaussiana amb pivotatge complet

$$\frac{\|\delta x\|_\infty}{\|x\|_\infty} \leq \epsilon n^{2+0.25 \ln n} \|C\|_\infty \|C^{-1}\|_\infty,$$

on ϵ és una fita de l'error relatiu d'arrodoniment en les operacions elementals.

a) Trobeu una fita d'aquest error relatiu en funció de ϵ , n , a i m .

Un sistema, equivalent en cert sentit, a l'anterior és

$$Ay^{(1)} = b, \quad Ay^{(j+1)} = y^{(j)} \quad (j = 1 \div m-1),$$

amb $y^{(j)} \in \mathbb{R}^n$ ($j = 1 \div m$).

b) Comproveu que $x = y^{(m)}$ i escriviu el dit sistema en forma matricial

$$By = d, \quad y, d \in \mathbb{R}^{nm}. \quad (*)$$

c) Calculeu la inversa de B .

d) Emprant la fita donada, trobeu una fita de l'error relatiu per al problema escrit en la forma $(*)$ en funció de ϵ , n , a i m .

e) Aplicació: Calculeu les fites trobades quan $\epsilon = 10^{-13}$, $n = 4$, $a = 2000$ i $m = 5$. És permisible la fita trobada en el primer cas? I la trobada en el segon?

38. Donats els sistemes

$$\left(\begin{array}{ccc|c} 2 & 0 & 3 & 1 \\ 0 & 4 & 2 & 2 \\ 0 & 1 & 6 & 3 \end{array} \right), \quad \left(\begin{array}{ccc|c} 1 & 2 & -2 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 \end{array} \right),$$

$$\left(\begin{array}{ccc|c} 2 & -1 & 1 & 1 \\ 2 & 2 & 2 & 1 \\ -1 & -1 & 2 & 1 \end{array} \right).$$

a) Discutiu la convergència dels mètodes iteratius de Jacobi i de Gauss-Seidel.

b) Calculeu les solucions aproximades al cap de 10 iteracions, partint del vector zero.

39. Sigui la matriu $n \times n$

$$A = \begin{pmatrix} 1 & 2 & 3 & \cdots & n-1 & n \\ n & 1 & 2 & \cdots & n-2 & n-1 \\ n-1 & n & 1 & \cdots & n-3 & n-2 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 3 & 4 & 5 & \cdots & 1 & 2 \\ 2 & 3 & 4 & \cdots & n & 1 \end{pmatrix};$$

això és,

$$a_{ij} = \begin{cases} n+j-i+1 & (j < i) \\ 1+j-i & (j \geq i) \end{cases}.$$

a) Demostreu que els vectors $v^{(j)}$, de components

$$v_k^{(j)} = q_j^k \text{ amb } q_j = e^{\frac{i2\pi(j-1)}{n}} \quad (k, j = 1 \div n),$$

són vectors propis d' A de valors propis respectius

$$\lambda_j = \sum_{k=1}^n k q_j^{k-1} \quad (j = 1 \div n) .$$

Volem resoldre un sistema $(A + aI)x = b$ pel mètode iteratiu de Jacobi.

b) Discutiu la convergència del mètode esmentat en funció d' a .

c) Per a $n = 4$, $a = 19$, $b = (1 \ 0 \ 0 \ 0)^\top$, trobeu $x^{(k)}$ tal que

$$\|b - (A + aI)x^{(k)}\|_\infty \leq 0.03 .$$

c) Comproveu la divergència del mètode quan $a = 0$.

d) Per a quin valor d' a és més ràpida la convergència?

40. Considerem la matriu per blocs

$$A = \begin{pmatrix} I & S \\ S^* & I \end{pmatrix} ,$$

on S és una matriu complexa $n \times n$.

a) Si volem resoldre un sistema $Ax = b$, doneu una condició necessària i suficient de convergència del mètode iteratiu de Jacobi.

b) Feu el mateix per al mètode iteratiu de Gauss-Seidel.

41. Discutiu, en funció d' a , la convergència dels mètodes iteratius de Jacobi i de Gauss-Seidel en sistemes lineals amb la matriu

$$A = \begin{pmatrix} 1 & & & & & a \\ -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \ddots \\ & & & & & -1 & 1 \end{pmatrix} .$$

42. Volem resoldre el sistema lineal

$$\begin{pmatrix} 1 & -a \\ -a & 1 \end{pmatrix} x = b ,$$

pel mètode iteratiu de sobrerelaxació amb un paràmetre ω .

a) Escriviu les iteracions corresponents.

b) Demostreu que el mètode convergeix si i només si les dues arrels de

$$\lambda^2 - [2(1 - \omega) + (\omega a)^2]\lambda + (1 - \omega)^2 = 0$$

són menors que 1, en mòdul.

c) Deduïu de l'apartat b) que el mètode convergeix si i només si $\omega \in (0, 2)$.

d) Demostreu que el valor òptim de ω és

$$\omega^* = \frac{2}{a^2}(1 - \sqrt{1 - a^2}) .$$

e) Aplicació: Si $a = \frac{1}{2}$ i usem el factor òptim de ω , quantes iteracions calen per tal d'assegurar que l'error inicial es redueix almenys en un factor 10^{-6} en la norma euclidiana?

43. Sigui B la matriu $n \times n$ d'elements

$$b_{ij} = \begin{cases} 1 & (\text{si } i - j = 1 \text{ o } j - i = 1) \\ 0 & (\text{altrament}) \end{cases} .$$

a) Demostreu que té els valors propis

$$\lambda_j = 2 \cos \frac{j\pi}{n+1} \quad (j = 1 \div n)$$

amb vectors propis associats $v^{(j)}$ de components

$$v_i^{(j)} = \sin \frac{ij\pi}{n+1} \quad (i, j = 1 \div n) .$$

Considerem els sistemes següents d'equacions lineals:

$$x_0 = p, \quad x_i = a(x_{i-1} - 2x_i + x_{i+1}) \quad (i = 1 \div n), \quad x_{n+1} = q,$$

per als valors positius d' a .

b) Estudieu la convergència dels mètodes iteratius corresponents, en funció d' a :

$$\begin{aligned} x_0^{(k+1)} &= p, \\ x_i^{(k+1)} &= a(x_{i-1}^{(k)} - 2x_i^{(k)} + x_{i+1}^{(k)}) \quad (i = 1 \div n), \\ x_{n+1}^{(k+1)} &= q. \end{aligned}$$

c) Demostreu que el mètode iteratiu de Jacobi aplicat a aquells sistemes convergeix per a tot valor positiu d' a .

44. Considerem el problema de valors a la frontera següent en una equació diferencial de segon ordre: determineu la funció $u(x)$ en l'interval $[0, 1]$ que compleix

$$\frac{d^2u}{dx^2}(x) - cu(x) = 0 \quad (c > 0) \quad \forall x \in (0, 1) ; \quad u(0) = \alpha, \quad u(1) = \beta .$$

- a) Trobeu la solució analítica del problema.

Una aproximació de la solució es pot trobar avaluant l'equació en els punts $x_i = \frac{i}{n}$ ($i = 0 \div n$), discretitzant la derivada segona en la forma usual

$$\frac{d^2u}{dx^2}(x_i) \simeq n^2[u(x_{i-1}) - 2u(x_i) + u(x_{i+1})]$$

i resolent finalment el sistema lineal resultant (discretitzat).

- b) Escriviu aquest problema discretitzat en les aproximacions u_i de $u(x_i)$.

- c) Estudieu la convergència dels mètodes iteratius de Jacobi i de Gauss-Seidel cap a la solució del problema discretitzat.

45. Sigui A una matriu regular $n \times n$ i X_0 una matriu arbitrària $n \times n$. Definim una successió de matrius, mitjançant la recurrència

$$X_{k+1} = X_k + X_k(I - AX_k) \quad (k \geq 0) .$$

- a) Demostreu que la successió esmentada convergeix a A^{-1} si i només si $\rho(I - AX_0) < 1$.

- b) Donats

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} , \quad X_0 = \begin{pmatrix} 1.9 & -0.9 \\ -0.9 & 0.9 \end{pmatrix} ,$$

trobeu una aproximació de la inversa d' A fent servir el mètode iteratiu anterior fins que $\rho(I - AX_k) < 10^{-6}$.

46. Localitzeu una zona del pla complex on es troben els valors propis de la matriu

$$A = \begin{pmatrix} -0.5 & 0.01 & 0 & -0.01 \\ 0.02 & 0 & 0.25 & 0.25 \\ -0.01 & 0.3 & 0 & 0.3 \\ 0 & -0.01 & 0.01 & 0.5 \end{pmatrix} .$$

47. a) Trobeu els valors propis de mòdul màxim i mínim de la matriu

$$A = \begin{pmatrix} 9 & 10 & 8 \\ 10 & 5 & -1 \\ 8 & -1 & 3 \end{pmatrix} .$$

- b) Trobeu ara l'altre valor propi.

48. La matriu següent

$$A = \begin{pmatrix} 14 & 7 & 6 & 9 \\ 7 & 9 & 4 & 6 \\ 6 & 4 & 9 & 7 \\ 9 & 6 & 7 & 15 \end{pmatrix}$$

té un valor propi prop de 4. Trobeu-lo amb 6 xifres decimals correctes.

49. Suposem que l'edat més gran a què poden arribar les femelles d'una certa població animal sigui de 15 anys. Dividim la població total de femelles en 3 classes de 5 anys de durada cadascuna i denotem per x_i el nombre de femelles en la classe i -ena. Siguin $a_i \geq 0$ les mitjanes dels nombres de filles que tenen les femelles que són a la classe i -ena ($i = 1, 2, 3$), i b_i ($0 < b_i \leq 1$) la fracció de femelles que pertanyen a la classe i -ena que s'espera que sobrevisquin i passin a la classe següent d'edat ($i = 1, 2$).

Podem escriure, per tant, la relació entre una distribució de femelles $x^{(k)}$ i la seva distribució al cap de 5 anys $x^{(k+1)}$:

$$x^{(k+1)} = Lx^{(k)}, \quad \text{amb } L = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & 0 & 0 \\ 0 & b_2 & 0 \end{pmatrix};$$

la matriu L s'anomena *matriu de Leslie*.

Suposem que inicialment hi ha només 1000 femelles en la primera de les 3 classes i que $a_1 = 0$, $a_2 = 1.68$, $a_3 = 1.28$, $b_1 = 0.5$ i $b_2 = 0.25$.

a) Trobeu la distribució al cap de $5k$ anys.

b) Determineu la distribució relativa en el límit (quan k tendeix a infinit); això és, els límits de les raons entre els nombres de les femelles de cada classe i el nombre total de femelles.

c) Relacioneu el comportament en el límit amb els valors i vectors propis d' A .

50. Modifiqueu el mètode de la potència a fi que pugui ser emprat en els casos següents:

i) el valor propi de mòdul màxim és de multiplicitat $m > 1$;

ii) els valors propis de mòdul màxim són λ_1 i $\lambda_2 = -\lambda_1$.

51. Suposem que A és una matriu 3×3 amb valors propis λ_1 i λ_2 , tals que $|\lambda_1| > |\lambda_2|$ i λ_1 és de multiplicitat 2. Suposem que A té només dos vectors propis normalitzats linealment independents $v^{(1)}$ i $v^{(2)}$ associats a λ_1 i λ_2 , respectivament. Es pot demostrar que existeix un vector no nul v tal que $(A - \lambda_1 I)v = v^{(1)}$ (aquest v s'anomena *vector propi generalitzat*).

a) Demostreu que v , $v^{(1)}$ i $v^{(2)}$ són linealment independents.

b) Expressau $A^k v$ com a combinació lineal dels vectors linealment independents d'a) ($k \geq 0$).

c) Sigui w un vector qualsevol tal que $w^\top v \neq 0$ i $w^\top v^{(1)} \neq 0$. Trobeu

$$\lim_{k \rightarrow \infty} \frac{w^\top x^{(k+1)}}{w^\top x^{(k)}} ,$$

on $(x^{(k)})_{k \geq 0}$ és la successió obtinguda en aplicar el mètode de la potència a la matriu A , partint d'un vector inicial

$$x^{(0)} = av + a_1 v^{(1)} + a_2 v^{(2)} ,$$

amb $a \neq 0$.

d) Sigui e_k l'error comès en aproximar el límit anterior pel terme k -è de la successió $w^\top x^{(k)}$ ($k \geq 0$). Trobeu

$$\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k} .$$

52. a) Reduïu les matrius següents a forma tridiagonal simètrica mitjançant matrius de Householder:

$$A = \begin{pmatrix} 6 & -1 & 1 & 1 \\ -1 & 5 & -1 & 1 \\ 1 & -1 & 4 & -1 \\ 1 & 1 & -1 & 3 \end{pmatrix} , \quad B = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 0 & 2 & 1 & 2 \\ 2 & 0 & 2 & 1 \end{pmatrix} .$$

b) Explicitau les equacions dels hiperplans respecte als quals s'han fet les reflexions.

53. Sigui A una matriu tridiagonal i B la matriu tridiagonal que s'obté a partir d' A , canviant tan sols el signe dels elements de la diagonal. Demostreu que λ és un valor propi de B si i només si $-\lambda$ és un valor propi d' A .

54. Sigui una matriu tridiagonal $n \times n$ $A = (a_{ij})$ amb $a_{ii} = a$ ($i = 1 \div n$) i $a_{i,i-1} = b$, $a_{i-1,i} = c$ ($i = 2 \div n$).

a) Doneu una fórmula recurrent per al càlcul del polinomi característic.

b) Demostreu que si n és imparell, $\lambda = a$ és un valor propi.

c) Calculeu els valors propis per a $n = 5$.

d) Trobeu els valors propis d' A en el cas general.

55. Quants valors propis positius tenen les matrius $(2n) \times (2n)$ del tipus

$$\begin{pmatrix} 3 & 2 & & & & \\ 2 & -3 & 1 & & & \\ & 1 & 3 & 2 & & \\ & & 2 & -3 & 1 & \\ & & & \ddots & \ddots & \ddots \\ & & & & \ddots & \ddots & \ddots \\ & & & & & 1 & 3 & 2 \\ & & & & & & 2 & -3 \end{pmatrix} ?$$

56. Calculeu els valors i vectors propis de les matrius

$$\begin{pmatrix} 3 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 3 & 2+i & 0 \\ 2-i & 2 & 1+i \\ 0 & 1-i & 2 \end{pmatrix},$$

trobant el polinomi característic per recurrència i fent servir el teorema de Sturm per tal de separar els valors propis.

57. Trobeu els valors i vectors propis de la matriu

$$A = \begin{pmatrix} 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 4 \\ 2 & 1 & 4 & 3 \\ 1 & 4 & 3 & 2 \end{pmatrix},$$

usant prèviament el mètode de Householder per reduir-la a forma tridiagonal simètrica.

58. Determineu els valors i vectors propis de la matriu

$$\begin{pmatrix} 3 & 4 & 0 & 0 \\ 2 & 3 & 4 & 0 \\ 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \end{pmatrix},$$

passant pel càlcul recurrent del polinomi característic.

59. Apliqueu els mètodes iteratius LR i QR al càlcul dels valors i vectors propis de la matriu

$$\begin{pmatrix} 7 & 6 \\ 3 & 4 \end{pmatrix}.$$

CAPÍTOL 3

INTERPOLACIÓ I APROXIMACIÓ DE FUNCIONS

3.1 INTERPOLACIÓ

3.1.1 Concepte d'interpolació

Siguin (x_k, y_k) ($k = 0 \div m$) parelles de valors reals (punts del pla) donats de manera que $x_k \neq x_i$, si $k \neq i$. Hom es pot preguntar si existeix alguna funció p , d'un tipus predeterminat, tal que $p(x_k) = y_k$ ($k = 0 \div m$).

La qüestió anterior s'anomena *problema d'interpolació* i al procés de recerca de la funció p se li dóna el nom d'*interpolació*.

La resolució d'aquest problema és d'utilitat en moltes situacions, en particular quan la procedència dels punts (x_k, y_k) ($k = 0 \div m$) és experimental. Considerem un procés en el qual, per a determinats valors d'una variable x , s'obté un resultat expressat per la variable y . Suposem que, coneguda experimentalment la resposta y_k obtinguda sota condicions x_k , ens interessa de trobar el resultat y que obtindríem en prendre condicions x no experimentades.

Podem pensar que els punts donats formen part de la gràfica d'una funció f que voldríem conèixer almenys aproximadament i de la qual únicament sabem que $f(x_k) = y_k$ ($k = 0 \div m$). Aquesta situació, freqüent en la ciència i la tècnica, porta al típic problema d'interpolació en el qual no aproximem nombres, sinó funcions.

EXEMPLE

El valor de l'acceleració de la gravetat en un punt de la superfície terrestre, al nivell del mar, depèn de la latitud. Experimentalment, s'ha comprovat la correspondència donada a la taula 3.1.

El problema consisteix a trobar el valor de la gravetat a Barcelona, que és situada a una latitud de $41^\circ 25'$.

Davant d'aquest exemple no resulta estrany que, a vegades, s'usi l'expressió "llegir entre línies" per referir-se a la interpolació.

En plantejar-se un problema d'interpolació s'han de fer tres preguntes:

1. De quin tipus (polinòmica, trigonomètrica, racional, etc) ha de ser la funció p bus-

cada?

La resposta a aquesta qüestió vindrà lligada a les “sospites” que tinguem respecte al comportament de f . Si f expressa algun fenomen periòdic, convindrà buscar p entre les funcions trigonomètriques. Si tenim raons per pensar que f té asímptotes, convindrà potser que p sigui de tipus racional. Si f respon a un comportament polinòmic, haurem de buscar p entre les funcions polinòmiques. Només tractarem aquest últim tipus d’interpolació: el de la *interpolació polinòmica*.

2. Elegit el conjunt de funcions, del qual hem d’extreure p , existeix la funció buscada? i, si existeix, és única?
3. És la funció p una bona aproximació de la funció f en els punts en què no coincideixen?

3.1.2 Interpolació polinòmica

Plantejament del problema

Donada una taula de $m + 1$ punts d’interpolació (x_k, f_k) ($k = 0 \div m$) amb $x_k \neq x_i$, si $k \neq i$, anomenarem *interpolació polinòmica* a la determinació d’un polinomi $p(x)$ de grau més petit o igual que N , tal que

$$p(x_k) = f_k \quad (k = 0 \div m) .$$

Quan f_k sigui el valor d’una funció f en x_k ($k = 0 \div m$), parlarem d’*interpolació polinòmica de la funció f* en les *abscisses d’interpolació x_k* ($k = 0 \div m$).

A continuació, es dona resposta, en aquest cas, a les tres preguntes amb què acabàvem l’apartat anterior.

Tipus de funció interpoladora

La funció p buscada formarà part del conjunt de polinomis de grau més petit o igual que N , per a un cert valor de N ; és a dir, $p(x)$ serà de la forma

$$p(x) = a_N x^N + a_{N-1} x^{N-1} + \cdots + a_1 x + a_0 ,$$

Latitud (en graus)	Acceleració gravetat (en m/s^2)
0	9.780350
30	9.793238
45	9.806154
60	9.819099
90	9.832072

Taula 3.1: Mesures de la gravetat a diverses latituds.

i, per tal de determinar-lo, caldrà conèixer els valors dels $N + 1$ coeficients $a_0, a_1, \dots, a_N$ (reals o complexos, segons el cas). En el cas que a_N sigui no nul, direm també que $p(x)$ té exactament grau N .

Donat un nombre α (en general, complex) i suposant $N \geq 1$, la *regla de Horner* ens permet efectuar el procés de *divisió sintètica*, consistent a trobar un polinomi $q(x)$ de grau menor o igual que $N - 1$ i un nombre r , tals que

$$p(x) = (x - \alpha)q(x) + r, \quad (3.1)$$

a través de la recurrència

$$b_{N-1} = a_N, \quad b_j = b_{j+1}\alpha + a_{j+1} \quad (j = N - 2 \div -1),$$

que dóna

$$q(x) = b_{N-1}x^{N-1} + b_{N-2}x^{N-2} + \dots + b_1x + b_0, \quad r = b_{-1}.$$

Se'n dedueixen dues conseqüències importants:

- a) $p(\alpha) = r = b_{-1}$ i, per tant, la regla de Horner és un procés d'avaluació de polinomis amb només N multiplicacions i sumes per a un polinomi de grau menor o igual que N .

A més, el procés de divisió sintètica aplicat a $q(x)$ per al nombre α

$$c_{N-2} = b_{N-1}, \quad c_j = c_{j+1}\alpha + b_{j+1} \quad (j = N - 3 \div -1),$$

dóna el valor de $p'(\alpha) = c_{-1}$, i repetint aquest procés

$$d_{N-3} = c_{N-2}, \quad d_j = d_{j+1}\alpha + c_{j+1} \quad (j = N - 4 \div -1),$$

obtenim el valor de $p^{(2)}(\alpha) = 2d_{-1}$, com resulta de derivar l'expressió (3.1) i avaluar-la en $x = \alpha$. Les derivades d'ordre superior s'obtenen repetint aquest procés.

- b) $p(\alpha) = 0$ si i només si $p(x) = (x - \alpha)q(x)$ o, en altres paraules, un nombre α és un *zero* del polinomi $p(x)$ si i només si $(x - \alpha)$ és un *factor* de $p(x)$. Com que el polinomi $q(x)$ és de grau menor o igual que $N - 1$, repetint aquest procés resulta que si un polinomi $p(x)$ de grau menor o igual que N té N zeros $\alpha_1, \dots, \alpha_n$ (és a dir, $p(\alpha_k) = 0$ ($k = 1 \div N$)), aleshores s'escriu com

$$p(x) = a_N(x - \alpha_1) \cdots (x - \alpha_N),$$

amb a_N coeficient del monomi de grau N . Com a conseqüència, un polinomi $p(x)$ de grau menor o igual que N no pot tenir més de N zeros, a menys que sigui idènticament nul. Aquesta propietat s'anomena també *propietat de Haar* per als polinomis i serà usada freqüentment.

Existència i unicitat del polinomi interpolador

Existeix un únic polinomi $p_m(x)$ de grau més petit o igual que $N = m$, tal que $p_m(x_k) = f_k$ ($k = 0 \div m$). El polinomi $p_m(x)$ rep el nom de *polinomi interpolador* de f en les abscisses x_k ($k = 0 \div m$).

Per veure això, considerem un polinomi $p_m(x) = a_0 + a_1x + \dots + a_mx^m$. Quan imposem les condicions d'interpolació, obtenim un sistema lineal de $m + 1$ equacions en les $m + 1$ incògnites $a_0, a_1, \dots, a_m$:

$$\left. \begin{array}{rcl} a_0 + a_1x_0 + \dots + a_mx_0^m & = & f_0 \\ a_0 + a_1x_1 + \dots + a_mx_1^m & = & f_1 \\ \dots\dots\dots & & \dots \\ a_0 + a_1x_m + \dots + a_mx_m^m & = & f_m \end{array} \right\}.$$

El determinant de la matriu del sistema s'anomena *determinant de Vandermonde* i té la forma

$$\begin{vmatrix} 1 & x_0 & \dots & x_0^m \\ 1 & x_1 & \dots & x_1^m \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 1 & x_m & \dots & x_m^m \end{vmatrix} = \prod_{k>i} (x_k - x_i);$$

aquest determinant és diferent de 0 (ja que $x_i \neq x_k$, si $i \neq k$). Aleshores, el sistema plantejat és compatible i determinat i, per tant, existeix una única solució: $a_0, a_1, \dots, a_m$, coeficients de l'únic polinomi interpolador.

Si bé el procés que s'acaba d'escriure és el primer mètode en què es pensa quan s'intenta trobar $p_m(x)$, resulta excessivament laboriós quan m no és petit. Convé llavors recórrer a altres mètodes, alguns dels quals s'exposen en l'apartat 3.1.3. És important notar que tots els mètodes conduiran al mateix polinomi interpolador (perquè és únic, com s'acaba de veure).

Error d'interpolació

Ens interessa tenir un criteri per "mesurar la proximitat" del polinomi $p_m(x)$ a la funció f .

Donat un interval obert (a, b) , el conjunt de les funcions diferenciables $n + 1$ vegades amb continuïtat a (a, b) es denotarà per $\mathcal{C}^{n+1}(a, b)$. Així mateix, denotarem per $\mathcal{C}^{n+1}([a, b])$ el conjunt de funcions que són restriccions a l'interval $[a, b]$ de funcions diferenciables $n + 1$ vegades amb continuïtat sobre un interval obert que conté $[a, b]$.

A continuació, es donen expressions de l'error d'interpolació per a funcions de $\mathcal{C}^{n+1}(a, b)$.

Si $f \in \mathcal{C}^{n+1}(a, b)$ i $x_k \in (a, b)$ ($k = 0 \div m$); aleshores, per a qualsevol $x \in (a, b)$, tenim l'expressió següent per a l'*error d'interpolació*:

$$f(x) - p_m(x) = \frac{f^{(m+1)}(\xi(x))}{(m+1)!} (x - x_0)(x - x_1) \dots (x - x_m); \quad (3.2)$$

on $\xi(x)$ depèn de x i pertany al mínim interval que conté les abscisses $x_0, x_1, \dots, x_m$ i x , que indicarem per $< x_0, \dots, x_m, x >$.

En efecte: si $x = x_k$ per a algun k , la fórmula (3.2) és certa i, si $x \neq x_k$ ($k = 0 \div m$), considerem la funció

$$\Phi(z) = f(z) - p_m(z) - a(x)(z - x_0)(z - x_1) \cdots (z - x_m),$$

amb $a(x)$ tal que $\Phi(x) = 0$. La funció Φ s'anul·la en les $m + 2$ abscisses: x_k ($k = 0 \div m$) i x ; aplicant $m + 1$ vegades el teorema de Rolle, es dedueix l'existència d'un valor $\xi(x)$ en què s'anul·la la derivada $m + 1$ de Φ :

$$\Phi^{(m+1)}(\xi(x)) = f^{(m+1)}(\xi(x)) - a(x)(m+1)! = 0;$$

així tenim una expressió per a $a(x)$ que, substituïda en la relació $\Phi(x) = 0$, permet trobar l'expressió de l'error (3.2).

3.1.3 Mètodes de càlcul del polinomi interpolador

Destacarem aquí els mètodes següents:

Mètode de Lagrange

Es pren com a expressió del polinomi interpolador la *fórmula d'interpolació de Lagrange*:

$$p_m(x) = \sum_{j=0}^m f_j l_j(x), \quad l_k(x) = \frac{\prod_{i \neq k} (x - x_i)}{\prod_{i \neq k} (x_k - x_i)} \quad (k = 0 \div m). \quad (3.3)$$

Els polinomis $l_k(x)$ reben el nom de *polinomis de Lagrange*. Observeu que el grau de $l_k(x)$ és igual a m i que $l_k(x_i) = \delta_{ik}$ ($i, k = 0 \div m$); per tant, $p(x_i) = f_i$ ($i = 0 \div m$), com es desitjava.

Tot i que aquest és el mètode d'interpolació més explícit (diversos exemples d'aplicació d'aquest mètode es troben en els problemes III.1, III.2 i III.3), els mètodes que segueixen demostren ser més eficaços pel que fa al nombre d'operacions requerides per obtenir el polinomi interpolador.

Mètodes d'Aitken i Neville

Ambdós mètodes obtenen el polinomi interpolador construint una successió de polinomis de graus creixents que van interpolant cada vegada en més abscisses.

Es basen en el principi següent:

Sigui $C = \{x_0, x_1, \dots, x_m\}$ i siguin $S \subset C$ i $T \subset C$ tals que S i T tenen el mateix nombre d'elements i que aquests coincideixen llevat de dos (això és, existeixen $x_j \in S$ i $x_i \in T$ tals que $x_i \notin T$ i $x_j \notin S$). Suposem coneguts els polinomis p_S i p_T tals que $p_S(x_k) = f_k$ (si $x_k \in S$) i $p_T(x_k) = f_k$ (si $x_k \in T$). Llavors es pot construir un nou polinomi $p_{S \cup T}$ tal que $p_{S \cup T}(x_k) = f_k$, per a $x_k \in S \cup T$, de la manera següent:

$$p_{S \cup T}(x) = \frac{(x_j - x)p_S(x) - (x_i - x)p_T(x)}{x_j - x_i}.$$

Els dos mètodes que estudiem construeixen el polinomi interpolador gràcies a successives aplicacions de la fórmula anterior, però seguint un esquema diferent.

Mètode d'Aitken. Es construeixen polinomis $p_{i,j}(x)$ ($i = j \div m$), de grau més petit o igual que j , que interpolen en les abscisses $\{x_0, x_1, \dots, x_{j-1}, x_i\}$ ($j = 0 \div m$); s'arriba al polinomi $p_m(x) \equiv p_{m,m}(x)$, de grau menor o igual que m , que interpolarà en $\{x_0, x_1, \dots, x_m\}$:

$$\begin{aligned} p_{i,0}(x) &= f_i \quad (i = 0 \div m) ; \\ p_{i,j+1}(x) &= \frac{(x_j - x)p_{i,j}(x) - (x_i - x)p_{j,j}(x)}{x_j - x_i} \\ &\quad (i = j + 1 \div n) \quad (j = 0 \div m - 1) . \end{aligned}$$

Mètode de Neville. Es construeixen polinomis $p_{i,j}(x)$ ($i = 0 \div m - j$), de grau menor o igual que j , que interpolen en les abscisses $\{x_i, x_{i+1}, \dots, x_{i+j}\}$ ($j = 0 \div m$), s'arriba al polinomi $p_m(x) \equiv p_{0,m}(x)$, de grau menor o igual que m , que interpolarà en $\{x_0, x_1, \dots, x_m\}$:

$$\begin{aligned} p_{i,0}(x) &= f_i \quad (i = 0 \div m) ; \\ p_{i,j+1}(x) &= \frac{(x_{i+j+1} - x)p_{i,j}(x) - (x_i - x)p_{i+1,j}(x)}{x_{i+j+1} - x_i} \\ &\quad (i = 0 \div m - j - 1) \quad (j = 0 \div m - 1) . \end{aligned}$$

Els esquemes de construcció per a $m = 3$ per a tots dos mètodes es troben a les taules 3.2 i 3.3.

$$\begin{array}{ccccccc} p_{0,0}(x) & & & & & & \\ & \searrow & & & & & \\ p_{1,0}(x) & \rightarrow & p_{1,1}(x) & & & & \\ & & & \searrow & & & \\ p_{2,0}(x) & \rightarrow & p_{2,1}(x) & \rightarrow & p_{2,2}(x) & & \\ & & & & & \searrow & \\ p_{3,0}(x) & \rightarrow & p_{3,1}(x) & \rightarrow & p_{3,2}(x) & \rightarrow & p_{3,3}(x) \end{array}$$

Taula 3.2: Esquema del mètode d'Aitken.

$$\begin{array}{ccccccc} p_{0,0}(x) & \rightarrow & p_{0,1}(x) & \rightarrow & p_{0,2}(x) & \rightarrow & p_{0,3}(x) \\ & \nearrow & & \nearrow & & \nearrow & \\ p_{1,0}(x) & \rightarrow & p_{1,1}(x) & \rightarrow & p_{1,2}(x) & & \\ & \nearrow & & \nearrow & & & \\ p_{2,0}(x) & \rightarrow & p_{2,1}(x) & & & & \\ & \nearrow & & & & & \\ p_{3,0}(x) & & & & & & \end{array}$$

Taula 3.3: Esquema del mètode de Neville.

Si bé aquests dos mètodes serveixen per trobar el polinomi interpolador, a la pràctica, s'usen per avaluar directament $p_m(x) \equiv p_{m,m}(x)$ per a un valor concret de x ja donat.

En aquest cas, les fórmules que descriuen els algorismes respectius representen expressions aritmètiques, no polinòmiques, i els esquemes triangulars de construcció són numèrics, no polinòmics.

Mètode de les diferències dividides de Newton

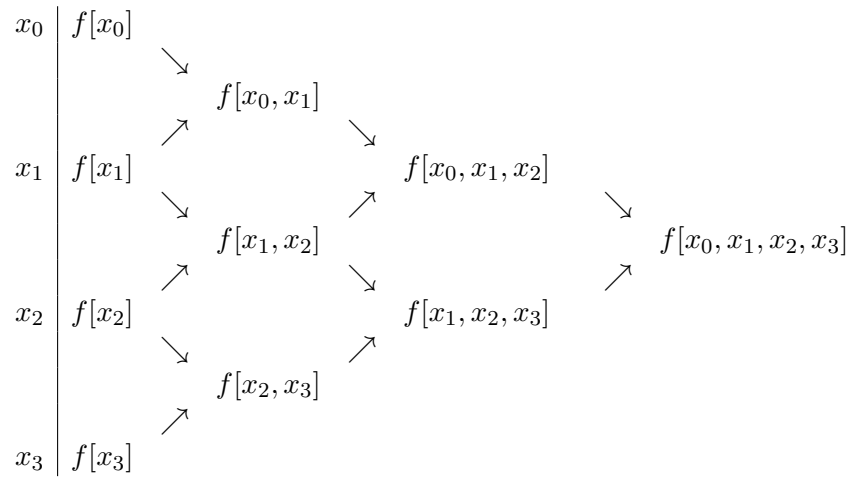
Expressem ara el polinomi interpolador de la manera següent:

$$p_m(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1) + \cdots + c_m(x - x_0)(x - x_1) \cdots (x - x_{m-1}) .$$

El mètode de les diferències dividides ens permet calcular els coeficients c_j ($j = 0 \div m$) mitjançant la construcció de les anomenades *diferències dividides*, definides per:

$$\begin{aligned} f[x_i] &= f_i \quad (i = 0 \div m) , \\ f[x_i, x_{i+1}, \dots, x_{i+j}, x_{i+j+1}] &= \frac{f[x_{i+1}, \dots, x_{i+j+1}] - f[x_i, \dots, x_{i+j}]}{x_{i+j+1} - x_i} \\ &\quad (i = 0 \div m - j) \quad (j = 0 \div m - 1) . \end{aligned}$$

L'esquema de construcció de les diferències dividides de Newton es dona a la taula 3.4 per a $m = 3$, i es pot trobar aplicat en diversos casos, en els problemes III.1-III.4.



Taula 3.4: Esquema de diferències dividides.

Usant la definició de les diferències dividides sobre $x \in [a, b]$, per a $x \neq x_k$ ($k = 0 \div m$), s'arriba, per inducció sobre m , a la fórmula següent:

$$\begin{aligned} f(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \cdots \\ &\quad + f[x_0, x_1, \dots, x_m](x - x_0)(x - x_1) \cdots (x - x_{m-1}) \\ &\quad + f[x_0, x_1, \dots, x_m, x](x - x_0)(x - x_1) \cdots (x - x_{m-1})(x - x_m) . \end{aligned} \quad (3.4)$$

També, per inducció sobre m , es comprova que el polinomi $p_m(x)$ de grau menor o igual que m ve també donat per la fórmula d'interpolació de Newton

$$\begin{aligned} p_m(x) &= f[x_0] + f[x_0, x_1](x - x_0) \\ &\quad + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \cdots \\ &\quad + f[x_0, x_1, \dots, x_m](x - x_0)(x - x_1) \cdots (x - x_{m-1}), \end{aligned} \quad (3.5)$$

que és, de fet, el polinomi $p_{0,m}(x)$ construït pel mètode de Neville i, com a conseqüència, és el polinomi interpolador de f en les abscisses $\{x_0, \dots, x_m\}$; és a dir, en (3.4),

$$c_j = f[x_0, x_1, \dots, x_j].$$

El mètode de les diferències dividides de Newton, a més a més de ser un interessant procediment per al càlcul explícit del polinomi interpolador, té l'avantatge següent, que comparteix amb els mètodes d'Aitken i de Neville: si s'afegeixen més punts d'interpolació, es pot aprofitar tota la feina feta. Per construir el nou polinomi interpolador només cal continuar l'esquema de construcció de diferències dividides i calcular els nous coeficients $c_{m+1}, c_{m+2}, \dots$, aprofitant així tots els càlculs previs.

Notem també que c_j és el coeficient que acompanya el terme x^j del polinomi interpolador p_j de f en les abscisses $x_0, x_1, \dots, x_j$ i, com que el polinomi interpolador no varia en permutar aquestes, tampoc no ho fa c_j ; com a conseqüència, la diferència dividida $f[x_0, x_1, \dots, x_j]$ és una funció simètrica de $x_0, x_1, \dots, x_j$ ($j = 0 \div m$). Per exemple: $f[x_0, x_1] = f[x_1, x_0]$, $f[x_0, x_1, x_2] = f[x_1, x_0, x_2]$, etc.

La fórmula (3.4) dona una nova expressió per a l'error d'interpolació

$$\begin{aligned} f(x) - p_m(x) \\ = f[x_0, x_1, \dots, x_m, x](x - x_0)(x - x_1) \cdots (x - x_{m-1})(x - x_m). \end{aligned}$$

Si les abscisses d'interpolació x_k ($k = 0 \div m$) i el propi x pertanyen a (a, b) i $f \in \mathcal{C}^{m+1}(a, b)$, comparant l'expressió anterior amb (3.2), tenim

$$f[x_0, x_1, \dots, x_m, x] = \frac{f^{(m+1)}(\xi(x))}{(m+1)!},$$

on $\xi(x)$ depèn de x i està contingut a $\langle x_0, x_1, \dots, x_m, x \rangle$.

Quan les abscisses en què s'interpolava són equidistants (o sigui, tals que les diferències $x_{k+1} - x_k$ ($k = 0 \div m-1$) són totes iguals), la fórmula d'interpolació de Newton admet una nova expressió emprant operadors en diferències finites, com es veurà en el capítol següent.

3.1.4 Interpolació de Taylor

Plantejament del problema

Es tracta d'un nou tipus d'interpolació, conceptualment diferent a l'anterior i que s'entendrà més endavant com el cas límit d'aquell quan totes les abscisses d'interpolació tendeixen a ser la mateixa.

Volem ara construir un polinomi que sigui una bona aproximació d'una funció f , prou derivable, en l'entorn d'una sola abscissa x_0 . A tal efecte, el buscarem de manera que no

només interpoli f en x_0 , sinó també que les seves derivades interpolin les derivades de f en x_0 , fins a un cert ordre n donat.

Més concretament, donada una funció f , prou derivable en x_0 , busquem un polinomi $p(x)$ de grau menor o igual que N tal que

$$p^{(j)}(x_0) = f^{(j)}(x_0) \quad (j = 0 \div n) ,$$

on $f^{(j)}$ i $p^{(j)}$ denoten les derivades j -enes de les funcions f i p , respectivament, si $j \geq 1$, i on prenem $f^{(0)} = f$ i $p^{(0)} = p$. Aquest és el *problema d'interpolació de Taylor*.

Existència, unicitat i error

Per tal d'estudiar l'existència, la unicitat i l'error del polinomi interpolador, farem ús del *teorema fonamental del càlcul*:

- Si $f \in \mathcal{C}^1(a, b)$ i $x, x_0 \in (a, b)$, aleshores

$$f(x) - f(x_0) = \int_{x_0}^x f'(s) \, ds . \quad (3.6)$$

La fórmula (3.6) ens resol el problema d'interpolació de Taylor de grau 0 de f en x_0 , essent $p_0(x) = f(x_0)$ l'única solució.

Per a $f \in \mathcal{C}^{n+1}(a, b)$ i $N = n$, fent successives integracions per parts a (3.6), s'obté la solució del problema d'interpolació de Taylor, donada per la *fórmula d'interpolació de Taylor*,

$$\begin{aligned} p_n(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f^{(2)}(x_0)}{2}(x - x_0)^2 + \dots \\ &\quad + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n , \end{aligned} \quad (3.7)$$

juntament amb l'*expressió integral de l'error de la interpolació de Taylor*,

$$\begin{aligned} R_n(x) &= f(x) - p_n(x) = \frac{1}{n!} \int_{x_0}^x (x - s)^n f^{(n+1)}(s) \, ds \\ &= \frac{(x - x_0)^{n+1}}{n!} \int_0^1 (1 - t)^n f^{(n+1)}(x_0 + t(x - x_0)) \, dt , \end{aligned} \quad (3.8)$$

per a $x_0, x \in (a, b)$. Aquesta última fórmula assegura la unicitat de $p_n(x)$.

El polinomi $p_n(x)$ solució rep el nom de *polinomi interpolador de Taylor de grau més petit o igual que n de f en x_0* .

Com que $(x - s)^n$ considerada com a funció de s no canvia de signe a $< x_0, x >$, es pot usar el *teorema del valor mitjà per a integrals*:

- Si g_1 i g_2 són funcions reals definides a $[0, 1]$, tals que g_1 és contínua i g_2 no canvia de signe, existeix algun $\tau \in [0, 1]$ de manera que es compleix

$$\int_0^1 g_1(t)g_2(t)dt = g_1(\tau) \int_0^1 g_2(t)dt .$$

Així, podem obtenir l'expressió de Lagrange de l'error d'interpolació de Taylor, equivalent a (3.8),

$$R_n(x) = f(x) - p_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} (x - x_0)^{n+1}, \quad (3.9)$$

on $\xi(x) \in \langle x_0, x \rangle$ és generalment desconegut.

OBSERVACIÓ IMPORTANT

El polinomi interpolador de Taylor pot ser considerat com a un límit de polinomis interpoladors en abscisses diferents. Així, si considerem el polinomi interpolador $p_m(x) = p_m(x; x_0, \dots, x_m)$ en $m + 1$ abscisses diferents $x_0, x_1, \dots, x_m$, escrit segons la fórmula d'interpolació de Newton (3.6)

$$\begin{aligned} p_m(x; x_0, x_1, \dots, x_m) &= f[x_0] + f[x_0, x_1](x - x_0) \\ &\quad + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots \\ &\quad + f[x_0, x_1, \dots, x_m](x - x_0)(x - x_1) \dots (x - x_{m-1}), \end{aligned}$$

resulta que l'error d'interpolació en abscisses diferents de la fórmula (3.2) tendeix a l'expressió (3.7), quan x_k tendeix a x_0 ($k = 1 \div m$). Per tant, si definim

$$p_m(x; x_0, x_0, \dots, x_0) = \lim_{x_k \rightarrow x_0} p_m(x; x_0, x_1, \dots, x_n),$$

resulta que $p_m(x; x_0, x_0, \dots, x_0) = p_n(x)$, on $p_n(x)$ denota el polinomi interpolador de Taylor de grau menor o igual que $n = m$. Per això, definirem també

$$f[x_0, x_0, \dots, x_0] \equiv \lim_{x_k \rightarrow x_0} f[x_0, x_1, \dots, x_j] = \frac{f^{(j)}(x_0)}{j!}.$$

EXEMPLES

En la taula 3.5 es donen exemples de polinomis interpoladors de Taylor en $x_0 = 0$, amb indicació dels residus R_n respectius, per a diverses funcions definides sobre els intervals (a, b) que s'hi ressenyen. Cal fer dues observacions sobre la notació usada en la taula:

- ξ representa una abscissa en l'interval $\langle 0, x \rangle$,
- s'hi ha fet servir la notació

$$\binom{\alpha}{j} = \frac{\alpha(\alpha-1) \dots (\alpha-j+1)}{j!}.$$

En el problema III.7 es pot trobar la deducció per a algun cas de la taula 3.5. Altres exemples es troben en els problemes III.8 i III.9.

Desenvolupaments de Taylor

Guaitant l'expressió (3.9), ens adonem que $(x - x_0)^{n+1}$ és un factor de $R_n(x)$ que va multiplicat per $\frac{f^{(n+1)}(\xi(x))}{(n+1)!}$ i que aquest restarà fitat en qualsevol entorn tancat de x_0 contingut a (a, b) , si $f \in \mathcal{C}^{n+1}(a, b)$.

A fi d'expressar d'una altra manera aquest fet, introduïrem les notacions següents:

(a, b)	$f(x) = p_n(x) + R_n(x)$
$(-1, 1)$	$(1+x)^{-1} = 1 - x + \cdots + (-1)^n x^n + (-1)^{n+1} \frac{x^{n+1}}{1-x}$
$(-1, 1)$	$(1+x)^\alpha = 1 + \alpha x + \cdots + \binom{\alpha}{n} x^n + \frac{\binom{\alpha}{n+1} x^{n+1}}{(1+\xi)^{n+1-\alpha}}$
$(-\infty, \infty)$	$e^x = 1 + x + \cdots + \frac{x^n}{n!} + e^\xi \frac{x^{n+1}}{(n+1)!}$
$(-\infty, \infty)$	$\sin x = x - \frac{x^3}{3!} + \cdots \pm \frac{x^n}{n!} \mp \cos \xi \frac{x^{n+2}}{(n+2)!} \quad (n \text{ senar})$
$(-\infty, \infty)$	$\cos x = 1 - \frac{x^2}{2!} + \cdots \pm \frac{x^n}{n!} \mp \cos \xi \frac{x^{n+2}}{(n+2)!} \quad (n \text{ parell})$
$(-1, 1)$	$\ln(1+x) = x - \frac{x^2}{2} + \cdots + (-1)^{n+1} \frac{x^n}{n} + (1+\xi)^{-1} \frac{x^{n+1}}{n+1}$

Taula 3.5: Polinomis interpoladors de Taylor i residus.

Donades f i g , funcions definides prop de x_0 , direm que f és *de l'ordre de g en x_0* si existeixen un entorn I de x_0 i una constant $C \geq 0$ tals que

$$|f(x)| \leq C |g(x)|,$$

per a tot $x \in I$. En símbols, escriurem

$$f(x) = \mathcal{O}(g(x)) \quad (x \rightarrow x_0).$$

Direm també que f és *d'ordre més petit que g en x_0* quan

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0.$$

En símbols, escriurem

$$f(x) = o(g(x)) \quad (x \rightarrow x_0).$$

En el nostre cas, si $f \in \mathcal{C}^{n+1}(a, b)$, podem escriure

$$R_n(x) = \mathcal{O}((x - x_0)^{n+1}) \quad (x \rightarrow x_0), \quad (3.10)$$

$$R_n(x) = o((x - x_0)^n) \quad (x \rightarrow x_0); \quad (3.11)$$

aquestes expressions indiquen que, donat $n \geq 0$ i un nombre ϵ (que serà la fita de l'error permès en el càlcul de $f(x)$), podem trobar un entorn $I = (x_0 - \delta, x_0 + \delta)$ de x_0 on l'error comès en avaluar $p_n(x)$ en lloc de $f(x)$ és menor que ϵ . Observem que, sobre aquest interval I , pot ser molt més econòmic avaluar $p_n(x)$ (sumes i productes) que no pas avaluar $f(x)$.

Pel fet de complir-se la relació (3.11), es diu que els polinomis interpoladors de Taylor formen un *desenvolupament asimptòtic*, quan $x \rightarrow x_0$, concepte que serà tractat en l'apartat 4.3.1.

En els exemples de polinomis interpoladors de Taylor donats en la taula 3.5 s'observa que

$$\lim_{n \rightarrow \infty} R_n(x) = 0 \quad \forall x \in (a, b) .$$

Això vol dir que fixat $x \in (a, b)$, podem obtenir una aproximació de $f(x)$ per un polinomi interpolador de Taylor $p_n(x)$ tan bona com vulguem, prenent n prou gran. Les funcions amb aquesta propietat reben el nom de *funcions analítiques* en l'interval (a, b) .

Aquestes funcions satisfan

$$f(x) = \lim_{n \rightarrow \infty} \sum_{j=0}^n \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j \quad \forall x \in (a, b) ,$$

que també se sol escriure com

$$f(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j \quad \forall x \in (a, b) .$$

La successió d'aquests polinomis s'anomena també *desenvolupament en sèrie de Taylor* de f en un entorn (a, b) de x_0 . Aquest concepte de sèrie serà també estudiat en l'apartat 4.3.1.

Seguint aquesta terminologia, el polinomi de Taylor de grau més petit o igual que n rep també el nom de *desenvolupament de Taylor d'ordre n* .

Cal observar que no totes les funcions indefinidament derivables en un punt són analítiques. Per exemple, la funció

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & (x \neq 0) \\ 0 & (x = 0) \end{cases}$$

és indefinidament derivable, però satisfà $f^{(j)}(0) = 0$ ($j \geq 0$), amb la qual cosa $R_n(x) = f(x) \quad \forall x \in \mathbb{R}$ ($n \geq 0$) i, per tant, no és analítica en $x_0 = 0$.

Finalment, fem notar un criteri molt elemental per fitar $R_n(x)$, el *criteri d'alternança dels residus*:

- Si $R_n(x)R_{n+1}(x) < 0$, aleshores $f(x) \in (p_n(x), p_{n+1}(x))$ i, per tant,

$$|R_n(x)| \leq \frac{|f^{(n+1)}(x_0)|}{(n+1)!} |x - x_0|^{n+1} ;$$

això és, l'error és més petit que el primer terme menyspreat.

3.1.5 Interpolació d'Hermite

Plantejament del problema

Siguin (x_k, f_k) ($k = 0 \div m$) $m + 1$ punts coneguts de la gràfica de $f(x)$ i suposem també coneguts els pendents $f'_k \equiv f'(x_k)$ ($k = 0 \div m$). Es planteja la qüestió següent: existeix algun polinomi p de grau més petit o igual que $N = 2m + 1$ (ja que amb $2m + 2$ condicions podem aspirar a determinar $2m + 2$ coeficients) tal que

$$p(x_k) = f_k, \quad p'(x_k) = f'_k \quad (k = 0 \div m) ?$$

Aquest problema d'interpolació, en el qual es busca un polinomi la gràfica del qual no tan sols passi per uns punts donats, sinó que també hi passi amb uns pendents determinats, s'anomena *problema d'interpolació d'Hermite*.

La resposta a la qüestió plantejada anteriorment és que existeix un únic polinomi amb les característiques exigides; s'anomena *polinomi interpolador d'Hermite* i el denotarem per $p_{2m+1}(x)$.

Expressió del polinomi interpolador d'Hermite i del seu error

El dit polinomi admet una expressió explícita anàloga a la donada pel mètode de Lagrange en el seu cas,

$$p_{2m+1}(x) = \sum_{k=0}^m f_k \Phi_k(x) + \sum_{k=0}^m f'_k \Psi_k(x),$$

amb

$$\begin{aligned} \Phi_k(x) &= (1 - 2l'_k(x_k)(x - x_k))l_k^2(x), \\ \Psi_k(x) &= (x - x_k)l_k^2(x) \quad (k = 0 \div m), \end{aligned}$$

on $l_k(x)$ ($k = 0 \div m$) són els polinomis de Lagrange que apareixen en (3.3).

Els polinomis Φ_k, Ψ_k ($k = 0 \div m$) es diuen *polinomis bàsics de la interpolació d'Hermite* perquè aquests polinomis compleixen:

$$\Phi_k(x_i) = \delta_{ki}, \quad \Phi'_k(x_i) = 0 \quad (k, i = 0 \div m);$$

$$\Psi_k(x_i) = 0, \quad \Psi'_k(x_i) = \delta_{ki} \quad (k, i = 0 \div m).$$

A fi de valorar el grau d'aproximació del polinomi a la funció, es disposa d'una expressió de l'error en el cas que f sigui $2m+2$ vegades diferenciable amb continuïtat sobre un interval I tal que $x_k \in I$ ($k = 0 \div m$):

$$f(x) - p_{2m+1}(x) = \frac{f^{(2m+2)}(\xi(x))}{(2m+2)!} (x - x_0)^2 (x - x_1)^2 \cdots (x - x_m)^2$$

per a $x \in I$ i on $\xi(x) \in \langle x_0, x_1, \dots, x_m, x \rangle$.

Interpolació d'Hermite generalitzada

La interpolació d'Hermite pot generalitzar-se al cas en què coneixem la funció f en les abscisses x_k ($k = 0 \div m$) amb més "profunditat", encara que aquesta "profunditat" no sigui la mateixa en totes elles.

Així, donades aquestes abscisses, diferents dues a dues, i derivades successives de la funció f en aquestes

$$x_k \quad , \quad f^{(j)}(x_k) \quad (j = 0 \div n_k) \quad (k = 0 \div m) \quad ,$$

existeix un únic polinomi p_N de grau més petit o igual que

$$N = \sum_{k=0}^m n_k + m$$

tal que

$$p_N^{(j)}(x_k) = f^{(j)}(x_k) \quad (j = 0 \div n_k) \quad (k = 0 \div m) \quad .$$

Pel que fa a l'error en aquest cas, es pot dir que, si $f \in \mathcal{C}^{N+1}(a, b)$ i $x_k \in (a, b)$ ($k = 0 \div m$), es compleix

$$f(x) - p_N(x) = \frac{f^{(N+1)}(\xi(x))}{(N+1)!} (x - x_0)^{n_0+1} (x - x_1)^{n_1+1} \dots (x - x_m)^{n_m+1}$$

per a $x \in (a, b)$ i on $\xi(x) \in \langle x_0, x_1, \dots, x_m, x \rangle$ (vegeu el problema III.13).

Mètode de les diferències dividides generalitzades

Malgrat l'innegable interès teòric de l'expressió que s'ha donat per al polinomi interpolador d'Hermite, en la pràctica i atenent a raons de comoditat, per al seu càlcul s'usa una generalització del mètode de les diferències dividides de Newton en la qual l'esquema triangular es construeix de la manera següent:

1. En la primera columna s'hi col·loquen els valors $f[x_i] = f_i$ ($i = 0 \div m$) de manera que cadascun hi aparegui dues vegades consecutives: $f[x_0], f[x_0], f[x_1], f[x_1], \dots, f[x_m], f[x_m]$.
2. En construir la segona columna, l'aparició d'abscisses repetides no ens permetrà usar la definició habitual de diferència dividida per al càlcul de $f[x_i, x_i]$ ($i = 0 \div m$). En aquest cas, usarem el fet que:

$$f[x_i, x_i] \equiv \lim_{x_k \rightarrow x_i} f[x_i, x_k] = f'(x_i) = f'_i \quad (i = 0 \div m) \quad .$$

3. La construcció de les altres columnes no comporta ja més problemes. Així, obtindrem un esquema triangular com el de la taula 3.6, on les quantitats subratllades són dades.

El polinomi interpolador d'Hermite és llavors

$$\begin{aligned} p(x) = & f[x_0] + f[x_0, x_0](x - x_0) + f[x_0, x_0, x_1](x - x_0)^2 \\ & + f[x_0, x_0, x_1](x - x_0)^2(x - x_1) + \dots \\ & + f[x_0, x_0, \dots, x_m, x_m](x - x_0)^2 \dots (x - x_{m-1})^2(x - x_m) \quad . \end{aligned}$$

x_0	$\underline{f[x_0] = f_0}$			
		$\underline{f[x_0, x_0] = f'_0}$		
x_0	$\underline{f[x_0] = f_0}$		$f[x_0, x_0, x_1]$	
		$f[x_0, x_1]$		$f[x_0, x_0, x_1, x_1] \quad \dots$
x_1	$\underline{f[x_1] = f_1}$		$f[x_0, x_1, x_1]$	$\vdots$
		$\underline{f[x_1, x_1] = f'_1}$	$\vdots$	
x_1	$\underline{f[x_1] = f_1}$	$\vdots$		
$\vdots$	$\vdots$			

Taula 3.6: Diferències dividides generalitzades.

Finalment, cal dir que les diferències dividides generalitzades es poden estendre també per tal de ser aplicades al càlcul del polinomi d'interpolació d'Hermite generalitzada.

Per tal d'il·lustrar el procés que s'acaba de descriure donem l'exemple següent (vegeu també el problema III.12).

EXEMPLE

Cerquem un polinomi de tercer grau la gràfica del qual passi pels punts (1,3) i (2,2) amb pendents respectius 1 i 4.

En aquest cas: $m = 1$, $x_0 = 1$, $x_1 = 2$, $f_0 = 3$, $f_1 = 2$, $f'_0 = 1$ i $f'_1 = 4$. Tindrem l'esquema de la taula 3.7 i el polinomi interpolador d'Hermite

$$\begin{aligned} p(x) &= 3 + 1(x-1) - 2(x-1)^2 + 7(x-1)^2(x-2) \\ &= 7x^3 - 30x^2 + 40x - 14. \end{aligned}$$

$x_0 = 1$	$\underline{f[x_0] = 3}$			
		$\underline{f[x_0, x_0] = 1}$		
$x_0 = 1$	$\underline{f[x_0] = 3}$		$f[x_0, x_0, x_1] = -2$	
		$f[x_0, x_1] = -1$		$f[x_0, x_0, x_1, x_1] = 7$
$x_1 = 2$	$\underline{f[x_1] = 2}$		$f[x_0, x_1, x_1] = 5$	
		$\underline{f[x_1, x_1] = 4}$		
$x_1 = 2$	$\underline{f[x_1] = 2}$			

Taula 3.7: Exemple de diferències dividides generalitzades.

3.2 APROXIMACIÓ DE FUNCIONS

3.2.1 Introducció al problema general d'aproximació

Tot el càlcul aproximat de funcions desenvolupat en la secció anterior ha estat basat en la interpolació com a eina fonamental; destaquem-ne els avantatges següents:

- El polinomi d'interpolació és fàcil de calcular i es disposa d'una fórmula explícita per a l'error d'interpolació que permet estimar els errors comesos.

- És molt útil per generar fórmules de derivació, integració, etc de funcions, tal com es veurà en el capítol següent.
- És especialment apropiada per al càlcul amb funcions donades per taules; és a dir, per a funcions ben conegudes sobre conjunts discrets d'abscisses, potser equiespaiades, on l'error d'arrodoniment dels valors de les funcions és menor que l'error propi d'interpolació.

Ara bé, el procés d'interpolació presenta diversos problemes en altres casos de càlcul de funcions, sobretot quan ens trobem en situacions totalment diferents de les del paràgraf anterior:

- Si tenim un conjunt discret de valors (x_k, y_k) ($k = 0 \div m$) que tenen errors d'arrodoniment apreciables, no és convenient trobar el polinomi de grau més petit o igual que m que interpoli exactament tots aquests valors, ja que el caràcter "oscil·lant" (amb diferents màxims i mínims) de l'error de la interpolació pot provocar que el polinomi interpolador sigui molt diferent de la funció interpolada fora de les abscisses d'interpolació. Per tal de disminuir l'efecte dels errors en les $m + 1$ abscisses, hom pot intentar trobar un polinomi p_n de grau menor o igual n , amb $n \leq m$, tal que els errors

$$e_k = y_k - p_n(x_k) \quad (k = 0 \div m)$$

siguin nuls. Si $n < m$ estem davant d'un problema *sobredeterminat*; és a dir, amb més equacions que incògnites i, per tant, sense solució en general. De fet, el modificarem imposant que els errors e_k siguin tan petits com sigui possible en algun sentit que determinarem més endavant. Aquest procés rep el nom d'*aproximació polinomial*. Si, per diverses raons, volguéssim "aproximar" les dades no necessàriament per un polinomi sinó per una funció f_n d'un tipus donat com

$$f_n(x) = a_0\varphi_0(x) + \dots + a_n\varphi_n(x) ,$$

ens trobaríem davant d'un problema totalment similar, si $n < m$: trobar els paràmetres $a_0, \dots, a_n$ de manera que els errors

$$e_k \equiv e_n(x_k) = y_k - f_n(x_k) \quad (k = 0 \div m)$$

siguin tan petits, en algun sentit, com es pugui.

- Un cas oposat a l'anterior es dona quan tenim una funció f que es coneix en tot x d'un interval I i volem tenir una manera eficient de calcular-la. El procediment estàndard de generació de taules i interpolació no és adequat en una calculadora o en un ordinador, per raons de memòria i eficiència de càlcul, i convé disposar d'una altra expressió calculable (un polinomi, una funció racional, etc) o expressions f_n de manera que la *funció error d'aproximació*

$$e_n(x) = f(x) - f_n(x)$$

sigui prou petita sobre I . Si $f_n(x) = p_n(x)$ (polinomi de grau menor o igual que n) ens trobem de nou amb el problema d'aproximació polinomial. En general, l'aproximació buscada té la forma

$$f_n(x) = F_n(a_0, \dots, a_n, \varphi_0(x), \dots, \varphi_n(x)) ,$$

però molt generalment es pren

$$f_n(x) = a_0\varphi_0(x) + \cdots + a_n\varphi_n(x) ,$$

on $\varphi_0, \dots, \varphi_n$ són funcions donades, fàcilment calculables. El problema es redueix llavors a cercar els paràmetres $a_0, \dots, a_n$ per tal que la funció error d'aproximació $e_n(x)$ sigui tan petita com es pugui en algun sentit determinat a priori.

Podem ara enunciar el *problema general d'aproximació*: donat un conjunt I d'abscisses d'aproximació i unes funcions bàsiques φ_j ($j = 0 \div n$) definides sobre I ; per a cada funció f definida també sobre I , cal cercar

$$f_n(x) = a_0\varphi_0(x) + \cdots + a_n\varphi_n(x)$$

de manera que la magnitud de l'error d'aproximació $e_n(x) = f(x) - f_n(x)$ sigui la més petita possible.

Així, per tal de tenir totalment determinat un problema d'aproximació, cal especificar el conjunt I d'abscisses d'aproximació, les funcions bàsiques i l'eina de mesura de la magnitud de les funcions error.

El conjunt I d'abscisses d'aproximació

Si I és finit ($I = \{x_0, \dots, x_m\}$), parlarem d'*aproximació discreta*; i, si I és un interval d'extremes a i b ($a < b$), parlarem d'*aproximació contínua*. Donar una funció f sobre un conjunt finit $I = \{x_0, \dots, x_m\}$ equival a donar $y_k \equiv f(x_k)$ ($k = 0 \div m$).

Les funcions bàsiques

Les funcions $\varphi_0, \dots, \varphi_n$, definides sobre I , poden triar-se de diverses maneres. El tipus de funcions escollides dependrà, com en el cas d'interpolació, de les "sospites" que tinguem respecte al comportament de f . Si f expressa algun fenomen periòdic, convindrà escollir entre les funcions trigonomètriques: per exemple,

$$\varphi_0(x) = 1 , \quad \varphi_1(x) = \sin x , \quad \varphi_2(x) = \cos x , \quad \dots , \quad \varphi_{2s}(x) = \cos 2sx ,$$

on $n = 2s$ i parlarem d'*aproximació trigonomètrica*. Si f respon a un comportament polinòmic, escollirem cada $\varphi_j(x) = p_j(x)$ entre els polinomis de grau j ($j = 0 \div n$) (prenent per exemple, $\varphi_j(x) = x^j$, encara que aquesta elecció no serà sempre la més adequada) i parlarem d'*aproximació polinomial*. El conjunt $\mathcal{F}_n$ de funcions f_n que es poden posar com a combinació lineal de $\varphi_0, \dots, \varphi_n$

$$f_n(x) = \sum_{j=0}^n a_j \varphi_j(x) , \quad x \in I ,$$

s'anomena també *espai vectorial generat per les funcions* $\varphi_0, \dots, \varphi_n$. Altres tipus d'aproximacions, que no són lineals en els paràmetres $a_0, \dots, a_n$, com són la racional i l'exponencial, no seran tractats aquí.

Mesura de la magnitud de les funcions error: normes funcionals

La magnitud de la funció error d'aproximació $e_n(x) = f(x) - f_n(x)$ en I pot també ser considerada de diferents maneres.

En el cas discret, e_n consisteix en $m + 1$ valors

$$e_k \equiv e_n(x_k) = f(x_k) - f_n(x_k) \quad (k = 0 \div m) ,$$

i mesurar e_n equival a mesurar el vector $(m+1)$ -dimensional de components e_k ($k = 0 \div m$), per mitjà de les normes introduïdes en l'apartat 2.1.2. Les dues normes més emprades per mesurar l'aproximació són:

- la *norma euclidiana*

$$\|e\|_2 = \left(\sum_{k=0}^m |e_k|^2 \right)^{\frac{1}{2}} ,$$

- la *norma del màxim*

$$\|e\|_\infty = \max_{k=0 \div m} |e_k| .$$

Quan es vol donar una importància diferent als errors e_k , s'introdueixen uns *pesos* positius w_k ($k = 0 \div m$) i es poden definir la *norma euclidiana* i la *norma del màxim ponderades*, respecte a aquesta col·lecció de pesos $w = \{w_k\}_{k=0 \div m}$,

$$\|e\|_{2,w} = \left(\sum_{k=0}^m w_k |e_k|^2 \right)^{\frac{1}{2}} , \quad \|e\|_{\infty,w} = \max_{k=0 \div m} w_k |e_k| .$$

En el cas continu, suposem $I = [a, b]$ i, per a cada funció e definida sobre l'interval I , definim:

- la *norma euclidiana*

$$\|e\|_2 = \left(\int_a^b |e(x)|^2 dx \right)^{\frac{1}{2}} ,$$

- la *norma del màxim*

$$\|e\|_\infty = \sup_{x \in I} |e(x)| .$$

Es pot comprovar que les definicions anteriors compleixen les propietats de norma sobre el conjunt $\mathcal{C}([a, b])$ de funcions contínues sobre l'interval $[a, b]$.

Anàlogament al cas discret, donada una *funció pes* $w \in \mathcal{C}([a, b])$ tal que $w(x) > 0$ sobre I , es poden definir:

- la *norma euclidiana ponderada*

$$\|e\|_{2,w} = \left(\int_a^b w(x) |e(x)|^2 dx \right)^{\frac{1}{2}} ,$$

- la norma del màxim ponderada

$$\|e\|_{\infty, w} = \sup_{x \in I} |e(x)| w(x) .$$

També s'admetran algunes *funcions pes* i *intervals singulars* com poden ser: $w(x) = \frac{1}{\sqrt{1-x^2}}$ a $[-1, 1]$, $w(x) = e^{-x}$ a $[0, \infty)$ i $w(x) = e^{-x^2}$ a $\mathbb{R}$.

Tant en el cas discret com en el continu, parlarem d'*aproximació per mínims quadrats* quan l'elecció dels paràmetres es faci per tal de minimitzar una norma euclidiana. En el cas de minimització de normes del màxim, parlarem d'*aproximació minimax*.

3.2.2 Aproximació per mínims quadrats

Definició del problema

Considerem un conjunt I d'abscisses d'aproximació i unes funcions bàsiques φ_j ($j = 0 \div n$). Per a cada funció f definida sobre I , cerquem $f_n^* \in \mathcal{F}_n$ (l'espai vectorial generat per aquelles) de manera que $\|f - f_n^*\|_2$ sigui mínima en $\mathcal{F}_n$; és a dir,

$$\|f - f_n^*\|_2 = \min_{f_n \in \mathcal{F}_n} \|f - f_n\|_2 ,$$

on $\| \cdot \|_2 \equiv \| \cdot \|_{2, w}$ representa aquí qualsevol de les normes euclidianes:

- a) en el cas discret, $I = \{x_0, \dots, x_m\}$, i si $e = (e_0, \dots, e_m)$,

$$\|e\|_{2, w} = \left(\sum_{k=0}^m w_k |e_k|^2 \right)^{\frac{1}{2}} ,$$

on $w = \{w_0, \dots, w_m\}$ és una col·lecció de pesos positius;

- b) en el cas continu, I és un interval de la recta real d'extremes a i b ,

$$\|e\|_{2, w} = \left(\int_a^b w(x) |e(x)|^2 dx \right)^{\frac{1}{2}} ,$$

on $w(x) > 0$ és una funció pes sobre I .

Productes escalars associats

La propietat fonamental de les normes euclidianes és que provenen de sengles *productes escalars*:

- a) en el cas discret:

$$(u, v) \equiv \sum_{k=0}^m w_k u_k v_k , \quad (3.12)$$

- b) en el cas continu:

$$(u, v) \equiv \int_a^b w(x) u(x) v(x) dx ; \quad (3.13)$$

en el sentit que es compleix en ambdós casos:

$$\|e\|_2^2 = (e, e) .$$

Els dits productes escalars satisfan les propietats:

- i) $(u, u) \geq 0$ i $(u, u) = 0$ si i només si $u = 0$;
- ii) $(u, v) = (v, u)$;
- iii) $(a_1 u_1 + a_2 u_2, v) = a_1(u_1, v) + a_2(u_2, v)$, per a qualssevol funcions u_1, u_2, u, v sobre I i nombres reals a_1, a_2 .

La introducció d'aquests productes escalars proporciona un procediment geomètric de resolució del problema que vindrà suggerit pel següent cas particular.

Cas particular i interpretació geomètrica

Si $I = \{x_0, x_1, x_2\}$, una funció real u sobre I ve donada pels tres valors reals $u = (u_0, u_1, u_2)$, on $u_k = u(x_k)$ ($k = 0, 1, 2$); si a més $w_0 = w_1 = w_2 = 1$, donades dues funcions u, v sobre I , el seu producte escalar coincideix amb el producte escalar habitual sobre $\mathbb{R}^3$

$$(u, v) = u_0 v_0 + u_1 v_1 + u_2 v_2 .$$

Identificarem d'ara endavant funcions sobre I amb vectors de $\mathbb{R}^3$.

Si considerem com a espai d'aproximació el subspai $\mathcal{V}_1$ generat per dos vectors $v^{(0)}, v^{(1)}$ de $\mathbb{R}^3$, suposats linealment independents,

$$\mathcal{V}_1 = \{a_0 v^{(0)} + a_1 v^{(1)}, a_0, a_1 \in \mathbb{R}\} ;$$

donat un vector $y \in \mathbb{R}^3$, el problema d'aproximació per mínims quadrats es redueix a trobar $v^* = a_0^* v^{(0)} + a_1^* v^{(1)}$ tal que

$$\|y - v^*\|_2 = \min_{v \in \mathcal{V}_1} \|y - v\|_2 = \min_{a_0, a_1 \in \mathbb{R}} \|y - a_0 v^{(0)} - a_1 v^{(1)}\|_2 .$$

La representació gràfica de la figura 3.1 ens indica que v^* ve caracteritzat per la *propietat de projecció ortogonal* que assegura que $y - v^*$ és un *vector ortogonal* a tot element de $\mathcal{V}_1$; això és,

$$(y - v^*, v) = 0 \quad \forall v \in \mathcal{V}_1 .$$

En efecte, pel teorema de Pitàgores, en ser $y - v^*$ ortogonal a $v - v^* \in \mathcal{V}_1$,

$$\|y - v\|_2^2 = \|y - v^*\|_2^2 + \|v - v^*\|_2^2 \geq \|y - v^*\|_2^2 ,$$

i, per tant, si $(y - v^*, v) = 0 \quad \forall v \in \mathcal{V}_1$,

$$\|y - v^*\|_2 = \min_{v \in \mathcal{V}_1} \|y - v\|_2 .$$

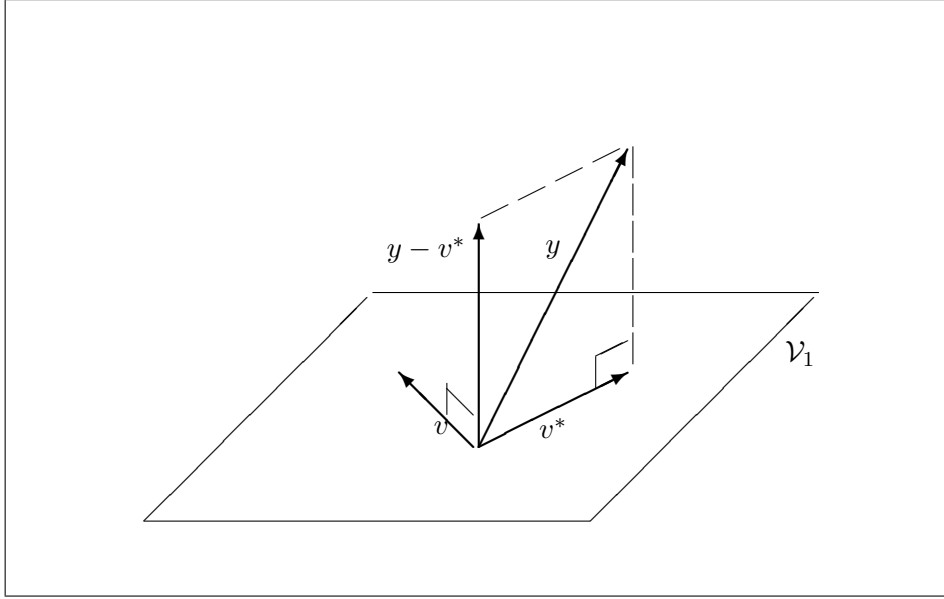
Equacions normals

Retornant al nostre problema general, sigui f_n^* una funció sobre I tal que

$$(f - f_n^*, f_n) = 0 \quad \forall f_n \in \mathcal{F}_n, \quad (3.14)$$

tenim

$$\begin{aligned} \|f - f_n\|_2^2 &= (f - f_n, f - f_n) = (f - f_n^* + f_n^* - f_n, f - f_n^* + f_n^* - f_n) \\ &= (f - f_n^*, f - f_n^*) + 2(f - f_n^*, f_n^* - f_n) + (f_n^* - f_n, f_n^* - f_n) \\ &= \|f - f_n^*\|_2^2 + \|f_n^* - f_n\|_2^2 \end{aligned}$$

Figura 3.1: Projecció ortogonal de y sobre $\mathcal{V}$.

i, per tant, $\|f - f_n\|_2 > \|f - f_n^*\|_2$, si $f_n \neq f_n^*$; és a dir, f_n^* és l'única funció de $\mathcal{F}_n$ que satisfà la condició d'aproximació per mínims quadrats

$$\|f - f_n^*\|_2 = \min_{f_n \in \mathcal{F}_n} \|f - f_n\|_2 .$$

Com que $\mathcal{F}_n$ està generat per φ_i ($i = 0 \div n$), la condició (3.14) equival al fet que els coeficients a_j^* ($j = 0 \div n$) de f_n^* en

$$f_n^*(x) = \sum_{j=0}^n a_j^* \varphi_j(x)$$

satisfacin les anomenades *equacions normals*:

$$\sum_{j=0}^n (\varphi_i, \varphi_j) a_j^* = (\varphi_i, f) \quad (i = 0 \div n) . \quad (3.15)$$

Aquest sistema es pot escriure, en forma matricial, com

$$Aa^* = b ,$$

on $A = ((\varphi_i, \varphi_j))_{i,j=0 \div n}$, $a^* = (a_j^*)_{j=0 \div n}$ i $b = ((\varphi_i, f))_{i=0 \div n}$.

La matriu A és *semidefinida positiva*; això és, simètrica i per a qualsevol vector $a = (a_0 \ a_1 \ \dots \ a_n)^\top$,

$$a^\top Aa = \left(\sum_{j=0}^n a_j \varphi_j, \sum_{l=0}^n a_l \varphi_l \right) = \left\| \sum_{j=0}^n a_j \varphi_j \right\|_2^2 = \|f_n\|_2^2 \geq 0 ,$$

on f_n ve donada per

$$f_n(x) = \sum_{j=0}^n a_j \varphi_j(x) .$$

Aquesta mateixa relació ens demostra que, si les funcions φ_j ($j = 0 \div n$) són linealment independents, $\det A \neq 0$ i viceversa: les equacions normals (3.15) tenen solució única per a qualsevol f si i només si les funcions φ_j ($j = 0 \div n$) són linealment independents.

Exemple: recta de regressió

Tenim un conjunt de dades (x_k, y_k) ($k = 0 \div m$), amb $m > 2$ corresponents a dues variables x i y i sospitem que satisfan una relació lineal del tipus $y = f_1(x) = a_0 + a_1x$, bé perquè la representació gràfica del núvol de punts en el pla així ens ho suggereix, o bé, per altres consideracions aportades pel context del problema real tractat.

A causa dels errors en les dades o perquè el model lineal $y = a_0 + a_1x$ no s'ajusta suficientment al núvol de punts, no trobarem una recta que uneixi tots els $m + 1$ punts, i ens haurem de conformar amb la recta que passi "més a prop de tots els punts".

En l'aproximació per mínims quadrats, s'imposa que sigui mínima la suma dels quadrats de les *desviacions* $d_k = y_k - a_0 - a_1x_k$ ($k = 0 \div m$):

$$\sum_{k=0}^m d_k^2 .$$

Aquest és un problema d'aproximació discreta, amb $I = \{x_0, \dots, x_m\}$, $\varphi_0(x) = 1$, $\varphi_1(x) = x$, en què hem suposat que totes les abscisses eren igualment importants i hem pres així tots els pesos iguals a 1 en el producte escalar; això és

$$(u, v) = \sum_{k=0}^m u_k v_k .$$

Les equacions normals

$$\begin{pmatrix} (\varphi_0, \varphi_0) & (\varphi_0, \varphi_1) \\ (\varphi_1, \varphi_0) & (\varphi_1, \varphi_1) \end{pmatrix} \begin{pmatrix} a_0^* \\ a_1^* \end{pmatrix} = \begin{pmatrix} (\varphi_0, f) \\ (\varphi_1, f) \end{pmatrix} ;$$

formen el sistema lineal de dues equacions i dues incògnites:

$$\begin{pmatrix} m+1 & \sum_{k=0}^m x_k \\ \sum_{k=0}^m x_k & \sum_{k=0}^m x_k^2 \end{pmatrix} \begin{pmatrix} a_0^* \\ a_1^* \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^m y_k \\ \sum_{k=0}^m x_k y_k \end{pmatrix} .$$

La solució d'aquest sistema és:

$$\begin{aligned} a_1^* &= \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} , \\ a_0^* &= \bar{y} - a_1^* \bar{x} , \end{aligned}$$

on la barra indica la mitjana:

$$\bar{x} = \frac{1}{m+1} \sum_{k=0}^m x_k , \quad \bar{y} = \frac{1}{m+1} \sum_{k=0}^m y_k ,$$

$$\overline{x^2} = \frac{1}{m+1} \sum_{k=0}^m x_k^2, \quad \overline{xy} = \frac{1}{m+1} \sum_{k=0}^m x_k y_k.$$

En particular, observem que el *punt mitjà* $(\bar{x}, \bar{y})$ pertany a la recta buscada $\bar{y} = a_0^* + a_1^* \bar{x}$. Tenint en compte aquest fet, podem buscar la recta $y = a_0 + a_1 x$ en la forma alternativa $y = c_0 + c_1(x - \bar{x}) = c_0 \psi_0(x) + c_1 \psi_1(x)$, on clarament $a_1 = c_1$ i $a_0 = c_0 - c_1 \bar{x}$; això és, prenent $\psi_0(x) = 1$ i $\psi_1(x) = x - \bar{x}$ com a funcions bàsiques de l'espai vectorial $\mathcal{P}_1$ dels polinomis de grau menor o igual que 1. Les equacions normals són aleshores

$$\begin{pmatrix} (\psi_0, \psi_0) & (\psi_0, \psi_1) \\ (\psi_1, \psi_0) & (\psi_1, \psi_1) \end{pmatrix} \begin{pmatrix} c_0^* \\ c_1^* \end{pmatrix} = \begin{pmatrix} (\psi_0, f) \\ (\psi_1, f) \end{pmatrix}.$$

Com que

$$(\psi_1, \psi_0) = \sum_{k=0}^m (x_k - \bar{x}) \cdot 1 = \sum_{k=0}^m x_k - (m+1)\bar{x} = 0,$$

les equacions normals queden reduïdes al sistema diagonal

$$\begin{pmatrix} m+1 & 0 \\ 0 & \sum_{k=0}^m (x_k - \bar{x})^2 \end{pmatrix} \begin{pmatrix} c_0^* \\ c_1^* \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^m y_k \\ \sum_{k=0}^m (x_k - \bar{x}) y_k \end{pmatrix}.$$

Usant

$$\sum_{k=0}^m (x_k - \bar{x}) \bar{y} = \bar{y} \sum_{k=0}^m (x_k - \bar{x}) = 0,$$

obtenim

$$a_1^* = c_1^* = \frac{\text{cov}(x, y)}{\text{cov}(x, x)}, \quad a_0^* = c_0^* - c_1^* \bar{x} = \bar{y} - a_1^* \bar{x},$$

on

$$\text{cov}(x, y) \equiv \frac{1}{m+1} \sum_{k=0}^m (x_k - \bar{x})(y_k - \bar{y})$$

rep el nom de *covariància* de les variables x i y , i

$$\text{cov}(x, x) \equiv \sigma^2(x) = \frac{1}{m+1} \sum_{k=0}^m (x_k - \bar{x})^2$$

rep el nom de *variància* de la variable x ; $\sigma(x)$ es diu *desviació típica* o *estàndard* de la variable x .

Cal observar que es disposa de dues expressions matemàticament equivalents per al pendent a_1^* de la recta buscada, però no numèricament equivalents, ja que en general hi haurà cancel·lacions més importants en la primera.

Dient ara $y^* = (y_0^*, \dots, y_m^*)$ on $y_k^* = a_0^* + a_1^* x_k$ ($k = 0 \div m$), si continuem pensant fermament que els punts (x_k, y_k) ($k = 0 \div m$) haurien d'estar sobre una recta, de la qual s'han "escapat" per errors en les dades, podem fer-los "regressar" a la recta $y = a_0^* + a_1^* x$ admetent ara com a punts vàlids els punts (x_k, y_k^*) ($k = 0 \div m$). És per això que la recta

$y = a_0^* + a_1^*x$ trobada rep el nom de *recta de regressió*. És clar que aquesta regressió serà més vàlida com més petit sigui

$$\sum_{k=0}^m d_k^{*2} = \sum_{k=0}^m (y_k - y_k^*)^2 = \|y - y^*\|_2^2 .$$

En termes de la variància de y , tenim

$$\sigma^2(y) = \frac{1}{m+1} \sum_{k=0}^m (y_k - \bar{y})^2 = \frac{\|y - \bar{y}\|_2^2}{m+1} .$$

Usant la propietat d'ortogonalitat de $y - y^*$ respecte a $y^* - \bar{y}$

$$\|y - \bar{y}\|_2^2 = \|y - y^*\|_2^2 + \|y^* - \bar{y}\|_2^2 ,$$

obtenim

$$\sigma^2(y) = \frac{\|y - y^*\|_2^2}{m+1} + \sigma^2(y^*) .$$

La regressió serà, doncs, més fiable com més a prop d'1 sigui el quocient

$$\begin{aligned} \rho_{xy} &= \frac{\sigma(y^*)}{\sigma(y)} = \frac{\|y^* - \bar{y}\|_2}{\|y - \bar{y}\|_2} \\ &= a_1^* \frac{\sigma(x)}{\sigma(y)} = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} , \end{aligned}$$

anomenat *coeficient de regressió*. Si ρ_{xy} no és a prop d'1, el núvol de punts no es distribueix com una recta i caldrà recórrer a altres tipus d'aproximació no lineal per tal d'ajustar-lo.

3.2.3 Resolució de les equacions normals

Una vegada reduït el problema d'aproximació per mínims quadrats, cal resoldre les equacions normals associades

$$Aa^* = b ,$$

on $A = (a_{ij})$ amb $a_{ij} = (\varphi_i, \varphi_j)$ ($i, j = 0 \div n$), $a^* = (a_0^* \dots a_n^*)^\top$ i $b = (b_i)$ amb $b_i = (\varphi_i, f)$ ($i = 0 \div n$).

Com que A és definida positiva, si les funcions φ_j ($j = 0 \div n$) són linealment independents, un mètode especialment adequat és el de Txoleski (vegeu l'apartat 2.2.3), que consisteix a factoritzar A en la forma $LDL^\top$ o $\mathcal{L}\mathcal{L}^\top$ amb $L, \mathcal{L}$ matrius triangulars inferiors, i que requereix $\frac{n^3}{6} + \mathcal{O}(n^2)$ operacions.

La feina preliminar de construcció de les equacions normals requereix generalment $\frac{1}{2}p(n+1)(n+4)$ operacions, on p és el nombre d'operacions necessàries per a cada producte escalar que normalment serà més gran que $n+1$. En el cas de l'aproximació discreta, $p = m+1 \geq n+1$, si I consta de $m+1$ elements; en el cas de l'aproximació contínua, cal calcular les integrals

$$\int_a^b \varphi_i(x) \varphi_j(x) w(x) dx .$$

En general, la major part del càlcul correspondrà a la formació de les equacions normals (3.15). Ara bé, aquesta part de càlcul està fortament condicionada per l'elecció de les

funcions bàsiques $\mathcal{F}_n$. En general, haurem de calcular tots els productes escalars possibles. A més, la matriu A pot estar mal condicionada com a conseqüència del fet que les funcions φ_j ($j = 0 \div n$) no siguin “poc independents” des d’un punt de vista numèric.

En el cas d’emprar una base de funcions ψ_j ($j = 0 \div n$) *ortogonals* (això és, $(\psi_i, \psi_j) = 0$, si $i \neq j$, i $(\psi_i, \psi_i) > 0$), el sistema (3.15) és diagonal (vegeu, per exemple, el problema III.15). Els *coeficients ortogonals* c_j ($j = 0 \div n$) de l’aproximació per mínims quadrats

$$f_n^*(x) = \sum_{j=0}^n c_j^* \psi_j(x)$$

s’obtenen, de forma immediata, fent

$$c_j^* = \frac{b_j}{a_{jj}} = \frac{(\psi_j, f)}{(\psi_j, \psi_j)} \quad (j = 0 \div n) .$$

Per aquesta raó, els mètodes de resolució estàndard de les equacions lineals (3.15) estan basats en l’ortogonalització de les funcions bàsiques φ_j ($j = 0 \div n$); és a dir, en l’expressió de les equacions normals en una base de funcions ortogonals.

A continuació veurem com es porten a terme aquests processos en 5 tipus d’aproximació per mínims quadrats.

- Aproximació discreta: ortogonalitzacions de Householder i de Gram-Schmidt.
- Sistemes lineals sobredeterminats: es redueix al cas anterior.
- Aproximació general: ortogonalització de Gram-Schmidt.
- Aproximació polinomial: polinomis ortogonals.
- Aproximació trigonomètrica: polinomis trigonomètrics ortogonals i interpolació trigonomètrica.

Cas d’aproximació discreta: ortogonalització de Householder i de Gram-Schmidt

Si considerem una aproximació per mínims quadrats ponderada amb els pesos $w_k > 0$ ($k = 0 \div m$), tenim:

$$a_{ij} = (\varphi_i, \varphi_j) = \sum_{k=0}^m w_k \varphi_i(x_k) \varphi_j(x_k) \quad (i, j = 0 \div n) ,$$

$$b_i = (\varphi_i, f) = \sum_{k=0}^m w_k \varphi_i(x_k) f(x_k) \quad (i = 0 \div n) ;$$

aleshores podem escriure la matriu A de les equacions normals en la forma

$$A = M^T W M$$

i el terme independent b en la forma

$$b = M^T W y ,$$

on $M = (m_{kj})$ és una matriu $(m+1) \times (n+1)$ amb $m_{kj} = \varphi_j(x_k)$ ($k = 0 \div m$) ($j = 0 \div n$), $W = (w_k)$ és la matriu diagonal $(m+1) \times (m+1)$ formada pels pesos w_k ($k = 0 \div m$) i $y = (y_k)$ és el vector de $m+1$ components amb $y_k = f(x_k)$ ($k = 0 \div m$).

Les equacions normals s'escriuen, en forma matricial, com

$$M^T W M a^* = M^T W y, \quad (3.16)$$

sistema que té solució única si els vectors

$$(\varphi_j(x_0) \quad \dots \quad \varphi_j(x_m))^T \quad (j = 0 \div n)$$

són linealment independents; això és, si el rang $\text{rg} M$ de la matriu M és $n+1$: $\text{rg} M = n+1$ i, per tant, $m \geq n$.

Suposem primerament que $W = I_{m+1}$; és a dir que tots els pesos són la unitat. Aleshores (3.16) queda en la forma

$$M^T M a^* = M^T y. \quad (3.17)$$

S'exposen a continuació dos mètodes que, seguint les indicacions de l'apartat 2.2.4, permeten trobar la factorització QR (generalitzada) de la matriu M i reduir així les equacions normals a un sistema triangular superior (vegeu el problema III.18).

a) *Mètode d'ortogonalització de Householder*

Trobem una matriu ortogonal $(m+1) \times (m+1)$ formada per composició de n matrius de Householder: $P = P_n P_{n-1} \dots P_1$ de manera que

$$PM = R = \begin{pmatrix} \tilde{R} \\ 0 \end{pmatrix} \begin{matrix} \} n+1 \\ \} m-n \end{matrix},$$

amb $\tilde{R}$ matriu $(n+1) \times (n+1)$ triangular superior, regular si $\text{rg} M = n+1$. En aquest cas $\tilde{R}^T$ és regular, i si diem $\tilde{P}y$ al vector format per les $n+1$ primeres components de Py , tenim que el sistema (3.17) és equivalent al sistema triangular $(n+1) \times (n+1)$

$$\tilde{R}a^* = \tilde{P}y. \quad (3.18)$$

b) *Mètode d'ortogonalització modificat de Gram-Schmidt*

Si $\text{rg} M = n+1$, trobem una matriu $(m+1) \times (n+1)$ Q tal que $Q^T Q = D$, amb D matriu diagonal $(n+1) \times (n+1)$, i una matriu triangular superior $(n+1) \times (n+1)$ R amb elements diagonals $r_{jj} = 1$ ($j = 0 \div n$) tals que $M = QR$. El sistema (3.17) és aleshores equivalent al sistema triangular $(n+1) \times (n+1)$

$$Ra^* = D^{-1} Q^T y. \quad (3.19)$$

Remarquem que no cal construir les equacions normals (3.16), sinó que es pot obtenir directament un dels sistemes triangulars superiors (3.18) i (3.19) (vegeu el problema III.18).

El cas de pesos no iguals a 1 es redueix al cas que s'acaba de veure prenent el sistema $M_1 a = y_1$ amb $M_1 = \sqrt{W} M$ i $y_1 = \sqrt{W} y$, que equival a multiplicar l'equació k -ena pel nombre $\sqrt{w_k}$ ($k = 0 \div m$); notem així la importància de fer servir pesos no iguals per donar més importància a algunes equacions respecte a les altres.

Cas de sistemes lineals sobredeterminats

A l'apartat anterior, hem trobat $a^* = (a_0^* \dots a_n^*)^\top$ de manera que, per a qualsevol altre $a = (a_0 \dots a_n)^\top$, tenim

$$\sum_{k=0}^m \left[f(x_k) - \sum_{j=0}^n a_j^* \varphi_j(x_k) \right]^2 \leq \sum_{k=0}^m \left[f(x_k) - \sum_{j=0}^n a_j \varphi_j(x_k) \right]^2 .$$

Recordant que $y_k = f(x_k)$ ($k = 0 \div m$) i $m_{kj} = \varphi_j(x_k)$ ($k = 0 \div m$) ($j = 0 \div n$), veiem que a^* és la solució del problema de minimització:

$$\|y - Ma^*\|_2 = \min_{a \in \mathbb{R}^{n+1}} \|y - Ma\|_2 .$$

Considerem ara el problema següent: donada una matriu M amb m_0 fileres i n_0 columnes i un vector y de m_0 components, trobar a de n_0 components solució del sistema

$$Ma = y . \quad (3.20)$$

Si $m_0 > n_0$, aquest sistema, en general, no té solució: hi ha més equacions (m_0) que incògnites (n_0) i rep el nom de *sistema lineal sobredeterminat*. Seguint la línia d'aquest capítol, podem intentar trobar un vector $a^* \in \mathbb{R}^{n_0}$ que minimitzi la norma euclidiana de $y - Ma$:

$$\|y - Ma^*\|_2 = \min_{a \in \mathbb{R}^{n_0}} \|y - Ma\|_2$$

que correspon al problema (3.19) i equival, per tant, a resoldre les *equacions normals*

$$M^\top Ma^* = M^\top y .$$

Aquestes equacions tenen solució única si i només si $\text{rg } M = n_0 \leq m_0$ i són, en aquest cas, equivalents a (3.18) i (3.19).

Per donar més importància a algunes equacions de (3.20) respecte a les altres, es pot introduir una *matriu de pesos* positius W diagonal per a la norma euclidiana, tal com ha estat tractat en cloure l'apartat anterior.

Cas d'aproximació general: ortogonalització de Gram-Schmidt

Si tenim una base de funcions φ_j ($j = 0 \div n$) de $\mathcal{F}_n$, el mètode (clàssic) d'ortogonalització de Gram-Schmidt proporciona un algorisme per al càlcul d'una base de funcions ortogonals ψ_j ($j = 0 \div n$) de $\mathcal{F}_n$, plantejant senzillament una factorització QR:

$$\begin{aligned} \varphi_0(x) &= r_{00}\psi_0(x) , \\ \varphi_1(x) &= r_{01}\psi_0(x) + r_{11}\psi_1(x) , \\ &\vdots \\ \varphi_n(x) &= r_{0n}\psi_0(x) + r_{1n}\psi_1(x) + \dots + r_{nn}\psi_n(x) ; \end{aligned} \quad (3.21)$$

o, equivalentment,

$$(\varphi_0(x) \quad \dots \quad \varphi_n(x)) = (\psi_0(x) \quad \dots \quad \psi_n(x)) \begin{pmatrix} r_{00} & r_{01} & \cdots & r_{0n} \\ & r_{11} & \cdots & r_{1n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}, \quad (3.22)$$

que es resol per substitució cap al davant:

$$\begin{aligned} \psi_0(x) &= \varphi_0(x), \quad d_0 = (\psi_0, \psi_0), \quad r_{00} = 1; \\ \psi_j(x) &= \varphi_j(x) - \sum_{i=0}^{j-1} r_{ij} \psi_i(x), \quad d_j = (\psi_j, \psi_j), \\ r_{ij} &= \frac{(\varphi_j, \psi_i)}{d_i} \quad (i = 0 \div j-1), \quad r_{jj} = 1 \quad (j = 0 \div n). \end{aligned} \quad (3.23)$$

Notem que la matriu $R = (r_{ij})$ així obtinguda és triangular superior amb la diagonal plena d'uns, ja que hem suposat φ_j ($j = 0 \div n$) linealment independents.

Introduint ara

$$\tilde{\psi}_j = \frac{\psi_j}{\sqrt{(\psi_j, \psi_j)}} = \frac{\psi_j}{\sqrt{d_j}} \quad (j = 0 \div n),$$

$$\tilde{r}_{ij} = \sqrt{d_i} r_{ij} \quad (0 \leq i \leq j \leq n),$$

obtenim un altre sistema triangular. Les funcions $\tilde{\psi}_j$ ($j = 0 \div n$) formen una *base de funcions ortonormals*:

$$(\tilde{\psi}_i, \tilde{\psi}_j) = \delta_{ij} \quad (i, j = 0 \div n),$$

i la matriu triangular superior $\tilde{R} = (\tilde{r}_{ij})$ té elements positius a la diagonal, no necessàriament unitaris,

$$\tilde{r}_{jj} = \sqrt{d_j} = \sqrt{(\psi_j, \psi_j)} \quad (j = 0 \div n).$$

En qualsevol d'aquestes dues bases de funcions ortogonals, les equacions normals (3.16) són diagonals i, per tant, resolubles immediatament.

