

(curs 7 - 54)

5.2.2. Liniarizarea datelor

- creșterea exponentială a unei populații poate fi dedusă atunci când rata de creștere este proporțională cu mărimea populației
- în condiții perfecte, atunci când mediul de creștere este neschimbat și când populația este mult sub capacitatea maximă a mediului, modelul este o reprezentare bună
- **modelul exponentzial**

$$y = c_1 e^{c_2 t} \quad (1)$$

nu poate fi interpolat direct de cele mai mici pătrate, deoarece c_2 nu apare liniar în ecuația modelului

- imediat ce punctele sunt substituite în model, dificultatea este clară: ecuațiile care trebuie rezolvate pentru a găsi coeficienții sunt neliniare și nu pot fi exprimate ca un sistem liniar $Ax = b$
- prin urmare, deducerea ecuațiilor normale este irelevantă
- există două moduri de a trata problema coeficienților neliniari
- modul mai dificil este de a minimiza direct eroarea în sensul celor mai mici pătrate, și anume, de a rezolva problema de tip cele mai mici pătrate neliniară
- vom reveni la această problemă în Secțiunea 5.4
- modul mai simplu este de a schimba problema

Aplicăm ln la (1)

$$\ln y = \ln c_1 e^{c_2 t} = \ln c_1 + c_2 t$$

- observăm că pentru modelul exponentzial, graficul lui $\ln y$ este un grafic liniar în t
- la prima vedere, pare că doar am schimbat o problemă cu o alta
- coeficientul c_2 este acum liniar în model, dar c_1 nu mai este
- totuși, dacă renotăm $k = \ln c_1$, putem scrie

$$\ln y = k + c_2 t. \quad (3)$$

- acum ambii coeficienți k și c_2 sunt liniari în model
- după rezolvarea ecuațiilor normale pentru a găsi cele mai bune valori k și c_2 , putem afla valoarea corespunzătoare $c_1 = e^k$ dacă dorim acest lucru
- trebuie să observăm că modul de a ieși din dificultatea coeficienților neliniari a fost să schimbăm problema
- problema inițială de tip cele mai mici pătrate a fost să interpolăm datele folosind (1)—și anume, de a găsi c_1 , c_2 care minimizează

$$(c_1 e^{c_2 t_1} - y_1)^2 + \dots + (c_1 e^{c_2 t_m} - y_m)^2, \quad (4)$$

suma pătratelor rezidualilor ecuațiilor $c_1 e^{c_2 t_i} = y_i$ pentru $i = 1, \dots, m$

- deocamdată, rezolvăm problema revizuită a minimizării erorii în sensul celor mai mici pătrate în „spațiul logaritm”—și anume, aceea de a găsi c_1 , c_2 care minimizează

$$\rightarrow (\ln c_1 + c_2 t_1 - \ln y_1)^2 + \dots + (\ln c_1 + c_2 t_m - \ln y_m)^2, \quad (5)$$

suma pătratelor rezidualilor ecuațiilor $\ln c_1 + c_2 t_i = \ln y_i$ pentru $i = 1, \dots, m$

- acestea sunt două minimizări diferite și au soluții diferite, ceea ce înseamnă că în general dau valori diferite pentru coeficienții c_1 , c_2
- care metodă este corectă pentru această problemă, metoda neliniară a celor mai mici pătrate din (4) sau versiunea liniarizată din (5)?
- prima este metoda celor mai mici pătrate, după cum am definit-o
- cea de-a doua nu este
- totuși, depinzând de contextul datelor, oricare dintre cele două poate fi alegerea mai naturală

-
- pentru a răspunde la întrebare, utilizatorul trebuie să se decidă care erori sunt cel mai important de minimizat, erorile în sensul inițial sau erorile în „spațiul logaritm”
 - de fapt, modelul logaritm este liniar, și s-ar putea argumenta că, doar după transformarea logaritmică a datelor astfel încât să aibă o relație liniară, este natural să evaluăm acuratețea modelului

Exemplul 1

- folosiți liniarizarea modelului pentru a găsi cea mai bună interpolare exponentială în sensul celor mai mici pătrate $y = c_1 e^{c_2 t}$ pentru numărul total de mașini existente în lume:

an	mașini ($\times 10^6$)
1950	53.05
1955	73.04
1960	98.31
1965	139.78
1970	193.48
1975	260.20
1980	320.39