Aquest mètode d'ortogonalització de Gram-Schmidt s'aplica tant en el cas d'aproximació contínua com discreta. En aquest últim cas, és matemàticament equivalent al d'ortogonalització modificat de Gram-Schmidt, encara que no numèricament.

En efecte, si tenim un problema d'aproximació discreta sobre un conjunt $I = \{x_0, \dots, x_m\}$, escrivint el sistema triangular (3.22) per a les abscisses $x = x_0, \dots, x_m$, obtenim el sistema

$$M = QR,$$

on les matrius $M = (m_{kj}) = (\varphi_j(x_k))$ i $Q = (q_{kj}) = (\psi_j(x_k))$ són de dimensió $(m+1) \times (n+1)$; R és una matriu $(n+1) \times (n+1)$ triangular superior. Aleshores, la condició d'ortogonalitat

$$(\psi_i, \psi_j) \equiv \sum_{k=0}^m w_k \psi_i(x_k) \psi_j(x_k) = d_j \delta_{ij} \quad (i, j = 0 \div n)$$

equivale a

$$Q^T W Q = D ,$$

on $W = \text{diag}(w_0, \dots, w_n)$ i $D = \text{diag}(d_0, \dots, d_m)$.

Si els pesos són tots iguals a 1, aleshores el mètode clàssic de Gram-Schmidt (3.23) proporciona una factorització QR (generalitzada) de la matriu M de dimensió $(m+1) \times (n+1)$ amb Q de dimensió $(m+1) \times (n+1)$ tal que $Q^T Q = D$ i R triangular superior amb uns a la diagonal. En canvi, la factorització amb $\tilde{\psi}_j, \tilde{r}_{ij}$ dona lloc a una factorització $A = \tilde{Q}^T \tilde{R}$ amb $\tilde{Q}^T \tilde{Q} = I_{n+1}$ i $\tilde{R}$ triangular superior amb elements positius a la diagonal: $\tilde{r}_{jj} > 0$.

Aquest mètode clàssic presenta problemes de cancel·lació si les funcions inicials φ_j ($j = 0 \div n$) són “poc independents”. En tal cas, els coeficients de la matriu R poden ser grans i la fórmula

$$\psi_j(x) = \varphi_j(x) - \sum_{i=0}^{j-1} r_{ij} \psi_i(x)$$

pot presentar problemes de cancel·lació. Tal cosa pot produir que la funció ψ_j calculada numèricament no sigui prou ortogonal a les funcions anteriors ψ_i ($i = 0 \div j-1$) ($j = 0 \div n$).

Observant que el valor (φ_j, ψ_i) no canvia si φ_j és modificat per

$$\varphi_j^{(i)} = \varphi_j - \sum_{l=0}^{i-1} b_l \psi_l \quad (i = 0 \div j-1)$$

i imposant que $\|\varphi_j^{(i)}\|$ sigui mínima, s'obté $b_l = r_{lj}$ ($l = 0 \div j-1$) . Per tal d'evitar les cancel·lacions es recomana de prendre

$$\varphi_j^{(i)} = \varphi_j - \sum_{l=0}^{i-1} b_l \psi_l$$

o, de manera recursiva,

$$\varphi_j^{(0)} = \varphi_j \quad (j = 0 \div n) ,$$

$$\varphi_j^{(i+1)} = \varphi_j^{(i)} - r_{ij} \psi_i \quad (j = i+1 \div n) \quad (i = 0 \div n) .$$

S'obté així el *mètode d'ortogonalització modificat de Gram-Schmidt*

$$\varphi_j^{(0)} = \varphi_j \quad (j = 0 \div n) ,$$

$$\psi_i(x) = \varphi_i^{(i)}(x) , \quad d_i = (\psi_i, \psi_i) ,$$

$$r_{ij} = \frac{(\varphi_j^{(i)}, \psi_i)}{d_i} , \quad \varphi_j^{(i+1)} = \varphi_j^{(i)} - r_{ij} \psi_i \quad (j = i+1 \div n) \quad (i = 0 \div n)$$

que fou ja introduït en l'apartat 2.2.4.

Cas d'aproximació polinomial: polinomis ortogonals

Considerem ara el cas en què φ_j és un polinomi de grau j (per exemple, $\varphi_j(x) = x^j$). L'espai vectorial $\mathcal{F}_n$ generat per les funcions φ_j ($j = 0 \div n$) és ara l'espai vectorial $\mathcal{P}_n$ de polinomis de grau menor o igual que n , i el problema d'aproximació per mínims quadrats polinomial de grau més petit o igual que n pren la forma que s'enuncia a continuació.

Donada $f : I \rightarrow \mathbb{R}$, cerquem $p_n^* \in \mathcal{P}_n$ tal que

$$\|f - p_n^*\| = \min_{p_n \in \mathcal{P}_n} \|f - p_n\|, \quad (3.24)$$

on $\|\cdot\| = \|\cdot\|_{2,w}$ és una norma euclidiana.

Recordem que I pot ser discret ($I = \{x_0, \dots, x_m\}, m \leq n$) o continu (un interval).

Escrivint $p_n^*(x)$ en la forma

$$p_n^*(x) = \sum_{j=0}^n a_j^* \varphi_j(x),$$

les equacions normals que solucionen el problema (3.24) d'aproximació per mínims quadrats

$$\sum_{j=0}^n (\varphi_i, \varphi_j) a_j^* = (\varphi_i, f) \quad (i = 0 \div n) \quad (3.25)$$

tenen solució única, a causa de la independència lineal de $\varphi_0(x), \dots, \varphi_n(x)$ sobre I ; notem que, en el cas discret, cal la condició $n \leq m$, atès que els polinomis satisfan la propietat de Haar introduïda a l'apartat 3.1.2.

La resolució de les equacions normals pot presentar en general molts problemes numèrics. És per això que cal usar mètodes d'ortogonalització. En el cas de polinomis, és molt eficient calcular els polinomis ortogonals a través de relacions de recurrència obtingudes per ortogonalització de Gram-Schmidt, tal com s'explica a continuació.

El primer polinomi (de grau 0) ha de ser una constant: $\psi_0(x) = A_0$, amb $A_0 \neq 0$. Sigui ara $j \geq 0$ i suposem que hem trobat ja els $j+1$ primers polinomis ortogonals $\psi_l(x)$ ($l = 0 \div j$). Cada polinomi $\psi_j(x) = A_j x^j + \dots$ té grau j ; això és, el seu *coeficient principal* A_j és no nul.

Considerem l'ortogonalitat tant en el cas discret com en el continu; és a dir, per als dos productes escalars (3.12) i (3.13). Per a tots dos es compleix la relació $(xu, v) = (u, xv) \forall u, v$. A més, com que els polinomis ortogonals $\psi_l(x)$ ($l = 0 \div j$) són linealment independents (en el cas discret, cal que $j \leq m$), tot polinomi $p_i(x)$ de grau $i \leq j$ s'expressa de manera única com a combinació lineal de $\psi_0(x), \dots, \psi_i(x)$:

$$p_i(x) = c_0 \psi_0(x) + \dots + c_i \psi_i(x).$$

Usant les relacions d'ortogonalitat, tenim que

$$(\psi_j, p_i) = 0, \text{ per a tot } p_i(x) \text{ de grau } i < j. \quad (3.26)$$

Aplicuem ara el mètode clàssic de Gram-Schmidt (3.23) al conjunt format per les funcions $\psi_0, \dots, \psi_j, \varphi_{j+1}$, on escollim $\varphi_{j+1}(x) = \alpha_j x \psi_j(x)$ (polinomi de grau $j+1$, si

$\alpha_j \neq 0$). Com que els primers $j+1$ polinomis ja són ortogonals, només cal trobar $\psi_{j+1}(x) = A_{j+1}x^{j+1} + \dots$, on prefixem un valor no nul per al coeficient principal A_{j+1} de $\psi_{j+1}(x)$.

Trobem

$$\psi_{j+1}(x) = \varphi_{j+1}(x) - \sum_{i=0}^j r_{i,j+1} \psi_i(x) ,$$

on

$$r_{i,j+1} = \frac{(\varphi_{j+1}, \psi_i)}{d_i} = \alpha_j \frac{(x\psi_j, \psi_i)}{d_i} = \alpha_j \frac{(\psi_j, x\psi_i)}{d_i} , \quad d_i = (\psi_i, \psi_i) .$$

De la relació (3.26), tenim

$$\begin{aligned} r_{i,j+1} &= 0 \quad (i = 0 \div j-2) \\ r_{j,j+1} &= \alpha_j \frac{(\psi_j, x\psi_j)}{d_j} \equiv \alpha_j \beta_j , \\ r_{j-1,j+1} &= \alpha_j \frac{(\psi_j, x\psi_{j-1})}{d_{j-1}} \equiv \gamma_j ; \end{aligned}$$

s'obté, doncs, la següent *recurrència dels polinomis ortogonals*:

$$\psi_0(x) = A_0 , \quad \psi_{j+1}(x) = \alpha_j(x - \beta_j)\psi_j(x) - \gamma_j\psi_{j-1}(x) \quad (j \geq 0) ; \quad (3.27)$$

amb

$$\begin{aligned} \alpha_j &= \frac{A_{j+1}}{A_j} \quad (j \geq 0) , \\ \beta_j &= \frac{(\psi_j, x\psi_j)}{(\psi_j, \psi_j)} \quad (j \geq 0) , \\ \gamma_j &= \alpha_j \frac{(\psi_j, x\psi_{j-1})}{(\psi_{j-1}, \psi_{j-1})} = \frac{\alpha_j}{\alpha_{j-1}} \frac{(\psi_j, \psi_j)}{(\psi_{j-1}, \psi_{j-1})} \quad (j \geq 1) , \end{aligned}$$

on hem pres, per conveni, $\psi_{-1}(x) = 0$.

L'última relació per als γ_j s'obté a partir de la relació recurrent (3.27), substituint $j+1$ per j , fent el producte escalar per $\psi_j(x)$ i aplicant (3.26). Obtenim així, de manera única, $\psi_{j+1}(x)$ a partir de $\psi_j(x)$ i $\psi_{j-1}(x)$ i el coeficient principal escollit $A_{j+1} = \alpha_j A_j \neq 0$.

Aquest procés es pot anar repetint indefinidament en el cas continu, però no en el cas discret, ja que calia que $j \leq m$. De fet, si $I = \{x_0, \dots, x_m\}$, de la unicitat dels polinomis ortogonals (un cop fixats els coeficients principals) tenim

$$\psi_{m+1}(x) = A_{m+1}(x - x_0) \cdots (x - x_m) ,$$

que és nul sobre I i

$$\|\psi_{m+1}\|^2 = \sum_{k=0}^m w_k \psi_{m+1}^2(x_k) = 0 .$$

COMENTARIS

1. Si cerquem polinomis ortogonals *mònics* (això és, amb coeficients principals $A_j = 1$ ($j \geq 0$)), aleshores prendrem $\alpha_j = 1$ ($j \geq 0$).

2. Si els pesos són *simètrics* respecte a l'abscissa mitjana $x = c$, aleshores $\beta_j = c$ ($j \geq 0$) i $\psi_j(c+x) = (-1)^j \psi_j(c-x)$. En el cas particular $c = 0$, tenim $\beta_j = 0$ ($j \geq 0$); aleshores, si j és parell, $\psi_j(x)$ és un polinomi parell ($\psi_j(-x) = \psi_j(x)$) i, si j és senar, $\psi_j(x)$ és un polinomi senar ($\psi_j(-x) = -\psi_j(x)$).

Usant ara la relació recurrent dels polinomis ortogonals (3.27), la solució del problema d'aproximació polinomial per mínims quadrats (3.24) ve donada per

$$p_n^*(x) = \sum_{j=0}^n c_j^* \psi_j(x) , \quad (3.28)$$

on els coeficients ortogonals c_j^* són les components de la solució de les equacions normals (3.25), que són diagonals i, per tant,

$$c_j^* = \frac{(\psi_j, f)}{(\psi_j, \psi_j)} \quad (j = 0 \div n) .$$

A més, l'expressió (3.28) juntament amb la recurrència (3.27) proporcionen un mètode eficaç per a l'avaluació de $p_n^*(x)$, anomenat *regla de Clenshaw* (vegeu el problema III.18):

$$\begin{aligned} q_{n+2} &= q_{n+1} = 0 , \\ q_j &= \alpha_j(x - \beta_j)q_{j+1} - \gamma_{j+1}q_{j+2} + c_j^* \quad (j = n \div 0) , \\ p_n^*(x) &= A_0 q_0 . \end{aligned}$$

En efecte,

$$\begin{aligned} p_n^*(x) &= \sum_{j=0}^n c_j^* \psi_j(x) = \sum_{j=0}^n [q_j - \alpha_j(x - \beta_j)q_{j+1} + \gamma_{j+1}q_{j+2}] \psi_j(x) \\ &= \sum_{j=0}^n [\psi_{j+1}(x) - \alpha_j(x - \beta_j)\psi_j(x) + \gamma_j\psi_{j-1}(x)] q_{j+1} + q_0 \psi_j(x) \\ &= A_0 q_0 . \end{aligned}$$

REMARCA

La interpolació d'una funció f en unes abscisses x_k ($k = 0 \div m$) per un polinomi de grau més petit o igual que m ,

$$p_m(x_k) = f(x_k) \quad (k = 0 \div m) ,$$

no és més que un cas particular del problema d'aproximació per mínims quadrats en el cas discret per a pesos arbitraris, en què busquem un polinomi $p_m^*(x)$ de manera que minimitzi

$$\|f - q_m\|_{2,w}^2 = \sum_{k=0}^m [f(x_k) - q_m(x_k)]^2 w_k$$

entre els polinomis $q_m(x)$ de grau més petit o igual que m . Ara bé, aquest mínim s'assoleix per al polinomi d'interpolació $p_m(x)$ atès que es fa zero per a ell.

Per tant, una altra manera de resoldre el problema d'interpolació, diferent de les exposades en l'apartat 3.1.3, consisteix a cercar una base de polinomis ortogonals $\psi_j(x)$ ($j = 0 \div m$) respecte al producte escalar discret

$$(u, v) = \sum_{k=0}^m w_k u(x_k) v(x_k) ,$$

i aleshores $p_m^*(x)$ ve donat explícitament per

$$p_m^*(x) = \sum_{j=0}^m c_j^* \psi_j(x) , \quad c_j^* = \frac{(\psi_j, f)}{(\psi_j, \psi_j)} \quad (j = 0 \div m) .$$

Exemples de famílies de polinomis ortogonals

1. ELS POLINOMIS DE LEGENDRE $P_j(t)$ ($j \geq 0$)

Si $I = [-1, 1]$ i $w(t) = 1$, aleshores els anomenats *polinomis de Legendre* definits per

$$P_0(t) = 1 , \quad P_j(t) = \frac{1}{2^j j!} \frac{d^j}{dt^j} [(t^2 - 1)^j] \quad (j \geq 1) ,$$

són polinomis ortogonals associats.

En efecte, si $i \neq j$, integrant per parts

$$(P_i, P_j) = \int_{-1}^1 P_i(t) P_j(t) dt ,$$

es veu que aquesta integral és nul·la; a més, per inducció, surt que

$$d_j = (P_j, P_j) = \frac{(2j)!}{(2^j j!)^2} \int_{-1}^1 (1 - t^2)^j dt = \frac{2}{2j+1} \quad (j \geq 0) .$$

Com que el coeficient principal és

$$A_j = \frac{2j(2j-1) \cdots (j+1)}{2^j j!} \quad (j \geq 0) ,$$

aleshores

$$\alpha_j = \frac{A_{j+1}}{A_j} = \frac{2j+1}{j+1} , \quad \gamma_j = \frac{\alpha_j d_j}{\alpha_{j-1} d_{j-1}} = \frac{j}{j+1} \quad (j \geq 0) .$$

Usant ara el fet que la funció pes $w(t) = 1$ és *simètrica* respecte a l'abscissa mitjana $c = 0$ de l'interval I , tenim que $\beta_j = 0$ ($j \geq 0$) i que els polinomis de Legendre satisfan

$$P_j(-t) = (-1)^j P_j(t)$$

i la recurrència:

$$\begin{aligned} P_0(t) &= 1 , \quad P_1(t) = t , \\ P_{j+1}(t) &= \frac{2j+1}{j+1} t P_j(t) - \frac{j}{j+1} P_{j-1}(t) \quad (j \geq 1) . \end{aligned}$$

2. ELS POLINOMIS DE GRAM $P_{j,m}(t)$ ($j = 0 \div m$)

Si $t_k = k$ i $w_k = 1$ ($k = 0 \div m$), els polinomis

$$P_{j,m}(t) = \sum_{i=0}^j (-1)^i \binom{j}{i} \binom{j+i}{i} \frac{t(t-1)\cdots(t-i+1)}{m(m-1)\cdots(m-i+1)}$$

són polinomis ortogonals associats: si $j \neq l$ ($j, l = 0 \div m$),

$$(P_{j,m}, P_{l,m}) = \sum_{k=0}^m P_{j,m}(x_k) P_{l,m}(x_k) = 0 ;$$

a més,

$$d_j = (P_{j,m}, P_{j,m}) = \frac{(m+j+1)!(m-j)!}{(2j+1)(m!)^2} \quad (j = 0 \div m) .$$

Com que el coeficient principal de $P_{j,m}(x)$ és

$$A_j = (-1)^j \binom{2j}{j} \frac{1}{m(m-1)\cdots(m-j+1)} ,$$

aleshores:

$$\alpha_j = \frac{A_{j+1}}{A_j} = -\frac{2(2j+1)}{(j+1)(m-j)} ,$$

$$\gamma_j = \frac{\alpha_j d_j}{\alpha_{j-1} d_{j-1}} = \frac{j(m+j+1)}{(j+1)(m-j)} .$$

Usant ara el fet que el conjunt d'abscisses d'aproximació i la funció pes són simètrics respecte a $\beta = \frac{m}{2}$, tenim que $\beta_j = \frac{m}{2}$ ($j = 0 \div m$) i els polinomis de Gram satisfan la recurrència:

$$\begin{aligned} P_{0,m}(t) &= 1 , \quad P_{1,m}(t) = 1 - \frac{2t}{m} , \\ (j+1)(m-j)P_{j+1,m}(t) &= (2j+1)(m-2t)P_{j,m}(t) \\ &\quad - j(m+j+1)P_{j-1,m}(t) \quad (j = 1 \div m-1) . \end{aligned}$$

3. ELS POLINOMIS DE LEGENDRE A L'INTERVAL $[a, b]$

Si $I = [a, b]$ i $w(x) = 1$, els polinomis ortogonals associats $\psi_j(x)$ ($j \geq 0$) s'obtenen dels polinomis de Legendre a través del canvi de variable afí:

$$x = a + \frac{b-a}{2}(t+1) \in [a, b] \Leftrightarrow t = \frac{2}{b-a}\left(x - \frac{a+b}{2}\right) \in [-1, 1] ;$$

és a dir,

$$\psi_j(x) \equiv P_j\left(\frac{2}{b-a}\left(x - \frac{a+b}{2}\right)\right) \quad (j \geq 0)$$

són polinomis ortogonals i satisfan la recurrència:

$$\begin{aligned} \psi_0(x) &= 1 , \quad \psi_1(x) = \frac{2}{b-a}\left(x - \frac{a+b}{2}\right) , \\ \psi_{j+1}(x) &= \frac{2j+1}{j+1} \frac{2}{b-a}\left(x - \frac{a+b}{2}\right) \psi_j(x) \\ &\quad - \frac{j}{j+1} \psi_{j-1}(x) \quad (j \geq 1) . \end{aligned}$$

Anàlogament, si $x_k = x_0 + kh$ i $w_k = 1$ ($k = 0 \div m$), aleshores

$$\psi_{j,m}(x) = P_{j,m}\left(\frac{x-x_0}{h}\right) \quad (j = 0 \div m)$$

són polinomis ortogonals. Per exemple, si $x_k = -1 + \frac{2k}{m}$ ($k = 0 \div m$), aleshores els polinomis $\psi_{j,m}(x) = (-1)^j P_{j,m}(\frac{m}{2}(x+1))$ satisfan

$$\psi_{j,m}(-x) = (-1)^j \psi_{j,m}(x) \quad (j = 0 \div m)$$

i la recurrència:

$$\begin{aligned} \psi_{0,m}(x) &= 1, \quad \psi_{1,m}(x) = x, \\ (j+1)(m-j)\psi_{j+1,m}(x) &= (2j+1)mx\psi_{j,m}(x) \\ &\quad - j(m+j+1)\psi_{j-1,m}(x) \quad (j = 1 \div m-1). \end{aligned}$$

Cas d'aproximació trigonomètrica: ortogonalitat dels polinomis trigonomètrics, interpolació trigonomètrica

Considerem el cas en què $\psi_0(\theta) = \frac{1}{2}$, $\psi_1(\theta) = \cos \theta$, $\psi_2(\theta) = \sin \theta$, ..., $\psi_{2n-1}(\theta) = \cos n\theta$ i $\psi_{2n}(\theta) = \sin n\theta$.

L'espai vectorial $\mathcal{T}_n$ generat per les funcions $\psi_j(\theta)$ ($j = 0 \div 2n$) es diu espai de les *sumes trigonomètriques* t_n de grau menor o igual que n

$$t_n(\theta) = \frac{a_0}{2} + \sum_{j=1}^n a_j \cos j\theta + \sum_{j=1}^n b_j \sin j\theta, \quad (3.29)$$

on suposem que els coeficients ortogonals a_0 , a_j i b_j ($j = 0 \div n$) són reals. És important tenir en compte que, per a qualsevol n , tota suma trigonomètrica $t_n(\theta)$ de grau menor o igual que n és 2π -periòdica; això és,

$$t_n(\theta) = t_n(\theta + 2\pi) \quad \forall \theta \in \mathbb{R}.$$

Usant ara la funció exponencial

$$e^{i\theta} = \cos \theta + i \sin \theta, \quad \cos \theta = \frac{e^{i\theta} + e^{-i\theta}}{2}, \quad \sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i},$$

on $i^2 = -1$, juntament amb la *fórmula de Moivre*

$$\cos j\theta + i \sin j\theta = e^{ij\theta} = (e^{i\theta})^j = (\cos \theta + i \sin \theta)^j,$$

s'obté una altra representació per a les sumes trigonomètriques (3.29), anomenada *desenvolupament de Fourier d'ordre n*

$$t_n(\theta) = \sum_{j=-n}^n c_j e^{ij\theta}, \quad (3.30)$$

on

$$c_0 = \frac{a_0}{2}, \quad c_j = \frac{1}{2}(a_j + ib_j), \quad c_{-j} = \bar{c}_j = \frac{1}{2}(a_j - ib_j) \quad (j = 1 \div n);$$

inversament,

$$\begin{aligned} a_0 &= 2c_0, & a_j &= c_j + c_{-j} = 2\Re(c_j), \\ b_j &= i(c_{-j} - c_j) = 2\Im(c_j) \\ & & (j &= 1 \div n). \end{aligned} \quad (3.31)$$

Aquests coeficients ortogonals dels desenvolupaments de Fourier reben el nom de *coeficients de Fourier*.

Notem que els coeficients de Fourier c_j ($j = -n \div n$) en (3.30) són complexos, però satisfan que $c_{-j} = \bar{c}_j$, si aquesta suma és real; això és, si prové d'una suma (3.29) amb coeficients reals.

Les sumes trigonomètriques de grau menor o igual que n , tant en la forma (3.29) com en la forma (3.30), reben també el nom de *polinomis trigonomètrics* de grau menor o igual que n , ja que es poden escriure en la forma

$$t_n(\theta) = p_n(\cos \theta, \sin \theta) \equiv \sum_{j+l \leq n} d_{j,l}(\cos \theta)^j (\sin \theta)^l,$$

i recíprocament (només cal usar la fórmula de Moivre).

El *problema d'aproximació trigonomètrica* de grau més petit o igual que n s'enuncia tot seguit tant per al cas continu com discret.

CAS CONTINU:

Donada $F : [0, 2\pi] \rightarrow \mathbb{R}$ contínua i tal que $F(0) = F(2\pi)$, cerquem $t_n^* \in \mathcal{T}_n$ tal que

$$\|F - t_n^*\| = \min_{t_n \in \mathcal{T}_n} \|F - t_n\|, \quad (3.32)$$

on $\|\cdot\|$ és la norma euclidiana sobre $[0, 2\pi]$

$$\|F\|^2 \equiv \int_0^{2\pi} F^2(\theta) d\theta.$$

CAS DISCRET:

Donat $m \geq 2n$ i donada $F : I_m \rightarrow \mathbb{R}$ sobre $m+1$ abscisses equiespaiades en qualsevol interval de la forma $[\phi, \phi + 2\pi]$ amb pas $h_m = \frac{2\pi}{m+1}$:

$$I_m = \{\theta_k = \theta_k^{(m)} = \phi + kh_m = \phi + \frac{2\pi k}{m+1} \quad (k = 0 \div m)\},$$

cerquem $t_n^* \in \mathcal{T}_n$ tal que

$$\|F - t_n^*\|_m = \min_{t_n \in \mathcal{T}_n} \|F - t_n\|_m, \quad (3.33)$$

on $\|\cdot\|_m$ és la norma euclidiana sobre I_m :

$$\|F\|_m^2 \equiv \sum_{k=0}^m F^2(\theta_k) = \sum_{k=0}^m F^2\left(\phi + \frac{2\pi k}{m+1}\right).$$

El gran avantatge d'aquesta aproximació és que les funcions

$$\{\psi_j(\theta)\}_{j \geq 0} = \{1, \cos \theta, \sin \theta, \cos 2\theta, \sin 2\theta, \dots\}$$

formen una base de funcions ortogonals en el cas continu, i que les funcions

$$\{\psi_j(\theta)\}_{j=0 \div 2n} = \{1, \cos \theta, \sin \theta, \cos 2\theta, \sin 2\theta, \dots, \cos 2n\theta, \sin 2n\theta\}$$

formen una base de funcions ortogonals en el cas discret si $2n \leq m$:

$$(\psi_j, \psi_l) = \int_0^{2\pi} \psi_j(\theta) \psi_l(\theta) d\theta = \begin{cases} 0 & \text{si } j \neq l, j, l \geq 0 \\ \frac{\pi}{2} & \text{si } j = l = 0 \\ \pi & \text{si } j = l > 0 \end{cases}, \quad (3.34)$$

$$(\psi_j, \psi_l)_m = \sum_{k=0}^m \psi_j(\theta_k) \psi_l(\theta_k) = \begin{cases} 0 & \text{si } j \neq l, j, l = 0 \div 2n \\ \frac{m+1}{4} & \text{si } j = l = 0 \\ \frac{m+1}{2} & \text{si } j = l = 1 \div 2n \end{cases}. \quad (3.35)$$

Per demostrar (3.34) només cal usar les fórmules trigonomètriques:

$$\begin{aligned} \cos A \cos B &= \frac{1}{2} [\cos(A - B) + \cos(A + B)], \\ \sin A \sin B &= \frac{1}{2} [\cos(A - B) - \cos(A + B)], \\ \sin A \cos B &= \frac{1}{2} [\sin(A - B) + \sin(A + B)]. \end{aligned} \quad (3.36)$$

La relació (3.35) s'obté també de la (3.36) i de la suma de la progressió geomètrica

$$\sum_{k=0}^m \cos kA + i \sum_{k=0}^m \sin kA = \sum_{k=0}^m e^{ikA} = \frac{e^{i(m+1)A} - 1}{e^{iA} - 1},$$

quan $e^{iA} \neq 1$.

Ara les equacions normals que resolten els problemes d'aproximació per mínims quadrats (3.32) i (3.33) són diagonals i, per tant, la solució t_n^* ve donada per

$$t_n^*(\theta) = \frac{a_0^*}{2} + \sum_{j=1}^n a_j^* \cos j\theta + \sum_{j=1}^n b_j^* \sin j\theta,$$

on a_0^* , a_j^* i b_j^* ($j = 1 \div n$) vénen donades per les expressions següents:

	Cas continu	Cas discret ($2n \leq m$)
$a_0^* = \frac{(\psi_0, F)}{(\psi_0, \psi_0)} =$	$\frac{1}{\pi} \int_0^{2\pi} F(\theta) d\theta,$	$\frac{2}{m+1} \sum_{k=0}^m F(\theta_k)$
$a_j^* = \frac{(\psi_{2j-1}, F)}{(\psi_{2j-1}, \psi_{2j-1})} =$	$\frac{1}{\pi} \int_0^{2\pi} F(\theta) \cos j\theta d\theta,$	$\frac{2}{m+1} \sum_{k=0}^m F(\theta_k) \cos j\theta_k$
$b_j^* = \frac{(\psi_{2j}, F)}{(\psi_{2j}, \psi_{2j})} =$	$\frac{1}{\pi} \int_0^{2\pi} F(\theta) \sin j\theta d\theta,$	$\frac{2}{m+1} \sum_{k=0}^m F(\theta_k) \sin j\theta_k$

Atès que l'aproximació $t_n^*(\theta)$ així obtinguda és 2π -periòdica, l'aproximació trigonomètrica és especialment útil per aproximar funcions $F : \mathbb{R} \rightarrow \mathbb{R}$ 2π -periòdiques (vegeu el problema III.15). En aquest cas, l'interval $[0, 2\pi]$ es pot substituir per qualsevol interval de la forma $[a, a + 2\pi]$ (per exemple, $[-\pi, \pi]$ és molt usual) sense que això alteri l'aproximació $t_n^*(\theta)$.

Una altra base important de funcions trigonomètriques ortogonals és la formada per les funcions $\psi_j(\theta) = \cos j\theta$ ($j \geq 0$). És ortogonal respecte al producte escalar continu a l'interval $I = [0, \pi]$

$$(F, G) = \int_0^\pi F(\theta)G(\theta)d\theta ,$$

amb $\|\psi_j\|^2 = \frac{\pi}{2}$, si $j > 0$ i $\|\psi_0\|^2 = \pi$.

Les funcions $\psi_j(\theta) = \cos j\theta$ ($j = 0 \div m$) són també ortogonals respecte al producte escalar discret

$$(F, G)_m = \sum_{k=0}^m F(\theta_k)G(\theta_k) , \quad \theta_k = \frac{(2k+1)\pi}{2(m+1)} \quad (k = 0 \div m) ,$$

amb $\|\psi_j\|_m^2 = \frac{m+1}{2}$ ($j = 1 \div m$) i $\|\psi_0\|_m^2 = m+1$.

3.2.4 Aproximació minimax

Definició del problema

Donem un interval fitat $I = [a, b]$ i unes funcions φ_j ($j = 0 \div n$); per a cada funció contínua f definida en I , cerquem $\hat{f}_n \in \mathcal{F}_n$ (l'espai generat per aquelles) de manera que

$$\|f - f_n\|_\infty \equiv \sup_{x \in I} |f(x) - f_n(x)|$$

sigui mínima en $\mathcal{F}_n$ per a $f_n = \hat{f}_n$; això és,

$$\|f - \hat{f}_n\|_\infty = \min_{f_n \in \mathcal{F}_n} \|f - f_n\|_\infty .$$

Cal dir aquí que només considerarem el cas de la norma del suprem sobre un interval, amb funció pes $w(x) = 1$, d'entre totes les normes del suprem introduïdes a l'apartat 3.2.1, i ens concentrarem en els casos d'aproximació polinomial i trigonomètrica, perquè tenim una condició especial sobre llurs zeros, la propietat de Haar:

a) Si $p_n(x) = a_0 + a_1x + \dots + a_nx^n$ és un polinomi de grau més petit o igual que n , no pot tenir més de n zeros a menys que sigui idènticament nul.

Aquesta propietat va ser deduïda a l'apartat 3.1.2, usant només la regla de Horner i, per això, també és vàlida per al cas de coeficients complexos a_j ($j = 0 \div n$) i zeros complexos.

b) Si

$$t_n(\theta) = \frac{a_0}{2} + \sum_{j=1}^n a_j \cos j\theta + \sum_{j=1}^n b_j \sin j\theta = \sum_{j=-n}^n c_j e^{ij\theta}$$

és un *polinomi trigonomètric* de grau més petit o igual que n , no pot tenir més de $2n$ zeros en cap interval $[a, a + 2\pi)$, a menys que sigui el polinomi nul.

Aquesta propietat és conseqüència d'a) aplicada sobre el polinomi de grau menor o igual que $2n$

$$p_{2n}(x) = \sum_{j=0}^n c_{n+j} x^j ,$$

ja que $t_n(\theta) = e^{-in\theta} p_{2n}(e^{i\theta}) = 0$ implica $p_{2n}(e^{i\theta}) = 0$.

Caracterització

Fent ús de la propietat de Haar, pot deduir-se el *teorema de Tchebixev* que caracteritza les aproximacions minimax polinomial i trigonomètrica:

- Donada $f : I \rightarrow \mathbb{R}$ contínua, sigui $\hat{p}_n(x)$ un polinomi de grau menor o igual que n tal que la funció $e_n(x) = f(x) - \hat{p}_n(x)$ tingui almenys $n+2$ *abscisses d'extrem* $a \leq \xi_0 < \xi_1 < \dots < \xi_{n+1} \leq b$ complint la *propietat d'equioscil·lació*

$$|e_n(\xi_j)| = \|e_n\|_\infty = \max_{x \in I} |e_n(x)| \quad (j = 0 \div n+1) ,$$

amb alternança de signes

$$e_n(\xi_0) = -e_n(\xi_1) = \dots = (-1)^{n+1} e_n(\xi_{n+1}) ;$$

aleshores, la norma del suprem sobre l'interval I d' e_n es minimitza dins de l'espai $\mathcal{P}_n$ de polinomis $p_n(x)$ de grau més petit o igual que n per a $p_n = \hat{p}_n$:

$$\|f - \hat{p}_n\|_\infty \leq \|f - p_n\|_\infty .$$

- Donada $F : I = [0, 2\pi] \rightarrow \mathbb{R}$ contínua amb $F(0) = F(2\pi)$, sigui $\hat{t}_n(\theta)$ un polinomi trigonomètric de grau menor o igual que n tal que la funció $e_n(x) = F(x) - \hat{t}_n(x)$ tingui almenys $2n+2$ *abscisses d'extrem* $0 \leq \eta_0 < \eta_1 < \dots < \eta_{2n+1} \leq 2\pi$ complint la *propietat d'equioscil·lació*

$$|e_n(\eta_j)| = \|e_n\|_\infty = \max_{x \in I} |e_n(x)| \quad (j = 0 \div 2n+1) ,$$

amb alternança de signes

$$e_n(\eta_0) = -e_n(\eta_1) = \dots = -e_n(\eta_{2n+1}) ;$$

aleshores, la norma del suprem sobre l'interval I d' e_n es minimitza dins de l'espai $\mathcal{T}_n$ de polinomis trigonomètrics t_n de grau més petit o igual que n per a $t_n = \hat{t}_n$; això és,

$$\|F - \hat{t}_n\|_\infty \leq \|F - t_n\|_\infty .$$

Ambdós resultats s'obtenen per reducció a l'absurd. Per exemple, en el primer, suposem que existeix un polinomi $p_n(x)$ tal que

$$\|f - p_n\|_\infty < \|f - \hat{p}_n\|_\infty ;$$

aleshores,

$$|f(\xi_j) - p_n(\xi_j)| < |f(\xi_j) - \hat{p}_n(\xi_j)| = \|e_n\|_\infty \quad (j = 0 \div n+1) .$$

Per tant, el polinomi $p_n - \hat{p}_n = (f - \hat{p}_n) - (f - p_n) \in \mathcal{P}_n$ té el mateix signe en ξ_j ($j = 0 \div n+1$) que la funció error $e_n = f - \hat{p}_n$. Així, $p_n - \hat{p}_n$ té almenys $n+1$ zeros en (a, b) , la qual cosa contradiu la propietat de Haar i ens porta a un absurd.

Es pot demostrar també que els problemes d'aproximació minimax presentats tenen solució única, aquesta és caracteritzada per la propietat d'oscil·lació uniforme per a la funció error, que hem donat ací com a condició suficient. Ara bé, el càlcul concret dels polinomis $\hat{p}_n(x)$ i dels polinomis trigonomètrics $\hat{t}_n(\theta)$ requereix, en general, l'ús d'un cert tipus de mètodes iteratius deguts a Remes. No obstant això, veurem a continuació alguns casos en què es poden calcular explícitament les aproximacions minimax.

EXEMPLES

1. Sigui $f \in C^2(a, b)$ tal que $f^{(2)}(x) \neq 0 \quad \forall x \in [a, b]$. Si anomenem c la solució única de

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

a (a, b) , aleshores l'aproximació lineal minimax a f sobre l'interval $[a, b]$ ve donada per $\hat{p}_1(x) = \hat{a}_0 + \hat{a}_1 x$, on

$$\hat{a}_1 = \frac{f(b) - f(a)}{b - a} = f'(c) , \quad 2\hat{a}_0 = f(a) + f(c) + \hat{a}_1(a + c)$$

(vegeu el problema III.19).

2. Sigui

$$F(\theta) = a_0 + \sum_{j=1}^{n+1} a_j \cos j\theta + \sum_{j=1}^{n+1} b_j \sin j\theta$$

un polinomi trigonomètric de grau menor o igual que $n+1$; aleshores, l'aproximació minimax trigonomètrica, de grau menor o igual que n , a F sobre $[0, 2\pi]$ ve donada per

$$\begin{aligned} \hat{t}_n(\theta) &= F(\theta) - a_{n+1} \cos(n+1)\theta - b_{n+1} \sin(n+1)\theta \\ &= a_0 + \sum_{j=1}^n a_j \cos j\theta + \sum_{j=1}^n b_j \sin j\theta \end{aligned}$$

(és a dir, es troba simplement suprimint els termes d'ordre $n+1$).

En efecte, notem que, amb aquesta elecció de $\hat{t}_n(\theta)$, la funció error és

$$e_n(\theta) = a_{n+1} \cos(n+1)\theta + b_{n+1} \sin(n+1)\theta = A \cos((n+1)\theta - \theta_0) ;$$

on A i θ_0 són les coordenades "polars" del punt (a_{n+1}, b_{n+1}) :

$$a_{n+1} = A \cos \theta_0 , \quad b_{n+1} = A \sin \theta_0 .$$

Com que $\cos((n+1)\theta - \theta_0)$ té exactament $2n+2$ abscisses d'extrem a $[0, 2\pi]$, en què pren alternativament els valors ± 1 , el teorema de Txebeixev assegura que

$$\|f - \hat{t}_n\|_\infty \leq \|f - t_n\|_\infty \quad \forall t_n \in \mathcal{T}_n .$$

3. Sigui ara

$$F(\theta) = a_0 + \sum_{j=1}^{n+1} a_j \cos j\theta$$

un polinomi trigonomètric de grau menor o igual que $n+1$, que és combinació lineal de només les funcions $\psi_j(\theta) = \cos j\theta$ ($j = 0 \div n+1$); aleshores, el polinomi

$$\hat{t}_n(\theta) = a_0 + \sum_{j=1}^n a_j \cos j\theta$$

dóna l'aproximació minimax trigonomètrica de grau més petit o igual que n sobre l'interval $[0, 2\pi]$. En ser F i $\hat{t}_n(\theta)$ *polinomis trigonomètrics en cosinus*, ambdues funcions queden totalment determinades pels seus valors sobre l'interval $[0, \pi]$, ja que $F(\pi + \theta) = F(\pi - \theta)$ i anàlogament per a $\hat{t}_n$; en altres paraules, es tracta de funcions simètriques respecte a $\theta = \pi$.

Polinomis de Txeixev

Aquest últim exemple permet trobar una estreta relació dels polinomis de Txeixev amb l'aproximació minimax.

Sobre l'interval $[0, \pi]$ té sentit pensar en el canvi de variable

$$x = \cos \theta \in [-1, 1] \Leftrightarrow \theta = \arccos x \in [0, \pi] ,$$

ja que la funció cosinus és bijectiva entre $[0, \pi]$ i $[-1, 1]$; els polinomis de Txeixev es troben fent aquest canvi sobre les funcions $\psi_j(\theta) = \cos j\theta$ ($j \geq 0$):

$$T_j(t) = \psi_j(\arccos t) = \cos(j \arccos t) \quad \forall t \in [-1, 1] .$$

Observem que $T_0(t) = 1$ i que $T_1(t) = t$.

Recordem que les funcions ψ_j ($j \geq 0$) formen una base de funcions trigonomètriques ortogonals respecte al producte escalar continu

$$(F, G) = \int_0^\pi F(\theta)G(\theta)d\theta ;$$

i que les $m+1$ primeres són també ortogonals respecte al producte escalar discret

$$(F, G)_m = \sum_{k=0}^m F(\theta_k)G(\theta_k), \quad \theta_k = \frac{(2k+1)\pi}{2(m+1)} \quad (k = 0 \div m) .$$

A més:

$$\|\psi_0\|_2^2 = \pi , \quad \|\psi_j\|_2^2 = \frac{\pi}{2} \quad (j > 0) ;$$

$$\|\psi_0\|_{2,m}^2 = m+1 , \quad \|\psi_j\|_{2,m}^2 = \frac{m+1}{2} \quad (j = 1 \div m) .$$

Usant les relacions trigonomètriques (3.36), les funcions ψ_j ($j \geq 0$) satisfan la relació de recurrència

$$\psi_{j+1} = 2\psi_1\psi_j - \psi_{j-1} \quad (j \geq 1) ;$$

així resulta que cada $T_j(t)$ és un polinomi de grau més petit o igual que j (que es pot definir per a tot $t \in \mathbb{R}$) i que s'anomena *polinomi de Txeixev* de grau j .

Aquests polinomis satisfan, doncs, les propietats següents:

- Si $t \in [-1, 1]$, $T_j(t) = \cos(j \arccos t)$ ($j \geq 1$).
- $T_0(t) = 1$, $T_1(t) = t$, $T_{j+1}(t) = 2tT_j(t) - T_{j-1}(t)$ ($j \geq 1$). Aquesta relació assegura que els polinomis de Txebixev tenen *paritat definida*:

$$T_j(-t) = (-1)^j T_j(t) \quad (j \geq 0) .$$

- El coeficient principal de $T_j(t)$ és 2^{j-1} : $T_j(t) = 2^{j-1}t^j + \dots$ ($j \geq 0$).
- Els polinomis $T_j(t)$ ($j \geq 0$) són ortogonals respecte al producte escalar continu

$$(f, g) = \int_{-1}^1 \frac{f(t)g(t)}{\sqrt{1-t^2}} dt .$$

A més:

$$\|T_0\|_2^2 = \pi , \quad \|T_j\|_2^2 = \frac{\pi}{2} \quad (j > 0) .$$

- Els polinomis $T_j(t)$ ($j = 0 \div m$) són ortogonals respecte al producte escalar discret

$$(f, g)_m = \sum_{k=0}^m f(t_k)g(t_k) ,$$

on $t_k = \cos \frac{2k+1}{m+1} \frac{\pi}{2}$ ($k = 0 \div m$) són els $m+1$ zeros de $T_{m+1}(t)$.

A més:

$$\|T_0\|_{2,m}^2 = m+1 , \quad \|T_j\|_{2,m}^2 = \frac{m+1}{2} \quad (j = 1 \div m) .$$

Aquesta propietat i l'anterior es dedueixen de la d'ortogonalitat de les funcions ψ_j , esmentades abans, via el canvi $t = \cos \theta$.

- Sigui $p_{n+1}(x) = a_0 + a_1x + \dots + a_{n+1}x^{n+1}$ un polinomi de grau més petit o igual que $n+1$; aleshores,

$$\hat{p}_n(x) = p_{n+1}(x) - \frac{a_{n+1}}{2^{n+1}} T_{n+1}(x)$$

dóna l'aproximació minimax de grau més petit o igual que n sobre l'interval $[-1, 1]$, atès que $\hat{p}_n \in \mathcal{P}_n$ i $T_{n+1}(t) = \cos((n+1) \arccos t)$ té exactament $n+2$ abscisses d'extrem $\xi_k = \cos \frac{k\pi}{n+1}$ ($k = 0 \div n+1$) sobre l'interval $[-1, 1]$, on pren alternativament els valors ± 1 .

Una altra manera d'expressar aquesta propietat és la següent:

- El polinomi mònic de Txebixev de grau $n+1$

$$\frac{1}{2^n} T_{n+1}(t)$$

minimitza la norma del suprem en l'interval $[-1, 1]$ sobre tots el polinomis mònic de grau $n+1$.

En efecte, si $p_{n+1}(t) = t^{n+1} - p_n(t)$ és un polinomi mònic de grau $n+1$, aleshores

$$\begin{aligned} \left\| \frac{T_{n+1}}{2^n} \right\|_{\infty} &= \left\| t^{n+1} - \left(t^{n+1} - \frac{T_{n+1}}{2^n} \right) \right\|_{\infty} \\ &\leq \|t^{n+1} - p_n\|_{\infty} = \|p_{n+1}\|_{\infty} . \end{aligned}$$

Economitització de Lanczos

Si podem aproximar una funció f per un polinomi en un interval $[a, b]$ (per exemple, trobant el seu polinomi de Taylor de grau n prop del punt a) de manera que

$$f(x) = b_0 + b_1(x-a) + b_2(x-a)^2 + \cdots + b_n(x-a)^n + r_n(x) \ , \ x \in [a, b] \ ;$$

llavors, fent el canvi afí de variables

$$\frac{x-a}{b-a} = \frac{t+1}{2} \quad (x \in [a, b] \Leftrightarrow t \in [-1, 1]) \ ,$$

podem traslladar-ho tot a l'interval $[-1, 1]$

$$g(t) = a_0 + a_1t + \cdots + a_nt^n + s_n(t) \ , \ t \in [-1, 1] \ ,$$

on

$$g(t) = f\left(\frac{a+b}{2} + \frac{b-a}{2}t\right) \ , \ s_n(t) = r_n\left(\frac{a+b}{2} + \frac{b-a}{2}t\right) \ ,$$

i aproximar

$$g_n(t) = a_0 + a_1t + \cdots + a_nt^n$$

per

$$g_{n-1}(t) = g_n(t) - \frac{a_n}{2^{n-1}}T_n(t) \ .$$

Tenim així

$$g(t) = g_{n-1}(t) + e_n(t) + s_n(t) \ , \ t \in [-1, 1] \ ,$$

amb $|e_n(t)| \leq \frac{a_n}{2^{n-1}}$, $t \in [-1, 1]$.

Aquest procés es pot tornar a fer sobre g_{n-1}

$$g(t) = g_{n-2}(t) + e_{n-1}(t) + e_n(t) + s_n(t)$$

i així successivament, sempre que els errors $e_j(t)$ ($j \leq n$) siguin prou petits respecte a la precisió exigida en els càlculs. L'avantatge d'aquestes noves expressions és que no requereixen tants càlculs com $g_n(t)$.

El procés anterior és equivalent a escriure $g_n(t)$ en la base ortogonal de polinomis de Tchebixev

$$g(t) = a_0 + a_1T_1(t) + a_2T_2(t) + \cdots + a_nT_n(t) + s_n(t) \ ;$$

aleshores, $e_j(t) = a_jT_j(t)$ ($j \leq n$). Per exemple, si $|s_n(t)| \leq C \ \forall t \in [-1, 1]$,

$$\begin{aligned} g(t) &= a_0 + a_1T_1(t) + \cdots + a_{n-2}T_{n-2}(t) \\ &\pm \left(\frac{|a_{n-1}|}{2^{n-2}} + \frac{|a_n|}{2^{n-1}} + C \right) \quad \forall t \in [-1, 1] \ . \end{aligned}$$

Interpolació de Txebeixev

La *interpolació de Txebeixev* de grau menor o igual que m a l'interval $[-1, 1]$ consisteix a interpolar una funció $g : [-1, 1] \rightarrow \mathbb{R}$ per un polinomi de grau més petit o igual que m en $m + 1$ *abscisses de Txebeixev*; això és, en els zeros de $T_{m+1}(t)$:

$$t_k = \cos \frac{(2k+1)\pi}{2(m+1)} \quad (k = 0 \div m) .$$

L'error comès és així

$$\begin{aligned} g(t) - p_m(t) &= \frac{g^{(m+1)}(\xi(t))}{(m+1)!} (t-t_0)(t-t_1) \cdots (t-t_m) \\ &= \frac{g^{(m+1)}(\xi(x))}{(m+1)!} \frac{T_{m+1}(t)}{2^m} \end{aligned}$$

on $\xi(x) \in (-1, 1)$.

Aquest error es pot fitar per

$$\|g - p_m\|_\infty \leq \frac{\|g^{(m+1)}\|_\infty}{2^m(m+1)!} ;$$

és a dir, en la interpolació de Txebeixev es fa mínima la norma de la funció $(t-t_0) \cdots (t-t_m)$ entre totes les possibles eleccions de $m + 1$ abscisses d'interpolació a l'interval $[-1, 1]$.

A més, com que els polinomis $T_j(t)$ ($j = 0 \div m$) són ortogonals respecte al producte escalar discret

$$(f, g)_m = \sum_{k=0}^m f(t_k)g(t_k) ,$$

els polinomis d'aproximació discreta per mínims quadrats vénen donats directament per

$$p_n^*(t) = \sum_{j=0}^n c_j^* T_j(t) \quad \text{amb} \quad c_j^* = \frac{(f, T_j)_m}{(T_j, T_j)_m} \quad (j = 0 \div n) .$$

El cas límit en què $n = m$ dona lloc a la interpolació de Txebeixev de grau més petit o igual que m , que s'expressa ara com

$$p_m^*(t) = \sum_{j=0}^m c_j^* T_j(t) \quad \text{amb} \quad c_j^* = \frac{(f, T_j)_m}{(T_j, T_j)_m} \quad (j = 0 \div m) .$$

Si la funció que volem interpolar està definida en un interval $[a, b]$; el canvi afí de variables

$$\frac{x-a}{b-a} = \frac{t+1}{2}$$

dóna les $m + 1$ abscisses de Txebeixev sobre l'interval $[a, b]$:

$$x_k = \frac{a+b}{2} + \frac{b-a}{2} t_k = \frac{a+b}{2} + \frac{b-a}{2} \cos \frac{(2k+1)\pi}{2(m+1)} \quad (k = 0 \div m) ;$$

ara, els polinomis ortogonals respecte al producte escalar

$$(f, g)_m = \sum_{k=0}^m f(x_k)g(x_k)$$

vénen donats per $T_j(-1 + \frac{2}{b-a}(x-a))$ ($j = 0 \div m$). La interpolació de Txebixev en l'interval $[a, b]$ es pot trobar també emprant la mateixa tècnica d'aproximació discreta per mínims quadrats (vegeu el problema III.22).

Convé destacar l'eficiència d'aquest mètode de mínims quadrats, quan es vol trobar polinomis d'interpolació de Txebixev, respecte a altres mètodes d'interpolació.

COMENTARIS BIBLIOGRÀFICS

Una de les referències més completes per a aproximació i interpolació és [Dav75] i també [Che66], [Ham73], [Hil74], [IK66], [RR78]. Una bona introducció a la interpolació es troba a [Hen64]. En molts llibres clàssics, l'enfocament donat del càlcul del polinomi interpolador està basat en l'ús de les diferències finites i també dels operadors lineals, conceptes que són introduïts en el proper capítol. Una referència bàsica sobre polinomis ortogonals és [Sze59], i diverses taules i mètodes numèrics per al seu càlcul es troben a [AS65], un dels manuals més complets sobre funcions matemàtiques. Els algorismes de Remes es poden trobar detallats en [Ral65].

PROBLEMES RESOLTS

Problema 3.1 a) Calculeu $f(3)$ per interpolació quadràtica de la taula

x_k	1	2	4	5
f_k	0	2	12	21

i) utilitzant els valors en $x = 1, 2, 4$;

ii) utilitzant els valors en $x = 2, 4, 5$.

b) Calculeu $f(3)$ per interpolació cúbica.

SOLUCIÓ:

Emprarem el mètode de Lagrange.

a) i) En aquest cas, $m = 2$:

$$x_0 = 1, f_0 = 0; \quad x_1 = 2, f_1 = 2; \quad x_2 = 4, f_2 = 12.$$

Els polinomis de Lagrange associats als punts de la taula són:

$$\begin{aligned} l_0(x) &= \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} = \frac{1}{3}(8-6x+x^2), \\ l_1(x) &= \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} = \frac{1}{-2}(4-5x+x^2), \\ l_2(x) &= \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} = \frac{1}{6}(2-3x+x^2), \end{aligned}$$

i el polinomi interpolador és

$$p_2(x) = f_0 l_0(x) + f_1 l_1(x) + f_2 l_2(x) = -x + x^2.$$

El polinomi interpolador en $x = 3$ val 6.

ii) En aquest cas, també $m = 2$:

$$x_0 = 2, f_0 = 2; \quad x_1 = 4, f_1 = 12; \quad x_2 = 5, f_2 = 21.$$

Els polinomis de Lagrange associats als punts de la taula són:

$$\begin{aligned} l_0(x) &= \frac{1}{6}(20-9x+x^2), \\ l_1(x) &= \frac{1}{-2}(10-7x+x^2), \\ l_2(x) &= \frac{1}{3}(8-6x+x^2), \end{aligned}$$

i el polinomi interpolador és

$$p_2(x) = \frac{1}{3}(8 - 9x + 4x^2) .$$

El polinomi interpolador en $x = 3$ val $\frac{17}{3} = 5.666\dots$

b) En aquest cas, $m = 3$:

$$x_0 = 1, f_0 = 0; \quad x_1 = 2, f_1 = 2; \quad x_2 = 4, f_2 = 12; \quad x_3 = 5, f_3 = 21.$$

Els polinomis de Lagrange associats als punts de la taula són:

$$\begin{aligned} l_0(x) &= \frac{1}{-12}(-40 + 38x - 11x^2 + x^3) , \\ l_1(x) &= \frac{1}{6}(-20 + 29x - 10x^2 + x^3) , \\ l_2(x) &= \frac{1}{-6}(-10 + 17x - 8x^2 + x^3) , \\ l_3(x) &= \frac{1}{12}(-8 + 14x - 7x^2 + x^3) , \end{aligned}$$

i el polinomi interpolador és

$$p_3(x) = \frac{1}{12}(-8 + 2x + 5x^2 + x^3) .$$

El polinomi interpolador en $x = 3$ val $\frac{35}{6} = 5.8333\dots$

Problema 3.2 Donada la següent taula de la funció $f(x) = e^x$:

x	0.0	0.2	0.4	0.6
$f(x)$	1.0000	1.2214	1.4918	1.8221

a) Trobeu valors aproximats de $\sqrt[3]{e}$ per interpolació lineal i cúbica, emprant els mètodes de Lagrange i de Newton.

b) Doneu fites respectives dels errors deguts a la interpolació. Compareu les dites fites amb l'error exacte, sabent que $\sqrt[3]{e} = 1.395612425\dots$

SOLUCIÓ:

a) Prendrem tots els punts per tal de fer la interpolació cúbica i només els punts centrals de la taula per a la lineal.

INTERPOLACIÓ LINEAL PEL MÈTODE DE LAGRANGE

$$x_0 = 0.2, f_0 = 1.2214; \quad x_1 = 0.4, f_1 = 1.4918.$$

Els polinomis de Lagrange associats als punts de la taula són:

$$\begin{aligned} l_0(x) &= \frac{1}{-0.2}(-0.4 + x) = 2 - 5x, \\ l_1(x) &= \frac{1}{0.2}(-0.2 + x) = 5x - 1, \end{aligned}$$

i el polinomi interpolador és

$$p_1(x) = 0.9510 + 1.3520x.$$

El polinomi interpolador en $x = \frac{1}{3}$ val $p_1(\frac{1}{3}) = \frac{4.205}{3} = 1.401666\dots$

INTERPOLACIÓ CÚBICA PEL MÈTODE DE LAGRANGE

$$x_0 = 0, f_0 = 1; \quad x_1 = 0.2, f_1 = 1.2214;$$

$$x_2 = 0.4, f_2 = 1.4918; \quad x_3 = 0.6, f_3 = 1.8221.$$

Els polinomis de Lagrange associats als punts de la taula són:

$$\begin{aligned} l_0(x) &= \frac{1}{-0.048}(-0.048 + 0.44x - 1.2x^2 + x^3), \\ l_1(x) &= \frac{1}{0.016}(0.24x - x^2 + x^3), \\ l_2(x) &= \frac{1}{-0.016}(0.12x - 0.8x^2 + x^3), \\ l_3(x) &= \frac{1}{0.048}(0.08x - 0.6x^2 + x^3), \end{aligned}$$

i el polinomi interpolador és

$$p_3(x) = \frac{1}{48}(48 + 48.128x + 22.86x^2 + 10.9x^3).$$

El polinomi interpolador en $x = \frac{1}{3}$ val $p_3(\frac{1}{3}) = \frac{113.0395}{81} \simeq 1.395549$.

INTERPOLACIONS LINEAL I CÚBICA PEL MÈTODE DE LES DIFERÈNCIES DIVIDIDES DE
NEWTON

La taula de diferències dividides de Newton és:

0.0	1.0000			
		1.107		
0.2	1.2214		0.6125	
		1.352	$\frac{0.68125}{3}$	
0.4	1.4918		0.74875	
		1.6515		
0.6	1.8221			

Els polinomis interpoladors seran llavors:

$$\begin{aligned} p_1(x) &= 1.2214 + 1.352(x - 0.2) , \\ p_3(x) &= 1 + 1.107x + 0.6125x(x - 0.2) \\ &\quad + \frac{0.68125}{3}x(x - 0.2)(x - 0.4) , \end{aligned}$$

que, en $x = \frac{1}{3}$, valen $\frac{4.205}{3}$ i $\frac{113.0395}{81}$, respectivament, igual que abans.

b) L'expressió per a l'error d'interpolació pot fitar-se per

$$\frac{e}{2} \left| \left(\frac{1}{3} - 0.2 \right) \left(\frac{1}{3} - 0.4 \right) \right| = \frac{e}{225} \simeq 0.012 ,$$

en la interpolació lineal, i per

$$\frac{e}{24} \left| \left(\frac{1}{3} - 0.0 \right) \left(\frac{1}{3} - 0.2 \right) \left(\frac{1}{3} - 0.4 \right) \left(\frac{1}{3} - 0.6 \right) \right| = \frac{e}{30375} \simeq 0.000089 ,$$

en la cúbica.

Comparant amb el resultat exacte donat 1.395612425..., els errors en les interpolacions lineal i cúbica són, respectivament:

$$\begin{aligned} p_1\left(\frac{1}{3}\right) - \sqrt[3]{e} &= 0.0061 , \\ p_3\left(\frac{1}{3}\right) - \sqrt[3]{e} &= 0.000063 . \end{aligned}$$

Les fites trobades són, doncs, aproximadament el doble i una vegada i mitja els valors dels errors exactes, respectivament.

Problema 3.3 Trobeu els polinomis d'interpolació en punts equidistants de graus 1, 2, 3 i 4 a les funcions:

- a) e^x , en $[0, 1]$;
- b) $\sin x$, en $[0, \pi/2]$;
- c) e^{x^2} , en $[0, 1]$;
- d) $\cos(\cos x)$, en $[0, \pi/2]$;

emprant els mètodes d'interpolació de Lagrange i de les diferències dividides de Newton.

Fiteu l'error comès en qualsevol punt dels intervals respectius.

SOLUCIÓ:

MÈTODE DE LAGRANGE

a) Per a la funció $f(x) = e^x$, que suposem coneguda en tot punt de l'interval $[0, 1]$ amb 6 decimals, obtindrem els polinomis d'interpolació.

En el cas $m = 1$, els valors de la funció en els punts $x_i = \frac{i}{1}$ ($i = 0 \div 1$) són:

$$f_0 = 1, \quad f_1 = 2.718282,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned} l_0(x) &= 1 - x, \\ l_1(x) &= x. \end{aligned}$$

El polinomi interpolador és

$$p_1(x) = 1 + 1.718282x.$$

En el cas $m = 2$, els valors de la funció en els punts $x_i = \frac{i}{2}$ ($i = 0 \div 2$) són:

$$f_0 = 1, \quad f_1 = 1.648721, \quad f_2 = 2.718282,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned} l_0(x) &= 1 - 3x + 2x^2, \\ l_1(x) &= 4x - 4x^2, \\ l_2(x) &= -x + 2x^2. \end{aligned}$$

El polinomi interpolador és

$$p_2(x) = 1 + 0.8766033x + 0.8416786x^2.$$

En el cas $m = 3$, els valors de la funció en els punts $x_i = \frac{i}{3}$ ($i = 0 \div 3$) són:

$$f_0 = 1, \quad f_1 = 1.395612, \quad f_2 = 1.947734, \quad f_3 = 2.718282,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned} l_0(x) &= 1 - 5.5x + 9x^2 - 4.5x^3, \\ l_1(x) &= 9x - 22.5x^2 + 13.5x^3, \\ l_2(x) &= -4.5x + 18x^2 - 13.5x^3, \\ l_3(x) &= x - 4.5x^2 + 4.5x^3. \end{aligned}$$

El polinomi interpolador és

$$p_3(x) = 1 + 1.01399x + 0.4256649x^2 + 0.2786264x^3.$$

En el cas $m = 4$, els valors de la funció en els punts $x_i = \frac{i}{4}$ ($i = 0 \div 4$) són:

$$f_0 = 1, f_1 = 1.284025, f_2 = 1.648721, f_3 = 2.117, f_4 = 2.718282,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned} l_0(x) &= 1 - 8.333333x + 23.33333x^2 - 26.66667x^3 + 10.66667x^4, \\ l_1(x) &= 16x - 69.33333x^2 + 96x^3 - 42.66667x^4, \\ l_2(x) &= -12x + 76x^2 - 128x^3 + 64x^4, \\ l_3(x) &= 5.333333x - 37.33333x^2 + 74.66667x^3 - 42.66667x^4, \\ l_4(x) &= -x + 7.333333x^2 - 16x^3 + 10.66667x^4. \end{aligned}$$

El polinomi interpolador és

$$p_4(x) = 1 + 0.998803x + 0.5097871x^2 + 0.140276x^3 + 0.06941567x^4.$$

b) Per a la funció $f(x) = \sin x$, calculada amb 7 decimals en l'interval $[0, \frac{\pi}{2}]$, obtindrem els polinomis interpoladors.

En el cas $m = 1$, els valors de la funció en els punts $x_i = \frac{i\pi}{2}$ ($i = 0 \div 1$) són:

$$f_0 = 0, f_1 = 1,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned} l_0(x) &= 1 - 0.6366198x, \\ l_1(x) &= 0.6366198x. \end{aligned}$$

El polinomi interpolador és

$$p_1(x) = 0.6366198x.$$

En el cas $m = 2$, els valors de la funció en els punts $x_i = \frac{i\pi}{4}$ ($i = 0 \div 2$) són:

$$f_0 = 0, f_1 = 0.7071068, f_2 = 1,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned} l_0(x) &= 1 - 1.909859x + 0.8105695x^2, \\ l_1(x) &= 2.546479x - 1.621139x^2, \\ l_2(x) &= -0.6366198x + 0.8105695x^2. \end{aligned}$$

El polinomi interpolador és

$$p_2(x) = 1.164013x - 0.3357489x^2.$$

En el cas $m = 3$, els valors de la funció en els punts $x_i = \frac{i\pi}{6}$ ($i = 0 \div 3$) són:

$$f_0 = 0, f_1 = 0.5, f_2 = 0.8660254, f_3 = 1,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned}l_0(x) &= 1 - 3.501409x + 3.647563x^2 - 1.161055x^3, \\l_1(x) &= 5.729578x - 9.118907x^2 + 3.483166x^3, \\l_2(x) &= -2.864789x + 7.295125x^2 - 3.483166x^3, \\l_3(x) &= 0.6366198x - 1.823781x^2 + 1.161055x^3.\end{aligned}$$

El polinomi interpolador és

$$p_3(x) = 1.020429x - 0.0654708x^2 - 0.1138719x^3.$$

En el cas $m = 4$, els valors de la funció en els punts $x_i = \frac{i\pi}{8}$ ($i = 0 \div 4$) són:

$$f_0 = 0, \quad f_1 = 0.3826834, \quad f_2 = 0.7071068, \quad f_3 = 0.9238795, \quad f_4 = 1,$$

i els polinomis de Lagrange associats als dits punts vénen donats per:

$$\begin{aligned}l_0(x) &= 1 - 5.305165x + 9.456644x^2 - 6.880327x^3 + 1.752061x^4, \\l_1(x) &= 10.18592x - 28.09974x^2 + 24.76918x^3 - 7.008244x^4, \\l_2(x) &= -7.639437x + 30.80164x^2 - 33.02557x^3 + 10.51237x^4, \\l_3(x) &= 3.395305x - 15.13063x^2 + 19.26492x^3 - 7.008244x^4, \\l_4(x) &= -0.6366198x + 2.972088x^2 - 4.128196x^3 + 1.752061x^4.\end{aligned}$$

El polinomi interpolador és

$$p_4(x) = 0.996317x + 0.01995143x^2 - 0.2035855x^3 + 0.02871423x^4.$$

Noteu que tots els polinomis de Lagrange d'aquest apartat s'obtenen substituint x per $\frac{2}{\pi}x \simeq 0.636619772x$ en les expressions dels respectius polinomis de Lagrange de l'apartat a).

c) Per a la funció $f(x) = \exp(x^2)$ en l'interval $[0, 1]$, calculada amb 6 xifres decimals exactes, ja no escrivim els polinomis de Lagrange atès que són els mateixos que en l'apartat a).

En el cas $m = 1$, els valors de la funció en els punts $x_i = \frac{i}{1}$ ($i = 0 \div 1$) són:

$$f_0 = 1, \quad f_1 = 2.718282.$$

El polinomi interpolador és

$$p_1(x) = 1 + 1.718282x.$$

En el cas $m = 2$, els valors de la funció en els punts $x_i = \frac{i}{2}$ ($i = 0 \div 2$) són:

$$f_0 = 1, \quad f_1 = 1.284025, \quad f_2 = 2.718282.$$

El polinomi interpolador és

$$p_2(x) = 1 - 0.5821802x + 2.300462x^2.$$

En el cas $m = 3$, els valors de la funció en els punts $x_i = \frac{i}{3}$ ($i = 0 \div 3$) són:

$$f_0 = 1, f_1 = 1.117519, f_2 = 1.559623, f_3 = 2.718282.$$

El polinomi interpolador és

$$p_3(x) = 1 + 0.2576477x - 0.3032243x^2 + 1.763858x^3.$$

En el cas $m = 4$, els valors de la funció en els punts $x_i = \frac{i}{4}$ ($i = 0 \div 4$) són:

$$f_0 = 1, f_1 = 1.064494, f_2 = 1.284025,$$

$$f_3 = 1.755055, f_4 = 2.718282.$$

El polinomi interpolador és

$$p_4(x) = 1 - 0.06771732x + 1.526342x^2 - 1.27888x^3 + 1.538537x^4.$$

d) Per a la funció $f(x) = \cos(\cos(x))$, calculada amb 6 xifres decimals exactes en l'interval $[0, \frac{\pi}{2}]$, tampoc no escriurem els polinomis de Lagrange, atès que apareixen en l'apartat b).

En el cas $m = 1$, els valors de la funció en els punts $x_i = \frac{i\pi}{2}$ ($i = 0 \div 1$) són:

$$f_0 = 0.5403023, f_1 = 1.$$

El polinomi interpolador és

$$p_1(x) = 0.5403023 + 0.2926526x.$$

En el cas $m = 2$, els valors de la funció en els punts $x_i = \frac{i\pi}{4}$ ($i = 0 \div 2$) són:

$$f_0 = 0.5403023, f_1 = 0.7602446, f_2 = 1.$$

El polinomi interpolador és

$$p_2(x) = 0.5403023 + 0.2674258x + 0.0160599x^2.$$

En el cas $m = 3$ els valors de la funció en els punts $x_i = \frac{i\pi}{6}$ ($i = 0 \div 3$) són:

$$f_0 = 0.5403023, f_1 = 0.6478593, f_2 = 0.8775826, f_3 = 1.$$

El polinomi interpolador és

$$p_3(x) = 0.5403023 - 0.05732768x + 0.6413111x^2 - 0.2664296x^3.$$

En el cas $m = 4$, els valors de la funció en els punts $x_i = \frac{i\pi}{8}$ ($i = 0 \div 4$) són:

$$f_0 = 0.5403023, f_1 = 0.602729, f_2 = 0.7602446,$$

$$f_3 = 0.927666, f_4 = 1.$$

El polinomi interpolador és

$$\begin{aligned} p_4(x) = & 0.5403023 - 0.02179655x + 0.525614x^2 \\ & - 0.1526548x^3 - 0.0347085x^4. \end{aligned}$$

MÈTODE DE LES DIFERÈNCIES DIVIDIDES DE NEWTON

a) Les taules de diferències dividides per a la funció $f(x) = e^x$ en l'interval $[0, 1]$ es donen a continuació per als valors de $m = 1, 2, 3, 4$, així com els polinomis interpoladors resultants:

$$\begin{array}{c|c} 0 & 1.000000 \\ & 1.718282 \\ 1 & 2.718282 \end{array}$$

$$p_1(x) = 1 + 1.718282x .$$

$$\begin{array}{c|c} 0 & 1.000000 \\ & 1.297443 \\ 0.5 & 1.648721 & 0.841679 \\ & 2.139121 \\ 1 & 2.718282 \end{array}$$

$$p_2(x) = 1 + 1.297443x + 0.841679x(x - \frac{1}{2}) .$$

$$\begin{array}{c|c} 0 & 1.000000 \\ & 1.186837 \\ 0.333333 & 1.395612 & 0.704291 \\ & 1.656365 & 0.278626 \\ 0.666666 & 1.947734 & 0.982918 \\ & 2.311643 \\ 1 & 2.718282 \end{array}$$

$$p_3(x) = 1 + 1.186837x + 0.704291x(x - \frac{1}{3}) + 0.278626x(x - \frac{1}{3})(x - \frac{2}{3}) .$$

$$\begin{array}{c|c} 0 & 1.000000 \\ & 1.136102 \\ 0.25 & 1.284025 & 0.645363 \\ & 1.458783 & 0.244400 \\ 0.5 & 1.648721 & 0.828663 & 0.069416 \\ & 1.873115 & 0.313815 \\ 0.75 & 2.117000 & 1.064025 \\ & 2.405127 \\ 1 & 2.718282 \end{array}$$

$$\begin{aligned} p_4(x) = & 1 + 1.136102x + 0.645363x(x - \frac{1}{4}) \\ & + 0.244400x(x - \frac{1}{4})(x - \frac{1}{2}) \\ & + 0.069416x(x - \frac{1}{4})(x - \frac{1}{2})(x - \frac{3}{4}) . \end{aligned}$$

b) Les taules de diferències dividides per a la funció $f(x) = \sin x$ en l'interval $[0, \frac{\pi}{2}]$ es donen a continuació per als valors de $m = 1, 2, 3, 4$, així com els polinomis interpoladors resultants:

$$\begin{array}{c|c} 0 & 0.000000 \\ & 0.636620 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$p_1(x) = 0.636620x .$$

$$\begin{array}{c|c} 0 & 0.000000 \\ & 0.900316 \\ \frac{\pi}{4} & 0.707107 & -0.335749 \\ & 0.372923 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$p_2(x) = 0.900316x - 0.335749x(x - \frac{\pi}{4}) .$$

$$\begin{array}{c|c} 0 & 0.000000 \\ & 0.954930 \\ \frac{\pi}{6} & 0.500000 & -0.244340 \\ & 0.699057 & -0.113872 \\ \frac{\pi}{3} & 0.866025 & -0.423210 \\ & 0.255873 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$p_3(x) = 0.954930x - 0.244340x(x - \frac{\pi}{6}) - 0.113872x(x - \frac{\pi}{6})(x - \frac{\pi}{3}) .$$

$$\begin{array}{c|c} 0 & 0.000000 \\ & 0.974495 \\ \frac{\pi}{8} & 0.382683 & -0.188895 \\ & 0.826137 & -0.135929 \\ \frac{\pi}{4} & 0.707107 & -0.349033 & 0.028714 \\ & 0.552007 & -0.090825 \\ \frac{3\pi}{8} & 0.923880 & -0.456034 \\ & 0.193839 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$\begin{aligned} p_4(x) = & 0.974495x - 0.188895x(x - \frac{\pi}{8}) \\ & -0.135929x(x - \frac{\pi}{8})(x - \frac{\pi}{4}) \\ & +0.028714x(x - \frac{\pi}{8})(x - \frac{\pi}{4})(x - \frac{3\pi}{8}) . \end{aligned}$$

c) Les taules de diferències dividides per a la funció $f(x) = \exp(x^2)$ en l'interval $[0, 1]$ es donen a continuació per als valors de $m = 1, 2, 3, 4$, així com els polinomis interpoladors resultants:

$$\begin{array}{c|c} 0 & 1.000000 \\ & 1.718282 \\ 1 & 2.718282 \end{array}$$

$$p_1(x) = 1 + 1.718282x .$$

$$\begin{array}{c|c} 0 & 1.000000 \\ & 0.568051 \\ 0.5 & 1.284025 & 2.300462 \\ & 2.868513 \\ 1 & 2.718282 \end{array}$$

$$p_2(x) = 1 + 0.568051x + 2.300462\left(x - \frac{1}{2}\right) .$$

$$\begin{array}{c|c} 0 & 1.000000 \\ & 0.352557 \\ 0.333333 & 1.117519 & 1.460634 \\ & 1.326313 & 1.763858 \\ 0.666666 & 1.559623 & 3.224493 \\ & 3.475975 \\ 1 & 2.718282 \end{array}$$

$$p_3(x) = 1 + 0.352557x + 1.460634x\left(x - \frac{1}{3}\right) + 1.763858x\left(x - \frac{1}{3}\right)\left(x - \frac{2}{3}\right) .$$

$$\begin{array}{c|c} 0 & 1.000000 \\ & 0.257978 \\ 0.25 & 1.064494 & 1.240292 \\ & 0.878124 & 1.028926 \\ 0.5 & 1.284025 & 2.011986 & 1.538537 \\ & 1.884117 & 2.567463 \\ 0.75 & 1.755055 & 3.937583 \\ & 3.852909 \\ 1 & 2.718282 \end{array}$$

$$\begin{aligned} p_4(x) = & 1 + 0.257978x + 1.240292x\left(x - \frac{1}{4}\right) \\ & + 1.028926x\left(x - \frac{1}{4}\right)\left(x - \frac{1}{2}\right) \\ & + 1.538537x\left(x - \frac{1}{4}\right)\left(x - \frac{1}{2}\right)\left(x - \frac{3}{4}\right) . \end{aligned}$$

d) Les taules de diferències dividides per a la funció $f(x) = \cos(\cos x)$ en l'interval $[0, \frac{\pi}{2}]$ es donen a continuació per als valors de $m = 1, 2, 3, 4$, així com els polinomis interpoladors resultants:

$$\begin{array}{c|c} 0 & 0.540302 \\ & 0.292653 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$p_1(x) = 0.540302 + 0.292653x .$$

$$\begin{array}{c|c} 0 & 0.540302 \\ & 0.280039 \\ \frac{\pi}{4} & 0.760245 & 0.016060 \\ & 0.305266 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$p_2(x) = 0.545302 + 0.280039x + 0.016060x(x - \frac{\pi}{4}) .$$

$$\begin{array}{c|c} 0 & 0.540302 \\ & 0.205419 \\ \frac{\pi}{6} & 0.647859 & 0.222804 \\ & 0.438739 & -0.266430 \\ \frac{\pi}{3} & 0.877583 & -0.195702 \\ & 0.233800 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$p_3(x) = 0.540302 + 0.205419x + 0.222804x(x - \frac{\pi}{6}) - 0.266430x(x - \frac{\pi}{6})(x - \frac{\pi}{3}) .$$

$$\begin{array}{c|c} 0 & 0.540302 \\ & 0.158968 \\ \frac{\pi}{8} & 0.602729 & 0.308304 \\ & 0.401110 & -0.234435 \\ \frac{\pi}{4} & 0.760245 & 0.032117 & -0.034708 \\ & 0.426335 & -0.288955 \\ \frac{3\pi}{8} & 0.927666 & -0.308299 \\ & 0.184197 \\ \frac{\pi}{2} & 1.000000 \end{array}$$

$$p_4(x) = 0.540302 + 0.158968x + 0.308304x(x - \frac{\pi}{8})$$

$$\begin{aligned}
& -0.234435x(x - \frac{\pi}{8})(x - \frac{\pi}{4}) \\
& -0.034708x(x - \frac{\pi}{8})(x - \frac{\pi}{4})(x - \frac{3\pi}{8}) .
\end{aligned}$$

ERRORS COMESOS

Seguidament passem a fitar els errors en tots els casos emprant la fórmula de l'error d'interpolació.

Fitem en un interval qualsevol $[a, b]$ el factor de l'error per a abscisses equidistants

$$\omega_m(x) = (x - x_0)(x - x_1) \cdots (x - x_m) .$$

El podem considerar escrit, si fem el canvi de variable $x = c + h_m u$, com

$$h_m^{m+1}(u + \frac{m}{2})(u + \frac{m}{2} - 1) \cdots (u - \frac{m}{2})$$

a l'interval $[-\frac{m}{2}, \frac{m}{2}]$, on hem fet $h_m = \frac{b-a}{m}$ ($m = 1, 2, 3, 4$) i $c = \frac{a+b}{2}$.

Definint

$$\bar{\omega}_m(u) \equiv (u + \frac{m}{2})(u + \frac{m}{2} - 1) \cdots (u - \frac{m}{2}) ,$$

el seu valor absolut, a l'interval $[-\frac{m}{2}, \frac{m}{2}]$, presenta màxims en els punts:

$$u^* = 0, \pm \frac{\sqrt{3}}{3}, \pm \frac{\sqrt{5}}{2}, \pm \sqrt{\frac{15 + \sqrt{145}}{10}} \quad (m = 1, 2, 3, 4) ;$$

els valors respectius d'aquests màxims resulten ser:

$$\Omega_m = |\bar{\omega}_m(u^*)| = \frac{1}{4}, \frac{2\sqrt{3}}{9}, 1, \frac{\sqrt{950 + 58\sqrt{145}}}{5\sqrt{5}} \quad (m = 1, 2, 3, 4) .$$

Així doncs, el factor $\omega_m(x)$ queda fitat per $h_m^{m+1}\Omega_m$; de manera que, si M_{m+1} és una fita de la funció $f^{(m+1)}$ a tot l'interval, l'error d'interpolació en qualsevol punt de l'interval esdevindrà fitat per

$$\frac{M_{m+1}}{(m+1)!} h_m^{m+1} \Omega_m .$$

a) Per a la funció $f(x) = e^x$ a l'interval $[0, 1]$ es pot prendre $M_{m+1} = e$, per a tots els valors de m , i $h_m = \frac{1}{m}$.

Les fites dels errors per als diferents valors de m seran, doncs:

$m = 1$	$\frac{e}{2}h_1^2\Omega_1 \simeq 0.341$
$m = 2$	$\frac{e}{6}h_2^3\Omega_2 \simeq 0.022$
$m = 3$	$\frac{e}{24}h_3^4\Omega_3 \simeq 0.0014$
$m = 4$	$\frac{e}{120}h_4^5\Omega_4 \simeq 0.000080$

b) Per a $f(x) = \sin x$ a $[0, \frac{\pi}{2}]$, $M_{m+1} = 1$ i $h_m = \frac{\pi}{2m}$:

$m = 1$	$\frac{1}{2}h_1^2\Omega_1 \simeq 0.31$
$m = 2$	$\frac{1}{6}h_2^3\Omega_2 \simeq 0.031$
$m = 3$	$\frac{1}{24}h_3^4\Omega_3 \simeq 0.0031$
$m = 4$	$\frac{1}{120}h_4^5\Omega_4 \simeq 0.00028$

c) Per a $f(x) = \exp(x^2)$ a $[0, 1]$, $h_m = \frac{1}{m}$ i tenim:

$$\begin{aligned}
 f'(x) &= 2xe^{x^2}, \\
 f^{(2)}(x) &= (2 + 4x^2)e^{x^2} \leq 6e \equiv M_2, \\
 f^{(3)}(x) &= (12x + 8x^3)e^{x^2} \leq 20e \equiv M_3, \\
 f^{(4)}(x) &= (12 + 48x^2 + 16x^4)e^{x^2} \leq 76e \equiv M_4, \\
 f^{(5)}(x) &= (120x + 160x^3 + 32x^5)e^{x^2} \leq 312e \equiv M_5;
 \end{aligned}$$

així,

$m = 1$	$M_2 \frac{1}{2} h_1^2 \Omega_1 \simeq 2.04$
$m = 2$	$M_3 \frac{1}{6} h_2^3 \Omega_2 \simeq 0.44$
$m = 3$	$M_4 \frac{1}{24} h_3^4 \Omega_3 \simeq 0.11$
$m = 4$	$M_5 \frac{1}{120} h_4^5 \Omega_4 \simeq 0.025$

d) Per a $f(x) = \cos(\cos x)$ a $[0, \frac{\pi}{2}]$, $h_m = \frac{\pi}{2m}$ i

$$\begin{aligned}
 f'(x) &= s(c)s, \\
 f^{(2)}(x) &= -c(c)s^2 + s(c)c, \\
 f^{(3)}(x) &= -3c(c)sc - s(c)(s^3 + s) = -3c(c)sc - s(c)s(s^2 + 1), \\
 f^{(4)}(x) &= c(c)(s^4 + 4s^2 - 3c^2) - s(c)(6s^2c + c) \\
 &= c(c)(s^4 + 7s^2 - 3) - 6s(c)(sc)s - s(c)c, \\
 f^{(5)}(x) &= c(c)(10s^3c + 15sc) + s(c)(s^5 + 10s^3 - 15sc^2 + s) \\
 &= c(c)sc(10s^2 + 15) + s(c)(s^5 + 25s^3 - 14s);
 \end{aligned}$$

on s'ha emprat $s(c) = \sin(\cos x)$, $c(c) = \cos(\cos x)$, $s = \sin x$ i $c = \cos x$.

Fitant $s(c)$ per $\sin 1$, $c(c)$ per 1 i sc per $\frac{1}{2}$, i fitant correctament els polinomis en s , podem prendre:

$$M_2 = 1 + \sin 1, \quad M_3 = \frac{3}{2} + 2 \sin 1, \quad M_4 = 5 + 4 \sin 1, \quad M_5 = \frac{25}{2} + 12 \sin 1.$$

Així,

$m = 1$	$M_2 \frac{1}{2} h_1^2 \Omega_1 \simeq 0.57$
$m = 2$	$M_3 \frac{1}{6} h_2^3 \Omega_2 \simeq 0.099$
$m = 3$	$M_4 \frac{1}{24} h_3^4 \Omega_3 \simeq 0.026$
$m = 4$	$M_5 \frac{1}{120} h_4^5 \Omega_4 \simeq 0.0064$

Problema 3.4 Disposem de la taula següent:

x	2.0	2.1	2.2	2.3	2.4	2.5
$J_0(x)$	0.2239	0.1666	0.1104	0.0555	0.0025	-0.0484

de la funció de Bessel d'ordre 0

$$J_0(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin t) dt ;$$

empreu el mètode de les diferències dividides de Newton per tal de trobar els valors de $J_0(2.15)$, $J_0(2.25)$ i $J_0(2.35)$ amb errors menors que $\frac{1}{2}10^{-3}$.

SOLUCIÓ:

Es proposa fer la següent tria de punts per als càlculs demanats: si es vol calcular $J_0(x)$ per a $x = 2.15, 2.25, 2.35$, prenem els punts consecutius de la taula $x_0, x_1, \dots, x_m$ de manera que $x_0 < x < x_1$.

L'expressió de l'error en la interpolació

$$J_0(x) - p_m(x) = \frac{J_0^{(m+1)}(\xi(x))}{(m+1)!} \omega_m(x)$$

ens permetrà trobar el valor de m per al qual es pot assegurar que aquest error sigui menor que $\frac{1}{2}10^{-3}$.

Una fita d'aquest error per a $x = 2.15, 2.25, 2.35$ es troba fent

$$\begin{aligned} |J_0(x) - p_m(x)| &\leq \frac{|J_0^{(m+1)}(\xi(x))|}{(m+1)!} |\omega_m(x)| \\ &\leq \frac{M_{m+1}}{(m+1)!} h^{m+1} \frac{1}{2} \frac{1}{2} \frac{3}{2} \dots \frac{2m-1}{2} . \end{aligned}$$

Troblem tot seguit fites per a les derivades de la funció J_0 :

$$\begin{aligned} J_0'(x) &= -\frac{1}{\pi} \int_0^\pi \sin(x \sin t) \sin t dt , \\ |J_0'(x)| &\leq \frac{1}{\pi} \int_0^\pi \sin t dt = \frac{2}{\pi} ; \\ J_0^{(2)}(x) &= \frac{1}{\pi} \int_0^\pi \cos(x \sin t) \sin^2 t dt , \\ |J_0^{(2)}(x)| &\leq \frac{1}{\pi} \int_0^\pi \sin^2 t dt = \frac{1}{2} ; \\ J_0^{(3)}(x) &= -\frac{1}{\pi} \int_0^\pi \sin(x \sin t) \sin^3 t dt , \\ |J_0^{(3)}(x)| &\leq \frac{1}{\pi} \int_0^\pi \sin^3 t dt = \frac{4}{3\pi} . \end{aligned}$$

Així, prenent $M_2 = \frac{1}{2}$ i $M_3 = \frac{4}{3\pi}$, tenim:

$$\begin{aligned} |J_0(x) - p_1(x)| &\leq \frac{1}{2 \cdot 2} (0.1)^2 \frac{1}{4} = \frac{1}{1600} , \\ |J_0(x) - p_2(x)| &\leq \frac{4}{3\pi \cdot 6} (0.1)^3 \frac{3}{8} = \frac{1}{12\pi} 10^{-3} < \frac{1}{2} 10^{-3} ; \end{aligned}$$

per tant, no sembla suficient de fer una interpolació lineal. En canvi, una interpolació quadràtica produeix un error sensiblement inferior a la fita demanada.

Construïm la taula de diferències dividides fins a diferències de segon ordre usant totes les abscisses llevat de la primera, que no es necessita,

2.1	0.1666	
	-0.562	
2.2	0.1104	0.065
	-0.549	
2.3	0.0555	0.095
	-0.530	
2.4	0.0025	0.105
	-0.509	
2.5	-0.0484	

Les aproximacions demanades, trobades per interpolació de segon grau, segons s'ha indicat, són:

$$\begin{aligned} J_0(2.15) &\simeq 0.1666 - 0.562(2.15 - 2.1) + 0.065(2.15 - 2.1)(2.15 - 2.2) \\ &= 0.1383 , \\ J_0(2.25) &\simeq 0.1104 - 0.549(2.25 - 2.2) + 0.095(2.25 - 2.2)(2.25 - 2.3) \\ &= 0.0827 , \\ J_0(2.35) &\simeq 0.0555 - 0.530(2.35 - 2.3) + 0.105(2.35 - 2.3)(2.35 - 2.4) \\ &= 0.0287 . \end{aligned}$$

Problema 3.5 Disposem d'una gràfica de $y = f(x)$. Una taula digitalitzadora dona les coordenades dels punts de la corba amb un error absolut (en x i y) menor que 0.02 dm. Cinc punts d'aquella gràfica donats per la taula són:

$x_k(\text{en dm})$	0.20	0.50	1.00	1.20	1.80
$y_k(\text{en dm})$	0.80	0.80	0.60	0.40	0.60

Trobeu una fita de l'error absolut del valor de y corresponent al punt $x = 1.50$ donat pel polinomi interpolador als punts de la gràfica, degut únicament als errors de les mesures.

SOLUCIÓ:

Per tal de trobar la fita demanada farem un estudi aproximat de la propagació dels errors de les dades en el valor del polinomi interpolador en el punt $x = 1.50$, suposat exacte, i sense comptar els errors en les operacions.

Es tria la fórmula d'interpolació de Lagrange per al seu càlcul,

$$p(x) = \sum_{i=0}^4 y_i l_i(x) = \sum_{i=0}^4 y_i \prod_{k \neq i} \frac{x - x_k}{x_i - x_k}.$$

Emprant les fórmules aproximades per a l'error en sumes, productes i divisions, trobarem una fita per a l'error absolut en $p(x)$ en funció de la fita de l'error absolut en la mesura de les coordenades dels punts $\epsilon = 0.02\text{dm}$:

$$\begin{aligned} \epsilon_a(p(x)) &= \sum_{i=0}^4 |l_i(x)| \epsilon_a(y_i) + \sum_{i=0}^4 |y_i| \epsilon_a\left(\prod_{k \neq i} \frac{x - x_k}{x_i - x_k}\right) \\ &= \sum_{i=0}^4 |l_i(x)| \epsilon_a(y_i) \\ &\quad + \sum_{i=0}^4 |l_i(x)| |y_i| \sum_{k \neq i} \left(\frac{\epsilon_a(x_k)}{|x - x_k|} + \frac{\epsilon_a(x_i)}{|x_i - x_k|} + \frac{\epsilon_a(x_k)}{|x_i - x_k|} \right) \\ &= \sum_{i=0}^4 |l_i(x)| \left[1 + |y_i| \sum_{k \neq i} \left(\frac{1}{|x - x_k|} + \frac{2}{|x_i - x_k|} \right) \right] \epsilon \end{aligned}$$

Substituint les coordenades pels valors de la taula, tenim

$$\begin{aligned} \epsilon_a(p(x)) &= 0.02 \cdot \\ &\quad \left[\frac{1.0 \cdot 0.5 \cdot 0.3 \cdot 0.3}{0.3 \cdot 0.8 \cdot 1.0 \cdot 1.6} \left[1 + 0.8 \left(\frac{1}{1.0} + \frac{1}{0.5} + \frac{1}{0.3} + \frac{1}{0.3} + \frac{2}{0.3} + \frac{2}{0.8} + \frac{2}{1.0} + \frac{2}{1.6} \right) \right] \right. \\ &\quad + \frac{1.3 \cdot 0.5 \cdot 0.3 \cdot 0.3}{0.3 \cdot 0.5 \cdot 0.7 \cdot 1.3} \left[1 + 0.8 \left(\frac{1}{1.3} + \frac{1}{0.5} + \frac{1}{0.3} + \frac{1}{0.3} + \frac{2}{0.3} + \frac{2}{0.5} + \frac{2}{0.7} + \frac{2}{1.3} \right) \right] \\ &\quad + \frac{1.3 \cdot 1.0 \cdot 0.3 \cdot 0.3}{0.8 \cdot 0.5 \cdot 0.2 \cdot 0.8} \left[1 + 0.6 \left(\frac{1}{1.3} + \frac{1}{1.0} + \frac{1}{0.3} + \frac{1}{0.3} + \frac{2}{0.8} + \frac{2}{0.5} + \frac{2}{0.2} + \frac{2}{0.8} \right) \right] \\ &\quad \left. + \frac{1.3 \cdot 1.0 \cdot 0.5 \cdot 0.3}{1.0 \cdot 0.7 \cdot 0.2 \cdot 0.6} \left[1 + 0.4 \left(\frac{1}{1.3} + \frac{1}{1.0} + \frac{1}{0.5} + \frac{1}{0.3} + \frac{2}{1.0} + \frac{2}{0.7} + \frac{2}{0.2} + \frac{2}{0.6} \right) \right] \right] . \end{aligned}$$

Un cop realitzats els càlculs resulta

$$\boxed{\epsilon_a(p(x)) \simeq 71\epsilon \simeq 0.35},$$

un error extremament gran, si observem que el valor aproximat trobat per a $p(1.50)$ és de 0.198.

Aquest error tan gran és degut principalment a l'error en les abscisses. Si aquestes fossin exactes, es pot comprovar fàcilment que

$$\epsilon_a(p(1.5)) \leq \epsilon \sum_{k=0}^4 |l_k(1.5)| = 5\epsilon = 2.5 \cdot 10^{-2},$$

apreciablement menor que el resultat obtingut anteriorment.

Problema 3.6 *Volem tabular la funció error*

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

en punts equidistants a l'interval $[0, 1]$ amb pas $h = 0.01$. Trobeu el nombre màxim t de xifres decimals arrodonides amb què s'han de donar els valors de la taula per tal que l'error degut a la interpolació per polinomis de graus 1, 3 i 5 (fent servir els punts més propers) no excedeixi l'error propagat en el càlcul del polinomi interpolador degut a l'error d'arrodoniment de les dades.

SOLUCIÓ:

Fitarem primer les derivades successives de la funció $f(x) = \operatorname{erf}(x)$ per tal de poder fitar després les expressions per als errors d'interpolació per a $x \in [0, 1]$:

$$\begin{aligned} |f(x) - p_1(x)| &\leq \frac{M_2}{2} (0.01)^2 \Omega_1, \\ |f(x) - p_3(x)| &\leq \frac{M_4}{24} (0.01)^4 \Omega_3, \\ |f(x) - p_5(x)| &\leq \frac{M_6}{720} (0.01)^6 \Omega_5; \end{aligned}$$

on M_{m+1} és una fita del valor absolut de $f^{(m+1)}$ a $[0, 1]$ i els Ω_m ($m = 1, 3, 5$) estan definits de forma anàloga als del problema III.3; però aquí, el màxim de ω_m se cerca en el subinterval central $[-\frac{1}{2}, \frac{1}{2}]$ en la variable u , allí definida.

En el cas d'interpolació de grau $m = 2s + 1$ a f en x , triarem les abscisses $x_0, \dots, x_{2s+1}$ de forma que x estigui en el subinterval central x_s, x_{s+1} de centre $c = \frac{x_s + x_{s+1}}{2}$. La variable u compleix, doncs, $x - c = hu$, amb $h = 0.01$.

Atesa la simetria, el màxim s'assoleix ara en el punt central $u = 0$ corresponent a $x = c$. Resulten així: $\Omega_1 = \frac{1}{4}$, $\Omega_3 = \frac{9}{16}$ i $\Omega_5 = \frac{225}{64}$.

Les primeres derivades de $f(x)$ són:

$$\begin{aligned} f'(x) &= \frac{2}{\sqrt{\pi}} e^{-x^2}, \\ f^{(2)}(x) &= -\frac{4}{\sqrt{\pi}} x e^{-x^2}, \\ f^{(3)}(x) &= \frac{2}{\sqrt{\pi}} (-2 + 4x^2) e^{-x^2}, \\ f^{(4)}(x) &= -\frac{2}{\sqrt{\pi}} (-12x + 8x^3) e^{-x^2}, \end{aligned}$$

$$\begin{aligned} f^{(5)}(x) &= \frac{2}{\sqrt{\pi}}(12 - 48x^2 + 16x^4)e^{-x^2}, \\ f^{(6)}(x) &= -\frac{2}{\sqrt{\pi}}(120x - 160x^3 + 32x^5)e^{-x^2}. \end{aligned}$$

Els polinomis $H_j(x)$ ($j \geq 0$) complint

$$f^{(j+1)}(x) = (-1)^j \frac{2}{\sqrt{\pi}} H_j(x) e^{-x^2},$$

o equivalentment

$$H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} e^{-x^2},$$

s'anomenen polinomis d'Hermite i són tractats en els problemes proposats.

Així les fites M_2 , M_4 i M_6 de les derivades segona, quarta i sisena, requerides en el càlcul, es poden trobar cercant els màxims i mínims locals d'aquestes funcions derivades en l'interval $[0, 1]$; aquests es donen en els zeros de les funcions derivades d'aquestes i, per tant, en els zeros de $H_2(x)$, $H_4(x)$ i $H_6(x)$, que són respectivament $\frac{\sqrt{2}}{2} = 0.7071\dots$, $\sqrt{\frac{3-\sqrt{6}}{2}} = 0.5246\dots$ i $0.4361\dots$

Les fites resultants per als errors e_m^I deguts a la interpolació seran:

$$\begin{aligned} \epsilon_1^I &\simeq \frac{|f^{(2)}(0.7071)|}{2} 10^{-4} \frac{1}{4} \simeq 0.12 \cdot 10^{-4}, \\ \epsilon_3^I &\simeq \frac{|f^{(4)}(0.5246)|}{24} 10^{-8} \frac{9}{16} \simeq 0.62 \cdot 10^{-9}, \\ \epsilon_5^I &\simeq \frac{|f^{(6)}(0.4361)|}{720} 10^{-12} \frac{225}{64} \simeq 0.19 \cdot 10^{-12}. \end{aligned}$$

En el cas d'interpolació lineal, l'error e_1^A propagat en

$$p_1(x) = f(x_0)l_0(x) + f(x_1)l_1(x)$$

serà fitat per

$$|l_0(x)|\epsilon + |l_1(x)|\epsilon \leq 2\epsilon = \epsilon_1^A,$$

si les dades tenen un error absolut fitat per ϵ , d'acord amb el problema III.5.

En el cas d'interpolació de grau $m = 3, 5$, l'error propagat e_m^A en

$$p_m(x) = \sum_{j=0}^m f(x_j)l_j(x),$$

degut a l'error d'arrodoniment en $f(x_k)$ ($k = 0 \div m$), està fitat per

$$\epsilon_m^A = \epsilon \sum_{j=0}^m |l_j(x)|,$$

que es pot demostrar que és menor que 2.5ϵ , si x pertany al subinterval central.

Sembla, doncs, adequat donar els valors de la funció amb $t = 4, 8, 12$ xifres decimals arrodonides, respectivament; llavors, una fita de l'error d'arrodoniment de les dades és $\epsilon = \frac{1}{2}10^{-t}$ i les fites ϵ_m^A no són sobrepassades per les fites ϵ_m^I trobades per a $m = 1, 3, 5$.

Problema 3.7 a) Doneu de forma explícita els polinomis de Taylor en el punt $x = 0$ a les funcions següents:

i) $\cosh x = (e^x + e^{-x})/2$, $\sinh x = (e^x - e^{-x})/2$;

ii) $(1+x)^\alpha$ ($\alpha = 2, \frac{1}{2}, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{3}$).

b) Trobeu expressions per als errors i fites d'aquestes.

c) Estudieu la convergència dels desenvolupaments de Taylor.

d) Calculeu les funcions donades en $x = 0.001$ amb un error menor que 10^{-20} .

SOLUCIÓ:

a) i) Prenem primerament $f(x) = \cosh x$. Les derivades successives d'aquesta funció són $f'(x) = \sinh x$, $f^{(2)}(x) = \cosh x$, $\dots$; és a dir, les derivades parelles són $f^{(2r)}(x) = \cosh x$ i les imparelles, $f^{(2r+1)}(x) = \sinh x$ ($r \geq 0$). En el punt $x = x_0 = 0$, tindrem, doncs,

$$f^{(2r)}(0) = 1, \quad f^{(2r+1)}(0) = 0 \quad (r \geq 0).$$

Tal cosa permet ja escriure els polinomis de Taylor demanats

$$p_{2s}(x) = p_{2s+1}(x) = 1 + \frac{1}{2}x^2 + \frac{1}{24}x^4 + \dots + \frac{x^{2s}}{(2s)!}.$$

Per a $f(x) = \sinh x$ és fàcil observar que els polinomis de Taylor són

$$p_{2s+1}(x) = p_{2s+2}(x) = x + \frac{1}{6}x^3 + \dots + \frac{x^{2s+1}}{(2s+1)!} \quad (s \geq 0).$$

ii) Ara $f(x) = (1+x)^\alpha$, la derivada j -ena és

$$f^j(x) = \alpha(\alpha-1)\dots(\alpha-j+1)(1+x)^{\alpha-j}$$

i els coeficients del polinomi de Taylor són

$$\frac{f^j(0)}{j!} = \frac{\alpha(\alpha-1)\dots(\alpha-j+1)}{j!} \equiv \binom{\alpha}{j}.$$

Els polinomis de Taylor resulten ser

$$p_n(x) = \sum_{j=0}^n \binom{\alpha}{j} x^j.$$

Per al cas $\alpha = 2$ tenim:

$$p_0(x) = 1, \quad p_1(x) = 1 + 2x, \quad p_n(x) = 1 + 2x + x^2 \quad (n \geq 2).$$

Per al cas $\alpha = \frac{1}{2}$, tenim que els coeficients del polinomi de Taylor són

$$\begin{aligned} \left(\frac{\frac{1}{2}}{j} \right) &= \frac{\frac{1}{2}(-\frac{1}{2}) \cdots (-\frac{2j-3}{2})}{j!} = (-1)^{j-1} \frac{1 \cdot 3 \cdot 5 \cdots (2j-3)}{2 \cdot 4 \cdot 6 \cdots (2j)} \\ &= (-1)^{j-1} \frac{(2j-3)!!}{(2j)!!} \quad (j \geq 0), \end{aligned}$$

i els polinomis de Taylor són

$$p_n(x) = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \cdots + (-1)^{n-1} \frac{(2n-3)!!}{(2n)!!} x^n.$$

Per al cas $\alpha = -\frac{1}{2}$, tenim que els coeficients del polinomi de Taylor són

$$\begin{aligned} \left(-\frac{\frac{1}{2}}{j} \right) &= \frac{(-\frac{1}{2})(-\frac{3}{2}) \cdots (-\frac{2j-1}{2})}{j!} = (-1)^j \frac{1 \cdot 3 \cdot 5 \cdots (2j-1)}{2 \cdot 4 \cdot 6 \cdots (2j)} \\ &= (-1)^j \frac{(2j-1)!!}{(2j)!!} \quad (j \geq 0), \end{aligned}$$

i els polinomis de Taylor són

$$p_n(x) = 1 - \frac{1}{2}x + \frac{3}{8}x^2 - \frac{5}{16}x^3 + \frac{35}{128}x^4 + \cdots + (-1)^n \frac{(2n-1)!!}{(2n)!!} x^n.$$

Per al cas $\alpha = \frac{1}{3}$, tenim que els coeficients del polinomi de Taylor són

$$\begin{aligned} \left(\frac{\frac{1}{3}}{j} \right) &= \frac{\frac{1}{3}(-\frac{2}{3}) \cdots (-\frac{3j-4}{3})}{j!} = (-1)^{j-1} \frac{2 \cdot 5 \cdot 8 \cdots (3j-4)}{3 \cdot 6 \cdot 9 \cdots (3j)} \\ &= (-1)^{j-1} \frac{(3j-4)!!!}{(3j)!!!} \quad (j \geq 0), \end{aligned}$$

i els polinomis de Taylor són

$$p_n(x) = 1 + \frac{1}{3}x - \frac{1}{9}x^2 + \frac{5}{81}x^3 - \frac{10}{243}x^4 + \cdots + (-1)^{n-1} \frac{(3n-4)!!!}{(3n)!!!} x^n.$$

Per al cas $\alpha = -\frac{1}{3}$, tenim que els coeficients del polinomi de Taylor són

$$\begin{aligned} \left(-\frac{\frac{1}{3}}{j} \right) &= \frac{(-\frac{1}{3})(-\frac{4}{3}) \cdots (-\frac{3j-2}{3})}{j!} = (-1)^j \frac{1 \cdot 4 \cdot 7 \cdots (3j-2)}{3 \cdot 6 \cdot 9 \cdots (3j)} \\ &= (-1)^j \frac{(3j-2)!!!}{(3j)!!!} \quad (j \geq 0), \end{aligned}$$

i els polinomis de Taylor són

$$p_n(x) = 1 - \frac{1}{3}x + \frac{2}{9}x^2 - \frac{14}{81}x^3 + \frac{35}{243}x^4 + \cdots + (-1)^n \frac{(3n-2)!!!}{(3n)!!!} x^n.$$

b) i) L'expressió de Lagrange per als errors dels polinomis de Taylor

$$R_n(x) = f(x) - p_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} x^{n+1}$$

permet trobar les fites

$$|R_n(x)| \leq \frac{M_{n+1}}{(n+1)!} x^{n+1},$$

essent M_{n+1} una fita del valor absolut de la derivada $(n+1)$ -ena de f en l'interval $< 0, x >$.

En el cas que $f(x) = \cosh x$, prenent $n = 2s + 1$, als polinomis de Taylor $p_{2s}(x) = p_{2s+1}(x)$ els correspon l'expressió següent per a l'error:

$$R_{2s}(x) = R_{2s+1}(x) = \frac{\cosh \xi(x)}{(2s+2)!} x^{2s+2}$$

i la fita

$$\frac{\cosh x}{(2s+2)!} x^{2s+2};$$

atès que les seves derivades parelles coincideixen amb la funció, que és parella i creixent per a valors positius de x i, per tant, pot fitar-se per $\cosh x$.

En l'altre cas, $f(x) = \sinh x$, tenim anàlogament

$$R_{2s+1}(x) = R_{2s+2}(x) = \frac{\cosh \xi(x)}{(2s+3)!} x^{2s+3}$$

que pot fitar-se, en valor absolut, per

$$\frac{\cosh x}{(2s+3)!} x^{2s+3}.$$

ii) Per a $f(x) = (1+x)^\alpha$, l'expressió de Lagrange per als errors

$$R_n(x) = \frac{1}{(1+\xi(x))^{n+1-\alpha}} \binom{\alpha}{n+1} x^{n+1}$$

pot fitar-se per

$$\binom{\alpha}{n+1} x^{n+1}$$

per a valors de $x \geq 0$ i per

$$\frac{1}{(1-|x|)^{n+1-\alpha}} \binom{\alpha}{n+1} x^{n+1}$$

per a valors de x a $(-1, 0)$. Aquesta darrera expressió esdevé arbitràriament gran a mesura que x tendeix a -1 per a cada n fix.

c) i) La convergència dels polinomis de Taylor a la funció està assegurada aquí en ambdós casos, per a qualsevol valor de x : és fàcil comprovar que les fites trobades per als errors tendeixen a zero quan s tendeix a infinit.

ii) La convergència en aquest cas és òbvia quan α és natural, atès que es tracta d'una suma finita. Si $x \in (0, 1)$, és també fàcil demostrar que la fita de l'error trobada tendeix a zero, quan n tendeix a infinit; però no es pot assolir el mateix resultat per a tots els valors a $(-1, 0)$ (encara que es pugui fer de manera anàloga a l'interval $(-\frac{1}{2}, 0)$), així doncs, cal seguir un altre camí per decidir sobre la convergència.

La fórmula integral és millor en aquest cas; prenent $y = -x$, tenim

$$\begin{aligned} R_n(x) &= (n+1) \binom{\alpha}{n+1} \int_0^{-y} \left(\frac{-y-t}{1+t} \right)^n (1+t)^{\alpha-1} dt \\ &= (-1)^{n+1} (n+1) \binom{\alpha}{n+1} \int_0^y u^n \left(\frac{1-y}{1-u} \right)^\alpha \frac{du}{1-u} \\ &= (-1)^{n+1} (n+1) \binom{\alpha}{n+1} \frac{(1-y)^\alpha}{(1-\eta)^{\alpha+1}} \int_0^y u^n du \\ &= \binom{\alpha}{n+1} \left(\frac{1+x}{1+\xi} \right)^\alpha \frac{x^{n+1}}{1+\xi}, \end{aligned}$$

on s'ha fet servir la variable $u = \frac{y+t}{1+t}$, $0 < \eta < y$, i $x < \xi = -\eta < 0$.

Clarament, la mateixa expressió és vàlida també per a $x \geq 0$, i tenim així

$$R_n(x) = \binom{\alpha}{n+1} \left(\frac{1+x}{1+\xi} \right)^\alpha \frac{x^{n+1}}{1+\xi}, \quad \xi \in (-1, x).$$

L'expressió trobada es pot fitar ara, per a $x \in (-1, 1)$,

$$\begin{aligned} |R_n(x)| &= \left| \binom{\alpha}{n+1} \right| \left| \frac{1+x}{1+\xi} \right|^\alpha \left| \frac{x^{n+1}}{1+\xi} \right| \\ &\leq \left| \binom{\alpha}{n+1} \right| \left(\frac{1+|x|}{1-|x|} \right)^\alpha \frac{|x|^{n+1}}{1-|x|}. \end{aligned}$$

D'aquesta fita, s'infereix la convergència dels polinomis de Taylor a la funció per a tot valor de $x \in (-1, 1)$.

d) i) Els càlculs de $\cosh 0.001$ i $\sinh 0.001$, aproximant amb polinomis de Taylor adequats, donen:

$$\begin{aligned} \cosh 0.001 &= 1 + \frac{10^{-6}}{2} + \frac{10^{-12}}{24} \pm \cosh 0.001 \frac{10^{-18}}{720} \\ &= 1.000000500000004166667 \pm 0.5 \cdot 10^{-20}, \end{aligned}$$

$$\begin{aligned}\sinh 0.001 &= 10^{-3} + \frac{10^{-9}}{6} + \frac{10^{-15}}{120} \pm \cosh 0.001 \frac{10^{-21}}{5040} \\ &= 0.001000000166666675000000 \pm 0.2 \cdot 10^{-24} .\end{aligned}$$

ii) El càlcul de $(1+x)^\alpha$ es porta a terme de la mateixa manera per als diferents valors de α demanats. Mostrem aquí només els càlculs per al cas $\alpha = \frac{1}{2}$:

$$\begin{aligned}(1+0.001)^{\frac{1}{2}} &= 1 + \frac{10^{-3}}{2} - \frac{10^{-6}}{8} + \frac{10^{-9}}{16} - \frac{5 \cdot 10^{-12}}{128} \\ &\quad + \frac{7 \cdot 10^{-15}}{256} - \frac{21 \cdot 10^{-18}}{1024} \pm \frac{33 \cdot 10^{-21}}{2048} \\ &= 1.0005001250625390898642 \pm 10^{-22} .\end{aligned}$$

Problema 3.8 Calculeu $e^{0.1}$, $\ln 1.1$, $1.1^{-3/2}$ i $\sin 0.1$ amb error absolut fitat per 10^{-10} , emprant desenvolupaments de Taylor en $x_0 = 0$.

SOLUCIÓ:

Partirem del coneixement dels desenvolupaments de Taylor de les funcions e^x , $\ln(1+x)$, $(1+x)^{-\frac{3}{2}}$ i $\sin x$ donats a la taula III.5. Els avaluarem en $x = 0.1$ per tal de trobar els valors demanats, tenint cura d'efectuar les operacions amb prou decimals (per exemple, 12), i fitarem l'error comès:

$$\begin{aligned}e^{0.1} &= 1 + 10^{-1} + \frac{10^{-2}}{2} + \frac{10^{-3}}{6} + \frac{10^{-4}}{24} + \frac{10^{-5}}{120} + \frac{10^{-6}}{720} \\ &\quad \pm e^{0.1} \frac{10^{-7}}{5040} \\ &= 1.10517091805 \pm 2 \cdot 10^{-11} .\end{aligned}$$

$$\begin{aligned}\ln 1.1 &= 10^{-1} - \frac{10^{-2}}{2} + \frac{10^{-3}}{3} - \frac{10^{-4}}{4} + \frac{10^{-5}}{5} - \frac{10^{-6}}{6} \\ &\quad + \frac{10^{-7}}{7} - \frac{10^{-8}}{8} + \frac{10^{-9}}{9} \pm \frac{10^{-10}}{10} \\ &= 0.09531017989 \pm 10^{-11} .\end{aligned}$$

$$\begin{aligned}1.1^{-\frac{3}{2}} &= 1 - \frac{3}{2}10^{-1} + \frac{3 \cdot 5}{2 \cdot 4}10^{-2} - \frac{3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6}10^{-3} \\ &\quad + \frac{3 \cdot 5 \cdot 7 \cdot 9}{2 \cdot 4 \cdot 6 \cdot 8}10^{-4} + \dots\end{aligned}$$

$$\begin{aligned}
& + \frac{3 \cdot 5 \cdots 21}{2 \cdot 4 \cdots 20} 10^{-10} \pm \frac{3 \cdot 5 \cdots 23}{2 \cdot 4 \cdots 22} 10^{-11} \\
= & 1 - \frac{3 \cdot 10^{-1}}{2} + \frac{15 \cdot 10^{-2}}{8} - \frac{35 \cdot 10^{-3}}{16} + \frac{315 \cdot 10^{-4}}{128} \\
& - \frac{693 \cdot 10^{-5}}{256} + \frac{3003 \cdot 10^{-6}}{1024} - \frac{6435 \cdot 10^{-7}}{2048} \\
& + \frac{109395 \cdot 10^{-8}}{32768} - \frac{230945 \cdot 10^{-9}}{65536} \\
& + \frac{969969 \cdot 10^{-10}}{262144} \pm \frac{2028117 \cdot 10^{-11}}{524288} \\
= & 0.86678417276 \pm 4 \cdot 10^{-11} . \\
\\
\sin 0.1 = & 10^{-1} - \frac{10^{-3}}{6} + \frac{10^{-5}}{120} \pm \frac{10^{-7}}{5040} \\
= & 0.099833416667 \pm 2 \cdot 10^{-11} .
\end{aligned}$$

Problema 3.9 Trobeu, com a mínim, els cinc primers termes no nuls dels desenvolupaments de Taylor en $x_0 = 0$ de les funcions següents:

- a) e^{3x} , $e^x \sin x^2$, $x^4(1+x^2)^{\frac{1}{2}}$, $\tan(3x)$, $\sin^3 x$;
- b) $\frac{\sin x}{x}$, $\frac{\cos x - 1}{x^2}$, $\frac{\sin x - x - x^3/6}{x^5}$, $\ln \left(\frac{1+x}{1-x} \right)^{\frac{1}{x}}$;
- c) $\frac{x^5}{\sin^3 x}$, $\frac{x(\cos x - 1)}{\sin x - \sin^3 x}$, $\frac{(1+x)^{\frac{1}{3}} - (1-x)^{\frac{1}{3}}}{(1+x)^{\frac{1}{3}} + (1-x)^{\frac{1}{3}}}$;
- d) $e^{\sin x}$, $e^{\sqrt{\cos x}}$, $\ln(\cos x)$.

SOLUCIÓ:

a) Emprant el desenvolupament de l'exponencial, tenim

$$\begin{aligned} e^{3x} &= 1 + 3x + \frac{3^2 x^2}{2!} + \frac{3^3 x^3}{3!} + \frac{3^4 x^4}{4!} + \frac{3^5 x^5}{5!} + \dots \\ &= 1 + 3x + \frac{9}{2}x^2 + \frac{9}{2}x^3 + \frac{27}{8}x^4 + \frac{81}{40}x^5 + \dots \end{aligned}$$

Fent el producte dels desenvolupaments d' e^x i de $\sin x^2$, trobem

$$\begin{aligned} e^x \sin x^2 &= (1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots)(x^2 - \frac{x^6}{3!} + \dots) \\ &= x^2 + x^3 + \frac{x^4}{2} + \frac{x^5}{6} - \frac{x^6}{8} + \dots \end{aligned}$$

De manera similar es troba

$$\begin{aligned} x^4(1+x^2)^{\frac{1}{2}} &= x^4(1 + \frac{1}{2}x^2 - \frac{1}{2 \cdot 4}x^4 + \frac{1 \cdot 3}{2 \cdot 4 \cdot 6}x^6 \\ &\quad - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6 \cdot 8}x^8 + \frac{1 \cdot 3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6 \cdot 8 \cdot 10}x^{10} - \dots) \\ &= x^4 + \frac{1}{2}x^6 - \frac{1}{8}x^8 + \frac{1}{16}x^{10} \\ &\quad - \frac{5}{128}x^{12} + \frac{7}{256}x^{14} - \dots \end{aligned}$$

Per tal de trobar el desenvolupament de $\tan 3x$ trobem primerament el de $\tan x$, partint del coneixement dels de $\sin x$ i de $\cos x$. Notem primerament que la funció tangent és imparella i, per tant, el seu desenvolupament està format només per monomis amb exponent imparell

$$\tan x = a_1x + a_3x^3 + a_5x^5 + a_7x^7 + a_9x^9 + a_{11}x^{11} + \dots$$

El desenvolupament donat compleix la relació

$$\begin{aligned} (1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \frac{x^{10}}{10!} + \dots) \tan x \\ = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \dots \end{aligned}$$

i, per tant, els coeficients del desenvolupament satisfan el sistema d'equacions lineals següent:

$$\begin{aligned} 1 &= a_1 \\ -\frac{1}{3!} &= a_3 - \frac{a_1}{2!} \\ \frac{1}{5!} &= a_5 - \frac{a_3}{2!} + \frac{a_1}{4!} \end{aligned}$$

$$\begin{aligned}
-\frac{1}{7!} &= a_7 - \frac{a_5}{2!} + \frac{a_3}{4!} - \frac{a_1}{6!} \\
\frac{1}{9!} &= a_9 - \frac{a_7}{2!} + \frac{a_5}{4!} - \frac{a_3}{6!} + \frac{a_1}{8!} \\
-\frac{1}{11!} &= a_{11} - \frac{a_9}{2!} + \frac{a_7}{4!} - \frac{a_5}{6!} + \frac{a_3}{8!} - \frac{a_1}{10!} \\
&\dots
\end{aligned}$$

que es pot escriure:

$$\begin{aligned}
a_1 &= 1 \\
6a_3 &= 3a_1 - 1 \\
120a_5 &= 60a_3 - 5a_1 + 1 \\
5040a_7 &= 2520a_5 - 210a_3 + 7a_1 - 1 \\
362880a_9 &= 181440a_7 - 15120a_5 + 504a_3 - 9a_1 + 1 \\
39916800a_{11} &= 19958400a_9 - 1663200a_7 + 55440a_5 \\
&\quad - 990a_3 + 11a_1 - 1 \\
&\dots
\end{aligned}$$

Així tenim el desenvolupament

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \frac{62}{2835}x^9 + \frac{1382}{155925}x^{11} + \dots$$

El desenvolupament demanat és, doncs,

$$\begin{aligned}
\tan 3x &= 3x + \frac{1}{3}3^3x^3 + \frac{2}{15}3^5x^5 + \frac{17}{315}3^7x^7 \\
&\quad + \frac{62}{2835}3^9x^9 + \frac{1382}{155925}3^{11}x^{11} + \dots \\
&= 3x + 9x^3 + \frac{162}{5}x^5 + \frac{4131}{35}x^7 \\
&\quad + \frac{15066}{35}x^9 + \frac{3022434}{1925}x^{11} + \dots
\end{aligned}$$

Per al càlcul del desenvolupament de $f(x) = \sin^3 x$, usarem una fórmula trigonomètrica que es dedueix de la fórmula de Moivre

$$\begin{aligned}
\sin^3 x &= \frac{3}{4} \sin x - \frac{1}{4} \sin 3x \\
&= \frac{3}{4} \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \dots \right) \\
&\quad - \frac{1}{4} \left(3x - \frac{3^3x^3}{3!} + \frac{3^5x^5}{5!} - \frac{3^7x^7}{7!} + \frac{3^9x^9}{9!} - \frac{3^{11}x^{11}}{11!} + \dots \right) \\
&= \frac{3}{4} \frac{8}{3!} x^3 - \frac{3}{4} \frac{3^4 - 1}{5!} x^5 + \frac{3}{4} \frac{3^6 - 1}{7!} x^7 \\
&\quad - \frac{3}{4} \frac{3^8 - 1}{9!} x^9 + \frac{3}{4} \frac{3^{10} - 1}{9!} x^9 - \dots \\
&= x^3 - \frac{1}{2}x^5 + \frac{13}{120}x^7 - \frac{41}{3024}x^9 + \frac{671}{604800}x^{11} - \dots
\end{aligned}$$

b) Els tres primers desenvolupaments demanats són molt senzills de trobar si es parteix dels desenvolupaments coneguts de $\sin x$ i $\cos x$.

Així, directament:

$$\begin{aligned}\frac{\sin x}{x} &= 1 - \frac{x^2}{6} + \frac{x^4}{120} - \frac{x^6}{5040} \\ &\quad + \frac{x^8}{362880} - \frac{x^{10}}{39916800} + \dots \\ \frac{\cos x - 1}{x^2} &= -\frac{1}{2} + \frac{x^2}{24} - \frac{x^4}{720} + \frac{x^6}{40320} \\ &\quad - \frac{x^8}{3628800} + \frac{x^{10}}{479001600} - \dots \\ \frac{\sin x - x + \frac{x^3}{6}}{x^5} &= \frac{1}{120} - \frac{x^2}{5040} + \frac{x^4}{362880} - \frac{x^6}{39916800} \\ &\quad + \frac{x^8}{6227020800} - \frac{x^{10}}{1307674368000} + \dots\end{aligned}$$

L'última funció d'aquest apartat es pot escriure com

$$\frac{1}{x}[\ln(1+x) - \ln(1-x)] ;$$

la dita funció té el desenvolupament

$$\ln\left(\frac{1+x}{1-x}\right)^{\frac{1}{x}} = 2\left(1 + \frac{x^2}{3} + \frac{x^4}{5} + \frac{x^6}{7} + \frac{x^8}{9} + \frac{x^{10}}{11} + \dots\right),$$

que es dedueix directament del de $\ln(1+x)$.

c) Notem primerament que el desenvolupament cercat per a la funció $f(x) = \frac{x^5}{\sin^3 x}$ ha de ser de la forma

$$f(x) = x^2 + a_4x^4 + a_6x^6 + a_8x^8 + a_{10}x^{10} + a_{12}x^{12} + \dots ;$$

partint del coneixement del desenvolupament de $\sin^3 x$ trobat a a), establirem un sistema d'equacions lineals per als seus coeficients imposant

$$\left(x^3 - \frac{x^5}{2} + \frac{13x^7}{120} - \frac{41x^9}{3024} + \frac{671x^{11}}{604800} - \dots\right)f(x) = x^5 .$$

Resulta:

$$\begin{aligned}a_4 - \frac{1}{2} &= 0 \\ a_6 - \frac{a_4}{2} + \frac{13}{120} &= 0 \\ a_8 - \frac{a_6}{2} + \frac{13a_4}{120} - \frac{41}{3024} &= 0 \\ a_{10} - \frac{a_8}{2} + \frac{13a_6}{120} - \frac{41a_4}{3024} + \frac{671}{604800} &= 0 \\ &\dots\end{aligned}$$

El desenvolupament trobat és llavors

$$\frac{x^5}{\sin^3 x} = x^2 + \frac{x^4}{2} + \frac{17x^6}{120} + \frac{457x^8}{15120} + \frac{3287x^{10}}{604800} + \dots$$

Per a $f(x) = \frac{x(\cos x - 1)}{\sin x - \sin^3 x}$ el desenvolupament té la mateixa forma que en el cas anterior

$$f(x) = x^2 + a_4x^4 + a_6x^6 + a_8x^8 + a_{10}x^{10} + a_{12}x^{12} + \dots$$

Els coeficients ara es troben imposant

$$\frac{\sin x - \sin^3 x}{x} f(x) = \cos x - 1 ;$$

això és,

$$\begin{aligned} (1 - \frac{7x^2}{6} + \frac{61x^4}{120} - \frac{547x^6}{5040} + \frac{703x^8}{51840} - \frac{44287x^{10}}{39916800} + \dots) f(x) \\ = \frac{x^2}{2} - \frac{x^4}{4!} + \frac{x^6}{6!} - \frac{x^8}{8!} + \frac{x^{10}}{10!} + \dots \end{aligned}$$

La resolució del sistema lineal corresponent dona el desenvolupament buscat. Els detalls del procés es deixen al lector.

El desenvolupament de

$$f(x) = \frac{(1+x)^{\frac{1}{3}} - (1-x)^{\frac{1}{3}}}{(1+x)^{\frac{1}{3}} + (1-x)^{\frac{1}{3}}} ,$$

que és de la forma

$$f(x) = a_1x + a_3x^3 + a_5x^5 + a_7x^7 + a_9x^9 + a_{11}x^{11} + \dots ,$$

es trobarà partint del desenvolupament

$$\begin{aligned} (1+x)^{\frac{1}{3}} &= 1 + \frac{1}{3}x - \frac{2}{3 \cdot 6}x^2 + \frac{2 \cdot 5}{3 \cdot 6 \cdot 9}x^3 - \frac{2 \cdot 5 \cdot 8}{3 \cdot 6 \cdot 9 \cdot 12}x^4 + \dots \\ &= 1 + \frac{x}{3} - \frac{x^2}{9} + \frac{5x^3}{81} - \frac{10x^4}{243} + \frac{22x^5}{729} - \frac{154x^6}{6561} + \frac{374x^7}{19683} \\ &\quad - \frac{935x^8}{59049} + \frac{21505x^9}{1594323} - \frac{223652x^{10}}{4782969} + \frac{147407x^{11}}{14398907} - \dots \end{aligned}$$

i llavors

$$\begin{aligned} (2 - \frac{2x^2}{9} - \frac{20x^4}{243} - \frac{308x^6}{6561} - \frac{1870x^8}{59049} - \frac{111826x^{10}}{4782969} \dots) f(x) \\ = \frac{2x}{3} + \frac{10x^3}{81} + \frac{44x^5}{729} + \frac{758x^7}{19683} + \frac{43010x^9}{1594323} + \frac{294814x^{11}}{14398907} + \dots \end{aligned}$$

La resolució del sistema lineal corresponent també es deixa al lector.

d) El desenvolupament de $f(x) = e^{\sin x}$ es podria trobar per composició dels desenvolupaments del cosinus i de l'exponencial; una altra manera d'operar consisteix a derivar i trobar després el desenvolupament

$$f(x) = a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 + a_6x^6 + \dots$$

imposant la relació resultant

$$f'(x) = f(x) \cos x ,$$

i emprant el desenvolupament de $\cos x$.

El sistema lineal corresponent és:

$$\begin{aligned} a_1 &= 1 \\ 2a_2 &= a_1 \\ 3a_3 &= a_2 - \frac{1}{2} \\ 4a_4 &= a_3 - \frac{a_1}{2} \\ 5a_5 &= a_4 - \frac{a_2}{2} + \frac{1}{24} \\ 6a_6 &= a_5 - \frac{a_3}{2} + \frac{a_1}{24} \\ &\dots \end{aligned}$$

i el desenvolupament buscat

$$e^{\sin x} = 1 + x + \frac{x^2}{2} - \frac{x^4}{8} - \frac{x^5}{15} - \frac{x^6}{240} + \dots$$

De manera similar, derivant ara $g(x) = \sqrt{\cos x}$, trobem

$$g'(x) = -\frac{1}{2} \frac{\sin x}{\sqrt{\cos x}} ;$$

això és,

$$2g'(x) = -g(x) \tan x .$$

Aquesta relació permet trobar el desenvolupament de la funció exponent $g(x) = \sqrt{\cos x}$ de $f(x) = e^{\sqrt{\cos x}}$, partint del desenvolupament conegut de $\tan x$; un cop conegut aquest, el de $f(x)$ es troba imposant la relació entre les derivades $f'(x) = g'(x)f(x)$.

Derivant $f(x) = \ln(\cos x)$, tenim

$$f'(x) = -\tan x ;$$

el desenvolupament buscat es troba simplement integrant terme a terme el desenvolupament conegut de $\tan x$ i imposant $f(0) = 0$:

$$f(x) = -\frac{1}{2}x^2 - \frac{1}{12}x^4 - \frac{1}{45}x^6 - \frac{17}{2520}x^8 - \frac{62}{28350}x^{10} - \frac{691}{935550}x^{12} + \dots$$

Problema 3.10 Considerem la funció

$$f_0(n) = [(\sqrt[n]{2} - 1)n - \ln 2]n.$$

Calculeu-la directament per a $n = 10^k$ ($k = 1 \div 10$) i expliqueu els resultats i trobeu $f_0(10^{10})$ amb 15 xifres significatives.

Feu el mateix amb les funcions

$$f_1(n) = \left(1 + \frac{1}{n}\right)^n - e, \quad f_j(n) = \sqrt[j]{n + \frac{1}{n}} - \sqrt[j]{n - \frac{1}{n}} \quad (j = 2, 3, 4).$$

$$\text{Useu } e = 2.718281828459045\dots, \ln 2 = 0.6931471805599453\dots$$

SOLUCIÓ:

En la taula 3.8 es mostren els resultats obtinguts treballant en doble precisió (amb aproximadament 16 xifres significatives). S'observa que, a mesura que augmenta k , els resultats obtinguts es van distorsionant per l'efecte de cancel·lació que s'observa en l'expressió de f_0 usada; notem, per exemple, que $\sqrt[n]{2} \simeq 1$ quan $n = 10^k$ tendeix a infinit. Una forma d'evitar la cancel·lació consisteix a desenvolupar $\sqrt[n]{2}$ per a valors de n grans.

k	$f_0(10^k)$
1	0.2458744
2	0.2407825
3	0.2402820
4	0.2402321
5	0.2402271
6	0.2402265
7	0.2402252
8	0.2406785
9	0.2600646
10	0.3449082

Taula 3.8: Càlcul directe de $f_0(10^k)$ ($k = 1 \div 10$).

En efecte,

$$\sqrt[n]{2} = e^{\frac{\ln 2}{n}} = \sum_{j=0}^{\infty} \frac{\ln^j 2}{j! n^j} = 1 + \frac{\ln 2}{n} + \frac{\ln^2 2}{2n^2} + \frac{\ln^3 2}{6n^3} + \dots$$

i llavors

$$f_0(n) \simeq \frac{\ln^2 2}{2} + \frac{\ln^3 2}{6n}$$

es pot usar per al càlcul de $f_0(10^{10})$ amb 20 xifres decimals exactes, fent

$$\begin{aligned} f_0(10^{10}) &= \frac{\ln^2 2}{2} + \frac{\ln^3 2}{6} 10^{-10} \pm 10^{-22} \\ &= 0.240226506964651(1 \pm \frac{1}{2} 10^{-15}). \end{aligned}$$

La fita donada per a l'error es dedueix del fet que

$$R_3(n) \equiv \sqrt[n]{2} - \sum_{j=0}^3 \frac{\ln^j 2}{j! n^j} = e^{\xi \frac{\ln^4 2}{4! n^4}},$$

amb $0 < \xi < \frac{\ln 2}{n}$ i, per tant,

$$\bar{R}_2(n) \equiv f_0(n) - \sum_{j=2}^3 \frac{\ln^j 2}{j! n^{j-2}} = e^{\xi \frac{\ln^4 2}{4! n^2}}$$

es pot fitar per

$$|\bar{R}_2(n)| \leq e^{\frac{\ln 2}{n} \frac{\ln^4 2}{4! n^2}} < 10^{-22},$$

si $n = 10^{10}$.

Per als altres casos demanats, s'obtenen els valors de la taula 3.9.

k	$f_1(10^k)$	$f_2(10^k)$	$f_3(10^k)$	$f_4(10^k)$
1	$-1.245394 \cdot 10^{-1}$	$3.162317 \cdot 10^{-2}$	$1.436316 \cdot 10^{-2}$	$8.891592 \cdot 10^{-3}$
2	$-1.346800 \cdot 10^{-2}$	$1.000000 \cdot 10^{-3}$	$3.094393 \cdot 10^{-4}$	$1.581139 \cdot 10^{-4}$
3	$-1.357896 \cdot 10^{-3}$	$3.162278 \cdot 10^{-5}$	$6.666667 \cdot 10^{-6}$	$2.811707 \cdot 10^{-6}$
4	$-1.359016 \cdot 10^{-4}$	$1.000000 \cdot 10^{-6}$	$1.436290 \cdot 10^{-7}$	$5.000000 \cdot 10^{-8}$
5	$-1.359127 \cdot 10^{-5}$	$3.162278 \cdot 10^{-8}$	$3.094393 \cdot 10^{-9}$	$8.891398 \cdot 10^{-10}$
6	$-1.359363 \cdot 10^{-6}$	$1.000089 \cdot 10^{-9}$	$6.670222 \cdot 10^{-11}$	$1.581279 \cdot 10^{-11}$
7	$-1.343270 \cdot 10^{-7}$	$3.370437 \cdot 10^{-11}$	$1.530873 \cdot 10^{-12}$	$2.996804 \cdot 10^{-13}$
8	$-3.011169 \cdot 10^{-8}$	$1.776534 \cdot 10^{-11}$	0	$8.882478 \cdot 10^{-14}$
9	$2.406785 \cdot 10^{-7}$	0	0	0
10	$-1.103034 \cdot 10^{-7}$	0	0	0

Taula 3.9: Càlcul directe de $f_j(10^k)$ ($j = 1 \div 4$) ($k = 1 \div 10$).

L'ús de desenvolupaments permetrà també fer els càlculs evitant l'efecte causat per les cancel·lacions produïdes i que s'observen clarament en la taula per als valors de k més grans.

Així,

$$\begin{aligned}
f_1(n) &= \left(1 + \frac{1}{n}\right)^n - e = \exp\left(n \ln\left(1 + \frac{1}{n}\right)\right) - e \\
&= \exp\left(1 - \frac{1}{2n} + \frac{1}{3n^2} - \dots\right) - e = e \exp\left(-\frac{1}{2n} + \frac{1}{3n^2} - \dots\right) - e \\
&= -\frac{e}{2n} + \frac{e}{24n^2} \dots, \\
f_1(10^{10}) &= -\frac{e}{2} 10^{-10} + \frac{e}{24} 10^{-20} \pm 10^{-30} \\
&= -1.359140914218197 \cdot 10^{-10} (1 \pm \frac{1}{2} 10^{-15}). \\
f_j(n) &= \sqrt[j]{n} \left[\left(1 + \frac{1}{n^2}\right)^{\frac{1}{j}} - \left(1 - \frac{1}{n^2}\right)^{\frac{1}{j}} \right]
\end{aligned}$$

$$\begin{aligned}
&= \sqrt[j]{n} \left[\frac{2}{jn^2} + \frac{(j-1)(2j-1)}{3j^3n^6} + \dots \right] ; \\
f_2(10^{10}) &= 10^5(10^{-20} \pm \frac{1}{2}10^{-60}) \\
&= 1.0000000000000000 \cdot 10^{-15} \pm \frac{1}{2}10^{-30} . \\
f_3(10^{10}) &= 10^{\frac{10}{3}}(\frac{2}{3}10^{-20} \pm 10^{-60}) \\
&= 1.43628979335459 \cdot 10^{-17}(1 \pm \frac{1}{2}10^{-15}) . \\
f_4(10^{10}) &= 10^{\frac{5}{2}}(\frac{1}{2}10^{-20} \pm 10^{-60}) \\
&= 1.58113883008419 \cdot 10^{-18}(1 \pm \frac{1}{2}10^{-15}) .
\end{aligned}$$

Problema 3.11 Donades les funcions $f_t(x) = \frac{xe^{tx}}{e^x-1}$, els polinomis de Bernoulli $B_j(t)$ ($j \geq 0$) vénen definits per $B_j(t) = f_t^{(j)}(0)$; així,

$$f_t(x) = B_0(t) + \frac{B_1(t)}{1!}x + \frac{B_2(t)}{2!}x^2 + \frac{B_3(t)}{3!}x^3 + \dots$$

Els valors $B_j \equiv B_j(0)$ ($j \geq 0$) reben el nom de nombres de Bernoulli i compleixen:

$$\frac{x}{e^x-1} = B_0 + \frac{B_1}{1!}x + \frac{B_2}{2!}x^2 + \frac{B_3}{3!}x^3 + \dots$$

a) Proveu la recurrència

$$B_0 = 1, \quad B_j = \sum_{l=0}^j \binom{j}{l} B_l \quad (j \geq 2) .$$

b) Proveu que

$$B_j(t+h) = \sum_{l=0}^j \binom{j}{l} B_l(t) h^{j-l} \quad (j \geq 0) ; \quad B_j(1) = B_j(0) \quad (j \geq 2) .$$

c) Establiu la recurrència

$$B_0(t) = 1, \quad B_1(t) = t - \frac{1}{2},$$

$$B'_j(t) = jB_{j-1}(t)$$

que, juntament amb $B_j(1) = B_j(0)$ ($j \geq 2$), determina completament els $B_j(t)$ ($j \geq 0$); proveu també que $B_j(t)$ és un polinomi mònic de grau j que compleix

$$(-1)^j B_j(1-t) = B_j(t) \quad (j \geq 0) .$$

d) Proveu, per a $r \geq 1$, que

$$B_{2r+1} = 0; \quad (-1)^r B_{2r-1}(t) > 0, \quad t \in (0, \frac{1}{2}) ;$$

$$(-1)^{r+1} B_{2r} > 0, \quad (-1)^r (B_{2r}(t) - B_{2r}) > 0, \quad t \in (0, 1) .$$

SOLUCIÓ:

a) De la definició dels polinomis de Bernoulli, tenim

$$e^{tx} = \frac{e^x - 1}{x} (B_0(t) + B_1(t)x + \frac{B_2(t)}{2!}x^2 \cdots) ;$$

és a dir,

$$(1 + tx + \frac{t^2 x^2}{2!} + \cdots) = (1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots) (B_0(t) + B_1(t)x + \frac{B_2(t)}{2!}x^2 + \cdots) .$$

Igualant els coeficients de x^{j-1} ($j \geq 1$), obtenim

$$\frac{t^{j-1}}{(j-1)!} = \sum_{l=0}^{j-1} \frac{1}{(j-l)!} \frac{B_l(t)}{l!} \quad (j \geq 1) ;$$

això és,

$$jt^{j-1} = \sum_{l=0}^{j-1} \binom{j}{l} B_l(t) \quad (j \geq 1) . \quad (*)$$

Fent $j = 1$, obtenim $B_0(t) = 1$, i fent $j \geq 2$ i $t = 0$, obtenim

$$\sum_{l=0}^{j-1} \binom{j}{l} B_l = 0 ,$$

on $B_l = B_l(0)$ ($l \geq 1$) i $B_0 = B_0(t) = 1$. Sumant B_j als dos membres de la igualtat obtenim la relació demanada. La dita relació se sol escriure per tal de recordar-la com $(B+1)_j = B_j$, ja que, si B és un nombre qualsevol, la fórmula del binomi de Newton dóna

$$(B+1)^j = \sum_{l=0}^j \binom{j}{l} B^l$$

i l'expressió trobada per als B_j s'obté "baixant els índexs" a aquesta.

b) Usant ara que $B_j(t+h) = f_{t+h}^{(j)}(0)$, tenim

$$\frac{xe^{(t+h)x}}{e^x - 1} = (B_0(t+h) + B_1(t+h)x + \frac{B_2(t+h)}{2!}x^2 + \dots);$$

però,

$$\begin{aligned} \frac{xe^{(t+h)x}}{e^x - 1} &= \frac{xe^{tx}}{e^x - 1} e^{hx} \\ &= (B_0(t) + B_1(t)x + \frac{B_2(t)}{2!}x^2 + \dots)(1 + hx + \frac{h^2}{2!}x^2 + \dots). \end{aligned}$$

Per la unicitat dels polinomis d'interpolació de Taylor d'un grau donat, podem igualar els coeficients de x^j en ambdues expressions i s'obté la recurrència. De (*) tenim $B_j(t+1) = B_j(t) + jt^{j-1}$ i, si $j \geq 2$ i $t = 0$, resulta $B_j(1) = B_j(0)$.

c) De b) es dedueix

$$\frac{B_j(t+h) - B_j(t)}{h} = jB_{j-1}(t) + h \sum_{l=0}^{j-2} B_l(t)h^{j-l-2};$$

per tant, passant al límit quan h tendeix a 0, s'obté $B'_j(t) = jB_{j-1}(t)$ ($j \geq 2$).

És clar que $B_0(t) = 1$. De la relació $B_0(t) + 2B_1(t) = 2t$ (obtinguda de (*) per a $j = 2$), s'obté ara $B_1(t) = t - \frac{1}{2}$. A més, $B_2(t) = t^2 - t + C_2$ satisfà $B_2(1) = B_2(0)$ ja que $\int_0^1 B_1(s)ds = 0$ i la constant C_2 es determina de manera que $\int_0^1 B_2(s)ds = 0$ i, per tant, $B_3(1) = B_3(0)$. És a dir, la recurrència:

$$B_0(t) = 1, \quad B_1(t) = t - \frac{1}{2};$$

$$B'_j(t) = jB_{j-1}(t), \quad B_j(1) = B_j(0) \quad (j \geq 2),$$

també determina unívocament els polinomis de Bernoulli.

Notem que $B_j(1) = B_j(0)$ ($j \geq 2$) és equivalent a $\int_0^1 B_{j-1}(t)dt = 0$, atesa la relació $B'_j(t) = jB_{j-1}(t)$, que implica immediatament per inducció el fet que $B_j(t) = t^j + \dots$, on els punts indiquen termes en t de grau inferior a j .

Finalment,

$$\begin{aligned} B_0(t) + B_1(t)x + \frac{B_2(t)}{2!}x^2 + \dots &= \frac{xe^{tx}}{e^x - 1} = \frac{(-x)e^{(1-t)(-x)}}{e^{-x} - 1} \\ &= B_0(1-t) + B_1(1-t)(-x) + \frac{B_2(1-t)}{2!}(-x)^2 + \dots; \end{aligned}$$

d'on, igualant termes, $B_j(t) = (-1)^j B_j(1-t)$ ($j \geq 0$). En particular, per tal de conèixer $B_j(t)$ en $t \in [0, 1]$, n'hi ha prou amb conèixer $B_j(t)$ per a $t \in [0, \frac{1}{2}]$: si j és senar, $B_j(t) = -B_j(1-t)$ i, si j és parell, $B_j(t) = B_j(1-t)$.

d) Si $j = 2r + 1$, $B_{2r+1}(t) = -B_{2r+1}(1-t)$ implica la relació

$$B_{2r+1}(0) = -B_{2r+1}(1) ;$$

ja sabem que es compleix $B_{2r+1}(0) = B_{2r+1}(1)$ si $r \geq 1$, aleshores

$$B_{2r+1} = B_{2r+1}(0) = B_{2r+1}(1) = 0 \quad (r \geq 1) .$$

Provarem ara, per inducció sobre $r \geq 1$:

$$(-1)^r B_{2r-1}(t) > 0, \quad t \in (0, \frac{1}{2});$$

$$(-1)^r (B_{2r}(t) - B_{2r}) > 0, \quad t \in (0, 1);$$

$$(-1)^{r+1} B_{2r} > 0.$$

Si $r = 1$ i $t \in (0, \frac{1}{2})$, $(-1)^r B_{2r-1}(t) = \frac{1}{2} - t > 0$. Suposem ara que $(-1)^r B_{2r-1}(t) > 0$ per a $t \in (0, \frac{1}{2})$, aleshores

$$(-1)^r (B_{2r}(t) - B_{2r}) = (2r-1) \int_0^t (-1)^r B_{2r-1}(s) ds > 0, \quad \text{si } t \in (0, \frac{1}{2}]$$

i, usant c), $(-1)^r (B_{2r}(t) - B_{2r}) = (-1)^r (B_{2r}(1-t) - B_{2r}) > 0$, si $t \in (\frac{1}{2}, 1)$.

Atès que $\int_0^1 B_{2r}(s) ds = 0$,

$$0 < (-1)^r \int_0^1 (B_{2r}(t) - B_{2r}) dt = (-1)^{r+1} B_{2r} .$$

Només cal provar ara $(-1)^{r+1} B_{2r+1}(t) > 0$, per a $t \in (0, \frac{1}{2})$.

Si $B_{2r+1}(t_1) = 0$ amb $t_1 \in (0, \frac{1}{2})$, voldria dir que $B_{2r+1}(t)$ s'anul·laria en $0, t_1, \frac{1}{2}$; pel teorema de Rolle, $B'_{2r+1}(t)$ s'anul·laria en t_2, t_3 complint l'ordenació $0 < t_2 < t_1 < t_3 < \frac{1}{2}$, i de nou, $B_{2r+1}^{(2)}(t)$ s'anul·laria en una abscissa $t_3 \in (0, \frac{1}{2})$; però, si $r \geq 1$,

$$B_{2r+1}^{(2)}(t) = (2r+1)B'_{2r}(t) = (2r+1)(2r)B_{2r-1}(t)$$

i, per tant, $B_{2r-1}(t_3) = 0$, $t_3 \in (0, \frac{1}{2})$, que entra en contradicció amb la suposició feta. Així doncs, el polinomi $B_{2r+1}(t)$ no s'anul·la en $(0, \frac{1}{2})$ i, com que és continu, no canvia de signe; el seu signe serà així igual al signe de $B'_{2r+1}(0) = (2r+1)B_{2r}(0)$, que és el de $(-1)^{r+1}$. En definitiva, tenim $(-1)^{r+1} B_{2r+1}(t) > 0$ per a $t \in (0, \frac{1}{2})$.

Com a il·lustració i per tal de facilitar el seu ús en el capítol següent, donem a continuació una taula dels primers nombres de Bernoulli $B_0, B_1, B_2, B_4, B_6, \dots, B_{30}$ (recordeu que, per d), $B_{2r+1} = 0$ $r \geq 1$):

$$\begin{array}{ll}
B_0 &= 1 & B_{16} &= -\frac{3617}{510} \\
B_1 &= -\frac{1}{2} & B_{18} &= \frac{43867}{798} \\
B_2 &= \frac{1}{6} & B_{20} &= -\frac{174611}{330} \\
B_4 &= -\frac{1}{30} & B_{22} &= \frac{854513}{138} \\
B_6 &= \frac{1}{42} & B_{24} &= -\frac{236364091}{2730} \\
B_8 &= -\frac{1}{30} & B_{26} &= \frac{8553103}{6} \\
B_{10} &= \frac{5}{66} & B_{28} &= -\frac{23749461029}{870} \\
B_{12} &= -\frac{691}{2730} & B_{30} &= \frac{8615841276005}{14322} \\
B_{14} &= \frac{7}{6} & & \dots
\end{array}$$

Problema 3.12 Calculeu $\tan 22.5^\circ$ fent servir interpolació d'Hermite en 0° i 45° . Fiteu l'error comès i compareu la fita trobada amb l'error exacte.

SOLUCIÓ:

Si definim $f(x) = \tan \pi x$ i treballem en radians, es demana el càlcul de $f(\frac{1}{8}) = \tan \frac{\pi}{8}$ per interpolació d'Hermite en $x_0 = 0$ i $x_1 = \frac{1}{4}$.

Usarem el mètode de les diferències dividides generalitzades repetint els punts x_0 i x_1 i aplicant la definició d'aquestes fent $f[x_0, x_0] \equiv f'(x_0)$ i $f[x_1, x_1] \equiv f'(x_1)$.

Recollim primer la informació requerida:

$$f(x_0) = 0, \quad f(x_1) = 1, \quad f'(x_0) = \pi, \quad f'(x_1) = 2\pi;$$

atès que $f'(x) = (1 + f^2(x))\pi$.

L'esquema resultant és el donat en la taula

$$\begin{array}{c|ccc}
0 & 0 & & \\
& \pi & & \\
0 & 0 & 4(4 - \pi) & \\
& 4 & & 16(3\pi - 8) \\
\frac{1}{4} & 1 & 8(\pi - 2) & \\
& 2\pi & & \\
\frac{1}{4} & 1 & &
\end{array}$$

El polinomi interpolador d'Hermite és, doncs,

$$p_3(x) = \pi x + 4(4 - \pi)x^2 + 16(3\pi - 8)x^2(x - \frac{1}{4})$$

i permet trobar l'aproximació

$$\tan(\frac{\pi}{8}) \simeq p_3(\frac{1}{8}) = 0.4018...$$

La fórmula trigonomètrica $\tan \frac{\theta}{2} = \frac{\sin \theta}{1 + \cos \theta}$ permet trobar el valor exacte

$$\tan(\frac{\pi}{8}) = \frac{\frac{1}{\sqrt{2}}}{1 + \frac{1}{\sqrt{2}}} = \frac{1}{1 + \sqrt{2}} = \sqrt{2} - 1 = 0.4142213562...$$

L'error comès es pot fitar emprant l'expressió de l'error en la interpolació d'Hermite

$$f(x) - p_3(x) = \frac{f^{(4)}(\xi)}{4!} x^2(x - \frac{1}{4})^2 \quad 0 < \xi < \frac{1}{4},$$

que en $x = \frac{1}{8}$ dóna

$$f(\frac{1}{8}) - p_3(\frac{1}{8}) = \frac{f^{(4)}(\xi)}{4!8^4};$$

donarem com a fita de l'error $\frac{M_4}{4!8^4}$, on M_4 és una fita de $f^{(4)}$ en l'interval $[0, \frac{1}{4}]$ que cerquem a continuació.

Tenim

$$f' = \pi(1 + f^2) > 0, \quad f^{(2)} = 2\pi f f' > 0, \quad f^{(3)} = 2\pi(f^{(2)}f + f'^2) > 0,$$

$$f^{(4)} = 2\pi(f^{(3)}f + 3f^{(2)}f') > 0, \quad f^{(5)} > 0;$$

així $f^{(4)}$, en ser creixent, pren el valor màxim sobre $[0, \frac{1}{4}]$ en $x = \frac{1}{4}$. Per tal de calcular-lo, trobem primer $f(\frac{1}{4}) = 1$, $f'(\frac{1}{4}) = 2\pi$, $f^{(2)}(\frac{1}{4}) = 4\pi^2$, $f^{(3)}(\frac{1}{4}) = 16\pi^3$; llavors,

$$M_4 \equiv f^{(4)}(\frac{1}{4}) = 80\pi^4$$

i la fita de l'error dóna 0.079, que és sensiblement més gran que l'error real 0.0124.

Problema 3.13 (Interpolació hermitiana generalitzada).

a) Demostreu l'existència i unicitat del polinomi $p_N(x)$ de grau menor o igual que N complint:

$$p_N^{(l)}(x_k) = f_k^{(l)} \quad (l = 0 \div n_k) \quad (k = 0 \div m),$$

amb $N = m + \sum_{k=0}^m n_k$.

b) Trobeu una expressió per als errors $f(x) - p_N(x)$ en $[a, b]$, si la funció $f \in \mathcal{C}^{N+1}(a, b)$.

c) Expliqueu amb tots els detalls com es generalitza el mètode de les diferències dividides en aquest cas.

d) Trobeu el polinomi $p_{11}(x)$ de grau 11 complint:

$$\begin{aligned} p_{11}^{(l)}(-1) &= f^{(l)}(-1) \quad (l = 0 \div 3) , \\ p_{11}^{(l)}(0) &= f^{(l)}(0) \quad (l = 0 \div 2) , \\ p_{11}^{(l)}(1) &= f^{(l)}(1) \quad (l = 0 \div 4) \end{aligned}$$

per a $f(x) = x^{12}$. Fiteu l'error d'interpolació a l'interval $[-1, 1]$.

SOLUCIÓ:

a) Les condicions d'interpolació d'Hermite generalitzada imposen un sistema d'equacions a complir pels coeficients del polinomi $p_N(x)$; l'existència i unicitat de la solució equival a l'existència i unicitat de la solució del sistema homogeni; això és, que l'únic polinomi que compleix

$$p_N^{(l)}(x_k) = 0 \quad (l = 0 \div n_k) \quad (k = 0 \div m) ,$$

sigui el polinomi 0. Però això és cert: si $p_N(x)$ fos no nul, atès que la condició $p_N^{(l)}(x_k) = 0$ ($l = 0 \div n_k$) equival al fet que $p_N(x)$ tingui el zero x_k de multiplicitat $n_k + 1$ (dit d'una altra manera, que $(x - x_k)^{n_k+1}$ sigui un factor de $p_N(x)$) ($k = 0 \div m$), $(x - x_0)^{n_0+1}(x - x_1)^{n_1+1} \dots (x - x_m)^{n_m+1}$ (de grau $N + 1$) seria un factor de $p_N(x)$ de grau màxim N ; d'on es dedueix que $p_N(x)$ ha de ser nul.

b) Per a cada $x \in (a, b)$, $x \neq x_k$ ($k = 0 \div m$), considerem la funció

$$F(z) = f(z) - p_N(z) - a(x)(z - x_0)^{n_0+1}(z - x_1)^{n_1+1} \dots (z - x_m)^{n_m+1} ,$$

on $a(x)$ es tria de forma que $F(x) = 0$. Així la funció F té els següents zeros: x, x_k (de multiplicitat $n_k + 1$) ($k = 0 \div m$); això és, $N + 1$ zeros (comptats amb llur multiplicitat). Una generalització simple del teorema de Rolle ens permet afirmar que la derivada $N + 1$ s'anul·la en un punt ξ de l'interval (a, b)

$$F^{(N+1)}(\xi) = f^{(N+1)}(\xi) - a(x)(N + 1)! = 0 ;$$

d'on, aïllant $a(x)$ i substituint-la en l'expressió inicial, tenim la fórmula de l'error en la interpolació d'Hermite generalitzada

$$f(z) - p_N(z) = \frac{f^{(N+1)}(\xi)}{(N + 1)!} (z - x_0)^{n_0+1} (z - x_1)^{n_1+1} \dots (z - x_m)^{n_m+1} .$$

c) El procés a seguir és quasi el mateix que en el cas d'interpolació usual (en què $n_k = 0$ ($k = 0 \div m$)). Recordem primerament que la interpolació de Taylor es pot entendre com una interpolació normal en què tots els punts d'interpolació s'han fet coincidir amb un

El polinomi interpolador demanat és

$$\begin{aligned} p_{11}(x) = & 1 - 12(x+1) + 66(x+1)^2 - 220(x+1)^3 \\ & + 165(x+1)^4 - 120(x+1)^4 x + 84(x+1)^4 x^2 \\ & - 34(x+1)^4 x^3 + 14(x+1)^4 x^3(x-1) \\ & - 4(x+1)^4 x^3(x-1)^2 + 4(x+1)^4 x^3(x-1)^3 \\ & + (x+1)^4 x^3(x-1)^4 . \end{aligned}$$

La fórmula de l'error trobada en b) ens diu que

$$x^{12} - p_{11}(x) = (x-1)^4 x^3 (x+1)^5 ,$$

exactament. Així, el polinomi interpolador també s'escriu

$$p_{11}(x) = x^{12} - (x-1)^4 x^3 (x+1)^5 .$$

El polinomi d'error $(x^2-1)(x+1)[(x-1)x(x+1)]^3$ es pot fitar pel producte de les fites de cada factor; això és, usant el problema III.3 sobre $[-1, 1]$, per $1 \cdot 2 \cdot \frac{2\sqrt{3}}{9} = \frac{4\sqrt{3}}{9}$. La millor fita es trobaria cercant el valor absolut màxim de l'error en els màxims i mínims relatius d'aquest.

Problema 3.14 Trobeu la constant que aproxima millor $f(x) = e^x$ en $[0, 1]$, segons les normes:

i) $\| \cdot \|_1$, ii) $\| \cdot \|_2$, iii) $\| \cdot \|_\infty$.

SOLUCIÓ:

L'aproximació per una constant és el cas més senzill d'aproximació; es pot tractar a partir de la mateixa definició del problema com la minimització de la funció

$$\Phi(c) = \min_{c \in \mathbb{R}} \|f - c\| .$$

Això és el que farem per a les tres normes suggerides.

i) La funció $\Phi_1(c) = \|f - c\|_1$ està definida per

$$\Phi_1(c) = \int_0^1 |e^x - c| dx .$$

Tenim, així,

$$\Phi_1(c) = \int_0^1 (e^x - c) dx = e - 1 - c , \text{ si } c \leq 1 ;$$

$$\begin{aligned}\Phi_1(c) &= \int_0^{\ln c} (c - e^x) dx + \int_{\ln c}^1 (e^x - c) dx = 2c \ln c - 3c + e + 1, \\ &\quad \text{si } 1 \leq c \leq e; \\ \Phi_1(c) &= \int_0^1 (c - e^x) dx = c - e + 1, \text{ si } c \geq e.\end{aligned}$$

La dita funció és decreixent en el tram $c \leq 1$ i creixent en $c \geq 1$; té, per tant, un mínim $\tilde{c}$ en $[1, e]$ que es pot trobar per anul·lació de la derivada:

$$\Phi_1'(\tilde{c}) = 2 \ln \tilde{c} - 1 = 0, \quad \tilde{c} = \sqrt{e}.$$

El dit mínim és la constant òptima en aquesta norma.

ii) Anàlogament, tenim

$$\Phi_2(c) = \int_0^1 (e^x - c)^2 dx = \frac{1}{2}(e^2 - 1) - 2(e - 1)c + c^2;$$

es tracta d'una paràbola en la variable c . El valor de $c = c^*$ en el mínim de la mateixa ens dóna la constant òptima cercada

$$c^* = e - 1,$$

que, en ser $\|\cdot\|_2$ una norma euclidiana, també s'obté de la resolució de l'equació normal associada

$$c^* = \frac{\int_0^1 1 \cdot e^x dx}{\int_0^1 1^2 dx} = e - 1.$$

iii) En aquest cas la funció que cal minimitzar és

$$\Phi_\infty(c) = \|e^x - c\|_\infty = \max\{c - 1, e - c\},$$

que resulta ser

$$\begin{aligned}\Phi_\infty(c) &= e - c, \text{ si } c \leq \frac{e+1}{2}; \\ \Phi_\infty(c) &= c - 1, \text{ si } c \geq \frac{e+1}{2}.\end{aligned}$$

La gràfica de la funció està formada per dues rectes: la primera amb pendent -1, la segona amb pendent 1 que coincideixen per a

$$c = \hat{c} = \frac{e+1}{2};$$

aquest és, doncs, el valor per al qual es dóna el mínim buscat de la funció Φ_∞ .

Problema 3.15 Trobeu la millor aproximació per mínims quadrats als punts

$$\left(-\frac{\pi}{2}, 1\right), (0, 0), \left(\frac{\pi}{2}, \frac{1}{2}\right), (\pi, 1)$$

i que sigui del tipus

$$t_1(\theta) = a + r \sin(\theta + \alpha) ,$$

amb a, r reals i $\alpha \in [0, 2\pi]$.

SOLUCIÓ:

Primer, desfem l'expressió del sinus de la suma i queda

$$t_1(\theta) = a + r \sin \alpha \cos \theta + r \cos \alpha \sin \theta ;$$

ara, anomenem $a_0 \equiv a$, $a_1 \equiv r \sin \alpha$ i $b_1 \equiv r \cos \alpha$.

Es tracta d'una aproximació discreta per mínims quadrats en les abscisses angulars $\theta_k = -\frac{\pi}{2} + kh$ ($k = 0 \div 3$) amb $h = \frac{\pi}{2}$; això és, cal trobar els coeficients a_0^* , a_1^* i b_1^* per tal que

$$\|f - t_1\|_2^2 = \sum_{k=0}^3 [f(\theta_k) - t_1(\theta_k)]^2$$

sigui mínima per a $a_0 = a_0^*$, $a_1 = a_1^*$ i $b_1 = b_1^*$.

La solució del problema és la solució de les equacions normals

$$\begin{pmatrix} (1, 1) & (1, \cos) & (1, \sin) \\ (\cos, 1) & (\cos, \cos) & (\cos, \sin) \\ (\sin, 1) & (\sin, \cos) & (\sin, \sin) \end{pmatrix} \begin{pmatrix} a_0^* \\ a_1^* \\ b_1^* \end{pmatrix} = \begin{pmatrix} (1, f) \\ (\cos, f) \\ (\sin, f) \end{pmatrix} ,$$

trobades fent servir el producte escalar associat a la norma euclidiana definida abans.

Les components diagonals de la matriu són, doncs,

$$\begin{aligned} (1, 1) &= \sum_{k=0}^3 1 \cdot 1 = 4 , \\ (\cos, \cos) &= \sum_{k=0}^3 \cos^2 \theta_k = 0 + 1 + 0 + 1 = 2 , \\ (\sin, \sin) &= \sum_{k=0}^3 \sin^2 \theta_k = 1 + 0 + 1 + 0 = 2 ; \end{aligned}$$

les components no diagonals resulten ser totes nul·les. Es tracta d'una aproximació trigonomètrica en abscisses equidistants $\theta_k = -\frac{\pi}{2} + \frac{k\pi}{2}$ ($k = 0 \div 3$).

Pel que fa a les components del terme independent, prenent

$$f_0 \equiv f(\theta_0) = 1 , \quad f_1 \equiv f(\theta_1) = 0 , \quad f_2 \equiv f(\theta_2) = \frac{1}{2} , \quad f_3 \equiv f(\theta_3) = 1 ,$$

tenim

$$\begin{aligned}(1, f) &= \sum_{k=0}^3 f_k = 1 + 0 + \frac{1}{2} + 1 = 2 , \\ (\cos, f) &= \sum_{k=0}^3 \cos \theta_k f_k = 0 + 0 + 0 - 1 = -1 , \\ (\sin, f) &= \sum_{k=0}^3 \sin \theta_k f_k = -1 + 0 + \frac{1}{2} + 0 = -\frac{1}{2} .\end{aligned}$$

La resolució de les equacions normals és immediata i dona:

$$a_0^* = \frac{1}{2} , \quad a_1^* = -\frac{1}{2} , \quad b_1^* = -\frac{1}{4} .$$

Els valors òptims, en el sentit de l'aproximació per mínims quadrats, dels coeficients a , r i α seran, doncs:

$$\begin{aligned}a^* &= a_0^* = \frac{1}{2} , \quad r^* = \sqrt{a_1^{*2} + b_1^{*2}} = \frac{\sqrt{5}}{4} = 0.559017... , \\ \alpha^* &= \arctan \frac{a_1}{b_1} + \pi = \arctan 2 + \pi = 4.24874...\end{aligned}$$

Problema 3.16 Determineu la família de polinomis ortogonals mòncics $\psi_j(x)$ ($j = 0 \div 4$) corresponent a la funció pes $w(x) = x^2 + 1$ a l'interval $[-1, 1]$.

SOLUCIÓ:

Els polinomis ortogonals respecte al producte escalar associat

$$(f, g) = \int_{-1}^1 (x^2 + 1) f(x) g(x) dx ,$$

es poden trobar fent servir la fórmula recurrent dels polinomis ortogonals amb $A_j = 1$ ($j \geq 0$):

$$\begin{aligned}\psi_{-1}(x) &= 0 , \quad \psi_0(x) = 1 , \\ \psi_{j+1}(x) &= \alpha_j(x - \beta_j)\psi_j(x) - \gamma_j\psi_{j-1}(x) \quad (j \geq 0) ,\end{aligned}$$

on

$$\begin{aligned}\alpha_j &= \frac{A_{j+1}}{A_j} = 1 , \\ \beta_j &= \frac{(\psi_j, x\psi_j)}{(\psi_j, \psi_j)} = 0 , \\ \gamma_j &= \frac{\alpha_j}{\alpha_{j-1}} \frac{(\psi_j, \psi_j)}{(\psi_{j-1}, \psi_{j-1})} = \frac{(\psi_j, \psi_j)}{(\psi_{j-1}, \psi_{j-1})} .\end{aligned}$$

L'anul·lació dels coeficients β_j ($j \geq 0$) es dedueix del fet que els polinomis ortogonals tenen paritat $\psi_j(-x) = (-1)^j \psi_j(x)$ ($j \geq 0$), atès que la funció pes $w(x) = x^2 + 1$ és parella en l'interval simètric $[-1, 1]$.

La fórmula recurrent queda, així, reduïda a la forma

$$\psi_{j+1}(x) = x\psi_j(x) - \gamma_j\psi_{j-1}(x) \quad (j \geq 0) .$$

Efectuant els càlculs corresponents i notant

$$I_{2r} \equiv \int_{-1}^1 x^{2r} dx = \frac{2}{2r+1} \quad (r \geq 0) ,$$

tenim

$$\begin{aligned} \psi_1(x) &= x\psi_0(x) = x ; \\ \gamma_1 &= \frac{(x\psi_1, \psi_0)}{(\psi_0, \psi_0)} = \frac{I_4 + I_2}{I_2 + I_0} = \frac{2}{5} , \\ \psi_2(x) &= x\psi_1(x) - \gamma_1\psi_0(x) = x^2 - \frac{2}{5} ; \\ \gamma_2 &= \frac{(x\psi_2, \psi_1)}{(\psi_1, \psi_1)} = \frac{I_6 + \frac{3}{5}I_4 - \frac{2}{5}I_2}{I_4 + I_2} = \frac{17}{70} , \\ \psi_3(x) &= x\psi_2(x) - \gamma_2\psi_1(x) = x^3 - \frac{9}{14}x ; \\ \gamma_3 &= \frac{(x\psi_3, \psi_2)}{(\psi_2, \psi_2)} = \frac{I_8 - \frac{3}{70}I_6 - \frac{55}{70}I_4 + \frac{9}{35}I_2}{I_6 + \frac{1}{5}I_4 - \frac{16}{25}I_2 + \frac{4}{25}} = \frac{185}{714} , \\ \psi_4(x) &= x\psi_3(x) - \gamma_3\psi_2(x) = x^4 - \frac{46}{51}x^2 + \frac{37}{357} . \end{aligned}$$

Els polinomis ortogonals mòncics cercats són, doncs,

$$1 , \quad x , \quad x^2 - \frac{2}{5} , \quad x^3 - \frac{9}{14}x , \quad x^4 - \frac{46}{51}x^2 + \frac{37}{357} .$$

Problema 3.17 a) Demostreu que la família de polinomis

$$Q_j(y) = T_j(2y - 1) \quad (j \geq 0)$$

és ortogonal respecte al pes $w(y) = (y(1-y))^{-\frac{1}{2}}$ a $[0, 1]$, on els polinomis $T_j(x)$ ($j \geq 0$) són els polinomis de Tchebixev.

b) Aproximeu y^{10} a l'interval $[0, 1]$ per un polinomi $p_4^*(y)$ de grau menor o igual que 4, de manera que l'error d'aproximació per mínims quadrats

$$\int_0^1 \frac{(y^{10} - p_4(y))^2}{(y(1-y))^{\frac{1}{2}}} dy$$

sigui mínim per a $p_4 = p_4^*$.

SOLUCIÓ:

a) Sabem que els polinomis de Tchebixev $T_j(x)$ ($j \geq 0$), donats per

$$T_j(x) = \cos(j \arccos x) \quad \forall x \in [-1, 1]$$

o també per la recurrència

$$T_0(x) = 1, \quad T_1(x) = x; \quad T_{j+1}(x) = 2xT_j(x) - T_{j-1}(x) \quad (j \geq 1),$$

són ortogonals respecte al pes $w(x) = \frac{1}{\sqrt{1-x^2}}$ a $[-1, 1]$.

En efecte, de

$$\begin{aligned} (T_j, T_l) &= \int_{-1}^1 \frac{\cos(j \arccos x) \cos(l \arccos x)}{\sqrt{1-x^2}} dx \\ &= \int_0^\pi \cos(j\theta) \cos(l\theta) d\theta, \end{aligned}$$

tenim

$$(T_j, T_l) = \begin{cases} 0 & \text{si } j \neq l \\ \pi & \text{si } j = l = 0 \\ \frac{\pi}{2} & \text{si } j = l \neq 0 \end{cases}.$$

Per al cas que ens ocupa i fent el canvi $x = 2y - 1$,

$$\begin{aligned} ((Q_j, Q_l)) &= \int_0^1 \frac{Q_j(y)Q_l(y)}{\sqrt{y(1-y)}} dy = \int_0^1 \frac{T_j(2y-1)T_l(2y-1)}{\sqrt{y(1-y)}} dy = \\ &= \int_{-1}^1 \frac{T_j(x)T_l(x)}{\sqrt{1-x^2}} dx = (T_j, T_l), \end{aligned}$$

d'on es desprèn la relació d'ortogonalitat enunciada.

b) L'aproximació polinomial per mínims quadrats demanada s'escriu com una combinació lineal dels polinomis ortogonals $Q_j(y)$ fins a grau 4

$$p_4^*(y) = \sum_{j=0}^4 c_j^* Q_j(y).$$

Els coeficients ortogonals corresponents són

$$c_j^* = \frac{((y^{10}, Q_j))}{((Q_j, Q_j))} \quad (j = 0 \div 4).$$

A continuació, portem a terme el seu càlcul.

Emprant els canvis $y = \frac{1+x}{2}$ i $x = \cos \theta$, tenim primer

$$\begin{aligned} ((y^{10}, Q_j)) &= \int_0^1 \frac{y^{10} Q_j(y)}{\sqrt{y(1-y)}} dy = \frac{1}{2^{10}} \int_{-1}^1 \frac{(1+x)^{10} T_j(x)}{\sqrt{1-x^2}} dx \\ &= \frac{1}{2^{10}} \int_0^\pi (1+\cos \theta)^{10} \cos j\theta d\theta \\ &= \int_0^\pi \cos^{20} \frac{\theta}{2} \cos j\theta d\theta . \end{aligned}$$

La integral resultant es pot avaluar fàcilment després d'expandir el terme $\cos^{20} \frac{\theta}{2}$ així:

$$\begin{aligned} \cos^{20} \frac{\theta}{2} &= \left(\frac{\exp(i\frac{\theta}{2}) + \exp(-i\frac{\theta}{2})}{2} \right)^{20} \\ &= \frac{1}{2^{20}} \sum_{l=0}^{20} \binom{20}{l} e^{i(l-10)\theta} = \frac{1}{2^{20}} \sum_{l=-10}^{10} \binom{20}{10+l} e^{il\theta} \\ &= \frac{1}{2^{20}} \binom{20}{10} + \frac{1}{2^{19}} \sum_{l=1}^{10} \binom{20}{10+l} \frac{e^{il\theta} + e^{-il\theta}}{2} \\ &= \frac{1}{2^{20}} \binom{20}{10} + \frac{1}{2^{19}} \sum_{l=1}^{10} \binom{20}{10+l} \cos l\theta . \end{aligned}$$

Tenim, doncs,

$$((y^{10}, Q_j)) = \frac{\pi}{2^{20}} \binom{20}{10-j} \quad (j = 0 \div 10) ;$$

$$c_0^* = \frac{1}{2^{20}} \binom{20}{10} ,$$

$$c_j^* = \frac{1}{2^{19}} \binom{20}{10-j} \quad (j = 1 \div 4) .$$

De

$$\frac{1}{2^{20}} \binom{20}{10} = \frac{46189}{8192} ,$$

en resulta:

$$c_0^* = \frac{1}{2^{20}} \binom{20}{10} = \frac{46189}{8192} ,$$

$$c_1^* = \frac{1}{2^{19}} \binom{20}{9} = \frac{10}{11} 2c_0^* = \frac{20995}{10240} ,$$

$$c_2^* = \frac{1}{2^{19}} \binom{20}{8} = \frac{9}{12} c_1^* = \frac{62985}{40960} ,$$

$$c_3^* = \frac{1}{2^{19}} \binom{20}{7} = \frac{8}{13} c_2^* = \frac{4845}{5120} ,$$

$$c_4^* = \frac{1}{2^{19}} \binom{20}{6} = \frac{7}{14} c_3^* = \frac{4845}{10240} .$$

Problema 3.18 Ajusteu per mínims quadrats la taula de dades:

x	0.25	0.50	0.75	1.00	1.25	1.50	1.75
y	0.40	0.50	0.90	1.28	1.60	1.66	2.02

a funcions dels tipus següents:

a) $y = a_0 + a_1x$,

b) $y = a_0 + a_1x + a_2x^2$,

c) $y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$,

d) $y = ax^\alpha$.

SOLUCIÓ:

a) La taula considerada (x_k, y_k) ($k = 0 \div 6$) permet escriure el sistema lineal sobre-determinat

$$a_0 + x_k a_1 = y_k \quad (k = 0 \div 6) ;$$

això és,

$$\begin{pmatrix} 1 & 0.25 \\ 1 & 0.50 \\ 1 & 0.75 \\ 1 & 1.00 \\ 1 & 1.25 \\ 1 & 1.50 \\ 1 & 1.75 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 0.40 \\ 0.50 \\ 0.90 \\ 1.28 \\ 1.60 \\ 1.66 \\ 2.02 \end{pmatrix} ,$$

que, en notació matricial, escrivim com $Ma = y$.

El mètode d'ortogonalització de Gram-Schmidt permet factoritzar la matriu $M = (m^{(1)} \ m^{(2)})$ del sistema en la forma QR, on Q és una matriu 7x2 amb columnes ortogonals ($D = Q^\top Q$ és diagonal), i R és una matriu 2x2 triangular superior amb uns a la diagonal:

$$\begin{aligned} q^{(1)} &= m^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} , \\ d_1 &= \|q^{(1)}\|_2^2 = 7 ; \\ r_{12} &= \frac{q^{(1)\top} m^{(2)}}{d_1} = \frac{7}{7} = 1 , \end{aligned}$$

$$q^{(2)} = m^{(2)} - r_{12}q^{(1)} = \begin{pmatrix} -0.75 \\ -0.50 \\ -0.25 \\ 0.00 \\ 0.25 \\ 0.50 \\ 0.75 \end{pmatrix},$$

$$d_2 = \|q^{(2)}\|_2^2 = 1.75.$$

Així doncs, resulta

$$M = QR = \begin{pmatrix} 1 & -0.75 \\ 1 & -0.50 \\ 1 & -0.25 \\ 1 & 0.00 \\ 1 & 0.25 \\ 1 & 0.50 \\ 1 & 0.75 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 7 & 0 \\ 0 & 1.75 \end{pmatrix}.$$

El sistema d'equacions normals associat al sistema $QRa = y$ és llavors $M^\top Ma^* = M^\top y$, equivalent a $Ra^* = D^{-1}Q^\top y$:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_0^* \\ a_1^* \end{pmatrix} = \begin{pmatrix} 1.194285714 \\ 1.125714286 \end{pmatrix},$$

de solució

$$a_0^* = 0.0685714282, \quad a_1^* = 1.125714286.$$

Com a il·lustració, usarem a continuació el mètode d'ortogonalització de Householder per tal de trobar un altre tipus de factorització QR de la matriu M .

Si u és un vector no nul, la matriu de Householder $P \equiv P(u) = I - \alpha uu^\top$ amb $\alpha = \frac{2}{u^\top u}$ és una matriu ortogonal simètrica: $P = P^\top = P^{-1}$ i, per a qualsevol vector a de norma euclidiana s , es compleix

$$P(a + se^{(1)})a = -se^{(1)}, \quad P(a - se^{(1)})a = se^{(1)},$$

on $e^{(1)} = (1 \ 0 \ \dots \ 0)^\top$.

En el nostre cas, si escrivim $M = (m^{(1)} \ m^{(2)})$ i prenem $a = m^{(1)}$, aleshores es compleix $s = \|m^{(1)}\|_2 = \sqrt{7}$. Si $u = m^{(1)} \pm se^{(1)}$, aleshores tenim $P(u)m^{(1)} = \mp se^{(1)}$; és a dir, la matriu $P_1 = P(u)$ envia el vector $m^{(1)}$ en un vector amb només la primera component no nul·la. Com que la primera component de $m^{(1)}$ és positiva, escollim $u = m^{(1)} + se^{(1)}$ per tal d'evitar cancel·lacions, i podem procedir a calcular $M_2 \equiv P_1 M = (P_1 m^{(1)} \ P_1 m^{(2)})$ i $P_1 y$:

$$s = \|m^{(1)}\|_2 = \sqrt{7},$$

$$\begin{aligned}
u &= m^{(1)} + \sqrt{7}e^{(1)} = \begin{pmatrix} 1 + \sqrt{7} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \\
\alpha &= \frac{2}{u^\top u} = \frac{1}{s(s+1)} = \frac{1}{\sqrt{7}(\sqrt{7}+1)} = \frac{1}{6} - \frac{\sqrt{7}}{42}, \\
P_1 m^{(1)} &= -\sqrt{7}e^{(1)} = \begin{pmatrix} -\sqrt{7} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \\
P_1 m^{(2)} &= m^{(2)} - \alpha(u^\top m^{(2)})u = m^{(2)} - \frac{9 - \sqrt{7}}{8}u \\
&= \frac{1}{8} \begin{pmatrix} -8\sqrt{7} \\ -5 + \sqrt{7} \\ -3 + \sqrt{7} \\ -1 + \sqrt{7} \\ 1 + \sqrt{7} \\ 3 + \sqrt{7} \\ 5 + \sqrt{7} \end{pmatrix}, \\
P_1 y &= y - \alpha(u^\top y)u = y - \left(\frac{199}{150} - \frac{139}{1050}\sqrt{7}\right)u \\
&= \frac{1}{1050} \begin{pmatrix} -1254\sqrt{7} \\ -868 + 139\sqrt{7} \\ -448 + 139\sqrt{7} \\ -49 + 139\sqrt{7} \\ 287 + 139\sqrt{7} \\ 350 + 139\sqrt{7} \\ 728 + 139\sqrt{7} \end{pmatrix}.
\end{aligned}$$

Tenim, així,

$$M_2 = P_1 M = \begin{pmatrix} -\sqrt{7} & -\sqrt{7} \\ 0 & \frac{-5+\sqrt{7}}{8} \\ 0 & \frac{-3+\sqrt{7}}{8} \\ 0 & \frac{-1+\sqrt{7}}{8} \\ 0 & \frac{1+\sqrt{7}}{8} \\ 0 & \frac{3+\sqrt{7}}{8} \\ 0 & \frac{5+\sqrt{7}}{8} \end{pmatrix} = \left(\begin{array}{c|c} -\sqrt{7} & -\sqrt{7} \\ \hline 0 & \overline{M}_2 \end{array} \right),$$

on $\overline{M}_2$ és ara una matriu 6x1; és a dir, un vector de 6 files que anomenem $\overline{m}^{(1)}$

$$\overline{M}_2 = \overline{m}^{(1)} = \frac{1}{8} \begin{pmatrix} -5 + \sqrt{7} \\ -3 + \sqrt{7} \\ -1 + \sqrt{7} \\ 1 + \sqrt{7} \\ 3 + \sqrt{7} \\ 5 + \sqrt{7} \end{pmatrix}$$

i que volem transformar en un vector amb només la primera component no nul·la, mitjançant una altra transformació $\overline{P}_2$ tal que

$$P_2 = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \overline{P}_2 \end{array} \right).$$

Fem, així,

$$s = \|\overline{m}^{(1)}\|_2 = \frac{\sqrt{7}}{2},$$

$$\overline{u} = \overline{m}^{(1)} - s\overline{e}^{(1)} = \frac{1}{8} \begin{pmatrix} -5 - 3\sqrt{7} \\ -3 + \sqrt{7} \\ -1 + \sqrt{7} \\ 1 + \sqrt{7} \\ 3 + \sqrt{7} \\ 5 + \sqrt{7} \end{pmatrix},$$

$$\alpha = \frac{2}{\overline{u}^\top \overline{u}} = \frac{1}{s(s + \frac{5-\sqrt{7}}{8})} = \frac{168 - 40\sqrt{7}}{133},$$

$$\overline{P}_2 \overline{m}^{(1)} = \frac{\sqrt{7}}{2} \overline{e}^{(1)} = \begin{pmatrix} \frac{\sqrt{7}}{2} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{aligned} \overline{P}_2 \overline{P}_1 \overline{y} &= \overline{P}_1 \overline{y} - \alpha (\overline{u}^\top \overline{P}_1 \overline{y}) \overline{u} = \overline{P}_1 \overline{y} - \left(\frac{10304 + 688\sqrt{7}}{9975} \right) \overline{u} \\ &= \begin{pmatrix} \frac{197}{350}\sqrt{7} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \end{aligned}$$

on les cinc darreres components de $\overline{P}_2 \overline{P}_1 \overline{y}$ no han estat calculades perquè no s'usaran.

Tenim, doncs,

$$M_3 = P_2 P_1 M = \sqrt{7} \begin{pmatrix} -1 & -1 \\ 0 & \frac{1}{2} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \tilde{R} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$P_2 P_1 y = \begin{pmatrix} -\frac{1254}{1050}\sqrt{7} \\ \frac{197}{350}\sqrt{7} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} \widetilde{P_2 P_1 y} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

i, per tant, el sistema d'equacions normals $M^\top M a^* = M^\top y$ és equivalent a $\tilde{R} a^* = \widetilde{P_2 P_1 y}$:

$$\sqrt{7} \begin{pmatrix} -1 & -1 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} a_0^* \\ a_1^* \end{pmatrix} = \sqrt{7} \begin{pmatrix} -\frac{1254}{1050} \\ \frac{197}{350} \end{pmatrix},$$

de solució:

$$a_0^* = \frac{12}{175} = 0.06857142857... \quad , \quad a_1^* = \frac{197}{175} = 1.125714286...$$

b) i c) Estem en el cas d'aproximació polinomial. De fet, el cas de l'apartat anterior era també d'aproximació polinomial i podríem aplicar els mateixos mètodes que allà. Resulta, però, més convenient generar polinomis ortogonals associats al producte escalar discret

$$(f, g) = \sum_{k=0}^6 f(x_k) g(x_k),$$

de norma euclidiana associada $\|f\|_2 = \sqrt{(f, f)}$, a través de la recurrència:

$$\psi_{-1}(x) = 0, \quad \psi_0(x) = 1,$$

$$\psi_{j+1}(x) = (x - \beta_j) \psi_j(x) - \gamma_j \psi_{j-1}(x) \quad (j \geq 0);$$

on

$$\beta_j = \frac{(\psi_j, x \psi_j)}{(\psi_j, \psi_j)} \quad (j \geq 0), \quad \gamma_j = \frac{(\psi_j, \psi_j)}{(\psi_{j-1}, \psi_{j-1})} \quad (j \geq 1).$$

Tindrem en compte les observacions següents:

1. Per comoditat, s'ha escollit que els polinomis ortogonals siguin mòncics: $\psi_j(x) = x^j + \dots$ ($j \geq 0$).

2. El conjunt d'abscisses d'aproximació consta de 7 punts equidistants amb pas $h = \frac{1}{4}$: $x_k = \frac{1}{4} + kh = (k+1)h$ ($k = 0 \div 6$); per tant, la família de polinomis ortogonals sobre aquesta xarxa es compon dels 7 polinomis $\psi_0(x), \psi_1(x), \dots, \psi_6(x)$.

3. En ser el conjunt d'abscisses d'aproximació simètric respecte a l'abscissa central $x_3 = 1$, es té $\beta_j = 1$ i $\psi_j(1+u) = (-1)^j \psi_j(1-u)$ ($j = 0 \div 6$). Així, si expressem els polinomis $\psi_j(x)$ en la variable $u = x - 1$, aquests seran senars (parells) en aquesta variable $x - 1$, si j és senar (parell) ($j = 0 \div 6$).

4. Usant aquests polinomis ortogonals $\psi_j(x)$ ($j = 0 \div 6$), la solució $p_n^*(x)$ del problema d'aproximació polinomial per mínims quadrats

$$\|f - p_n^*\|_2 = \min_{p_n \in \mathcal{P}_n} \|f - p_n\|_2 ,$$

on $\mathcal{P}_n$ denota l'espai vectorial dels polinomis de grau menor o igual que n , ve donada per

$$p_n^*(x) = \sum_{j=0}^n c_j^* \psi_j(x) , \quad c_j^* = \frac{(\psi_j, f)}{(\psi_j, \psi_j)} \quad (j = 0 \div n) .$$

Per tant, només cal conèixer els polinomis ortogonals $\psi_0(x), \dots, \psi_n(x)$ (en el nostre cas, per a $n = 2, 4$); tampoc no és necessari conèixer-los explícitament: n'hi ha prou de conèixer els valors $\psi_j(x_k)$ ($k = 0 \div 6$) ($j = 0 \div n$) que, a més, es troben successivament usant la fórmula recurrent.

Calculem en la taula següent els valors $\psi_j(x_k)$ ($k = 0 \div 6$), així com $d_j = (\psi_j, \psi_j)$, c_j^* i γ_j , començant amb $\psi_0(x) = 1$, $\psi_1(x) = x - 1$, i fent $\psi_{j+1}(x) = (x - 1)\psi_j(x) - \gamma_j\psi_{j-1}(x)$ ($j = 1 \div 3$):

k	0	1	2	3	4	5	6	d_j	c_j^*	γ_j
x_k	0.25	0.50	0.75	1.00	1.25	1.50	1.75			
y_k	0.40	0.50	0.90	1.28	1.60	1.66	2.02			
$\psi_0(x_k)$	1	1	1	1	1	1	1	7	$\frac{209}{175}$	
$\psi_1(x_k)$	$-\frac{3}{4}$	$-\frac{2}{4}$	$-\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{7}{4}$	$\frac{175}{175}$	$\frac{1}{4}$
$\psi_2(x_k)$	$\frac{5}{16}$	0	$-\frac{3}{16}$	$-\frac{4}{16}$	$-\frac{3}{16}$	0	$\frac{5}{16}$	$\frac{21}{64}$	$-\frac{525}{525}$	$\frac{3}{16}$
$\psi_3(x_k)$	$-\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	0	$-\frac{3}{32}$	$-\frac{3}{32}$	$\frac{3}{32}$	$\frac{27}{512}$	$-\frac{32}{75}$	$\frac{9}{56}$
$\psi_4(x_k)$	$\frac{9}{448}$	$-\frac{21}{448}$	$\frac{3}{448}$	$\frac{18}{448}$	$\frac{3}{448}$	$-\frac{21}{448}$	$\frac{9}{448}$	$\frac{99}{14336}$	$\frac{1856}{825}$	

Podem dir així que les millors aproximacions per mínims quadrats de f per polinomis de grau 2 i de grau 4, respectivament, són

$$\begin{aligned} p_2^*(x) &= c_0^* + c_1^* \psi_1(x) + c_2^* \psi_2(x) , \\ p_4^*(x) &= c_0^* + c_1^* \psi_1(x) + c_2^* \psi_2(x) + c_3^* \psi_3(x) + c_4^* \psi_4(x) , \end{aligned}$$

on

$$\begin{aligned} \psi_0(x) &= 1 , \quad \psi_1(x) = x - 1 , \quad \psi_2(x) = (x - 1)\psi_1(x) - \gamma_1\psi_0(x) , \\ \psi_3(x) &= (x - 1)\psi_2(x) - \gamma_2\psi_1(x) , \quad \psi_4(x) = (x - 1)\psi_3(x) - \gamma_3\psi_2(x) , \end{aligned}$$

amb

$$\gamma_1 = \frac{1}{4} , \quad \gamma_2 = \frac{3}{16} , \quad \gamma_3 = \frac{9}{56} .$$

Una manera eficaç d'avaluar $p_2^*(x)$ i $p_4^*(x)$ consisteix a aplicar la regla de Clenshaw. Per exemple, per al càlcul de $p_2^*(x)$ aquesta dona:

$$q_2 = c_2^* , \quad q_1 = (x - 1)q_2 + c_1^* , \quad p_2^*(x) = q_0 = (x - 1)q_1 - \gamma_1 q_2 + c_0^* .$$

Notem ara que la solució de l'apartat a) també la tenim resolta aquí, i ve donada per

$$p_1^*(x) = c_0^* + c_1^* \psi_1(x) = \frac{209}{175} + \frac{197}{175}(x-1) .$$

A més, els càlculs efectuats per tal de calcular una aproximació $p_n^*(x)$ són molt útils si es vol calcular després una aproximació $p_{n'}^*(x)$, amb $n' > n$.

Finalment, calculem els errors $e_j \equiv \|f - p_j^*\|_2$ de les aproximacions $p_j^*(x)$ ($j = 0 \div 4$). Per això, calculem primer

$$e_0^2 \equiv \|f - p_0^*\|_2^2 = \sum_{k=0}^6 (f(x_k) - c_0^*)^2 = 2.2701714... ;$$

usant ara que $f - p_1^*$ és ortogonal a $p_1^* - p_0^*$, el teorema de Pitàgores dona

$$e_0^2 \equiv \|f - p_0^*\|_2^2 = \|f - p_1^*\|_2^2 + \|p_1^* - p_0^*\|_2^2 = e_1^2 + \|p_1^* - p_0^*\|_2^2 ;$$

$$e_1^2 = e_0^2 - \|p_1^* - p_0^*\|_2^2 = e_0^2 - c_1^{*2} \|\psi_1\|_2^2 = 0.052514... ,$$

$$e_2^2 = e_1^2 - \|p_2^* - p_1^*\|_2^2 = e_1^2 - c_2^{*2} \|\psi_2\|_2^2 = 0.049295... ,$$

$$e_3^2 = e_2^2 - \|p_3^* - p_2^*\|_2^2 = e_2^2 - c_3^{*2} \|\psi_3\|_2^2 = 0.039695... ,$$

$$e_4^2 = e_3^2 - \|p_4^* - p_3^*\|_2^2 = e_3^2 - c_4^{*2} \|\psi_4\|_2^2 = 0.0047445...$$

Observem que les aproximacions de graus 2 i 3 no fan baixar substancialment l'error de l'aproximació per mínims quadrats. Això mostra que l'aproximació de grau 1 és ja prou bona en aquest cas, sobretot si tenim en compte els pocs càlculs que requereix.

d) A fi de trobar una expressió lineal d'aproximació, prenem les variables $Y = \ln y$ i $X = \ln x$ i llavors fem $a_0 = \ln A$ i $a_1 = \alpha$; en resulta la relació $Y = a_0 + a_1 X$.