- datele descriu numărul de automobile care se află în exploatare în lume într-un anumit an
- definim variabila de timp t în termeni de ani din 1950
- rezolvând problema de tip cele mai mici pătrate liniară, obținem $k \approx 3.9896$, $c_2 \approx 0.06152$
- deoarece $c_1 \approx e^{3.9896} \approx 54.03$, modelul este $y = 54.03e^{0.06152t}$
- valoarea REMP-ului modelului liniarizat în spațiul logaritm este ≈ 0.0357 , în timp ce REMP-ul modelului exponential inițial este ≈ 9.56
- cel mai bun model împreună cu datele sunt reprezentate grafic în Figura 1

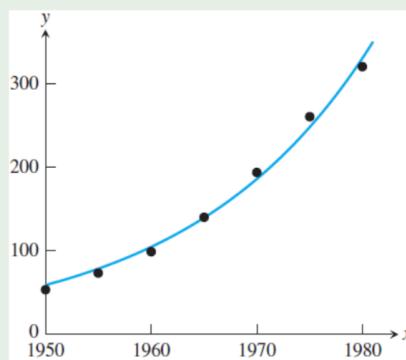


Figura 1: Interpolare exponentială pentru numărul total de mașini existente în lume, folosind liniarizarea. Cea mai bună interpolare în sensul celor mai mici pătrate este $y = 54.03e^{0.06152t}$. Comparați cu Figura 7.

```
% Nr de masini
data = [53.05; 73.04; 98.31; 139.78; 193.48; 260.20; 320.39];
```

```
% Anii
years = [1950; 1955; 1960; 1965; 1970; 1975; 1980];
```

```
Problema_5.m Problema_7.m Exemplu_1.curs.m +
1 % Datele pentru numărul de mașini și anii corespunzători
2 data = [53.05; 73.04; 98.31; 139.78; 193.48; 260.20; 320.39];
3 years = [1950; 1955; 1960; 1965; 1970; 1975; 1980];
4
5 t0 = years;
6 y0 = data;
7
8 % Construirea matricei A și a vectorului b
9 A = [t0.^1, t0.^2];
10 b = log(y0);
11
12 % Aproximarea modelului folosind metoda celor mai mici pătrate
13 x = (A' * A) \ (A' * b);
14 c1 = exp(x(1));
15 c2 = x(2);
16
17 r = b - A * x;
18
19 REMP = norm(r) / (sqrt(length(r)));
20
21 % Calcularea valorii aproximative y_aprox pentru modelul exponential
22 y_aprox = c1 * exp(c2 * t0);
23
24 % Calcularea reziduurilor și a REMP pentru modelul exponential
25 r1 = y0 - y_aprox;
26 REMP2 = norm(r1) / sqrt(length(r1));
```

Command Window

c1 =

54.0315

c2 =

0.0615

REMP =

0.0357

REMP2 =

9.5552

Exemplul 2

Exemplul 2

- numărul de tranzistoare din CPU-urile Intel de la începutul anilor 1970 sunt date în tabelul următor; interpolați modelul $y = c_1 e^{c_2 t}$ pentru aceste date

CPU	an	tranzistoare
4004	1971	2,250
8008	1972	2,500
8080	1974	5,000
8086	1978	29,000
286	1982	120,000
386	1985	275,000
486	1989	1,180,000
Pentium	1993	3,100,000
Pentium II	1997	7,500,000
Pentium III	1999	24,000,000
Pentium 4	2000	42,000,000
Itanium	2002	220,000,000
Itanium 2	2003	410,000,000

- parametrii vor fi găsiți folosind liniarizarea modelului (2)
- liniarizarea modelul ne dă

$$\ln y = k + c_2 t.$$

- vom considera că $t = 0$ corespunde anului 1970
- înlocuind datele în modelul liniarizat, obținem

$$\begin{aligned} k + c_2(1) &= \ln 2250 \\ k + c_2(2) &= \ln 2500 \\ k + c_2(4) &= \ln 5000 \\ k + c_2(8) &= \ln 29000, \end{aligned} \quad (6)$$

și aşa mai departe

- ecuația matricială este $Ax = b$, unde $x = [k, c_2]^T$,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 8 \\ \vdots & \vdots \\ 1 & 33 \end{bmatrix}, \quad b = \begin{bmatrix} \ln 2250 \\ \ln 2500 \\ \ln 5000 \\ \ln 29000 \\ \vdots \\ \ln 410000000 \end{bmatrix}. \quad (7)$$

- ecuațiile normale $A^T Ax = A^T b$ sunt

$$\begin{bmatrix} 13 & 235 \\ 235 & 5927 \end{bmatrix} \begin{bmatrix} k \\ c_2 \end{bmatrix} = \begin{bmatrix} 176.90 \\ 3793.23 \end{bmatrix},$$

care au soluția $k \approx 7.197$ și $c_2 \approx 0.3546$, ceea ce conduce la $c_1 = e^k \approx 1335.3$

- curba exponentială $y = 1335.3e^{0.3546t}$ este prezentată în Figura 2 împreună cu datele

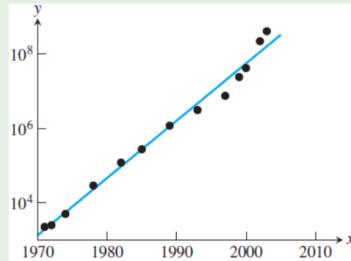


Figura 2: Graficul semilog al legii lui Moore. Numărul de tranzistoare dintr-un CPU pe an.

- timpul de dublare pentru această lege este $\ln 2/c_2 \approx 1.95$ ani
- Gordon C. Moore, cofondator Intel, a prezis în 1965 că pe parcursul deceniului următor, puterea de procesare se va dubla la fiecare 2 ani
- în mod surprinzător, această rată exponențială a continuat timp de 40 de ani

- un alt exemplu important cu coeficienți neliniari este modelul **legea de putere** $y = c_1 t^{c_2}$
- acest model poate fi de asemenea simplificat utilizând liniarizarea, luând logaritmul ambelor părți:

$$\begin{aligned} \ln y &= \ln c_1 + c_2 \ln t \\ &= k + c_2 \ln t. \end{aligned} \quad (8)$$

- înlocuirea datelor în acest model ne va da

$$\begin{aligned} k + c_2 \ln t_1 &= \ln y_1 \\ &\vdots \\ k + c_2 \ln t_m &= \ln y_m, \end{aligned} \quad (9)$$

ceea ce rezultă în forma matricială

$$A = \begin{bmatrix} 1 & \ln t_1 \\ \vdots & \vdots \\ 1 & \ln t_m \end{bmatrix}, \quad b = \begin{bmatrix} \ln y_1 \\ \vdots \\ \ln y_m \end{bmatrix}. \quad (10)$$

- ecuațiile normale permit determinarea lui k și c_2 , și $c_1 = e^k$

Exemplul 3

- folosiți liniarizarea pentru a interpola dependența greutate–înălțime folosind un model de tip lege de putere
- greutatea și înălțimea medie a copiilor de 2–11 ani sunt date în tabelul de mai jos:

vârstă (ani)	înălțimea (m)	greutatea (kg)
2	0.9120	13.7
3	0.9860	15.9
4	1.0600	18.5
5	1.1300	21.3
6	1.1900	23.5
7	1.2600	27.2
8	1.3200	32.7
9	1.3800	36.0
10	1.4100	38.6
11	1.4900	43.7

- urmând strategia precedentă, legea de putere rezultată pentru greutate versus înălțime este $W = 16.3H^{2.42}$

- această relație este reprezentată grafic în Figura 3
- deoarece greutatea este un indicator al volumului, coeficientul $c_2 \approx 2.42$ poate fi considerat ca fiind „dimensiunea efectivă” a corpului uman

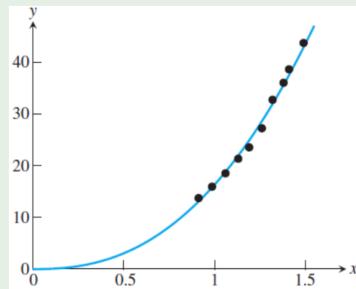


Figura 3: Legea de putere pentru dependența greutate–înălțime pentru copii de 2–11 ani. Cea mai bună formulă este $W = 16.3H^{2.42}$.