Considerem ara la taula en les variables (X, Y)

X	Y
-1.386294	-0.916290
-0.693147	-0.693147
-0.287682	-0.105361
0.000000	0.246860
0.223144	0.470004
0.405465	0.506817
0.559616	0.703098

Fent els mateixos càlculs de l'apartat a), utilitzant el mètode de Gram-Schmidt, tenim

$$A = QR = \begin{pmatrix} 1 & -1.217880 \\ 1 & -0.524733 \\ 1 & -0.119268 \\ 1 & 0.168414 \\ 1 & 0.391558 \\ 1 & 0.573879 \\ 1 & 0.728030 \end{pmatrix} \begin{pmatrix} 1 & -0.168414 \\ 0 & 1 \end{pmatrix} ,$$

$$D = \begin{pmatrix} 7 & 0 \\ 0 & 2.813847552 \end{pmatrix} .$$

El sistema d'equacions normals té llavors la solució:

$$a_0^* = 0.18114 , \quad a_1^* = 0.89577 ;$$

així els valors corresponents d' A i α seran:

$$A^* = \exp(a_0^*) = 1.19859 , \quad \alpha^* = a_1^* = 0.89577 .$$

Problema 3.19 Trobeu l'aproximació minimax de la forma $x^2 + \hat{a}_1x + \hat{a}_0$ a la funció $\sin x$ a l'interval $[0, 1]$.

SOLUCIÓ:

L'aproximació cercada no és més que una aproximació minimax polinomial de grau més petit o igual que 1 a la funció $f(x) = \sin x - x^2$:

$$\hat{p}_1(x) = \hat{a}_0 + \hat{a}_1x .$$

El teorema de caracterització de Txebeixev permet trobar l'aproximació cercada imposant la propietat d'equioscil·lació que assegura que la funció error $e_1(x) = f(x) - \hat{a}_0 - \hat{a}_1x$ presenta 3 abscisses ξ_0 , ξ_1 i ξ_2 en les quals assoleix valors extrems $\pm \|e_1\|_\infty$ amb alternança de signes:

$$e_1(\xi_0) = -e_1(\xi_1) = e_1(\xi_2) = \pm \|e_1\|_\infty ;$$

a més, com que $f^{(2)}(x) = -\sin x - 2$ no canvia de signe sobre $[0, 1]$, ξ_0 i ξ_2 coincideixen amb els extrems de l'interval, $a = 0$, $b = 1$. L'altra abscissa ξ_1 es caracteritza per l'anul·lació de la derivada primera de l'error,

$$e_1'(\xi_1) = f'(\xi_1) - \hat{a}_1 = 0 .$$

De la propietat d'equioscil·lació en els extrems $e_1(a) = e_1(b)$, en surt la fórmula per al pendent de la recta d'aproximació minimax

$$\hat{a}_1 = \frac{f(b) - f(a)}{b - a} ,$$

i, de la propietat d'equioscil·lació en ξ_1 i a ,

$$\hat{a}_0 = \frac{1}{2}[f(a) + f(\xi_1) - (a + \xi_1)\hat{a}_1] .$$

Per al cas proposat, tenim que

$$\hat{a}_1 = \sin 1 - 1 = -0.158529 ;$$

l'abscissa ξ_1 compleix l'equació

$$\cos \xi_1 - 2\xi_1 + 1 - \sin 1 = 0 ,$$

que, usant els mètodes del capítol V, té la solució $\xi_1 = 0.5145272\dots$

Així,

$$\hat{a}_0 = \frac{1}{2}(\sin \xi_1 - \xi_1^2 - \hat{a}_1 \xi_1) = 0.154476\dots$$

Problema 3.20 Trobeu el polinomi d'aproximació minimax de grau 3 a la funció $f(x) = x^4$ a l'interval $[0, 10]$. Fiteu l'error comès a tot l'interval.

SOLUCIÓ:

L'obtenció d'aquest polinomi és molt simple si ens basem en el fet que el polinomi mònic de grau $n + 1$ amb norma del màxim mínima a $[-1, 1]$ és el polinomi mònic de Txebeixev

$$\tilde{T}_{n+1}(t) = \frac{1}{2^n} T_{n+1}(t) .$$

En l'interval $[0, 10]$, el polinomi de Txebeixev es pot trobar aplicant la transformació afí

$$t \in [-1, 1] \longrightarrow x = 5(t + 1) \in [0, 10] ,$$

d'inversa

$$x \in [0, 10] \longrightarrow t = \frac{x - 5}{5} \in [-1, 1] ,$$

a l'anterior polinomi $T_{n+1}(t)$; obtenim així el polinomi

$$T_{n+1}\left(\frac{x - 5}{5}\right) .$$

El polinomi mònic associat

$$\tilde{T}_{n+1}\left(\frac{x - 5}{5}\right) = \frac{5^{n+1}}{2^n} T_{n+1}\left(\frac{x - 5}{5}\right)$$

serà consegüentment el polinomi mònic de norma mínima a $[0, 10]$; per tant, el polinomi

$$x^{n+1} - \tilde{T}_{n+1}(x)$$

és el polinomi d'aproximació minimax de grau més petit o igual que n a x^{n+1} .

En el cas en qüestió, cal fer només $n = 3$.

Tenim:

$$\begin{aligned} T_4(t) &= 8t^4 - 8t^2 + 1, \\ T_4\left(\frac{x-5}{5}\right) &= \frac{8(x^4 - 20x^3 + 150x^2 - 500x + 625)}{625} \\ &\quad - \frac{8(x^2 - 10x + 25)}{25} + 1 \\ &= \frac{8x^4 - 160x^3 + 1000x^2 - 2000x + 625}{625}, \\ \tilde{T}_4\left(\frac{x-5}{5}\right) &= x^4 - (20x^3 - 125x^2 + 250x - \frac{625}{8}). \end{aligned}$$

El polinomi d'aproximació minimax cercat és, doncs,

$$\hat{p}_3(x) = 20x^3 - 125x^2 + 250x - \frac{625}{8}.$$

L'error comès estarà fitat per $\|\tilde{T}_4\|_\infty$ a $[0, 10]$, que resulta ser

$$\frac{625}{8} \|T_4(\frac{x-5}{5})\|_{\infty, [0, 10]} = \frac{625}{8} \|T_4(t)\|_{\infty, [-1, 1]} = \frac{625}{8}.$$

Problema 3.21 Utilitzeu l'economització de Lanczos per tal de calcular un polinomi de quart grau que approximi amb un error de magnitud menor que $2 \cdot 10^{-4}$ la funció

$$f(x) = \int_0^x \frac{e^t - 1}{t} dt, \quad \forall x \in [-1, 1].$$

SOLUCIÓ:

Trobem primer el desenvolupament de Taylor de f prop de $x_0 = 0$.

Tenim $f(0) = 0$ i

$$f'(x) = \frac{e^x - 1}{x} \equiv \frac{F(x)}{x} = \frac{1}{x} \left(x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + R_n(x) \right),$$

on

$$R_n(x) = \frac{F^{(n+1)}(\xi(x))}{(n+1)!} x^{n+1} = \frac{e^{\xi(x)}}{(n+1)!} x^{n+1}, \quad \xi(x) \in \langle 0, x \rangle.$$

Així,

$$f'(x) = 1 + \frac{x}{2!} + \frac{x^2}{3!} + \cdots + \frac{x^{n-1}}{n!} + \frac{e^{\xi(x)}}{(n+1)!} x^n, \quad \xi(x) \in \langle 0, x \rangle;$$

per tant,

$$f(x) = x + \frac{x^2}{2 \cdot 2!} + \frac{x^3}{3 \cdot 3!} + \cdots + \frac{x^n}{n \cdot n!} + \int_0^x \frac{e^{\xi(t)}}{(n+1)!} t^n dt, \quad \xi(t) \in \langle 0, t \rangle.$$

Emprant el teorema del valor mitjà per a integrals la resta del desenvolupament es pot escriure també com

$$\frac{e^\eta}{(n+1)!} \int_0^x t^n dt = \frac{e^\eta}{(n+1)(n+1)!} x^{(n+1)}, \quad \eta \in \langle 0, x \rangle.$$

Per al càlcul de f podem triar una primera possibilitat que consisteix a aproximar-la pel seu desenvolupament de Taylor fins a un cert ordre n de manera que produeixi uns errors fitats per $2 \cdot 10^{-4}$.

La resta del desenvolupament trobat fins a ordre n es pot fitar per $\frac{e}{(n+1)(n+1)!}$ a l'interval $[-1, 1]$. $n = 6$ és el valor més petit per al qual la dita fita és menor que l'error permès i es té

$$f(x) = x + \frac{x^2}{2 \cdot 2!} + \frac{x^3}{3 \cdot 3!} + \cdots + \frac{x^6}{6 \cdot 6!} \pm 7.7 \cdot 10^{-5}, \quad x \in [-1, 1].$$

La segona possibilitat intenta economitjar els càlculs a fer, tractant de trobar un polinomi d'aproximació a la funció de grau més baix que 6 que permeti calcular-la amb un error menor que el permès. Es fa servir a continuació l'economitjació de Lanczos per assolir-ho.

Es redueix primer el desenvolupament de grau 6 trobat a un polinomi de grau 5, restant-li el polinomi de grau 6 que ho pugui aconseguir i que sigui el de norma mínima; això és, el múltiple del polinomi de Txeixev $T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$ que tingui el mateix coeficient principal.

Com que el coeficient principal de $\frac{x^6}{6 \cdot 6!}$ és $\frac{1}{4320}$, caldrà restar-li

$$\frac{1}{2^5 \cdot 4320} T_6(x) = \frac{1}{138240} T_6(x),$$

que produeix un error fitat per $0.072 \cdot 10^{-4}$ a tot l'interval.

Notem que aquesta modificació no alterarà el coeficient de x^5 del desenvolupament de f , atesa la paritat dels polinomis de Txeixev. Així, si de manera anàloga li restem també un múltiple adequat del polinomi de Txeixev $T_5(x) = 16x^5 - 20x^3 + 5x$ a $f(x) - \frac{1}{138240} T_6(x)$, aconseguirem un polinomi de grau 4 que potser encara permetrà fer el càlcul de la funció f amb un error menor que el permès.

Tenim, així,

$$f(x) = x + \frac{x^2}{2 \cdot 2!} + \frac{x^3}{3 \cdot 3!} + \frac{x^4}{3 \cdot 4!} + \frac{x^5 - \frac{1}{16} T_5(x)}{5 \cdot 5!} + \frac{x^6 - \frac{1}{32} T_6(x)}{6 \cdot 6!} + \frac{1}{16 \cdot 600} T_5(x) + \frac{1}{32 \cdot 4320} T_6(x) \pm 0.77 \cdot 10^{-4}, \quad x \in [-1, 1].$$

Fitem ara els errors d'economització, tenint en compte que $\|T_j\|_\infty = 1$,

$$\left| \frac{1}{16 \cdot 600} T_5(x) + \frac{1}{32 \cdot 4320} T_6(x) \right| < 1.12 \cdot 10^{-4} ;$$

com que l'error del desenvolupament de Taylor inicial era més petit que $0.77 \cdot 10^{-4}$, l'error total encara és menor que l'error permès.

Utilitzant les expressions de $T_5(x)$ i $T_6(x)$, resulta el desenvolupament economitat cercat

$$\int_0^x \frac{e^t - 1}{t} dt = \frac{1}{138240} (1 + 138168 + 34542x^2 + 1056x^3 + 1488x^4) \pm 2 \cdot 10^{-4}, \quad x \in [-1, 1].$$

Notem que amb un altre pas d'economització no seria possible assegurar que els errors estiguessin encara dins els límits permesos.

Problema 3.22 a) La funció $\omega_m(x) = (x - x_0)(x - x_1) \cdots (x - x_m)$ és un factor de l'error en la interpolació polinomial en les abscisses x_k ($k = 0 \div m$). Si fem interpolació de Txeixev en un interval $[a, b]$ qualsevol, trobeu $\|\omega_m\|_\infty$ en funció de m , a i b .

b) Volem interpolare la funció $f(x) = \frac{1}{1+2x}$ a l'interval $[4, 5]$. Doneu una estimació del nombre de punts d'interpolació que cal a fi que l'error d'interpolació sigui menor que 10^{-6} en tot l'interval.

c) Compareu aquest error amb el que es produiria amb interpolació equidistant amb el mateix nombre de punts que els trobats a b).

SOLUCIÓ:

a) La interpolació de Txeixev de grau m a l'interval $[-1, 1]$ té lloc en les anomenades abscisses de Txeixev: els zeros del polinomi de grau $m + 1$ de Txeixev

$$T_{m+1}(t) = \cos((m+1) \arccos t) ;$$

és a dir, en els punts

$$t_k = \cos \frac{(2k+1)\pi}{2(m+1)} \quad (k = 0 \div m) .$$

En un altre interval $[a, b]$, considerarem com a interpolació de Txeixev la que té lloc en les abscisses de Txeixev que s'obtenen per la transformació de les de l'interval $[-1, 1]$ mitjançant el canvi afi

$$x = \frac{b-a}{2}t + \frac{a+b}{2} ;$$

això és, en els punts

$$x_k = \frac{b-a}{2} \cos \frac{(2k+1)\pi}{2(m+1)} + \frac{a+b}{2} \quad (k = 0 \div m) .$$

El factor $(t - t_0)(t - t_1) \cdots (t - t_m)$ de l'error en la interpolació de Txebeixev en $[-1, 1]$ té, doncs, els mateixos zeros que $T_{m+1}(t)$; atès que el coeficient principal de $T_{m+1}(t)$ és 2^m , tenim

$$(t - t_0)(t - t_1) \cdots (t - t_m) = \frac{1}{2^m} T_{m+1}(t) .$$

A més,

$$\omega_m(t) = \left(\frac{b-a}{2} \right)^{m+1} (t - t_0)(t - t_1) \cdots (t - t_m) .$$

Notem finalment que la relació $\|T_{m+1}\|_\infty = 1$ ens permet trobar la norma de ω_m :

$$\|\omega_m\|_{\infty, [a, b]} = \left(\frac{b-a}{2} \right)^{m+1} \frac{1}{2^m} = 2 \left(\frac{b-a}{4} \right)^{m+1} .$$

b) L'expressió anterior ens permet donar una fórmula de fitació de l'error en la interpolació de Txebeixev en un interval $[a, b]$ qualsevol a una funció f , suposada $m+1$ vegades diferenciable amb continuïtat

$$|f(x) - p_m(x)| \leq 2 \frac{M_{m+1}}{(m+1)!} \left(\frac{b-a}{4} \right)^{m+1} ,$$

on M_{m+1} és una fita superior de $|f^{(m+1)}|$ a $[a, b]$.

La derivada $m+1$ de la funció $f(x) = \frac{1}{1+2x}$ és

$$f^{(m+1)}(x) = (-1)^{m+1} 2^{m+1} (m+1)! \frac{1}{(1+2x)^{m+2}} ;$$

el seu valor absolut, en ser decreixent a l'interval $[a, b] = [4, 5]$, pot fitar-se pel seu valor en l'extrem inferior $x = 4$; així, per a $x \in [4, 5]$,

$$|f(x) - p_m(x)| \leq 2 \cdot 2^{m+1} \frac{1}{9^{m+2}} \frac{1}{4^{m+1}} = \frac{4}{18^{m+2}} .$$

Per a $m = 3$, aquesta fita val aproximadament $2.1 \cdot 10^{-6}$, i, per a $m = 4$, $1.17 \cdot 10^{-7}$. Resulta així que, fent una interpolació de Txebeixev en 5 punts ($m = 4$), podem estar segurs que l'error d'aquesta serà menor que 10^{-6} en tot l'interval $[4, 5]$.

c) Si portéssim a terme una interpolació equidistant en les 5 abscisses $x_k = -4 + kh_4$ ($k = 0 \div 4$), amb $h_4 = \frac{1}{4}$, l'error d'interpolació seria menor que

$$\frac{M_5}{5!} h_4^5 \Omega_4 \simeq 2.13 \cdot 10^{-7} ,$$

on

$$\Omega_4 = \frac{\sqrt{950 + 58\sqrt{145}}}{5\sqrt{5}} \simeq 3.63$$

està introduïda en el problema III.3.

Tindríem així un error d'interpolació menor que 10^{-6} , encara que dues vegades més gran que l'error d'interpolació de Txebeixev en 5 punts. El més gran avantatge de la interpolació de Txebeixev és, però, que requereix menys càlcul que l'equidistant.

PROBLEMES PROPOSATS

1. Trobeu el polinomi d'interpolació a la taula

x_k	0	1	2	4
f_k	1	5	10	24

emprant els mètodes de: a) Lagrange, b) les diferències dividides de Newton, c) Aitken i d) Neville.

2. Repetiu l'exercici anterior amb la taula

x_k	0	1	2	4	8
f_k	1	5	10	24	50

Es poden aprofitar els càlculs fets amb tots quatre mètodes del problema anterior? Quins dels quatre mètodes permeten aprofitar els càlculs fets en l'exercici anterior?

3. Volem tabular la *funció de Bessel d'ordre 0*

$$J_0(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin t) dt$$

en punts equidistants de l'interval $[0, 1]$. Quin pas de tabulació s'ha d'emprar a fi i efecte que tots els valors de $J_0(x)$ ($x \in [0, 1]$) obtinguts per interpolació lineal (emprant els dos punts de la taula més propers) tinguin un error degut a la interpolació menor que $\frac{1}{2}10^{-6}$? I si fem interpolació cúbica (emprant els quatre punts més propers)?

4. Interpolem la funció e^x a $[0, 1]$ per un polinomi de grau n en abscisses equidistants. Doneu el valor de n mínim per tal que l'error d'interpolació sigui menor que 10^{-6} en qualsevol punt de $[0, 1]$. Feu el mateix prenent com a fita d'error 10^{-12} .
5. Sigui $f \in C^\infty(\mathbb{R})$ amb $|f^{(l)}(x)| \leq 4^l$ ($l \geq 0$), $\forall x \in [0, 1]$. Volem construir polinomis d'interpolació $p_m(x)$ ($m \geq 0$) en abscisses equidistants. Sabem que, degut als errors d'arrodoniment, f s'avalua amb un error menor que 10^{-5} i anomenem el polinomi obtingut per $\bar{p}_m(x)$. Quin valor de m fa que la fita de l'error

$$\max_{x \in [0, 1]} |f(x) - \bar{p}_m(x)|$$

sigui mínima?

6. Considerem la funció

$$f(x) = \frac{1}{1+x}$$

i sigui $p_m(x)$ el polinomi d'interpolació a f en les abscisses $1, a, a^2, \dots, a^m$ amb $0 < a < 1$ ($m \geq 0$); demostreu que

$$\lim_{m \rightarrow \infty} p_m(0) = f(0) .$$

7. Segui $p_m(x)$ el polinomi interpolador a la funció

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

en els punts equidistants $x_k^{(m)} = -5 + \frac{10k}{m}$ ($k = 0 \div m$) $m \geq 0$. Verifiquen la divergència de la successió $(p_m(4))_{m \geq 0}$ (*fenomen de Runge*).

8. a) Trobeu el polinomi de Taylor $p_n(x)$ de grau n de la funció $f(x) = e^x$ per a qualsevol n .

b) Doneu expressions per als errors $f(x) - p_n(x)$ i fiteu-les.

c) Per a quins valors de n pot assegurar-se que el polinomi de Taylor donarà $f(x)$ amb 8 xifres decimals correctes, per a tot $x \in [-1, 1]$?

Repetiu els tres apartats per a $f(x) = \sin x$ a $[0, \pi/2]$.

9. Considereu els desenvolupaments de Taylor de les funcions següents en $x_0 = 0$ en els intervals que s'indiquen:

i) $\ln(1+x)$ a $(-1, 1)$;

ii) $(1+x)^{-1}$ a $(-1, 1)$;

iii) $\cos x$ a $(-\pi/2, \pi/2)$.

Responen per a cada funció a les qüestions següents:

a) Quants termes hauríem de prendre per tal d'assegurar una precisió en l'aproximació de 8 xifres decimals correctes, per a qualsevol valor de x en l'interval donat?

b) Per a quins valors de x tenim aquella precisió, si prenem només els tres primers termes no nuls del desenvolupament?

10. a) Trobeu els desenvolupaments de Taylor de les funcions següents en qualsevol x_0 :

i) e^x , $\cosh x$, $\sinh x$, a^x ($a > 0$);

ii) $\sin x$, $\cos x$;

iii) $\ln x$, $\log x$, x^α .

b) Feu explícits aquests desenvolupaments per a $x_0 = 2$ i $x_0 = \pi$.

11. Mitjançant l'ús de desenvolupaments de Taylor en $x_0 = 0$, trobeu els límits, quan x tendeix a 0, de les funcions següents:

$$\frac{\sin x^2}{\cos x - 1}, \quad \frac{x^4}{(e^x - 1)^2}, \quad \left(\frac{1+x}{1-x}\right)^{\frac{1}{x}}.$$

12. Volem calcular la funció

$$f(x) = \frac{-2x^3 + 3x - 3 \sin x \cos x}{3x^3},$$

en $\bar{x} = 0.1$.

- a) Amb quina precisió s'ha de conèixer x prop de $\bar{x}$ per tal que l'error propagat a $f(x)$ sigui menor que 10^{-3} ?
- b) Doneu una altra fórmula per calcular $f(\bar{x})$ que sigui millor des del punt de vista numèric, si es tenen en compte també els errors en els càlculs.
13. Trobeu els desenvolupaments de Taylor de grau n en $x_0 = 0$ de les següents funcions, expressant-los en termes dels nombres de Bernoulli:
- a) $\cot x (= i \frac{e^{2ix} + 1}{e^{2ix} - 1})$;
- b) $\tan x (= \cot x - 2i(1 + \frac{2}{e^{4ix} - 1}))$;
- c) $\frac{x}{\sin x} (= \frac{2ix}{e^{ix} - 1} - \frac{2ix}{e^{2ix} - 1})$;
- d) $\ln \frac{\sin x}{x}$.

14. Els *nombres d'Euler* $(E_j)_{j \geq 0}$ es defineixen per

$$\frac{1}{\cos x} = E_0 - \frac{E_2}{2!}x^2 - \frac{E_4}{4!}x^4 + \dots$$

- a) Proveu que es compleixen les relacions

$$E_0 = 1, \quad \sum_{r=0}^s \binom{2s}{2r} E_{2r} = 0 \quad (s \geq 1).$$

- b) Calculeu el desenvolupament de Taylor en $x_0 = 0$ de $\ln(\cos x)$.

15. Proveu que els coeficients c_j ($j \geq 0$) del desenvolupament de Taylor

$$(1 - x - x^2)^{-1} = c_0 + c_1x + c_2x^2 + \dots$$

verifiquen la relació $c_{j+1} = c_j + c_{j-1}$ ($j \geq 1$); són els anomenats *nombres de Fibonacci*.

16. Els *polinomis de Legendre* $(P_j(t))_{j \geq 0}$ poden definir-se a partir de la seva *funció generatriu* $F(t, x) = (1 - 2tx + x^2)^{-1}$. Els dits polinomis són llavors els coeficients del desenvolupament de Taylor de $F(t, x)$ respecte a x en $x = x_0$:

$$F(t, x) = P_0(t) + P_1(t)x + P_2(t)x^2 + \cdots, \quad |x| < 1, \quad |t| \leq 1.$$

- a) Comproveu que

$$P_0(t) = 1, \quad P_1(t) = t, \quad P_2(t) = \frac{1}{2}(3t^2 - 1), \quad P_3(t) = \frac{1}{2}(5t^3 - 3t).$$

- b) Demostreu que $P_j(-t) = (-1)^j P_j(t)$ ($j \geq 0$).
 c) Per a $j \geq 1$, trobeu les relacions de recurrència:

$$(j+1)P_{j+1}(t) = (2j+1)tP_j(t) - jP_{j-1}(t),$$

$$2P_j(t) = P'_{j+1}(t) - 2tP'_j(t) + P'_{j-1}(t),$$

$$(2j+1)P_j(t) = P'_{j+1}(t) - P'_{j-1}(t).$$

17. Els polinomis de Tchebixev $(T_j(t))_{j \geq 0}$ es poden definir també per la *funció generatriu*

$$\frac{1 - tx}{1 - 2tx + x^2} = T_0(t) + T_1(t)x + T_2(t)x^2 + \cdots, \quad |x| < 1, \quad |t| \leq 1.$$

- a) Trobeu $T_j(t)$ ($j = 0, 1, 2, 3, 4$).
 b) Demostreu que $T_j(-t) = (-1)^j T_j(t)$ ($j \geq 0$).
 c) Proveu que $T_{j+1}(t) = 2tT_j(t) - T_{j-1}(t)$ ($j \geq 1$).
 d) Demostreu que $T_j(t) = \cos(j \arccos t)$, $|t| < 1$ ($j \geq 0$).

18. Els *polinomis d'Hermite* $(H_j(t))_{j \geq 0}$ es poden definir també per la funció generatriu

$$e^{2tx - x^2} = H_0(t) + \frac{H_1(t)}{1!}x + \frac{H_2(t)}{2!}x^2 + \cdots, \quad |x| < 1, \quad t \in \mathbb{R}.$$

- a) Trobeu $H_j(t)$ ($j = 0, 1, 2, 3, 4$).
 b) Demostreu que $H_j(-t) = (-1)^j H_j(t)$ ($j \geq 0$).
 c) Proveu les relacions de recurrència:

$$H_{j+1}(t) = 2tH_j(t) - 2jH_{j-1}(t), \quad H'_j(t) = 2jH_{j-1}(t) \quad (j \geq 1).$$

- d) Proveu que

$$H_j(t) = (-1)^j e^{t^2} \frac{d^j}{dx^j} e^{-t^2}.$$

19. Segui

$$f(x) = x^9 \sin \frac{1}{x}, \quad \text{si } x \neq 0; \quad f(0) = 0.$$

a) Proveu que el polinomi de Taylor de grau n ($0 \leq n \leq 8$) és idènticament nul.

b) Si

$$R_4(x) = x^9 \sin \frac{1}{x} - \left(x^8 - \frac{x^6}{3!} + \frac{x^4}{5!} \right),$$

proveu que $|R_4(x)| \leq \frac{x^2}{7!}$ i que la fórmula aproximada

$$x^9 \sin \frac{1}{x} \simeq x^8 - \frac{x^6}{3!} + \frac{x^4}{5!}$$

és especialment apropiada per a x complint $\frac{1}{\sqrt{72}} \leq |x| \leq \frac{1}{\sqrt{42}}$; és a dir, que és millor que les fórmules

$$x^9 \sin \frac{1}{x} \simeq x^8 - \frac{x^6}{3!}, \quad x^9 \sin \frac{1}{x} \simeq x^8 - \frac{x^6}{3!} + \frac{x^4}{5!} - \frac{x^2}{7!}, \quad \text{etc.}$$

20. (*Interpolació d'Hermite*) a) Demostreu que el polinomi p_{2m+1} , de grau menor o igual que $2m+1$, complint:

$$p_{2m+1}(x_k) = f_k, \quad p'_{2m+1}(x_k) = f'_k \quad (k = 0 \div m),$$

existeix sempre i és únic.

b) Demostreu que

$$p_{2m+1}(x) = \sum_{i=0}^m f_i \Phi_i(x) + \sum_{i=0}^m f'_i \Psi_i(x),$$

amb

$$\Phi_i(x) = (1 - 2l'_i(x)(x - x_i))l_i^2(x) \text{ i } \Psi_i(x) = (x - x_i)l_i^2(x),$$

essent $l_i(x)$ el polinomi de Lagrange associat al punt x_i ($i = 0 \div m$).

c) Demostreu també:

$$\begin{aligned} \sum_{i=0}^m \Phi_i(x) &= 1, \\ \sum_{i=0}^m \Phi_i(x)x_i^j + j \sum_{i=0}^m \Psi_i(x)x_i^{j-1} &= x^j \quad (j = 1 \div 2m+1), \\ \sum_{i=0}^m \Phi_i(x)x_i^{2m+2} + (2m+2) \sum_{i=0}^m \Psi_i(x)x_i^{2m+1} \\ &= x^{2m+2} - (x - x_0)^2 \cdots (x - x_m)^2. \end{aligned}$$

d) Trobeu el polinomi interpolador d'Hermite $p_5(x)$ a $f(x) = e^x$ en els punts $x_0 = -1$, $x_1 = 0$ i $x_2 = 1$, emprant les fórmules d'a) i diferències dividides generalitzades. Fiteu l'error $|f(x) - p_5(x)|$ a l'interval $[-1, 1]$.

21. a) Donada la funció $f(x) = \frac{1}{1+x}$ en l'interval $[0, 1]$, calculeu les normes

$$\|f\|_p = \left(\int_0^1 |f(x)|^p dx \right)^{\frac{1}{p}} \quad (p \geq 1)$$

i $\|f\|_\infty$.

- b) Comproveu que

$$\lim_{p \rightarrow 1} \|f\|_p = \|f\|_1, \quad \lim_{p \rightarrow \infty} \|f\|_p = \|f\|_\infty.$$

22. Sigui w una funció pes a $[a, b]$ normalitzada per $\int_a^b w(x) dx = 1$. Considereu les normes $\|f\|_\infty$ i $\|f\|_{2,w}$ a $[a, b]$.

a) Existeix alguna funció f tal que $\|f\|_\infty < \|f\|_{2,w}$?

b) Existeix alguna funció f tal que $\|f\|_\infty > \|f\|_{2,w}$?

En cas d'existència, doneu-ne un exemple.

23. Trobeu a, r reals i $\alpha \in [0, 2\pi]$ tals que l'expressió

$$y(x) = a + r \sin(x + \alpha)$$

aproximi de manera òptima, en el sentit dels mínims quadrats, als punts:

$$\left(-\frac{\pi}{2}, 1\right), (0, 0), \left(\frac{\pi}{2}, \frac{1}{2}\right), (\pi, 1).$$

24. Resoleu per mínims quadrats els següents sistemes lineals sobredeterminats:

$$\left(\begin{array}{cc|c} 1 & 1 & 5 \\ 1 & 2 & 9 \\ 1 & 3 & 12 \\ 1 & 4 & 16 \end{array} \right), \quad \left(\begin{array}{ccc|c} 1 & 1 & 1 & 5 \\ 1 & 1 & 2 & 9 \\ 1 & 1 & 3 & 12 \\ 1 & 1 & 4 & 16 \end{array} \right).$$

25. Sabem que els pesos atòmics de l'oxigen i del nitrogen són aproximadament $O = 16$ i $N = 14$; utilitzeu els pesos moleculars dels sis òxids de nitrogen donats a continuació per tal d'ajustar-los per mínims quadrats

Compost	NO	N_2O	NO_2	N_2O_3	N_2O_5	N_2O_4
Pes molecular	30.006	44.013	46.006	76.012	108.010	92.011

26. Els processos termodinàmics adiabàtics de sistemes físics caracteritzats per la pressió P , el volum V i la temperatura T (els gasos, per exemple) segueixen una llei del tipus $PV^\gamma = C$, on C és constant al llarg del procés. Ajusteu per mínims quadrats els valors de C i de γ en un procés adiabàtic segons la taula de mesures experimentals següent:

P (atm)	1.62	1.00	0.75	0.62	0.52	0.46
V (litres)	0.5	1.0	1.5	2.0	2.5	3.0

27. Hom suposa que el cometa Tentax, descobert l'any 1968, és un objecte del Sistema Solar. En un cert sistema de coordenades polars (r, φ) , centrat en el Sol, s'han mesurat experimentalment les següents posicions del cometa:

r	2.70	2.00	1.61	1.20	1.02
φ	48°	67°	83°	108°	126°

Les lleis de Kepler garanteixen que el cometa es mourà en una òrbita el·líptica, parabòlica o hiperbòlica (si es menyspreen les pertorbacions dels planetes), que, en les dites coordenades polars, tindrà per equació

$$r = \frac{p}{1 + e \cos \varphi} ,$$

on p és un paràmetre i e l'excentricitat. Ajusteu per mínims quadrats els valors de p i e , a partir de les mesures fetes.

28. Considerem un vector $y = (y_1 \cdots y_m)^T$ i una matriu M de dimensió $m \times n$, amb $(m \geq n \geq 2)$, i rang n . Notem les columnes de M per $m^{(j)}$ ($j = 1 \div n$). Per tal de resoldre el sistema $Ma = y$, es considera el mètode iteratiu

$$a^{(k+1)} = \frac{1}{m} \sum_{j=1}^m b^{(k,j)} ,$$

on $b^{(k,j)}$ és la projecció d' $a^{(k)}$ sobre l'hiperplà d'equació $m^{(j)T}a = y_j$ ($j = 1 \div n$).

- Demostreu que els valors propis de la matriu $M^T M$ pertanyen a l'interval $(0, m)$.
- Escriviu el mètode iteratiu anterior en la forma $a^{(k+1)} = Ba^{(k)} + c$ i demostreu que convergeix a la solució del sistema d'equacions normals corresponent al sistema $Ma = y$, partint de qualsevol aproximació $a^{(0)}$ inicial.
- Si $a^{(0)}$ no és ja solució del sistema d'equacions normals, es pot donar la convergència del mètode en un nombre finit d'iteracions?

29. Determineu les rectes que aproximem la corba $y(x) = \sin(\pi x)$ fent que:

- la norma euclidiana discreta de l'error en la xarxa

$$\{-0.5, -0.25, 0, 0.25, 0.5\}$$

sigui mínima;

ii) la norma euclidiana de l'error en l'interval $[-0.5, 0.5]$ sigui mínima.

30. Donada una taula de la funció f en punts equidistants $x_k = x_0 + kh$ ($k = -2 \div 2$), volem aproximar la derivada de la funció f en el punt x_0 per la derivada en aquest punt de la funció f^* de la forma

$$f^*(x) = \sum_{j=0}^4 a_j^*(x - x_0)^j$$

tal que

$$\sum_{k=-2}^2 (f(x_k) - f^*(x_k))^2$$

sigui mínima entre totes les funcions d'aquell tipus.

Doneu l'expressió d'aquesta derivada aproximada en funció de $f(x_k)$ ($k = -2 \div 2$) i de h .

31. Sigui w una funció pes a l'interval $[a, b]$. El seu moment d'ordre j es defineix per

$$\mu_j = \int_a^b w(x)x^j dx .$$

Demostreu que el sistema d'equacions

$$-\begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_{n-1} \\ \mu_1 & \mu_2 & \cdots & \mu_n \\ \vdots & \vdots & \cdots & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-2} \end{pmatrix} \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{n-1} \end{pmatrix} = \begin{pmatrix} \mu_n \\ \mu_{n+1} \\ \vdots \\ \mu_{2n-1} \end{pmatrix}$$

té com a solució els coeficients del polinomi mònic de grau n

$$\varphi_n(x) = x^n + \sum_{j=0}^{n-1} d_j x^j$$

de la família de polinomis ortogonals associada al pes w a $[a, b]$. La matriu del sistema anterior rep el nom de *matriu de Hankel*.

32. En l'espai vectorial dels polinomis a coeficients reals es defineix el producte escalar

$$\langle p, q \rangle = \sum_{j \geq 0} a_j b_j ,$$

essent

$$p(x) = \sum_{j \geq 0} a_j x^j , \quad q(x) = \sum_{j \geq 0} b_j x^j .$$

Existeix algun interval $[a, b]$ i algun pes w en el dit interval tals que

$$\langle p, q \rangle = \int_a^b p(x)q(x)w(x)dx \text{ ?}$$

33. Considerem la norma euclidiana associada al pes w a $[a, b]$. Demostreu que, d'entre tots els polinomis mònics de grau n , el polinomi mònic de grau n de la família de polinomis ortogonals respecte al producte escalar associat és el que té norma mínima.

34. Els polinomis d'Hermite queden definits per la recurrència:

$$H_0(x) = 1, H_1(x) = x, H_{j+1}(x) = 2xH_j(x) - 2jH_{j-1}(x).$$

- a) Calculeu-ne els deu primers.
 b) Escriuiu els polinomis $1, x, \dots, x^9$ en funció d'ells.
 c) Demostreu que els polinomis d'Hermite es poden definir també segons la fórmula

$$H_j(x) = (-1)^j \exp(-x^2) \frac{d^j}{dx^j} (\exp(-x^2)), (j \geq 0).$$

- d) Vegeu que són ortogonals respecte al producte escalar corresponent al pes $w(x) = \exp(-x^2)$ a $\mathbb{R}$.

35. a) Demostreu que els polinomis de Laguerre, definits per

$$L_j(x) = e^x \frac{d^j}{dx^j} (x^j e^{-x}),$$

són ortogonals respecte al producte escalar de pes $w(x) = e^{-x}$ a l'interval $[0, \infty)$.

- b) Calculeu-ne els 6 primers.
 c) Trobeu la relació de recurrència:

$$L_0(x) = 1, L_1(x) = 1 - x, \\ L_{j+1}(x) = (2j + 1 - x)L_j(x) - j^2 L_{j-1}(x) \quad (j \geq 1).$$

36. a) Determineu els polinomis ortogonals $\psi_j(x)$ ($j = 0 \div 4$) corresponents a la funció pes $x^2 + x^4$ a l'interval $[-1, 1]$.

- b) Donada $f(x) = \sin(\pi x)$, trobeu el polinomi $p_4^*(x)$ que fa mínim

$$\int_{-1}^1 (x^2 + x^4)(f(x) - p_4(x))^2 dx$$

entre els polinomis $p_4(x)$ de grau menor o igual que 4.

37. a) Sigui $f(x) = \cos x$, obteniu la combinació de polinomis de Legendre $P_j(x)$ ($j \geq 0$) de la forma

$$p_4(x) = \sum_{j=0}^4 c_j P_j(x)$$

que fa que sigui mínim l'error quadràtic

$$\int_{-1}^1 (f(x) - p_4(x))^2 dx .$$

- b) Quant val aquest mínim?

38. Trobeu el polinomi de primer grau obtingut per interpolació de Txebeixev de la funció $f(x) = \frac{1}{1+3x}$ a l'interval $(-1, 1)$. Quina és la norma del màxim de l'error en el dit interval?

39. Trobeu el polinomi d'interpolació de Txebeixev de grau 4 a la funció $f(x) = \ln(1+x)$ a l'interval $[2, 4]$,

i) emprant el mètode de les diferències dividides de Newton,

ii) trobant els coeficients ortogonals de Txebeixev.

Quin dels dos mètodes us sembla més eficient?

40. Donada la funció $f(\theta) = \theta(\pi - \theta)$ a $[0, \pi]$, determineu el polinomi en cosinus

$$f_n^*(\theta) = \sum_{j=0}^n c_j^* \cos(j\theta)$$

que minimitza la norma euclidiana de l'error

$$\|f - f_n\|_2 = \left(\int_0^\pi (f(\theta) - f_n(\theta))^2 d\theta \right)^{\frac{1}{2}} .$$

41. a) Sigui la funció f definida a l'interval $[-\pi, \pi]$ per

$$f(\theta) = \begin{cases} -1 & (-\pi < \theta < 0) \\ 1 & (0 < \theta < \pi) \\ 0 & \theta = -\pi, 0, \pi \end{cases} .$$

Calculeu els coeficients a_0, a_j, b_j ($j = 1 \div n$) de la funció

$$t_{2n-1}^*(\theta) = a_0 + \sum_{j=1}^{2n-1} a_j \cos(j\theta) + \sum_{j=1}^{2n-1} b_j \sin(j\theta)$$

que minimitzen la norma euclidiana de l'error.

b) Demostreu, per inducció, que

$$t_{2n-1}^*(\theta) = \frac{2}{\pi} \int_0^\theta \frac{\sin 2n\varphi}{\sin \varphi} d\varphi$$

i deduíu-ne que t_{2n-1}^* té el primer màxim local positiu a $\frac{\pi}{2n}$.

c) Vegeu també que

$$\lim_{n \rightarrow \infty} t_{2n-1}^* \left(\frac{\pi}{2n} \right) = \frac{2}{\pi} \int_0^\pi \frac{\sin \varphi}{\varphi} d\varphi .$$

d) Avalueu numèricament aquest últim resultat i discutiu-lo.

42. Trobeu el polinomi d'aproximació minimax de grau 4 a la funció $f(x) = x^5 - 3x^3$ a l'interval $[-3, 3]$ i fiteu l'error comès a tot l'interval.

43. Trobeu l'aproximació minimax polinomial de grau 1 a $f(x) = e^x$ a l'interval $[0, 1]$ i fiteu l'error comès en tot l'interval.

44. a) Aproximeu la funció $f(x) = \frac{1}{x+2}$ a l'interval $[-1, 1]$ per un polinomi de grau 1 de les formes següents:

i) per interpolació en els extrems,

ii) per interpolació de Txebeixev,

iii) minimitzant la norma del màxim de l'error.

b) Doneu fites per als errors comesos en tots els casos i compareu els resultats trobats tot justificant-los teòricament.

45. Trobeu l'aproximació minimax de la forma ax^2 a la funció $f(x) = x^4$ a l'interval $[0, 1]$.

46. Utilitzeu el desenvolupament de Taylor de

$$f(x) = \int_0^x \frac{\sin t}{t} dt$$

fins a grau 9 a fi de trobar un polinomi de grau mínim que approximi la funció donada sobre l'interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ amb un error menor que 10^{-4} .

CAPÍTOL 4

DERIVACIÓ, INTEGRACIÓ I SUMACIÓ

4.1 DERIVACIÓ NUMÈRICA

4.1.1 Introducció

Si bé hi ha regles ben conegudes per derivar les funcions més usuals, no sempre poden ser emprades (per exemple, en funcions donades per taules de valors o numèricament), o no és convenient de fer-ne ús (per exemple, en funcions amb expressions analítiques massa complicades). En aquests casos haurem de recórrer a tècniques numèriques que, partint dels valors de la dita funció en diverses abscisses, ens permetran calcular una aproximació al valor d'alguna de les seves derivades en una abscissa propera.

4.1.2 Derivades primeres

Fórmules de derivació interpolatòria i errors

La *derivació numèrica* d'una funció f diferenciable en $a \in \mathbb{R}$ consta de dues etapes:

- Construcció del polinomi interpolador $p_m(x)$ a la funció f en una família d'abscisses $x_0, x_1, \dots, x_m$ (que convindrà prendre prop d' a).
- Derivació del polinomi $p_m(x)$ i avaluació en a , donant la fórmula de derivació numèrica:

$$f'(a) \simeq p'_m(a) .$$

Les fórmules així obtingudes reben el nom de *fórmules de derivació interpolatòries* (vegeu el problema IV.1).

Si notem per $e_m(x)$ la funció d'error d'interpolació $f(x) - p_m(x)$ de la funció f pel polinomi $p_m(x)$ i suposem que f és $m + 1$ vegades diferenciable amb continuïtat en un interval I que conté els punts x_k ($k = 0 \div m$), aleshores, per a $x \in I$ disposem de les expressions següents per a l'error d'interpolació:

$$\begin{aligned} e_m(x) &= F_{m+1}(x)\omega_m(x) , \\ F_{m+1}(x) &= \frac{f^{(m+1)}(\xi(x))}{(m+1)!} = f[x_0, \dots, x_m, x] , \end{aligned}$$

on $\omega_m(x) = (x - x_0) \cdots (x - x_m)$ i $\xi(x) \in \langle x_0, \dots, x_m, x \rangle$ (vegeu les fórmules (3.2) i (3.6)). Resulta així que l'error en la derivació numèrica ve donat per

$$e'_m(a) = f'(a) - p'_m(a) = F'_{m+1}(a)\omega_m(a) + F_{m+1}(a)\omega'_m(a) ;$$

aquesta fórmula presenta el problema del càlcul de $F'_{m+1}(a)$, si a és arbitrari, atès que la funció F_{m+1} no és coneguda en general (vegeu-ne un exemple en el problema IV.2).

Ara bé, si a pertany al conjunt d'abscisses $x_0, \dots, x_m$ (és a dir, si $a = x_k$ per a algun $k = 0 \div m$), aleshores tenim

$$e'_m(x_k) = \frac{f^{(m+1)}(\xi_k)}{(m+1)!} \prod_{i \neq k} (x_k - x_i) ,$$

on $\xi_k = \xi(x_k) \in \langle x_0, \dots, x_m \rangle$.

Exemples

Si $m = 1$, $x_0 = a$, $x_1 = a + h$ i $f \in \mathcal{C}^2([a, a + h])$:

$$f'(a) = \frac{f(a + h) - f(a)}{h} + \frac{f^{(2)}(\xi)}{2}h , \quad \xi \in \langle a, a + h \rangle .$$

Si $m = 2$, $x_0 = a - h$, $x_1 = a$, $x_2 = a + h$ i $f \in \mathcal{C}^3([a - h, a + h])$:

$$f'(a) = \frac{f(a + h) - f(a - h)}{2h} - \frac{f^{(3)}(\xi)}{6}h^2 , \quad \xi \in \langle a - h, a + h \rangle .$$

4.1.3 Derivades d'ordre superior

Fórmules de derivació interpolatòria

El procediment anterior es pot repetir per calcular $f^{(2)}(a), \dots, f^{(d)}(a)$, sempre que $d \leq m$. Així, si coneixem la funció f en les abscisses $x_0, \dots, x_m$ pròximes a a , derivant d vegades el polinomi interpolador $p_m(x)$ i avaluant en a , tenim

$$f^{(d)}(a) \simeq d!f[x_0, x_1, \dots, x_d] .$$

Per tal d'obtenir expressions de l'error és convenient usar els errors d'interpolació de Taylor per a la fórmula trobada.

Exemple

Si $m = 2$, $x_0 = a - h$, $x_1 = a$, $x_2 = a + h$,

$$f^{(2)}(a) \simeq 2f[x_0, x_1, x_2] = \frac{f(a + h) - 2f(a) + f(a - h)}{h^2} ;$$

si suposem $f \in \mathcal{C}^4([a - h, a + h])$, obtenim una expressió de l'error fent servir

$$f(a + h) = f(a) + f'(a)h + \frac{f^{(2)}(a)}{2}h^2 + \frac{f^{(3)}(a)}{3!}h^3 + \frac{f^{(4)}(\xi_1)}{4!}h^4 ,$$

$$f(a-h) = f(a) - f'(a)h + \frac{f^{(2)}(a)}{2}h^2 - \frac{f^{(3)}(a)}{3!}h^3 + \frac{f^{(4)}(\xi_2)}{4!}h^4 ,$$

on $\xi_1 \in \langle a, a+h \rangle$ i $\xi_2 \in \langle a-h, h \rangle$.

Sumant-les s'obté

$$f^{(2)}(a) = \frac{f(a+h) - 2f(a) + f(a-h)}{h^2} + R(h) ,$$

amb

$$R(h) = -\frac{f^{(4)}(\xi_1) + f^{(4)}(\xi_2)}{24}h^2 ,$$

i així, $|R(h)| \leq \frac{M}{12}h^2$ si $|f^{(4)}(x)| \leq M$ per a $x \in \langle a-h, a+h \rangle$; és a dir,

$$R(h) = \mathcal{O}(h^2) \quad (h \rightarrow 0) . \quad (4.1)$$

Ara tindrem en compte el resultat general següent que serà utilitzat sovint:

- Si F és una funció real i contínua sobre un interval I , donades les abscisses de I , $\xi_1, \xi_2, \dots, \xi_m$ i els coeficients positius $\alpha_1, \alpha_2, \dots, \alpha_m$, existeix $\xi \in \langle \xi_1, \xi_2, \dots, \xi_m \rangle$ tal que

$$\sum_{k=1}^m \alpha_k F(\xi_k) = F(\xi) \sum_{k=1}^m \alpha_k$$

(teorema del valor mitjà per a sumes).

Aplicant aquesta propietat general a la funció $f^{(4)}$, s'obté la següent expressió per a l'error

$$R(h) = -\frac{f^{(4)}(\xi)}{12}h^2 , \quad \xi \in \langle a-h, a+h \rangle . \quad (4.2)$$

No sempre podrem obtenir una expressió explícita per a $R(h)$ com en (4.2), i tot sovint ens haurem de conformar amb una expressió asimptòtica com en (4.1).

Suposant $f \in \mathcal{C}^6([a-h, a+h])$, podríem obtenir

$$R(h) = -\frac{f^{(4)}(a)}{12}h^2 + \mathcal{O}(h^4) \quad (h \rightarrow 0) ,$$

$$R(h) = -\frac{f^{(4)}(a)}{12}h^2 - \frac{2f^{(6)}(\xi)}{6!}h^4 , \quad \xi \in \langle a-h, a+h \rangle ,$$

i també desenvolupaments anàlegs, si f fos encara més derivable.

Observacions

1. L'error de les fórmules de derivació numèrica trobades depèn de termes en h , h^2 , etc, on $h = \max |a - x_k|$; per tant, decreix en decreixent h . Això no vol dir que convé escollir h molt petit, ja que en les fórmules esmentades es produeix aleshores una notable cancel·lació de termes (vegeu el problema IV.15). En aquesta situació convindrà prendre h prudencialment petit i fer ús del mètode de Richardson d'extrapolació repetida, tal com veurem en l'apartat 4.4.2.

2. La idea fonamental emprada ací és que si $p(x)$ és propera a $f(x)$ en a , aleshores $p^{(d)}(x)$ serà propera a $f^{(d)}$ en a . Malhauradament amb molta freqüència això no és cert, puix que dues funcions poden ser properes i tenir pendents molt diferents (vegeu la figura 4.1).

3. A l'apartat 4.5.3, usant operadors, s'exposarà una manera sistemàtica de generació de fórmules generals de derivació numèrica.

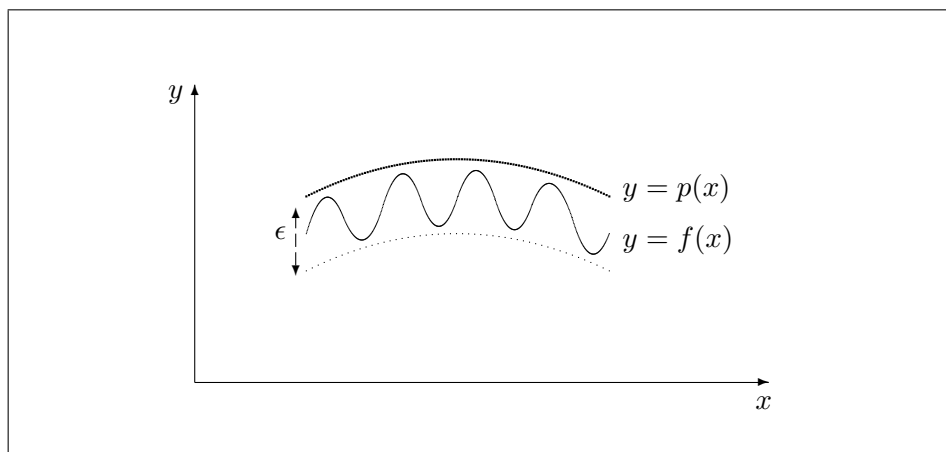


Figura 4.1: Funcions properes amb derivades allunyades.

4.2 INTEGRACIÓ NUMÈRICA

4.2.1 Introducció

Donada una funció f definida sobre un interval fitat $[a, b]$, volem ara calcular

$$J(f) = \int_a^b f(x) dx ,$$

suposant que aquesta integral tingui sentit per a la funció f (per exemple, si f és contínua sobre $[a, b]$). La *quadratura* o *integració numèrica* consisteix a donar fórmules aproximades per al càlcul de la integral $J(f)$ de f ; aquestes fórmules poden ser de gran utilitat quan la integral no es pot calcular per mètodes analítics (per exemple, usant la regla de Barrow), o bé, quan no convé usar-los perquè resulten complexos de fer servir, i ens conformem a conèixer $J(f)$ amb una precisió finita donada.

A tal fi, aproximarem f per un polinomi interpolador $p(x)$ i calcularem exactament $J(p)$, obtenint així fórmules d'integració numèrica.

4.2.2 Integració amb abscisses donades

Fórmules d'integració interpolatòria

Siguin $a \leq x_0 < x_1 < \dots < x_m \leq b$ $m + 1$ abscisses del segment $[a, b]$, i considerem el polinomi $p_m(x)$ de grau menor o igual que m verificant

$$p_m(x_k) = f(x_k) \quad (k = 0 \div m) ;$$

aleshores aproximem $J(f)$ per $J(p_m)$

$$\int_a^b f(x)dx \simeq \int_a^b p_m(x)dx .$$

Així, integrant la fórmula d'interpolació de Lagrange

$$p_m(x) = \sum_{k=0}^m f_k l_k(x) , \quad \text{on } l_k(x) = \prod_{i \neq k} \frac{x - x_i}{x_k - x_i} ,$$

trobem la fórmula d'integració numèrica

$$\int_a^b f(x)dx \simeq \sum_{k=0}^m W_k f_k , \quad W_k = \int_a^b l_k(x)dx \quad (k = 0 \div m) . \quad (4.3)$$

Com que la dita fórmula es troba per integració d'un polinomi interpolador, rep el nom de *fórmula d'integració interpolatòria de $m + 1$ abscisses*.

Els coeficients W_k ($k = 0 \div m$), anomenats *pesos* de la fórmula d'integració, depenen de l'interval $[a, b]$ i de les abscisses $x_0, \dots, x_m$, però no de f . Per la unicitat del polinomi interpolador, la fórmula (4.3) és exacta per a qualsevol polinomi de grau més petit o igual que m , la qual cosa ens proporciona una manera de calcular els pesos W_k ($k = 0 \div m$) sense necessitat d'integrar $l_k(x)$ ($k = 0 \div m$): s'imposa que la fórmula (4.3) sigui exacta per als monomis $1, x, x^2, \dots, x^m$ (per exemple), i es resol el sistema lineal resultant (*mètode dels coeficients indeterminats*).

Exemple: fórmula de Simpson

Volem trobar els pesos d'integració w_{-1} , w_0 i w_1 per tal que la fórmula d'integració numèrica

$$\int_{-1}^1 g(t)dt \simeq w_{-1}g_{-1} + w_0g_0 + w_1g_1$$

sigui exacta per a tots els polinomis de grau més petit o igual que 2, on $g_k = g(k)$ ($k = -1, 0, 1$); després, volem calcular l'error comès en aplicar-la a polinomis de grau més petit o igual que 4.

Imposant la exactitud de la fórmula per a $g(t) = 1, t, t^2$, s'obté el sistema

$$\left. \begin{array}{rrcr} w_{-1} & + & w_0 & + & w_1 & = & 2 \\ -w_{-1} & & & + & w_1 & = & 0 \\ w_{-1} & & & + & w_1 & = & \frac{2}{3} \end{array} \right\} ,$$

de solució $w_1 = w_{-1} = \frac{1}{3}$, $w_0 = \frac{4}{3}$, donant lloc a la fórmula d'integració numèrica

$$\int_{-1}^1 g(t)dt \simeq \frac{1}{3}(g_{-1} + 4g_0 + g_1) . \quad (4.4)$$

Aquesta fórmula es pot traslladar a qualsevol interval $[a, b]$, via el canvi de variables

$$t = 2\frac{x-a}{b-a} - 1 \quad \text{o, equivalentment,} \quad x = \frac{b-a}{2}t + \frac{a+b}{2} ,$$

donant lloc a la *fórmula de Simpson*

$$\int_a^b f(x)dx \simeq \frac{b-a}{6} \left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right] , \quad (4.5)$$

que també s'escriu com

$$\int_{c-h}^{c+h} f(x)dx \simeq \frac{h}{3} [f(c-h) + 4f(c) + f(c+h)] , \quad (4.6)$$

on $c = \frac{a+b}{2}$ i $h = \frac{b-a}{2}$.

Sobre els polinomis de grau 3, les fórmules de Simpson (4.5) i (4.6) resulten ser també exactes: si $g(t) = t^3$, ambdós membres de (4.4) són nuls.

Si prenem $g(t) = t^4$ en (4.4), obtenim

$$\int_{-1}^1 g(t)dt - \frac{1}{3}(g_{-1} + 4g_0 + g_1) = \frac{2}{5} - \frac{2}{3} = -\frac{4}{15} = -\frac{24}{90} = -\frac{g^{(4)}(\xi)}{90} \quad (4.7)$$

i, prenent anàlogament $f(x) = x^4$ en 4.5,

$$\int_{c-h}^{c+h} f(x)dx - \frac{h}{3} (f(c-h) + 4f(c) + f(c+h)) = -\frac{f^{(4)}(\xi)}{90} h^5 . \quad (4.8)$$

Les abscisses ξ que apareixen en (4.7) i (4.8) són arbitràries, atès que $g^{(4)}$ i $f^{(4)}$ són funcions constants.

Error de les fórmules d'integració interpolatòria

Retornant a la fórmula (4.3), observem que l'error d'aquesta és la integral de l'error d'interpolació; si $f \in \mathcal{C}^{m+1}([a, b])$, prenent l'expressió de l'error d'interpolació (3.2), s'obté

$$\begin{aligned} E_m &= \int_a^b f(x)dx - \sum_{k=0}^m W_k f_k \\ &= \int_a^b \frac{f^{(m+1)}(\xi(x))}{(m+1)!} (x-x_0) \cdots (x-x_m) dx , \end{aligned}$$

amb $\xi(x) \in (a, b)$.

Aquesta fórmula dona l'expressió general per a l'*error de la fórmula d'integració interpolatòria de $m+1$ abscisses*.

Si $|f^{(m+1)}(x)| \leq M_{m+1}$ per a tot $x \in [a, b]$, podrem escriure

$$|E_m| \leq \frac{M_{m+1}}{(m+1)!} \int_a^b |(x-x_0) \cdots (x-x_m)| dx .$$

Una manera d'obtenir fites més precises consisteix a desenvolupar per Taylor les fórmules de quadratura donades, tal com es fa en l'apartat que segueix.

Error de la fórmula de Simpson

Continuem amb la fórmula de Simpson i definim

$$E_S(h) = \int_{c-h}^{c+h} f(x)dx - \frac{h}{3}[f(c-h) + 4f(c) + f(c+h)] ;$$

aleshores $E_S(0) = E'_S(0) = E_S^{(2)}(0) = 0$ i

$$E_S^{(3)}(h) = -\frac{h}{3}[f^{(3)}(c+h) - f^{(3)}(c-h)] .$$

Si $f \in \mathcal{C}^4([c-h, c+h])$, prenem

$$F(h) = \begin{cases} \frac{f^{(3)}(c+h)-f^{(3)}(c-h)}{2h} & (\text{si } h \neq 0) , \\ f^{(4)}(c) & (\text{si } h = 0) . \end{cases}$$

Resulta que F és contínua, ja que $F(h) \rightarrow F(0)$ (quan $h \rightarrow 0$); a més, pel teorema del valor mitjà, per a qualsevol $\xi \in (0, h]$, $F(\xi) = f^{(4)}(\eta)$, per a algun $\eta \in (c-h, c+h)$.

Usant l'expressió integral de l'error d'interpolació de Taylor (3.8),

$$E_S(h) = \frac{1}{2} \int_0^h (h-t)^2 E_S^{(3)}(t) dt = -\frac{1}{3} \int_0^h (h-t)^2 t^2 F(t) dt .$$

Com que $G(t) = (h-t)^2 t^2$ és una funció contínua que no canvia de signe en $(0, h)$, el teorema del valor mitjà per a integrals ens permet treure $F(t)$ fora del signe integral i així,

$$E_S(h) = -\frac{1}{3} F(\xi) \int_0^h (h-t)^2 t^2 dt = -\frac{f^{(4)}(\eta)}{90} h^5 , \quad \eta \in (c-h, c+h) ,$$

que justament coincideix amb l'expressió (4.8).

Fórmules del rectangle i del trapezi

Prenent $h = b-a$, anàlogament a la fórmula de Simpson, obtenim: la *fórmula del rectangle*

$$\int_a^b f(x)dx = hf\left(\frac{a+b}{2}\right) + \frac{f^{(2)}(\xi)}{24} h^3 , \quad \xi \in (a, b) ,$$

i la *fórmula del trapezi*

$$\int_a^b f(x)dx = \frac{h}{2}[f(a) + f(b)] - \frac{f^{(2)}(\xi)}{12} h^3 , \quad \xi \in (a, b) , \quad (4.9)$$

fent servir interpolació només en l'abscissa mitjana i en les abscisses extremes, respectivament.

Per a una representació gràfica, vegeu la figura 4.2.

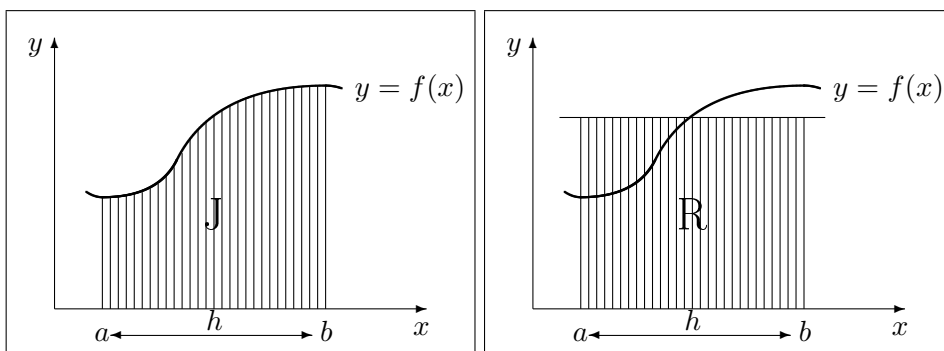


Figura 4.2: L'àrea ratllada J indica el valor de la integral. Les àrees de les zones ratllades en R , T i S corresponen als valors donats per les fórmules del rectangle, trapezi i Simpson, respectivament.

Fórmules de Newton-Cotes

Les fórmules del trapezi i de Simpson són un cas particular de les dites *fórmules (tancades) de Newton-Cotes* que es comenten tot seguit.

Si prenem $m + 1$ abscisses equidistants a l'interval $[a, b]$

$$x_k = a + kh \quad (k = 0 \div m), \quad h = \frac{b - a}{m}$$

en (4.3), obtenim la *fórmula de Newton-Cotes de $m + 1$ abscisses*:

$$\int_a^b f(x) dx \simeq h \sum_{k=0}^m \alpha_k f_k, \quad \alpha_k = \int_0^m \prod_{i \neq k} \frac{t - i}{k - i} dt \quad (k = 0 \div m),$$

on s'ha usat la notació $f_k = f(a + kh)$ ($k = 0 \div m$).

Els coeficients α_k ($k = 0 \div m$) no depenen ni de l'interval $[a, b]$ ni de la funció f ; només depenen del grau m . L'expressió de l'error de la fórmula de Newton-Cotes de $m + 1$ abscisses ve donada per

$$E_m = \int_a^b f(x) dx - h \sum_{k=0}^m \alpha_k f_k = K_m \frac{f^{(p+1)}(\xi)}{(p+1)!} h^{p+2}, \quad \xi \in (a, b), \quad (4.10)$$

on

$$\begin{aligned} K_m &= \int_0^m \pi_m(t) dt, \quad p = m \quad (\text{si } m \text{ senar}), \\ K_m &= \int_0^m t \pi_m(t) dt, \quad p = m + 1 \quad (\text{si } m \text{ parell}); \end{aligned}$$

essent $\pi_m(t) = t(t-1) \cdots (t-m)$.

Així, les fórmules de Newton-Cotes de $m + 1$ abscisses són exactes, fins i tot, per a tots els polinomis de grau $m + 1$, quan m és parell (o equivalentment, quan el nombre d'abscisses és senar) tal com succeïa amb la fórmula de Simpson ($m = 2$) en l'exemple anterior.

Notem que K_m es pot obtenir també aplicant (4.10) a un polinomi de grau $m + 1$ (resp. $m + 2$) si m és senar (resp. parell), ja que aleshores $f^{(p+1)}$ és constant.

Altres fórmules d'integració interpolatòries

Les fórmules interpolatòries que s'acaben de trobar partien només del polinomi interpolador a la funció que s'integra. Unes altres fórmules d'integració numèrica es poden deduir fent servir tipus diferents d'interpolació; això és, interpolació de Taylor, interpolació d'Hermite i, fins i tot, interpolació d'Hermite generalitzada. Alguns exemples d'aquestes fórmules i llurs aplicacions es poden trobar en els problemes IV.4 i IV.5.

4.2.3 Regles (compostes) d'integració numèrica

Les fórmules d'integració numèrica no s'apliquen normalment sobre tot l'interval $[a, b]$, sinó sobre subinterval·ls de $[a, b]$, donant lloc així a les *regles (compostes) d'integració numèrica*.

Regla dels trapezis

Si dividim $[a, b]$ en M parts iguals i, en cada una d'elles, apliquem la fórmula del trapezi, obtenim la *regla dels trapezis*

$$T(h) = \frac{h}{2} [f(a) + 2f(a+h) + 2f(a+2h) + \cdots + 2f(b-h) + f(b)] , \quad (4.11)$$

obtinguda en descompondre la integral inicial com la suma de les integrals en les M parts d'amplada $h = \frac{b-a}{M}$ en què s'ha dividit l'interval $[a, b]$:

$$J(f) = \int_a^b f(x)dx = \sum_{k=0}^{M-1} \int_{x_k}^{x_{k+1}} f(x)dx = \sum_{k=0}^{M-1} J_k(f) ,$$

amb $x_k = a + kh$ ($k = 0 \div M$).

Com que

$$J_k(f) = \int_{x_k}^{x_{k+1}} f(x)dx = \frac{h}{2} [f(x_k) + f(x_{k+1})] - \frac{f^{(2)}(\xi_k)}{12} h^3 ,$$

on $\xi_k \in (x_k, x_{k+1})$, aleshores

$$J(f) - T(h) = -\frac{1}{12} \sum_{k=0}^{M-1} f^{(2)}(\xi_k) h^3 = -\frac{b-a}{12M} \sum_{k=0}^{M-1} f^{(2)}(\xi_k) h^2 ,$$

on suposem que $f \in \mathcal{C}^2([a, b])$; tenint ara en compte el teorema del valor mitjà per a sumes, existeix $\xi \in (a, b)$ tal que

$$\int_a^b f(x)dx - T(h) = -\frac{b-a}{12} f^{(2)}(\xi) h^2 . \quad (4.12)$$

Regla de Simpson

Dividint ara $[a, b]$ en $2M$ parts iguals i aplicant, en cada interval de longitud $\frac{b-a}{M} = 2h$, la fórmula de Simpson, obtenim la *regla de Simpson*,

$$S(h) = \frac{h}{3} [f(a) + 4f(a+h) + 2f(a+2h) + 4f(a+3h) + \dots + 2f(b-2h) + 4f(b-h) + f(b)] ,$$

amb el següent error associat, suposant $f \in \mathcal{C}^4([a, b])$,

$$\int_a^b f(x)dx - S(h) = -\frac{b-a}{180} f^{(4)}(\xi) h^4 , \quad (4.13)$$

on $\xi \in (a, b)$.

Fórmula d'Euler-Maclaurin

Una manera força útil d'obtenir fórmules d'integració numèrica és per mitjà de la integració per parts. Així, per exemple, escrivim la integral

$$J(f) = \int_a^b f(x)dx$$

com

$$\int_a^b f(x)b_1'(x)dx ,$$

on $b_1(x)$ és un polinomi de grau 1 tal que $b_1'(x) = 1$ ($b_1(x) = x + c_1$); aleshores

$$J(f) = f(x)b_1(x) \Big|_a^b - \int_a^b f'(x)b_1(x)dx .$$

Si ara introduïm $b_2(x)$ tal que $b_2'(x) = b_1(x)$ ($b_2(x) = \frac{(x+c_1)^2}{2} + c_2$), resulta

$$J(f) = f(x)b_1(x) \Big|_a^b - f'(x)b_2(x) \Big|_a^b + \int_a^b f^{(2)}(x)b_2(x)dx . \quad (4.14)$$

Si prenem $c_1 = -\frac{a+b}{2}$, $c_2 = -\frac{(b-a)^2}{8}$, aleshores $b_2(x) = \frac{1}{2}(x-a)(x-b)$ s'anul·la sobre a i b i obtenim

$$J(f) = \frac{b-a}{2} [f(a) + f(b)] + \frac{1}{2} \int_a^b f^{(2)}(x)(x-a)(x-b)dx .$$

Com que $(x-a)(x-b) < 0$ sobre (a, b) , si $f \in \mathcal{C}^2([a, b])$ retrobem la fórmula del trapezi amb l'error corresponent (4.12):

$$\begin{aligned} J(f) - \frac{b-a}{2} [f(a) + f(b)] &= \frac{1}{2} \int_a^b f^{(2)}(x)(x-a)(x-b)dx \\ &= -\frac{f^{(2)}(\xi)}{12} (b-a)^3 . \end{aligned}$$

El procediment de (4.14) es pot iterar per tal d'obtenir una expressió asimptòtica per a l'error de la regla dels trapezis. Per simplificar, podem considerar que l'interval $[a, b]$ sigui l'interval $[0, 1]$; més concretament, fent el canvi $x = a + th$ i introduint $g(t) = f(a + th)$, tenim

$$\int_a^b f(x)dx = h \int_0^1 g(t)dt ,$$

amb $h = b - a$.

Prenem $b_1(t) = t - \frac{1}{2}$ com abans (ara l'abscissa mitjana és $\frac{1}{2}$) i $b_2(t), b_3(t), \dots$, de manera que

$$b'_{j+1}(t) = b_j(t) \quad (j \geq 1) , \quad (4.15)$$

i amb les constants adequades per tal que

$$\int_0^1 b_j(t)dt = 0 \quad (j \geq 1) \quad (4.16)$$

o, equivalentment, $b_{j+1}(0) = b_{j+1}(1)$ ($j \geq 1$).

Repetides integracions per parts donen lloc a

$$\begin{aligned} \int_0^1 g(t)dt &= \frac{1}{2}(g(0) + g(1)) \\ &\quad - b_2(0)(g'(1) - g'(0)) + b_3(0)(g^{(2)}(1) - g^{(2)}(0)) + \dots \\ &\quad + (-1)^{(n+1)}b_n(0)(g^{(n-1)}(1) - g^{(n-1)}(0)) \\ &\quad + (-1)^{(n+2)} \int_0^1 g^{(n)}(t)b_n(t)dt , \end{aligned} \quad (4.17)$$

per a $g \in \mathcal{C}^n([0, 1])$.

En el problema III.11, es demostra que

$$(-1)^{r+1}b_{2r}(0) > 0 , \quad b_{2r+1}(0) = 0 \quad (r \geq 1) , \quad (4.18)$$

$$(-1)^r(b_{2r}(t) - b_{2r}(0)) > 0 \quad \forall t \in (0, 1) \quad (r \geq 1) . \quad (4.19)$$

Prenent ara $n = 2s + 1$ i integrant per parts l'últim sumand de (4.17) de signe negatiu, s'obté

$$\begin{aligned} \int_0^1 g^{(2s+1)}(t)b_{2s+1}(t)dt &= [b_{2s+2}(t) - b_{2s+2}(0)]g^{(2s+1)}(t) \Big|_0^1 \\ &\quad - \int_0^1 (b_{2s+2}(t) - b_{2s+2}(0))g^{(2s+2)}(t)dt \\ &= b_{2s+2}(0)g^{(2s+2)}(\xi) , \quad \xi \in (0, 1) , \end{aligned}$$

on s'ha emprat (4.15), (4.16) i (4.19).

Posem la fórmula (4.17) en termes dels polinomis *mònics* $B_0(t) = 1$, $B_j(t) = j!b_j(t) = t^j + \dots$ ($j \geq 1$) (això és, polinomis amb coeficients principals unitaris). Aquests reben el

nom de *polinomis de Bernoulli* i els valors $B_j \equiv B_j(0)$ ($j \geq 0$) són anomenats *nombres de Bernoulli* (vegeu de nou el problema III.11 per generar-los):

$$\int_0^1 g(t)dt = \frac{1}{2}(g(0) + g(1)) \quad (4.20)$$

$$\begin{aligned} & - \sum_{r=1}^s \frac{B_{2r}}{(2r)!} [g^{(2r-1)}(1) - g^{(2r-1)}(0)] \\ & - \frac{B_{2s+2}}{(2s+2)!} g^{(2s+2)}(\xi) , \quad \xi \in (0, 1) , \end{aligned} \quad (4.21)$$

fórmula vàlida per a tota $g \in \mathcal{C}^{2s+2}([0, 1])$.

Desfent el canvi $g(t) = f(a+th)$ i reordenant els termes, obtenim l'*expressió asimptòtica de l'error de la fórmula del trapezi*

$$\begin{aligned} \frac{h}{2}[f(a) + f(a+h)] &= \int_a^{a+h} f(x)dx \\ &+ \sum_{r=1}^s h^{2r} \frac{B_{2r}}{(2r)!} [f^{(2r-1)}(a+h) - f^{(2r-1)}(a)] \\ &+ \frac{B_{2s+2}}{(2s+2)!} f^{(2s+2)}(\xi) h^{2s+3} , \quad \xi \in (a, a+h) , \end{aligned} \quad (4.22)$$

fórmula vàlida per a $f \in \mathcal{C}^{2s+2}([a, a+h])$.

Per a una subdivisió de l'interval $[a, b]$ en M parts iguals, i aplicant en cada segment $[x_k, x_{k+1}]$ la fórmula del trapezi amb $h = \frac{b-a}{M}$ ($k = 0 \div M-1$), s'obté de manera anàloga al procediment de (4.12) l'*expressió asimptòtica de l'error de la regla dels trapezis*

$$\begin{aligned} T(h) &= \int_a^b f(x)dx + \sum_{r=1}^s h^{2r} \frac{B_{2r}}{(2r)!} [f^{(2r-1)}(b) - f^{(2r-1)}(a)] \\ &+ \frac{B_{2s+2}}{(2s+2)!} (b-a) f^{(2s+2)}(\xi) h^{2s+2} , \quad \xi \in (a, b) , \end{aligned} \quad (4.23)$$

vàlida per a $f \in \mathcal{C}^{2s+2}([a, b])$.

Les fórmules (4.21), (4.22) i (4.23) reben el nom de *fórmules d'Euler-Maclaurin*. En aplicar-les, per exemple la (4.23), convé saber fitar la resta

$$R_s = \frac{B_{2s+2}}{(2s+2)!} (b-a) f^{(2s+2)}(\xi) h^{2s+2} , \quad \xi \in (a, b) .$$

Clarament,

$$|R_s| \leq \frac{|B_{2s+2}|}{(2s+2)!} (b-a) \sup_{x \in [a, b]} |f^{(2s+2)}(x)| h^{2s+2} .$$

Ara bé, segons (4.18) els nombres de Bernoulli alternen el signe (això és, $B_{2s}B_{2s+2} < 0$ ($s \geq 0$)). Si $f^{(2s+2)}$ i $f^{(2s+4)}$ no s'anul·len en (a, b) i tenen el mateix signe, aleshores $R_s R_{s+1} < 0$ i, per tant, $|R_s|$ és més petit que el valor absolut del primer terme menyspreat, segons el criteri d'alternança dels residus de l'apartat 3.1.4,

$$|R_s| \leq \frac{|B_{2s+2}|}{(2s+2)!} |f^{(2s+1)}(b) - f^{(2s+1)}(a)| h^{2s+2} .$$

Cal advertir que els nombres de Bernoulli creixen ràpidament i, per tant, no cal esperar que $R_s \rightarrow 0$ quan $s \rightarrow \infty$. A la pràctica, convé prendre el nombre de termes s en (4.23) de manera que $|R_s|$ sigui menor que la tolerància ϵ demanada, sempre que això sigui possible.

4.2.4 Integració gaussiana

Les fórmules d'integració interpolatòria de $m + 1$ abscisses $x_0, x_1, \dots, x_m$, obtingudes integrant el polinomi interpolador en aquestes abscisses, són exactes per als polinomis de grau més petit o igual que m ; això passa per a qualsevol tria feta de les abscisses dins de l'interval d'integració. Veurem ara que una tria adequada d'aquestes $m + 1$ abscisses ens proporcionarà fórmules d'integració numèrica de $m + 1$ abscisses, exactes per a polinomis de grau més petit o igual que $2m + 1$, que rebran el nom de *fórmules gaussianes*.

Exemple motivador

La regla dels trapezis ens proporcionarà el primer exemple d'aquestes fórmules, en ser aplicada convenientment sobre polinomis trigonomètrics.

Així, sigui

$$t_n(\theta) = \frac{a_0}{2} + \sum_{j=1}^n a_j \cos j\theta + \sum_{j=1}^n b_j \sin j\theta = \sum_{j=-n}^n c_j e^{ij\theta}$$

un polinomi trigonomètric de grau més petit o igual que n , i calculem

$$J(t_n) = \int_0^{2\pi} t_n(\theta) d\theta = \pi a_0 = 2\pi c_0 ,$$

usant la regla dels trapezis amb pas $h = \frac{2\pi}{M}$ per a $M > n$.

Atesa la periodicitat de $t_n(\theta)$,

$$t_n(\theta + 2\pi) = t_n(\theta) \quad \forall \theta \in \mathbb{R} ;$$

per a qualsevol $\phi \in \mathbb{R}$ es compleix

$$J(t_n) = \int_{\phi}^{2\pi+\phi} t_n(\varphi) d\varphi = \int_0^{2\pi} t_n(\phi + \theta) d\theta = \pi a_0 .$$

L'aplicació de la regla dels trapezis sobre aquesta última integral dona el resultat exacte

$$\begin{aligned} T\left(\frac{2\pi}{M}\right) &= \frac{2\pi}{M} \sum_{k=0}^{M-1} t_n\left(\phi + \frac{2\pi k}{M}\right) \\ &= \frac{2\pi}{M} \sum_{j=-n}^n c_j e^{ij\phi} \sum_{k=0}^{M-1} e^{i\frac{2\pi jk}{M}} = 2\pi c_0 = J(t_n) , \end{aligned}$$

on hem usat que

$$\sum_{k=0}^{M-1} e^{i\frac{2\pi jk}{M}} = \begin{cases} M & (j = 0) , \\ 0 & (0 < |j| \leq n < M) , \end{cases}$$

sabent que $e^{i\frac{2\pi j}{M}} \neq 1$ ($0 < |j| \leq n < M$).

Si ara considerem un polinomi trigonomètric en cosinus de grau més petit o igual que n

$$t_n(\theta) = \frac{a_0}{2} + \sum_{j=1}^n a_j \cos j\theta ,$$

tenim, apart de la periodicitat $t_n(\theta) = t_n(\theta + 2\pi)$, la simetria respecte a $\theta = \pi$: $t_n(\theta) = t_n(2\pi - \theta)$; per tant, si $M > n$,

$$\begin{aligned} J(t_n) &= \int_0^\pi t_n(\theta) d\theta = \frac{1}{2} \int_0^{2\pi} t_n(\theta) d\theta = \frac{\pi}{2} a_0 \\ &= \frac{1}{2} T\left(\frac{2\pi}{M}\right) = \frac{\pi}{M} \sum_{k=0}^{M-1} t_n\left(\phi + \frac{2\pi k}{M}\right) . \end{aligned}$$

Prenem ara $M = 2(m+1) > n$ i $\phi = \frac{\pi}{M}$ per tal que el conjunt d'abscisses sigui simètric respecte a π . Aleshores $\phi + \frac{2\pi k}{M} = \frac{2k+1}{m+1} \frac{\pi}{2}$ ($k = 0 \div 2m+1$) i els valors $t_n\left(\frac{(2k+1)\pi}{2(m+1)}\right)$ apareixen dues vegades: si $l = 2m+1-k$, quan $k = 0 \div m$, tenim que $l = 2m+1 \div m+1$, i

$$t_n\left(\frac{(2k+1)\pi}{2(m+1)}\right) = t_n\left(2\pi - \frac{(2k+1)\pi}{2(m+1)}\right) = t_n\left(\frac{(2l+1)\pi}{2(m+1)}\right) .$$

Obtenim així la següent fórmula d'integració numèrica de $m+1$ abscisses:

$$\int_0^\pi F(\theta) d\theta \simeq \frac{\pi}{m+1} \sum_{k=0}^m F\left(\frac{(2k+1)\pi}{2(m+1)}\right) ,$$

exacta per als polinomis trigonomètrics en cosinus de grau més petit o igual que $2m+1$.

Fent ara el canvi $t = \cos \theta$, usual en l'aproximació emprant polinomis de Txebeixev (vegeu l'apartat 3.2.4), aleshores $f(t) = F(\arccos t)$ és un polinomi de grau més petit o igual que n si F és un polinomi trigonomètric en cosinus de grau més petit o igual que n (recordeu la fórmula de Moivre), i tenim la següent fórmula d'integració numèrica de $m+1$ abscisses:

$$\int_{-1}^1 \frac{f(t)}{\sqrt{1-t^2}} dt \simeq \frac{\pi}{m+1} \sum_{k=0}^m f\left(\cos \frac{(2k+1)\pi}{2(m+1)}\right) , \quad (4.24)$$

exacta per als polinomis de grau més petit o igual que $2m+1$, anomenada *fórmula de Gauss-Txebeixev* (vegeu un exemple de la seva aplicació en el problema IV.7).

Les fórmules anteriors també es poden escriure com:

$$\begin{aligned} \int_0^\pi F(\theta) d\theta &\simeq \frac{\pi}{m+1} \sum_{k=0}^m F(\theta_k) , \\ \int_{-1}^1 \frac{f(t)}{\sqrt{1-t^2}} dt &\simeq \frac{\pi}{m+1} \sum_{k=0}^m f(t_k) , \end{aligned}$$

on θ_k ($k = 0 \div m$) són els zeros de $\psi_{m+1}(\theta) = \cos((m+1)\theta)$ i t_k ($k = 0 \div m$), els de funció $T_{m+1}(t) = \cos((m+1) \arccos t) = \psi_{m+1}(\arccos t)$. Recordem, de l'apartat 3.2.4, que $\psi_{m+1}(\theta)$ i $T_{m+1}(t)$ formen part de les famílies $\psi_j(\theta)$ ($j \geq 0$) i $T_j(t)$ ($j \geq 0$) ortogonals respecte als productes escalars

$$(F, G) = \int_0^\pi F(\theta)G(\theta)d\theta, \quad (f, g) = \int_{-1}^1 \frac{f(t)g(t)}{\sqrt{1-t^2}}dt,$$

respectivament.

De l'exactitud de (4.24) per als polinomis de grau més petit o igual que $2m+1$, en resulta que $\psi_j(\theta) = \cos j\theta$ i $T_j(t) = \psi_j(\arccos t)$ ($j = 0 \div m$) són ortogonals respecte als productes escalars discrets

$$\begin{aligned}(F, G)_m &= \sum_{k=0}^m F(\theta_k)G(\theta_k), \\ (f, g)_m &= \sum_{k=0}^m f(t_k)g(t_k),\end{aligned}$$

tal com ja s'ha demostrat en el capítol III, amb

$$\begin{aligned}(\psi_j, \psi_j) = (T_j, T_j) &= \frac{\pi}{m+1}(\psi_j, \psi_j)_m = \frac{\pi}{m+1}(T_j, T_j)_m \\ &= \begin{cases} \pi & (j=0), \\ \frac{\pi}{2} & (j=1 \div m). \end{cases}\end{aligned}$$

Fórmules gaussianes

Veurem que l'elecció de les abscisses x_k ($k = 0 \div m$) com a zeros d'un polinomi $\psi_{m+1}(x)$, de grau $m+1$, d'una família de polinomis ortogonals portarà sempre a fórmules de quadratura exactes per als polinomis de grau més petit o igual que $2m+1$.

Així, sigui $w : [a, b] \rightarrow \mathbb{R}$ una funció pes positiva i contínua sobre l'interval $[a, b]$ i sigui $\psi_{m+1}(x) = A_{m+1}x^{m+1} + \dots$ el polinomi ortogonal de grau $m+1$, amb coeficient principal A_{m+1} , associat al producte escalar

$$(f, g) = \int_a^b w(x)f(x)g(x)dx.$$

Aquest polinomi $\psi_{m+1}(x)$ té $m+1$ zeros simples x_k ($k = 0 \div m$) (per tant, diferents dos a dos), que es troben a l'interval (a, b) .

En efecte, si $\psi_{m+1}(x)$ només canviés de signe en i abscisses $\alpha_1, \dots, \alpha_i$ de $[a, b]$, amb $1 \leq i \leq m$, aleshores el polinomi

$$q_i(x)\psi_{m+1}(x) \equiv (x - \alpha_1) \cdots (x - \alpha_i)\psi_{m+1}(x),$$

de grau $m+i+1$, no canviaria de signe sobre (a, b) i, per tant, la integral

$$\int_a^b w(x)q_i(x)\psi_{m+1}(x)dx = (q_i, \psi_{m+1})$$

seria no nul·la, en clara contradicció amb el fet que $\psi_{m+1}(x)$ és ortogonal a qualsevol polinomi de grau menor o igual que m .

Considerem ara la següent fórmula d'integració numèrica de $m + 1$ abscisses, avaluada sobre els zeros de $\psi_{m+1}(x)$:

$$\int_a^b w(x)f(x)dx \simeq \sum_{k=0}^m W_k f(x_k) . \quad (4.25)$$

D'entrada, imposant exactitud per als polinomis de grau més petit o igual que m , obtenim com en (4.3) que els pesos W_k vénen donats per

$$W_k = \int_a^b l_k(x)w(x)dx , \quad l_k(x) = \prod_{i \neq k} \frac{x - x_i}{x_k - x_i} \quad (k = 0 \div m) . \quad (4.26)$$

Comprovarem ara que, amb aquesta elecció d'abscisses i pesos, aquesta fórmula és exacta també per als polinomis de grau més petit o igual que $2m + 1$. Les fórmules així obtingudes s'anomenen *fórmules gaussianes de $m + 1$ abscisses*.

Sigui $p_{2m+1}(x)$ un polinomi de grau més petit o igual que $2m + 1$ i siguin $q_m(x)$ i $r_m(x)$ els polinomis de grau més petit o igual que m que són respectivament el quocient i el residu obtinguts en fer la divisió de $p_{2m+1}(x)$ pel polinomi $\psi_{m+1}(x)$ de grau $m + 1$

$$p_{2m+1}(x) = q_m(x)\psi_{m+1}(x) + r_m(x) .$$

El polinomi $q_m(x)$ serà, doncs, ortogonal a $\psi_{m+1}(x)$: $(q_m, \psi_{m+1}) = 0$; per tant,

$$\begin{aligned} \int_a^b w(x)p_{2m+1}(x)dx &= \int_a^b w(x)q_m(x)\psi_{m+1}(x) \\ &\quad + \int_a^b w(x)r_m(x)dx \\ &= \int_a^b w(x)r_m(x)dx = \sum_{k=0}^m W_k r_m(x_k) , \end{aligned}$$

ja que (4.25) és exacta per al polinomi $r_m(x)$ de grau més petit o igual que m .

Usant ara que $\psi_{m+1}(x_k) = 0$ ($k = 0 \div m$), obtenim l'exactitud de (4.25) per a $p_{2m+1}(x)$:

$$\begin{aligned} \sum_{k=0}^m W_k p_{2m+1}(x_k) &= \sum_{k=0}^m W_k q_m(x_k)\psi_{m+1}(x_k) + \sum_{k=0}^m W_k r_m(x_k) \\ &= \int_a^b w(x)p_{2m+1}(x)dx . \end{aligned}$$

Error de les fórmules gaussianes

Per a funcions $f \in \mathcal{C}^{2m+2}([a, b])$ podem donar una expressió per a l'error de les fórmules gaussianes. Per això, considerem el polinomi interpolador d'Hermite $p_{2m+1}(x)$ a f en les abscisses x_k ($k = 0 \div m$) (de grau més petit o igual que $2m + 1$); d'una banda, tenim que la fórmula gaussiana és exacta per a aquest polinomi i, d'altra banda, disposem d'una fórmula d'error d'interpolació per a tot $x \in [a, b]$

$$f(x) - p_{2m+1}(x) = \frac{f^{(2m+2)}(\xi(x))}{(2m+2)!} \omega_m^2(x) ,$$

on $\xi(x) \in \langle x_0, x_1, \dots, x_m, x \rangle \subset [a, b]$ i

$$\omega_m(x) = (x - x_0)(x - x_1) \cdots (x - x_m) = \frac{\psi_{m+1}(x)}{A_{m+1}} .$$

Multiplicant per $w(x)$ i integrant entre a i b , s'obté l'expressió següent de l'error de la fórmula gaussiana de $m + 1$ abscisses:

$$\begin{aligned} \int_a^b w(x)f(x)dx - \sum_{k=0}^m W_k f(x_k) \\ = \frac{f^{(2m+2)}(\xi)}{(2m+2)!} \frac{1}{A_{m+1}^2} \int_a^b w(x)\psi_{m+1}^2(x)dx , \end{aligned} \quad (4.27)$$

on $\xi \in (a, b)$ i A_{m+1} indica el coeficient principal del polinomi ortogonal $\psi_{m+1}(x)$ triat.

Cal fer notar que el factor de l'error

$$\frac{(\psi_{m+1}, \psi_{m+1})}{A_{m+1}^2} = \int_a^b w(x)\omega_m^2(x)dx$$

es pot obtenir per aplicació de la fórmula gaussiana a $f(x) = x^{2m+2}$, atès que aleshores l'altre factor és la unitat

$$\frac{f^{(2m+2)}(\xi)}{(2m+2)!} = 1 .$$

Pesos de les fórmules gaussianes

Els pesos W_k , que es poden trobar per (4.26), admeten altres expressions equivalents.

Emprant l'exactitud de (4.25) per als polinomis $l_k^2(x)$ de grau $2m$, obtenim

$$W_k = \int_a^b w(x)l_k^2(x)dx > 0 \quad (k = 0 \div m) ,$$

d'on resulta que tots els pesos d'una fórmula gaussiana són positius.

Imposant ara l'exactitud per a

$$f(x) = \frac{\psi_{m+1}(x)}{x - x_k} = A_{m+1} \prod_{i \neq k} (x - x_i) ,$$

aleshores $f(x_i) = 0$ ($i \neq k$) i $f(x_k) = \psi'_{m+1}(x_k)$, donant lloc a

$$W_k = \frac{1}{\psi'_{m+1}(x_k)} \int_a^b w(x) \frac{\psi_{m+1}(x)}{x - x_k} dx \quad (k = 0 \div m) ;$$

anàlogament, per a

$$f(x) = \frac{\psi_{m+1}^2(x)}{(x - x_k)^2} = A_{m+1}^2 \prod_{i \neq k} (x - x_i)^2 ,$$

s'obté

$$W_k = \frac{1}{(\psi'_{m+1}(x_k))^2} \int_a^b w(x) \frac{\psi_{m+1}^2(x)}{(x - x_k)^2} dx > 0 \quad (k = 0 \div m) .$$

Finalment, si substituïm en la fórmula (4.27) el polinomi de grau $2m + 2$

$$f(x) = \frac{\psi_{m+1}(x)}{x - x_k} \psi_{m+2}(x) ,$$

i tenim en compte el fet que

$$\left(\frac{\psi_{m+1}}{x - x_k}, \psi_{m+2} \right) = 0 ,$$

arribem a l'expressió

$$W_k = - \frac{A_{m+2}(\psi_{m+1}, \psi_{m+1})}{A_{m+1}\psi'_{m+1}(x_k)\psi_{m+2}(x_k)} .$$

Usant, per acabar, la relació de recurrència dels polinomis ortogonals, obtenim

$$W_k = \frac{A_{m+1}(\psi_m, \psi_m)}{A_m\psi'_{m+1}(x_k)\psi_m(x_k)} .$$

Exemples de fórmules gaussianes

1. Tornem a la fórmula de Gauss-Txebixev (4.24). El polinomi ortogonal de grau $m + 1$ i coeficient principal $A_{m+1} = 2^m$ corresponent és el polinomi de Txebixev $T_{m+1}(t) = \cos((m + 1) \arccos t)$.

Els seus zeros $x_k \equiv t_k = \cos \theta_k$, $\theta_k = \frac{(2k+1)\pi}{2(m+1)}$ ($k = 0 \div m$) són les abscisses de la fórmula; llavors, els pesos de la fórmula de Gauss-Txebixev són

$$\begin{aligned} W_k &= - \frac{\pi}{T'_{m+1}(t_k)T_{m+2}(t_k)} = - \frac{\pi}{\frac{(-1)^k(m+1)}{\sin \theta_k}(-1)^{k+1} \sin \theta_k} \\ &= \frac{\pi}{m+1} \quad (k = 0 \div m) , \end{aligned}$$

tal com havíem vist; la fórmula amb l'error corresponent queda finalment així

$$\begin{aligned} \int_{-1}^1 \frac{f(t)}{\sqrt{1-t^2}} dt &= \frac{\pi}{m+1} \sum_{k=0}^m f\left(\cos \frac{(2k+1)\pi}{2(m+1)}\right) \\ &\quad + \frac{\pi}{2^{2m+1}(2m+2)!} f^{(2m+2)}(\xi) , \quad \xi \in (-1, 1) . \end{aligned}$$

2. Si considerem ara la funció pes $w(x) = 1$ a l'interval $[-1, 1]$, el polinomi ortogonal de grau $m + 1$ i coeficient principal

$$A_{m+1} = \frac{(2m+2)!}{2^{m+1}[(m+1)!]^2}$$

és el polinomi de Legendre

$$P_{m+1}(t) = \frac{1}{2^{m+1}(m+1)!} \frac{d^{m+1}}{dt^{m+1}} [(t^2 - 1)^{m+1}] ,$$

introduït en l'apartat 3.2.3. Els seus zeros $t_0, t_1, \dots, t_m$ estan distribuïts simètricament respecte a l'origen i la fórmula gaussiana, dita de *Gauss-Legendre*, que amb l'error corresponent queda (vegeu el problema IV.8)

$$\int_{-1}^1 f(t) dt = \sum_{k=0}^m W_k f(t_k) + \frac{2^{2m+3}[(m+1)!]^4}{(2m+3)[(2m+2)!]^3} f^{(2m+2)}(\xi) ,$$

amb $\xi \in (-1, 1)$ i

$$W_k = \frac{2}{(1 - t_k^2)[P'_{m+1}(t_k)]^2} \quad (k = 0 \div m) .$$

Aquesta fórmula de Gauss-Legendre pot ser estesa a qualsevol interval $[a, b]$, mitjançant el canvi

$$x = \frac{b-a}{2}t + \frac{a+b}{2} .$$

4.3 SUMACIÓ NUMÈRICA

4.3.1 Introducció

Considerem seguidament les sumes (finites)

$$S_n = \sum_{j=0}^n a_j ; \quad (4.28)$$

on els a_j ($j = 0 \div n$) són els $n + 1$ primers termes de la successió $(a_j)_{j \geq 0}$. Les sumes S_j ($j \geq 0$) s'anomenen *sumes parcials* de la dita successió.

Sovint els termes de la successió es podran escriure en la forma

$$a_j = f(j) \quad (j \geq 0) ,$$

essent $f : [0, \infty) \rightarrow \mathbb{R}$ (també podria prendre valors complexos).

La successió $(S_j)_{j \geq 0}$ rep el nom de *sèrie* i, si és convergent a un valor S , escriurem

$$S = \sum_{j=0}^{\infty} a_j$$

i direm que S és la *suma de la sèrie de terme general* a_j .