- concentrația de medicament y din fluxul sanguin este bine descrisă de modelul

$$y = c_1 t e^{c_2 t} \quad (11)$$

unde t denotă timpul scurs de la administrarea medicamentului

- caracteristicile modelului sunt date de o creștere rapidă pe măsură ce medicamentul intră în fluxul sanguin, urmată de o ușoară descreștere exponentială
- tempul de înjumătățire** al medicamentului este timpul scurs de când are loc concentrația maximă până când aceasta scade la jumătate din acea valoare
- modelul poate fi liniarizat prin aplicarea logaritmului natural în ambele părți, ceea ce ne dă

$$\begin{aligned} \ln y &= \ln c_1 + \ln t + c_2 t \\ k + c_2 t &= \ln y - \ln t, \end{aligned}$$

unde am notat $k = \ln c_1$

- aceasta ne conduce la ecuația matricială $Ax = b$, unde

$$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \quad b = \begin{bmatrix} \ln y_1 - \ln t_1 \\ \vdots \\ \ln y_m - \ln t_m \end{bmatrix}. \quad (12)$$

- ecuațiile normale se pot rezolva pentru a găsi k și c_2 , și $c_1 = e^k$

Exemplul 4

- interpolați modelul (11) pentru nivelul măsurat al concentrației medicamentului norfluoxetine din fluxul sanguin al unui pacient, dat în tabelul următor:

ora	concentrația (ng/ml)
1	8.0
2	12.3
3	15.5
4	16.8
5	17.1
6	15.8
7	15.2
8	14.0

- rezolvând ecuațiile normale, obținem $k \approx 2.28$ și $c_2 \approx -0.215$, și $c_1 \approx e^{2.28} \approx 9.77$
- cea mai bună versiune a modelului este $y = 9.77te^{-0.215t}$, reprezentată grafic în Figura 4
- din acest model, ora la care se înregistrează concentrația maximă și timpul de înjumătățire al medicamentului pot fi estimate

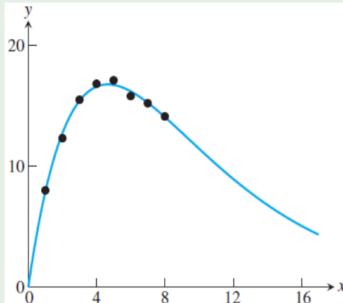


Figura 4: Reprezentarea grafică a concentrației de medicament din fluxul sanguin. Modelul (11) arată descreșterea exponențială după creșterea initială.

- este important să înțelegem că liniarizarea modelului schimbă problema de tip cele mai mici pătrate
- soluția obținută va minimiza REMP-ul pentru problema liniarizată, nu neapărat pentru problema inițială, care în general va avea un set diferit de parametri optimi
- dacă apar în model neliniar, atunci parametrii nu pot fi calculați din ecuațiile normale, și avem nevoie de tehnici neliniare de a rezolva problema de tip cele mai mici pătrate inițială
- acest lucru poate fi făcut folosind metoda Gauss–Newton din Secțiunea 5.4, unde vom discuta din nou problema numărului de mașini existente în lume și vom compara interpolarea realizată de modelul exponențial atât în forma liniarizată, cât și în forma neliniarizată

5.3. Factorizarea QR

→ LU folosită pt. a rezolva ecuații matriciale

→ factorizarea este utilă, deoarece oferă o reprezentare compactă a pasilor eliminării gauziene

→ Factorizarea QR → modalitate de a rezolva probleme de tip cele mai mici pătrate, care este supersetă ecuațiilor normale

5.3.1. Ortogonalizarea Gramm–Schmidt și cele mai mici pătrate

- metoda Gram–Schmidt ortogonalizează o mulțime de vectori
- dându-se o mulțime de vectori m -dimensionali, scopul este de a găsi un sistem de coordonate ortogonal pentru subspațiul liniar generat de această mulțime de vectori
- mai precis, dându-se n vectori liniar independenți, metoda determină n vectori unitari perpendiculari care generează același subspațiu ca vectorii date
- lungimea unitară este în funcție de norma euclidiană sau 2-norma, care este folosită în cuprinsul Capitolului 5
- fie A_1, \dots, A_n vectori liniar independenți din \mathbb{R}^m
- rezultă $n \leq m$
- metoda Gram–Schmidt începe prin împărțirea lui A_1 la lungimea lui, pentru a-l face vector unitar

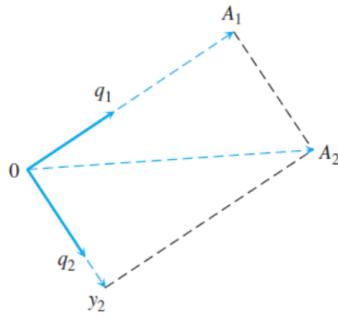


Figura 5: Ortogonalizarea Gram–Schmidt. Vectorii dați sunt A_1 și A_2 , iar ieșirea este mulțimea ortonormală constând din q_1 și q_2 . Al doilea vector ortogonal q_2 este format prin scăderea proiecției lui A_2 în direcția lui q_1 din A_2 , urmată de normalizare.