Això vol dir que, per a qualsevol $\epsilon > 0$, podem trobar un índex n , dependent de ϵ , tal que $|S - S_n| \leq \epsilon$; o també

$$S = S_n \pm \epsilon . \quad (4.29)$$

El problema de la *sumació numèrica* és donar mètodes per aproximar sumes del tipus (4.28) o, si existeix, el seu límit quan n tendeix a infinit. Des del punt de vista numèric, el cas límit es redueix a calcular S_n amb $n = n(\epsilon)$ prou gran a fi que es compleixi (4.29), donat un error permisible ϵ . Ara bé, si n és molt gran, el càlcul directe de (4.28) és molt costós i cal cercar mètodes alternatius.

També considerarem sumes els termes de les quals són funcions d'una variable real (o complexa)

$$S_n(x) = \sum_{j=0}^n f_j(x) ,$$

que donen lloc, en el cas límit $n \rightarrow \infty$, a les anomenades *sèries de funcions*

$$S(x) = \sum_{j=0}^{\infty} f_j(x) .$$

Aquestes sèries tenen sentit només per a aquells valors de x per als quals la sèrie numèrica corresponent és convergent. El conjunt d'abscisses en què la sèrie de funcions convergeix rep el nom de *regió de convergència*.

Un cas molt important és el de les *sèries de potències* del tipus

$$S(x) = \sum_{j=0}^{\infty} a_j (x - x_0)^j ,$$

que sorgeixen en fer desenvolupaments de Taylor d'una funció prop d'una abscissa x_0 . En efecte, si $f : I \rightarrow \mathbb{R}$ és una funció indefinidament diferenciable sobre un interval I , podem escriure, per a qualsevol $x_0 \in I$,

$$f(x) = p_n(x) + R_n(x) \quad (n \geq 0) ,$$

on

$$p_n(x) = \sum_{j=0}^n \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j$$

és el polinomi de Taylor de grau més petit o igual que n de f en x_0 i $R_n(x)$, l'error de la interpolació de Taylor que admet les expressions (3.8) i (3.9).

Recordem que, per a qualsevol n fixat,

$$\lim_{x \rightarrow x_0} \frac{|R_n(x)|}{|x - x_0|^n} = 0 .$$

La sèrie de potències formada per la successió de polinomis de Taylor d'una funció f indefinidament diferenciable prop de x_0 va rebre el nom de *desenvolupament en sèrie de Taylor de f prop de x_0* en l'apartat 3.1.4. De fet, es tracta d'un cas particular dels anomenats desenvolupaments asimptòtics que s'introdueixen a continuació.

Direm que

$$f(x) = \sum_{j=0}^n a_j (x - x_0)^j + R_n(x)$$

és un *desenvolupament asimptòtic d'una funció f prop de x_0* quan

$$\lim_{x \rightarrow x_0} \frac{|R_n(x)|}{|x - x_0|^n} = 0 .$$

En tal cas, escriurem

$$f(x) \sim \sum_{j=0}^{\infty} a_j (x - x_0)^j \quad (x \rightarrow x_0) .$$

De manera anàloga, direm que

$$f(x) = \sum_{j=0}^n \frac{a_j}{x^j} + R_n(x)$$

és un *desenvolupament asimptòtic de f prop d'infinít* i escriurem

$$f(x) \sim \sum_{j=0}^{\infty} \frac{a_j}{x^j} \quad (x \rightarrow \infty)$$

quan

$$\lim_{x \rightarrow \infty} |x|^n |R_n(x)| = 0.$$

Donada una abscissa x , el desenvolupament asimptòtic d'una funció f

$$f(x) = S_n(x) + R_n(x)$$

no és necessàriament convergent a $f(x)$, quan n tendeix a infinit; això és: els residus $R_n(x)$ no sempre convergeixen a 0.

En el cas dels desenvolupaments de Taylor, si es complia la condició de convergència

$$\lim_{n \rightarrow \infty} R_n(x) = 0 \quad \forall x \in (a, b),$$

dèiem que la funció f era analítica a l'interval (a, b) . Diversos exemples de funcions analítiques es troben a la taula 3.5.

Un desenvolupament asimptòtic pot ser emprat per al càlcul d'una funció, encara que no es compleixi la condició de convergència

$$\lim_{n \rightarrow \infty} R_n(x) = 0.$$

En aquest cas, l'error comès no serà arbitràriament petit.

Si es disposa d'una expressió per a cada $R_j(x)$ ($j \geq 0$), caldrà trobar un valor n per al qual $|R_n(x)|$ sigui menor que l'error permès. Això no sempre serà possible; en tal cas, interessarà trobar n que faci mínim $|R_n(x)|$ i ens haurem de conformar a tenir $f(x)$ calculat amb aquest error (vegeu el problema IV.11).

Les sèries divergents que apareixen en l'aplicació de desenvolupaments asimptòtics reben el nom de *sèries semiconvergentes* perquè permeten també el càlcul de les dites funcions, encara que no amb un error arbitràriament petit.

4.3.2 Fites dels residus de les sèries

Donada una sèrie convergent

$$S = \sum_{j=0}^{\infty} a_j,$$

on a_j pot ser eventualment del tipus $f(j)$ o $f_j(x)$, cal primer saber estimar els seus *residus*

$$R_n = S - S_n = \sum_{j=n+1}^{\infty} a_j$$

per poder aproximar correctament S mitjançant el càlcul directe de S_n ; si n resulta ser molt gran, s'haurà de cercar un altre mètode a fi d'aproximar S_n o S .

Per exemple, si

$$|a_{j+1}| \leq \rho |a_j| \quad (j \geq n),$$

amb $\rho < 1$; aleshores, per inducció,

$$|a_{j+l}| \leq \rho^l |a_j| \quad (l \geq 1) \quad (j \geq n)$$

i obtenim la fita del residu

$$|R_n| \leq \sum_{j=n+1}^{\infty} |a_j| \leq \left(\sum_{j=0}^{\infty} \rho^j \right) |a_{n+1}| = \frac{|a_{n+1}|}{1-\rho} ;$$

és a dir, la magnitud del residu és menor o igual que la magnitud del primer terme menyspreat dividida pel factor $1-\rho$.

Si, en canvi, tenim $a_j = f(j)$, amb $f: [0, \infty) \rightarrow \mathbb{R}$ decreixent (almenys sobre $[n, \infty)$); aleshores, si $x \in [j, j+1]$ ($j \geq n$), es té $f(j) \geq f(x) \geq f(j+1)$ i, per tant,

$$f(j) = \int_j^{j+1} f(j) dx \geq \int_j^{j+1} f(x) dx \geq f(j+1) .$$

Així, resulta

$$R_{n-1} = \sum_{j=n}^{\infty} f(j) \geq \int_n^{\infty} f(x) dx \geq R_n ;$$

en particular, s'assegura que el residu

$$\sum_{j=n+1}^{\infty} f(j)$$

és convergent si i només si la integral

$$\int_{n+1}^{\infty} f(x) dx$$

és convergent.

Recordem el criteri d'alternança dels residus de l'apartat 3.1.4: si els residus R_n i R_{n+1} tenen signes diferents, aleshores

$$S \in \langle S_n, S_{n+1} \rangle , \quad |R_n| \leq |a_{n+1}| ;$$

això és: la magnitud del residu és menor que la del primer terme menyspreat. En particular, si tenim una *sèrie alternada* a partir del terme n ,

$$a_j a_{j+1} < 0 \quad (j \geq n+1) ,$$

que sigui monòtona decreixent, en mòdul, cap a 0 a partir del terme n ,

$$|a_{j+1}| \leq |a_j| \quad (j \geq n+1) , \quad \lim_{j \rightarrow \infty} a_j = 0 ;$$

aleshores

$$\operatorname{sgn}(a_j + a_{j+1}) = \operatorname{sgn}(a_{j+2} + a_{j+3}) \quad (j \geq n+1)$$

i, per tant,

$$R_{j-1} = (a_j + a_{j+1}) + (a_{j+2} + a_{j+3}) + \cdots$$

té el mateix signe que a_j ($j \geq n+1$). En resulta, doncs, que

$$R_n R_{n+1} < 0 , \quad |R_n| \leq |a_{n+1}| .$$

Recopilarem ara aquestes fites de R_n en els criteris de fitació següents:

1. *Criteri de comparació amb una sèrie geomètrica*

Si existeix $\rho < 1$ tal que $|a_{j+1}| \leq \rho |a_j|$ ($j \geq n$), aleshores

$$|R_n| \leq \frac{|a_{n+1}|}{1-\rho} \leq \frac{\rho}{1-\rho} |a_n| .$$

2. *Criteri integral*

Si $a_j = f(j)$ amb $f : [0, \infty) \rightarrow \mathbb{R}$ decreixent sobre $[n, \infty)$, aleshores

$$R_n \leq \int_n^\infty f(x) dx \leq R_{n-1} .$$

En particular, sota la condició $|a_j| \leq f(j)$ ($j \geq n$),

$$|R_n| \leq \int_n^\infty f(x) dx .$$

3. *Criteri d'alternança de residus*

Si $R_n R_{n+1} < 0$, aleshores

$$S = S_n + R_n = S_{n+1} + R_{n+1} \in \langle S_n, S_{n+1} \rangle , \quad |R_n| \leq |a_{n+1}| .$$

4. *Criteri per a sèries alternades*

Si $a_j a_{j+1} < 0$, $|a_{j+1}| \leq |a_j|$ ($j \geq n+1$) i $a_j \rightarrow 0$ ($j \rightarrow \infty$); aleshores

$$S \in \langle S_n, S_{n+1} \rangle , \quad |R_n| \leq |a_{n+1}| .$$

És clar que aquests criteris s'apliquen de forma directa també per al càlcul de sumes finites.

4.3.3 Mètodes de sumació numèrica

Usant les fites del residu R_n , el càlcul de la sèrie $S = S_n + R_n$ es redueix al càlcul de S_n per a n tal que la magnitud de R_n sigui prou petita, d'acord amb la precisió demanada. Ara bé, el càlcul directe de S_n pot arribar a ser molt costós, per a sèries anomenades *lentament convergents*. Presentarem primer un exemple d'aquest tipus de sèries i, a continuació, donarem alguns mètodes que permeten aproximar S_n (i S) d'una forma alternativa més eficaç al càlcul directe, i que denominarem *mètodes d'acceleració de la sumació*.

EXAMPLE

Si volem sumar la sèrie

$$S = \sum_{j=1}^{\infty} \frac{1}{j^2}$$

amb un error més petit que 10^{-9} mitjançant el càlcul directe de les sumes parcials, el criteri integral aconsella prendre més de 10^9 termes.

La fórmula d'Euler-Maclaurin

Recordem la fórmula d'Euler-Maclaurin (4.23) per a una funció qualsevol f de $\mathcal{C}^{2s+2}([a, b])$: si dividim $[a, b]$ en n parts iguals d'amplada $h = \frac{b-a}{n}$, aleshores

$$\begin{aligned} T(h) &= \int_a^b f(x)dx + \sum_{r=1}^s h^{2r} \frac{B_{2r}}{(2r)!} [f^{(2r-1)}(b) - f^{(2r-1)}(a)] \\ &\quad + \frac{B_{2s+2}}{(2s+2)!} (b-a) f^{(2s+2)}(\xi) h^{2s+2}, \quad \xi \in (a, b), \end{aligned}$$

on els coeficients B_{2r} són els nombres de Bernoulli i

$$\begin{aligned} T(h) &= h \left[\frac{f(a)}{2} + \sum_{j=1}^{n-1} f(a+jh) + \frac{f(b)}{2} \right] \\ &= h \sum_{j=0}^n f(a+jh) - \frac{h}{2} [f(a) + f(a+nh)]. \end{aligned}$$

La dita fórmula pot ser escrita en aquesta altra manera, i l'anomenarem *fórmula d'Euler-Maclaurin per a sumes*

$$\begin{aligned} \sum_{j=0}^n f(a+jh) &= \frac{1}{h} \int_a^{a+nh} f(x)dx + \frac{1}{2} [f(a) + f(a+nh)] \\ &\quad + \sum_{r=1}^s h^{2r-1} \frac{B_{2r}}{(2r)!} [f^{(2r-1)}(a+nh) - f^{(2r-1)}(a)] \\ &\quad + R_s, \end{aligned}$$

on

$$R_s = nh^{2s+2} \frac{B_{2s+2}}{(2s+2)!} f^{(2s+2)}(\xi), \quad \xi \in (a, a+nh). \quad (4.30)$$

Per l'alternança de signe dels nombres de Bernoulli (vegeu (4.18)), si $f \in \mathcal{C}^{2s+4}([a, a+nh])$ i $f^{(2s+2)} f^{(2s+4)} \geq 0$ en $(a, a+nh)$, aleshores $R_s R_{s+1} < 0$ i, pel criteri d'alternança dels residus,

$$|R_s| \leq h^{2s+1} \frac{|B_{2s+2}|}{(2s+2)!} |f^{(2s+1)}(a+nh) - f^{(2s+1)}(a)|. \quad (4.31)$$

La *integral impròpia*

$$\int_a^\infty f(x)dx$$

és convergent quan existeix

$$J(f) = \lim_{n \rightarrow \infty} \int_a^{a+nh} f(x)dx.$$

Fent ús d'integrals impròpies convergents, en el cas que

$$\lim_{x \rightarrow \infty} f^{(2r-1)}(x) = 0 \quad (r = 1 \div s),$$

la fórmula d'Euler-Maclaurin per a sumes finites es pot generalitzar al càlcul de sumes de sèries

$$\sum_{j=0}^{\infty} f(a+jh) = \frac{1}{h} \int_a^{\infty} f(x)dx + \frac{f(a)}{2} - \sum_{r=1}^s h^{2r-1} \frac{B_{2r}}{(2r)!} f^{(2r-1)}(a) + R_s ,$$

anomenada *fórmula d'Euler-Maclaurin per a sèries*.

Una expressió per al residu R_s , anàloga a (4.30) no és ara vàlida, perquè $n \rightarrow \infty$ i s'ha de recórrer a (4.17) per obtenir-ne una expressió correcta (vegeu el problema IV.10). Si $f^{(2s+2)} f^{(2s+4)} > 0$ a $[a, \infty)$ i

$$\lim_{x \rightarrow \infty} f^{(2s+1)}(x) = 0 ,$$

la fórmula (4.31) ens diu que $|R_s|$ és més petit que la magnitud del primer terme menyspreat

$$|R_s| \leq h^{2s+1} \frac{|B_{2s+2}|}{(2s+2)!} |f^{(2s+1)}(a)| .$$

Aquesta fórmula es fa servir, prenent $h = 1$, per calcular sumes de sèries de forma molt eficient en un gran nombre de casos.

Finalment, notem que el desenvolupament de la fórmula d'Euler-Maclaurin no és sempre convergent; és a dir, no s'ha de complir necessàriament que $R_s \rightarrow 0$, quan $s \rightarrow \infty$. De fet, la magnitud dels nombres de Bernoulli B_{2s+2} creix indefinidament quan s tendeix a infinit, de manera que el seu quocient amb $(2s+2)!$ decreix potencialment en s ; així que, si $f^{(2s+1)}$ té un comportament de creixement més fort que aquell quocient, tindrem que R_s creixerà indefinidament. Molt comunament, en variar s de 0 a ∞ , s'observa que la magnitud de R_s decreix primer, per tornar a créixer després; es tracta d'un exemple típic de sèries semiconvergentes. Per a aquest tipus de sèries divergents, interessa prendre s de forma que la magnitud de R_s sigui com més petita millor; ara bé, a diferència de les sèries convergents, el valor més petit de R_s és fixat i no arbitràriament petit, com en aquelles. És a dir, pot passar que, si la precisió demanada en la suma és massa gran, no pugui ser assolida per a cap valor de s (vegeu-ne un exemple en el problema IV.11).

Mètode de comparació

Per tal de calcular la suma S d'una sèrie lentament convergent

$$\sum_{j=0}^{\infty} a_j ,$$

cerquem una altra sèrie

$$\sum_{j=0}^{\infty} b_j$$

tal que sigui convergent a un valor conegut T i que la sèrie

$$\sum_{j=0}^{\infty} (a_j - b_j)$$

que sigui més ràpidament convergent; així, n'hi ha prou de calcular la suma de la darrera sèrie, més ràpidament convergent que la inicial, i després sumar T al resultat obtingut,

$$S = T + \sum_{j=0}^{\infty} (a_j - b_j) .$$

Caldrà, doncs, conèixer alguns exemples de sumes de sèries per tal de poder aplicar aquest mètode de comparació al càlcul de les sumes desconegudes d'altres sèries.

Els exemples més obvis de sumes finites (i també, de sèries) calculables exactament corresponen a les anomenades *sumes telescòpiques*; això és, sumes

$$\sum_{j=m}^n b_j ,$$

on $b_j = \Delta T_j = T_{j+1} - T_j$, ja que, aleshores

$$\sum_{j=m}^n b_j = T_{n+1} - T_m .$$

Així, si considerem les *funcions factorials*, per a $\alpha \in \mathbb{Z}$,

$$x^{(\alpha)} = \begin{cases} x(x-1) \cdots (x-\alpha+1) & (\alpha > 0) \\ 1 & (\alpha = 0) \\ \frac{1}{(x+1)(x+2) \cdots (x-\alpha)} & (\alpha < 0) \end{cases}$$

resulta que $\Delta x^{(\alpha)} = \alpha x^{(\alpha-1)}$ i, per tant, si $\alpha \neq -1$,

$$x^{(\alpha)} = \frac{\Delta x^{(\alpha+1)}}{\alpha+1} , \quad \sum_{j=m}^n j^{(\alpha)} = \frac{(n+1)^{(\alpha+1)} - m^{(\alpha+1)}}{\alpha+1} .$$

En particular, si $\alpha < -1$,

$$\sum_{j=m}^{\infty} j^{(\alpha)} = -\frac{m^{(\alpha+1)}}{\alpha+1} .$$

Vegeu el problema IV.9 per a una aplicació pràctica.

4.4 EXTRAPOLACIÓ

4.4.1 Introducció

Per a molts problemes matemàtics, l'objectiu dels quals consisteix en el càlcul d'un valor v , convé desenvolupar la seva resolució numèrica en dues etapes:

- **DISCRETITZACIÓ:** Es calculen aproximacions numèriques $F(h)$ al valor exacte v , depenents d'un paràmetre h , anomenat *pas de discretització*. Per exemple, en el càlcul numèric de derivades i integrals, h és la separació entre abscisses consecutives. En el cas que les quantitats aproximades depenguin d'una variable entera n de forma que el valor v es troba quan n tendeix a ∞ , prendrem $h = \frac{1}{n}$.

- PAS AL LÍMIT: Es considera el límit de les aproximacions $F(h)$, quan h tendeix a 0.

En el context numèric, el pas al límit consisteix a realitzar els càlculs amb passos h cada vegada més petits, cosa que comporta dos tipus de dificultats:

1. El nombre d'operacions augmenta enormement.
2. Apareixen importants errors d'arrodoniment per excés d'operacions, o per cancel·lació de termes en restes de quantitats molt properes, que desfiguren el valor de $F(h)$.

4.4.2 Mètode de Richardson d'extrapolació repetida

A fi d'evitar totes aquestes dificultats, si coneixem el desenvolupament asimptòtic de $F(h)$ quan h tendeix a 0, el *mètode de Richardson* proporciona, a partir d'una successió $F(h_1)$, $F(h_2)$, ... de valors de F calculats directament, noves successions més ràpidament convergents cap a v (per a diverses aplicacions al càlcul de derivades i integrals, consulteu els problemes IV.12, IV.13 i IV.15).

Així, si sabem que el desenvolupament asimptòtic de $F(h)$ és de la forma

$$F(h) = v + a_1 h^{p_1} + a_2 h^{p_2} + \dots, \quad 0 < p_1 < p_2 < \dots,$$

prenem un nombre q qualsevol més gran que 1 i definim la successió d'extrapolacions següent:

$$F_1(h) = F(h), \quad F_{j+1}(h) = F_j(h) + \frac{F_j(h) - F_j(qh)}{q^{p_j} - 1} \quad (j \geq 1).$$

Aleshores, es prova per inducció que

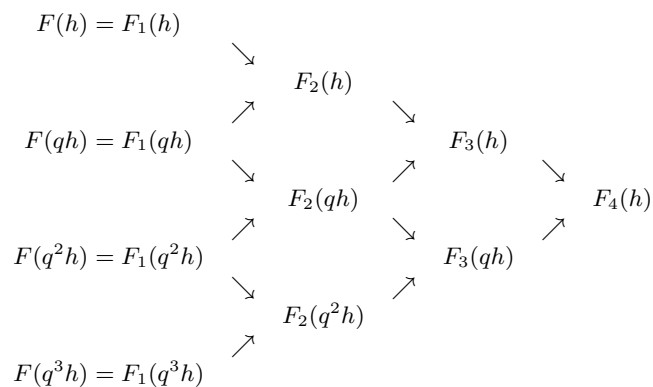
$$F_j(h) = v + a_j^{(j)} h^{p_j} + a_{j+1}^{(j)} h^{p_{j+1}} + \dots,$$

amb la qual cosa, els termes que contaminen v són ara molt més petits, si prenem el pas h prou petit.

Per tal de calcular $F_{j+1}(h)$ necessitem conèixer $F_j(h)$ i $F_j(qh)$; per calcular aquests últims, $F_{j-1}(h)$, $F_{j-1}(qh)$ i $F_{j-1}(q^2h)$, i així successivament. És per això que la forma més còmoda de procedir al càlcul dels $F_j(h)$ ($j \geq 1$) és mitjançant un esquema triangular com el que es mostra en la taula 4.1.

OBSERVACIONS

1. Per tal d'utilitzar aquest algorisme només cal conèixer els exponents $p_1, p_2, \dots$ i no els coeficients $a_1, a_2, \dots$.
2. Coneguts $F(h_1), F(h_2), \dots$, es pretén calcular $F(0)$; és a dir, el valor de F en una abscissa fora de $\langle h_1, h_2, \dots \rangle$. Per aquesta raó aquest procés rep el nom d'*extrapolació*.
3. El pas d'extrapolació de $F_j(h)$ a $F_{j+1}(h)$ es porta a terme afegint a $F_j(h)$ la diferència entre $F_j(h)$ i $F_j(qh)$ dividida per $q^{p_j} - 1$. Per aquesta raó, se'l sol anomenar *extrapolació del tipus $\frac{\Delta}{q^{p_j}-1}$* .



Taula 4.1: Esquema del mètode de Richardson.

4.5 CÀLCUL AMB OPERADORS

4.5.1 Introducció

Els mètodes basats en el càlcul formal amb operadors es revelen com a eines eficaces i elegants per a la construcció de fórmules (d'interpolació, derivació numèrica, integració numèrica ...) amb abscisses equidistants, exactes per a polinomis fins a un cert grau i, per tant, aproximades per a altres tipus de funcions.

Definicions

Sobre el conjunt $\mathcal{P}$ de polinomis en una variable (que, de fet, és un espai vectorial), definim els operadors següents, mitjançant llur aplicació a una funció polinomial f en un punt x qualssevol:

$$(Ef)(x) = f(x + h) \text{ (Operador desplaçament cap endavant).}$$

$$(\Delta f)(x) = f(x + h) - f(x) \text{ (Operador diferència cap endavant).}$$

$$(\nabla f)(x) = f(x) - f(x - h) \text{ (Operador diferència cap endarrera).}$$

$$(\delta f)(x) = f(x + \frac{h}{2}) - f(x - \frac{h}{2}) \text{ (Operador diferència centrada).}$$

$$(\mu f)(x) = \frac{1}{2}(f(x + \frac{h}{2}) + f(x - \frac{h}{2})) \text{ (Operador mitjana).}$$

$$(Df)(x) = f'(x) \text{ (Operador de derivació).}$$

$$(Jf)(x) = \int_x^{x+h} f(t)dt \text{ (Operador d'integració).}$$

Tots aquests operadors, llevat de l'operador D , depenen d'un pas h . Si és necessària una notació més explícita respecte al pas h usat (per exemple, quan treballem amb més d'un pas), escriurem $E_h f$, $\Delta_h f$, etc.

Els cinc primers operadors E , Δ , ∇ , δ i μ poden definir-se també per a qualsevol funció $f: \mathbb{R} \rightarrow \mathbb{R}$ i, fins i tot, per a funcions tabulades sobre abscisses equidistants, $\dots$, x_{-2} , x_{-1} , x_0 , x_1 , $\dots$, sempre que h sigui un múltiple enter de $x_{k+1} - x_k$ per a E , Δ i ∇ , i de $2(x_{k+1} - x_k)$ per a δ i μ . Els operadors D i J es poden definir sobre funcions més generals que els polinomis (funcions derivables i integrables, respectivament); però, aleshores requereixen un procés de pas al límit.

Donats dos operadors L_1 i L_2 , el seu *producte (per composició)* $L_1 L_2$ i la seva *suma* $L_1 + L_2$ vénen donats per

$$(L_1 L_2)f = L_1(L_2 f) \ , \quad (L_1 + L_2)f = L_1 f + L_2 f \ .$$

En particular, les potències L^j ($j \geq 0$) d'un operador es defineixen inductivament per $L^0 = 1$, on 1 és l'*operador identitat* (és a dir, $1f = f \ \forall f \in \mathcal{P}$), $L^j = L L^{j-1}$ ($j \geq 1$).

4.5.2 Propietats dels operadors

Donem a continuació propietats molt útils dels operadors definits:

1. Si L designa un qualsevol dels operadors definits, $L: \mathcal{P} \rightarrow \mathcal{P}$ és un operador lineal; això és,

$$L(\alpha f + \beta g) = \alpha Lf + \beta Lg \ ,$$

on α, β són coeficients i $f, g \in \mathcal{P}$.

2. Si $f \in \mathcal{P}$ és de grau N , aleshores Δf , ∇f , δf i Df són de grau $N - 1$ (si $N \geq 1$), Ef i μf de grau N , i Jf de grau més petit o igual que N . En particular, $L^j f = 0$ (si $j \geq N + 1$ i $L = \Delta, \nabla, \delta, D$).

3. Tota suma del tipus

$$\sum_{j=0}^{\infty} a_j L^j = a_0 + a_1 L + a_2 L^2 + a_3 L^3 + \dots \ ,$$

amb $L = \Delta, \nabla, \delta, D$, i amb a_j ($j \geq 0$) coeficients qualssevol, es redueix a una suma finita en aplicar-la sobre un polinomi i, per tant, la *sèrie d'operadors*

$$\sum_{j=0}^{\infty} a_j L^j$$

és un operador ben definit sobre $\mathcal{P}$.

A continuació, en donem alguns exemples.

Si $t \in \mathbb{R}$ i $L = \Delta, \nabla, \delta, D$:

$$\begin{aligned} (1 + L)^t &\equiv \sum_{j=0}^{\infty} \binom{t}{j} L^j \\ &= 1 + tL + \frac{t(t-1)}{2} L^2 + \frac{t(t-1)(t-2)}{6} L^3 + \dots \ , \\ e^{tL} &\equiv \sum_{j=0}^{\infty} \frac{t^j}{j!} L^j \end{aligned}$$

$$\begin{aligned}
&= 1 + tL + \frac{t^2}{2}L^2 + \frac{t^3}{6}L^3 + \dots, \\
\ln(1 + L) &\equiv \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} L^j \\
&= L - \frac{1}{2}L^2 + \frac{1}{3}L^3 - \dots, \\
\arg \sinh L &\equiv \sum_{r=0}^{\infty} \frac{1}{2r+1} \left(-\frac{1}{2}\right)^r L^{2r+1} \\
&= L - \frac{1}{2} \frac{L^3}{3} + \frac{1 \cdot 3}{2 \cdot 4} \frac{L^5}{5} - \dots, \\
\sinh(tL) &\equiv \frac{1}{2}(e^{tL} - e^{-tL}) = \sum_{r=0}^{\infty} \frac{t^{2r+1}}{(2r+1)!} L^{2r+1} \\
&= tL + \frac{t^3}{3!}L^3 + \frac{t^5}{5!}L^5 + \dots, \\
\cosh(tL) &\equiv \frac{1}{2}(e^{tL} + e^{-tL}) = \sum_{r=0}^{\infty} \frac{t^{2r}}{(2r)!} L^{2r} \\
&= 1 + \frac{t^2}{2!}L^2 + \frac{t^4}{4!}L^4 + \frac{t^6}{6!}L^6 + \dots
\end{aligned}$$

Els *operadors inversos* (respecte al producte per composició) de $1 + L$ i e^{tL} s'indican, respectivament, per $(1 + L)^{-1}$ i e^{-tL} . En notació multiplicativa, se sol escriure:

$$\frac{1}{1 + L} \equiv (1 + L)^{-1}, \quad \frac{1}{e^{tL}} \equiv e^{-tL}, \quad \sqrt{1 + L} = (1 + L)^{\frac{1}{2}}, \quad \text{etc.}$$

4. Els set operadors $E, \Delta, \delta, \nabla, D, \mu$ i J *commuten entre ells*. Això és que, si L_1 i L_2 són dos d'aquests operadors, $L_1 L_2 = L_2 L_1$; per exemple, $JD = DJ = \Delta$.

5. Disposem de la relació:

$$(E^t f)(x) = ((1 + \Delta)^t f)(x) = f(x + th) \quad \forall t \in \mathbb{R} \quad \forall f \in \mathcal{P}.$$

En efecte, aquesta relació és clarament certa per a $t \in \mathbb{N}$. Vegem ara que també és vàlida per a $t \in \mathbb{R}$. Això és conseqüència del fet que, si $x_0, x_1, x_2, \dots$ formen una xarxa equidistant amb $x_{k+1} - x_k = h$ ($k \geq 0$), aleshores

$$f[x_0, x_1] = \frac{1}{h}(\Delta f)(x_0), \quad f[x_0, x_1, x_2] = \frac{1}{2h^2}(\Delta^2 f)(x_0),$$

i, per inducció,

$$f[x_0, x_1, \dots, x_j] = \frac{1}{j!h^j}(\Delta^j f)(x_0).$$

Si $f \in \mathcal{P}$ té grau m , fent servir la variable t tal que $x = x_0 + th$ tenim $x_0 + th - x_k = (t - k)h$ i, usant la fórmula d'interpolació proporcionada pel mètode de les diferències dividides, obtenim la relació buscada:

$$\begin{aligned} f(x_0 + th) &= \sum_{j=0}^m f[x_0, x_1, \dots, x_j] th(t-1)h \cdots (t-j+1)h \\ &= \sum_{j=0}^m \binom{t}{j} (\Delta^j f)(x_0) = ((1 + \Delta)^t f)(x_0) . \end{aligned}$$

Notem que l'operador $\ln E = \ln(1 + \Delta)$ té també sentit i apareix definit a la propietat 3.

6. Es compleixen les relacions entre els operadors donades en la taula 4.2, deduïdes a partir de les definicions i emprant la *fórmula de Taylor usant operadors*

$$E = e^{hD} .$$

	E	Δ	δ	∇	D
E	E	$1 + \Delta$	$1 + \frac{1}{2}\delta^2 + \delta(1 + \frac{1}{4}\delta^2)^{\frac{1}{2}}$	$\frac{1}{1-\nabla}$	e^{hD}
Δ	$E - 1$	Δ	$\delta(1 + \frac{1}{4}\delta^2)^{\frac{1}{2}} + \frac{1}{2}\delta^2$	$\frac{\nabla}{1-\nabla}$	$e^{hD} - 1$
δ	$E^{\frac{1}{2}} - E^{-\frac{1}{2}}$	$\frac{\Delta}{(1+\Delta)^{\frac{1}{2}}}$	δ	$\frac{\nabla}{(1-\nabla)^{\frac{1}{2}}}$	$2 \sinh \frac{h}{2} D$
∇	$1 - E^{-1}$	$\frac{\Delta}{1+\Delta}$	$\delta(1 + \frac{1}{4}\delta^2)^{\frac{1}{2}} - \frac{1}{2}\delta^2$	∇	$1 - e^{-hD}$
D	$\frac{1}{h} \ln E$	$\frac{1}{h} \ln(1 + \Delta)$	$\frac{2 \arg \sinh(\frac{1}{2}\delta)}{h}$	$-\frac{\ln(1-\nabla)}{h}$	D
μ	$\frac{1}{2}(E^{\frac{1}{2}} + E^{-\frac{1}{2}})$	$\frac{1+\frac{1}{2}\Delta}{(1+\Delta)^{\frac{1}{2}}}$	$(1 + \frac{1}{4}\delta^2)^{\frac{1}{2}}$	$\frac{1-\frac{1}{2}\nabla}{(1-\nabla)^{\frac{1}{2}}}$	$\cosh(\frac{h}{2} D)$

Taula 4.2: Relacions entre operadors.

Per exemple, $\mu = (1 + \frac{1}{4}\delta^2)^{\frac{1}{2}}$, $D = \frac{2 \arg \sinh(\frac{1}{2}\delta)}{h}$, etc.

7. Tenim l'expressió següent per a l'operador J :

$$J = h\Delta(\ln(1 + \Delta))^{-1} = h(1 + \frac{1}{2}\Delta - \frac{1}{12}\Delta^2 + \frac{1}{24}\Delta^3 - \frac{19}{720}\Delta^4 + \cdots) .$$

Notem que, a partir de $JD = \Delta$ i $D = \frac{1}{h} \ln(1 + \Delta)$, resulta la relació $J \ln(1 + \Delta) = h\Delta$ de la qual es desprèn la primera igualtat donada. Per calcular el desenvolupament de J en funció de Δ , s'escriu

$$J = h \sum_{j=0}^{\infty} c_j \Delta^j$$

i s'imposa l'equació satisfeta per J , $J \ln(1 + \Delta) = h\Delta$, obtenint

$$h(c_0 + c_1\Delta + c_2\Delta^2 + c_3\Delta^3 + \dots)(\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \dots) = h\Delta ;$$

d'on, es poden determinar els successius coeficients

$$c_0 = 1 , \quad c_1 = \frac{1}{2} , \quad c_2 = -\frac{1}{12} , \quad c_3 = \frac{1}{24} , \quad \dots$$

4.5.3 Aplicacions del càlcul amb operadors

Generem ara de manera sistemàtica tota una sèrie de fórmules per mitjà de les relacions establertes entre els operadors definits. Aquestes fórmules, que utilitzaran sempre abscisses equidistants, seran exactes només per a polinomis; però, per a les altres funcions, a més de ser aproximades, tot sovint donaran un desenvolupament asimptòtic de l'error comès.

Interpolació

A partir de la relació $E^t = (1 + \Delta)^t$ per a $t \in \mathbb{R}$, obtenim

$$f(x_0 + th) = \sum_{j=0}^{\infty} \binom{t}{j} (\Delta^j f)(x_0) , \quad t, x_0 \in \mathbb{R} ,$$

exacta per a qualsevol polinomi $f \in \mathcal{P}$, essent a més la suma finita.

Si fixem $m \geq 0$, obtenim una fórmula exacta per a polinomis f de grau més petit o igual que m

$$f(x_0 + th) = \sum_{j=0}^m \frac{t(t-1) \cdots (t-j+1)}{j!} (\Delta^j f)(x_0) , \quad t, x_0 \in \mathbb{R} ;$$

d'on, dient $x = x_0 + th$ ($t = \frac{x-x_0}{h}$), $\Delta^j f_0 = (\Delta^j f)(x_0)$ ($j = 0 \div m$), $x_k = x_0 + kh$, $f_k = f(x_k)$ ($k = 0 \div m$), obtenim la fórmula d'interpolació següent que aproxima el valor de $f(x)$, suposant coneguts $f_0, f_1, \dots, f_m$:

$$f(x) \simeq p_m(x) = \sum_{j=0}^m \frac{\Delta^j f_0}{j! h^j} (x - x_0) \cdots (x - x_{j-1}) ,$$

anomenada *fórmula d'interpolació de Newton cap al davant*, exacta per a polinomis de grau més petit o igual que m .

Com ja hem vist

$$\frac{\Delta^j f_0}{j! h^j} = f[x_0, x_1, \dots, x_j]$$

i, per tant, $p_m(x)$ és el polinomi de grau menor o igual que m que interpola f en les abscisses $x_0, x_1 = x_0 + h, \dots, x_m = x_0 + mh$.

Igualment, usant la relació $E^{-t} = (1 - \nabla)^t$, obtenim

$$f(x) \simeq p_m(x) = \sum_{j=0}^m \frac{\nabla^j f_0}{j! h^j} (x - x_0) \cdots (x - x_{-(j-1)}) ,$$

anomenada *fórmula d'interpolació de Newton cap al darrera*, exacta per a polinomis de grau més petit o igual que m ; aquí el polinomi $p_m(x)$ interpola f en les abscisses $x_0, x_{-1}, \dots, x_{-m}$, ja que

$$\frac{\nabla^j f_0}{j!h^j} = f[x_{-j}, x_{-(j-1)}, \dots, x_0] .$$

Altres fórmules d'interpolació, per a xarxes centrades, es troben en el problema IV.14.

Derivació numèrica

A partir de

$$\begin{aligned} D &= \frac{1}{h} \ln(1 + \Delta) = -\frac{1}{h} \ln(1 - \nabla) \\ &= \frac{2}{h} \arg \sinh\left(\frac{\delta}{2}\right) = \frac{2}{h} \mu \left(1 + \frac{\delta^2}{4}\right)^{-\frac{1}{2}} \arg \sinh\left(\frac{\delta}{2}\right) , \end{aligned}$$

obtenim les fórmules de derivació numèrica següents, exactes per als polinomis de grau més petit o igual que m , si tallem després dels termes en Δ^m , δ^m i ∇^m , respectivament:

$$\begin{aligned} f'_0 &\simeq \frac{1}{h} (\Delta f_0 - \frac{1}{2} \Delta^2 f_0 + \frac{1}{3} \Delta^3 f_0 - \dots) , \\ f'_0 &\simeq \frac{1}{h} (\nabla f_0 + \frac{1}{2} \nabla^2 f_0 + \frac{1}{3} \nabla^3 f_0 + \dots) , \\ f'_0 &\simeq \frac{1}{h} (\delta f_0 - \frac{1}{2} \frac{1}{2^2 \cdot 3} \delta^3 f_0 + \frac{1 \cdot 3}{2 \cdot 4} \frac{1}{2^4 \cdot 5} \delta^5 f_0 - \dots) , \\ f'_0 &\simeq \frac{1}{h} (\mu \delta f_0 - \frac{1}{6} \mu \delta^3 f_0 + \frac{1}{30} \mu \delta^5 f_0 - \dots) \\ &= \frac{1}{2h} [(f_1 - f_{-1}) - \frac{1}{6} (\delta^2 f_1 - \delta^2 f_{-1}) \\ &\quad + \frac{1}{30} (\delta^4 f_1 - \delta^4 f_{-1}) - \dots] ; \end{aligned}$$

on $f'_0 \equiv f'(x_0)$.

La primera és una fórmula de derivació cap al davant i la segona, cap al darrera. Les dues últimes són fórmules centrades; l'última té, a més a més, l'avantatge que no fa intervenir valors de la funció en abscisses intermèdies de la xarxa equiespaiada amb pas h : per exemple, la tercera conté $\delta f_0 = f_{\frac{1}{2}} - f_{-\frac{1}{2}}$ i requereix valors que no coneixem en una taula de f feta amb pas h .

Per tal de calcular derivades d'ordre superior, és molt comú emprar les relacions:

$$\begin{aligned} D^{2r} &= \left[\frac{2}{h} \arg \sinh\left(\frac{\delta}{2}\right) \right]^{2r} , \\ D^{2r+1} &= \mu \left(1 + \frac{1}{4} \delta^2\right)^{-\frac{1}{2}} \left[\frac{2}{h} \arg \sinh\left(\frac{\delta}{2}\right) \right]^{2r+1} . \end{aligned}$$

Així, per exemple:

$$f_0^{(2)} \simeq \frac{1}{h^2} (\delta^2 f_0 - \frac{1}{12} \delta^4 f_0 + \frac{1}{90} \delta^6 f_0 - \dots) ,$$

$$\begin{aligned}
f_0^{(3)} &\simeq \frac{1}{h^3}(\mu\delta^3 f_0 - \frac{1}{4}\mu\delta^5 f_0 + \frac{7}{120}\mu\delta^7 f_0 - \dots) , \\
f_0^{(4)} &\simeq \frac{1}{h^4}(\delta^4 f_0 - \frac{1}{6}\delta^6 f_0 + \frac{7}{240}\delta^8 f_0 - \dots) , \\
f_0^{(5)} &\simeq \frac{1}{h^5}(\mu\delta^5 f_0 - \frac{1}{3}\mu\delta^7 f_0 + \dots) .
\end{aligned}$$

Notem finalment que totes les expressions asimptòtiques obtingudes permeten l'aplicació del mètode de Richardson (vegeu el problema IV.15).

Integració numèrica

Notant que

$$(E^{m-1}Jf)(x) = (JE^{m-1}f)(x) = \int_{x+(m-1)h}^{x+mh} f(t)dt ,$$

resulta que

$$\begin{aligned}
J_m f_0 &\equiv \int_{x_0}^{x_0+mh} f(t)dt = [(1 + E + \dots + E^{m-1})J]f_0 \\
&= [(E^m - 1)(E - 1)^{-1}J]f_0 ,
\end{aligned}$$

per a qualsevol polinomi $f \in \mathcal{P}$. Desenvolupant en potències de Δ , s'obté

$$\begin{aligned}
J_m f_0 &= h[(1 + \Delta)^m - 1]\Delta^{-1}[\Delta(\ln(1 + \Delta))^{-1}]f_0 \\
&= h \left[\Delta^{m-1} + m\Delta^{m-2} + \binom{m}{2}\Delta^{m-3} + \dots \right. \\
&\quad \left. + \binom{m}{m-2}\Delta + \binom{m}{m-1} \right] [c_0 + c_1\Delta + c_2\Delta^2 + \dots]f_0 \\
&= h [d_0^{(m)} + d_1^{(m)}\Delta + d_2^{(m)}\Delta^2 + d_3^{(m)}\Delta^3 + \dots]f_0 ,
\end{aligned}$$

on

$$d_j^{(m)} = \sum_{l=0}^{\min(j, m-1)} \binom{m}{m-l-1} c_{j-l} = \sum_{l=0}^{\min(j, m-1)} \binom{m}{l+1} c_{j-l} ;$$

podem trobar aquests coeficients, emprant la propietat 7 dels operadors

$$\Delta[\ln(1 + \Delta)]^{-1} = c_0 + c_1\Delta + c_2\Delta^2 + \dots ,$$

amb $c_0 = 1$, $c_1 = \frac{1}{2}$, $c_2 = -\frac{1}{12}$, $c_3 = \frac{1}{24}$, $c_4 = -\frac{19}{720}$, $c_5 = \frac{3}{160}$, ...

En particular, en els casos $m = 1, 2$, obtenim:

$$\begin{aligned}
J_1 f_0 &= \int_{x_0}^{x_0+h} f(t)dt \simeq h(f_0 + \frac{1}{2}\Delta f_0 - \frac{1}{12}\Delta^2 f_0 \\
&\quad + \frac{1}{24}\Delta^3 f_0 - \frac{19}{720}\Delta^4 f_0 + \dots) , \\
J_2 f_0 &= \int_{x_0}^{x_0+2h} f(t)dt \simeq 2h(f_0 + \Delta f_0 + \frac{1}{6}\Delta^2 f_0 \\
&\quad - \frac{1}{180}\Delta^4 f_0 + \frac{1}{180}\Delta^5 f_0 + \dots) .
\end{aligned}$$

Les fórmules d'integració numèrica així obtingudes per a les integrals $J_m f_0$ ($m \geq 1$) són exactes per als polinomis de grau menor o igual que n si tallem després del terme en $\Delta^n f_0$. Si $n = m$, les fórmules obtingudes depenen linealment només dels valors $f_0, f_1, \dots, f_m$, i són, per tant, les anomenades fórmules de Newton-Cotes d'ordre m .

Per exemple, si $m = 1, 2$, obtenim:

$$\begin{aligned} \int_{x_0}^{x_0+h} f(t) dt &\simeq \frac{h}{2}(f_0 + f_1) - \frac{h}{12}(\Delta^2 f_0 - \frac{1}{2}\Delta^3 f_0 + \dots), \\ \int_{x_0}^{x_0+2h} f(t) dt &\simeq \frac{h}{3}(f_0 + 4f_1 + f_2) - \frac{h}{90}(\Delta^4 f_0 - \Delta^5 f_0 + \dots). \end{aligned}$$

Aquestes fórmules són les del trapezi i de Simpson amb termes correctors, respectivament. Notem també que, substituint $\Delta^2 f_0, \Delta^4 f_0$, per $h^2 f^{(2)}(\xi), h^4 f^{(4)}(\xi)$ en els primers termes correctors, retrobem l'expressió de l'error d'integració numèrica d'ambdues fórmules.

La fórmula d'Euler-Maclaurin es dedueix en el problema IV.16, usant també tècniques de càlcul amb operadors.

COMENTARIS BIBLIOGRÀFICS

La presentació feta del polinomi interpolador, eina fonamental per a la derivació i integració numèriques, és similar a la que es troba a [Hen64]. Una bona col·lecció de fórmules de derivació i integració numèriques es troben a [AS65]. La referència [IK66] conté les expressions de l'error en la derivació numèrica i en les fórmules (tancades) de Newton-Cotes presentades, així com les corresponents a altres fórmules de Newton-Cotes, anomenades obertes; el teorema del valor mitjà per a sumes i la convergència de les regles d'integració numèrica també s'hi troben. Més complet és [DR67] que abasta una part molt important de mètodes d'integració numèrica, dels quals destaquem els d'*integració adaptable*, que proporcionen estratègies segons el comportament de la funció que s'integra en diverses zones de l'interval d'integració (vegeu també [RR78]). Per a tècniques d'*integració múltiple*, consulteu [Str71] i, per a la integració gaussiana, [SS66]. Els criteris de fitació dels residus de les sèries es troben en els llibres estàndards d'anàlisi, per exemple [Apo57]. Uns altres mètodes de sumació, diferents dels presentats, es poden trobar a [DB74], [DM73] i [RR78]. La fórmula d'Euler-Maclaurin, amb la mateixa notació per als nombres de Bernoulli, es troba a [SB80], que també tracta amb detall l'extrapolació en el marc de la integració numèrica. Altres referències sobre extrapolació són [DB74], [Hen64] i [RR78]. La generació, via operadors lineals de fórmules d'interpolació, derivació i integració numèriques, és bastant clàssica i es troba, per exemple, a [DB74], [Fro69], [Hil74] i [Sch67]. Aquest darrer aporta, a més, molts exercicis resolts sobre la resta de temes d'aquest llibre.

PROBLEMES RESOLTS

Problema 4.1 a) Trobeu una fórmula de derivació interpolatòria per al càlcul aproximat de $f'(a)$, per derivació del polinomi interpolador en els punts $a + h$, $a + \frac{h}{2}$, $a + \frac{h}{4}$ i a .

b) Suposant que $f \in C^4([a, a + h])$, doneu una expressió per a l'error comès.

c) Calculeu la derivada de la funció $f(x) = \cosh x^2$ en $x = a = 1$, emprant la fórmula trobada en a) amb $h = 0.1$ i fiteu l'error comès.

d) Compareu l'error exacte amb la fita d'error trobada.

SOLUCIÓ:

a) Comencem trobant el polinomi interpolador a partir de la taula de diferències dividides

$$\begin{array}{l|l}
 x_0 = a & f_0 \\
 & \frac{4(f_1 - f_0)}{h} \\
 x_1 = a + \frac{h}{4} & f_1 \\
 & \frac{4(f_2 - f_1)}{h} \quad \frac{8(f_2 - 2f_1 + f_0)}{h^2} \\
 & \frac{8(f_3 - 6f_2 + 8f_1 - 3f_0)}{3h^3} \\
 x_2 = a + \frac{h}{2} & f_2 \\
 & \frac{8(f_3 - 3f_2 + 2f_1)}{3h^2} \\
 & \frac{2(f_3 - f_2)}{h} \\
 x_3 = a + h & f_3
 \end{array}$$

El polinomi interpolador que en resulta és

$$\begin{aligned}
 p_3(x) &= f[x_0] + f[x_0, x_1](x - a) + f[x_0, x_1, x_2](x - a)(x - a - \frac{h}{4}) \\
 &\quad + f[x_0, x_1, x_2, x_3](x - a)(x - a - \frac{h}{4})(x - a - \frac{h}{2}) .
 \end{aligned}$$

La derivada d'aquest polinomi en $x = a$ és l'aproximació demanada de $f'(a)$

$$\begin{aligned}
 p'_3(a) &= f[x_0, x_1] + f[x_0, x_1, x_2](-\frac{h}{4}) + f[x_0, x_1, x_2, x_3](-\frac{h}{4})(-\frac{h}{2}) \\
 &= f[x_0, x_1] - \frac{h}{4}f[x_0, x_1, x_2] + \frac{h^2}{8}f[x_0, x_1, x_2, x_3] \\
 &= \frac{4}{h}(f_1 - f_0) - \frac{h}{4} \frac{8}{h^2}(f_2 - 2f_1 + f_0) \\
 &\quad + \frac{h^2}{8} \frac{8}{3h^3}(f_3 - 6f_2 + 8f_1 - 3f_0) \\
 &= \frac{f_3 - 12f_2 + 32f_1 - 21f_0}{3h} .
 \end{aligned}$$

Així, queda

$$f'(a) \simeq \frac{1}{3h}[f(a + h) - 12f(a + \frac{h}{2}) + 32f(a + \frac{h}{4}) - 21f(a)] .$$

b) L'error en la interpolació s'expressa com

$$f(x) - p_3(x) = \frac{f^{(4)}(\xi(x))}{4!} (x-a)(x-a-\frac{h}{4})(x-a-\frac{h}{2})(x-a-h) ,$$

i l'error de la fórmula de derivació trobada,

$$e'_3(a) = f'(a) - p'_3(a) = \frac{f^{(4)}(\xi)}{4!} (-\frac{h}{4})(-\frac{h}{2})(-h) = -\frac{f^{(4)}(\xi)}{192} h^3 ,$$

on $\xi \in \langle a, a+h \rangle$.

c) Prenent $h = 0.1$, resulta

$$\begin{aligned} f'(1) &\simeq \frac{1}{3 \cdot 0.1} [f(1.1) - 12f(1.05) + 32f(1.025) - 21f(1)] \\ &\simeq 2.351006625 . \end{aligned}$$

Per tal de trobar una fita de l'error cal fitar primerament la derivada quarta de f en l'interval $[1, 1.1]$; tenim

$$\begin{aligned} f'(x) &= 2x \sinh x^2 , \\ f^{(2)}(x) &= 4x^2 \cosh x^2 + 2 \sinh x^2 , \\ f^{(3)}(x) &= 8x^3 \sinh x^2 + 12x \cosh x^2 , \\ f^{(4)}(x) &= (16x^4 + 12) \cosh x^2 + 48x^2 \sinh x^2 . \end{aligned}$$

Com que $f^{(4)}$ és una funció creixent i positiva, donem com a fita d'aquesta el seu valor en $x = 1.1$ que resulta ser $M_4 = 153.408$.

La fita trobada per a l'error és llavors

$$\frac{M_4}{192} 10^{-3} \simeq 0.8 \cdot 10^{-3} .$$

d) Emprant directament l'expressió de la derivada de f en $x = a = 1$, tenim

$$f'(a) = 2a \sinh a^2 = 2 \sinh 1 = 2.3504023873 ;$$

l'error real és, doncs, $0.60424 \cdot 10^{-3}$, no gaire menor que la fita trobada en c).

Problema 4.2 Disposem de les dades següents d'una funció f :

x	0.4	0.5
$f(x)$	1.554284	1.561136
$f'(x)$	0.243031	-0.089618

a) Trobeu l'abscissa del màxim de f a $[0.4, 0.5]$, aproximant-la pel màxim del polinomi interpolador d'Hermite $p_3(x)$ a la taula donada de f .

b) Suposant que $f \in \mathcal{C}^4([x_0, x_1])$, trobeu la següent expressió per a la derivada e'_3 de l'error en la interpolació d'Hermite en dos punts $x_0 < x_1$:

$$e'_3(x) \equiv f'(x) - p'_3(x) = \frac{f^{(4)}(\eta(x))}{3!}(x - x_0)(x - x_1)(x - \xi) ,$$

on $\xi \in (x_0, x_1)$ i $\eta(x) \in (x_0, x_1, x)$.

c) Fiteu l'error en l'abscissa del màxim degut a la interpolació, sabent que $|f^{(4)}(x)| < 10^3$ i $|f^{(2)}(x)| > 1 \ \forall x \in (0.4, 0.5)$.

SOLUCIÓ:

a) El polinomi interpolador d'Hermite es troba a partir de la taula de diferències dividides generalitzades següent, on $f[x_i, x_i] \equiv f'(x_i)$:

0.4	1.554284		
		0.243031	
0.4	1.554284	-1.7451	
		0.068520	1.637
0.5	1.561136	-1.5814	
		-0.089618	
0.5	1.561136		

l'escrivim així

$$\begin{aligned} p_3(x) &= d + c(x - x_0) + b(x - x_0)^2 + a(x - x_0)^2(x - x_1) \\ &= d + c\delta + b\delta^2 - a(x_1 - x_0)\delta^2 + a\delta^3 , \end{aligned}$$

amb $a = 1.637$, $b = -1.7451$, $c = 0.243031$ i $d = 1.554284$, fent $\delta = x - x_0$. El màxim es dóna en el zero del polinomi derivat en l'interval $(0.4, 0.5)$. La derivada del polinomi resulta ser

$$p'_3(x) = c - 2[a(x_1 - x_0) - b]\delta + 3a\delta^2 = C - 2B\delta + A\delta^2;$$

amb $C = c = 0.243031$, $B = a(x_1 - x_0) - b = 1.9088$ i $A = 3a = 4.911$. Dels dos zeros d'aquest polinomi

$$\delta = \frac{B \pm \sqrt{B^2 - AC}}{A} = \frac{C}{B \mp \sqrt{B^2 - AC}}$$

triem el més petit, perquè $f'(x_0) > 0$ i $f'(x_1) < 0$; llavors el màxim cercat serà

$$x_M = x_0 + \frac{C}{B + \sqrt{B^2 - AC}} = 0.4700 .$$

b) L'error d'interpolació e_3 i la seva derivada e'_3 s'anul·len en els punts d'interpolació x_0 i x_1 . Aplicant el teorema de Rolle, existeix $\xi \in (x_0, x_1)$ tal que $e'_3(\xi) = 0$. La fórmula proposada per a $e'_3(x)$ és clarament certa per a $x = x_0, x_1, \xi$ i qualsevol valor de η triat.

Per als altres valors de x , definim per a $z \in [x_0, x_1]$ la funció

$$F(z) = e'_3(z) - a(x)(z - x_0)(z - x_1)(z - \xi) ,$$

on

$$a(x) = \frac{e'_3(x)}{(x - x_0)(x - x_1)(x - \xi)} .$$

És clar que $F \in \mathcal{C}^3([x_0, x_1])$ i s'anul·la en els quatre punts diferents x, x_0, x_1, ξ . Aplicant successivament el teorema de Rolle a F , F' i $F^{(2)}$, existeix $\eta(x) \in]x_0, x_1, x >$ tal que $F^{(3)}(\eta(x)) = 0$; és a dir,

$$f^{(4)}(\eta(x)) - 3!a(x) = 0$$

i, per tant,

$$0 = F(x) = e'_3(x) - \frac{f^{(4)}(\eta(x))}{3!}(x - x_0)(x - x_1)(x - \xi) .$$

Fitant ara $|e'_3(x)|$, obtenim, per a $x \in [x_0, x_1]$,

$$|e'_3(x)| \leq \frac{M_4}{6} |(x - x_0)(x - x_1)(x_1 - x_0)| \leq \frac{M_4}{24} (x_1 - x_0)^3 ,$$

amb $M_4 = \max_{\eta \in [x_0, x_1]} |f^{(4)}(\eta)|$.

Per exemple, en les condicions de l'apartat c), $|e'_3(x)| < \frac{1}{24} \equiv \epsilon$. És a dir, en tots els punts de l'interval $(0.4, 0.5)$, la derivada s'avalua amb un error fitat per ϵ .

c) Siguin $\bar{x}_M$ i x_M els màxims de f i $p_3(x)$, respectivament, en (x_0, x_1) . Aleshores, pel teorema del valor mitjà,

$$e'_3(x_M) = f'(x_M) = f^{(2)}(\zeta)(x_M - \bar{x}_M) , \quad \zeta \in]x_M, \bar{x}_M > .$$

Per tant, si fem

$$m_2 = \min_{\zeta \in [x_0, x_1]} |f^{(2)}(\zeta)| ,$$

obtenim la fita demanada

$$\begin{aligned} |x_M - \bar{x}_M| &\leq \frac{|e'_3(x_M)|}{m_2} \leq \frac{M_4}{6m_2} |(x_M - x_0)(x_M - x_1)(x_M - \xi)| \\ &< \frac{10^3}{6} (0.07)(0.03)(0.07) \simeq 2.5 \cdot 10^{-2} . \end{aligned}$$

Problema 4.3 a) Determineu la fórmula d'integració numèrica

$$\int_{-1}^1 g(t) dt \simeq \sum_{i=0}^1 W_i g(t_i)$$

que és exacta per a tots els polinomis de grau més petit o igual que 3.

b) Si $g \in \mathcal{C}^4([-1, 1])$, trobeu una expressió per a l'error comès.

SOLUCIÓ:

a) S'han de determinar quatre paràmetres: les abscisses $t_0, t_1 \in [-1, 1]$ i els pesos w_0, w_1 .

Atesa la linealitat de la fórmula respecte a g , el fet d'imposar exactitud a $\mathcal{P}_3$ és equivalent a fer-ho per a la base de polinomis $\{1, t, t^2, t^3\}$:

$$\begin{array}{l|l} g(t) = 1 & 2 = w_0 + w_1 \\ g(t) = t & 0 = w_0 t_0 + w_1 t_1 \\ g(t) = t^2 & \frac{2}{3} = w_0 t_0^2 + w_1 t_1^2 \\ g(t) = t^3 & 0 = w_0 t_0^3 + w_1 t_1^3 \end{array} .$$

Aquest sistema no lineal és senzill de resoldre, resulta:

$$t_0 = -\frac{\sqrt{3}}{3}, \quad t_1 = \frac{\sqrt{3}}{3}; \quad w_0 = w_1 = 1 .$$

La fórmula d'integració numèrica cercada és, doncs,

$$\int_{-1}^1 g(t) dt \simeq g\left(-\frac{\sqrt{3}}{3}\right) + g\left(\frac{\sqrt{3}}{3}\right) .$$

b) Aquesta fórmula és exacta per a tots els polinomis de grau menor o igual 3, però no ho és per a tots els de grau 4. En efecte, prenent

$$g(t) = (t - t_0)^2(t - t_1)^2 = \left(t^2 - \frac{1}{3}\right)^2 = t^4 - \frac{2}{3}t^2 + \frac{1}{9} ,$$

tenim

$$\int_{-1}^1 (t - t_0)^2(t - t_1)^2 dt = \frac{8}{45}$$

i l'aproximació donaria zero, ja que $g(t_i) = 0$ ($i = 0, 1$).

Suposant que $g \in \mathcal{C}^2([-1, 1])$, considerem el polinomi $p_3(x)$ d'interpolació d'Hermite a f en les abscisses t_0, t_1 . De la fórmula de l'error a la interpolació d'Hermite es té

$$g(t) - p_3(t) = \frac{g^{(4)}(\xi(t))}{4!} (t - t_0)^2(t - t_1)^2, \quad \xi(t) \in \langle t_0, t_1, t \rangle .$$

Atès que el polinomi d'interpolació d'Hermite s'integra exactament (en ser de grau menor o igual que 3), trobem l'error d'integració numèrica fent servir el teorema del valor mitjà per a integrals:

$$E_3 \equiv \int_{-1}^1 g(t) dt - g\left(-\frac{\sqrt{3}}{3}\right) - g\left(\frac{\sqrt{3}}{3}\right)$$

$$\begin{aligned}
&= \int_{-1}^1 \frac{g^{(4)}(\xi(t))}{4!} (t-t_0)^2 (t-t_1)^2 dt \\
&= \frac{g^{(4)}(\xi)}{4!} \int_{-1}^1 (t-t_0)^2 (t-t_1)^2 dt \\
&= \frac{1}{135} g^{(4)}(\xi) , \quad \xi \in (-1, 1) .
\end{aligned}$$

Noteu que no hem fet més que redescobrir, usant únicament mètodes elementals, la fórmula de Gauss-Legendre de dos punts que apareix deduïda, en general, en el problema IV.8.

Problema 4.4 Volem trobar fórmules d'integració en $m+1$ punts equidistants de la forma

$$\int_a^b f(x) dx \simeq \sum_{k=0}^m [A_k f(x_k) + B_k f'(x_k)] ,$$

exactes per a tots els polinomis de grau més petit o igual que $2m+1$.

a) Demostreu que les dites fórmules es poden trobar integrant el polinomi interpolador d'Hermite de f en els $m+1$ punts de la fórmula.

b) Feu explícites aquestes fórmules per a $m=1, 2$, donant així mateix una expressió per als errors comesos.

c) Aplicació: Calculeu

$$\int_0^1 e^{e^x} dx ,$$

fitant l'error comès, emprant totes les fórmules trobades a b).

SOLUCIÓ:

a) Els coeficients d'aquestes fórmules es poden trobar per integració del polinomi interpolador d'Hermite, atès que la fórmula ha de ser exacta per a aquest polinomi, que té grau més petit o igual que $2m+1$ i

$$p_{2m+1}(x_k) = f(x_k) , \quad p'_{2m+1}(x_k) = f'(x_k) \quad (k = 0 \div m) .$$

Així, resulta

$$\begin{aligned}
\int_a^b f(x) dx &\simeq \sum_{k=0}^m [A_k f(x_k) + B_k f'(x_k)] \\
&= \sum_{k=0}^m [A_k p_{2m+1}(x_k) + B_k p'_{2m+1}(x_k)] \\
&= \int_a^b p_{2m+1}(x) dx .
\end{aligned}$$

b) Trobarem primerament les fórmules emprant l'expressió del polinomi interpolador d'Hermite en funció dels polinomis bàsics

$$\begin{aligned}\Phi_k(x) &= (1 - 2l'_k(x_k)(x - x_k))l_k^2(x) , \\ \Psi_k(x) &= (x - x_k)l_k^2(x) \quad (k = 0 \div m) .\end{aligned}$$

Recordem que aquests polinomis compleixen

$$\begin{aligned}\Phi_k(x_i) &= \delta_{ki} , \quad \Phi'_k(x_i) = 0 \quad (k, i = 0 \div m) ; \\ \Psi_k(x_i) &= 0 , \quad \Psi'_k(x_i) = \delta_{ki} \quad (k, i = 0 \div m) ;\end{aligned}$$

i que el polinomi interpolador d'Hermite s'expressa com

$$p_{2m+1}(x) = \sum_{k=0}^m f(x_k)\Phi_k(x) + \sum_{k=0}^m f'(x_k)\Psi_k(x) .$$

Substituint aquesta expressió en l'integrand, trobem

$$\begin{aligned}A_k &= \int_a^b \Phi_k(x)dx = \int_a^b (1 - 2l'_k(x_k)(x - x_k))l_k^2(x)dx \\ &= h \int_0^m (1 - 2\frac{d\ell_k}{ds}(k)(s - k))\ell_k^2(s)ds \equiv h\alpha_k^{(m)} , \\ B_k &= \int_a^b \Psi_k(x)dx = \int_a^b (x - x_k)l_k^2(x)dx \\ &= h^2 \int_0^m (s - k)\ell_k^2(s)ds \equiv h^2\beta_k^{(m)} \\ &\quad (k = 0 \div m) ,\end{aligned}$$

on

$$\ell_k(s) = \prod_{i \neq k} \frac{s - i}{k - i} , \quad \frac{d\ell_k}{ds}(k) = \sum_{i \neq k} \frac{1}{k - i} .$$

Convé remarcar que els coeficients $\alpha_k^{(m)}$ i $\beta_k^{(m)}$ ($k = 0 \div m$) depenen tan sols de k i de m , però no depenen ni dels intervals, ni dels passos triats.

L'expressió de l'error en aquestes fórmules es troba per integració de l'expressió de l'error en la interpolació d'Hermite

$$\int_a^b e_{2m+1}(x)dx = \int_a^b \frac{f^{(2m+2)}(\xi(x))}{(2m+2)!} \omega_m^2(x)dx ;$$

aplicant el teorema del valor mitjà per a integrals, resulta

$$\begin{aligned}E_{2m+1} &= \int_a^b e_{2m+1}(x)dx = \frac{f^{(2m+2)}(\xi)}{(2m+2)!} \int_a^b \omega_m^2(x)dx = \\ &= \frac{f^{(2m+2)}(\xi)}{(2m+2)!} h^{2m+3} \int_0^m \prod_{i=0}^m (s - i)^2 ds \\ &= f^{(2m+2)}(\xi) h^{2m+3} \delta_m , \quad \xi \in (a, b) .\end{aligned}$$

Notem també que el factor δ_m no depèn més que de m .

Calculem els coeficients $\alpha_k^{(m)}$ i $\beta_k^{(m)}$ ($k = 0 \div m$), així com el factor δ_m de l'error, per a $m = 1, 2$.

En el cas $m = 1$, tenim:

$$\begin{aligned}\alpha_0^{(1)} &= \int_0^1 (1 + 2(s - 0)) \frac{(s - 1)^2}{(0 - 1)^2} ds = \frac{1}{2}, \\ \alpha_1^{(1)} &= \int_0^1 (1 - 2(s - 1)) \frac{(s - 0)^2}{(1 - 0)^2} ds = \frac{1}{2}, \\ \beta_0^{(1)} &= \int_0^1 (s - 0) \frac{(s - 1)^2}{(0 - 1)^2} ds = \frac{1}{12}, \\ \beta_1^{(1)} &= \int_0^1 (s - 1) \frac{(s - 0)^2}{(1 - 0)^2} ds = -\frac{1}{12}, \\ \delta_1 &= \frac{1}{24} \int_0^1 s^2 (s - 1)^2 ds = \frac{1}{720}.\end{aligned}$$

En el cas $m = 2$, tenim:

$$\begin{aligned}\alpha_0^{(2)} &= \int_0^2 (1 - 2(-\frac{1}{2} - 1)(s - 0)) \frac{(s - 1)^2 (s - 2)^2}{(0 - 1)^2 (0 - 2)^2} ds = \frac{7}{15}, \\ \alpha_1^{(2)} &= \int_0^2 (1 - 2(1 - 1)(s - 1)) \frac{(s - 0)^2 (s - 2)^2}{(1 - 0)^2 (1 - 2)^2} ds = \frac{16}{15}, \\ \alpha_2^{(2)} &= \int_0^2 (1 - 2(\frac{1}{2} + 1)(s - 2)) \frac{(s - 0)^2 (s - 1)^2}{(2 - 0)^2 (2 - 1)^2} ds = \frac{7}{15}, \\ \beta_0^{(2)} &= \int_0^2 (s - 0) \frac{(s - 1)^2 (s - 2)^2}{(0 - 1)^2 (0 - 2)^2} ds = \frac{1}{15}, \\ \beta_1^{(2)} &= \int_0^2 (s - 1) \frac{(s - 0)^2 (s - 2)^2}{(1 - 0)^2 (1 - 2)^2} ds = 0, \\ \beta_2^{(2)} &= \int_0^2 (s - 2) \frac{(s - 0)^2 (s - 1)^2}{(2 - 0)^2 (2 - 1)^2} ds = -\frac{1}{15}, \\ \delta_2 &= \frac{1}{720} \int_0^2 s^2 (s - 1)^2 (s - 2)^2 ds = \frac{1}{4725}.\end{aligned}$$

Les fórmules cercades són, doncs:

$$\begin{aligned}\int_{x_0}^{x_1} f(x) dx &\simeq \frac{h}{2} [f(x_0) + f(x_1)] \\ &\quad + \frac{h^2}{12} [f'(x_0) - f'(x_1)] \\ &\quad + \frac{h^5}{720} f^{(4)}(\xi), \quad \xi \in (x_0, x_1); \end{aligned}$$

$$\begin{aligned} \int_{x_0}^{x_2} f(x) dx &\simeq \frac{h}{15} [7f(x_0) + 16f(x_1) + 7f(x_2)] + \frac{h^2}{15} [f'(x_0) - f'(x_2)] \\ &\quad + \frac{h^7}{4725} f^{(6)}(\xi) , \quad \xi \in (x_0, x_2) . \end{aligned}$$

c) Per a la funció $f(x) = e^{e^x}$ en $[0, 1]$, aquestes fórmules d'integració donen, respectivament, les aproximacions:

$$\int_0^1 e^{e^x} dx \simeq \frac{1}{2}(e + e^e) + \frac{1}{12}(e - e^e e) = 5.73 ,$$

$$\int_0^1 e^{e^x} dx \simeq \frac{1}{30}(7e + 16e^{\sqrt{e}} + 7e^e) + \frac{1}{60}(e - e^e e) = 6.3025 .$$

Els errors comesos es poden fitar tenint en compte les expressions següents de les derivades de f :

$$\begin{aligned} f'(x) &= e^{e^x} e^x , \\ f^{(2)}(x) &= e^{e^x} (e^{2x} + e^x) , \\ f^{(3)}(x) &= e^{e^x} (e^{3x} + 3e^{2x} + e^x) , \\ f^{(4)}(x) &= e^{e^x} (e^{4x} + 6e^{3x} + 7e^{2x} + e^x) , \\ f^{(5)}(x) &= e^{e^x} (e^{5x} + 10e^{4x} + 25e^{3x} + 15e^{2x} + e^x) , \\ f^{(6)}(x) &= e^{e^x} (e^{6x} + 15e^{5x} + 65e^{4x} + 90e^{3x} + 31e^{2x} + e^x) . \end{aligned}$$

Les derivades quarta i sisena estan fitades, doncs, per:

$$M_4 = e^e (e^4 + 6e^3 + 7e^2 + e) < 3479 ,$$

$$M_6 = e^e (e^6 + 15e^5 + 65e^4 + 90e^3 + 31e^2 + e) < 124538 ,$$

respectivament.

Les fites trobades dels errors seran:

$$|E_3| < \frac{1}{720} M_4 < 4.832... < 4.9 ,$$

$$|E_5| < \frac{1}{4725 \cdot 2^7} M_6 < 0.20592... < 0.21 .$$

Problema 4.5 El període d'un pèndol simple de longitud l , deixat lliure des d'un angle inicial α amb la direcció vertical en un lloc de la terra on l'acceleració de la gravetat val g , és

$$T = 4\sqrt{\frac{l}{g}} \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{1 - K^2 \sin^2 \varphi}} , \quad K = \sin \frac{\alpha}{2} .$$

- a) Trobeu el desenvolupament de Taylor de T com a funció de K .
 b) Descobriu-hi la fórmula aproximada $T \simeq 2\pi\sqrt{\frac{l}{g}}$, vàlida per a petites oscil·lacions.
 c) Quin error es comet usant la fórmula de b) quan $\alpha = 5^\circ$?

SOLUCIÓ:

a) El desenvolupament de la integral en funció de K es troba integrant terme a terme el desenvolupament de l'integrand de la forma que segueix:

$$\begin{aligned} \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{1 - K^2 \sin^2 \varphi}} &= \int_0^{\frac{\pi}{2}} (1 - K^2 \sin^2 \varphi)^{-\frac{1}{2}} d\varphi \\ &= \int_0^{\frac{\pi}{2}} \sum_{j=0}^{\infty} \binom{-\frac{1}{2}}{j} (-K^2 \sin^2 \varphi)^j d\varphi \\ &= \sum_{j=0}^{\infty} (-1)^j \binom{-\frac{1}{2}}{j} \left(\int_0^{\frac{\pi}{2}} \sin^{2j} \varphi d\varphi \right) K^{2j} \\ &= \frac{\pi}{2} + \frac{\pi}{2} \sum_{j=1}^{\infty} \left(\frac{(2j-1)!!}{(2j)!!} \right)^2 K^{2j} . \end{aligned}$$

Fent-lo explícit i substituint-lo en l'expressió del període del pèndol trobem

$$T = 2\pi\sqrt{\frac{l}{g}} \left(1 + \frac{K^2}{4} + \frac{9K^4}{64} + \frac{25K^6}{256} + \dots \right) .$$

b) Notem que el primer terme del desenvolupament dona l'expressió aproximada quan es tracta de petites oscil·lacions del pèndol $T \simeq 2\pi\sqrt{\frac{l}{g}}$.

c) L'error relatiu que es comet emprant la fórmula aproximada de b) vindrà donat pel desenvolupament

$$\frac{K^2}{4} + \frac{9K^4}{64} + \frac{25K^6}{256} + \dots ,$$

que pot ser majorat per una sèrie geomètrica de raó $K^2 = \sin^2 2.5^\circ$, resultant la fita aproximada

$$\frac{K^2}{4} \frac{1}{1 - K^2} = \frac{1}{4} \tan^2 2.5^\circ = 0.4766... \cdot 10^{-3} < 0.5 \cdot 10^{-3} .$$

Problema 4.6 Considerem les integrals del tipus

$$\int_{-1}^1 (1 + x^4) f(x) dx .$$

a) Trobeu una fórmula de tres abscisses que sigui exacta per a tot polinomi de grau més petit o igual que 5 i doneu una expressió per a l'error comès si $f \in \mathcal{C}^6([-1, 1])$.

b) Aplicació: Calculeu la integral per a $f(x) = \sin^2 x$, fitant l'error comès.

SOLUCIÓ:

a) La fórmula cercada és necessàriament gaussiana; és a dir, té com a abscisses els zeros del polinomi $\psi_3(x)$ ortogonal respecte al producte escalar

$$(f, g) = \int_{-1}^1 (1 + x^4) f(x) g(x) dx$$

i com a pesos aquells que fan la fórmula exacta per a tots els polinomis de grau més petit o igual que 2.

Els polinomis ortogonals es poden trobar per recurrència; però, com que només necessitem calcular-los fins a grau 3, se'n pot fer el càlcul més directament.

La simetria de l'interval d'integració i la paritat de la funció pes impliquen la paritat dels polinomis ortogonals: $\psi_j(-x) = (-1)^j \psi_j(x)$ ($j \geq 0$). Així, si cerquem els polinomis ortogonals mòncics, tenim

$$\psi_0(x) = 1, \quad \psi_1(x) = x, \quad \psi_2(x) = x^2 - \alpha, \quad \psi_3(x) = x^3 - \beta x.$$

El polinomi $\psi_3(x)$, atès que no té termes parells, és automàticament ortogonal a $\psi_0(x)$ i $\psi_2(x)$. El coeficient indeterminat β d'aquest polinomi es troba imposant la seva ortogonalitat amb $\psi_1(x)$, resultant-ne

$$\beta = \frac{(x^3, x)}{(x, x)} = \frac{49}{75}.$$

El polinomi ortogonal $\psi_3(x)$ té, doncs, els zeros $x_0 = -\sqrt{\frac{49}{75}}$, $x_1 = 0$ i $x_2 = \sqrt{\frac{49}{75}}$ i, tenint en compte que $\psi_3(x)$ és mònic, la fórmula gaussiana cercada té la forma

$$\begin{aligned} \int_{-1}^1 (1 + x^4) f(x) dx &= W_0 f\left(-\sqrt{\frac{49}{75}}\right) + W_1 f(0) + W_2 f\left(\sqrt{\frac{49}{75}}\right) \\ &\quad + \frac{f^{(6)}(\xi)}{6!} (\psi_3, \psi_3). \end{aligned}$$

En trobarem ara els pesos, imposant exactitud per als polinomis de grau més petit o igual que 2, i n'explicitarem l'expressió de l'error.

El sistema lineal corresponent a imposar l'exactitud de la fórmula per a la base $\{1, x, x^2\}$ és

$$\begin{array}{lcl} f(x) = 1 & \left| \begin{array}{l} \frac{12}{5} \\ 0 \\ \frac{20}{21} \end{array} \right. & = \begin{array}{l} W_0 + W_1 + W_2 \\ -\sqrt{\frac{49}{75}} W_0 + \sqrt{\frac{49}{75}} W_2 \\ \frac{49}{75} W_0 + \frac{49}{75} W_2 \end{array} \end{array}.$$

La resolució del sistema dóna $W_0 = W_2 = \frac{250}{343}$ i $W_1 = \frac{1616}{1715}$.

Calculant primerament

$$(\psi_3, \psi_3) = \int_{-1}^1 (1+x^4) \psi_3^2(x) dx = \frac{15856}{259875} ,$$

trobem l'expressió de l'error.

Resumint, la fórmula gaussiana cercada és

$$\begin{aligned} \int_{-1}^1 (1+x^4) f(x) dx &= \frac{250}{343} f\left(-\sqrt{\frac{49}{75}}\right) + \frac{1616}{1715} f(0) + \frac{250}{343} f\left(\sqrt{\frac{49}{75}}\right) \\ &\quad + \frac{991}{11694375} f^{(6)}(\xi) , \quad \xi \in (-1, 1) . \end{aligned}$$

b) La fórmula dona l'aproximació següent de la integral:

$$J = \int_{-1}^1 (1+x^4) \sin^2 x \, dx \simeq \frac{500}{343} \sin^2 \sqrt{\frac{49}{75}} = 0.76222...$$

Per tal de fitar l'error comès

$$\left| \frac{991}{11694375} f^{(6)}(\xi) \right| \leq \frac{991}{11694375} M_6 ,$$

cal trobar una fita M_6 de la derivada sisena de la funció. De la relació $\sin^2 x = \frac{1}{2}(1 - \cos 2x)$, es troben fàcilment les derivades successives de la funció i tenim

$$f^{(6)}(x) = 2^5 \cos 2x ,$$

que es pot fitar per $M_6 = 32$.

Resulta, finalment,

$$J \simeq 0.76222... \pm \frac{991 \cdot 32}{11694375} = 0.76222... \pm 2.72 \cdot 10^{-3} .$$

De fet, la integral demanada es pot calcular exactament, descomponent-la com

$$\begin{aligned} J &= \int_0^1 (1 - \cos 2x) dx + \int_0^1 x^4 dx - \int_0^1 x^4 \cos 2x \, dx \\ &= \frac{6}{5} - \frac{1}{4} \sin 2 + \frac{1}{2} \cos 2 = 0.7646022... ; \end{aligned}$$

finalment,

$$J - 0.76222... \simeq 2.38 \cdot 10^{-3} ,$$

que mostra la gran similitud entre la fita estimada de l'error i l'error real.

Problema 4.7 El període d'un pèndol simple de longitud l , deixat lliure des d'un angle inicial α amb la direcció vertical en un lloc de la terra on l'acceleració de la gravetat val g , és

$$T = 4\sqrt{\frac{l}{g}} \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{1 - K^2 \sin^2 \varphi}}, \quad K = \sin \frac{\alpha}{2}.$$

Calculeu T amb un error relatiu menor que 10^{-4} per a $\alpha = 30^\circ$, emprant una fórmula de Gauss-Txebixev adient.

SOLUCIÓ:

La integral proposada es pot convertir, mitjançant el canvi $t = \sin \varphi$, en una integral a l'interval $[-1, 1]$ en què apareix el pes $w(t) = \frac{1}{\sqrt{1-t^2}}$ de les fórmules de Gauss-Txebixev en aquest interval,

$$\begin{aligned} T &= 4\sqrt{\frac{l}{g}} \int_0^1 \frac{dt}{\sqrt{1-t^2}\sqrt{1-K^2t^2}} \\ &= 2\sqrt{\frac{l}{g}} \int_{-1}^1 \frac{dt}{\sqrt{1-t^2}\sqrt{1-K^2t^2}}, \end{aligned}$$

ateses les paritats del pes i de la funció

$$f(t) = \frac{1}{\sqrt{1-K^2t^2}}.$$

La fórmula de Gauss-Txebixev de $m+1$ abscisses s'escriu ara així

$$\begin{aligned} T &= 2\sqrt{\frac{l}{g}} \left[\frac{\pi}{m+1} \sum_{k=0}^m \frac{1}{\sqrt{1-K^2 \cos^2(\frac{(2k+1)\pi}{2(m+1)})}} \right. \\ &\quad \left. + \frac{\pi}{2^{2m+1}(2m+2)!} f^{(2m+2)}(\xi) \right], \end{aligned}$$

amb $\xi \in (-1, 1)$ i $K = \sin 15^\circ = 0.258819045\dots$

Per a $m = 0, 1, 2$, tenim, respectivament,

$$\begin{aligned} T &= 2\pi\sqrt{\frac{l}{g}} \left[1 + \frac{1}{4} f^{(2)}(\xi) \right], \\ T &= 2\pi\sqrt{\frac{l}{g}} \left[\frac{1}{2} \left(\frac{1}{\sqrt{1-\frac{1}{2}K^2}} + \frac{1}{\sqrt{1-\frac{1}{2}K^2}} \right) + \frac{1}{192} f^{(4)}(\xi) \right], \\ T &= 2\pi\sqrt{\frac{l}{g}} \left[\frac{1}{3} \left(\frac{1}{\sqrt{1-\frac{3}{4}K^2}} + 1 + \frac{1}{\sqrt{1-\frac{3}{4}K^2}} \right) + \frac{1}{23040} f^{(6)}(\xi) \right]. \end{aligned}$$

A fi d'avaluar els errors per trobar un valor adequat de m cal fitar les derivades parelles corresponents de la funció f :

$$\begin{aligned} f(t) &= (1 - K^2 t^2)^{-\frac{1}{2}} , \\ f'(t) &= K^2 t f^3(t) , \\ f^{(2)}(t) &= K^2 (1 + 2K^2 t^2) f^5(t) , \\ f^{(3)}(t) &= 3K^4 t (3 + 2K^2 t^2) f^7(t) , \\ f^{(4)}(t) &= 3K^4 (3 + 24K^2 t^2 + 8K^4 t^4) f^9(t) , \\ f^{(5)}(t) &= 15K^6 t (15 + 40K^2 t^2 + 8K^4 t^4) f^{11}(t) , \\ f^{(6)}(t) &= 45K^6 (5 + 90K^2 t^2 + 120K^4 t^4 + 16K^6 t^6) f^{13}(t) . \end{aligned}$$

Per tant,

$$M_{2r} \equiv \max_{t \in [-1,1]} |f^{(2r)}(t)| = f^{(2r)}(1) \quad (r = 1, 2, 3) .$$

Així, obtenim

$$M_2 \simeq 0.0903 , \quad M_4 \simeq 0.0854 , \quad M_6 \simeq 0.2457 .$$

Les fites dels errors corresponents són, per a $m = 0, 1, 2$, respectivament,

$$\frac{M_2}{4} \leq 226 \cdot 10^{-4} , \quad \frac{M_4}{192} \leq 4.5 \cdot 10^{-4} , \quad \frac{M_6}{23040} \leq 0.11 \cdot 10^{-4} .$$

La fórmula adequada és, doncs, la de $m = 2$ que dona l'aproximació següent de T :

$$\begin{aligned} T &= 2\pi \sqrt{\frac{l}{g}} \left[\frac{1.0261082 + 1 + 1.0261082}{3} \pm 1.1 \cdot 10^{-5} \right] \\ &= 2\pi \sqrt{\frac{l}{g}} (1.017405 \pm 2 \cdot 10^{-5}) . \end{aligned}$$

Problema 4.8 a) Trobeu les relacions de recurrència següents per als polinomis de Legendre, per a $t \in \mathbb{R}$ i $j \geq 1$:

- i) $P'_j(t) - tP'_{j-1}(t) = jP_{j-1}(t)$,
- ii) $tP'_j(t) - P'_{j-1}(t) = jP_j(t)$,
- iii) $P'_{j+1}(t) - P'_{j-1}(t) = (2j+1)P_j(t)$,
- iv) $(t^2 - 1)P'_j(t) = jtP_j(t) - jP_{j-1}(t)$.

b) Deduïu la fórmula de Gauss-Legendre de $m+1$ punts

$$\int_a^b f(x)dx = \frac{b-a}{2} \sum_{k=0}^m w_k f(x_k) + E_{m+1}(f) ,$$

amb

$$x_k = \frac{b-a}{2}t_k + \frac{a+b}{2}, \quad w_k = \frac{2}{(1-t_k)^2[P'_{m+1}(t_k)]^2},$$

on $t_k \in (-1, 1)$ ($k = 0 \div m$) són els zeros del polinomi de Legendre $P_{m+1}(t)$ i, si la funció $f \in \mathcal{C}^{(2m+2)}([a, b])$, l'error de la fórmula anterior ve donat per l'expressió

$$E_{m+1}(f) = \frac{(b-a)^{2m+3}[(m+1)!]^4}{(2m+3)[(2m+2)!]^3} f^{(2m+2)}(\xi), \quad \xi \in (a, b).$$

c) Feu explícites les fórmules de Gauss-Legendre d'1, 2 i 3 punts sobre $[-1, 1]$.

SOLUCIÓ:

Recordem del capítol anterior que els polinomis de Legendre estan definits per

$$P_0(t) = 1, \quad P_j(t) = \frac{1}{2^j j!} \frac{d^j}{dt^j} [(t^2 - 1)^j] \quad (j \geq 1)$$

i són ortogonals respecte al pes $w(t) = 1$ a l'interval $[-1, 1]$:

$$(P_j, P_l) = \int_{-1}^1 P_j(t) P_l(t) dt = \frac{2}{2j+1} \delta_{jl} \quad (j, l \geq 0).$$

A més, tenen coeficient principal

$$A_j = \frac{(2j)!}{2^j (j!)^2},$$

i satisfan la relació de recurrència

$$\begin{aligned} P_0(t) &= 1, \\ P_1(t) &= t, \\ (j+1)P_{j+1}(t) &= (2j+1)tP_j(t) - jP_{j-1}(t) \quad (j \geq 1). \quad (*) \end{aligned}$$

a) D'ara endavant suposarem $j \geq 1$.

Derivant la definició de $P_j(t)$, obtenim:

$$\begin{aligned} P'_j(t) &= \frac{1}{2^j j!} \frac{d^{j+1}}{dt^{j+1}} [(t^2 - 1)^j] = \frac{1}{2^j j!} \frac{d^j}{dt^j} [2jt(t^2 - 1)^{j-1}] \\ &= \frac{1}{2^{j-1}(j-1)!} \left(t \frac{d^j}{dt^j} [(t^2 - 1)^{j-1}] + j \frac{d^{j-1}}{dt^{j-1}} [(t^2 - 1)^{j-1}] \right) \\ &= tP'_{j-1}(t) + jP_{j-1}(t); \end{aligned}$$

on hem usat que

$$\frac{d^j}{dt^j} [tg(t)] = t \frac{d^j}{dt^j} [g(t)] + j \frac{d^{j-1}}{dt^{j-1}} [g(t)].$$

Ara les altres relacions s'obtenen ràpidament. Així, derivant la relació de recurrència (*) i substituint $P'_{j+1}(t)$ en la relació i) per a l'índex $j+1$, s'obté ii). Sumant la relació ii)

a la relació i) per a l'índex $j + 1$, surt iii). Finalment, iv) es troba aïllant $P'_{j-1}(t)$ en i) i ii) i igualant.

b) Començarem suposant $[a, b] = [-1, 1]$. Per l'ortogonalitat dels polinomis de Legendre respecte a la funció $w(t) = 1$ en $[-1, 1]$, tenim la fórmula gaussiana de $m + 1$ abscisses ($m \geq 0$):

$$\int_{-1}^1 g(t) dt = \sum_{k=0}^m w_k g(t_k) + E_{m+1}(g) ,$$

on $t_k \in (-1, 1)$ ($k = 0 \div m$) són els zeros del polinomi de Legendre $P_{m+1}(t)$, i

$$\begin{aligned} E_{m+1}(g) &= \frac{(P_{m+1}, P_{m+1})}{A_{m+1}^2} \frac{g^{(2m+2)}(\eta)}{(2m+2)!} \\ &= \frac{2^{2m+3}[(m+1)!]^4}{(2m+3)[(2m+2)!]^3} g^{(2m+2)}(\eta) , \quad \eta \in (-1, 1) , \end{aligned}$$

ja que $g \in \mathcal{C}^{2m+2}([-1, 1])$. Els pesos w_k ($k = 0 \div m$) es poden obtenir imposant que la fórmula gaussiana sigui exacta per a polinomis de grau menor o igual que m ; això és, que $E_{m+1}(t^i) = 0$ ($i = 0 \div m$).

Ara bé, tenim també l'expressió següent per als pesos:

$$w_k = \frac{A_{m+1}(P_m, P_m)}{A_m P'_{m+1}(t_k) P_m(t_k)} = \frac{2}{(m+1) P_m(t_k) P'_{m+1}(t_k)} ;$$

aplicant la relació iv), obtenim

$$(t_k^2 - 1) P'_{m+1}(t_k) = -(m+1) P_m(t_k) ,$$

i, per tant,

$$w_k = \frac{2}{(1 - t_k^2) [P'_{m+1}(t_k)]^2} \quad (k = 0 \div m) .$$

El cas general d'una integral

$$\int_a^b f(x) dx$$

es redueix al cas anterior mitjançant el canvi

$$x = \frac{b-a}{2} t + \frac{a+b}{2} ;$$

aleshores

$$\int_a^b f(x) dx = \frac{b-a}{2} \int_{-1}^1 g(t) dt ,$$

on

$$g(t) = f\left(\frac{b-a}{2} t + \frac{a+b}{2}\right) , \quad t \in [-1, 1] ,$$

compleix

$$g^{(2m+2)}(\eta) = \left(\frac{b-a}{2}\right)^{2m+2} f^{(2m+2)}(\xi) , \quad \xi = \frac{b-a}{2} \eta + \frac{a+b}{2} \in (a, b) .$$

c) Aplicant l'apartat b), trobem les fórmules de Gauss-Legendre sobre $[-1, 1]$ per a $m = 0, 1, 2$:

$$m = 0: \quad P_1(t) = t ,$$

$$t_0 = 0 ,$$

$$P'_1(t_0) = 1 ;$$

$$\int_{-1}^1 g(t)dt = 2g(0) + \frac{1}{3}g^{(2)}(\eta) .$$

$$m = 1: \quad P_2(t) = \frac{1}{2}(3t^2 - 1) ,$$

$$t_1 = -t_0 = \frac{\sqrt{3}}{3} ,$$

$$P'_2(t_1) = -P'_2(t_0) = \sqrt{3} ;$$

$$\int_{-1}^1 g(t)dt = g\left(\frac{-\sqrt{3}}{3}\right) + g\left(\frac{\sqrt{3}}{3}\right) + \frac{g^{(4)}(\eta)}{135} .$$

$$m = 2: \quad P_3(t) = \frac{5}{2}t(t^2 - \frac{3}{5}) ,$$

$$t_1 = 0 , \quad t_2 = -t_0 = \frac{\sqrt{15}}{5} ,$$

$$P'_3(t_1) = -\frac{3}{2} , \quad P'_3(t_2) = P'_3(t_0) = 3 ;$$

$$\int_{-1}^1 g(t)dt = \frac{5}{9}g\left(-\frac{\sqrt{15}}{5}\right) + \frac{8}{9}f(0) + \frac{5}{9}g\left(\frac{\sqrt{15}}{5}\right) + \frac{g^{(6)}(\eta)}{15750} .$$

Problema 4.9 Calculeu, usant el mètode de comparació,

$$S = \sum_{j=1}^{\infty} \frac{1}{j^2}$$

amb 4 xifres exactes.

SOLUCIÓ:

Considerem primer la sèrie telescòpica

$$\sum_{j=1}^{\infty} \frac{1}{j(j+1)} = \sum_{j=0}^{\infty} \frac{1}{(j+1)(j+2)} = \sum_{j=0}^{\infty} j^{(-2)} = 0^{(-1)} = 1 ;$$

així,

$$S = 1 + \sum_{j=1}^{\infty} \left(\frac{1}{j^2} - \frac{1}{j(j+1)} \right) = 1 + \sum_{j=1}^{\infty} \frac{1}{j^2(j+1)} \equiv 1 + S_1 .$$

Prenent ara la segona sèrie telescòpica

$$\begin{aligned} \sum_{j=2}^{\infty} \frac{1}{(j-1)j(j+1)} &= \sum_{j=0}^{\infty} \frac{1}{(j+1)(j+2)(j+3)} \\ &= \sum_{j=0}^{\infty} j^{(-3)} = \frac{0^{(-2)}}{2} = \frac{1}{4} ; \end{aligned}$$

tenim,

$$\begin{aligned} S &= 1 + \frac{1}{2} + \sum_{j=2}^{\infty} \frac{1}{j^2(j+1)} \\ &= 1 + \frac{1}{2} + \frac{1}{4} + \sum_{j=2}^{\infty} \left(\frac{1}{j^2(j+1)} - \frac{1}{(j-1)j(j+1)} \right) \\ &= \frac{7}{4} - \sum_{j=2}^{\infty} \frac{1}{j^2(j^2-1)} . \end{aligned}$$

Considerant

$$\begin{aligned} \sum_{j=3}^{\infty} \frac{1}{(j-2)(j-1)j(j+1)} \\ &= \sum_{j=0}^{\infty} \frac{1}{(j+1)(j+2)(j+3)(j+4)} \\ &= \sum_{j=0}^{\infty} j^{(-4)} = \frac{0^{(-3)}}{3} = \frac{1}{18} , \end{aligned}$$

tindrem

$$\begin{aligned} S &= \frac{7}{4} - \frac{1}{12} - \sum_{j=3}^{\infty} \frac{1}{j^2(j^2-1)} \\ &= \frac{5}{3} - \frac{1}{18} - \sum_{j=3}^{\infty} \left(\frac{1}{j^2(j^2-1)} - (j-3)^{(-4)} \right) \\ &= \frac{29}{18} + 2 \sum_{j=3}^{\infty} \frac{1}{j^2(j-1)(j+1)(j-2)} . \end{aligned}$$

Per a aquesta darrera sèrie, els residus es poden fitar pel criteri integral

$$\begin{aligned} R_n &= 2 \sum_{j=n+1}^{\infty} \frac{1}{j^2(j-1)(j^2-j-2)} < 2 \sum_{j=n+1}^{\infty} \frac{1}{(j-1)^5} \\ &< 2 \int_n^{\infty} \frac{dx}{(x-1)^5} = \frac{1}{2(n-1)^4} ; \end{aligned}$$

si $n = 6$, $0 \leq R_6 \leq 0.8 \cdot 10^{-3}$ i

$$\begin{aligned} S &= \frac{29}{18} + 2 \sum_{j=3}^6 \frac{1}{j^2(j^2-1)(j-2)} + R_6 \\ &= 1.64456 + R_6 = 1.64496 \pm 0.4 \cdot 10^{-3} . \end{aligned}$$

Notem que el resultat exacte és $S = \frac{\pi^2}{6} = 1.644934\dots$, que és un cas particular del resultat

$$\zeta(2s) = \sum_{j=1}^{\infty} \frac{1}{j^{2s}} = \frac{(2\pi)^{2s}}{2(2s)!} |B_{2s}| \quad (s \geq 1) ,$$

on ζ és l'anomenada funció *zeta de Riemann* i els coeficients B_{2s} són els nombres de Bernoulli.

Problema 4.10 La constant d'Euler es defineix com

$$\gamma = \lim_{n \rightarrow \infty} F(n) ,$$

essent

$$F(n) = 1 + \frac{1}{2} + \dots + \frac{1}{n-1} + \frac{1}{2n} - \ln n .$$

Calculeu γ amb 10 xifres correctes.

SOLUCIÓ:

Per tal de calcular

$$F(N) = \sum_{j=1}^{N-1} \frac{1}{j} + \frac{1}{2N} - \ln N = \sum_{j=1}^N \frac{1}{j} - \frac{1}{2N} - \ln N ,$$

triem $M < N$ i apliquem la fórmula d'Euler-Maclaurin per a sumes amb $a = M \geq 1$, $n = N - M$, $h = 1$ i $f(x) = \frac{1}{x}$.

Com que $f^{(j)}(x) = (-1)^j j! x^{-(j+1)}$, la fórmula es converteix en

$$\begin{aligned} F(N) - F(M) &= \sum_{j=M}^N \frac{1}{j} - \ln N + \ln M - \frac{1}{2} \left(\frac{1}{M} + \frac{1}{N} \right) \\ &+ \sum_{r=1}^s \frac{B_{2r}}{2r} \left(\frac{1}{M^{2r}} - \frac{1}{N^{2r}} \right) + R_s \end{aligned}$$

on

$$R_s = (N - M) B_{2s+2} \frac{1}{\xi^{2s+3}} , \quad \xi \in (M, N) .$$

Els nombres B_{2r} ($r \geq 1$) que apareixen en la fórmula són els nombres de Bernoulli (consulteu el problema III.11):

$$B_{2r} = \frac{1}{6}, -\frac{1}{30}, \frac{1}{42}, -\frac{1}{30}, \dots \quad (r \geq 1),$$

que satisfan la propietat $B_{2r}B_{2r+2} < 0$ ($r \geq 1$). Per tant, $R_s R_{s+1} < 0$ i, pel criteri d'alternança dels residus, la magnitud de R_s és menor o igual que la del primer terme menyspreat

$$|R_s| \leq \frac{|B_{2s+2}|}{2s+2} \left(\frac{1}{M^{2s+2}} - \frac{1}{N^{2s+2}} \right) \leq \frac{|B_{2s+2}|}{2s+2} \frac{1}{M^{2s+2}}.$$

Així, fent tendir N a infinit en l'expressió trobada de $F(N) - F(M)$, tenim

$$\begin{aligned} \gamma &= F(M) + \sum_{r=1}^s \frac{B_{2r}}{2r} \frac{1}{M^{2r}} \pm \frac{|B_{2s+2}|}{2s+2} \frac{1}{M^{2s+2}} \\ &= F(M) + \frac{1}{12M^2} - \frac{1}{120M^4} + \frac{1}{252M^6} - \frac{1}{240M^8} + \dots \\ &\quad \pm \frac{|B_{2s+2}|}{2s+2} \frac{1}{M^{2s+2}}. \end{aligned}$$

Per comoditat de càlcul, prenem $M = 10$; notem que el terme en M^8 ja és menor que l'error permès per al càlcul de γ

$$\frac{1}{240M^8} < \frac{1}{2} 10^{-10}.$$

Així,

$$\begin{aligned} \gamma &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{20} - \ln 10 \\ &\quad + \frac{1}{1200} - \frac{1}{1200000} + \frac{1}{252000000} \pm \frac{1}{2} 10^{-10} \\ &= 0.57721566494 \pm \frac{1}{2} 10^{-10}. \end{aligned}$$

Problema 4.11 La funció d'Airy

$$\text{Ai}(x) = \frac{1}{\pi} \int_0^\infty \cos\left(\frac{t^3}{3} + xt\right) dt,$$

admet el desenvolupament asimptòtic següent:

$$\text{Ai}(x) \sim \frac{1}{2\sqrt{\pi x^{\frac{1}{4}}}} e^{-\zeta} \sum_{j=0}^{\infty} (-1)^j \frac{c_j}{\zeta^j} \quad (x \rightarrow \infty),$$

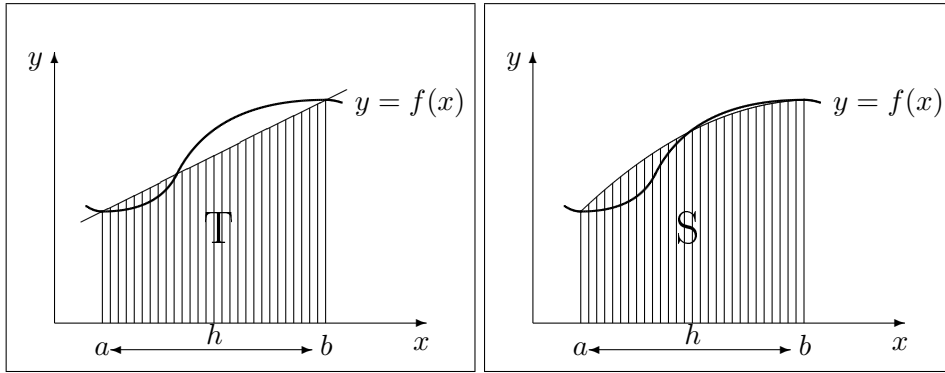


Figura 4.3: Comportament dels termes d'una sèrie semiconvergent.

amb

$$\zeta = \frac{2}{3}x^{\frac{3}{2}},$$

i

$$c_0 = 1, \quad c_j = \frac{(2j+1)(2j+3)\cdots(6j-1)}{216^j j!} \quad (j \geq 1).$$

Fent servir que l'error d'aproximació de la funció d'Airy pels primers termes del desenvolupament és menor que la magnitud del primer terme menyspreat, trobeu $\text{Ai}(4)$ amb el menor error que permet el dit desenvolupament.

SOLUCIÓ:

Per a $x = 4$ resulta $\zeta = \frac{16}{3}$ i

$$\text{Ai}(4) \sim B \sum_{j=0}^n a_j,$$

amb

$$B = \frac{1}{2\sqrt{2\pi}} e^{-\frac{16}{3}}, \quad a_j = (-1)^j \left(\frac{3}{16}\right)^j c_j \quad (j \geq 0).$$

Cercarem aquell valor n que fa mínim $|a_{n+1}|$, aleshores calcularem

$$\text{Ai}(4) = B \left(\sum_{j=0}^n a_j \pm |a_{n+1}| \right).$$

Per tal de trobar el mínim dels $|a_{j+1}|$ calculem

$$\begin{aligned} \left| \frac{a_{j+1}}{a_j} \right| &= \frac{3}{16} \frac{c_{j+1}}{c_j} = \frac{3}{16} \frac{(6j+5)(6j+3)(6j+1)}{216(2j+1)(j+1)} \\ &= \frac{(6j+5)(6j+1)}{384(j+1)}; \end{aligned}$$

notem que, si $|a_{n+1}|$ és mínim, complirà

$$\left| \frac{a_{n+1}}{a_n} \right| < 1, \quad \left| \frac{a_{n+2}}{a_{n+1}} \right| > 1.$$

La condició

$$\left| \frac{a_{j+1}}{a_j} \right| < 1$$

equivale a

$$36j^2 - 348j - 379 < 0;$$

com que els zeros del polinomi $36x^2 - 348x - 379$ són aproximadament -0.988 i 10.65, resulta que $n = 10$ és el nombre natural més gran per al qual

$$\left| \frac{a_{n+1}}{a_n} \right| < 1,$$

i a_{11} és el menor dels termes, en valor absolut.

Finalment,

$$\text{Ai}(4) = B \left(\sum_{j=0}^{10} a_j \pm |a_{11}| \right) = 0.0009515652 \pm 3 \cdot 10^{-9}.$$

Notem que, usant aquesta sèrie semiconvergent, no es pot trobar el valor demanat amb precisió arbitrària, atès que l'error mínim d'aproximació és el trobat. No serviria de res calcular més termes del desenvolupament: l'error comès seria encara més gran. Vegeu la figura 4.3.

Problema 4.12 Considerem la integral

$$I = \int_0^{\frac{\pi}{2}} \sin x \, dx = 1.$$

a) Escriviu la fórmula d'Euler-Maclaurin per a la dita integral, prenent termes fins a ordre 8 en el pas h .

b) Aproximeu la integral per la regla dels trapezis, dividint l'interval en M subinterval ($M = 2, 4, 8, 16, 32, 64$), treballant amb 9 xifres significatives.

c) Realitzeu una taula d'extrapolació repetida de Richardson, basada en la fórmula trobada en a), a partir dels resultats trobats a b).

d) Obteniu els primers termes de les fórmules asimptòtiques d'error per a les tres primeres fórmules extrapolades, quan el pas h tendeix a zero.

SOLUCIÓ:

a) Per a $f(x) = \sin x$ i $[a, b] = [0, \frac{\pi}{2}]$, els factors $f^{(2r-1)}(b) - f^{(2r-1)}(a)$ que apareixen en la fórmula d'Euler-Maclaurin resulten ser iguals a $(-1)^r$.

Tenim, doncs, el següent desenvolupament asimptòtic de la fórmula dels trapezis en aquest cas:

$$\begin{aligned} T(h) &= I + \sum_{r=1}^{\infty} (-1)^r \frac{B_{2r}}{(2r)!} h^{2r} \\ &= I - \frac{h^2}{12} - \frac{h^4}{720} - \frac{h^6}{30240} - \frac{h^8}{1209600} - \dots \end{aligned}$$

b) Per a $M = 2, 4, 8, 16, 32, 64$, s'ha de calcular

$$T(h) = \frac{h}{2} (f_0 + 2f_1 + \dots + 2f_{M-1} + f_M) ,$$

amb $h = \frac{\pi}{2M}$ i $f_k = \sin \frac{k\pi}{2M}$ ($k = 0 \div M$). Notem que els càlculs fets per tal d'avaluar qualsevol $T(h)$ es poden aprofitar per al següent $T(\frac{h}{2})$.

Els resultats obtinguts es donen en la taula següent:

M	$T(h)$
2	0.948059449
4	0.987115801
8	0.996785172
16	0.999196681
32	0.999799194
64	0.999949800

c) Tenim una expressió asimptòtica de la forma

$$T(h) = I + a_1 h^2 + a_2 h^4 + a_3 h^6 + a_4 h^8 + \dots , \quad a_r = (-1)^r \frac{B_{2r}}{(2r)!} \neq 0 .$$

Són, doncs, adequades les extrapolacions dels tipus

$$\frac{\Delta}{2^{p_j} - 1} , \quad p_j = 2j \quad (j \geq 1) ;$$

és a dir,

$$\frac{\Delta}{3} , \quad \frac{\Delta}{15} , \quad \frac{\Delta}{63} , \quad \frac{\Delta}{255} , \quad \frac{\Delta}{1023} , \quad \dots$$

L'avaluació efectiva d'integrals extrapolant la regla dels trapezis tal com s'indica anteriorment rep el nom de *méthode de Romberg*.

El mètode de Romberg porta aleshores a terme aquestes extrapolacions, segons l'esquema

$$T_1(h) = T(h) , \quad T_{j+1}(h) = T_j(h) + \frac{T_j(h) - T_j(2h)}{4^j - 1} \quad (j \geq 1) ;$$

els resultats obtinguts es recullen en la taula següent:

T_1	Δ	$T_2 = T_1 + \frac{\Delta}{3}$	Δ	$T_3 = T_2 + \frac{\Delta}{15}$
0.948059449				
0.987115801	0.039056352	1.000134585		
0.996785172	0.009669371	1.000008296	-0.000126289	0.999999876
0.999196681	0.002411509	1.000000517	-0.000007779	0.999999998
0.999799194	0.000602513	1.000000032	-0.000000485	1.000000000
0.999949800	0.000150706	1.000000002	-0.000000030	1.000000000

d) Les expressions asimptòtiques dels errors en les diverses fórmules extrapolades són de la forma

$$T_j(h) - I = a_j^{(j)} h^{2j} + a_{j+1}^{(j)} h^{2j+2} + \dots,$$

on els coeficients es troben mitjançant la recurrència

$$a_l^{(j+1)} = a_l^{(j)} \left(1 + \frac{1 - 4^l}{4j - 1}\right) = \frac{4^j - 4^l}{4j - 1} a_l^{(j)} \quad (l \geq j) \quad (j \geq 1).$$

Això es veu substituint les expressions asimptòtiques de $T_j(h)$ i de $T_j(2h)$ en la fórmula d'extrapolació. Noteu que s'aconsegueix la finalitat de l'extrapolació repetida, que és fer $a_j^{(j+1)} = 0$.

Així, els factors de variació dels diversos coeficients seran:

$$\begin{aligned} \frac{a_l^{(2)}}{a_l^{(1)}} &= -\frac{4^l - 4}{3} = 0, -4, -20, -84, -340, \dots \quad (l \geq 1); \\ \frac{a_l^{(3)}}{a_l^{(2)}} &= -\frac{4^l - 16}{15} = 0, -\frac{16}{5}, -16, -\frac{336}{5}, \dots \quad (l \geq 2); \\ \frac{a_l^{(4)}}{a_l^{(3)}} &= -\frac{4^l - 64}{63} = 0, -\frac{64}{21}, -\frac{320}{21}, \dots \quad (l \geq 3). \end{aligned}$$

Les fórmules extrapolades tindran, doncs, els errors asimptòtics següents, quan h tendeix a 0:

$$\begin{aligned} T_2(h) - I &= \frac{h^4}{180} + \frac{h^6}{1512} + \frac{h^8}{14400} + \dots, \\ T_3(h) - I &= -\frac{2h^6}{945} - \frac{h^8}{900} - \dots, \\ T_4(h) - I &= \frac{16h^8}{4725} + \dots \end{aligned}$$

Problema 4.13 Supposem el desenvolupament asimptòtic per a la regla dels trapezis

$$T(h) = \int_a^b f(x) dx + a_1 h^2 + a_2 h^4 + \dots + a_r h^{2r} + \dots,$$

amb $a_r \neq 0$ ($r \geq 1$).

Considerem les fórmules de quadratura $T_j(\frac{h_0}{2^{j-1}})$ obtingudes pel mètode de Richardson aplicat a la regla dels trapezis $T_1(h) \equiv T(h)$ amb passos $h_0, \frac{h_0}{2}, \dots, \frac{h_0}{2^{j-1}}$, on $h_0 = b - a$.

a) Trobeu les dites fórmules $T_j(\frac{h_0}{2^{j-1}})$ per a $j = 2, 3$ i identifiqueu-les amb fórmules de Newton-Cotes.

b) Demostreu que la fórmula $T_j(h)$ té un error asimptòtic de la forma $a_j^{(j)}h^{2j} + \dots$ amb $a_j^{(j)} \neq 0$.

c) Vegeu, recolzant-vos en el resultat de b), que cap d'aquestes fórmules extrapolades coincidirà amb fórmules de Newton-Cotes per a $j \geq 4$.

SOLUCIÓ:

a) Per tal de deduir aquestes fórmules, considerem per més simplicitat només les absisses $x_k = a + kh$ ($k = 0 \div 4$) amb $h = \frac{b-a}{4}$, atès que són suficients per poder dur a terme els tres primers passos del mètode de Richardson amb $p_j = 2j$ ($j \geq 1$) i $q = 2$.

Si agafem $T_1(h_0)$ i $T_1(\frac{h_0}{2})$ i fem la primera extrapolació (del tipus $\frac{\Delta}{3}$), trobem

$$\begin{aligned} T_2\left(\frac{h_0}{2}\right) &= T_1\left(\frac{h_0}{2}\right) + \frac{T_1\left(\frac{h_0}{2}\right) - T_1(h_0)}{3} \\ &= \frac{h_0}{4}(f_0 + 2f_2 + f_4) + \frac{\frac{h_0}{4}(f_0 + 2f_2 + f_4) - \frac{h_0}{2}(f_0 + f_4)}{3} \\ &= \frac{h_0}{6}(f_0 + 4f_2 + f_4), \end{aligned}$$

que és la fórmula de Simpson amb pas $h = \frac{h_0}{2}$. Se'n dedueix, doncs, que $T_2(h)$ és la regla de Simpson $S(h)$.

Si considerem ara $T_2(\frac{h_0}{2}) = S(\frac{h_0}{2})$ i $T_2(\frac{h_0}{4}) = S(\frac{h_0}{4})$ i fem la subsegüent extrapolació (del tipus $\frac{\Delta}{15}$) tindrem

$$\begin{aligned} T_3\left(\frac{h_0}{4}\right) &= T_2\left(\frac{h_0}{4}\right) + \frac{T_2\left(\frac{h_0}{4}\right) - T_2\left(\frac{h_0}{2}\right)}{15} \\ &= \frac{h_0}{12}(f_0 + 4f_1 + 2f_2 + 4f_3 + f_4) \\ &\quad + \frac{\frac{h_0}{12}(f_0 + 4f_1 + 2f_2 + 4f_3 + f_4) - \frac{h_0}{6}(f_0 + 4f_2 + f_4)}{15} \\ &= \frac{h_0}{90}(7f_0 + 32f_1 + 12f_2 + 32f_3 + 7f_4), \end{aligned}$$

que és la fórmula de Newton-Cotes per a $n = 4$.

b) El procés d'extrapolació modifica els coeficients del desenvolupament asimptòtic de l'error segons la recurrència

$$a_l^{(j+1)} = \frac{4^j - 4^l}{4^j - 1} a_l^{(j)} \quad (l \geq j),$$

així s'aconsegueix l'objectiu del mètode de fer $a_j^{(j+1)} = 0$ i que els coeficients $a_l^{(j+1)}$ ($l > j$) de l'error asimptòtic de T_{j+1} siguin no nuls, si ho són els coeficients corresponents de T_j . Com que els coeficients $a_l^{(1)}$ ($l \geq 1$) són suposats no nuls, deduïm que els errors asimptòtics són com s'indica en l'enunciat.

c) Per a $j \geq 1$, la fórmula trobada per a T_j utilitza $2^{j-1} + 1$ abscisses i té un error asimptòtic que comença amb termes de l'ordre $\mu_j = 2j$ en h . Aquesta fórmula s'hauria de comparar amb la de Newton-Cotes del mateix nombre d'abscisses $m + 1 = 2^{j-1} + 1$. Si $j = 1$, l'expressió de l'error de la fórmula de Newton-Cotes de 2 punts és de l'ordre de $\nu_1 = 2$ en h . Si $j \geq 2$, el nombre d'abscisses $2^{j-1} + 1$ és senar i l'ordre de l'error de la fórmula de Newton-Cotes corresponent és $\nu_j = 2^{j-1} + 2$. La coincidència d'aquests valors es dona per a $j = 1, 2, 3$:

$$\mu_1 = \nu_1 = 2, \quad \mu_2 = \nu_2 = 4, \quad \mu_3 = \nu_3 = 6;$$

però no es dona per a $j > 3$. Això demostra que les fórmules extrapolades no poden coincidir amb les de Newton-Cotes quan $j > 3$.

Problema 4.14 a) Deduïu la fórmula de Gauss cap al davant d'interpolació en les abscisses equidistants $\{x_0, x_1, x_{-1}, \dots, x_s, x_{-s}\}$:

$$\begin{aligned} f(x_0 + th) = & f_0 + \sum_{r=1}^s \left[\binom{t+r-1}{2r-1} \delta^{2r-1} f_{\frac{1}{2}} + \binom{t+r-1}{2r} \delta^{2r} f_0 \right] \\ & + h^{2s+1} f^{(2s+1)}(\xi(t)) \binom{t+s}{2s} \end{aligned}$$

i dedueix la fórmula de Gauss cap al darrera d'interpolació en les abscisses equidistants $\{x_0, x_{-1}, x_1, \dots, x_{-s}, x_s\}$:

$$\begin{aligned} f(x_0 + th) = & f_0 + \sum_{r=1}^s \left[\binom{t+r-1}{2r-1} \delta^{2r-1} f_{-\frac{1}{2}} + \binom{t+r}{2r} \delta^{2r} f_0 \right] \\ & + h^{2s+1} f^{(2s+1)}(\xi(t)) \binom{t+s}{2s}. \end{aligned}$$

b) Deduïu, fent la mitjana de les fórmules de Gauss cap al davant i cap al darrera, la fórmula d'interpolació de Stirling

$$\begin{aligned} f(x_0 + th) = & f_0 + \sum_{r=1}^s \left[\binom{t+r-1}{2r-1} \mu \delta^{2r-1} f_0 + \frac{t}{2r} \binom{t+r-1}{2r-1} \delta^{2r} f_0 \right] \\ & + h^{2s+1} f^{(2s+1)}(\xi(t)) \binom{t+s}{2s}. \end{aligned}$$

c) Aplicació: Feu servir la fórmula d'interpolació de Stirling per tal de trobar $\sin 1.05$, amb 9 xifres decimals correctes, a partir de la taula de $\sin(1+x)$ en les abscisses $x_k = \frac{k}{10}$ ($k = -10, 10$).

SOLUCIÓ:

a) En les abscisses indicades, la fórmula de les diferències dividides de Newton pren la forma següents:

$$\begin{aligned} f(x_0 + th) &= f_0 + \sum_{r=1}^s \left[f[x_0, \dots, x_{r-1}, x_{-(r-1)}, x_r] \cdot \right. \\ &\quad (x - x_0) \cdots (x - x_{r-1})(x - x_{-(r-1)}) \\ &\quad + f[x_0, \dots, x_{r-1}, x_{-(r-1)}, x_r, x_{-r}] \cdot \\ &\quad \left. (x - x_0) \cdots (x - x_{r-1})(x - x_{-(r-1)})(x - x_r) \right] \\ &\quad + \frac{f^{(2s+1)}(\xi(t))}{(2s+1)!} (x - x_0) \cdots (x - x_s)(x - x_{-s}) . \end{aligned}$$

Atesa llur simetria, les diferències dividides es poden escriure, fent servir els operadors Δ i δ , en la forma:

$$\begin{aligned} f[x_0, \dots, x_{r-1}, x_{-(r-1)}, x_r] &= f[x_{-(r-1)}, \dots, x_{r-1}, x_r] \\ &= \frac{\Delta^{2r-1} f_{-(r-1)}}{(2r-1)! h^{2r-1}} = \frac{\delta^{2r-1} f_{\frac{1}{2}}}{(2r-1)! h^{2r-1}} , \\ f[x_0, \dots, x_{r-1}, x_{-(r-1)}, x_r, x_{-r}] &= f[x_{-r}, x_{-(r-1)}, \dots, x_{r-1}, x_r] \\ &= \frac{\Delta^{2r} f_{-r}}{(2r)! h^{2r}} = \frac{\delta^{2r} f_0}{(2r)! h^{2r}} . \end{aligned}$$

Els polinomis en x que hi apareixen s'escriuen com a polinomis en la variable $t = \frac{x-x_0}{h}$:

$$\begin{aligned} (x - x_0) \cdots (x - x_{r-1})(x - x_{-(r-1)}) &= h^{2r-1} (t + r - 1) \cdots (t - r + 1) \\ &= h^{2r-1} (2r - 1)! \binom{t + r - 1}{2r - 1} , \\ (x - x_0) \cdots (x - x_{r-1})(x - x_{-(r-1)})(x - x_r) &= h^{2r} (t + r - 1) \cdots (t - r + 1)(t - r) \\ &= h^{2r} (2r)! \binom{t + r - 1}{2r} ; \end{aligned}$$

$$\begin{aligned} (x - x_0) \cdots (x - x_s)(x - x_{-s}) &= h^{2s+1} (t + s) \cdots (t - s) \\ &= h^{2s+1} (2s + 1)! \binom{t + s}{2s + 1} . \end{aligned}$$

Substituint aquestes expressions en la primera expressió trobada, obtenim la fórmula de Gauss cap al davant de l'enunciat.

De manera anàloga, es dedueix la fórmula de Gauss cap al darrera.

b) En fer la mitjana d'ambdues fórmules apareixen els càlculs següents per al primer i segon termes del sumatori de la mateixa:

$$\begin{aligned} \frac{1}{2}(\delta^{2r-1}f_{\frac{1}{2}} + \delta^{2r-1}f_{-\frac{1}{2}}) &= \mu\delta^{2r-1}f_0, \\ \frac{1}{2} \left[\binom{t+r-1}{2r} + \binom{t+r}{2r} \right] &= \frac{1}{2(2r)!} \cdot \\ &\quad [(t+r-1)\cdots(t-r+1)(t-r) + (t+r)(t+r-1)\cdots(t-r+1)] \\ &= \frac{t}{2r} \binom{t+r-1}{2r-1}. \end{aligned}$$

Així, en resulta la fórmula d'interpolació de Stirling de l'enunciat, que fem explícita seguidament:

$$\begin{aligned} f(x_0 + th) &= f_0 + \mu\delta f_0 t + \frac{1}{2}\delta^2 f_0 t^2 \\ &\quad + \frac{1}{3!}\mu\delta^3 f_0(t+1)t(t-1) + \frac{1}{4!}\delta^4 f_0(t+1)t^2(t-1) \\ &\quad + \frac{1}{5!}\mu\delta^5 f_0(t+2)(t+1)t(t-1)(t-2) \\ &\quad + \frac{1}{6!}\delta^6 f_0(t+2)(t+1)t^2(t-1)(t-2) + \cdots \\ &\quad + \frac{f^{(2s+1)}(\xi(t))}{(2s+1)!} h^{2s+1}(t+s)\cdots(t-s). \end{aligned}$$

c) Pel que fa a l'aplicació demanada, notem que l'error permès s'assoleix per a $s = 3$. Fitem, per això, el terme d'error en aquest cas en què $t = \frac{1}{2}$ i $h = 0.1$:

$$\begin{aligned} &\left| \frac{f^{(7)}(\xi)}{7!} 10^{-7} \left(\frac{1}{2} + 3\right) \cdots \left(\frac{1}{2} - 3\right) \right| \\ &< \frac{1}{5040} 10^{-7} \frac{7}{2} \frac{5}{2} \frac{3}{2} \frac{1}{2} \frac{1}{2} \frac{3}{2} \frac{5}{2} M_7 = \frac{5}{2048} 10^{-7} M_7, \end{aligned}$$

on M_7 és una fita de la derivada setena de $f(x) = \sin(1+x)$ en $[-0.3, 0.3]$ que podem fitar per 1; la fita de l'error resultant és menor que $\frac{1}{2}10^{-9}$.

En resum, per a $s = 3$ i treballant amb valors de la funció donats amb més de 10 decimals, hauríem de trobar el valor demanat amb 9 decimals correctes.

La fórmula que hem d'aplicar, per a $t = \frac{1}{2}$, és

$$f(0.05) = f_0 + \frac{1}{2}\mu\delta f_0 + \frac{1}{8}\delta^2 f_0$$

$$\begin{aligned}
& -\frac{1}{16}\mu\delta^3 f_0 - \frac{1}{128}\delta^4 f_0 \quad (*) \\
& + \frac{3}{256}\mu\delta^5 f_0 + \frac{1}{1024}\delta^6 f_0 \\
& \pm \frac{1}{2}10^{-9} .
\end{aligned}$$

A fi de calcular còmodament els termes de la forma $\mu\delta^{(2r-1)}f_0$ i $\delta^{2r}f_0$ ($r = 1 \div 3$), cal adonar-se que els primers són les mitjanes dels termes

$$\delta^{(2r-1)}f_{\frac{1}{2}} = \Delta^{(2r-1)}f_{-(r-1)} , \quad \delta^{(2r-1)}f_{-\frac{1}{2}} = \Delta^{(2r-1)}f_{-r}$$

de la taula de diferències finites, i que els segons són els termes $\Delta^{2r}f_{-r}$ de la taula de diferències finites.

Això és, la taula de diferències finites que hem d'aplicar es pot escriure també com

x_{-3}	f_{-3}						
		$\delta f_{-\frac{5}{2}}$					
x_{-2}	f_{-2}		$\delta^2 f_{-2}$				
		$\delta f_{-\frac{3}{2}}$		$\delta^3 f_{-\frac{3}{2}}$			
x_{-1}	f_{-1}		$\delta^2 f_{-1}$		$\delta^4 f_{-1}$		
		$\delta f_{-\frac{1}{2}}$		$\delta^3 f_{-\frac{1}{2}}$		$\delta^5 f_{-\frac{1}{2}}$	
x_0	f_0		$\delta^2 f_0$		$\delta^4 f_0$		$\delta^6 f_0$
		$\delta f_{\frac{1}{2}}$		$\delta^3 f_{\frac{1}{2}}$		$\delta^5 f_{\frac{1}{2}}$	
x_1	f_1		$\delta^2 f_1$		$\delta^4 f_1$		
		$\delta f_{\frac{3}{2}}$		$\delta^3 f_{\frac{3}{2}}$			
x_2	f_2		$\delta^2 f_2$				
		$\delta f_{\frac{5}{2}}$					
x_3	f_3						

Escrivim, doncs, aquesta taula de diferències finites per a la funció $f(x) = \sin(1+x)$ en les abscisses $x_k = \frac{k}{10}$ ($k = -3 \div 3$):

0.7	0.64421768724			
		0.07313840366		
0.8	0.71735609090		-0.00716758493	
		0.06597081873		-0.00065915861
0.9	0.78332690963		-0.00782674355	
		0.05814407518		-0.00058095638
1.0	0.84147098481		-0.00840769993	
		0.04973637525		-0.00049694942
1.1	0.89120736006		-0.00890464935	
		0.04083172591		-0.00040797711
1.2	0.93203908597		-0.00931262646	
		0.03151909945		
1.3	0.96355818542			
		0.00007820223		
			0.00000580472	
		0.00008400696		-0.0000083937
			0.00000496535	
		0.00008897231		

De l'aplicació de la fórmula d'interpolació d'Stirling trobada (*), en resulta l'aproximació

$$\begin{aligned}
 \sin 1.05 &= 0.84147098481 \\
 &+ \frac{1}{4}(0.04973637525 + 0.05814407518) \\
 &+ \frac{1}{8}(-0.00840769993) \\
 &- \frac{1}{32}(-0.00049694942 - 0.00058095638) \\
 &- \frac{1}{128}0.00008400696 \\
 &+ \frac{3}{512}(0.00000496535 + 0.00000580472) \\
 &+ \frac{1}{1024}(-0.00000083937) \pm \frac{1}{2}10^{-9} \\
 &= 0.86742322546 \pm \frac{1}{2}10^{-9},
 \end{aligned}$$

resultat correcte amb la precisió demanada, comparant amb el resultat exacte $\sin 1.05 = 0.86742322559\dots$

Problema 4.15 Considerem la fórmula de derivació per al càlcul de derivades segones de f en x_0

$$f_0^{(2)} \simeq \frac{1}{h^2}(\delta^2 - \frac{1}{12}\delta^4)f_0.$$

a) Expressen aquesta fórmula en funció dels valors de la taula de f , $\{(x_k = x_0 + kh, f_k = f(x_k))\}_{k=-2 \div 2}$.

b) Demostreu que aquesta fórmula es troba per extrapolació de la fórmula

$$f_0^{(2)} \simeq \frac{1}{h^2}\delta^2 f_0$$

i trobeu l'error de truncament d'aquella, si $f \in \mathcal{C}^6([x_0 - 2h, x_0 + 2h])$.

c) Si f s'avalua amb un error fitat per ϵ , quin seria el valor del pas h^* que hem d'emprar a fi que sigui mínima la suma de les fites d'error degudes a la discretització i a l'avaluació de la fórmula?

d) Aplicació: Trobeu la derivada segona de la funció $f(x) = \sin(1+x)$ en $x_0 = 0$ amb els passos $h = 10^{-3}, 10^{-2}, 10^{-1}, 1$ i compareu amb el pas h^* , suposant que la funció es calcula amb 6 xifres decimals correctes.

SOLUCIÓ:

a) Tenim

$$\delta^2 f_0 = f_1 - 2f_0 + f_{-1}, \quad \delta^4 f_0 = f_2 - 4f_1 + 6f_0 - 4f_{-1} + f_{-2};$$

la fórmula donada s'expressa, per tant, així:

$$\begin{aligned} f_0^{(2)} &\simeq \frac{1}{h^2} \left(f_1 - 2f_0 + f_{-1} - \frac{f_2 - 4f_1 + 6f_0 - 4f_{-1} + f_{-2}}{12} \right) \\ &= \frac{1}{12h^2} (-f_2 + 16f_1 - 30f_0 + 16f_{-1} - f_{-2}) . \end{aligned}$$

b) Anomenant $D_1^2 f_0(h) = \frac{1}{h^2} \delta^2 f_0$, es té

$$D_1^2 f_0(h) = \frac{1}{h^2} (f_1 - 2f_0 + f_{-1}) .$$

A partir dels desenvolupaments de Taylor fins a ordre 5, es pot trobar una expressió per a l'error d'aquesta aproximació de $f_0^{(2)}$:

$$\begin{aligned} D_1^2 f_0(h) &= \frac{1}{h^2} \cdot \\ &\left[f_0 + hf'_0 + \frac{h^2}{2} f_0^{(2)} + \frac{h^3}{3!} f_0^{(3)} + \frac{h^4}{4!} f_0^{(4)} + \frac{h^5}{5!} f_0^{(5)} + \frac{h^6}{6!} f^{(6)}(\xi_+) \right. \\ &\quad \left. - 2f_0 \right. \\ &\quad \left. + f_0 - hf'_0 + \frac{h^2}{2} f_0^{(2)} - \frac{h^3}{3!} f_0^{(3)} + \frac{h^4}{4!} f_0^{(4)} - \frac{h^5}{5!} f_0^{(5)} + \frac{h^6}{6!} f^{(6)}(\xi_-) \right] \\ &= f_0^{(2)} + \frac{h^2}{12} f_0^{(4)} + \frac{h^4}{360} f^{(6)}(\xi) , \end{aligned}$$

amb $\xi_+, \xi_-, \xi \in (x_0 - h, x_0 + h)$, on s'ha aplicat el teorema del valor mitjà per a sumes, atesa la continuïtat de $f^{(6)}$.

Suposant que $f_0^{(4)} \neq 0$ i considerant la mateixa fórmula amb pas $2h$

$$D_1^2 f_0(2h) = \frac{1}{4h^2} (f_2 - 2f_0 + f_{-2}) ,$$

l'extrapolació adequada és la del tipus $\frac{\Delta}{3}$, que dona com a fórmula extrapolada

$$\begin{aligned} D_2^2 f_0(h) &= D_1^2 f_0(h) + \frac{1}{3} (D_1^2 f_0(h) - D_1^2 f_0(2h)) \\ &= \frac{1}{h^2} \left[f_1 - 2f_0 + f_{-1} + \frac{1}{3} (f_1 - 2f_0 + f_{-1} - \frac{f_2 - 2f_0 + f_{-2}}{4}) \right] \\ &= \frac{1}{12h^2} (-f_2 + 16f_1 - 30f_0 + 16f_{-1} - f_{-2}) = \frac{1}{h^2} \left(\delta^2 - \frac{1}{12} \delta^4 \right) f_0 . \end{aligned}$$

L'error d'aquesta fórmula extrapolada es troba fent servir la fórmula d'error trobada abans. Es té, per tant,

$$\begin{aligned} D_2^2 f_0(h) - f_0^{(2)} &= \frac{h^4}{360} f^{(6)}(\xi) + \frac{1}{3} \left[\frac{h^4}{360} f^{(6)}(\xi) - \frac{16h^4}{360} f^{(6)}(\eta) \right] \\ &= -\frac{h^4}{90} f^{(6)}(\zeta) , \end{aligned}$$

amb $\xi, \eta, \zeta \in (x_0 - 2h, x_0 + 2h)$, on s'ha aplicat de nou el teorema del valor mitjà per a sumes.

c) Notant per e_A i e_T els errors absoluts d'avaluació i discretització, en trobem les fites següents:

$$\begin{aligned} |e_A| &\leq \frac{1 + 16 + 30 + 16 + 1}{12h^2} \epsilon = \frac{16\epsilon}{3h^2}, \\ |e_T| &\leq \frac{h^4}{90} M_6, \quad \text{on } M_6 = \max_{x \in [x_0 - 2h, x_0 + 2h]} |f^{(6)}(x)|. \end{aligned}$$

La suma de les fites d'error trobades es la funció d' h següent:

$$\epsilon(h) = \frac{16\epsilon}{3h^2} + \frac{h^4}{90} M_6;$$

aquesta funció és mínima per al pas cercat $h = h^*$ que anul·la la seva derivada, $\epsilon'(h^*) = 0$; això és,

$$\frac{32\epsilon}{3h^{*3}} = \frac{4h^{*3}}{90} M_6, \quad h^* = \sqrt[6]{\frac{240\epsilon}{M_6}}.$$

d) Calculem les taules de diferències finites de la funció $f(x) = \sin(1 + x)$ amb els passos donats, treballant amb 6 xifres decimals. Els termes $\delta^2 f_0$ i $\delta^4 f_0$ amb els quals es troben les aproximacions de la derivada segona corresponents hi apareixen subratllats:

-0.002	0.840389			
		0.000541		
-0.001	0.840930		-0.000000	
		0.000541		0.000000
0.000	0.841471		<u>-0.000001</u>	<u>0.000000</u>
		0.000540		0.000000
0.001	0.842011		-0.000001	
		0.000539		
0.002	0.842550			

$$f_0^2 \simeq D_2^2 f_0(10^{-3}) = 10^6(-0.000001) = -1.$$

-0.02	0.830497			
		0.005529		
-0.01	0.836026		-0.000085	
		0.005445		-0.000001
0.00	0.841471		<u>-0.000084</u>	<u>0.000000</u>
		0.005361		-0.000001
0.01	0.846832		-0.000085	
		0.005276		
0.02	0.852108			

$$f_0^2 \simeq D_2^2 f_0(10^{-2}) = 10^4(-0.000084) = -0.84 .$$

-0.2	0.717356			
		0.065971		
-0.1	0.783327		-0.007827	
		0.058144		-0.000581
0.0	0.841471		<u>-0.008408</u>	<u>0.000085</u>
		0.049736		-0.000496
0.1	0.891297		-0.008904	
		0.040832		
0.2	0.932939			

$$f_0^2 \simeq D_2^2 f_0(10^{-1}) = 10^2(-0.008408 - \frac{1}{12}0.000085) = -0.841508 .$$

-2	-0.841471			
		0.841471		
1	0.000000		0.000000	
		0.841471		-0.773645
0	0.841471		<u>-0.773645</u>	<u>0.711285</u>
		0.067826		-0.062359
1	0.909297		-0.836004	
		-0.768177		
2	0.141120			

$$f_0^2 \simeq D_2^2 f_0(1) = -0.773645 - \frac{1}{12}0.711285 = -0.832919 .$$

Restant ara el valor exacte de la derivada $-\sin 1 = -0.841471$., trobem els errors:

$$D_2^2 f_0(h) - f_0^2 \simeq -0.16, 0.001471, -0.000037, 0.008552,$$

per a $h = 10^{-3}, 10^{-2}, 10^{-1}, 1$, respectivament.

Si ara fem el mateix càlcul amb el valor

$$h^* = \sqrt[6]{\frac{240\epsilon}{M_6}} \simeq 0.23 ,$$

on hem pres $\epsilon = \frac{1}{2}10^{-6}$ i hem fitat la derivada sisena per $M_6 = 1$; tenim

-0.46	0.514136			
		0.181999		
-0.23	0.696135		-0.036663	
		0.145336		-0.007655
0.0	0.841471		<u>-0.044318</u>	<u>0.002334</u>
		0.101018		-0.005321
0.23	0.942489		-0.049639	
		0.051379		
0.46	0.993868			

$$f_0^2 \simeq D_2^2 f_0(0.23) = \frac{1}{0.23^2}(-0.044318 - \frac{1}{12}0.002334) = -0.841446 .$$

Aquesta aproximació dóna un error de 0.000025, menor que els anteriors, tal com s'esperava.

Problema 4.16 a) Deduïu la fórmula d'Euler-Maclaurin mitjançant el càlcul formal amb operadors.

b) Expliciteu la dita fórmula fins als termes d'ordre 14 en el pas h .

SOLUCIÓ:

a) Fixat h , definim l'operador T que, aplicat a una funció f en x_0 , dóna l'aproximació de la regla dels trapezis amb pas h

$$Tf_0 = h(\frac{1}{2}f_0 + f_1 + f_2 + \dots + f_{M-1} + \frac{1}{2}f_M)$$

a la integral definida de f entre x_0 i $x_0 + Mh$.

Volem trobar una relació entre l'operador T i l'operador que dóna la integral definida que, aplicant la regla de Barrow i usant notació multiplicativa per als operadors, és

$$\frac{E^M - 1}{D} \equiv (E^M - 1)D^{-1} .$$

L'operador T es pot escriure també en funció de l'operador E

$$T = h(\frac{1}{2} + E + E^2 + \dots + E^{M-1} + \frac{1}{2}E^M) .$$

Operant en aquesta relació, fent servir la fórmula de Taylor usant operadors $E = e^{hD}$, trobem la seva relació amb l'operador integral definida i potències de l'operador D :

$$\begin{aligned} T &= h\left(-\frac{1}{2} + \frac{E^M - 1}{E - 1} + \frac{1}{2}E^M\right) = h(E^M - 1)\left(\frac{1}{E - 1} + \frac{1}{2}\right) \\ &= h(E^M - 1)\frac{E + 1}{2(E - 1)} = h(E^M - 1)\frac{hD}{2}\frac{E + 1}{E - 1}\frac{1}{hD} \\ &= h(E^M - 1)\left[1 + \sum_{j=1}^{\infty} c_j(hD)^j\right]\frac{1}{hD} \\ &= \frac{E^M - 1}{D} + h\sum_{j=1}^{\infty} c_j(E^M - 1)(hD)^{j-1} , \end{aligned}$$

on c_j ($j \geq 1$) són els coeficients del desenvolupament de Taylor de la funció

$$G(t) = \frac{t e^t + 1}{2 e^t - 1} = \frac{t}{2} \frac{e^{\frac{t}{2}} + e^{-\frac{t}{2}}}{e^{\frac{t}{2}} - e^{-\frac{t}{2}}} = \frac{t}{e^t - 1} + \frac{t}{2} = 1 + \sum_{j=1}^{\infty} c_j t^j .$$

Així, tenim la fórmula d'Euler-Maclaurin que relaciona la regla dels trapezis amb la integral definida i dóna uns termes correctius en funció de diferències entre les derivades imparells de la funció en els extrems de l'interval d'integració:

$$\begin{aligned} T(h) \equiv T f_0 &= \left[\frac{E^M - 1}{D} + h \sum_{j=1}^{\infty} c_j (E^M - I)(hD)^{j-1} \right] f_0 \\ &= \int_{x_0}^{x_0+Mh} f(x) dx \\ &\quad + \sum_{r=1}^{\infty} h^{2r} c_{2r} \left[f^{(2r-1)}(x_0 + Mh) - f^{(2r-1)}(x_0) \right] , \end{aligned}$$

on s'ha fet ús de la paritat de la funció G : $G(-t) = -G(t)$ i, per tant, $c_{2r-1} = 0$ ($r \geq 1$).

b) Els nombres de Bernoulli es poden definir pel desenvolupament

$$\frac{t}{e^t - 1} = \sum_{j=0}^{\infty} \frac{B_j}{j!} t^j ,$$

i satisfan la recurrència

$$(j+1)B_j = -\frac{1}{j+1} \sum_{i=0}^{\infty} \binom{j+1}{i} B_i ;$$

aquesta s'escriu també $(B+1)_{j+1} = B_{j+1}$ (vegeu el problema III.11).

Els coeficients c_j ($j \geq 1$) estan relacionats amb els nombres de Bernoulli, segons

$$c_0 = B_0 = 1 , \quad c_1 = B_1 + \frac{1}{2} = 0 , \quad c_{2r+1} = 0 , \quad c_{2r} = \frac{B_{2r}}{(2r)!} .$$

La relació de recurrència anterior, dóna els valors següents:

$$\begin{aligned} B_j &= 1, -\frac{1}{2}, \frac{1}{6}, 0, -\frac{1}{30}, 0, \frac{1}{42}, 0, -\frac{1}{30}, 0, \frac{5}{66}, \\ &\quad 0, -\frac{691}{2730}, 0, \frac{7}{6}, \dots \quad (j \geq 0) ; \end{aligned}$$

d'on

$$\begin{aligned} c_{2r} = \frac{B_{2r}}{(2r)!} &= \frac{1}{12}, -\frac{1}{720}, \frac{1}{30240}, -\frac{1}{1209600}, \frac{1}{47900160}, \\ &\quad -\frac{691}{2730 \cdot 12!}, \frac{7}{6 \cdot 14!}, \dots \quad (r \geq 1) . \end{aligned}$$

La fórmula d'Euler-Maclaurin, expressada fins al terme demanat, és

$$\begin{aligned}
 T(h) = & \int_{x_0}^{x_0+Mh} f(x)dx \\
 & + \frac{h^2}{12}(f'(x_0+Mh) - f'(x_0)) \\
 & - \frac{h^4}{720}(f^{(3)}(x_0+Mh) - f^{(3)}(x_0)) \\
 & + \frac{h^6}{30240}(f^{(5)}(x_0+Mh) - f^{(5)}(x_0)) \\
 & - \frac{h^8}{1209600}(f^{(7)}(x_0+Mh) - f^{(7)}(x_0)) \\
 & + \frac{h^{10}}{47900160}(f^{(9)}(x_0+Mh) - f^{(9)}(x_0)) \\
 & - \frac{691h^{12}}{2730 \cdot 12!}(f^{(11)}(x_0+Mh) - f^{(11)}(x_0)) \\
 & + \frac{7h^{14}}{6 \cdot 14!}(f^{(13)}(x_0+Mh) - f^{(13)}(x_0)) + \dots
 \end{aligned}$$

PROBLEMES PROPOSATS

1. Sigui f una funció definida en un interval que contingui les abscisses a i b ; volem aproximar el valor de la seva derivada en l'abscissa a mitjançant una fórmula del tipus

$$f'(a) \simeq \sum_{k=0}^m a_k f(x_k)$$

on les abscisses x_k ($k = 0 \div m$) són equiespaiades en l'interval $[a, b]$.

Trobeu els coeficients a_k ($k = 0 \div m$) de manera que la fórmula sigui exacta per a les funcions següents $f_j(x) = x^j$ ($j = 0 \div m$), amb $[a, b] = [0, 1]$.

2. a) Determineu la fórmula de derivació numèrica

$$g'(0) \simeq w_{-1}g(-1) + w_0g(0) + w_1g(1)$$

de manera que sigui exacta per a tots els polinomis de grau menor o igual que 2.

b) És exacta també per als polinomis de grau menor o igual que 3?

c) Usant els apartats anteriors, deduïu una fórmula de derivació de la forma

$$f'(c) \simeq a_{-1}f(c-h) + a_0f(c) + a_1f(c+h)$$

de manera que sigui exacta per a tots els polinomis de grau menor o igual que 3.

3. a) Escriviu de manera explícita una fórmula de derivació per al càlcul de $f'(a)$, deduïda per derivació del polinomi d'interpolació en les abscisses a , $a+h$, $a+2h$, $a+3h$ i $a+4h$.
- b) Doneu una expressió exacta per a l'error, si $f \in C^5([a, a+4h])$.
- c) Trobeu una expressió asimptòtica per a l'error, quan f és suficientment diferenciable.
4. Repetiu el problema anterior, canviant les abscisses d'interpolació per les noves abscisses $a-2h$, $a-h$, a , $a+h$ i $a+2h$.
5. a) Calculeu numèricament $f^{(2)}(0.6)$ per a $f(x) = \sin x$, emprant les fórmules:

$$f^{(2)}(a) \simeq \frac{1}{4h^2}[f(a+2h) - 2f(a) + f(a-2h)] ,$$

$$f^{(2)}(a) \simeq \frac{1}{h^2}[f(a+h) - 2f(a) + f(a-h)] ,$$

amb passos $h = 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005$ i fent servir els valors de la funció sinus amb 8 xifres decimals correctes.

b) Comenteu els errors comesos atenent tant a la fórmula emprada com als passos utilitzats.

c) Trobeu fites dels errors d'arrodoniment i dels errors de truncament de les fórmules de derivació usades, donant finalment el valor (òptim, en teoria) del pas h que minimitza la fita d'error total.

6. Determineu les abscisses de la fórmula d'integració numèrica

$$\int_0^1 f(x)dx \simeq 0.72f(x_0) + 0.28f(x_1)$$

de manera que sigui exacta per a tots els polinomis de grau menor o igual que 2.

7. Determineu els pesos i les abscisses de la fórmula d'integració

$$\int_{-1}^1 g(t)dt \simeq w_0g(t_0) + w_1g(t_1)$$

per tal que sigui exacta per a tots els polinomis del grau més alt possible.

8. Suposant que tots els pesos són iguals ($w_{-1} = w_0 = w_1 = w$), trobeu el valor w d'aquests i les abscisses de la fórmula d'integració numèrica

$$\int_{-1}^1 g(t)dt \simeq w_{-1}g(t_{-1}) + w_0f(t_0) + w_1(t_1) ,$$

per tal que sigui exacta per a tots els polinomis de grau més petit o igual que 3.

9. a) Determineu els pesos de la fórmula d'integració següent per tal que sigui exacta per a tots els polinomis de grau menor o igual que 3:

$$\int_a^b f(x)dx = A_0f(a) + A_1f(b) + C_0f^{(2)}(a) + C_1f^{(2)}(b) .$$

b) Demostreu que també és exacta per als polinomis de grau 4.

10. Determineu les incògnites W_0, W_1, W_2 i c de la fórmula d'integració numèrica següent (de tipus Lobatto):

$$\int_{-1}^1 g(t)dt \simeq W_0f(0) + W_1[g(-1) + g(1)] + W_2[g(-c) + g(c)]$$

per tal que tingui el grau de precisió més alt possible.

11. a) Trobeu totes les fórmules de Newton-Cotes (tancades) de $m+1$ abscisses $m = 1 \div 8$ per al càlcul de

$$\int_{x_0}^{x_m} f(x) dx .$$

b) Demostreu que les dites fórmules són exactes per a tots els polinomis de grau més petit o igual que m (si m és senar) i que $m+1$ (si m és parell).

c) Suposant que f és suficientment diferenciable, feu explícita la fórmula de l'error d'integració numèrica per a $m = 1 \div 8$, sabent que és de la forma

$$\frac{f^{(m+1)}(\xi)}{(m+1)!} \int_{x_0}^{x_m} \omega_m(x) dx \quad (\text{si } m \text{ és senar}) ,$$

$$\frac{f^{(m+2)}(\xi)}{(m+2)!} \int_{x_0}^{x_m} x \omega_m(x) dx \quad (\text{si } m \text{ és parell}) ,$$

on $\omega_m(x) = (x - x_0) \cdots (x - x_m)$ i $\xi \in (x_0, x_m)$.

12. a) Avalueu la integral

$$\int_{-1}^1 \frac{dx}{1+x^2} ,$$

emprant les fórmules de Newton-Cotes de $m+1$ abscisses ($m = 1 \div 8$) i adoneu-vos que els errors comesos van decreixent mentre m augmenta.

b) Feu ara els mateixos càlculs amb la integral

$$\int_{-5}^5 \frac{dx}{1+x^2} ;$$

què s'hi observa?, podríeu explicar-ho?

13. a) Trobeu les *fórmules de Newton-Cotes obertes* de $m-1$ abscisses (exactes per a tots els polinomis de grau menor o igual que $m-2$):

$$\int_{x_0}^{x_m} f(x) dx \simeq \sum_{k=1}^{m-1} A_k f(x_k) ,$$

on les abscisses x_k ($k = 0 \div m$) són equidistants. Feu-ho per a $m = 2 \div 8$.

b) Aplicació: Calculeu amb elles les integrals

$$\int_0^1 \tanh(sx) dx \quad (s = 1, 2, 4, 8) ,$$

on

$$\tanh z = \frac{e^z - e^{-z}}{e^z + e^{-z}} ,$$

i comenteu-ne els resultats.

14. a) Trobeu les fórmules d'integració interpolatòria en $m + 1$ abscisses equidistants ($m = 2, 3, 4, 5$) per a les integrals de la forma

$$\int_{-1}^1 w(x)f(x)dx ,$$

i les funcions pes següents:

i) $w(x) = x$,

ii) $w(x) = |x|$,

iii) $w(x) = |x|^{\frac{1}{2}}$,

iv) $w(x) = \ln(1 + x)$,

v) $w(x) = (1 - x^2)^{-\frac{1}{2}}$.

b) Trobeu expressions per als errors comesos quan f és prou diferenciable.

c) Aplicació: Calculeu amb elles les integrals

$$\int_{-1}^1 w(x)e^{-x^2}dx$$

i fiteu els errors comesos.

15. Considerem una funció f suficientment diferenciable.

a) Doneu, de forma explícita, les *fórmules d'integració numèrica de Taylor*

$$\int_a^b f(x)dx \simeq \sum_{j=0}^n C_j f^{(j)}(c) ,$$

trobades per integració numèrica del polinomi d'interpolació de Taylor de grau més petit o igual que n a f en una abscissa c de l'interval $[a, b]$.

b) Expliciteu la dita fórmula quan $c = a$ i quan $c = b$.

c) Doneu una expressió per a l'error en tots els casos.

16. Trobeu la integral

$$\int_0^1 \frac{\cos x - 1}{x^2} dx$$

per desenvolupament de Taylor de l'integrand amb un error menor que 10^{-10} .

17. Calculeu amb 5 xifres decimals correctes

$$\int_{10}^{\infty} (x^3 + x)^{-\frac{1}{2}} dx .$$

18. Calculeu amb 5 xifres decimals correctes

$$\int_0^{0.1} (1 - 0.1 \sin t)^{\frac{1}{2}} dt .$$

19. a) Partint del coneixement del desenvolupament

$$(1 - z)^{-1} = 1 + z + z^2 + z^3 + \cdots + z^n + \frac{z^{n+1}}{1 - z} \quad (|z| < 1) ,$$

trobeu el desenvolupament de Taylor en $x = 0$ d' $\arctan x$.

- b) Trobeu de forma anàloga el d' $\arcsin x$.

20. a) Trobeu de forma explícita la funció inversa de

$$\tanh y = \frac{e^y - e^{-y}}{e^y + e^{-y}}$$

i digueu on està definida. Aquesta funció inversa s'anomena *argument de la tangent hiperbòlica* i s'escriu $\arg \tanh$.

- b) Trobeu el desenvolupament de Taylor de

$$F(x) = \int_0^x \arg \tanh t \, dt$$

prop de $x_0 = 0$ i estudieu-ne la convergència.

- c) Digueu quants termes del desenvolupament calen per tal de calcular $F(0.1)$ i $F(0.9)$ amb un error menor que $\frac{1}{2}10^{-15}$.

21. Sigui

$$F(x) = \int_0^{\frac{\pi}{2}} \cos(x \sin t) dt .$$

- a) Trobeu el seu desenvolupament en sèrie de Taylor en $x_0 = 0$.

- b) Calculeu $F(2)$ amb un error menor que 10^{-6} .

- c) Si volguéssim calcular $F(20)$, quants termes caldrien a fi d'obtenir una precisió semblant?

- d) Amb almenys quantes xifres s'haurien de fer els càlculs en l'apartat c)?

22. La longitud d'una el·lipse de semieix major a i d'excentricitat e (entre 0 i 1) ve donada per

$$L = 4a \int_0^{\frac{\pi}{2}} \sqrt{1 - e^2 \sin^2 \varphi} d\varphi .$$

- a) Trobeu el desenvolupament de Taylor de L com a funció d' e .
 b) Si considerem L aproximada pel primer terme del desenvolupament, per a quins valors d' e l'error comès és menor que l' u per cent?
 c) Notem que $L = 4a$ per a $e = 1$; per a quins valors d' e menors que 1 podem aproximar L per $4a$ amb un error menor que l' u per cent?

Indicació: Useu

$$\int_0^{\frac{\pi}{2}} \sin^{2r} \varphi d\varphi = \frac{\pi}{2} \frac{(2r-1)!!}{(2r)!!} .$$

23. Volem trobar fórmules d'integració en $m+1$ abscisses equidistants de la forma

$$\int_a^b \frac{f(x)}{\sqrt{x}} dx \simeq \sum_{k=0}^m [A_k f(x_k) + B_k f'(x_k)] ,$$

exactes per a tots els polinomis de grau més petit o igual que $m+1$.

- a) Feu explícites aquestes fórmules per a $m = 1, 2, 3$, donant així mateix una expressió per als errors comesos.

- b) Aplicació: Calculeu

$$\int_0^1 \frac{e^x}{\sqrt{x}} dx ,$$

fitant l'error comès, emprant totes les fórmules trobades en a).

24. Calculeu

$$\int_0^\pi \cosh(3x) dx ,$$

- a) directament, donant-ne el resultat amb 6 xifres decimals correctes;
 b) emprant la regla dels trapezis amb un pas h adequat a fi que l'error sigui menor que 10^{-6} ;
 c) emprant la regla de Simpson amb les mateixes condicions que en b).

25. Calculeu

$$\int_0^\pi \frac{\cos x - 1}{x^2} dx ,$$

amb un error menor que 10^{-10} , emprant les regles dels trapezis i de Simpson.

26. a) Trobeu les regles (compostes) a partir de les fórmules (simples) d'integració numèrica de Newton-Cotes tancades de $m+1$ abscisses ($m = 2, 3, 4, 5$).
 b) Trobeu expressions per als errors d'aquestes regles, suposant f prou diferenciable.
 c) Repetiu l'apartat a) per a les fórmules de Newton-Cotes obertes.

27. a) Considerem les aproximacions donades per la regla dels trapezis $T(h)$ per a valors del pas h tendint a 0, al càlcul de la integral J d'una funció contínua f en un interval fitat $[a, b]$. Demostreu que

$$\lim_{h \rightarrow 0} T(h) = J .$$

- b) Generalitzeu aquest resultat a totes les fórmules de quadratura

$$Q_m(f) = \sum_{k=0}^m A_k f(x_k) , \text{ amb } \sum_{k=0}^m A_k = b - a .$$

- c) Deduïu de b) que el resultat és cert per a totes les regles compostes de Newton-Cotes tancades i obertes.

28. a) Trobeu una fórmula d'integració de dues abscisses per a integrals del tipus

$$\int_{-1}^1 (1+x^2)f(x)dx ,$$

que sigui exacta per a tots els polinomis de grau més petit o igual que 3.

- b) Doneu una fita de l'error comès en la seva aplicació al càlcul de

$$\int_{-1}^1 (1+x^2)x^4 dx .$$

29. Considerem la fórmula d'integració gaussiana

$$\int_a^b w(x)f(x)dx \simeq \sum_{k=0}^m W_k f(x_k) .$$

Demostreu que, si $\{\psi_j(x)\}_{j \geq 0}$ és una família de polinomis ortogonals respecte al producte escalar continu associat al pes w a $[a, b]$, llavors $\{\psi_j(x)\}_{j=0 \div m}$ és una família de polinomis ortogonals respecte al producte escalar discret

$$(f, g)_d = \sum_{k=0}^m W_k f(x_k) g(x_k) .$$

30. a) Calculeu exactament

$$\int_0^\pi \cos^3 \theta d\theta ,$$

emprant una fórmula de Gauss-Txebixev adequada.

- b) Calculeu

$$\int_0^\pi e^{\cos x} dx ,$$

mitjançant una fórmula de Gauss-Txebixev amb un nombre adequat d'abscisses per tal que l'error sigui més petit que 10^{-8} .

31. Calculeu les sumes finites següents i analitzeu llur comportament quan n tendeix a infinit:

$$\sum_{j=1}^n j, \quad \sum_{j=1}^n r^j, \quad \sum_{j=1}^n \ln \frac{j+1}{j}, \quad \sum_{j=1}^n \frac{1}{4j^2 + 4j + 1}.$$

32. a) Mitjançant l'ús de funcions factorial, calculeu i analitzeu el comportament de les sumes següents quan n tendeix a infinit:

$$\sum_{j=1}^n j^k \quad (k = 2, 3, 4, 5, 6, 7), \quad \sum_{j=1}^n (3j^2 + 3j^5 - 7j^7).$$

- b) Estudieu el cas general

$$\sum_{j=1}^n P_k(j),$$

essent $P_k(x)$ un polinomi de grau k .

33. a) Useu el mètode de comparació per al càlcul de les sumes de les sèries següents amb un error menor que 10^{-6} :

$$\sum_{j=1}^{\infty} \frac{1}{j^2 + k} \quad (k = 1, 2, 3), \quad \sum_{j=1}^{\infty} \frac{1}{j^2 + j + 1},$$

$$\sum_{j=1}^{\infty} \frac{1}{j^4 + 1}, \quad \sum_{j=1}^{\infty} \frac{1}{j^8 + 1}, \quad \sum_{j=1}^{\infty} \frac{j^2 + 3j + 2}{j^4 + 1}.$$

- b) Estudieu el cas general

$$\sum_{j=1}^{\infty} \frac{P_r(j)}{Q_s(j)} \quad (s \geq r + 2),$$

essent $P_r(x)$ i $Q_s(x)$ polinomis de graus r i s , respectivament.

34. a) Sigui

$$S = \sum_{j=1}^{\infty} a_j \quad (a_j > 0).$$

Definim

$$c_j = a_j + 2a_{2j} + 4a_{4j} + 8a_{8j} + \cdots.$$

Proveu que

$$S = \sum_{j=1}^{\infty} (-1)^{j+1} c_j, \quad a_j = c_j - 2c_{2j}.$$

b) Utilitzeu el resultat anterior per calcular

$$\sum_{j=1}^{\infty} \frac{1}{j^3}, \quad \sum_{j=1}^{\infty} \frac{1}{(2j-1)^2},$$

amb 6 xifres decimals correctes.

35. Usant la fórmula d'Euler-Maclaurin per a sumes, calculeu les sumes del problema 32.

36. a) Calculeu, amb un error fitat per una milionèsima,

$$\sum_{j=1}^{\infty} \frac{1}{j^{\frac{4}{3}}},$$

usant la fórmula d'Euler-Maclaurin per a sèries.

b) Quants termes s'haurien de sumar directament per assolir la mateixa precisió?

37. Calculeu, amb un error fitat per una milionèsima,

$$\sum_{j=123}^{369} \frac{1}{j^2 + 1},$$

usant la fórmula d'Euler-Maclaurin per a sumes.

38. La funció *zeta de Riemann*

$$\zeta(x) = \sum_{j=1}^{\infty} \frac{1}{j^x} \quad (x > 1)$$

es pot aproximar sumant prèviament els 9 primers termes i aplicant a la resta la fórmula d'Euler-Maclaurin per a sèries fins als termes que contenen derivades cinquenes.

a) Proveu que l'error de l'aproximació feta és menor que 10^{-10} per a qualsevol x .

b) Feu una taula de $\zeta(x)$ per a x entre 2 i 25 amb pas $h = 1$ i una altra per a x entre 1.01 i 2 amb pas $h = 0.01$.

c) Utilitzeu les taules trobades per repetir la suma de les sèries del problema 33.

39. Calculeu amb 6 xifres decimals correctes les sumes

$$\sum_{j=1}^{\infty} [\exp(\frac{1}{j^3}) - 1], \quad \sum_{j=1}^{\infty} \tan^2 \frac{1}{j}, \quad \sum_{j=1}^{\infty} \sin^2 \frac{1}{j},$$

emprant les taules de la funció zeta de Riemann del problema anterior.

40. Definim la funció *digamma* per

$$\begin{aligned}\Psi(1+x) &= \sum_{j=1}^{\infty} \frac{x}{j(j+x)} - \gamma \\ &= \sum_{j=1}^{\infty} \left(\frac{1}{j} - \frac{1}{j+x} \right) - \gamma \quad (x \neq -1, -2, \dots),\end{aligned}$$

on γ és la constant d'Euler, introduïda en el problema IV.10.

a) Comproveu la fórmula recurrent

$$\Psi(k+x) = \sum_{j=1}^{k-1} \frac{1}{j+x} + \Psi(1+x).$$

b) Tabuleu $\Psi(1+x)$, $\Psi'(1+x)$, $\Psi^{(2)}(1+x)$ i $\Psi^{(3)}(1+x)$, amb 7 xifres decimals correctes i pas $h = 0.01$ en l'interval $[0, 1]$, usant la fórmula d'Euler-Maclaurin per a sumes.

c) Tabuleu $\Psi(1+x)$, $\Psi'(1+x)$, $\Psi^{(2)}(1+x)$ i $\Psi^{(3)}(1+x)$, amb 7 xifres decimals correctes i pas $h = 0.1$ en l'interval $[-10, 10]$.

41. Useu les taules del problema anterior per sumar les sèries següents amb 6 xifres decimals correctes:

$$\begin{aligned}\sum_{j=1}^{\infty} \frac{1}{(2j+1)(j+1)^2}, \quad \sum_{j=1}^{\infty} \frac{1}{(4j+2)(4j+1)^2(4j+3)^3}, \\ \sum_{j=1}^{\infty} \frac{2j+5}{(2j+7)j^4}.\end{aligned}$$

42. Calculeu amb 6 xifres decimals correctes

$$P = \prod_{j=1}^{\infty} \left(1 + j^{-\frac{3}{2}}\right),$$

usant la fórmula d'Euler-Maclaurin per a sèries, aplicada a la sèrie $\ln P$.

43. Dibuixem un triangle equilàter circumscrit a una circumferència; dibuixem a continuació una circumferència que passi pels seus vèrtexs; dibuixem després el quadrat circumscrit a aquesta última circumferència, i així successivament, amb tots els polígons regulars. Estudieu el comportament asimptòtic del radi de les successives circumferències quan el nombre de costats tendeix a infinit.

44. a) Trobeu desenvolupaments asimptòtics associats a les funcions expressades per les integrals següents:

$$\int_x^\infty \frac{e^{-t}}{t} dt, \quad \int_x^\infty \frac{\sin t}{t} dt, \quad \int_x^\infty \frac{\cos t}{t} dt, \quad \int_x^\infty \sin t^2 dt.$$

$$\int_x^\infty \frac{e^{-t}}{t^\alpha} dt, \quad \int_x^\infty \frac{\sin t}{t^\alpha} dt, \quad \int_x^\infty \frac{\cos t}{t^\alpha} dt, \quad \int_x^\infty \sin t^\alpha dt, \quad (\alpha \geq 1).$$

$$\int_x^\infty \exp(-\frac{t^2}{2}) dt, \quad \int_0^\infty \frac{e^t}{1+xt} dt.$$

- b) Amb quina precisió es poden calcular en funció de x ?
 c) Aplicació: Calculeu-les en $x = 10, 100, 1000$.
45. Una funció F , relacionada amb les integrals de Fresnel usades en Òptica, té el desenvolupament asimptòtic següent:

$$F(x) = 1 + \sum_{j=1}^{\infty} (-1)^j \frac{(4j-1)!!}{(\pi x^2)^{2j}}.$$

Sabem que la magnitud de l'error en prendre n termes és menor que la del terme $n+1$.

- a) Quin és el valor òptim de n per al càlcul de $F(6)$? I per al càlcul de $F(3)$?
 b) Fiteu els errors comesos en ambdós casos.

46. Les funcions

$$E_k(x) = \int_1^\infty \frac{e^{-xt}}{t^k} dt$$

admeten les expressions següents

$$\text{i)} \quad \frac{e^{-x}}{x} \left(1 - \frac{k}{x} + \frac{k(k+1)}{x^2} - \frac{k(k+1)(k+2)}{x^3} + \dots \right)$$

$$\text{ii)} \quad \frac{(-x)^{k-1}}{(k-1)!} [-\ln x + \Psi(k)] - \sum_{j=0}^{\infty} \frac{(-x)^j}{(j-k+1)j!},$$

on Ψ és la funció digamma del problema 40 que compleix

$$\Psi(1) = -\gamma = -0.5772156649\dots, \quad \Psi(k) = -\gamma + \sum_{j=1}^{k-1} \frac{1}{j} \quad (k > 1).$$

Per a $n = 2, 6, 10$, indiqueu fins a quin valor de $x > 0$ és millor usar l'expressió i) que la ii).

47. Considerem la *funció gamma d'Euler*

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

que compleix

$$\Gamma(z+1) = z\Gamma(z)$$

i, per a valors enters positius de z ,

$$\Gamma(z) = (z-1)! .$$

La *fórmula de Stirling* dona el desenvolupament asimptòtic següent per al logaritme de la *funció gamma d'Euler*,

$$\ln \Gamma(x) = \left(x - \frac{1}{2}\right) \ln x - x + \frac{1}{2} \ln 2\pi + \sum_{r=1}^{\infty} \frac{B_{2r}}{2r(2r-1)x^{2r-1}} .$$

- Per a cada x , quin és el nombre de termes que cal sumar per tal de cometre menys errors?
- Calculeu $999999! = \Gamma(1000000)$ amb 5 xifres significatives correctes.
- Quin error mínim té la fórmula quan s'aplica al càlcul de $2! = \Gamma(3)$?

48. La funció error

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

admet els desenvolupaments següents:

$$\frac{2}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j+1}}{(2j+1)j!} , \quad \frac{2}{\sqrt{\pi}} \exp(-x^2) \sum_{j=0}^{\infty} \frac{2^j}{(2j+1)!!} x^{2j+1} ,$$

$$1 - \frac{1}{\sqrt{\pi}x} \exp(-x^2) \left(1 + \sum_{j=1}^{\infty} (-1)^j \frac{(2j-1)!!}{(2x^2)^j} \right) .$$

Comenteu quin d'ells és millor emprar per als diferents valors de x .

49. a) Calculeu aproximacions de la derivada de la funció $f(x) = \ln x$ en $x = 2$ emprant les fórmules

$$f'(a) \simeq \frac{1}{h}(f(a+h) - f(a)) ,$$

$$f'(a) \simeq \frac{1}{2h}(f(a+h) - f(a-h)) ,$$

amb passos $h = 0.1, 0.01, 0.001, 0.0001, 0.00001$.

- b) Usant l'expressió asimptòtica de l'error de truncament de les fórmules de derivació numèrica d'a), deduïu-ne expressions millors, quan el pas h tendeix a 0.
- c) Apliqueu les fórmules trobades per extrapolació a b) al càlcul de la derivada proposada en a) i comenteu els resultats.
50. a) Calculeu la derivada aproximada de la funció $f(x) = e^{x^2}$, en $x = 2$, mitjançant extrapolació repetida de Richardson, a partir de la taula dels valors de la funció en les abscisses $x_i = i/4$ ($i = 0 \div 16$), donats amb 8 xifres decimals correctes arrodonides.
- b) Estudieu els efectes dels errors d'arrodoniment sobre les diferents aproximacions trobades.
51. a) Trobeu una fórmula per al càlcul de $f^{(n)}(a)$ basada en la derivació del polinomi interpolador de grau més petit o igual que $n + 1$ en les abscisses $x_0 = a, x_1, x_2, \dots, x_{n+1}$.
- b) Expliciteu-la per a $n = 1, 2, 3, 4$ i x_i ($i = 0 \div n + 1$) equidistants amb pas h .
- c) Aplicació: Calculeu, amb les fórmules trobades a b), les 4 primeres derivades de la funció $f(x) = \cos(\cos(\sin x))$ en $x = a = \frac{\pi}{4}$, emprant els passos $h = 0.1, 0.01$.
- d) Extrapoleu els resultats obtinguts.

52. Considerem el càlcul de la integral

$$\int_0^1 \frac{\sin x}{x} dx ,$$

emprant la regla dels trapezis $T(h)$ amb passos h cada cop més petits.

- a) Calculeu $T(h)$ per a $h = 2^{-k}$ ($k = 0 \div 20$), treballant amb simple precisió (o, equivalentment, amb 6 xifres decimals i arrodoniment); a partir de quin valor de k es inútil continuar fent els càlculs perquè els errors d'arrodoniment són ja superiors als de truncament?
- b) Des d'un punt de vista teòric, suposem que f s'avalua amb un error relatiu fitat per ϵ . Trobeu la fita de l'error relatiu total, que s'obté sumant les fites dels errors de truncament i d'arrodoniment i el valor del pas (òptim, des d'aquest punt de vista) que minimitza aquesta fita de l'error.
53. a) Calculeu la integral

$$J = \int_1^2 \ln x \, dx ,$$

emprant la regla dels trapezis amb passos $h = 0.1, 0.01, 0.001$.

- b) Avalueu els errors comesos $E_T(h) = T(h) - J$, comparant amb el resultat obtingut directament.

c) Ajusteu els resultats d'a) a una expressió asimptòtica de l'error de la forma

$$E_T(h) = a_1 h^2 + a_2 h^4 + \dots$$

d) Extrapoleu els resultats anteriors i trobeu una millor aproximació del valor de J . Com seria l'expressió asimptòtica de l'error d'aquests valors extrapolats?

54. a) Calculeu

$$\int_0^{\frac{1}{2}} e^x dx$$

de les maneres següents:

i) cercant la funció primitiva,

ii) mitjançant la regla dels trapezis amb passos $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$,

iii) extrapolant els resultats obtinguts en b).

b) Trobeu els quocients entre els errors trobats i comenteu-ne el resultat.

55. Suposem que la funció f s'avalua amb un error absolut fitat per ϵ en l'interval $[a, b]$. Demostreu que l'error en l'avaluació de les aproximacions a la integral donades pel mètode de Romberg a la integral de f en $[a, b]$ és menor que $2(b-a)\epsilon$, independentment del pas usat.

56. a) Deduïu la *fórmula de Gauss cap al davant* d'interpolació equidistant en un nombre senar d'abscisses $\{x_0, x_1, x_{-1}, \dots, x_s, x_{-s}, x_{s+1}\}$

$$\begin{aligned} f(x_0 + th) = & \sum_{r=0}^s \left[\binom{t+r}{2r+1} \delta^{2r+1} f_{\frac{1}{2}} + \binom{t+r-1}{2r} \delta^{2r} f_0 \right] \\ & + h^{2s+2} f^{(2s+2)}(\xi(t)) \binom{t+s}{2s+1}. \end{aligned}$$

b) A partir de la fórmula anterior, deduïu la *fórmula d'Everett*

$$\begin{aligned} f(x_0 + th) = & \sum_{r=0}^s \left[\binom{t+r}{2r+1} \delta^{2r} f_1 - \binom{t+r-1}{2r+1} \delta^{2r} f_0 \right] \\ & + h^{2s+1} f^{(2s+1)}(\xi(t)) \binom{t+s}{2s+1}, \end{aligned}$$

i la *fórmula de Bessel*

$$\begin{aligned} f(x_0 + th) = & \sum_{r=0}^s \left[\binom{t+r-1}{2r} \mu \delta^{2r} f_{\frac{1}{2}} \right. \\ & \left. + \frac{1}{2r+1} \left(t - \frac{1}{2}\right) \binom{t+r-1}{2r} \delta^{2r+1} f_{\frac{1}{2}} \right] \\ & + h^{2s+1} f^{(2s+1)}(\xi(t)) \binom{t+s}{2s+1}. \end{aligned}$$

c) Aplicació: Feu servir les fórmules d'Everett i de Bessel per tal de trobar $\sin 1.05$, amb 9 xifres decimals correctes, a partir de la taula de $\sin(1+x)$ en les abscisses $x_k = \frac{k}{10}$ ($k = -10, 10$).

57. a) Deriveu les fórmules d'interpolació de Newton amb diferències cap al davant i cap al darrera a fi de trobar fórmules de derivació en l'abscissa x_0 i l'error comès corresponent.

b) Deriveu la fórmula de Stirling del problema IV.14 a fi de trobar una fórmula de derivació en l'abscissa central de la xarxa x_0 i l'error comès.

c) Aplicació: Calculeu amb les dites fórmules la derivada de la funció $f(x) = \sin(1+x)$ en $x = 0$, fent servir la taula esmentada en el problema anterior.

d) Repetiu els apartats anteriors, afegint-hi el càlcul de fórmules per a derivades segones i terceres.

58. a) Deduïu la *regla dels trapezis amb termes correctius fins a ordre 5*

$$\begin{aligned} \int_{x_0}^{x_M} f(x) dx &\simeq \frac{h}{2}(f_0 + 2f_1 + 2f_2 + \cdots + 2f_{M-1} + f_M) \\ &\quad - \frac{h}{12}(\mu\delta f_M - \mu\delta f_0) + \frac{11h}{720}(\mu\delta^3 f_M - \mu\delta^3 f_0) \\ &\quad - \frac{191h}{60480}(\mu\delta^5 f_M - \mu\delta^5 f_0), \end{aligned}$$

integrant sobre l'interval $[x_i, x_{i+1}]$ la fórmula de Bessel de grau 5 d'interpolació en les abscisses x_{i-k} ($k = -2 \div 3$), i efectuant la suma per a $i = 0 \div M-1$.

b) Deduïu la *regla de Simpson amb termes correctius fins a ordre 6*

$$\begin{aligned} \int_{x_0}^{x_{2M}} f(x) dx &\simeq \frac{h}{3}(f_0 + 4f_1 + 2f_2 + \cdots \\ &\quad + 2f_{2M-2} + 4f_{2M-1} + f_{2M}) \\ &\quad - \frac{h}{90}(\delta^4 f_1 + \delta^4 f_3 + \cdots + \delta^4 f_{2M-1}) \\ &\quad + \frac{h}{756}(\delta^6 f_1 + \delta^6 f_3 + \cdots + \delta^6 f_{2M-1}), \end{aligned}$$

integrant sobre l'interval $[x_{2r}, x_{2r+1}]$ la fórmula de Stirling de grau 6 d'interpolació en les abscisses x_{i-k} ($k = -3 \div 3$), i sumant per a $r = 0 \div 2M-1$.

c) Deduïu la *fórmula de Gregory amb termes correctius fins a ordre 6*

$$\begin{aligned} \int_{x_0}^{x_M} f(x) dx &\simeq \frac{h}{2}(f_0 + 2f_1 + 2f_2 + \cdots + 2f_{M-1} + f_M) \\ &\quad - \frac{h}{12}(\nabla f_M - \Delta f_0) - \frac{h}{24}(\nabla^2 f_M + \Delta^2 f_0) \\ &\quad - \frac{19h}{720}(\nabla^3 f_M - \Delta^3 f_0) - \frac{3h}{160}(\nabla^4 f_M + \Delta^4 f_0) \\ &\quad - \frac{863h}{60480}(\nabla^5 f_M - \Delta^5 f_0), \end{aligned}$$

a partir de la fórmula d'Euler-Maclaurin, substituint-hi les derivades successives en les abscisses x_0 i x_M per les derivades corresponents de les fórmules de Newton amb diferències cap endavant i cap endarrera, respectivament.

CAPÍTOL 5

EQUACIONS NO LINEALS

5.1 EQUACIONS EN UNA VARIABLE

5.1.1 Introducció

Anomenem *equació no lineal* una equació del tipus $f(x) = 0$ (a vegades l'expressarem per $x = g(x)$) en la qual f és una funció real de variable real no lineal. La funció f pot ésser polinòmica (cas que estudiarem en la segona part d'aquest capítol), transcendent (apareixen en la seva expressió funcions exponencials, logarítmiques i trigonomètriques) i, fins i tot, pot ocórrer que no es disposi d'una expressió explícita de $f(x)$, però que es coneguin les regles per al seu càlcul. Un exemple d'aquest cas, freqüent en matemàtica aplicada, es dona quan $f(x)$ és el valor de la solució d'una equació diferencial, després d'un temps determinat, partint d'unes condicions inicials representades per x .

Anomenem *arrel* o *solució* d'una equació no lineal $f(x) = 0$ un valor α tal que $f(\alpha) = 0$. En aquest cas es diu també que α és un *zero* de f .

En general, les arrels d'una equació no lineal no poden ésser calculades de forma exacta. L'objectiu d'aquest capítol és oferir mètodes numèrics que ens permetin obtenir aproximacions numèriques de les arrels.

En qualsevol procés de càlcul d'arrels d'una equació no lineal es poden distingir tres fases:

- LOCALITZACIÓ

Interessa tenir un cert coneixement de la zona on es troben les arrels. En general, aquesta informació s'obtindrà sigui mitjançant la confecció de taules de la funció, sigui a partir d'un estudi analític o, fins i tot, a partir d'una representació gràfica aproximada, quan la complexitat de la funció no ho impedeixi. En molts casos, les equacions provenen d'un problema tècnic o científic, el coneixement del qual pot ajudar a la localització de les arrels.

- SEPARACIÓ

A vegades dues arrels diferents d'una equació són molt pròximes. En aquests casos, abans d'aplicar qualsevol mètode numèric, convé separar les arrels; això és, determinar intervals que continguin una i sols una arrel de l'equació.

- APROXIMACIÓ NUMÈRICA

El seu objectiu, en general, consistirà a construir una successió de valors que convergeixi cap a l'arrel buscada. Aquesta construcció es farà, normalment, de manera iterativa partint de valors inicials que suposarem suficientment pròxims a l'arrel cercada: a partir de $x_0, \dots, x_m$, obtindrem $x_{m+1} = G(x_m, \dots, x_0)$ i, d'una manera més general, $x_{k+1} = G(x_k, \dots, x_{k-m})$.

A continuació es presenten alguns mètodes d'aproximació d'arrels. Més endavant dedicarem un apartat a l'estudi de la seva convergència.

5.1.2 Mètodes iteratius d'aproximació de solucions

Mètode de bisecció

Un cop trobats uns valors a i b en què la funció f canvia de signe, el teorema de Bolzano assegura que, si f és contínua en $[a, b]$ i $f(a)f(b) < 0$, llavors existeix $\alpha \in (a, b)$ tal que $f(\alpha) = 0$. De fet, en pot existir més d'una, d'aquestes arrels, però nosaltres suposarem que n'hi ha una, entenent que prèviament s'hauran efectuat processos de localització i separació. El mètode de bisecció construeix llavors una successió d'interval·ls encaixats,

$$(a_0, b_0) \supset (a_1, b_1) \supset \dots \supset (a_k, b_k) \supset \dots,$$

de manera que sempre continguin l'arrel buscada i que l'amplitud de cada un sigui la meitat de l'anterior. Quan l'amplitud de l'interval sigui prou petita d'acord amb la precisió desitjada per a l'arrel, podrem considerar com a una bona aproximació d'aquesta un qualsevol dels seus extrems.

Per a la construcció de la successió d'interval·ls es parteix d' a_0 i b_0 tals que $f(a_0)f(b_0) < 0$ i es considera l'abscissa mitjana de (a_0, b_0) , $c = \frac{1}{2}(a_0 + b_0)$. Si $f(c) = 0$, c és l'arrel buscada. Altrament s'agafarà com a (a_1, b_1) l'interval (a_0, c) o el (c, b_0) segons que $f(a_0)f(c) < 0$ o $f(c)f(b_0) < 0$, respectivament. El procés iteratiu es continua anàlogament (vegeu la figura 5.1).

Aquest mètode, tot i que convergeix sempre cap a l'arrel buscada, té l'inconvenient de no aprofitar, llevat del signe, les característiques

concretes de la funció f , raó per la qual és bastant lent.

Mètode de Newton

Suposem ara que f és derivable en un entorn de l'arrel buscada.

L'objectiu d'aquest mètode consisteix a construir una successió $(x_k)_{k \geq 0}$, convergent cap a l'arrel. Per fer-ho es parteix d'un valor x_0 , es tira la tangent a la corba $y = f(x)$ pel punt $(x_0, f(x_0))$ i es busca el seu punt de tall amb l'eix $y = 0$. L'abscissa d'aquest punt serà x_1 i així successivament (vegeu la figura 5.2).

En general,

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (k \geq 0),$$

suposant que f' no s'anul·la en els diferents x_k .

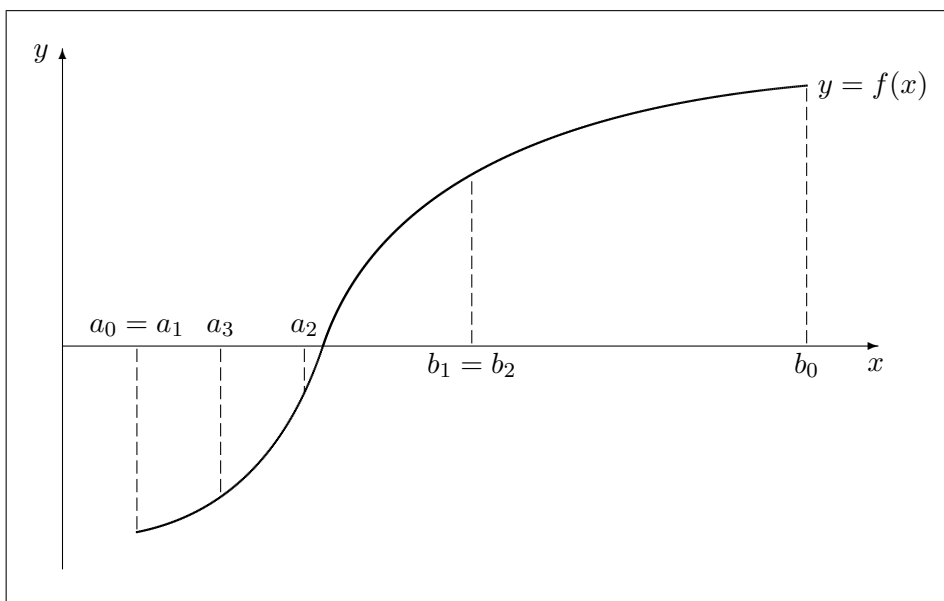


Figura 5.1: Mètode de bisecció.

La convergència d'aquest mètode no està assegurada per a qualsevol x_0 i és convenient escollir l'aproximació inicial x_0 tan pròxima com sigui possible a l'arrel. Ara bé, tal com es veurà més endavant, la seva convergència sol ser bastant ràpida; per això, és un dels mètodes més usats.

El mètode de Newton presenta problemes quan el zero α de f que es busca és múltiple; això és, tal que $f'(\alpha) = 0$. Tot i així pot ésser modificat de manera trivial per tal d'adaptar-lo a aquest cas (vegeu el problema V.3).

El criteri general per acabar la construcció d'aquesta successió és que $|x_k - x_{k-1}|$ sigui suficientment petit. En aquest cas, x_k es podrà considerar com a una bona aproximació de l'arrel buscada.

Cal fer notar aquí, encara que el que es diu val per a qualsevol mètode, que el criteri per acabar les iteracions no té perquè ser sempre que la correcció $|x_k - x_{k-1}|$ sigui més petita que un valor δ donat. Sigui α la solució exacta i suposem que $f'(\alpha)$ té un valor relativament petit i que els errors en el càlcul de f estan fitats per ϵ , llavors no es poden demanar errors en la determinació de α més petits que $\frac{\epsilon}{|f'(\alpha)|}$. Si $\frac{\epsilon}{|f'(\alpha)|} > \delta$, llavors el criteri adequat per acabar les iteracions no serà que $|x_k - x_{k-1}| < \delta$ sinó que $|f(x_k) - f(x_{k-1})| < 2\epsilon$.

Mètode de la secant

L'objectiu del mètode de la secant consisteix en la construcció d'una successió $x_0, x_1, x_2, \dots, x_k, \dots$ que convergeixi cap a l'arrel buscada. Per fer aquesta construcció es parteix de dos valors x_0 i x_1 i es tira la recta secant (interpolació lineal) a la corba $y = f(x)$ pels punts $(x_0, f(x_0))$ i $(x_1, f(x_1))$, prenent-se x_2 com l'abscissa del punt de tall d'aquesta recta amb l'eix $y = 0$. El procés continua trobant x_3 a partir de x_1 i x_2 i així successivament.

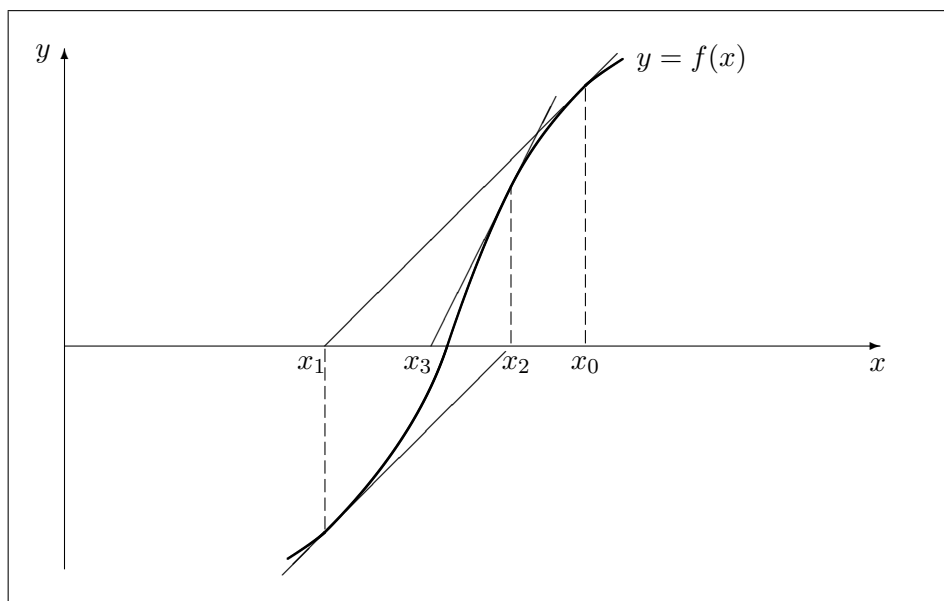


Figura 5.2: Mètode de Newton.

(vegeu la figura 5.3); en general:

$$x_{k+1} = x_k - f(x_k) \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} \quad (k \geq 1) .$$

De fet, és com el mètode de Newton, però agafant la secant com a aproximació de la tangent.

Si bé la convergència no està assegurada (convé escollir x_0 i x_1 tan pròxims com sigui possible a l'arrel), quan es dona, és bastant ràpida (com es veurà més endavant).

L'elecció entre la utilització del mètode de la secant i la del mètode de Newton depèn bàsicament de la quantitat de feina necessària per al càlcul de $f'(x_k)$. Es pot dir que, si l'esforç per a aquest càlcul és més gros que 0.44 vegades l'esforç de càlcul de $f(x_k)$, resulta més eficient usar el mètode de la secant (vegeu el problema V.5).

Una variant del mètode de la secant que recorda el de bisecció és el dit *mètode de la regula falsi*. En aquest es parteix de dos valors x_0 i x_1 tals que $f(x_0)f(x_1) < 0$ i es troba x_2 pel mètode de la secant. En el càlcul del valor següent x_3 no s'utilitzaran necessàriament x_1 i x_2 (com en el mètode de la secant), sinó x_0 i x_2 o x_1 i x_2 segons que $f(x_0)f(x_2) < 0$ o $f(x_1)f(x_2) < 0$, respectivament, i així successivament (vegeu la figura 5.4). Tot i que aquest mètode és més lent que el de la secant, té l'avantatge que sempre és convergent.

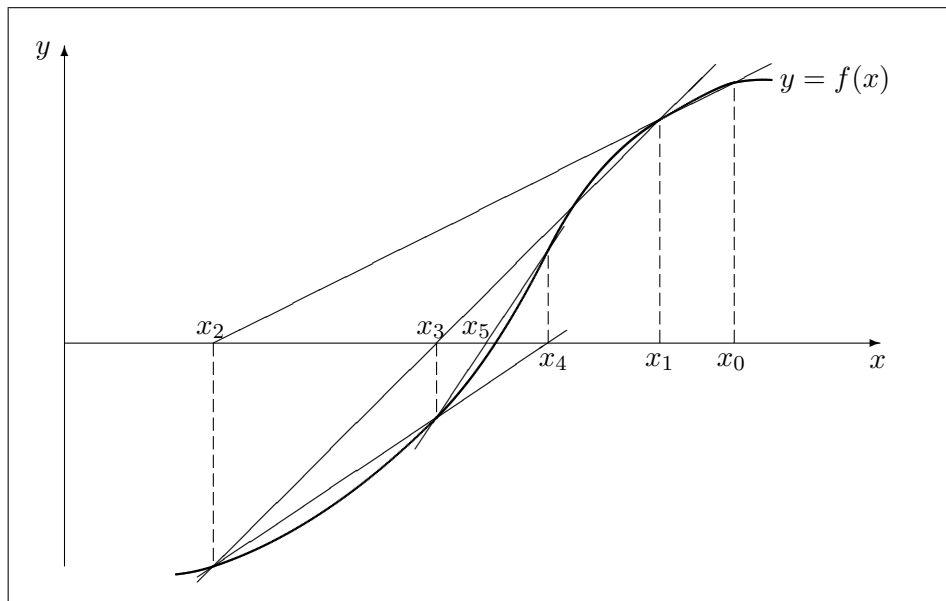


Figura 5.3: Mètode de la secant.

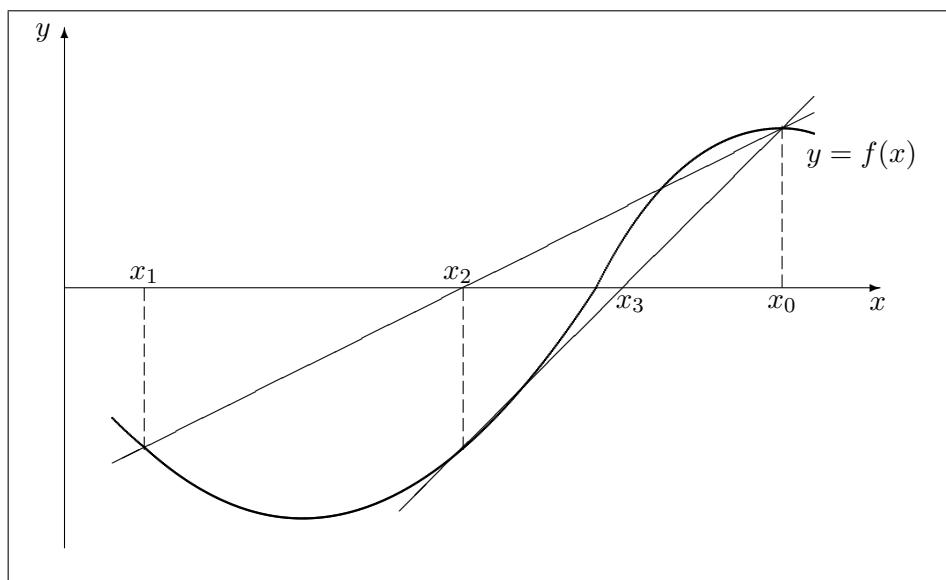


Figura 5.4: Mètode de la regula falsi.

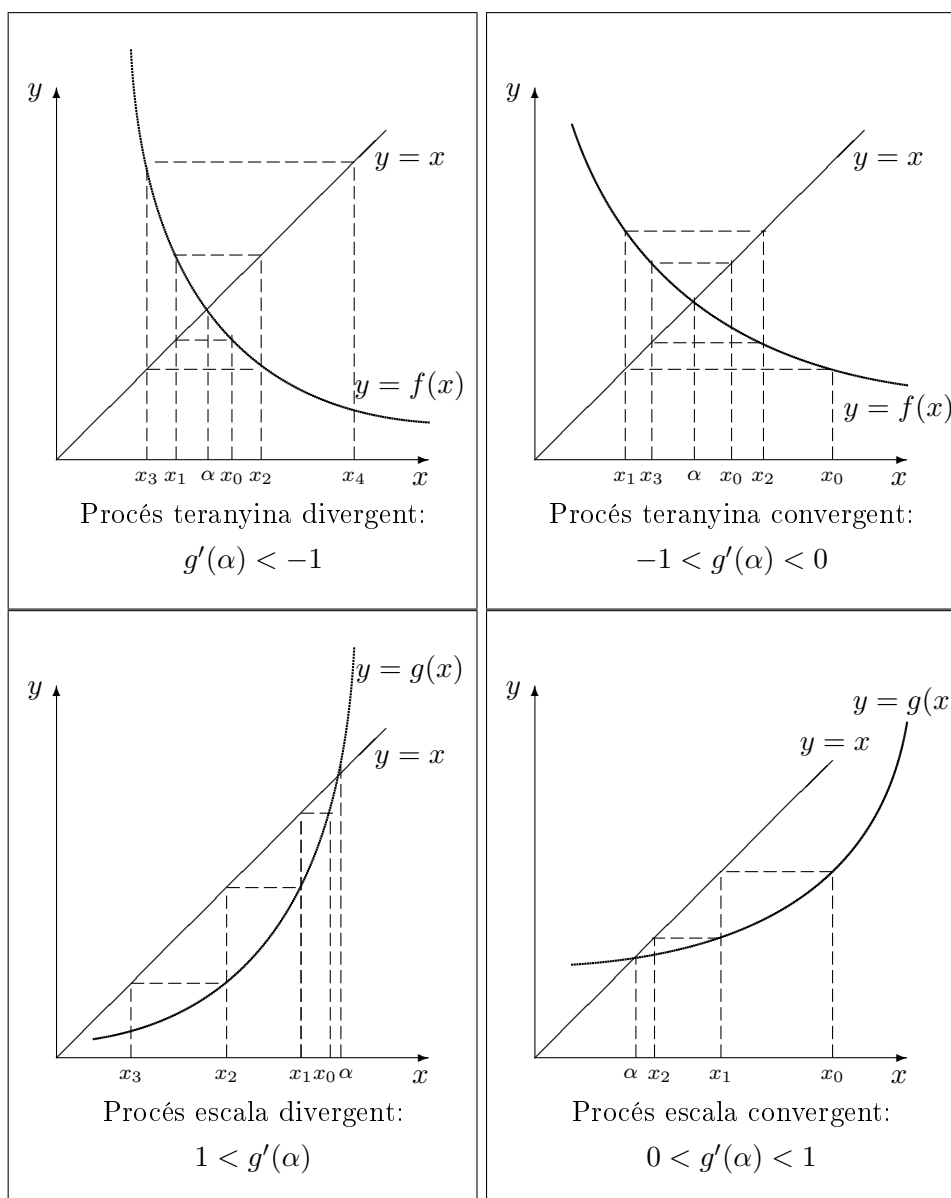


Figura 5.5: Processos escala i teranyina.

Mètode d'iteració simple

En aquest cas escriurem l'equació que volem resoldre de la forma $x = g(x)$; a una arrel α d'una equació d'aquest tipus, $\alpha - g(\alpha) = 0$, l'anomenarem *punt fix de g* . Novament, pretenem generar una successió de valors tendent a l'arrel: $x_0, x_1, x_2, \dots, x_k, \dots$. La construcció d'aquesta successió es realitza fent *iteracions simples* $x_{k+1} = g(x_k)$ ($k \geq 0$), a partir d'un x_0 escollit pròxim a l'arrel buscada.

La convergència, que no està assegurada, depèn de l'elecció de x_0 i de la funció g . En el cas que g sigui derivable, un criteri perquè, triant x_0 suficientment pròxim a l'arrel α ,

tinguem convergència és que $|g'| \leq r < 1$ en un entorn del valor α (en aquest cas direm que la funció g és *contractiva* prop del punt fix α); aquest entorn existeix si g' és contínua i $|g'(\alpha)| < 1$.

Les quatre representacions de la figura 5.5 il·lustren la convergència del mètode segons el valor de la derivada de la funció g en α .

5.1.3 Ordre de convergència i constant asimptòtica de l'error

Direm que una successió $(x_k)_{k \geq 0}$, convergent a un límit α , té *ordre de convergència almenys* p si existeixen $N \geq 0$ i $C > 0$ tals que

$$|x_{k+1} - \alpha| \leq C |x_k - \alpha|^p$$

per a qualsevol $k \geq N$. Quan $p = 1$ cal exigir que $C < 1$.

En particular, en el cas que existeixi

$$L = \lim_{k \rightarrow \infty} \frac{x_{k+1} - \alpha}{(x_k - \alpha)^p},$$

la successió $(x_k)_{k \geq 0}$ tindrà ordre de convergència almenys p (si $p = 1$, haurem d'exigir a més que $|L| < 1$). Si $L \neq 0$, direm que la successió $(x_k)_{k \geq 0}$ té *ordre de convergència* p i que L és la *constant asimptòtica de l'error*. Per a $p = 1, 2, 3$, parlarem de convergència almenys *lineal*, *quadràtica*, *cúbica*, respectivament.

Direm que un mètode iteratiu $x_{k+1} = G(x_k, x_{k-1}, \dots, x_{k-m})$ ($k \geq m$) és almenys d'ordre p prop d'un punt fix α de G : $\alpha = G(\alpha, \dots, \alpha)$, quan existeixi un entorn $(\alpha - \epsilon, \alpha + \epsilon)$ de manera que, per a qualsevol elecció de valors inicials $x_0, \dots, x_m$ en $(\alpha - \epsilon, \alpha + \epsilon)$, la successió $(x_k)_{k \geq 0}$ tingui almenys ordre de convergència p .

Si tenim un mètode d'ordre de convergència almenys p i no tenim en compte els errors d'arrodoniment (és a dir, donats $x_k, \dots, x_{k-m}$, suposem que podem calcular exactament $x_{k+1} = G(x_k, \dots, x_{k-m})$), llavors, anomenant ϵ_k a $|x_k - \alpha|$ i suposant que $\epsilon_i \leq C\epsilon_{i-1}^p$ ($i = 1 \div k$), es compleix:

$$\begin{aligned} \epsilon_k &\leq C^k \epsilon_0 \quad (p = 1), \\ \epsilon_k &\leq C^{1+p+\dots+p^{k-1}} \epsilon_0^{p^k} = \frac{(\overline{C}\epsilon_0)^{p^k}}{\overline{C}}, \quad (p > 1), \end{aligned}$$

amb

$$\overline{C} = C^{\frac{1}{p-1}}.$$

Així, com més alt sigui l'ordre de convergència, més ràpidament convergirà $(x_k)_{k \geq 0}$ a α .

De totes maneres, en l'estudi de l'eficiència d'un mètode, no s'ha de tenir en compte tan sols el seu ordre de convergència, sinó també la quantitat de càlcul necessari en cada iteració $x_{k+1} = G(x_k, \dots, x_{k-m})$ ($k \geq m$).

Tot seguit recollim algunes consideracions sobre ordre de convergència:

- I. Considerem l'equació $x = g(x)$ i sigui α una solució.

1. Si g és contractiva en un entorn de α , el mètode d'iteració simple $x_{k+1} = g(x_k)$ ($k \geq 0$) és almenys d'ordre 1.
 2. Si g és m vegades derivable i $g'(\alpha) = \dots = g^{(m-1)}(\alpha) = 0$, llavors el mètode d'iteració simple $x_{k+1} = g(x_k)$ ($k \geq 0$) és almenys d'ordre m ; si $g^{(m)}(\alpha) \neq 0$, és d'ordre m .
- II. Considerem ara l'equació $f(x) = 0$ i sigui α un zero de f : $f(\alpha) = 0$.
 1. Si $f'(\alpha) \neq 0$, el mètode de Newton té convergència almenys quadràtica.
 2. Si $f'(\alpha) = 0$, el mètode de Newton té convergència lineal.
 3. Si $f'(\alpha) \neq 0$, el mètode de la secant té ordre de convergència almenys $p = \frac{1+\sqrt{5}}{2}$.
 4. Si $f^{(2)}(\alpha) \neq 0$, el mètode de la regula falsi té convergència almenys lineal. En comparació amb el mètode de la secant, s'assegura la convergència a canvi d'una pèrdua de velocitat d'aproximació.
 5. Independentment del comportament de f es pot considerar que el mètode de bisecció té convergència lineal ja que, a cada pas, reduïm a la meitat l'interval d'error on es troba el zero buscat.

5.1.4 Acceleració de la convergència

Observem que els mètodes proposats per trobar arrels d'equacions no lineals tenen com a objectiu comú la construcció d'una successió de valors $(x_k)_{k \geq 0}$ convergent (en el millor dels casos) cap a l'arrel buscada. La rapidesa amb què es produeix aquesta convergència depèn del mètode usat i de la pròpia equació. Els anomenats *mètodes d'acceleració de la convergència* tenen com a objectiu l'obtenció, a partir de la successió $(x_k)_{k \geq 0}$, d'una nova successió $(x'_k)_{k \geq 0}$ més ràpidament convergent cap a l'arrel. Veurem dos d'aquests mètodes.

Mètode d'acceleració d'Aitken

Suposem que la successió $(x_k)_{k \geq 0}$ de límit α , amb $x_k \neq \alpha$, és tal que l'error es comporta asimptòticament com una successió geomètrica de raó més petita que 1, en valor absolut; és a dir,

$$x_{k+1} - \alpha = (L + \delta_k)(x_k - \alpha) ,$$

amb

$$\lim_{k \rightarrow \infty} \delta_k = 0 , \quad |L| < 1 ;$$

llavors, la successió $(x'_k)_{k \geq 0}$ donada per

$$x'_k = x_k - \frac{(\Delta x_k)^2}{\Delta^2 x_k} = x_k - \frac{(x_{k+1} - x_k)^2}{x_{k+2} - 2x_{k+1} + x_k} \quad (k \geq 0) ,$$

està ben definida per a k suficientment gran i compleix que

$$\lim_{k \rightarrow \infty} \frac{x'_k - \alpha}{x_k - \alpha} = 0 .$$

Per tant, $(x'_k)_{k \geq 0}$ convergeix cap a α més ràpidament que $(x_k)_{k \geq 0}$; aquest mètode permet, doncs, accelerar la convergència de mètodes amb convergència lineal.

Mètode d'acceleració de Steffensen

Si tenim una iteració simple $x_{k+1} = g(x_k)$ ($k \geq 0$), el *mètode d'acceleració de Steffensen* construeix la iteració simple donada per $y_{k+1} = G(y_k)$, amb

$$G(x) = x - \frac{(g(x) - x)^2}{g(g(x)) - 2g(x) + x} \quad (k \geq 0).$$

Es compleix:

1. Si $G(\alpha) = \alpha$, llavors $g(\alpha) = \alpha$.
2. Si $g(\alpha) = \alpha$ i $g'(\alpha) \neq 1$, llavors $G(\alpha) = \alpha$.
3. Si $(x_k)_{k \geq 0}$ té convergència almenys lineal, llavors $(y_k)_{k \geq 0}$ té convergència almenys quadràtica.
4. Si $(x_k)_{k \geq 0}$ té ordre de convergència almenys $p > 1$, llavors $(y_k)_{k \geq 0}$ té ordre de convergència almenys $2p - 1$.

En aplicar aquest mètode al cas de convergència lineal, obtenim una acceleració de convergència ja que $(y_k)_{k \geq 0}$ convergeix més ràpidament que $(x_k)_{k \geq 0}$. Quan l'ordre de convergència del mètode iteratiu inicial sigui $p > 1$, l'aplicació del mètode no té avantatges: el mètode d'acceleració de Steffensen aconsegueix accelerar la convergència fins a ordre $2p - 1$ mentre que, prenent senzillament $\bar{G}(x) = g(g(x))$ (càlcul d'altra banda necessari en l'avaluació de $G(x)$ pel mètode de Steffensen), es pot generar una successió d'iterats d'ordre p^2 . En el problema V.2 es presenta una altra expressió de la iteració simple de Steffensen utilitzable quan tenim equacions del tipus $f(x) = 0$.

La diferència fonamental entre els mètodes d'acceleració d'Aitken i de Steffensen aplicats a un mètode d'iteració simple, rau en el fet que, en el mètode d'Aitken, l'acceleració es produeix sobre la successió d'iterats ja construïda d'entrada, mentre que, en el mètode de Steffensen, es construeix directament la successió accelerada. En general, el mètode de Steffensen ofereix millors resultats que el d'Aitken ja que, per al càlcul dels successius termes de $(y_k)_{k \geq 0}$, utilitza valors que, de fet, han estat ja accelerats.

5.1.5 Classificació de mètodes iteratius

S'acaben de presentar diversos mètodes iteratius per al càlcul aproximat d'arrels d'equacions. Tots ells tenen en comú la construcció d'una successió $x_0, x_1, x_2, \dots, x_k, \dots$ que, sota determinades condicions, tendeix a l'arrel α buscada. L'objectiu d'aquest apartat és donar una visió global d'aquests procediments, obrint l'horitzó a altres mètodes que, si bé tenen el mateix fonament que els presentats, no són tan usats per la dificultat de càlcul que comporten.

Els mètodes iteratius d'aproximació de zeros d'una funció f es poden classificar atenent a dos trets que els caracteritzen:

1. L'eina que s'usa per a l'estudi de la funció:
 - El desenvolupament de Taylor: calculem x_{k+1} a partir del coneixement de la funció i de les seves derivades en un punt x_k . Es parlarà de *mètodes de Taylor*.

- La interpolació: calculem x_{k+1} a partir del coneixement de la funció en els iterats anteriors $x_k, x_{k-1}, \dots$. Es parlarà de *mètodes d'interpolació*.
2. La funció que es desenvolupa o s'interpola: si $g = f^{-1}$ (noteu que la inversa de f existirà sempre prop de zeros simples de f), serà equivalent cercar α tal que $f(\alpha) = 0$ a buscar α tal que $g(0) = \alpha$. Així podrem donar dos enfocaments del problema segons que usem el desenvolupament de Taylor o la interpolació sobre la funció f o la funció g :
- Si usem la funció f , es parlarà de *mètodes directes*.
 - Si usem la funció $g = f^{-1}$, es parlarà de *mètodes inversos*.

Aquestes característiques classifiquen els mètodes iteratius tractats ací en 4 famílies: mètodes de Taylor directes, mètodes de Taylor inversos, mètodes d'interpolació directes i mètodes d'interpolació inversos.

Mètodes de Taylor directes

El càlcul de x_{k+1} a partir de x_k , en aquests mètodes, es basa en el procediment següent que, finalment, es resumeix en una fórmula iterativa:

- Es pren com a aproximació de $f(x)$ el seu polinomi de Taylor $p_n(x)$ de grau més petit o igual que n en x_k .
- Es resol l'equació $p_n(x) = 0$. La solució obtinguda (si n'hi ha més d'una, es pren la més propera a x_k) serà el valor x_{k+1} .
- Es repeteix el procés a partir de x_{k+1} fins a obtenir la precisió desitjada.

Els diferents mètodes d'aquesta família apareixen segons el valor de n que es prengui:

- Per a $n = 1$, resulta el mètode de Newton ja estudiat.
- Per a $n = 2$, resulta la iteració simple

$$x_{k+1} = x_k - \frac{2f(x_k)}{f'(x_k) \pm \sqrt{f'^2(x_k) - 2f(x_k)f^{(2)}(x_k)}}.$$

- Per a $n > 2$, l'equació $p_n(x) = 0$ no permet un càlcul senzill de x_{k+1} .

Mètodes de Taylor inversos

El càlcul de x_{k+1} a partir de x_k , en aquests mètodes, es basa en el procediment següent que, finalment, es resumeix en una fórmula iterativa:

- Es pren com a aproximació de $g(y) = f^{-1}(y)$ el seu polinomi de Taylor $q_n(y)$ de grau més petit o igual que n en $y_k = f(x_k)$. Les derivades successives de g s'obtenen derivant repetidament la relació $g(f(x)) = x$ (vegeu el problema V.1).

- Es pren $x_{k+1} = q_n(0)$. Serà una nova aproximació del valor α buscat.
- Es repeteix el procés a partir de x_{k+1} fins a obtenir la precisió desitjada.

Els diferents mètodes d'aquesta família apareixen segons el valor de n que es prengui:

- Per a $n = 1$, resulta novament el mètode de Newton.
- Per a $n = 2$, resulta el *mètode de Trebixev*, contemplat detalladament en el problema V.1.
- Per a $n = 3$, resulta el mètode iteratiu:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} - \frac{f^{(2)}(x_k)f^2(x_k)}{2f'^3(x_k)} - \frac{3f^{(2)2}(x_k) - f^{(3)}(x_k)f'(x_k)}{6f'^5(x_k)}f^3(x_k).$$

Si bé l'ordre d'aquest mètode és alt, l'excés de càlcul de cada iteració imposa una seriosa dificultat pel que fa a la seva eficiència.

- Per a $n > 3$, les expressions que apareixen no permeten un càlcul senzill de x_{k+1} .

Mètodes d'interpolació directes

El càlcul de x_{k+1} a partir de $x_k, x_{k-1}, \dots, x_0$, en aquests mètodes, es basa en el procediment següent, que, finalment, es resumeix en una fórmula iterativa:

- Es consideren $m + 1$ aproximacions del valor α buscat: $x_k, x_{k-1}, \dots, x_{k-m}$.
- Es construeix el polinomi interpolador $p_m(x)$ de la funció f en les abscisses $x_k, x_{k-1}, \dots, x_{k-m}$.
- Es resol l'equació $p_m(x) = 0$. La solució obtinguda més propera a x_k serà el valor x_{k+1} .
- Es repeteix el procés a partir de $x_{k+1}, x_k, \dots, x_{k-m+1}$ fins a obtenir la precisió desitjada.

Els diferents mètodes d'aquesta família apareixen segons el valor de m que es prengui:

- Per a $m = 1$, resulta el mètode de la secant ja estudiat.
- Per a $m = 2$, resulta el *mètode de Muller-Traub* que s'exposa en la segona part d'aquest capítol ja que generalment s'usa per al càlcul de zeros de polinomis.
- Per a $m > 2$, l'equació $p_m(x) = 0$ no permet un càlcul senzill de x_{k+1} .

Mètodes d'interpolació inversos

El càlcul de x_{k+1} a partir de $x_k, x_{k-1}, \dots, x_0$, en aquests mètodes, es basa en el procediment següent que, finalment, es resumeix en una fórmula iterativa:

- Es consideren $m + 1$ aproximacions del valor α buscat: $x_k, x_{k-1}, x_{k-2}, \dots, x_{k-m}$.
- Es considera la funció $g = f^{-1}$.
- Es construeix el polinomi interpolador $q_m(y)$ de la funció g en $f(x_k), f(x_{k-1}), \dots, f(x_{k-m})$.
- Es pren $x_{k+1} = q_m(0)$. Serà una nova aproximació del valor $\alpha = g(0)$ buscat.
- Es repeteix el procés a partir de $x_{k+1}, x_k, \dots, x_{k-m+1}$ fins a obtenir la precisió desitjada.

Els diferents mètodes d'aquesta família apareixen segons el valor de m que es prengui:

- Per a $m = 1$, resulta novament el mètode de la secant.
- Per a $m = 2$, resulta el mètode que es tracta en el problema V.4.
- Per a $m > 2$, l'expressió de $q_m(y)$ no permet un càlcul senzill de x_{k+1} .

5.2 EQUACIONS POLINOMIALS

5.2.1 Introducció

El teorema fonamental de l'àlgebra ens assegura que, donat un polinomi de grau n amb coeficients complexos

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 ,$$

amb $a_n \neq 0$ i $a_i \in \mathbb{C}$ ($i = 0 \div n$), existeixen n zeros complexos $\alpha_1, \alpha_2, \dots, \alpha_n$ (repetits segons llur multiplicitat) i, llavors,

$$p(x) = a_n(x - \alpha_1)(x - \alpha_2) \dots (x - \alpha_n) .$$

És també conegut que, si $a_0, a_1, \dots, a_n$ són reals, els zeros complexos de $p(x)$ apareixen en parelles conjugades.

Per a l'obtenció dels zeros de polinomis de grau 2, 3 o 4, es coneixen fórmules generals; en canvi, se sap que aquestes fórmules no existeixen per a graus superiors a 4. Així, si salvem els casos trivials en què la factorització del polinomi aparegui com a evident, convindrà que utilitzem mètodes numèrics per a la recerca dels zeros. Fins i tot per a polinomis de tercer i quart grau, l'ús de les fórmules no és senzill i hom prefereix els procediments numèrics.

En emprar els mètodes numèrics en polinomis, gaudirem d'uns avantatges particulars: primerament, la facilitat en l'avaluació de $p(x)$ i de les seves derivades i, segona, l'existència de regles especials per a la fitació i separació dels seus zeros. A més dels mètodes estudiats d'aproximació de zeros, existeixen mètodes específics per a polinomis. També convé subratllar que, conegut un zero d'un polinomi, podem cercar la resta com a zeros d'un polinomi de grau menor. Tot seguit estudiem aquests procediments específics per a polinomis.

5.2.2 Avaluació i deflació de polinomis

L'avaluació de polinomis, i també la de llurs derivades, pot ser feta de forma eficient mitjançant la regla de Horner, com s'explica en l'apartat III.1.2.

La regla de Horner ens ofereix també la possibilitat d'obtenir el polinomi $p_1(x)$ que sorgeix en eliminar de $p(x)$ un zero α_1 ; això és,

$$p(x) = p_1(x)(x - \alpha_1) .$$

La resta de zeros de $p(x)$ la cercarem calculant els zeros de $p_1(x)$, el grau del qual és una unitat menor que el de $p(x)$. Gràcies a aquest procediment, anomenat *deflació de polinomis*, cada vegada que obtinguem un zero d'un polinomi podrem buscar la resta de zeros treballant amb un polinomi d'un grau menor. Si α_2 és un zero del nou polinomi $p_1(x)$, podrem realitzar una nova deflació i continuar treballant amb un polinomi $p_2(x)$ d'un grau menor, i així successivament.

Ara bé, com que els zeros $\alpha_1, \alpha_2, \dots$ només es coneixen aproximadament, $p_1(x)$ està afectat per errors que es propagaran en calcular $p_2(x)$, etc, amb la qual cosa pot ocórrer que els darrers zeros $\dots, \alpha_{n-1}, \alpha_n$ que calculem tinguin un error molt gran respecte als zeros exactes del polinomi $p(x)$.

Per tal que aquests errors en els zeros calculats no s'amplifiquin considerablement, convindrà calcular els zeros $\alpha_1, \alpha_2, \dots$ de menor a major en mòdul $|\alpha_1| \leq |\alpha_2| \leq \dots$; és a dir, s'ha de calcular el zero de mòdul més petit de cada polinomi intermedi $p_i(x)$.

Aquest procés també es pot realitzar en sentit contrari; és a dir, calculant el zero de major valor absolut de cada polinomi intermedi $p_i(x)$, sempre i quan la deflació es realitzi cap al darrera.

Donat el polinomi

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 ,$$

amb a_n i a_0 no nuls, la *deflació cap al darrera* parteix del polinomi

$$q(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_n ,$$

que compleix $p(x) = x^n q(\frac{1}{x})$ per a $x \neq 0$ i, per tant, $p(\alpha) = 0$ si i només si $q(\frac{1}{\alpha}) = 0$. A continuació, en lloc de realitzar la deflació estàndard, realitza la deflació

$$q_1(x) = \frac{q(x)}{x - \frac{1}{\alpha}} = b_0 x^{n-1} + \dots + b_{n-1} .$$

El polinomi resultant

$$p_1(x) = x^{n-1} q_1\left(\frac{1}{x}\right) = b_{n-1} x^{n-1} + \dots + b_0$$

compleix que

$$p_1(x) = \frac{p(x)}{1 - \frac{x}{\alpha}} = -\alpha \frac{p(x)}{x - \alpha} .$$

Remarquem que els processos de deflació comporten errors numèrics, molts cops importants, en l'obtenció dels polinomis intermedis i, per tant, en les aproximacions dels zeros $\alpha_2, \alpha_3, \dots$ trobades. És per això que, després de calcular-les, convé purificar-les amb algun procés iteratiu sobre el polinomi original $p(x)$. Normalment, són suficients poques iteracions.

5.2.3 Fitació de zeros de polinomis

Es coneixen molts resultats sobre fitació de zeros de polinomis. Tot seguit n'enunciem dos d'útils quan els zeros són reals.

Regla de Laguerre-Thibault

Si, en dividir $p(x)$ per $x - b$ amb $b > 0$, tots els coeficients del quocient i el residu són positius, llavors b és fita superior dels zeros reals de $p(x)$.

Regla de Newton

Si $p(x)$ és un polinomi de grau n i, per a un nombre real b , $p(b)$, $p'(b)$, $\dots$, $p^{(n)}(b)$ són positius, llavors b és una fita superior dels zeros reals de $p(x)$.

Aquestes regles es poden utilitzar per trobar una fita inferior dels zeros positius. Donat $p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$, amb $a_0 \neq 0$, considerem $q(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_n$. Si C és fita superior dels zeros positius de $q(x)$, $D = \frac{1}{C}$ serà fita inferior dels zeros positius de $p(x)$.

Si volem trobar la fita superior i la fita inferior dels zeros negatius de $p(x)$, també podem emprar les esmentades regles cercant la fita inferior i la fita superior, respectivament, dels zeros positius del polinomi $r(x) = p(-x)$.

Uns altres resultats sobre fitació dels zeros α_i , ara tant reals com complexos, d'un polinomi $p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$ són:

$$\begin{aligned} |\alpha_i| &\leq \max \left\{ \left| \frac{a_0}{a_n} \right|, 1 + \left| \frac{a_1}{a_n} \right|, \dots, 1 + \left| \frac{a_{n-1}}{a_n} \right| \right\} , \\ |\alpha_i| &\leq \max \left\{ 1, \sum_{k=0}^{n-1} \left| \frac{a_k}{a_n} \right| \right\} , \\ |\alpha_i| &\leq \max \left\{ \left| \frac{a_0}{a_1} \right|, 2 \left| \frac{a_1}{a_2} \right|, \dots, 2 \left| \frac{a_{n-1}}{a_n} \right| \right\} , \\ |\alpha_i| &\leq \sum_{k=0}^{n-1} \left| \frac{a_k}{a_{k+1}} \right| , \\ |\alpha_i| &\leq 2 \max \left\{ \left| \frac{a_{n-1}}{a_n} \right|, \sqrt{\left| \frac{a_{n-2}}{a_n} \right|}, \sqrt[3]{\left| \frac{a_{n-3}}{a_n} \right|}, \dots, \sqrt[n]{\left| \frac{a_0}{a_n} \right|} \right\} . \end{aligned}$$

5.2.4 Separació de zeros reals de polinomis. Mètode de Sturm

Successions de Sturm

Una successió $\{f_0, f_1, \dots, f_m\}$ de funcions reals contínues, definides sobre un interval $[a, b]$, és una *successió de Sturm* per a $f = f_0$ sobre $[a, b]$ si

1. f_0 és diferenciable en $[a, b]$.
2. f_m no té zeros reals en $[a, b]$.
3. Si $f_0(\alpha) = 0$, llavors $f_1(\alpha)f'_0(\alpha) > 0$.

4. Si $f_i(\alpha) = 0$, llavors $f_{i+1}(\alpha)f_{i-1}(\alpha) < 0$ ($i = 1 \div m - 1$).

La importància de les successions de Sturm rau en el *teorema de Sturm* següent:

- Sigui $\{f_0 = f, f_1, \dots, f_m\}$ una successió de Sturm per a f sobre $[a, b]$ amb $f(a)f(b) \neq 0$; aleshores, el nombre de zeros reals de f en l'interval (a, b) és igual a $V(a) - V(b)$, on $V(x)$ és el nombre de canvis de signe de $\{f_0(x), f_1(x), \dots, f_m(x)\}$.

Així doncs, si coneixem una successió de Sturm per a una funció f , podem separar tots els seus zeros reals de manera anàloga al mètode de bisecció (vegeu el problema V.7). Tot seguit s'indica com construir una successió de Sturm per a un polinomi $p(x)$.

Sigui $p(x)$ un polinomi real de grau n . Definim una successió de polinomis $(p_i(x))_{i=0, \dots, m}$, amb $m \leq n$, de la manera següent:

$$\begin{aligned} p_0(x) &= p(x), \\ p_1(x) &= p'(x), \\ p_{i-1}(x) &= q_i(x)p_i(x) - c_i p_{i+1}(x) \quad (i = 1 \div m - 1), \\ p_{m-1}(x) &= q_m(x)p_m(x), \end{aligned} \tag{5.1}$$

on $q_i(x)$ i $c_i p_{i+1}(x)$ són el quocient i el residu, aquest canviat de signe, de la divisió de $p_{i-1}(x)$ entre $p_i(x)$, i c_i és una constant positiva arbitrària que generalment es tria de manera que no apareguin coeficients fraccionaris en $p_{i+1}(x)$. La construcció d'aquesta successió acaba quan el residu és zero: $p_{m+1}(x) \equiv 0$. Com que l'algorisme usat és essencialment l'algorisme d'Euclides per al càlcul del màxim comú divisor de $p_0(x)$ i $p_1(x)$, tindrem que $p_m(x) = \text{m.c.d.}(p_0(x), p_1(x))$.

Si tots els zeros de $p(x)$ són simples, $p_m(x)$ és un polinomi que no s'anul·la sobre la recta real i $\{p_0(x), \dots, p_m(x)\}$ és una successió de Sturm per a $p(x)$ sobre qualsevol interval de la recta real. Si $p(x)$ té zeros reals múltiples, la dita successió no és de Sturm però encara es compleix la tesi del teorema de Sturm (vegeu el problema V.7). Així podem enunciar el resultat següent:

- Siguin $p_0(x) = p(x)$, $p_1(x)$, $\dots$, $p_m(x)$ construïts mitjançant la recurrència expressada en (5.1), amb $p(a)p(b) \neq 0$; aleshores, el nombre de zeros reals diferents del polinomi $p(x)$ en (a, b) és igual a $V(a) - V(b)$, on $V(x)$ indica el nombre de canvis de signe de

$$\{p_0(x), p_1(x), \dots, p_m(x)\}.$$

Com a conseqüència d'aquest resultat, calculant $V(x)$ per a diferents valors de x , podem determinar intervals de la recta real que continguin un únic zero de $p(x)$. Aquest mètode de separació de zeros de polinomis rep el nom de *mètode de Sturm*.

5.2.5 Mètodes numèrics per al càlcul de zeros de polinomis

En plantejar-nos el problema de calcular zeros de polinomis, hem de tenir en compte els dos avantatges que els polinomis ens ofereixen:

- La possibilitat d'avaluació ràpida d'un polinomi i de la seva derivada mitjançant la regla de Horner.

- L'ús de la deflació que, conegut un zero d'un polinomi, permet cercar els altres zeros treballant amb un polinomi d'un grau més petit que l'inicial.

Els mètodes numèrics per al càlcul de zeros de polinomis es poden classificar en dos grups: els mètodes generals de recerca de zeros i els mètodes específics per a polinomis. El primer grup està format pels mètodes estudiats a l'apartat 5.1.2: bisecció, secant, Newton (que, en aquest cas, rep el nom de Birge-Vieta), regla falsi, iteració simple, ..., aplicats ara a funcions polinòmiques. Tot seguit se'n presenten alguns del segon grup.

Mètode de Laguerre

El *mètode de Laguerre* és un mètode iteratiu l'objectiu del qual consisteix a generar una successió de valors (reals o complexos) $z_0, z_1, \dots, z_k, \dots$ que convergeixi cap a un zero que pot ésser real o complex. Per a la construcció d'aquesta successió s'utilitza la fórmula següent:

$$z_{k+1} = z_k - \frac{np(z_k)}{p'(z_k) \pm \sqrt{H(z_k)}} ,$$

on $p(x)$ és el polinomi sobre el qual treballem, n el seu grau i

$$H(x) = (n-1)[(n-1)(p'(x))^2 - np(x)p^{(2)}(x)] .$$

El signe del denominador es pren de manera que el faci més gran, en mòdul.

Aquest mètode és d'alt ordre de convergència: per a zeros reals és cúbicament convergent i té l'avantatge que, quan tots els zeros de $p(x)$ són reals, convergeix sigui quina sigui l'aproximació inicial real z_0 que s'hagi escollit. En aquest cas es diu que és *globalment convergent*.

Mètode de Bernoulli

Sigui

$$\begin{aligned} p(x) &= a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \\ &= a_n (x^n + b_{n-1} x^{n-1} + \dots + b_1 x + b_0) , \end{aligned}$$

amb

$$b_i = \frac{a_i}{a_n} \quad (i = 0 \div n-1) .$$

Una *matriu de Frobenius* associada a aquest polinomi (vegeu el problema II.10) és la matriu

$$B = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -b_0 & -b_1 & -b_2 & \dots & -b_{n-2} & -b_{n-1} \end{pmatrix} .$$

Es compleix que

$$\det(B - xI) = (-1)^n (x^n + b_{n-1} x^{n-1} + \dots + b_1 x + b_0) .$$

Per tant, trobar els zeros de $p(x)$ equival a trobar els valors propis de la matriu B . Com a conseqüència, si apliquem a la matriu B el mètode de la potència exposat en el capítol II, obtindrem el zero de $p(x)$ de mòdul màxim, suposat real i únic: $|\alpha_1| > |\alpha_2| \geq \dots \geq |\alpha_n|$. Així doncs, donats $x_0, x_1, \dots, x_{n-1}$ arbitraris, es compleix que

$$B \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ amb } x_n = -b_0x_0 - b_1x_1 - \dots - b_{n-1}x_{n-1} ;$$

és a dir,

$$x_n = -\frac{a_0x_0 + a_1x_1 + \dots + a_{n-1}x_{n-1}}{a_n} .$$

Repetint successivament aquest procés, obtenim la fórmula iterativa corresponent al *mètode de Bernoulli*

$$x_{k+n} = -\frac{a_0x_k + a_1x_{k+1} + \dots + a_{n-1}x_{k+n-1}}{a_n} \quad (k \geq 0) .$$

Sota condicions anàlogues a les comentades en el mètode de la potència del capítol II, tenim

$$\lim_{i \rightarrow \infty} \frac{x_{i+1}}{x_i} = \alpha_1 ,$$

on α_1 és el zero de $p(x)$ de major valor absolut. La velocitat de convergència depèn essencialment de $|\frac{\alpha_2}{\alpha_1}|$ i pot ésser molt lenta quan $|\alpha_2| \sim |\alpha_1|$.

Aplicant el mètode de la potència inversa a la matriu $B - aI$, podem trobar el zero real més pròxim al valor a , suposat únic en mòdul. En el cas $a = 0$, cercarem el zero més petit en valor absolut i l'algorisme és fàcilment expressable, ja que equival a buscar l'invers del zero més gran en valor absolut del polinomi

$$q(x) = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n .$$

Així doncs, quan $a = 0$, donats $y_0, \dots, y_{n-1}$ arbitraris, calculem y_{k+n} , ($k \geq 0$), mitjançant la fórmula

$$y_{k+n} = -\frac{a_ny_k + a_{n-1}y_{k+1} + \dots + a_1y_{k+n-1}}{a_0} .$$

Sota condicions anàlogues a les del mètode de la potència, es compleix ara que

$$\lim_{i \rightarrow \infty} \frac{y_{i+1}}{y_i} = \frac{1}{\alpha_n} ,$$

on α_n és el zero de $p(x)$ de mòdul més petit.

Mètode del quocient-diferència (Q-D)

El *mètode del quocient-diferència* és una generalització del mètode de Bernoulli que genera, sota condicions adequades, successions convergents a tots els zeros del polinomi $p(x) = a_nx^n + a_{n-1}x^{n-1} + \dots + a_0$. El procediment es basa en la construcció de la taula 5.1.

Els elements $q_k^{(i)}, d_k^{(i)}$ ($k \geq 2 - i$) ($i = 1 \div n$) de la taula es van calculant, per files successives, mitjançant les anomenades *regles dels rombes*:

	$-\frac{a_{n-1}}{a_n}$		0		0	$\dots$	0		0	
0		$\frac{a_{n-2}}{a_{n-1}}$		$\frac{a_{n-3}}{a_{n-2}}$		$\dots$		$\frac{a_0}{a_1}$		0
	$q_1^{(1)}$		$q_0^{(2)}$		$q_{-1}^{(3)}$	$\dots$	$q_{-n+3}^{(n-1)}$		$q_{-n+2}^{(n)}$	
0		$d_1^{(1)}$		$d_0^{(2)}$		$\dots$		$d_{-n+3}^{(n-1)}$		0
	$q_2^{(1)}$		$q_1^{(2)}$		$q_0^{(3)}$	$\dots$	$q_{-n+4}^{(n-1)}$		$q_{-n+3}^{(n)}$	
0		$d_2^{(1)}$		$d_1^{(2)}$		$\dots$		$d_{-n+4}^{(n-1)}$		0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$

Taula 5.1: Esquema del mètode de quocient-diferència.

- $q_{k+1}^{(i)} = (d_k^{(i)} - d_{k+1}^{(i-1)}) + q_k^{(i)}$, gràficament:

$$\begin{array}{ccc}
 & q_k^{(i)} & \\
 & + & \\
 d_{k+1}^{(i-1)} & & d_k^{(i)} \\
 + & & \\
 & q_{k+1}^{(i)} &
 \end{array}$$

- $d_{k+1}^{(i)} = \frac{q_k^{(i+1)}}{q_{k+1}^{(i)}} d_k^{(i)}$, gràficament:

$$\begin{array}{ccc}
 & d_k^{(i)} & \\
 & \times & \\
 q_{k+1}^{(i)} & & q_k^{(i+1)} \\
 \times & & \\
 & d_{k+1}^{(i)} &
 \end{array}$$

És a dir, per al càlcul d'un element de la taula, és suficient de tenir en compte el rombe format per ell i els tres elements superiors. Si el rombe està centrat en la columna de les $q^{(i)}$ la suma dels elements situats en la part superior-dreta és igual a la dels elements situats en la part inferior-esquerra. Si el rombe està centrat en la columna de les $d^{(i)}$ es pot afirmar el mateix, però substituint sumes per productes.

Es compleix el resultat següent: si l'esquema que s'acaba de descriure es pot construir (això és, no s'anul·la cap $q_{k+1}^{(i)}$), aleshores

$$\lim_{k \rightarrow \infty} q_k^{(i)} = \alpha_i \quad \text{i} \quad \lim_{k \rightarrow \infty} d_k^{(i)} = 0$$

per a tot índex i tal que $|\alpha_{i+1}| < |\alpha_i| < |\alpha_{i-1}|$.

Cal notar que la successió $(q_k^{(1)})_{k \geq 1}$ és una successió de quocients associada al mètode de Bernoulli i que la velocitat de convergència d'aquestes successions $(q_k^{(i)})_{k \geq 2-i}$ depèn dels valors $|\frac{\alpha_i}{\alpha_{i+1}}|$, essent lenta quan aquests quocients són propers a 1. Un tal cas fa aconsellable la utilització d'aquest mètode només per obtenir una primera aproximació

dels zeros del polinomi. Aquestes aproximacions s'hauran de purificar després, aplicant un mètode més ràpidament convergent, com el de Newton o el de la secant, per exemple.

L'oscil·lació en els valors dels elements $d_k^{(i)}$ d'una mateixa columna k indica l'existència d'una parella de zeros complexos conjugats. En tal cas, convindrà prendre les successions $q_{k+1}^{(i)} + q_k^{(i+1)}$ i $q_k^{(i)} q_k^{(i+1)}$. Aquestes convergeixen respectivament a nombres $-r$ i s tals que $x^2 + rx + s$ és un factor quadràtic del polinomi $p(x)$. Resolent l'equació de segon grau $x^2 + rx + s = 0$, obtindrem els zeros buscats. El problema V.8 il·lustra aquestes consideracions.

Mètode de Muller-Traub

Aquest mètode és utilitzable per calcular zeros de qualsevol funció; tot i així, el seu ús més generalitzat correspon al càlcul de zeros de polinomis. Aquest fet ha aconsellat que fos inclòs en aquest apartat.

El mètode de Muller-Traub està basat en l'esquema següent: donades tres aproximacions x_{k-2}, x_{k-1}, x_k d'un zero α d'una funció f que, en el cas que tractem, serà un polinomi, es calcula el polinomi $p_2(x)$, de grau 2, que interpola f en x_{k-2}, x_{k-1}, x_k i es pren x_{k+1} com el zero de $p_2(x)$ més pròxim a x_k .

Així, en el pas k -è, donats x_{k-2}, x_{k-1}, x_k , calculem (vegeu l'apartat 3.1.3):

$$\begin{aligned} f[x_k] &= f(x_k), \\ f[x_{k-1}, x_k] &= \frac{f[x_k] - f[x_{k-1}]}{x_k - x_{k-1}}, \\ f[x_{k-2}, x_{k-1}, x_k] &= \frac{f[x_{k-1}, x_k] - f[x_{k-2}, x_{k-1}]}{x_k - x_{k-2}}, \end{aligned}$$

(observeu que els valors $f[x_{k-1}]$ i $f[x_{k-2}, x_{k-1}]$ són ja coneguts del pas anterior); llavors, el polinomi $p_2(x)$ ve donat per

$$p_2(x) = f[x_k] + f[x_{k-1}, x_k](x - x_k) + f[x_{k-2}, x_{k-1}, x_k](x - x_k)(x - x_{k-1}),$$

que podem expressar com $p_2(x) = a_k(x - x_k)^2 + 2b_k(x - x_k) + c_k$ amb

$$\begin{aligned} a_k &= f[x_{k-2}, x_{k-1}, x_k], \\ b_k &= \frac{1}{2}(f[x_{k-1}, x_k] + f[x_{k-2}, x_{k-1}, x_k](x_k - x_{k-1})), \\ c_k &= f[x_k], \end{aligned}$$

i la fórmula iterativa és

$$x_{k+1} = x_k - \frac{c_k}{b_k \pm \sqrt{b_k^2 - a_k c_k}},$$

amb el signe escollit de manera que el denominador sigui màxim en mòdul.

Convé fer dues observacions:

1. Si el procés iteratiu no pot continuar perquè $a_k = b_k = 0$, cal tornar a començar, partint d'uns altres valors inicials x_0, x_1, x_2 .

2. Encara que es comenci amb valors x_0, x_1, x_2 reals, es poden obtenir, si $b_k^2 < a_k c_k$, iterats complexos. Treballant amb aritmètica complexa, aquest mètode serveix també per calcular zeros no reals de f .

L'ordre d'aquest mètode és almenys $p = 1.84\dots$, essent p l'únic zero positiu del polinomi $x^3 - x^2 - x - 1$. Per tant, si el càlcul de la successió $(x_k)_{k \geq 0}$ és possible i es parteix de valors propers al zero buscat, aquest mètode és convergent. El problema V.6 n'ofereix un bon exemple d'ús.

Mètode de Bairstow

Donat un polinomi $p(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$ (suposem, sense perdre generalitat, que $a_n = 1$), l'objectiu del mètode de Bairstow consisteix a obtenir factors quadràtics de la forma $x^2 + rx + s$ del polinomi $p(x)$. Els zeros d'aquests factors ho seran també de $p(x)$.

En general, donats r i s , podem escriure

$$p(x) = (x^2 + rx + s)(x^{n-2} + b_{n-3}x^{n-3} + \dots + b_0) + lx + m,$$

on l i m són funcions de r i s : $l(r, s), m(r, s)$. En aquests termes, cal trobar r i s solucions del sistema d'equacions no lineals

$$\left. \begin{aligned} l(r, s) &= 0 \\ m(r, s) &= 0 \end{aligned} \right\}.$$

Per obtenir aquests valors serà suficient resoldre el sistema anterior. Un mètode de resolució és el de Newton en dues variables que serà estudiat en l'apartat 5.3.3 i que, en aquest cas, adopta la forma especialment manejable que es descriu tot seguit.

A partir d'una parella adequada de valors inicials r_0, s_0 , es genera una parella de successions $(r_k)_{k \geq 0}, (s_k)_{k \geq 0}$ convergents cap a la parella de valors de r, s buscats.

La construcció d'aquesta successió es realitza de la manera següent:

$$\left. \begin{aligned} r_{k+1} &= r_k + \Delta r_k \\ s_{k+1} &= s_k + \Delta s_k \end{aligned} \right\},$$

essent $\Delta r_k, \Delta s_k$ les solucions del sistema

$$\left. \begin{aligned} c_0 \Delta r_k + c_1 \Delta s_k &= b_{-1} \\ (c_{-1} - b_{-1}) \Delta r_k + c_0 \Delta s_k &= b_{-2} \end{aligned} \right\},$$

on els b_j, c_j es generen per les fórmules recurrents:

$$\begin{aligned} b_{n-1} &= 0, & b_{n-2} &= 1, & b_j &= a_{j+2} - r_k b_{j+1} - s_k b_{j+2} & (j = n-3 \div -2), \\ c_{n-1} &= 0, & c_{n-2} &= 1, & c_j &= b_j - r_k c_{j+1} - s_k c_{j+2} & (j = n-3 \div -1). \end{aligned}$$

Els valors r_0, s_0 s'han d'escollir tan pròxims com sigui possible als valors buscats, d'aquesta elecció dependrà la convergència del mètode. Si, per algun procediment aproximat, s'han obtingut valors reals $\hat{\alpha}_1, \hat{\alpha}_2$, pròxims als zeros α_1, α_2 , convindrà partir de $r_0 = -\hat{\alpha}_1 - \hat{\alpha}_2, s_0 = \hat{\alpha}_1 \hat{\alpha}_2$. Si els valors aproximats de què disposem són dos complexos conjugats $\hat{\alpha} + \hat{\beta}i, \hat{\alpha} - \hat{\beta}i$ convindrà, evidentment, prendre $r_0 = -2\hat{\alpha}$ i $s_0 = \hat{\alpha}^2 + \hat{\beta}^2$. El procés iteratiu s'acabarà quan $\Delta r_k, \Delta s_k$ siguin suficientment petits. Llavors les arrels de l'equació $x^2 + r_k x + s_k = 0$ seran aproximacions de dos zeros del polinomi $p(x)$. En el problema V.8 s'ofereix un exemple d'utilització d'aquest mètode.

Mètode de Graeffe

En l'exposició d'aquest mètode limitarem la nostra atenció al tractament de polinomis amb zeros reals simples. Cal, però, tenir en compte que és igualment útil en el cas més general. Considerem un polinomi de grau n , $p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$, amb zeros $\alpha_1, \alpha_2, \dots, \alpha_n$, que suposarem ordenats segons $|\alpha_1| > |\alpha_2| > \dots > |\alpha_n|$. Si aquests zeros estan *molt separats*, en el sentit que $|\frac{\alpha_k}{\alpha_{k-1}}|$ ($k = 2 \div n$) siguin suficientment petits, llavors

$$\alpha_1 \simeq -\frac{a_{n-1}}{a_n}, \quad \alpha_2 \simeq -\frac{a_{n-2}}{a_{n-1}}, \quad \dots, \quad \alpha_n \simeq -\frac{a_0}{a_1}.$$

Donat un polinomi qualsevol amb tots els zeros diferents en mòdul, aquests no estaran molt separats, en general; però, elevant-los a una potència convenient, es pot aconseguir separar-los tant com es desitgi. Aquesta és la idea central del *mètode de Graeffe*: a partir d'un polinomi $p(x)$ de zeros $\alpha_1, \alpha_2, \dots, \alpha_n$, obtenir un nou polinomi $p_1(x)$ de zeros $\alpha_1^2, \alpha_2^2, \dots, \alpha_n^2$. Aplicant reiteradament aquest mètode s'aniran obtenint polinomis $p_k(x)$ de zeros $\alpha_1^{2^k}, \alpha_2^{2^k}, \dots, \alpha_n^{2^k}$, fins que, considerant-los suficientment separats, puguem calcular

$$\alpha_i^{2^k} \simeq -\frac{a_{n-i}^{(k)}}{a_{n-i+1}^{(k)}} \quad (i = 1 \div n),$$

on els $a_j^{(k)}$ ($j = 0 \div n$) són els coeficients de $p_k(x)$. A partir d'aquí no hi ha cap dificultat en el càlcul de α_i , tenint en compte que la indeterminació del signe es resol substituint $+\alpha_i$ i $-\alpha_i$ en el polinomi original $p(x)$. Així doncs, el nucli del mètode de Graeffe rau en la construcció recurrent del polinomi $p_{k+1}(x)$, de coeficients $a_j^{(k+1)}$, a partir del polinomi $p_k(x)$, de coeficients $a_j^{(k)}$. Aquesta construcció es basa en la relació $p_{k+1}(x^2) = (-1)^n p_k(x) p_k(-x)$, de la qual es desprenen les expressions que ens donen els coeficients de $p_{k+1}(x)$ a partir dels de $p_k(x)$:

$$\begin{aligned} a_0^{(k+1)} &= (-1)^n a_0^{(k)^2}, \\ a_1^{(k+1)} &= (-1)^{n-1} \left(a_1^{(k)^2} - 2a_0^{(k)} a_2^{(k)} \right), \\ a_2^{(k+1)} &= (-1)^{n-2} \left(a_2^{(k)^2} - 2a_1^{(k)} a_3^{(k)} + 2a_0^{(k)} a_4^{(k)} \right), \\ &\vdots \\ a_j^{(k+1)} &= (-1)^{n-j} \left(a_j^{(k)^2} - 2a_{j-1}^{(k)} a_{j+1}^{(k)} + \dots + (-1)^l 2a_{j-l}^{(k)} a_{j+l}^{(k)} \right), \\ &\vdots \\ a_n^{(k+1)} &= a_n^{(k)^2}, \end{aligned}$$

on $l = \min(j, n-j)$ en l'expressió del coeficient $a_j^{(k+1)}$.

5.3 SISTEMES NO LINEALS

5.3.1 Introducció

En aquesta secció s'exposen, per a la resolució de sistemes d'equacions no lineals, dos mètodes numèrics que són generalitzacions de mètodes ja estudiats en el cas d'una equació

no lineal: el mètode d'iteració simple i el de Newton. Per comoditat, treballarem únicament amb sistemes no lineals de dues equacions amb dues incògnites que escriurem com

$$\left. \begin{array}{l} x_1 = g_1(x_1, x_2) \\ x_2 = g_2(x_1, x_2) \end{array} \right\} \quad \circ \quad \left. \begin{array}{l} f_1(x_1, x_2) = 0 \\ f_2(x_1, x_2) = 0 \end{array} \right\} ,$$

segons convingui.

5.3.2 Mètode d'iteració simple en diverses variables

Considerem el sistema

$$\left. \begin{array}{l} x_1 = g_1(x_1, x_2) \\ x_2 = g_2(x_1, x_2) \end{array} \right\}$$

que, usant la notació vectorial

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} , \quad g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \end{pmatrix} = \begin{pmatrix} g_1(x_1, x_2) \\ g_2(x_1, x_2) \end{pmatrix} ,$$

s'expressa en la forma $x = g(x)$.

L'objectiu del *mètode d'iteració simple en diverses variables* rau en la construcció d'una successió de vectors $x^{(0)}, x^{(1)}, x^{(2)}, \dots$ que convergeixi cap a la solució buscada. Partint d'un $x^{(0)}$ adequat, la construcció de l'esmentada successió es fa segons la fórmula recurrent $x^{(k+1)} = g(x^{(k)})$.

Sobre la convergència d'aquest mètode es disposa del següent resultat:

- Sigui R un conjunt tancat qualsevol de $\mathbb{R}^2$; si $g(x) \in R$, per a tot $x \in R$, i existeix una constant real $C < 1$ tal que, per a alguna norma vectorial $\| \cdot \|$, es compleix que

$$\|g(x^{(1)}) - g(x^{(2)})\| \leq C \|x^{(1)} - x^{(2)}\| ,$$

per a $x^{(1)} \in R$ i $x^{(2)} \in R$; aleshores, el sistema $x = g(x)$ té una solució única a R i, per a qualsevol $x^{(0)} \in R$, la successió $x^{(0)}, x^{(1)}, x^{(2)}, \dots$, obtinguda per la recurrència $x^{(k+1)} = g(x^{(k)})$ ($k \geq 0$), convergeix cap a la solució.

Quan una funció g compleix les hipòtesis del teorema anterior, es diu que és *contractiva* (de constant C) dins de R . De manera anàloga al cas d'una variable, una condició suficient per tal que g sigui contractiva en R és que sigui contínuament diferenciable i que la matriu jacobiana $Dg(x)$ de g , compleixi que

$$\|Dg(x)\| \leq C < 1$$

per a tot $x \in R$, essent $\| \cdot \|$ una norma matricial consistent amb la norma vectorial considerada.

5.3.3 Mètode de Newton en diverses variables

Considerem un sistema escrit en la forma

$$\left. \begin{array}{l} f_1(x_1, x_2) = 0 \\ f_2(x_1, x_2) = 0 \end{array} \right\} ,$$

on suposem que les funcions f_1 i f_2 són diferenciables amb continuïtat.

Usant la notació vectorial

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix},$$

el sistema s'expressa en la forma $f(x) = 0$.

Suposarem que la matriu jacobiana és regular en un entorn de la solució,

$$\det(Df(x)) = \det \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) \end{pmatrix} \neq 0.$$

L'objectiu del *mètode de Newton en diverses variables* consisteix a generar una successió de vectors $x^{(0)}, x^{(1)}, x^{(2)}, \dots$, que convergeixi cap a la solució buscada. Partint d'un vector $x^{(0)}$ adequat, la construcció de l'esmentada successió es fa segons la fórmula recurrent

$$x^{(k+1)} = x^{(k)} - (Df(x^{(k)}))^{-1} f(x^{(k)}),$$

que és una generalització a dimensió dos del mètode de Newton per a una equació amb una variable.

El mètode de Newton s'acostuma a formular de la manera següent:

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)},$$

amb $\Delta x^{(k)}$ complint el sistema lineal

$$Df(x^{(k)})\Delta x^{(k)} = -f(x^{(k)}).$$

Així, en comptes de calcular a cada iteració la inversa de la matriu jacobiana, tan sols cal resoldre un sistema lineal.

Quan el càlcul de les derivades parcials de f_1 i de f_2 comporti dificultats (per exemple, quan es desconeguin les expressions analítiques de f_1 i f_2 i es disposi únicament d'un algorisme per al seu càlcul), serà necessari aproximar-les pels seus quocients incrementals:

$$\begin{aligned} \frac{\partial f_1}{\partial x_1}(x_1, x_2) &\simeq \frac{f_1(x_1 + h_1, x_2) - f_1(x_1, x_2)}{h_1}, \\ \frac{\partial f_1}{\partial x_2}(x_1, x_2) &\simeq \frac{f_1(x_1, x_2 + h_2) - f_1(x_1, x_2)}{h_2}, \end{aligned}$$

i, anàlogament, per a la funció f_2 (es poden usar també altres expressions basades en les fórmules de derivació introduïdes en el capítol precedent). Els valors h_1 i h_2 de les expressions anteriors s'han d'escollir de manera convenient. Hi ha una variant del mètode de Newton, que s'anomena *mètode de Steffensen*, que consisteix a prendre $h_1 = f_1(x^{(k)})$ i $h_2 = f_2(x^{(k)})$ per al càlcul aproximat de $Df(x^{(k)})$ que s'ha de portar a terme en el pas k -è del procés iteratiu.

COMENTARIS BIBLIOGRÀFICS

Una de les referències més completes sobre resolució d'equacions no lineals és [Tra64], que inclou a més una àmplia bibliografia. Unes altres referències generals són [M.66], [RR78]. El mètode de Newton és estudiat en pràcticament tots els llibres de càlcul numèric i en altres camps: per a una base teòrica, es pot consultar [Die68], [M.66] i, per a questions pràctiques, [Hen64], [SB80]. Els mètodes de la secant i regula falsi es tracten de manera detallada en [Hil74], [M.66], [RR78]. Per a diversos resultats de convergència global dels mètodes de Newton i d'iteració simple, destaquem [Hen64], [IK66], [M.66]. Per a la generació de mètodes via interpolació, així com per a l'estudi de llurs ordres de convergència, es pot consultar [Dur61], [Hen64], [Hil74], [Hou53], [IK66], [RR78], [M.66] per a la seva acceleració. [CdB72] presenta alguns programes en llenguatge Fortran.

Diverses fites per als zeros dels polinomis es troben en [SB80] i, sobretot, en [Hen74], el qual també estudia molt extensament la localització i aproximació de zeros complexos. El mètode de Laguerre es detalla en [RR78], els de Bernoulli, Q-D i Muller-Traub en [Hen64], i els de Bairstow i Graeffe, en [Hil74]. [IK66], [SB80] i, sobretot, [Wil64] tracten la sensibilitat dels zeros dels polinomis respecte a llurs coeficients.

Per a sistemes d'equacions no lineals, podeu consultar, per exemple, [IK66], [OR70], [M.66], [SB80].

PROBLEMES RESOLTS

Problema 5.1 Sigui f una funció 3 vegades diferenciable amb continuïtat de la qual busquem un zero simple α .

a) Deduïu la fórmula corresponent al mètode iteratiu que obté x_{k+1} avaluant en 0 el polinomi de Taylor de grau més petit o igual que 2 de la funció inversa g de f en $f(x_k)$ (mètode de Txebixev).

b) Proveu que el mètode descrit té ordre almenys 3 per a zeros simples i trobeu la constant asimptòtica de l'error quan l'ordre sigui exactament 3.

SOLUCIÓ:

a) Sigui g la funció inversa de f a la vora de α (existeix ja que α és un zero simple de f).

Seguint l'enunciat, considerem

$$x_{k+1} = \sum_{j=0}^2 \frac{g^{(j)}(f(x_k))}{j!} (-f(x_k))^j .$$

Observem que $g^{(0)}(f(x_k)) = g(f(x_k)) = x_k$, ja que $g(f(x)) = x$.

Derivant dues vegades aquesta última expressió, s'obté

$$g'(f(x))f'(x) = 1 , \quad g^{(2)}(f(x))f'^2(x) + g'(f(x))f^{(2)}(x) = 0 ;$$

i, com a conseqüència, les expressions següents per a les derivades primera i segona de la funció inversa g en $f(x_k)$ en termes de les derivades corresponents de f en x_k :

$$g'(f(x_k)) = \frac{1}{f'(x_k)} , \quad g^{(2)}(f(x_k)) = -\frac{f^{(2)}(x_k)}{f'^3(x_k)} .$$

Per tant, el mètode de Txebixev s'escriu com

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} - \frac{f^{(2)}(x_k)f^2(x_k)}{2f'^3(x_k)} .$$

b) Tot seguit demostrarem que l'ordre de convergència d'aquest mètode, en calcular zeros simples, és almenys 3 i trobarem la constant asimptòtica de l'error en el cas que sigui 3. Considerem l'expressió del zero buscat

$$\begin{aligned} \alpha &= g(f(x_k)) - g'(f(x_k))f(x_k) + \frac{g^{(2)}(f(x_k))}{2}f^2(x_k) \\ &\quad - \frac{g^{(3)}(\xi)}{6}f^3(x_k) , \quad \xi \in \langle 0, f(x_k) \rangle . \end{aligned}$$

Tenint present que els tres primers termes del segon membre constitueixen la fórmula d'iteració, podem escriure

$$x_{k+1} - \alpha = \frac{g^{(3)}(\xi)}{6} f^3(x_k) .$$

Trobem ara les expressions de $f(x_k)$ i de $g^{(3)}(\xi)$ en funció de les derivades de f .

Per una banda,

$$f(x_k) = f(\alpha) + f'(\eta)(x_k - \alpha) = f'(\eta)(x_k - \alpha)$$

amb $\eta \in]\alpha, x_k[$.

Per l'altra, derivant tres vegades l'expressió $g(f(x)) = x$, obtenim que

$$g^{(3)}(f(x))f'^3(x) + 3g^{(2)}(f(x))f'(x)f^{(2)}(x) + g'(f(x))f^{(3)}(x) = 0 .$$

Reemplaçant les expressions de $g'(f(x))$ i de $g^{(2)}(f(x))$ obtingudes anteriorment i aïllant $g^{(3)}(f(x))$, tenim

$$g^{(3)}(f(x)) = \frac{3f^{(2)2}(x) - f'(x)f^{(3)}(x)}{f'^5(x)} .$$

Substituint aquestes expressions en la de $x_{k+1} - \alpha$ i tenint en compte que $\xi = f(g(\xi))$, obtenim

$$x_{k+1} - \alpha = \frac{3f^{(2)2}(g(\xi)) - f'(g(\xi))f^{(3)}(g(\xi))}{6f'^5(g(\xi))} f'^3(\eta)(x_k - \alpha)^3$$

i, per tant,

$$\lim_{x_k \rightarrow \alpha} \frac{x_{k+1} - \alpha}{(x_k - \alpha)^3} = \lim_{x_k \rightarrow \alpha} \frac{3f^{(2)2}(g(\xi)) - f'(g(\xi))f^{(3)}(g(\xi))}{6f'^5(g(\xi))} f'^3(\eta) .$$

Ara bé, quan $x_k \rightarrow \alpha$, tenim $\eta \rightarrow \alpha$, $f(x_k) \rightarrow 0$, $\xi \rightarrow 0$ i $g(\xi) \rightarrow \alpha$; aleshores,

$$\lim_{x_k \rightarrow \alpha} \frac{x_{k+1} - \alpha}{(x_k - \alpha)^3} = \frac{3f^{(2)2}(\alpha) - f'(\alpha)f^{(3)}(\alpha)}{6f'^2(\alpha)} .$$

Amb aquesta darrera expressió queda demostrat que, per al càlcul de zeros simples, l'ordre de convergència del mètode de Txeixev és almenys 3 i que, en el cas que sigui exactament 3, la constant asimptòtica de l'error ve donada per l'expressió

$$\frac{3f^{(2)2}(\alpha) - f'(\alpha)f^{(3)}(\alpha)}{6f'^2(\alpha)} .$$

Problema 5.2 a) Trobeu l'ordre de convergència del mètode de Steffensen

$$x_{k+1} = x_k - \frac{f^2(x_k)}{f(x_k + f(x_k)) - f(x_k)} ,$$

en el càlcul de zeros simples de funcions dues vegades diferenciables amb continuïtat.

b) Aplicació: Comproveu numèricament que, en usar aquest mètode per al càlcul del zero de $f(x) = \cos x - x$, l'ordre de convergència és el trobat en l'apartat anterior.

SOLUCIÓ:

a) Escrivim el mètode de la forma

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

amb

$$g(x_k) = \frac{f(x_k + f(x_k)) - f(x_k)}{f(x_k)} .$$

Noteu que $g(x_k)$ és una aproximació per derivació numèrica de $f'(x_k)$ amb pas $f(x_k)$. Usant el desenvolupament de Taylor de la funció f a la vora de x_k

$$f(x_k + f(x_k)) = f(x_k) + f'(x_k)f(x_k) + \frac{f^{(2)}(\xi_1)}{2}f^2(x_k) ,$$

amb $\xi_1 \in \langle x_k, x_k + f(x_k) \rangle$, s'obté

$$g(x_k) = f'(x_k) + \frac{1}{2}f^{(2)}(\xi_1)f(x_k) .$$

Considerant ara els desenvolupaments de f i f' en un entorn del zero simple buscat α :

$$\begin{aligned} f(x_k) &= f'(\xi_2)(x_k - \alpha) \\ \text{i } f'(x_k) &= f'(\alpha) + f^{(2)}(\xi_3)(x_k - \alpha) , \end{aligned}$$

amb $\xi_2, \xi_3 \in \langle \alpha, x_k \rangle$ i, substituint en l'expressió de $g(x_k)$, obtenim

$$g(x_k) = f'(\alpha) + f^{(2)}(\xi_3)(x_k - \alpha) + \frac{1}{2}f^{(2)}(\xi_1)f'(\xi_2)(x_k - \alpha) .$$

Llavors, tenint present que

$$f(x_k) = f'(\alpha)(x_k - \alpha) + \frac{f^{(2)}(\xi_4)}{2}(x_k - \alpha)^2 ,$$

amb $\xi_4 \in \langle \alpha, x_k \rangle$, podem escriure:

$$\begin{aligned} x_{k+1} - \alpha &= \frac{(x_k - \alpha)g(x_k) - f(x_k)}{g(x_k)} \\ &= \frac{[f^{(2)}(\xi_3) + \frac{1}{2}f^{(2)}(\xi_1)f'(\xi_2)](x_k - \alpha)^2 - \frac{f^{(2)}(\xi_4)}{2}(x_k - \alpha)^2}{f'(\alpha) + [f^{(2)}(\xi_3) + \frac{1}{2}f^{(2)}(\xi_1)f'(\xi_2)](x_k - \alpha)} . \end{aligned}$$

Com que, quan $k \rightarrow \infty$, $x_k \rightarrow \alpha$ i $\xi_i \rightarrow \alpha$ per a $i = 1, 2, 3, 4$, tindrem que

$$\lim_{k \rightarrow \infty} \frac{x_{k+1} - \alpha}{(x_k - \alpha)^2} = \frac{f^{(2)}(\alpha)}{2f'(\alpha)}(1 + f'(\alpha)) .$$

Així doncs, la convergència es quadràtica si $f^{(2)}(\alpha) \neq 0$ i $f'(\alpha) \neq -1$. En cas contrari, es almenys cúbica.

b) Considerem la funció concreta $f(x) = \cos x - x$. Podrem escriure la recurrència donada pel mètode de Steffensen de la forma $x_{k+1} = G(x_k)$ on

$$G(x) = x - \frac{(\cos x - x)^2}{\cos(\cos x) - 2\cos x + x}.$$

Prenent $x_0 = 1$ i aplicant la recurrència descrita, s'obté

$$\begin{aligned} x_1 &= 0.7280103655, \\ x_2 &= 0.7390669669, \\ x_3 &= 0.739085133167, \\ x_4 &= 0.739085133216, \\ x_5 &= 0.739085133216 \simeq \alpha; \end{aligned}$$

de manera que

$$\begin{aligned} e_0 &= 0.260915, \\ e_1 &= -0.01107477, \\ e_2 &= -0.0000181663, \\ e_3 &= -0.000000000049; \end{aligned}$$

i, per tant,

$$\frac{e_1}{e_0^2} = -0.163, \quad \frac{e_2}{e_1^2} = -0.148, \quad \frac{e_3}{e_2^2} = -0.148.$$

Notem que la constant asimptòtica de l'error és

$$\frac{f^{(2)}(\alpha)}{2f'(\alpha)}(1 + f'(\alpha)) = -\frac{\alpha \sin \alpha}{2(1 + \sin \alpha)} = -0.1487371708,$$

valor al qual ha de tendir la successió

$$\left(\frac{e_{k+1}}{e_k^2} \right)_{k \geq 0}.$$

Aquesta tendència mostra la convergència quadràtica del mètode.

Problema 5.3 Demostreu que el mètode de Newton modificat per a zeros múltiples

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

té ordre de convergència 2 en calcular un zero α de multiplicitat m de f sempre que f sigui $m + 1$ vegades diferenciable amb continuïtat i que $f^{(m+1)}(\alpha) \neq 0$.

SOLUCIÓ:

Si f és $m + 1$ vegades diferenciable amb continuïtat, tenint en compte que $f(\alpha) = f'(\alpha) = \dots = f^{(m-1)}(\alpha) = 0$ i $f^{(m)}(\alpha) \neq 0$, resulten les expressions:

$$\begin{aligned} f(x) &= \frac{f^{(m)}(\alpha)}{m!}(x - \alpha)^m + \frac{f^{(m+1)}(\xi_1)}{(m+1)!}(x - \alpha)^{m+1}, \\ f'(x) &= \frac{f^{(m)}(\alpha)}{(m-1)!}(x - \alpha)^{m-1} + \frac{f^{(m+1)}(\xi_2)}{m!}(x - \alpha)^m, \end{aligned}$$

amb $\xi_1, \xi_2 \in \langle x, \alpha \rangle$.

Per tant,

$$x_{k+1} - \alpha = x_k - \alpha - \frac{m \left[\frac{f^{(m)}(\alpha)}{m!}(x_k - \alpha)^m + \frac{f^{(m+1)}(\xi_1)}{(m+1)!}(x_k - \alpha)^{m+1} \right]}{\frac{f^{(m)}(\alpha)}{(m-1)!}(x_k - \alpha)^{m-1} + \frac{f^{(m+1)}(\xi_2)}{m!}(x_k - \alpha)^m},$$

amb $\xi_1, \xi_2 \in \langle x_k, \alpha \rangle$.

Operant en l'expressió anterior s'obté

$$\begin{aligned} \frac{x_{k+1} - \alpha}{(x_k - \alpha)^2} &= \frac{\frac{f^{(m+1)}(\xi_2)}{m!} - \frac{mf^{(m+1)}(\xi_1)}{(m+1)!}}{\frac{f^{(m)}(\alpha)}{(m-1)!} + \frac{f^{(m+1)}(\xi_2)}{m!}(x_k - \alpha)} \\ &= \frac{f^{(m+1)}(\xi_2) - \frac{m}{m+1}f^{(m+1)}(\xi_1)}{mf^{(m)}(\alpha) + f^{(m+1)}(\xi_2)(x_k - \alpha)}. \end{aligned}$$

Quan $k \rightarrow \infty$, $x_k \rightarrow \alpha$ i $\xi_i \rightarrow \alpha$ per a $i = 1, 2$ i, per tant,

$$\lim_{k \rightarrow \infty} \frac{x_{k+1} - \alpha}{(x_k - \alpha)^2} = \frac{f^{(m+1)}(\alpha) - \frac{m}{m+1}f^{(m+1)}(\alpha)}{mf^{(m)}(\alpha)} = \frac{f^{(m+1)}(\alpha)}{m(m+1)f^{(m)}(\alpha)} \neq 0.$$

Així doncs, l'ordre de convergència és 2 i la constant asimptòtica de l'error és

$$\frac{f^{(m+1)}(\alpha)}{m(m+1)f^{(m)}(\alpha)}.$$

Problema 5.4 a) Deduïu la fórmula recurrent del mètode d'interpolació inversa iterada de 3 punts per al càlcul de zeros d'una funció f .

b) Trobeu l'ordre de convergència i la constant asimptòtica de l'error del mètode anterior en calcular zeros simples per al cas més general.

c) Aplicació: Calculeu els zeros de la funció

$$f(x) = \frac{1}{x} - 2 \ln x - 0.5,$$

amb error menor que 10^{-6} , a partir de les aproximacions inicials següents $x_0 = 0.5$, $x_1 = 1$, $x_2 = 1.5$.

SOLUCIÓ:

a) Recordem l'algorisme del mètode d'interpolació inversa iterada de 3 punts per al càlcul de zeros d'una funció f :

- Es parteix de tres aproximacions x_0 , x_1 , x_2 .
- S'interpol·la la funció g , inversa de f , mitjançant un polinomi $q_2(y)$ que passa pels punts $(f(x_0), x_0)$, $(f(x_1), x_1)$, $(f(x_2), x_2)$.
- Es pren $x_3 = q_2(0)$ i es repeteix el procés amb els valors x_1 , x_2 , x_3 com a aproximacions del zero α buscat.

Tot seguit detallem el càlcul de x_{k+1} a partir de x_{k-2} , x_{k-1} , x_k . Per tal de simplificar la notació, escriurem $f_i = f(x_i)$.

Aplicant el mètode de les diferències dividides de Newton en els punts (f_{k-2}, x_{k-2}) , (f_{k-1}, x_{k-1}) , (f_k, x_k) , s'obté el polinomi interpolador de la funció inversa de f

$$q_2(y) = x_k + g[f_k, f_{k-1}](y - f_k) + g[f_k, f_{k-1}, f_{k-2}](y - f_k)(y - f_{k-1}) .$$

Prenent $x_{k+1} = q_2(0)$ i operant, s'obté

$$x_{k+1} = x_k - g[f_k, f_{k-1}]f_k + g[f_k, f_{k-1}, f_{k-2}]f_k f_{k-1} ,$$

on, tal com s'ha explicat en el capítol III, el càlcul de

$$\begin{aligned} g[f_k, f_{k-1}] &= \frac{x_k - x_{k-1}}{f_k - f_{k-1}} , \\ g[f_k, f_{k-1}, f_{k-2}] &= \frac{g[f_k, f_{k-1}] - g[f_{k-1}, f_{k-2}]}{f_k - f_{k-2}} \end{aligned}$$

es fa a través d'un esquema triangular del tipus següent:

$$\begin{array}{c|cc} f_k & x_k & \\ & g[f_k, f_{k-1}] & \\ f_{k-1} & x_{k-1} & g[f_k, f_{k-1}, f_{k-2}] \\ & g[f_{k-1}, f_{k-2}] & \\ f_{k-2} & x_{k-2} & \end{array} .$$

Notem que, havent calculat x_{k+1} , quan es comenci una nova iteració amb x_{k+1} , x_k , x_{k-1} , es podrà aprofitar la part de l'esquema triangular que té com a vèrtex $g[f_k, f_{k-1}]$, i caldrà afegir-hi només una nova filera per sobre.

b) Per tal de trobar l'ordre de convergència i la constant asimptòtica de l'error del mètode descrit a l'apartat anterior, partim de l'expressió de l'error que es produeix en interpolar la funció $g(y)$ pel polinomi $q_2(y)$ (consulteu l'apartat 3.1.2):

$$g(y) - q_2(y) = \frac{g^{(3)}(\eta_k(y))}{3!}(y - f_k)(y - f_{k-1})(y - f_{k-2}) ,$$

on $\eta_k(y) \in (f_k, f_{k-1}, f_{k-2}, y)$.

Si avaluem aquesta expressió per a $y = 0$ i observem que $g(0) = \alpha$ i $q_2(0) = x_{k+1}$, obtenim

$$\alpha - x_{k+1} = -\frac{g^{(3)}(\eta_k)}{3!}f_k f_{k-1} f_{k-2} ,$$

amb $\eta_k = \eta_k(0)$.

Prement $e_i = x_i - \alpha$ i considerant el polinomi interpolador de Taylor de grau més petit o igual que 1 de la funció f en α avaluat en els punts x_k, x_{k-1}, x_{k-2} amb els errors corresponents, s'obté

$$f_i = f'(\alpha)e_i + \frac{f^{(2)}(\xi_i)}{2}e_i^2 ,$$

amb $\xi_i \in (x_i, \alpha)$ ($i = k, k-1, k-2$).

Substituint aquestes expressions en la fórmula de l'error, podem escriure

$$e_{k+1} = \frac{g^{(3)}(\eta_k)}{3!} \prod_{i=k-2}^k \left(f'(\alpha) + \frac{f^{(2)}(\xi_i)}{2}e_i \right) e_i .$$

Quan $k \rightarrow \infty$ i suposada la convergència $x_k \rightarrow \alpha$, tenim $\xi_i \rightarrow \alpha$ i $\eta_k \rightarrow 0$. Per tant,

$$\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k e_{k-1} e_{k-2}} = \frac{g^{(3)}(0)}{6} f'^3(\alpha) .$$

En el problema V.1 es dedueix l'expressió de $g^{(3)}(f(x))$ en funció de les derivades fins a ordre 3 de f en x

$$g^{(3)}(f(x)) = \frac{3f^{(2)2}(x) - f'(x)f^{(3)}(x)}{f'^5(x)} ,$$

que, per a $x = \alpha$, queda

$$g^{(3)}(0) = \frac{3f^{(2)2}(\alpha) - f'(\alpha)f^{(3)}(\alpha)}{f'^5(\alpha)} .$$

Així, podrem escriure

$$\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k e_{k-1} e_{k-2}} = \frac{3f^{(2)2}(\alpha) - f'(\alpha)f^{(3)}(\alpha)}{6f'^2(\alpha)} \equiv K .$$

Si indiquem l'ordre de convergència per p i la constant asimptòtica de l'error per L , tindrem

$$\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k^p} = L .$$

De manera que, per a $k \rightarrow \infty$:

$$e_{k+1} \sim Le_k^p, \quad e_k \sim Le_{k-1}^p, \quad e_{k-1} \sim Le_{k-2}^p;$$

així mateix, per a $k \rightarrow \infty$,

$$e_{k+1} \sim Ke_k e_{k-1} e_{k-2} \quad (k \rightarrow \infty).$$

Posant e_{k+1} , e_{k-1} i e_{k-2} en funció d' e_k i substituint en la darrera expressió, obtindrem

$$Le_k^p \sim K \frac{e_k^{1+\frac{1}{p}+\frac{1}{p^2}}}{L^{\frac{2}{p}+\frac{1}{p^2}}} \quad (k \rightarrow \infty).$$

Així doncs, caldrà que

$$p = 1 + \frac{1}{p} + \frac{1}{p^2}, \quad K = L^{1+\frac{2}{p}+\frac{1}{p^2}}$$

o, si es vol,

$$p^3 - p^2 - p - 1 = 0, \quad L = K^{\frac{p}{p^2+1}}.$$

Per tant, l'ordre de convergència del mètode proposat serà l'únic zero real del polinomi

$$x^3 - x^2 - x - 1,$$

que val, aproximadament, 1.84; la constant asimptòtica de l'error serà llavors

$$L = K^{\frac{p}{p^2+1}}.$$

Finalment, obtenim

$$p \simeq 1.84, \quad L = \left(\frac{3f^{(2)2}(\alpha) - f'(\alpha)f^{(3)}(\alpha)}{6f'^2(\alpha)} \right)^{\frac{p}{p^2+1}}.$$

c) Apliquem ara el mètode d'interpolació inversa iterada al càlcul del zero de la funció $f(x) = \frac{1}{x} - 2\ln x - 0.5$, partint dels valors $x_0 = 0.5$, $x_1 = 1$ i $x_2 = 1.5$. A cada iteració farem el càlcul de $g[f_k, f_{k-1}]$ i de $g[f_k, f_{k-1}, f_{k-2}]$ mitjançant un esquema triangular, tenint ben present que, per a la seva construcció, podrem aprofitar alguns valors de l'esquema triangular corresponent a la iteració anterior. Detallem, a continuació, les successives iteracions.

En la primera iteració, l'esquema triangular de diferències dividides

f_i	x_i	$g[f_i, f_{i+1}]$	$g[f_i, f_{i+1}, f_{i+2}]$
-0.6442635	1.5		
0.5	1	-0.4369623	
2.886294	0.5	-0.2095299	0.0644183

ens permet calcular $x_3 \simeq 1.19773$ amb $f(x_3) = -0.0259436$.

En la segona iteració, l'esquema triangular de diferències dividides

f_i	x_i	$g[f_i, f_{i+1}]$	$g[f_i, f_{i+1}, f_{i+2}]$
-0.0259436	1.19773	-0.4888569	0.0986696
-0.6442635	1.5	-0.4369623	
0.5	1		

ens permet calcular $x_4 \simeq 1.186697$ amb $f(x_4) \simeq 0.0003284$.

En la tercera iteració, l'esquema triangular de diferències dividides

f_i	x_i	$g[f_i, f_{i+1}]$	$g[f_i, f_{i+1}, f_{i+2}]$
0.0003284	1.186697	-0.4199677	0.1068726
-0.0259436	1.19773	-0.4888569	
-0.6442635	1.5		

ens permet calcular $x_5 \simeq 1.186834$ amb $f(x_5) \simeq 3.58 \cdot 10^{-7}$.

Vegem ara que l'error d'aproximació de x_5 és ja més petit que 10^{-6} .

Com que

$$f(x_5) = f'(\xi)(x_5 - \alpha), \quad \xi \in \langle x_5, \alpha \rangle,$$

podem expressar

$$\epsilon = |e_5| = \left| \frac{f(x_5)}{f'(\xi)} \right|.$$

Un breu estudi de la funció f ens permet assegurar que $\alpha \in (1, 2)$ i que, per tant, $\xi \in (1, 2)$ i $1.25 < |f'(\xi)| < 3$.

En resulta, aleshores,

$$\epsilon_5 < \frac{3.58 \cdot 10^{-7}}{1.25} < 10^{-6}.$$

Així doncs, el zero buscat, amb error més petit que 10^{-6} , és

$$\alpha \simeq 1.186834.$$

Problema 5.5 Compareu, segons la seva eficiència, els mètodes de Newton i de la secant en calcular un zero simple α d'una funció f tal que $f^{(2)}(\alpha) \neq 0$.

SOLUCIÓ:

Volem calcular un zero d'una funció f amb una aproximació ϵ usant els mètodes de Newton i de la secant a partir d'un valor aproximat amb error ϵ_0 . Considerarem que el cost d'una avaluació de la funció f és la unitat i indicarem per c el cost d'una avaluació de la funció derivada f' . Prendrem com a nuls els costos corresponents a les operacions aritmètiques enfront dels d'avaluació de les funcions en aplicar les fórmules iteratives.

Per a aplicar la fórmula iterativa del mètode de Newton cal avaluar f i la seva derivada f' ; en canvi, per al mètode de la secant, només cal avaluar la funció f .

De les característiques de tots dos mètodes, es dedueix que:

- N iteracions de Newton tenen un cost de $N(1+c)$ i assolixen una precisió aproximada de $\epsilon_0^{2^N}$, ja que el mètode de Newton té ordre 2 en aquest cas.
- S iteracions de la secant tenen un cost de S i assolixen una precisió aproximada de $\epsilon_0^{p^S}$, on $p = \frac{1+\sqrt{5}}{2}$ és l'ordre del mètode de la secant en aquest cas.

Noteu que les constants asimptòtiques s'han considerat iguals a 1; el resultat seria essencialment el mateix si no es fes aquesta suposició.

Per obtenir, aproximadament, un mateix error ϵ , caldrà que es compleixi $N \ln 2 \simeq S \ln p$. Així doncs, el mètode de Newton serà més eficient que el de la secant quan el cost total de l'aplicació d'aquell sigui inferior al de l'aplicació d'aquest. Això passarà per als valors de c que compleixin:

$$N(1+c) \leq S \simeq N \frac{\ln 2}{\ln p},$$

o sigui, $c \leq c_0 = \frac{\ln 2}{\ln p} - 1 \simeq 0.44$.

Per tant, podem afirmar que el mètode de Newton és més eficient que el mètode de la secant si el cost d'una avaluació de f' és menor que c_0 vegades el cost d'una avaluació de f .

Problema 5.6 Calculeu, pel mètode de Muller-Traub, tots els zeros del polinomi $p(x) = 128x^4 - 256x^3 + 160x^2 - 32x + 1$.

SOLUCIÓ:

Tal com s'ha exposat en la part teòrica, el mètode de Muller-Traub es basa en la construcció d'una successió de valors $x_0, x_1, x_2, \dots$ que tendeixi cap al zero buscat α . Aquesta construcció es porta a terme fent els passos següents:

- Es parteix de tres abscisses (per exemple, $x_0 = -1$, $x_1 = 1$, $x_2 = 0$).
- Es construeix el polinomi de grau més petit o igual que 2 que passa per $(x_0, p(x_0))$, $(x_1, p(x_1))$, $(x_2, p(x_2))$.
- Es pren x_3 com el zero d'aquest polinomi que sigui més pròxim a x_2 .
- El càlcul de x_4 es fa de manera similar però partint dels valors x_1, x_2, x_3 .
- Els altres termes de la successió es van calculant de la mateixa manera, partint sempre de les tres darreres abscisses calculades.
- El procés s'interromp quan dos termes consecutius de la successió disten menys d'un error admissible prefixat que, en el nostre cas, establirem en 10^{-7} .

Un estudi del polinomi $p'(x)$ en un entorn de cadascun dels zeros buscats ens indica que $1 < |p'(\xi)|$. Així, tenint en compte que

$$|x_k - \alpha| = \left| \frac{p(x_k)}{p'(\xi)} \right| ,$$

amb $\xi \in \langle x_k, \alpha \rangle$, $|p(x_k)|$ és una fita de l'error de l'aproximació x_k . Usarem aquesta fita per tal d'estimar la precisió dels resultats α_i que obtinguem.

Aplicant l'algorisme de Muller-Traub al polinomi $p(x)$ obtenim la successió de valors:

$$\begin{aligned} x_0 &= -1 , \\ x_1 &= 1 , \\ x_2 &= 0 , \\ x_3 &= 0.003484363 , \\ x_4 &= 0.032763594 , \\ x_5 &= 0.038144149 , \\ x_6 &= 0.038060056 , \\ x_7 &= 0.038060234 , \\ x_8 &= 0.038060234 . \end{aligned}$$

Així, $\alpha_1 \simeq 0.038060234$, amb $|p(\alpha_1)| \simeq 5.3 \cdot 10^{-9}$.

Conegut el zero α_1 , podem procedir, per la regla de Horner, a una deflació del polinomi $p(x)$

$$p(x) = (x - \alpha_1)p_1(x) + p(\alpha_1) ,$$

amb

$$p_1(x) = 128x^3 - 251.12829x^2 + 150.442x - 26.274142 .$$

En fer una nova iteració de Muller-Traub sobre $p_1(x)$ obtenim:

$$\begin{aligned}x_0 &= 1 , \\x_1 &= 0.5 , \\x_2 &= 0 , \\x_3 &= 0.43103863 , \\x_4 &= 0.33287561 , \\x_5 &= 0.31182922 , \\x_6 &= 0.30862015 , \\x_7 &= 0.30865827 , \\x_8 &= 0.30865826 .\end{aligned}$$

El valor $\tilde{\alpha}_2 = x_8$ és un zero aproximat del polinomi $p_1(x)$ els coeficients del qual estan afectats d'un error degut al procés de deflació de $p(x)$ que s'ha realitzat amb un zero α_1 que no era exacte. Per aquesta raó convé purificar $\tilde{\alpha}_2$, aplicant el mètode de Muller-Traub al polinomi inicial $p(x)$, prenent com a valors de partida els x_6, x_7 i $\tilde{\alpha}_2 = x_8$ calculats en el procés que s'ha portat a terme sobre $p_1(x)$.

Obtenim $\alpha_2 \simeq 0.30865828$, amb $|p(\alpha_2)| \simeq 3.3 \cdot 10^{-8}$.

Procedint a una nova deflació

$$p_1(x) = (x - \alpha_2)p_2(x) + p_1(\alpha_2) ,$$

s'obté

$$p_2(x) = 128x^2 - 211.620030x + 85.123725 .$$

Com que és de segon grau, el càlcul dels zeros de $p_2(x)$ no té cap dificultat. Es troben: $\tilde{\alpha}_3 = 0.69134177$ i $\tilde{\alpha}_4 = 0.96193972$.

Novament convindrà purificar $\tilde{\alpha}_3$ i $\tilde{\alpha}_4$ dels errors que s'hagin pogut acumular al llarg de les dues deflacions realitzades. Igual que hem fet per a $\tilde{\alpha}_2$, aquest refinament es porta a terme aplicant el mètode de Muller-Traub al polinomi inicial $p(x)$. S'obté: $\alpha_3 = 0.69134172$, amb $|p(\alpha_3)| \simeq 3.3 \cdot 10^{-8}$ i $\alpha_4 = 0.96193977$, amb $|p(\alpha_4)| \simeq 7.8 \cdot 10^{-8}$.

Problema 5.7 a) Indiqueu com es pot estendre el mètode de Sturm al cas d'un polinomi real amb zeros reals múltiples.

b) Aplicació: Separeu els zeros reals de

$$p(x) = x^6 - 2x^5 + x^4 - 4x^2 + 8x - 4 .$$

SOLUCIÓ:

a) Donat un polinomi real $p(x) = a_n x^n + \dots + a_0$, construïm la successió

$$\begin{aligned} p_0(x) &= p(x) , \\ p_1(x) &= p'(x) , \\ p_{i-1}(x) &= q_i(x)p_i(x) - c_i p_{i+1}(x) \quad (i = 1 \div m) , \end{aligned}$$

amb l'ajuda de l'algorisme d'Euclides.

Aquesta successió acaba quan $p_{m+1}(x) \equiv 0$. Clarament $m \leq n$, ja que $p_m(x) = \text{m.c.d.}(p(x), p'(x))$.

Si tots els zeros reals de $p(x)$ són simples, llavors $\{p_0(x), \dots, p_m(x)\}$ és una successió de Sturm per a $p(x)$ sobre qualsevol interval de la recta real i es compleix, pel teorema de Sturm, que si $p(a)p(b) \neq 0$, el nombre de zeros reals de $p_0(x)$ en (a, b) és $V(a) - V(b)$, essent $V(x)$ el nombre de canvis de signe en la successió $\{p_0(x), \dots, p_m(x)\}$.

Si $p(x)$ té zeros reals múltiples, llavors $p_m(x)$ s'anul·la en alguns punts de la recta real. Com que $p_m(x)$ divideix $p_0(x)$ i $p_1(x)$, és fàcil veure per inducció que divideix tots els polinomis $p_i(x)$ ($i = 0 \div m$). Si considerem els polinomis

$$f_i(x) = \frac{p_i(x)}{p_m(x)} \quad (i = 0 \div m) ,$$

resulta que $\{f_0(x), \dots, f_m(x)\}$ és una successió de Sturm per a $f_0(x)$ sobre qualsevol interval de la recta real ja que les condicions 1, 2 i 4 són immediates (vegeu l'apartat 5.2.4) i, per tal de demostrar la condició 3, tan sols hem de derivar

$$p_0(x) = f_0(x)p_m(x)$$

per obtenir

$$p'_0(x) = f'_0(x)p_m(x) + f_0(x)p'_m(x) .$$

Si $f_0(\alpha) = 0$, llavors

$$\begin{aligned} p'_0(\alpha) &= f'_0(\alpha)p_m(\alpha) , \\ f_1(\alpha)f'_0(\alpha) &= \frac{p_1(\alpha)}{p_m(\alpha)} \frac{p'_0(\alpha)}{p_m(\alpha)} = \left(\frac{p'_0(\alpha)}{p_m(\alpha)} \right)^2 > 0 , \end{aligned}$$

ja que, si $p'_0(\alpha) = 0$, també passa que $p_m(\alpha) = 0$ i, en la fracció, s'eliminen els factors anul·ladors.

Pel teorema de Sturm, si $f_0(a)f_0(b) \neq 0$, el nombre de zeros reals de $f_0(x)$ en (a, b) és igual a $W(a) - W(b)$, essent $W(x)$ el nombre de canvis de signe de la successió

$$\{f_0(x), \dots, f_m(x)\} .$$

Noteu que, si $p(x) \neq 0$, llavors $p_m(x) \neq 0$ i $V(x) = W(x)$. D'altra banda, com que $f_0(x)$ té els mateixos zeros que $p_0(x)$, però simples, el nombre de zeros reals de $f_0(x)$ en (a, b) és igual al nombre de zeros reals diferents del polinomi $p_0(x)$ en (a, b) , si $p_0(a)p_0(b) \neq 0$. Així doncs, es compleix el resultat següent:

- Si $p(a)p(b) \neq 0$, el nombre de zeros reals diferents de $p(x)$ en (a, b) és igual a $V(a) - V(b)$.

b) Tot seguit apliquem aquest resultat al polinomi

$$p(x) = x^6 - 2x^5 + x^4 - 4x^2 + 8x - 4 .$$

En primer lloc construïm la successió de polinomis associada. Comencem amb $p_0(x) = p(x)$, $p_1(x) = p'(x)$. Per al càlcul de $p_2(x)$, dividim $p_0(x)$ entre $p_1(x)$, obtenint

$$p_0(x) = \left(\frac{1}{6}x - \frac{1}{18}\right)p_1(x) - \frac{2}{9}x^4 + \frac{2}{9}x^3 - \frac{24}{9}x^2 + \frac{56}{9}x - \frac{32}{9} .$$

Fent $p_2(x) = x^4 - x^3 + 12x^2 - 28x + 16$, resulta que

$$p_0(x) = \left(\frac{1}{6}x - \frac{1}{18}\right)p_1(x) - \frac{2}{9}p_2(x) .$$

El procés es pot continuar:

- Dividint $p_1(x)$ entre $p_2(x)$, s'obté

$$p_1(x) = (6x - 4)p_2(x) - 72p_3(x) ,$$

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

- Dividint $p_2(x)$ entre $p_3(x)$, s'obté

$$p_2(x) = (x + 2)p_3(x) - 3p_4(x) ,$$

$$\text{amb } p_4(x) = -5x^2 + 11x - 6 .$$

- Dividint $p_3(x)$ entre $p_4(x)$, s'obté

$$p_3(x) = \left(-\frac{x}{5} + \frac{4}{25}\right)p_4(x) - \frac{1}{25}p_5(x) ,$$

$$\text{amb } p_5(x) = -x + 1 .$$

- Dividint $p_4(x)$ entre $p_5(x)$, s'obté

$$p_4(x) = (5x - 6)p_5(x)$$

i s'acaba el procés.

Observeu que, usant les relacions de recurrència entre els polinomis $p_i(x)$, per calcular $\{p_0(x), \dots, p_5(x)\}$ per a un valor x donat, fan falta tan sols 10 sumes i 13 multiplicacions.

Com que $p_5(x) = -x + 1 = \text{m.c.d.}(p(x), p'(x))$, resulta que 1 és un zero doble de $p(x)$. Per tal de localitzar els altres zeros reals estudiem els signes de la successió de polinomis per a alguns valors x . S'obté la taula següent:

x	$-\infty$	-2	-1	0	1	$1.192\dots$	1.25	1.5	2	$+\infty$
$p_0(x)$	+	+	-	-	0	-	-	+	+	+
$p_1(x)$	-	-	-	+	0	-	-	+	+	+
$p_2(x)$	+	+	+	+	0	0	+	+	+	+
$p_3(x)$	-	-	-	-	0	+	+	+	+	+
$p_4(x)$	-	-	-	-	0	+	-	-	-	-
$p_5(x)$	+	+	+	+	0	-	-	-	-	-
$V(x)$	4	4	3	3	-	2	2	1	1	1

HISTÒRIA DE LA TAULA:

Hem començat amb $-\infty, +\infty$ i, com que $V(-\infty) - V(+\infty) = 3$, hi ha tres zeros reals diferents, un dels quals ja sabem que és $x = 1$. Hem assajat després amb 0 i, com que $V(-\infty) - V(0) = 1$, hi ha un zero real en l'interval $(-\infty, 0)$. Provant amb -1 i -2 , observem que aquest zero real negatiu estarà en $(-2, -1)$, ja que $V(-2) - V(-1) = 1$. Temptejant amb 2, 1.5 i 1.25, detectem un zero real en $(1.25, 1.5)$. Així queden separats els tres zeros reals.

Cal fer dues observacions:

1. En ser $x = 1$ zero múltiple, es compleix que $p_i(1) = 0$ ($i = 0 \div 5$).
2. Si provem amb $\bar{x} = 1.192143\dots$, per al qual $p_2(x)$ s'anul·la, apareix un 0 en la columna de signes. En aquests casos s'ha d'ignorar aquest valor i calcular el nombre de canvis de signe segons els elements anterior i posterior. Així, si tenim $-, 0, +$, hi ha un canvi de signe, però, si tenim $-, 0, -$, no hi ha cap canvi de signe.

Problema 5.8 Volem calcular tots els zeros del polinomi

$$p(x) = x^4 - 2x^3 - 14x^2 - 2x - 15.$$

Useu el mètode Q-D per a obtenir-ne una primera aproximació i els mètodes de Birge-Vieta i Bairstow per precisar-ne els resultats.

SOLUCIÓ:

La resolució d'aquest exercici es fa en tres apartats.

a) APROXIMACIÓ DELS ZEROS PEL MÈTODE Q-D

L'algorisme descrit en la part teòrica del capítol ens permet construir la taula següent:

	$q_k^{(1)}$	$d_k^{(1)}$	$q_k^{(2)}$	$d_k^{(2)}$	$q_k^{(3)}$	$d_k^{(3)}$	$q_k^{(4)}$	
	2		0		0		0	
0		7		0.1429		7.5		0
	9		-6.8571		7.3571		-7.5	
0		-5.3333		-0.1533		-7.6456		0
	3.6667		-1.6771		-0.1352		0.1456	
0		2.4394		-0.0124		8.2347		0
	6.1061		-4.1288		8.1118		-8.0890	
0		-1.6495		0.0243		-8.2115		0
	4.4566		-2.4551		-0.1240		0.1225	
0		0.9087		0.0012		8.1120		0
	5.3653		-3.3625		7.9867		-7.9894	
0		-0.5695		-0.0029		-8.1147		0

En aquesta taula podem observar que els valors de les columnes $d^{(1)}$ i $d^{(2)}$ es van fent petits, la qual cosa indica que les successions $(q_k^{(1)})_{k \geq 1}$ i $(q_k^{(2)})_{k \geq 0}$ tendeixen cap a zeros reals de $p(x)$. Podem considerar els valors 5.3653 i -3.3625 com a aproximacions d'aquests zeros.

El comportament oscil·lant de la columna $d^{(3)}$ ens indica l'existència d'un parell de zeros complexos conjugats. El càlcul d'aquests zeros es fa a partir de la determinació del corresponent factor quadràtic de $p(x)$. Aquest factor és del tipus $x^2 + rx + s$, on

$$r = -\lim_{k \rightarrow \infty} (q_{k+1}^{(3)} + q_k^{(4)}) ,$$

$$s = \lim_{k \rightarrow \infty} q_k^{(3)} q_k^{(4)} ,$$

i es calcularà pel mètode de Bairstow, a partir d'un factor amb valors aproximats de r i s .

Amb les dades que apareixen a la taula, observem que:

- Els primers termes de la successió $(q_{k+1}^{(3)} + q_k^{(4)})_{k \geq -2}$ són:

$$-0.142857, 0.010417, 0.0228, -0.0015, -0.0027, \dots$$

- Els primers termes de la successió $(q_k^{(3)} q_k^{(4)})_{k \geq -2}$ són:

$$0, 1.0714, 1.0938, 0.9938, 0.9908, \dots$$

Així, doncs, podrem prendre com a aproximacions de r i de s els valors 0.00273 i 0.99081, respectivament.

b) REFINAMENT DELS ZEROS REALS PEL MÈTODE DE BIRGE-VIETA

Fent servir la iteració

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

a partir dels valors aproximats calculats pel mètode Q-D, obtenim:

$$\begin{array}{l|l} x_0 = 5.3652 & x_0 = -3.362527 \\ x_1 = 5.054466 & x_1 = -3.070510 \\ x_2 = 5.001458 & x_2 = -3.003374 \\ x_3 = 5.000001 & x_3 = -3.000008 \\ x_4 = 5.000000 & x_4 = -3.000000 . \end{array}$$

Per tant, els dos zeros reals de $p(x)$ són 5 i -3 .

Ara podríem aplicar el mètode de deflació a $p(x)$ i obtenir, sense gaire dificultat, els zeros restants, però no ho farem per tal de mostrar un exemple d'aplicació del mètode de Bairstow.

c) CÀLCUL DELS ZEROS COMPLEXOS PEL MÈTODE DE BAIRSTOW

Partint del factor aproximat $x^2 + 0.0027x + 0.9908$ i usant l'algorisme de Bairstow, descrit en l'apartat 5.2.5, tenim $r_0 = 0.0027$ i $s_0 = 0.9908$.

En la primera iteració, els valors:

$$\begin{aligned} b_1 &= -2.0027, & c_1 &= -2.00546, \\ b_0 &= -14.98534255, & c_0 &= -15.97067764, \\ b_{-1} &= 0.025234896, & c_{-1} &= 2.055864669, \\ b_{-2} &= -0.152441640, \end{aligned}$$

ens permeten construir el sistema

$$\left. \begin{aligned} -15.97067764 \quad \Delta r_0 &- & 2.00546 \quad \Delta s_0 &= & 0.025234896 \\ 2.030629773 \quad \Delta r_0 &- & 15.97067764 \quad \Delta s_0 &= & -0.152441640 \end{aligned} \right\},$$

la solució del qual és

$$\begin{aligned} \Delta r_0 &= -2.735 \cdot 10^{-3}, \\ \Delta s_0 &= 9.197347 \cdot 10^{-3}; \end{aligned}$$

de manera que $r_1 = r_0 + \Delta r_0 = -5 \cdot 10^{-6}$ i $s_1 = s_0 + \Delta s_0 = 1.000007347$.

En la segona iteració, els valors

$$\begin{aligned} b_1 &= -1.999995, & c_1 &= -1.99999, \\ b_0 &= -15.00001735, & c_0 &= -16.00003470, \\ b_{-1} &= -6.5306 \cdot 10^{-5}, & c_{-1} &= 1.999859388, \\ b_{-2} &= 1.2755967 \cdot 10^{-4}, \end{aligned}$$

ens permeten construir el sistema

$$\left. \begin{aligned} -16.00003470 \quad \Delta r_1 &- & 1.99999 \quad \Delta s_1 &= & -6.5306 \cdot 10^{-5} \\ 1.999924694 \quad \Delta r_1 &- & 16.00003470 \quad \Delta s_1 &= & 1.2755967 \cdot 10^{-4} \end{aligned} \right\},$$

la solució del qual és

$$\begin{aligned} \Delta r_1 &= 5 \cdot 10^{-6}, \\ \Delta s_1 &= -7.3474 \cdot 10^{-6}; \end{aligned}$$

de manera que $r_2 = r_1 + \Delta r_1 = 0$ i $s_2 = s_1 + \Delta s_1 = 1$.

En fer una tercera iteració s'obté $\Delta r_2 = \Delta s_2 = 0$, de manera que els valors definitius de r i s són, respectivament, 0 i 1. El factor quadràtic buscat és $x^2 + 1$, que té per zeros i , $-i$.

Així, els zeros del polinomi

$$p(x) = x^4 - 2x^3 - 14x^2 - 2x - 15 = (x - 5)(x + 3)(x^2 + 1)$$

són 5, -3, i , $-i$.

Problema 5.9 a) Siguin $p(x)$ i $q(x)$ dos polinomis i ϵ un número real, petit en valor absolut. Com es poden aproximar els zeros de $\bar{p}(x) = p(x) + \epsilon q(x)$, coneguts els de $p(x)$ i suposats aquests simples o dobles?

b) Apliqueu la resposta a la pregunta anterior al càlcul aproximat dels zeros del polinomi $\bar{p}(x) = p(x) + \epsilon q(x)$, amb

$$p(x) = x^3 - 9x^2 + 15x + 25, \quad q(x) = -x^2, \quad \epsilon = 10^{-6}.$$

SOLUCIÓ:

a) Observeu que la qüestió que es planteja és com queden pertorbats els zeros d'un polinomi en pertorbar lleugerament els seus coeficients.

Sigui α zero de $p(x)$ i $\bar{\alpha}$ zero de $\bar{p}(x)$ i suposem que $\bar{\alpha}$ prové de l'evolució seguida per α en augmentar el paràmetre ϵ , partint de 0. Considerarem sempre que $q(\alpha) \neq 0$ ja que sinó α seria també zero de $\bar{p}(x)$ i no estaria afectat per la pertorbació ($\bar{\alpha} = \alpha$). Sigui $\delta = \bar{\alpha} - \alpha$, ens proposem aproximar $\delta(\epsilon)$ del qual tan sols coneixem que $\delta(0) = 0$.

Considerem el desenvolupament de Taylor de $\bar{p}(x)$ a la vora de α , avaluat a $\bar{\alpha}$

$$0 = \bar{p}(\bar{\alpha}) = \bar{p}(\alpha + \delta) = \bar{p}(\alpha) + \bar{p}'(\alpha)\delta + \frac{\bar{p}^{(2)}(\alpha)}{2}\delta^2 + \frac{\bar{p}^{(3)}(\alpha)}{3!}\delta^3 + \dots,$$

i substituïm $\bar{p}(x)$ per $p(x) + \epsilon q(x)$

$$\begin{aligned} 0 &= \epsilon q(\alpha) + (p'(\alpha) + \epsilon q'(\alpha))\delta + (p^{(2)}(\alpha) + \epsilon q^{(2)}(\alpha))\frac{\delta^2}{2} \\ &\quad + (p^{(3)}(\alpha) + \epsilon q^{(3)}(\alpha))\frac{\delta^3}{3!} + \dots \end{aligned}$$

Ara convé distingir el cas en què α és un zero simple de $p(x)$ del cas en què és doble.

CAS DE ZERO SIMPLE

Per obtenir una aproximació de δ , prenem els dos primers termes del desenvolupament de Taylor,

$$0 \simeq \epsilon q(\alpha) + (p'(\alpha) + \epsilon q'(\alpha))\delta,$$

i aïllem δ :

$$\delta \simeq -\frac{\epsilon q(\alpha)}{p'(\alpha) + \epsilon q'(\alpha)} = -\frac{\epsilon q(\alpha)}{p'(\alpha)} \left(1 + \epsilon \frac{q'(\alpha)}{p'(\alpha)}\right)^{-1} = -\epsilon \frac{q(\alpha)}{p'(\alpha)} + \mathcal{O}(\epsilon^2).$$

Si haguéssim agafat fins a termes en δ^2 en el desenvolupament de Taylor, hauríem obtingut el mateix resultat, una vegada desenvolupada l'expressió de δ .

Si el grau de $p(x)$ és 1 i el de $q(x)$ és 0, l'expressió obtinguda de δ és exacta i s'il·lustra en la figura 5.6. En aquesta figura, com que

$$\tan \varphi = p'(\alpha) = \frac{\epsilon q(\alpha)}{-\delta},$$

resulta que

$$\delta = -\frac{\epsilon q(\alpha)}{p'(\alpha)},$$

com estava previst.

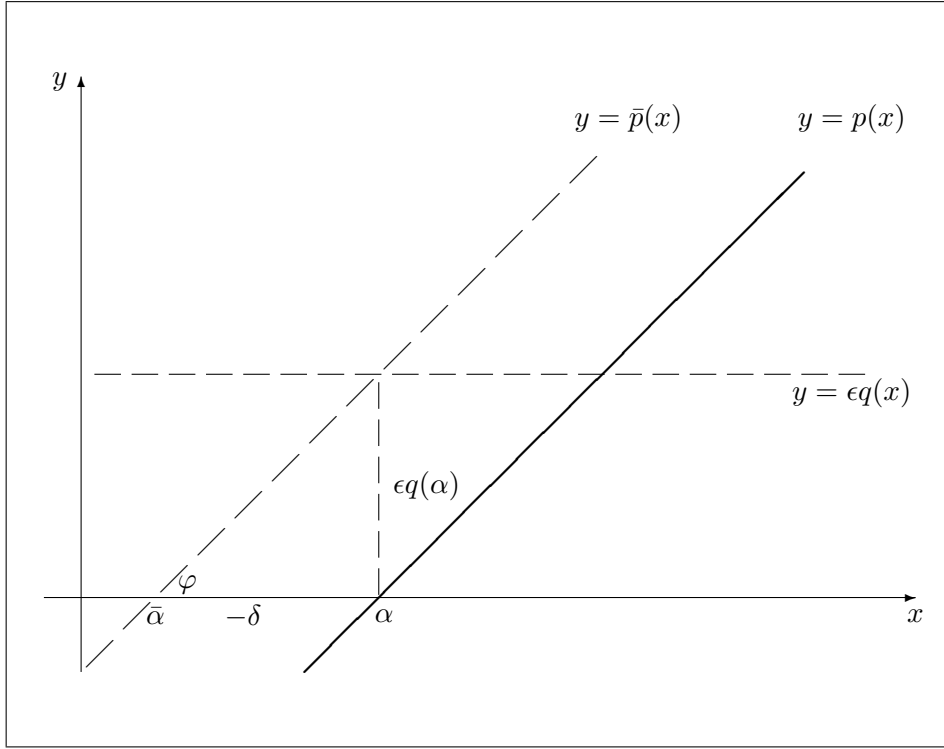


Figura 5.6: Funció lineal desplaçada per constant.

CAS DE ZERO DOBLE

Si, en aquest cas, agaféssim solament els dos primers termes del desenvolupament de Taylor, no s'aportaria cap més dada per a la determinació de δ que l'anul·lació de $p(\alpha)$ i de $p'(\alpha)$, de manera que actuaríem com si el polinomi $p(x)$ fos idènticament nul i, en realitat, estaríem cercant un zero de $q(x)$: δ sortiria igual que l'increment que obtindríem en aplicar el mètode de Newton a $q(x)$ a partir de α . Per tal d'evitar aquesta pèrdua d'informació sobre $p(x)$, agafarem els tres primers termes del desenvolupament

$$0 \simeq \epsilon q(\alpha) + \epsilon q'(\alpha)\delta + (p^{(2)}(\alpha) + \epsilon q^{(2)}(\alpha))\frac{\delta^2}{2}.$$

Així doncs,

$$\delta \simeq \frac{-\epsilon q'(\alpha) \pm H(\alpha)}{p^{(2)}(\alpha) + \epsilon q^{(2)}(\alpha)},$$

amb

$$H(\alpha) = (\epsilon^2 q'^2(\alpha) - 2\epsilon q(\alpha)(p^{(2)}(\alpha) + \epsilon q^{(2)}(\alpha)))^{\frac{1}{2}}.$$

Usant desenvolupaments de Taylor, trobem les expressions

$$H(\alpha) = \sqrt{-2q(\alpha)p^{(2)}(\alpha)}\epsilon^{\frac{1}{2}} + \mathcal{O}(\epsilon^{\frac{3}{2}}),$$

$$\frac{1}{p^{(2)}(\alpha) + \epsilon q^{(2)}(\alpha)} = \frac{1}{p^{(2)}(\alpha)} - \epsilon \frac{q^{(2)}(\alpha)}{p^{(2)2}(\alpha)} + \mathcal{O}(\epsilon^2);$$

substituint-les en l'expressió de δ , tenim

$$\delta = \pm \sqrt{\frac{-2q(\alpha)}{p^{(2)}(\alpha)}} \epsilon^{\frac{1}{2}} - \frac{q'(\alpha)}{p^{(2)}(\alpha)} \epsilon + \mathcal{O}(\epsilon^{\frac{3}{2}}) .$$

Així, prenent els termes dominants,

$$\delta \simeq \pm \sqrt{\frac{-2q(\alpha)}{p^{(2)}(\alpha)}} \epsilon^{\frac{1}{2}} - \frac{q'(\alpha)}{p^{(2)}(\alpha)} \epsilon .$$

Aquí convé fer tres comentaris:

- El doble signe de δ indica que el zero doble α de $p(x)$, en ésser pertorbat, evoluciona (*es bifurca*) cap a dos zeros.
- Considerem el cas en què el grau de $p(x)$ sigui 2 i el de $q(x)$ sigui 0. Llavors l'expressió trobada per a δ és exacta i podem fer un estudi d'allò que passa:
 - $p(x)$ es representa per una paràbola amb un punt de contacte α amb l'eix d'abscisses, la concavitat o convexitat de la qual ve donada pel signe de la constant $p^{(2)}(x)$.
 - $q(x)$ és una constant que gràficament pertorba $p(x)$, traslladant la paràbola en sentit vertical segons els signes de $q(x)$ i de ϵ .

Si suposem $\epsilon > 0$ (per a $\epsilon < 0$ seria anàleg), podem descriure en el quadre següent el comportament dels zeros:

signe de $p^{(2)}(\alpha)$	signe de $q(\alpha) = C$	δ és ...	α es bifurca a dos zeros ...	Figura
+	+	imaginària	complexos	a)
+	–	real	reals	b)
–	+	real	reals	c)
–	–	imaginària	complexos	d)

- Donat el cas que α fos un zero d'ordre k de $p(x)$, es podria fer un raonament semblant i s'obtingria una expressió aproximada del tipus

$$0 \simeq \epsilon p_k(\alpha + \delta) + \frac{\delta^k p^{(k)}(\alpha)}{k!} ,$$

on $p_k(x)$ és el polinomi de Taylor de grau k corresponent a la funció $q(x)$ en un entorn de α :

$$p_k(x) = q(\alpha) + q'(\alpha)(x - \alpha) + \dots + \frac{q^{(k)}(\alpha)}{k!}(x - \alpha)^k .$$

En aquest cas, una aproximació de la δ buscada, vindria donada per

$$\delta = \sqrt[k]{\frac{-k!q(\alpha)}{p^{(k)}(\alpha)}} \epsilon^{\frac{1}{k}} + \mathcal{O}(\epsilon^{\frac{2}{k}}) .$$

Notem que, per obtenir δ amb la mateixa precisió que en el cas de zeros simples, hauríem de considerar k termes en el seu desenvolupament.

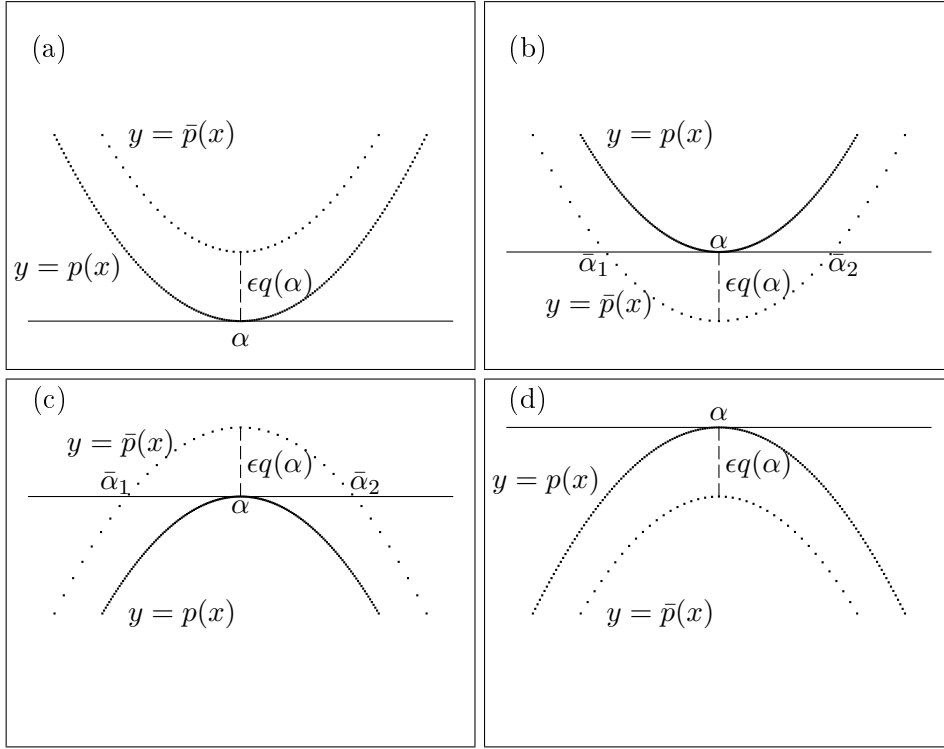


Figura 5.7: Funció quadràtica desplaçada per una constant.

b) Apliquem ara l'estudi anterior al càlcul aproximat dels zeros del polinomi $\bar{p}(x) = p(x) + \epsilon q(x)$, amb $p(x) = x^3 - 9x^2 + 15x + 25$, $q(x) = -x^2$ i $\epsilon = 10^{-6}$.

Obtenim

$$p(x) = (x+1)(x-5)^2, \quad \bar{p}(x) = x^3 - (9+10^{-6})x^2 + 15x + 25.$$

El zero $\alpha_1 = -1$ de $p(x)$ passa a un zero $\bar{\alpha}_1 = -1 + \delta_1$ de $\bar{p}(x)$ i el zero doble $\alpha_2 = 5$ de $p(x)$ es bifurca en dos zeros $\bar{\alpha}_2 = 5 + \delta_2$ i $\bar{\alpha}_3 = 5 + \delta_3$, on:

$$\delta_1 \simeq -\epsilon \frac{q(\alpha_1)}{p'(\alpha_1)} = -10^{-6} \frac{q(-1)}{p'(-1)} = 2.778 \cdot 10^{-8},$$

$$\delta_2 \simeq \epsilon^{\frac{1}{2}} \sqrt{\frac{-2q(\alpha_2)}{p^{(2)}(\alpha_2)}} - \epsilon \frac{q'(\alpha_2)}{p^{(2)}(\alpha_2)} = 2.04207 \cdot 10^{-3},$$

$$\delta_3 \simeq -2.04207 \cdot 10^{-3}.$$

Així doncs:

$$\begin{aligned} \bar{\alpha}_1 &\simeq -1 + 2.778 \cdot 10^{-8}, & \bar{p}(\bar{\alpha}_1) &\simeq 9 \cdot 10^{-11}, \\ \bar{\alpha}_2 &\simeq 5 + 2.04207 \cdot 10^{-3}, & \bar{p}(\bar{\alpha}_2) &\simeq 8.5 \cdot 10^{-9}, \\ \bar{\alpha}_3 &\simeq 5 - 2.04207 \cdot 10^{-3}, & \bar{p}(\bar{\alpha}_3) &\simeq -8.5 \cdot 10^{-9}. \end{aligned}$$

Problema 5.10 Apliqueu el mètode de Newton amb dues variables per tal de calcular la solució del sistema no lineal:

$$\left. \begin{aligned} x_1 &= \sin(x_1 + x_2) \\ x_2 &= \cos(x_1 - x_2) \end{aligned} \right\}$$

prop de $x_1 = 1$, $x_2 = 1$. Acabeu el procés quan el vector residual, resultant de restar els dos membres de cada equació, sigui menor que 10^{-10} , en $\|\cdot\|_\infty$.

SOLUCIÓ:

Escrivim el sistema de la forma:

$$f \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1 - \sin(x_1 + x_2) \\ x_2 - \cos(x_1 - x_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

El mètode de Newton amb dues variables es concreta en la fórmula iterativa:

$$\begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \end{pmatrix} = \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \end{pmatrix} - \left[Df \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \end{pmatrix} \right]^{-1} f \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \end{pmatrix}.$$

En el nostre cas,

$$Df \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 - \cos(x_1 + x_2) & -\cos(x_1 + x_2) \\ \sin(x_1 - x_2) & 1 - \sin(x_1 - x_2) \end{pmatrix}$$

i el càlcul de la inversa dóna

$$\left[Df \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right]^{-1} = \frac{1}{\Delta(x_1, x_2)} \begin{pmatrix} 1 - \sin(x_1 - x_2) & \cos(x_1 + x_2) \\ -\sin(x_1 - x_2) & 1 - \cos(x_1 + x_2) \end{pmatrix},$$

on

$$\Delta(x_1, x_2) = 1 - \sin(x_1 - x_2) - \cos(x_1 + x_2) + 2 \sin(x_1 - x_2) \cos(x_1 + x_2).$$

Així doncs, la iteració s'escriu:

$$\begin{aligned} x_1^{(k+1)} &= x_1^{(k)} - \frac{(1 - \sin(x_1^{(k)} - x_2^{(k)}))(x_1^{(k)} - \sin(x_1^{(k)} + x_2^{(k)}))}{\Delta(x_1^{(k)}, x_2^{(k)})} \\ &\quad - \frac{\cos(x_1^{(k)} + x_2^{(k)})(x_2^{(k)} - \cos(x_1^{(k)} - x_2^{(k)}))}{\Delta(x_1^{(k)}, x_2^{(k)})}, \\ x_2^{(k+1)} &= x_2^{(k)} + \frac{\sin(x_1^{(k)} - x_2^{(k)})(x_1^{(k)} - \sin(x_1^{(k)} + x_2^{(k)}))}{\Delta(x_1^{(k)}, x_2^{(k)})} \\ &\quad - \frac{(1 - \cos(x_1^{(k)} + x_2^{(k)}))(x_2^{(k)} - \cos(x_1^{(k)} - x_2^{(k)}))}{\Delta(x_1^{(k)}, x_2^{(k)})}. \end{aligned}$$

Fent els càlculs s'obté la taula de valors següent:

k	$x_1^{(k)}$	$x_2^{(k)}$
0	1	1
1	0.9359511522	1
2	0.9350846651	0.9980207933
3	0.9350820642	0.9980200582
4	0.9350820641	0.9980200582

Es pot comprovar que

$$\left\| f \begin{pmatrix} x_1^{(4)} \\ x_2^{(4)} \end{pmatrix} \right\|_{\infty} \simeq 10^{-11}$$

i, per tant, la solució del sistema, amb la precisió desitjada, és

$$x_1^{(4)} = 0.9350820641 \text{ , } x_2^{(4)} = 0.9980200582 \text{ .}$$

PROBLEMES PROPOSATS

1. Demostreu que la funció

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

té un únic zero.

2. Demostreu que els polinomis de la forma $x^n + x + 1$ no tenen cap zero real, si n és parell, i en tenen exactament un, si n és senar.

3. a) Discutiu el nombre de solucions de l'equació

$$M = x - e \sin x \quad (\text{equació de Kepler el·líptica}) ,$$

per a tots els valors de $M \in [0, 2\pi)$ i $e \in (0, 1)$.

- b) Feu el mateix amb l'equació

$$M = e \sinh x - x \quad (\text{equació de Kepler hiperbòlica}) ,$$

per a tots els valors de $M > 0$ i $e > 1$.

4. Usant el mètode de bisecció, trobeu totes les arrels de les equacions següents amb un error més petit que 10^{-2} :

$$\text{i) } \sin x = \frac{1}{2}x^2 , \quad \text{ii) } e^{-x} = x^4 , \quad \text{iii) } \ln x = x - 1 .$$

5. Què s'obté quan posem un número qualsevol en la pantalla d'una calculadora i premem reiteradament la tecla $\boxed{\cos}$ (en radians)? Responen a la mateixa pregunta si, en lloc de prémer $\boxed{\cos}$, polsem $\boxed{\arccos}$. Interpreteu-ne els resultats.

6. Volem fer servir la fórmula d'iteració $x_{k+1} = 2^{x_k-1}$ per tal de resoldre l'equació $2x = 2^x$. Estudieu per a quins valors de x_0 convergeix i, donat el cas, a quin límit ho fa.

7. Expliqueu el comportament en el límit de la iteració $x_{k+1} = g(x_k)$ amb $g(x) = 1.4 \cos x$, partint d'un valor x_0 qualsevol.

8. Considerem la iteració simple $x_0 = 0$, $x_{k+1} = g(x_k)$ amb

$$g(x) = \frac{\epsilon - 2x^2 - x^3}{3},$$

essent ϵ suficientment petit.

- a) Digueu si és convergent o no i, en cas afirmatiu, cap a quin límit.
 b) Aplicació: Feu els càlculs per a $\epsilon = 0.1$ i trobeu, en aquest cas, l'ordre de convergència i la constant asimptòtica de l'error.
9. Es vol calcular la solució de l'equació $e^x = 3x$, usant iteració simple amb diferents funcions d'iteració:

$$g_1(x) = \frac{e^x}{3}, \quad g_2(x) = \ln(3x), \quad g_3(x) = \frac{e^x - x}{2}, \quad g_4(x) = e^x - 2x.$$

- a) Quines d'aquestes són útils?
 b) Amb quina d'elles calen menys iteracions per tal d'obtenir el resultat amb una precisió donada, partint del mateix valor inicial? Comproveu-ho numèricament, partint del valor aproximat $x_0 = 0.6$.
 c) Trobeu una funció d'iteració significativament millor que les indicades i empreu-la per al cas proposat a b).
10. Siguin f_1 i f_2 funcions derivables amb continuïtat a l'interval $[a, b]$ i suposem que una d'elles, per exemple f_1 , té inversa. Volem trobar una arrel simple de l'equació $f_1(x) = f_2(x)$.
 Considerem el mètode iteratiu $f_1(x_{k+1}) = f_2(x_k)$.
 a) Sota quines condicions és localment convergent?
 b) Aplicació numèrica: Calculeu l'arrel de $(1+x)\sin x = 1$ pertanyent a $[\frac{\pi}{2}, \pi]$.
11. a) Trobeu funcions d'iteració per al càlcul de $\sqrt{a}$ que tinguin ordres de convergència 1 i 2, respectivament.
 b) Empreu-les per tal d'obtenir $\sqrt{19}$ amb 8 xifres decimals correctes.

12. El *mètode de la corda* fa servir la iteració

$$x_{k+1} = x_k - mf(x_k) \quad (k \geq 0)$$

per al càlcul de zeros d'una funció f donada.

- a) Per a quins valors de m és localment convergent cap a un zero simple α de f ? Què passa si α és un zero múltiple de f ?
 b) Trobeu l'ordre de convergència en tots els casos.

c) Indiqueu quins problemes pràctics hi ha quan es volen assolir els ordres més elevats.

13. Calculeu pel mètode de Newton totes les solucions de les equacions

$$\text{i) } x = \cos x, \quad \text{ii) } x^3 = 10 \sin x, \quad \text{iii) } x^2 = e^{-x},$$

amb 8 xifres decimals i comproveu la convergència quadràtica del mètode en aquests casos.

14. a) Emprant el mètode de Newton, trobeu els punts fixos de la funció $g(x) = e^x - 2$ amb un error menor que 10^{-6} .

b) En quins intervals de valors inicials podem assegurar que el mètode convergeix, i cap a quina solució?

15. Siguin a, b reals amb $b > a$ i $f \in \mathcal{C}^2([a, b])$ complint $f'(x) \neq 0$, $f^{(2)}(x) \neq 0 \forall x \in [a, b]$; suposem que la funció f pren signes diferents en a i b ($f(a)f(b) < 0$) i que les correccions que el mètode de Newton fa d'aquests punts són menors que l'amplada de l'interval:

$$\left| \frac{f(a)}{f'(a)} \right| < b - a \quad \text{i} \quad \left| \frac{f(b)}{f'(b)} \right| < b - a.$$

Demostreu que el mètode de Newton, aplicat a trobar els zeros de f a $[a, b]$, convergeix per a qualsevol $x_0 \in [a, b]$.

16. Donada f de classe $\mathcal{C}^2$ amb un únic punt d'inflexió i tal que

$$\lim_{x \rightarrow -\infty} f(x) = -\infty, \quad \lim_{x \rightarrow \infty} f(x) = \infty.$$

Quina estratègia cal seguir, segons els diferents valors de c , per garantir la convergència del mètode de Newton aplicat a l'obtenció d'una solució de $f(x) = c$? I per a obtenir totes les solucions?

17. a) Demostreu que el mètode de Newton aplicat a la funció $f(x) = \frac{1}{x} - a$ permet calcular $\frac{1}{a}$ sense fer divisions.

b) Quina relació exacta existeix entre $e_{k+1} = x_{k+1} - \frac{1}{a}$ i $e_k = x_k - \frac{1}{a}$ ($k \geq 0$)?

c) Si $a = 0.4$ i $e_0 = -0.2$, per a quins valors de k tindrem $|e_k| \leq 10^{-20}$?

d) Doneu, en funció d' a , els valors de x_0 que fan que el mètode sigui convergent.

18. a) Trobeu el procediment més simple que, fent ús del mètode de Newton, calculi iterativament $\frac{1}{\sqrt[r]{a}}$ per a $r \in \mathbb{N}$.
 b) Aplicació: Calculeu $\sqrt[3]{0.1}$.
19. a) Demostreu que la funció $h(x) = \frac{f(x)}{f'(x)}$ té els mateixos zeros que f , però simples.
 b) Quins avantatges i inconvenients presenta l'aplicació del mètode de Newton a la funció h enfront de la seva aplicació a f ?
 c) Apliqueu el mètode de Newton per tal de trobar el zero de la funció $f(x) = (e^x - \pi)^3$ directament i fent servir la funció h corresponent.
20. Determineu per a quin valor d' $a > 0$ la funció $f(x) = \sin x - 2e^{-ax}$ té un únic zero més petit que π de multiplicitat 2. Calculeu llavors aquest zero emprant un mètode iteratiu d'ordre 2.
21. Considerem l'equació $f(x) = e^{ax} - x = 0$.
 a) Demostreu que existeix a_0 tal que, si $a > a_0$, f no té zeros reals i, si $a \leq a_0$, sí que en té. Trobeu a_0 i digueu quin és el nombre de zeros reals en funció d' a .
 b) Demostreu que, per a tot $a < a_0$ i per a tot valor inicial real, llevat d'un punt com a màxim, el mètode de Newton és convergent.
 c) Calculeu tots els zeros reals per a $a = \frac{1}{4}$.
22. Considerem l'equació $ax + b = \sin x$, on a i b són paràmetres reals. Determineu els valors d' a per als quals, qualsevol que sigui el valor de b , l'equació considerada té exactament 3 zeros. Feu el mateix per tal que tingui exactament 5 i 7 zeros respectivament, per a tot valor de b .
23. Trobeu tots els zeros de la funció $f(x) = \sin x - 2 - 0.3x$ amb un error més petit que 10^{-8} , usant el mètode de Newton.
24. a) Trobeu el primer punt de tall positiu α entre les corbes $y = \cos x$ i $y = e^{-x}$.
 b) Calculeu l'àrea compresa enmig de les esmentades corbes entre 0 i α .
25. a) Modifiqueu el mètode de Txebeixev

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} - \frac{f^{(2)}(x_k)f^2(x_k)}{2f'^3(x_k)},$$

per tal que tingui ordre almenys 3 quan l'usem per trobar zeros de multiplicitat coneguda m .

b) Aplicació: Calculeu, amb un error menor que 10^{-10} , el zero de $f(x) = (x \ln x - 1)^2$ que es troba prop de $x_0 = 1.8$ i comproveu numèricament l'ordre de convergència.

26. Sigui f una funció de la qual es vol calcular un zero simple.

a) Deduïu la fórmula corresponent al mètode iteratiu que obté x_{k+1} , a partir de x_k , avaluant en $y = 0$ el polinomi de Taylor de grau més petit o igual que 3 associat a la funció inversa local g de f prop del punt $f(x_k)$.

b) Trobeu l'ordre mínim del mètode i el coeficient asimptòtic de l'error quan es dona aquest ordre mínim.

27. Considerem una variant del mètode de Newton consistent a emprar aquest mètode, però usant el valor calculat de la derivada q vegades abans de calcular una nova derivada:

$$x_{rq+1} = x_{rq} - \frac{f(x_{rq})}{f'(x_{rq})} ; \quad x_{rq+i+1} = x_{rq+i} - \frac{f(x_{rq+i})}{f'(x_{rq})} \quad (i = 1 \div q - 1) .$$

a) Trobeu l'ordre del mètode iteratiu consistent a prendre, com a una iteració d'aquest, q iteracions del mètode donat.

b) Compareu la seva eficiència amb la del mètode de Newton.

c) Aplicació: Feu $q = 2$ i comproveu numèricament l'ordre trobat en el càlcul del zero de $f(x) = \cos x - x$.

28. Considerem el mètode iteratiu donat per la *fórmula de Halley*

$$x_{k+1} = x_k - \frac{2f(x_k)f'(x_k)}{2f'^2(x_k) - f(x_k)f^{(2)}(x_k)} .$$

Demostreu que té ordre de convergència almenys 3 per a zeros simples.

29. Donades les equacions

$$\text{i) } \sin x = e^{-x} , \quad \text{ii) } \tan x = \frac{x}{2} ,$$

volem trobar les solucions menors, en valor absolut, que 2π . Useu el mètode de la secant per trobar-les amb errors menors que 10^{-6} .

30. a) Useu el mètode de la secant per tal de calcular el zero de la funció $f(x) = \cos x - x$ amb un error menor que 10^{-12} .

b) Deduïu analíticament quant valen

$$\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k e_{k-1}}, \quad \lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k^{\frac{1+\sqrt{5}}{2}}}.$$

c) Comproveu numèricament els límits anteriors.

31. Trobeu l'ordre i la constant asimptòtica de l'error del mètode de la secant, quan s'aplica a la cerca de zeros de multiplicitat senar.

32. a) Expliciteu el mètode iteratiu consistent a aproximar el terme $f'(x_k)$ que apareix en el mètode de Newton per la derivada en el punt x_k de la paràbola que passa pels punts $(x_k, f(x_k))$, $(x_{k-1}, f(x_{k-1}))$ i $(x_{k-2}, f(x_{k-2}))$.

b) Quin és el seu mínim ordre de convergència per a zeros simples?

c) Compareu la seva eficiència amb la del mètode de Newton, en el cas de cercar zeros simples α amb $f^{(2)}(\alpha) \neq 0$.

d) Apliqueu-lo a l'obtenció del zero de $f(x) = \cos x - x$.

33. Dissenyem un mètode per obtenir zeros simples de la funció f així: quan coneixem x_k i x_{k-1} , trobem el polinomi $q_3(y)$ d'interpolació d'Hermite a la funció inversa local g de f

$$q_3(f(x)) = x, \quad q_3'(f(x)) = g'(f(x)) \quad (x = x_k, x_{k-1}),$$

després prenem, com a nova aproximació del zero buscat, $x_{k+1} = q_3(0)$.

a) Expliciteu aquest mètode i obteniu el seu ordre de convergència mínim per a la cerca de zeros simples.

b) Generalització: Trobeu també l'ordre mínim (en funció de n i m), d'un mètode anàleg a l'anterior, però fent interpolació d'Hermite generalitzada a g fins a les derivades d'ordre n en els punts $f(x_k), \dots, f(x_{k-m})$.

34. Considerant f de classe $\mathcal{C}^3$, volem construir un mètode iteratiu per tal de trobar zeros simples α de f ; per això, conegudes les aproximacions x_k i x_{k-1} , trobem x_{k+1} com el zero més proper a x_k del polinomi interpolador a f en x_{k-1} i x_k segons les condicions:

$$p_2(x_{k-1}) = f(x_{k-1}), \quad p_2(x_k) = f(x_k), \quad p_2'(x_k) = f'(x_k).$$

a) Trobeu x_{k+1} en funció de x_k i x_{k-1} .

b) Trobeu una expressió de l'error $f(x) - p_2(x)$.

c) Fent $e_k = x_k - \alpha$ i suposant que $f^{(3)}(\alpha) \neq 0$, demostreu que

$$\lim_{k \rightarrow \infty} \frac{e_{k+1}}{e_k^2 e_{k-1}} = \frac{f^{(3)}(\alpha)}{6f'(\alpha)}.$$

d) Trobeu el seu ordre i la seva constant asimptòtica de l'error.

35. El mètode de la secant es pot deduir del de Newton substituint $f'(x_k)$ per la derivada primera en x_k de la recta d'interpolació en els punts d'abscisses x_k i x_{k-1} . Prenent com a mètode de partida el de Txebixev i substituint $f^{(2)}(x_k)$ per la derivada segona en x_k del polinomi interpolador d'Hermite en els punts d'abscisses x_k i x_{k-1} , s'obté un nou mètode iteratiu per al càlcul de zeros de funcions.

a) Expliciteu la fórmula d'iteració d'aquest mètode.

b) Trobeu-ne l'ordre de convergència mínim per a zeros simples.

c) Compareu la seva eficiència amb la del mètode de la secant i la del mètode de Txebixev en el cas més general.

36. Com varia el zero $\alpha = 2$ de

$$f(x) = e^{6x}(x^3 + 2x^2 - x - 14) + e^{2x}(x^4 - 5x^3 + x^2 + 20)$$

si canviem 6 per $6 + \epsilon$ i 20 per $20 + \delta$, amb ϵ i δ prou petits? Apliqueu-ho al cas: $\epsilon = 10^{-6}$ i $\delta = 10^{-5}$.

37. Per resoldre l'equació $f(x) = 0$, s'aplica el mètode iteratiu

$$x_{k+1} = x_k - g(x_k)f(x_k).$$

a) Quines condicions ha de complir g per tal que l'ordre de convergència cap a un zero simple de f sigui el més gran possible? Comproveu que el mètode de Newton les verifica.

b) En el cas que sigui impossible o molt costosa l'avaluació de f' en el mètode de Newton, podem aproximar $f'(x_k)$ per

$$\frac{f(x_k + h_k) - f(x_k)}{h_k},$$

amb $h_k > 0$. Si l'algorisme d'avaluació de f comet un error relatiu fitat per ϵ i $|f^{(2)}(x)| \leq M_2 \forall x$, trobeu el valor de h_k per al qual la fita de l'error comès en l'aproximació de $f'(x_k)$ sigui mínima (suposant $f(x_k + h_k) \simeq f(x_k)$).

c) Demostreu que, si en l'expressió recomanada a a) per a $g(x_k)$ se substitueix $f'(x_k)$ per l'aproximació feta a b) amb el valor òptim de h_k trobat allà, el mètode iteratiu obtingut té *convergència superlineal*; és a dir, que

$$\lim_{k \rightarrow \infty} \frac{x_{k+1} - \alpha}{x_k - \alpha} = 0.$$

38. El mètode iteratiu $x_{k+1} = g(x_k)$ ($k \geq 0$), que calcula una solució α de $x = g(x)$, té una funció d'iteració g que s'avalua amb un error absolut fitat per δ .
- a) Si $|g'(x)| \leq M < 1$, expresseu una fita de l'error ϵ_k del valor calculat $f(x_k)$ de x_k en funció de l'error inicial $\epsilon_0 = |x_0 - \alpha|$.
- b) Estimeu el nombre màxim d'iteracions que convindria fer en usar el mètode donat.
39. Apliquem el mètode de Newton a un polinomi $p(x)$ amb tots els zeros reals a fi de trobar el seu zero α més gran. Si iniciem les iteracions en un valor $x_0 > \alpha$, podem assegurar que és convergent a α ? Per què?
40. Demostreu que, donats un polinomi $p(x)$ de grau n i un valor a tal que $p'(a) \neq 0$, existeix com a mínim un zero de $p(x)$ dins el cercle $\mathcal{C}$ definit per

$$\mathcal{C} = \left\{ z \in \mathbb{C} \mid |z - a| \leq n \left| \frac{p(a)}{p'(a)} \right| \right\}.$$

Per a la demostració convé que seguiu els passos següents:

- a) Utilitzeu el desenvolupament de Taylor de $p(x)$ prop d' a .
- b) Si α_i ($i = 1 \div n$) són els zeros de $p(x)$, determineu un polinomi $q(x)$ que tingui els zeros

$$\beta_i = \frac{1}{\alpha_i - a} \quad (i = 1 \div n).$$

- c) Vegeu que, si un polinomi

$$q(x) = b_n x^n + b_{n-1} x^{n-1} + \cdots + b_1 x + b_0$$

té els zeros β_i ($i = 1 \div n$), es compleix que

$$\frac{b_{n-1}}{b_n} = -(\beta_1 + \beta_2 + \cdots + \beta_n).$$

- d) Deduïu-ne el resultat proposat inicialment.

41. La *regla de Descartes* assegura que, si ν és el nombre de canvis de signe dels coeficients d'un polinomi $p(x)$ i μ és el nombre de zeros positius de $p(x)$, aleshores $\mu \leq \nu$ i $\nu - \mu$ és parell.
- a) Usant la regla de Descartes, demostreu que el polinomi de tercer grau $p(x) = x^3 - x^2 - x - 1$ té un únic zero real positiu.
- b) Calculeu-lo pel mètode de Birge-Vieta.

42. Considerem l'equació polinomial $p(x) = 16x^3 + 12x^2 - 8x - 1 = 0$.
- Fiteu les seves arrels amb els mètodes donats a l'apartat 5.2.3.
 - Quantes arrels positives i negatives té?
 - Determineu les seves tres arrels.
43. a) Separeu les arrels de $p(x) = 36x^4 - 60x^3 - 29x^2 + 9x + 2 = 0$, usant el mètode de Sturm.
- b) Trobeu les dites arrels amb 6 xifres decimals correctes pel mètode de Muller-Traub.
44. Useu el mètode de Laguerre, partint de $x_0 = 1000$, per calcular un zero del polinomi $p(x) = x^4 + 2x^3 + 3x^2 + 4x + 5$.
45. a) Determineu, pel mètode de Bernoulli, el zero de mòdul més petit del polinomi $p(x) = 32x^3 - 48x^2 + 18x - 1$ amb 3 xifres significatives.
- b) Refineu-lo fins a 6 xifres significatives emprant el mètode de Birge-Vieta o el de la secant.
46. a) Calculeu, amb 3 xifres significatives i pel mètode de Bernoulli, el zero més gran en mòdul del polinomi
- $$p(x) = x^4 - 4x^3 - 2x^2 - 12x + 9 ,$$
- usant les condicions inicials $x_k = 0$ ($k = 0 \div 2$), $x_3 = 1$.
- b) Observeu que la convergència en l'apartat a) és lenta. Utilitzeu el mètode Q-D com a mètode alternatiu.
47. Determineu, pel mètode de Bairstow, els zeros complexos del polinomi $p(x) = x^3 - x - 1$, a partir del factor quadràtic aproximat $x^2 + x$.
48. Calculeu, emprant el mètode o la combinació de mètodes que cregueu més convenient, tots els zeros dels polinomis següents:
- $x^3 - 2x - 5$,
 - $x^3 - 16x^2 - 3$,
 - $x^4 - 3x^2 + 2x - 1$,
 - $x^4 + 4x^2 - 3x - 1$,
 - $x^6 + 5x^3 + 7x + 1$.

49. Determineu el zero de mòdul més gran dels polinomis:

- a) $x^3 - 20x^2 - 3x + 18$,
- b) $x^4 - 3x^3 - 60x^2 + 150x + 300$,
- c) $10x^3 - 21x^2 - 40x + 84$.

50. Considerem el polinomi $p(x) = x^3 - 3x + 2$. Com varien els seus zeros en pertorbar-lo substituint el coeficient -3 per $-3 + 10^{-6}$?

51. Els punts 2-periòdics d'una funció f són aquells valors de x que compleixen $x = f(f(x))$. Notem que els punts 2-periòdics satisfan el sistema d'equacions

$$\left. \begin{array}{l} x_1 = f(x_2) \\ x_2 = f(x_1) \end{array} \right\} .$$

Per a la seva resolució, podem usar un mètode iteratiu de tipus Jacobi

$$x_1^{(k+1)} = f(x_2^{(k)}) , \quad x_2^{(k+1)} = f(x_1^{(k)}) ,$$

o de tipus Gauss-Seidel

$$x_1^{(k+1)} = f(x_2^{(k)}) , \quad x_2^{(k+1)} = f(x_1^{(k+1)}) ,$$

per analogia amb la resolució iterativa de sistemes lineals.

- a) Doneu condicions necessàries de convergència dels mètodes esmentats, suposant coneguda la derivada de f en la solució.
- b) Com s'hauria de variar el mètode en el cas que no es complissin les dites condicions?
- c) Aplicació: Trobeu els punts 2-periòdics de $f(x) = 1.4 \cos x$, partint de $x_0 = 1$.

52. Trobeu, usant mètodes iteratius de tipus Jacobi i Gauss-Seidel a partir de $x_0 = y_0 = 1$, la solució del sistema no lineal

$$\left. \begin{array}{l} x = \sin(x + y) \\ y = \cos(x - y) \end{array} \right\} .$$

53. Donat un sistema no lineal del tipus

$$\left. \begin{array}{l} x_1 = f_1(x_1, x_2, x_3) \\ x_2 = f_2(x_1, x_2, x_3) \\ x_3 = f_3(x_1, x_2, x_3) \end{array} \right\} ,$$

escrivim el sistema com $x = f(x)$, amb

$$x = (x_1, x_2, x_3) , \quad f = (f_1, f_2, f_3) ,$$

i cerquem una solució $\alpha = (\alpha_1, \alpha_2, \alpha_3)$.

a) Doneu condicions necessàries de convergència dels mètodes iteratius de tipus Jacobi i Gauss-Seidel, en funció dels coeficients de la matriu $A = Df(\alpha)$.

b) Aplicació: Estudieu el cas en què resulti

$$A = \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ -0.3 & 0.4 & -0.2 \\ 0.6 & -0.1 & 0.2 \end{pmatrix}.$$

54. a) Trobeu, emprant el mètode de Newton amb tres variables, la solució del sistema no lineal

$$\left. \begin{array}{lcl} x & = & 2 \cos(-x + y + z) \\ y & = & 2 \cos(x - y + z) \\ z & = & 2 \cos(x + y - z) \end{array} \right\},$$

prop de $x = y = z = 1$.

b) Comproveu la convergència quadràtica del mètode, emprant la norma $\| \cdot \|_{\infty}$.

BIBLIOGRAFIA

- [Apo57] T. M. Apostol. *Mathematical analysis*. Addison-Wesley, Reading, Mass., 1957. En castellà: Reverté, Barcelona, 1982.
- [AS65] M. Abramowitz and I. A. Stegun, editors. *Handbook of mathematical functions*. Dover, N.Y., 1965.
- [CdB72] S.D. Conte and C. de Boor. *Elementary numerical analysis, an algorithmic approach*. McGraw-Hill, N.Y., 1972. En castellà: 1974.
- [Che66] E.W. Cheney. *Introduction to approximation theory*. McGraw-Hill, N.Y., 1966.
- [Cia82] P. G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson, París, 1982.
- [Cra46] H. Cramér. *Mathematical methods of statistics*. Princeton Univ. Press, Princeton, N.J., 1946.
- [Dav75] P. J. Davis. *Interpolation and approximation*. Dover, N.Y., 1966, 1975.
- [DB74] G. Dahlquist and Å. Björck. *Numerical Methods*. Prentice Hall, Englewood Cliffs, N.J., 1974.
- [Die68] J. Dieudonné. *Calcul infinitésimal*. Hermann, París, 1968. En castellà: Omega, Barcelona, 1971.
- [DM73] B. Demidovitch and I. Maron. *Eléments de calcul numérique*. Mir, Moscou, 1973. En castellà: Paraninfo, Madrid, 1977.
- [DR67] P. J. Davis and P. Rabinowitz. *Numerical integration*. Blaisdell, Londres, 1967.
- [Dur61] E. Durand. *Solutions numériques des équations algébriques*. Volumes 1,2. Masson, París, 1960, 1961.
- [Fel50] W. Feller. *An introduction to the probability theory and its applications*. Wiley, N.Y., 1950.
- [Fro69] C. E. Froberg. *Introduction to numerical analysis*. Addison-Wesley, Reading, Mass., 1969. En castellà: Vicens Vives, 1977.
- [Gas66] N. Gastinel. *Analyse numérique linéaire*. Hermann, París, 1966. En castellà: Reverté, 1975.

- [Ham70] S. J. Hammarling. *Latent roots and latent vectors*. Adam Hilger, Londres, 1970.
- [Ham73] R. W. Hamming. *Numerical methods for scientists and engineers*. McGraw-Hill, N.Y., 2a edició, 1973.
- [Hen64] P. Henrici. *Elements of numerical analysis*. Wiley, N.Y., 1964. En castellà: Trillas, Mèxic, 1968.
- [Hen74] P. Henrici. *Applied and computational complex analysis*. Wiley, N.Y., 1974.
- [Hil74] F. B. Hildebrand. *Introduction to numerical analysis*. McGraw-Hill, N.Y., 2a edició, 1974.
- [Hou53] A. S. Householder. *Principles of numerical analysis*. McGraw-Hill, N.Y., 1953.
- [Hou64] A. S. Householder. *The theory of matrices in numerical analysis*. Blaisdell, N.Y., 1964.
- [HS74] M. W. Hirsch and S. Smale. *Differential equations, dynamical systems, and linear algebra*. Academic Press, N.Y., 1974. En castellà: Alianza Universidad, Madrid, 1984??
- [IK66] E. Isaacson and H. B. Keller. *Analysis of numerical methods*. Wiley, N.Y., 1966.
- [Jac77] D. A. H. Jacobs, editor. *The state of the art in numerical analysis*. Academic Press, N.Y., 1977.
- [Knu69] D. E. Knuth. *The art of computer programming*. Addison-Wesley, Reading, Mass., 1969.
- [LH74] C. L. Lawson and R. J. Hanson. *Solving least squares problems*. Prentice-Hall, Englewood Cliffs, N.J., 1974.
- [M.66] Ostrowski A. M. *Solution of equations and systems of equations*. Academic Press, N.Y., 2a edició, 1966.
- [Moo66] R. E. Moore. *Interval analysis*. Prentice-Hall, Englewood Cliffs, N.J., 1966.
- [OR70] J. Ortega and W. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic Press, N.Y., 1970.
- [Par80] B. N. Parlett. *The symmetric eigenvalue problem*. Prentice-Hall, Englewood Cliffs, N.J., 1980.
- [Que71] M. Queysanne. *Algebre*. Librairie Armand Colin, París, 1971. En castellà: Vicens-Vives, Barcelona, 1973.
- [Ral65] A. Ralston. *A first course in numerical analysis*. McGraw-Hill, N.Y., 1965. En castellà: Limusa-Wiley, Mèxic, 1970.
- [Ric81] J. R. Rice. *Matrix computation and mathematical software*. McGraw-Hill, N.Y., 1981.

- [RR78] A. Ralston and P. Rabinowitz. *A first course in numerical analysis*. McGraw-Hill, N.Y., 2a edició, 1978.
- [RW67] A. Ralston and H. S. Wilf, editors. *Mathematical methods for digital computers*. Wiley, N.Y., 1967.
- [SB80] J. Stoer and R. Bulirsch. *Introduction to numerical analysis*. Springer, Berlín, 1980.
- [Sch67] F. Scheid. *Numerical analysis, including 775 solved problems*. Schaum, N.Y., 1967. En castellà: 1972.
- [SS66] A. H. Stroud and D. Secrest. *Gaussian quadrature formulas*. Prentice-Hall, Englewood Cliffs, N.J., 1966.
- [Ste73] G. W. Stewart. *Introduction to matrix computations*. Academic Press, N.Y., 1973.
- [Str71] A. H. Stroud. *Approximate calculation of multiple integrals*. Prentice-Hall, Englewood Cliffs, N.J., 1971.
- [Sze59] G. Szego. *Orthogonal polynomials*. AMS, N.Y., 1959.
- [Tra64] J.F. Traub. *Iterative methods for the solution of equations*. Prentice-Hall, Englewood Cliffs, N.J., 1964.
- [Wil64] J. H. Wilkinson. *Rounding errors in algebraic processes*. Prentice-Hall, Englewood Cliffs, N.J., 1964.
- [Wil65] J. H. Wilkinson. *The algebraic eigenvalue problem*. Clarendon, Oxford, 1965.
- [WR71] J. H. Wilkinson and C. Reinsch. *Handbook for automatic computation. Vol. 2: Linear algebra*. Springer, Berlín, 1971.
- [YG72] D. M. Young and R. T. Gregory. *A survey of numerical mathematics*. Addison-Wesley, Reading, Mass., 1972.
- [You71] D. M. Young. *Iterative solution of large linear systems*. Academic Press, N.Y., 1971.

ÍNDIX DE SÍMBOLS I ALFABÈTIC

- $(A)_k$, submatriu principal, 49
- $(\ , \)_m$, producte escalar discret sobre I_M , 182
- (u, v) , producte escalar de u i v , 164
- $\langle x_0, \dots, x_m, x \rangle$, mínim interval que conté $x_0, \dots, x_m, x$, 148
- A^{-1} , matriu inversa d' A , 49
- A_j , coeficient principal d'un polinomi, 174
- B_J , matriu d'iteració del mètode iteratiu de Jacobi, 72
- $B_j(t)$, polinomis de Bernoulli, 223
- B_j , nombres de Bernoulli, 223
- B_{GS} , matriu d'iteració del mètode iteratiu de Gauss-Seidel, 72
- B_ω , matriu d'iteració del mètode iteratiu de sobrerelaxació, 72
- D , matriu diagonal, 54
- D , operador de derivació, 289
- E , operador desplaçament cap endavant, 289
- E_S , error de la fórmula de Simpson, 268
- E_j , nombres d'Euler, 253
- E_m , error d'una fórmula d'integració interpolatòria, 267
- F funció distribució, 7
- $H_j(x)$, polinomi d'Hermite, 209
- I , matriu identitat, 49
- J , operador d'integració, 289
- $J(f)$, integral de f , 265
- J_n funcions de Bessel, 42
- K_m , factor d' E_m , 269
- L , matriu triangular inferior de la factorització LU, 58
- L , operador lineal, 290
- $L_j(x)$, polinomis de Laguerre, 259
- M^* , matriu adjunta de M , 49
- $M^\top$, matriu transposada de M , 49
- M_{m+1} , fita de la funció f^{m+1} , 202
- $P(u) = I - \alpha uu^\top$, matriu de Householder associada al vector u , 61
- $P_{j,m}(t)$, polinomis de Gram, 178
- Q , matriu ortogonal de la factorització QR, 60
- R , matriu triangular superior de la factorització LR, 87
- R , matriu triangular superior de la factorització QR, 60
- $R_n(x)$, error de la interpolació de Taylor, 153
- $R_n(x)$, residu d'un desenvolupament, 281
- S , suma d'una sèrie, 280
- S_j , suma parcial d'una sèrie, 280
- $T_j(t)$, polinomis de Txeixev, 185
- U , matriu triangular superior de la factorització LU, 58
- W , matriu de pesos, 170
- W_k, w_k , pesos de fórmules d'integració numèrica, 266
- Δ , operador diferència cap endavant, 289
- Δ_k , determinant principal, 49
- Δ_k , determinant principal d'ordre k , 90
- Γ , funció gamma d'Euler, 344
- Φ_k, Ψ_k , polinomis bàsics de la interpolació d'Hermite, 157

- Ψ , funció digamma, 342
- α , zero d'una funció, 349
- $\bar{\omega}_m(u) = (u + \frac{m}{2})(u + \frac{m}{2} - 1) \cdots (u - \frac{m}{2})$, 202
- $\bar{x}$ valor exacte de x , 1
- δ , operador diferència centrada, 289
- δ_{ij} , delta de Kronecker, 49
- ϵ_A , fita error relatiu de representació en punt flotant per arrodoniment, 2
- ϵ_T , fita error relatiu de representació en punt flotant per tall, 2
- $\epsilon_a(x)$ fita de l'error absolut de x , 1
- $\epsilon_r(x)$ fita de l'error relatiu de x , 1
- $\frac{\Delta}{q^{p_j-1}}$, tipus d'extrapolació, 288
- γ , constant d'Euler, 315
- $\hat{f}_n$, millor aproximació minimax de f , 182
- λ , valor propi, 51
- $\binom{\alpha}{j}$, nombre combinatori, 154
- Ai, funció d'Airy, 317
- $\text{cov}(x, y)$, covariància de x i y , 167
- $\det A$, determinant d' A , 49
- $\text{erf}(x)$ funció d'error, 8
- $\text{fl}(x)$, representació en punt flotant de x , 2
- $\text{fl}_A(x)$, representació en punt flotant de x per arrodoniment, 3
- $\text{fl}_T(x)$, representació en punt flotant de x per tall, 3
- μ mitjana, 7
- μ , operador mitjana, 289
- $\mu(A)$, nombre de condició de la matriu A , 69
- ∇ , diferència cap endarrera, 289
- $\omega_m(x) = (x - x_0)(x - x_1) \cdots (x - x_m)$, 202
- $\overline{M}$, matriu conjugada de M , 49
- $\bar{x}$, mitjana de x , 167
- ψ_j , funcions bàsiques ortogonals, 169
- ρ funció densitat, 7
- $\rho(A)$, radi espectral d' A , 51
- ρ_{xy} , coeficient de regressió, 168
- σ desviació estàndard, 7
- $\sigma(x)$, desviació típica de x , 167
- σ^2 variància, 7
- φ_j , funcions bàsiques d'aproximació, 161
- ζ , funció zeta de Riemann, 315
- $e^{(j)}$, element j -è de la base canònica, 62
- $e_a(x)$ error absolut de x , 1
- $e_n(x)$, funció error d'aproximació, 161
- $e_r(x)$ error relatiu de x , 1
- $f[x_i, \dots, x_{i+j}]$, diferència dividida de Newton, 151
- f_n^* , millor aproximació per mínims quadrats de f , 163
- g_n , factor d'amplificació en l'eliminació gaussiana, 70
- $l_k(x)$, polinomi de Lagrange, 149
- $o(g(x))$, funció d'ordre més petit que g , 155
- p_A , polinomi característic de la matriu A , 51
- p_{ij} , polinomis interpolador del mètode d'Aitken i Neville, 150
- $t_n(\theta)$, suma trigonomètrica, 179
- $u^\perp$, subspai ortogonal al vector u , 61
- v , vector propi, 51
- w , funció pes, 163
- w_k , pesos, 162
- $\mathcal{M}_{m,n}$, espai vectorial de les matrius complexes $m \times n$, 49
- $\mathcal{F}_n$, espai vectorial generat per funcions bàsiques, 163
- $\mathcal{O}(g(x))$, funció del mateix ordre que g , 155
- $\mathcal{P}_n$, espai vectorial dels polinomis de grau menor o igual que n , 174
- $\mathcal{T}_n$, espai vectorial de sumes trigonomètriques, 179
- abscisses d'extrem, 183, 185, 186, 245
- abscisses d'interpolació, 146
- acceleració de la convergència, 356, 357
- acceleració de la sumació, mètode de, 284
- acceleració de mètodes iteratius, 372
- Airy, funció d', 317
- Aitken, acceleració d', 89
- Aitken, mètode d', 149, 150, 152, 251
- Aitken, mètode d'acceleració d', 356, 357
- aleatòria, variable, 7-9
- algorisme, 30, 32

- analítica, funció, 156
- anàlisi d'interval·ls, 10
- aproximació discreta, 169
- aproximació numèrica d'arrels, 350
- aproximació per mínims quadrats, 233
- aproximació contínua, 161, 169, 172
- aproximació discreta, 161, 166, 169, 172, 173
- aproximació minimax, 163, 182–186, 245–247, 261
- aproximació per mínims quadrats, 163–166, 168–170, 174, 176, 181, 233, 234
- aproximació per mínims quadrats, 174
- aproximació polinomial, 160, 161, 173
- aproximació polinomial per mínim quadrats, 243
- aproximació polinomial per mínims quadrats, 243
- aproximació trigonomètrica, 161, 169, 179, 180, 182
- aproximació trigonomètrica equidistant, 233
- arrel, 349
- arrodoniment, 2
- Bairstow, mètode de, 368, 372, 387, 404
- banda, matriu, 128
- Barrow, regla de, 265
- base de funcions ortogonals, 169, 171, 172, 180
- base de funcions ortonormals, 172
- base de funcions trigonomètriques ortogonals, 182, 185
- Bernoulli, mètode de, 364, 365, 367, 372, 404
- Bernoulli, nombres de, 223, 224, 253, 273, 274, 285, 286, 315, 316, 331
- Bernoulli, polinomis de, 223, 273
- Bessel, funcions de, 42
- Bessel, funció de, 205, 251
- Bessel, fórmula de, 347
- bidiagonal, matriu, 132
- bifurcació de zeros, 392
- Birge-Vieta, mètode de, 364, 387, 403, 404
- bisecció, mètode de, 121, 350–352, 356, 363, 364, 396
- bit, 16
- cancel·lació, 3, 24, 25, 63, 111, 239
- característic, polinomi, 51
- característica, equació, 51, 86
- Clenshaw, regla de, 176, 243
- coeficient principal d'un polinomi, 174, 175, 177, 178, 186
- coeficients indeterminats, mètode dels, 266
- coeficients ortogonals, 169, 176, 179, 180
- comparació, mètode de, 287, 313
- compatible, sistema, 50
- condicions de frontera, 124
- conjunt d'abscisses d'aproximació, 161, 163, 173
- conjunt d'abscisses d'aproximació simètric, 178
- consistent amb una norma vectorial, norma matricial, 53, 55, 70
- constant asimptòtica de l'error, 355, 373
- convergència cúbica, 355, 364
- convergència global, 372
- convergència lineal, 114, 355–357
- convergència quadràtica, 355–357, 376
- convergència superlineal, 402
- corda, mètode de la, 397
- cosinus, fórmula del, 46
- covariància, 167
- criteri d'alternança de residus, 284
- criteri d'alternança dels residus, 274, 283
- criteri d'alternància de residus, 285
- criteri d'alternància dels residus, 156, 316
- criteri de comparació amb una sèrie geomètrica, 284
- criteri integral, 284, 314
- criteri per a sèries alternades, 284
- Crout, mètode de, 56, 60, 91, 127
- Danilevski, mètode de, 105
- definida positiva, matriu, 50, 53, 59, 73, 93, 129, 131, 132, 168
- deflació cap al darrera de polinomis, 361

- deflació d'un polinomi, 383
- deflació d'una matriu, 74, 76
- deflació de matrius, 89
- deflació de polinomis, 73, 361
- deflació enrera de polinomis, 364
- densitat, funció, 7
- derivació interpolatòria, fórmula de, 262, 263, 297
- derivació numèrica, 262, 294
- derivació numèrica, fórmula d', 294
- derivació numèrica, fórmula de, 289
- derivades d'ordre superior, 263, 295
- Descartes, regla de, 403
- desenvolupament asimptòtic, 155, 281, 282, 288, 293, 317
- desenvolupament asimptòtics, 282
- desenvolupament de Taylor, 155, 282
- desviació estàndard, 167
- desviació típica, 167
- desviacions, 166
- desviació estàndard, 7, 9
- determinant, 50
- determinant de matrius Hessenberg, recurrència de, 86
- determinant de matrius tridiagonals, recurrència de, 85
- determinant principal, 49
- determinants principals, 58, 90, 94
- determinants, càlcul de, 67
- determinat, sistema, 50
- diagonal, matriu, 50, 54, 55, 59, 79, 81, 88, 134
- diagonalitzable, matriu, 50, 54, 55, 76, 88
- diferències dividides, 151, 323
- diferències dividides generalitzades, 158, 159, 228
- diferències dividides generalitzades, taula de, 230
- diferències dividides, esquema de, 154
- diferències dividides, taula de, 206, 297
- direcció pròpia, 51
- discretització, 287
- discretitzat, problema, 124, 125, 141
- distribució normal, 7, 9
- distribució uniforme, 7
- distribució, funció, 7, 9
- distribució, funció de, 7
- divisió sintètica, 147
- Doolite, mètode de, 59
- Doolittle, mètode de, 56, 60, 127
- dígit, 2
- equacions normals, 164–172, 174, 176, 181, 233, 234, 239, 242, 245
- equació característica, 51, 86, 106, 109
- equació diferencial lineal, 124
- equació diferencial lineal, solució d'una, 124
- equioscil·lació, propietat de, 183
- error absolut, 1
- error absolut, fita d', 1
- error cap al darrera, anàlisi de l', 5, 69, 70
- error d'aproximació, funció, 161, 162
- error d'arrodoniment, 3, 5, 68, 69, 77
- error d'avaluació, 326, 328
- error d'interpolació, 148
- error d'interpolació de Taylor, expressió de Lagrange de l', 154, 212, 213
- error de derivació numèrica, expressió asimptòtica de l', 264
- error de l'aproximació per mínims quadrats, 236, 244
- error de la interpolació d'Hermite, 278
- error de la derivació interpolatòria, 262–264
- error de la fórmula de Gauss-Legendre, 311
- error de la fórmula de Newton-Cotes, 269
- error de la fórmula de Simpson, 268
- error de la fórmula del trapezi, expressió asimptòtica de l', 273
- error de la fórmules d'integració interpolatòria, 267
- error de la integració interpolatòria, 267
- error de la integració numèrica, 266, 296
- error de la interpolació d'Hermite, 301
- error de la interpolació de Taylor, 281
- error de la interpolació de Taylor, expressió integral de l', 153
- error de la regla dels trapezis, 273

- error de la regla dels trapezis, expressió
asimptòtica de l', 272, 273
- error de la solució, 56, 58, 68
- error de les dades, 58, 68
- error de les fórmules gaussianes, 277
- error de truncament, 3, 5, 326, 328
- error en la interpolació d'Hermite
generalitzada, 229
- error en la solució, 69, 70
- error estàndard, fórmula de propagació
de l', 9
- error maximal, fórmula de propagació de
l', 4-6
- error relatiu, 1
- error relatiu en les operacions
aritmètiques, fita de l', 70
- error relatiu, fita, 3, 5
- error relatiu, fita d', 1
- error, anàlisi de l', 56
- error, fita relativa de la norma del vector
d', 68
- error, fita relativa del vector d', 98
- error, funció d', 8
- error, fórmula de propagació de l', 4, 9,
69
- error, norma del vector d', 69
- error, tractament estadístic de l', 7, 8
- errors en la deflació, 74
- errors d'arrodoniment, 69
- escala, procés, 354
- escassa, matriu, 89
- estrictament diagonal dominant, matriu,
50, 54, 72
- Euclides, algorisme d', 363
- euclidiana, norma vectorial, 53, 61
- Euler, constant d', 315
- Euler, nombres d', 253
- Euler-Maclaurin per a sumes, fórmula d',
285, 286, 315, 341, 342
- Euler-Maclaurin, fórmula d', 271, 273,
285, 286, 296, 318, 330, 331, 348
- Everett, fórmula d', 347
- exactitud d'una fórmula, 289, 293
- exactitud d'una fórmula d'integració
numèrica, 266, 267, 269, 274-279
- exactitud d'una fórmula d'interpolació,
293
- exactitud d'una fórmula de derivació
numèrica, 294
- exponent, 2
- exponent modificat, 16
- expressió asimptòtica d'una fórmula de
derivació numèrica, 295
- expressió asimptòtica de l'error de
derivació numèrica, 264
- expressió integral de l'error
d'interpolació de Taylor, 268
- extrapolació, 288, 296, 326
- extrapolació repetida, 288
- extrapolació, tipus d', 288, 327
- extrapolació, tipus de, 319
- factor d'un polinomi, 147
- factor òptim de sobrerelaxació, 140
- factorial, funció, 287
- factorització QR, 130, 132
- Fibonacci, nombres de, 253
- fitació de zeros, 360
- fitació de zeros de polinomis, 362
- fitació dels residus de les sèries, 296
- fites per a zeros de polinomis, 372
- fonts d'error, 1
- Fresnel, integrals de, 343
- Frobenius, forma normal de, 105, 107
- Frobenius, matriu de, 104, 105, 107, 364
- funcions bàsiques, 161, 167, 169, 171
- funció analítica, 282
- funció contractiva, 355, 356, 370
- funció d'error d'aproximació, 245
- funció d'ordre més petit, 155
- funció del mateix ordre, 155
- funció digamma, 342
- funció error, 208, 344
- funció error de la interpolació, 263
- funció gamma d'Euler, 344
- funció pes, 163, 178, 182, 276, 279
- funció pes simètrica, 177, 178
- funció pes singular, 163
- funció zeta de Riemann, 341
- fórmula d'Euler-Maclaurin per a sèries,
286
- Gauss cap al darrera, fórmula de, 322

- Gauss cap al davant, fórmula de, 322, 324, 346
- Gauss, mètode de, 56, 58, 67, 68, 90, 92, 126, 130, 132, 137
- Gauss-Jordan, mètode de, 68, 130
- Gauss-Legendre, fórmula de, 280, 302, 311, 312
- Gauss-Legendre, pesos de la fórmula de, 312
- Gauss-Seidel, mètode iteratiu de, 71
- Gauss-Seidel, mètode iteratiu de, 71, 72
- Gauss-Seidel, mètode iteratiu de, 99, 138, 139, 141
- Gauss-Seidel, mètode iteratiu de tipus, 405
- Gauss-Txebixev, fórmula, 275
- Gauss-Txebixev, fórmula de, 279, 309, 339
- gaussiana amb pivotatge, eliminació, 67
- gaussiana, eliminació, 56, 57, 59, 70, 126, 138
- gaussiana, error d'una fórmula, 277, 278
- gaussiana, exactitud d'una fórmula, 278
- gaussiana, fórmula, 274–278, 307, 308
- gaussianes, fórmules, 274
- gaussians, mètodes, 58, 69
- Gerschgorin, teorema de, 54, 110
- Givens, mètode de, 81
- Givens, mètode de, 88, 89
- globalment convergent, mètode, 364
- Graeffe, mètode de, 369, 372
- Gram, polinomis de, 178
- Gram-Schmidt, mètode
 - d'ortogonalització de, 173
- Gram-Schmidt, mètode
 - d'ortogonalització de, 130, 169, 171, 172, 174, 175, 238, 244
- Gram-Schmidt, mètode
 - d'ortogonalització modificat de, 173
- Gram-Schmidt, mètode
 - d'ortogonalització modificat de, 170, 172
- Gregory, fórmula de, 348
- Haar, propietat de, 147
- Haar, propietat de, 174, 183, 184
- Halley, fórmula de, 400
- Hankel, matriu de, 258
- heptadiagonal, matriu, 128
- Hermite generalitzada, error en la interpolació d', 229
- Hermite generalitzada, interpolació d', 158, 228–230, 270
- Hermite generalitzada, polinomi d'interpolació d', 159
- Hermite generalitzada, polinomi interpolador d', 401
- Hermite, error de la interpolació d', 157
- Hermite, error en la interpolació, 299
- Hermite, error en la interpolació d', 158, 228
- Hermite, funció generatriu dels polinomis de, 254
- Hermite, interpolació d', 157, 158, 227, 270
- Hermite, polinomi d'interpolació d', 227
- Hermite, polinomi interpolador d', 157–159, 255, 299, 301–303, 401, 402
- Hermite, polinomi interpolador de, 278
- Hermite, polinomis bàsics de la interpolació d'Hermite, 303
- Hermite, polinomis d', 209
- Hermite, problema d'interpolació d', 157
- Hermite, recurrència dels polinomis d', 254, 259
- hermítica, matriu, 50, 55, 73
- Hessenberg inferior, matriu, 50
- Hessenberg superior, matriu, 50, 74, 81, 86, 88, 137
- Hessenberg, reducció a matriu, 89
- Horner, regla de, 32, 361, 363, 383
- Hotteling, deflació de, 74
- Householder, deflació de, 74, 111
- Householder, matriu, 118, 122
- Householder, matriu de, 61, 62, 66, 67, 76, 82, 83, 111, 143, 170, 239
- Householder, matrius de, 67
- Householder, mètode d'ortogonalització de, 67
- Householder, mètode d'ortogonalització,

- 64
- Householder, mètode d'ortogonalització de, 130, 169, 170, 239
- Householder, mètode de, 88, 89, 144
- Hölder, norma de, 52, 53
- independent, vector de termes, 50, 57, 58
- indepents, variable aleatòries, 9
- integració adaptable, 296
- integració numèrica, fórmula d', 275
- integració interpolatòria, fórmula d', 265, 266, 270, 274
- integració numèrica, 265, 295
- integració numèrica, fórmula d', 266, 270, 271, 274–276, 289, 296
- integració numèrica, fórmula de, 265, 301
- integral impròpia, 285
- interpolació, 145, 293, 358
- interpolació directe, mètode d', 358, 359
- interpolació invers, mètode de, 358, 360
- interpolació lineal, 351
- interpolació polinòmica, 146
- interpolació polinòmica d'una funció, 146
- interpolació, abscisses d', 146
- interpolació, concepte d', 145
- interpolació, error d', 148, 149, 152, 154, 159, 160
- interpolació, fórmula d', 289, 294
- interpolació, problema d', 145
- interpolació, punts d', 146
- interpolador, polinomi, 149, 151, 152, 154
- interpolador, càlcul del polinomi, 149
- interpolador, polinomi, 148, 160
- interval singular, 163
- interval, funcions entre, 7
- interval, operació aritmètica entre, 7
- inverses, càlcul de la matrius, 67
- iteració simple, 354, 357, 358, 364
- iteració simple en diverses variables, mètode d', 370
- iteració simple, mètode d', 354, 397
- iteració simple, ordre de convergència del mètode d', 356
- iteració, matriu d', 70, 72
- Jacobi amb llinars, mètode cíclic de, 81
- Jacobi, mètode clàssic de, 81
- Jacobi, mètode cíclic de, 81
- Jacobi, mètode de, 81, 82
- Jacobi, mètode iteratiu de, 72
- Jacobi, mètode de, 79
- Jacobi, mètode clàssic de, 115
- Jacobi, mètode de, 89, 115
- Jacobi, mètode iteratiu de, 71, 72, 99, 138, 139, 141
- Jacobi, mètode iteratiu de tipus, 405
- Jacobi, variant del mètode de, 80
- Kronecker, delta de, 49
- Lagrange, polinomis de, 196
- Lagrange, fórmula d'interpolació de, 149, 207, 266
- Lagrange, mètode de, 149, 190, 191, 193, 251
- Lagrange, polinomi de, 149
- Lagrange, polinomis de, 190, 191, 194–197
- Laguerre, mètode de, 364, 372, 404
- Laguerre, ordre de convergència del mètode de, 364
- Laguerre, polinomis de, 259
- Laguerre, recurrència dels polinomis d', 259
- Laguerre-Thibault, regla de, 362
- Lanczos, economitxació de, 187, 247, 248
- Legendre, polinomi de, 280
- Legendre, polinomis de, 177, 179, 310, 312
- Legendre, recurrència dels polinomis d', 254
- Legendre, zeros dels polinomis de, 311
- Leslie, matriu de, 142
- Lobatto, fórmula d'integració de tipus, 334
- localització d'arrels, 349, 350
- localització de valors propis, 89
- localització de zeros complexos, 372
- LR, factorització, 87
- LR, mètode, 89
- LR, mètode iteratiu, 87, 88, 144
- LU, factorització, 58, 68, 70, 87, 88, 127, 129, 130, 137

LU, mètode, 56, 58, 59, 127, 128

límit central, teorema de, 9

mal condicionada, matriu, 69

mantissa, 2

matricial, sèrie, 135, 136

matriu banda, 128

matriu bidiagonal, 132

matriu d'iteració, 70, 72, 101

matriu d'un sistema, 50

matriu de Householder, 143

matriu definida positiva, 50, 53, 59, 73, 93, 129, 131, 132, 168

matriu diagonal, 50, 54, 55, 59, 79, 81, 88, 134

matriu diagonalitzable, 50, 54, 55, 76, 88

matriu escassa, 89

matriu estrictament diagonal dominant, 50, 54, 72

matriu heptadiagonal, 128

matriu hermítica, 50, 55, 73

matriu Hessenberg inferior, 50

matriu Hessenberg superior, 50, 74, 81, 86, 88, 137

matriu mal condicionada, 69

matriu ortogonal, 50, 55, 60, 61, 69, 74, 75, 79–82

matriu pentadiagonal, 50, 128, 132

matriu per blocs, 131, 139

matriu quasi-tridiagonal, 129

matriu regular, 50, 54, 58, 61, 88

matriu semidefinida positiva, 165

matriu simètrica, 50, 55, 61, 64, 74–77, 79, 81, 83, 88, 128, 129

matriu singular, 49

matriu triangular, 135

matriu triangular inferior, 50, 58, 59, 71, 87

matriu triangular superior, 50, 53, 56–58, 60, 65–67, 71, 87, 88

matriu tridiagonal, 50, 73, 128, 132, 137, 143

matriu tridiagonal per blocs, 73, 128

matriu tridiagonal simètrica, 81–84, 88, 143, 144

matriu unitària, 50, 53, 55, 69, 134

matrius semblants, 50, 54

mesuraments incorrectes, 1

mitjana, 7–9, 167

Moivre, fórmula de, 180, 275

Muller-Traub, mètode de, 359, 367, 372, 382, 404

mètode de Householder, 117

mètode de substitució cap al darrera, 56, 57

mètode de substitució cap al davant, 56

Neville, mètode de, 149, 150, 152, 251

Newton amb dues variables, mètode de, 406

Newton amb dues variables, mètode de, 394

Newton cap al darrera, fórmula d'interpolació de, 294

Newton cap al davant, fórmula d'interpolació de, 293

Newton en dues variables, mètode de, 368, 370

Newton modificat per a zeros múltiples, mètode de, 376

Newton, mètode de, 401, 402

Newton, fórmula d'interpolació de, 152

Newton, fórmula d'interpolació de les diferències dividides de, 291

Newton, fórmula de les diferències dividides, 323

Newton, mètode de, 352

Newton, mètode de, 350–352, 356, 358, 359, 364, 367, 371, 372, 382, 398, 399, 403

Newton, mètode de les diferències dividides, 152, 205

Newton, mètode de les diferències dividides de, 151, 158, 191, 193, 198, 251, 260

Newton, ordre del mètode de, 356

Newton, regla de, 362

Newton, variant del mètode de, 371, 400

Newton-Cotes obertes, fórmules de, 335

Newton-Cotes, expressió de l'error de la fórmula de, 269

Newton-Cotes, fórmula de, 269, 296, 321,

- 322
- Newton-Cotes, fórmules de, 296
- Newton-Cotes, fórmules de, 335
- nombre d'operacions, 89, 128–132
- nombre de condició, 69, 98, 135–137
- nombre de condició euclidià, 69
- norma matricial subordinada a una
norma vectorial, 55
- norma de Hölder, 52, 53
- norma del màxim, 53, 162, 163, 246
- norma del màxim ponderada, 162, 163
- norma euclidiana, 162–164, 171, 174,
180, 233
- norma euclidiana ponderada, 163
- norma matricial, 53, 55
- norma matricial consistent amb una
norma vectorial, 53, 70, 97
- norma matricial subordinada a una
norma vectorial, 53, 70, 95,
134–136
- norma multiplicativa, 53
- norma suma de mòduls, 53
- norma vectorial euclidiana, 53, 61
- operacions, nombre d', 57
- operacions, nombre d', 56, 62–64, 67, 69,
84, 85, 88
- operació, 56
- operació aritmètica, 3–5
- operació aritmètica entre intervals, 7
- operador de derivació, 289
- operador d'integració, 289
- operador desplaçament cap endavant,
289
- operador diferència cap endarrera, 289
- operador diferència cap endavant, 289
- operador diferència centrada, 289
- operador identitat, 290
- operador mitjana, 289
- operadors, fórmules amb, 296
- operadors, propietats dels, 290
- ordinador vectorial, 129
- ordre de convergència, 355, 356, 372
- ortogonal, matriu, 50, 55, 60, 61, 69, 74,
75, 79–82
- ortogonal, vector, 164
- paritat definida, 186
- pentadiagonal, matriu, 50, 128, 132
- pesos, 162, 163, 166, 170, 173, 176
- pesos d'integració, 266, 277
- pesos de fórmules gaussianes, 278
- pesos de Gauss-Txebixev, 279
- pesos simètrics, 176
- pesos, matriu de, 170, 171
- pivot, 57, 58, 68, 70, 130
- pivotatge, 58
- pivotatge complet, 58, 70, 126, 138
- pivotatge maximal per columnes, 58, 70,
126, 137, 138
- polinomi característic, 51, 104, 105, 108,
123
- polinomi interpolador, 148, 149, 151,
152, 154, 160, 262, 263, 265,
266, 270, 274
- polinomi interpolador, càlcul del, 149
- polinomi mònic, 175, 186
- polinomi mònic de Txebixev, 186
- polinomi ortogonal, 307
- polinomi trigonomètric, 182–185, 274
- polinomi trigonomètric en cosinus, 185,
275
- polinomis ortogonals, 169, 174, 175,
177–179, 186, 189, 242
- polinomis ortogonals mònics, 234, 235
- polinomis ortogonals, fórmula recurrent
de, 243
- polinomis ortogonals, fórmula recurrent
dels, 234, 235
- polinomis ortogonals, relació recurrent
dels, 175, 176
- polinomis trigonomètrics, 180
- polinomis trigonomètrics ortogonals,
169, 179
- polinòmica d'una funció, interpolació,
146
- polinòmica, interpolació, 146
- potència desplaçada, mètode de la, 78
- potència inversa desplaçada, mètode de
la, 78
- potència inversa, mètode de la, 78, 86,
87, 113
- potència, mètode de, 89

- potència, mètode de la, 76, 113, 142, 143
- potència, variants del mètode de la, 78
- precisió doble, 16
- precisió simple, 16
- precisió doble, 132
- problema general d'aproximació, 159, 161
- producte d'operadors, 290
- producte escalar, 163, 164, 169, 174, 175, 233
- producte escalar continu, 182, 185, 186
- producte escalar discret, 177, 182, 185, 186, 242
- projecció ortogonal, propietat de, 164
- propagació d'errors en sistemes lineals, 89
- propagació de l'error estàndard, fórmula de, 9
- propagació de l'error maximal, fórmula de, 4–6
- propagació de l'error, fórmula de, 4, 9
- punt fix, 354, 355, 398
- punt mitjà, 167
- punts 2-periòdics, 405

- Q-D, mètode, 372
- QR, factorització, 60, 64, 68, 87–89, 170, 171, 173, 238, 239
- QR, mètode, 60, 67, 69, 89
- QR, mètode iteratiu, 87, 88, 144
- quadratura, 265
- quasi-tridiagonal, matriu, 129
- quocient diferència (Q-D), mètode de, 387
- quocient-diferència (Q-D), mètode del, 365

- radi espectral, 51, 55, 70, 72, 101, 103, 110, 111
- Rayleigh, quocient de, 77
- Rayleigh, quocients de, 114
- rectangle, fórmula del, 268
- recurrència inestable, 40
- reflexió respecte a un hiperplà, 61, 62
- regió de convergència, 281
- regla de Horner, 147, 182
- regles d'integració numèrica, convergència, 296
- regles de fitació de zeros de polinomis, 362
- regles de localització de zeros de polinomis, 362
- regressió, coeficient de, 168
- regressió, recta de, 168
- regula falsi, mètode de la, 352, 353, 364, 372
- regula falsi, ordre del mètode de la, 356
- regular, matriu, 50, 54, 58, 61, 88
- relaxació, factor de, 72
- relaxació, factor òptim de, 73
- Remes, algorismes de, 189
- Remes, mètodes de, 184
- representació de nombres, 10
- representació en punt flotant, 2
- residu d'un desenvolupament, 282
- residu d'una sèrie, 282–284, 286
- residus vectorials, 137
- resolució directa de sistemes lineals, 88
- resolució iterativa de sistemes lineals, 88
- Richardson, mètode de, 288
- Richardson, mètode de, 264, 318, 321
- Riemann, zeta de, 315
- Romberg, mètode de, 319, 346
- rombes, regles dels, 365
- Ruffini, regla de, 25
- Runge, fenomen de, 252

- Schur, lema de, 53
- secant, mètode de la, 352
- secant, mètode de, 121, 382
- secant, mètode de la, 351–353, 356, 359, 364, 367, 372
- secant, ordre de convergència del mètode de, 368
- secant, ordre del mètode de la, 356
- secant, variant del mètode de la, 352
- semblants, matrius, 50, 54
- semidefinida positiva, matriu, 165
- separació d'arrels, 349, 350
- separació de zeros, 360, 362
- Sherman-Morrison, fórmula de, 133
- simetria respecte a un hiperplà, 61

- Simpson amb termes correctius, regla de, 347
- Simpson, error de la fórmula de, 268
- Simpson, fórmula de, 266–269, 296, 321
- Simpson, regla de, 271, 321, 338
- simètrica, matriu, 50, 55, 61, 64, 74–77, 79, 81, 83, 88, 128, 129
- singular, matriu, 49
- sistema compatible, 50
- sistema determinat, 50
- sistema triangular, 58, 60, 69
- sistema triangular inferior, 56
- sistema triangular superior, 56, 57
- sistema, solució aproximada d'un, 69
- sistema, solució d'un, 50, 51, 58, 68, 70
- sistema, solució directa d'un, 56
- sistema, solució iterativa d'un, 56
- sistemes d'equacions no lineals, 372
- sobredeterminat, problema, 160
- sobredeterminat, sistema lineal, 169, 171, 238
- sobredeterminats, sistemes lineals, 256
- sobrerelaxació, mètode iteratiu de, 72
- sobrerelaxació, mètode iteratiu, 139
- sobrerelaxació, mètode iteratiu de, 72, 73
- solució aproximada d'un sistema, 69
- solució d'un sistema, 50, 51, 58, 68, 70
- solució d'una equació no lineal, 349
- solució directa d'un sistema, 56
- solució iterativa d'un sistema, 56
- Steffensen, mètode d'acceleració de, 357
- Steffensen, mètode de, 371, 374
- Steffensen, ordre de convergència del mètode de, 375
- Stirling, fórmula d'interpolació de, 323, 324
- Stirling, fórmula de, 344
- Sturm, mètode de, 362, 363, 404
- Sturm, successió de, 120, 362, 363
- Sturm, teorema de, 120, 144
- submatriu principal, 49
- subspai ortogonal a un vector, 61
- subspai propi, 51
- substitució cap al darrera, mètode de, 56, 57, 60, 93
- substitució cap al davant, mètode de, 56
- suma d'operadors, 290
- suma d'una sèrie, 280, 284, 286, 287
- suma parcial, 280
- suma telescòpica, 287
- suma trigonomètrica, 179, 180
- sumació numèrica, 280
- sumació numèrica, mètodes de, 284
- sumació, mètodes de, 296
- Sylvester, criteri de, 53, 90, 91, 94
- sèrie alternada, 283
- sèrie d'operadors, 290
- sèrie de funcions, 281
- sèrie de potències, 281
- sèrie lentament convergent, 284, 287
- sèrie matricial, 135, 136
- sèrie més ràpidament convergent, 287
- sèrie semiconvergent, 282, 286, 318
- sèrie telescòpica, 314
- tall, 2
- Taylor directe, mètode de, 358
- Taylor directe, ordre de convergència d'un mètode de, 359
- Taylor invers, mètode de, 358
- Taylor usant operadors, fórmula de, 292
- Taylor, desenvolupament de, 154, 156, 210, 214, 215, 247–249, 252–254, 261, 306, 357, 390
- Taylor, desenvolupament de , 155
- Taylor, desenvolupament en sèrie de, 156, 281
- Taylor, expressió de Lagrange de l'error d'interpolació de, 154, 212, 213
- Taylor, expressió integral de l'error de la interpolació de, 153
- Taylor, fórmula d'interpolació de, 153
- Taylor, fórmules d'integració numèrica de, 336
- Taylor, interpolació de, 152
- Taylor, mètode de, 357
- Taylor, polinomi de, 210–212, 252, 255, 358, 400
- Taylor, polinomi interpolador de, 153, 154
- Taylor, problema d'interpolació de, 153

- Taylor, residu de desenvolupament de, 154, 155
- teorema central del límit, 10
- teorema de Bolzano, 350
- teorema de Pitàgores, 164
- teorema de Rolle, 149
- teorema de Sturm, 89, 363, 384
- teorema de Txebeixev, 183, 185, 245
- teorema del valor mitjà per a sumes, 296
- teorema del valor mitjà, 268
- teorema del valor mitjà per a integrals, 153, 268, 302
- teorema del valor mitjà per a sumes, 264, 270
- teorema fonamental del càlcul, 153
- teranyina, procés, 354
- $\text{tr } A$, traça de la matriu A , 54
- transformacions de semblança, 81, 82, 84
- transformació de semblança, 50, 54, 55, 74
- transformació de semblança, 104, 118, 122
- trapezi, fórmula del, 268–270, 272, 273, 296
- trapezis amb termes correctius, regla dels, 347
- trapezis, regla dels, 270, 274, 318, 320, 330, 331, 338, 339, 345, 346
- traça, 54
- triangular inferior, matriu, 50, 58, 59, 71, 87
- triangular inferior, sistema, 56
- triangular superior, matriu, 50, 53, 56–58, 60, 65–67, 71, 87, 88
- triangular superior, sistema, 56, 57, 126
- triangular, matriu, 135
- triangular, sistema, 58, 60, 69
- triangularització, 92
- tridiagonal per blocs, matriu, 73, 128
- tridiagonal simètric, sistema, 121
- tridiagonal simètrica, matriu, 81–84, 88, 117, 120, 122, 143, 144
- tridiagonal, matriu, 50, 73, 128, 132, 137, 143
- trigonomètrica, aproximació, 161, 169, 179, 180, 182
- Txebeixev, abscisses de, 249
- Txebeixev, coeficients ortogonals de, 260
- Txebeixev, error en la interpolació de, 250
- Txebeixev, funció generatriu dels polinomis de, 254
- Txebeixev, interpolació de, 188, 249, 250, 260, 261
- Txebeixev, modificació del mètode de, 402
- Txebeixev, mètode de, 359, 373, 399
- Txebeixev, ordre del mètode de, 374
- Txebeixev, polinomi, 246
- Txebeixev, polinomi de, 122, 123, 248, 249, 279
- Txebeixev, polinomi mònic, 246
- Txebeixev, polinomi mònic de, 123
- Txebeixev, polinomis de, 185, 187, 235, 236
- Txoleski, factorització de, 59, 92, 94, 128, 133
- Txoleski, mètode de, 90, 128, 168
- Txolesky, mètode de, 56
- unitària, matriu, 50, 53, 55, 69, 134
- valor propi, 51, 54, 73, 74, 88
- Vandermonde, determinant de, 148
- variància, 7
- vector de termes independents, 50, 57, 58
- vector propi, 51, 52, 54, 73, 74
- vector propi generalitzat, 142
- vector propi per l'esquerra, 54
- vector propi per la dreta, 54
- vectorial instrucció, 129
- vectors propis ortonormals d' A , base de, 55
- vectors propis per l'esquerra, base de, 54
- vectors propis per la dreta, base de, 54
- Wielandt, deflació de, 74
- xifra, 2
- zero, 349
- zero d'un polinomi, 147, 182
- zeros d'un polinomi de Txebeixev, 186