- definim

$$y_1 = A_1 \text{ și } q_1 = \frac{y_1}{\|y_1\|_2}. \quad (13)$$

- pentru a determina al doilea vector unitar, scădem din A_2 proiecția lui A_2 în direcția lui q_1 , și normalizăm rezultatul:

$$y_2 = A_2 - q_1(q_1^T A_2) \text{ și } q_2 = \frac{y_2}{\|y_2\|_2}. \quad (14)$$

- atunci $q_1^T y_2 = q_1^T(A_2 - q_1(q_1^T A_2)) = q_1^T A_2 - q_1^T q_1 A_2 = 0$, deci q_1 și q_2 sunt ortogonali, după cum se arată în Figura 5

- la pasul j , definim

$$y_j = A_j - q_1(q_1^T A_j) - q_2(q_2^T A_j) - \cdots - q_{j-1}(q_{j-1}^T A_j) \text{ și } q_j = \frac{y_j}{\|y_j\|_2}. \quad (15)$$

- este clar că q_j este ortogonal pe fiecare dintre vectorii anteriori q_i pentru $i = 1, \dots, j-1$, deoarece (15) implică

$$\begin{aligned} q_i^T y_j &= q_i^T A_j - q_i^T q_1 q_1^T A_j - \cdots - q_i^T q_{j-1} q_{j-1}^T A_j \\ &= q_i^T A_j - q_i^T q_i q_i^T A_j = 0, \end{aligned}$$

- unde, prin ipoteza de inducție, vectorii q_i sunt ortogonali doi câte doi pentru $i < j$
- geometric, (15) corespunde scăderii din A_j a proiecțiilor lui A_j pe vectorii ortogonali anterior determinați q_i , $i = 1, \dots, j-1$
- ceea ce rămâne este ortogonal pe vectorii q_i și, după împărțirea cu lungimea lui pentru a deveni un vector unitar, va fi folosit ca q_j
- prin urmare, mulțimea $\{q_1, \dots, q_n\}$ constă din vectori ortogonali doi câte doi care generează același subspațiu al lui \mathbb{R}^m ca $\{A_1, \dots, A_n\}$
- rezultatul ortogonalizării Gram–Schmidt poate fi scris în formă matricială prin introducerea unei notății noi pentru produsele scalare implicate în calculul de mai sus
- definim $r_{jj} = \|y_j\|_2$ și $r_{ij} = q_i^T A_j$
- atunci (13) și (14) pot fi scrise astfel:

$$\begin{aligned} A_1 &= r_{11} q_1 \\ A_2 &= r_{12} q_1 + r_{22} q_2, \end{aligned}$$

- și cazul general (15) se traduce în

$$A_j = r_{1j} q_1 + \cdots + r_{j-1,j} q_{j-1} + r_{jj} q_j.$$

- prin urmare, rezultatul ortogonalizării Gram–Schmidt poate fi scris în formă matricială astfel:

$$(A_1 | \cdots | A_n) = (q_1 | \cdots | q_n) \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{nn} & & & \end{bmatrix}, \quad (16)$$

sau $A = QR$, unde considerăm A ca fiind matricea constând din coloanele A_j

- vom numi aceasta **factorizarea QR redusă**; versiunea completă va fi detaliată mai jos
- presupunerea că vectorii A_j sunt liniar independenți garantează faptul că intrările de pe diagonala principală r_{jj} sunt nenule
- în schimb, dacă A_j se află în subspațiul generat de A_1, \dots, A_{j-1} , atunci proiecțiile pe acești din urmă vectori dau însumate tot vectorul A_j , și $r_{jj} = \|y_j\|_2 = 0$

Exemplul 5

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$$

$$y_1 = A = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$r_{11} = \|y_1\|_2 = \sqrt{1^2 + 2^2 + 2^2} = \sqrt{9} = 3$$

$$q_1 = \frac{y_1}{\|y_1\|_2} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

- pentru a găsi al doilea vector unitar, luăm

$$y_2 = A_2 - q_1 q_1^T A_2 = \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} 2 = \begin{bmatrix} -14/3 \\ 5/3 \\ 1/3 \end{bmatrix}$$

și

$$q_2 = \frac{y_2}{\|y_2\|_2} = \frac{1}{5} \begin{bmatrix} -14/3 \\ 5/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} -14/15 \\ 1/15 \\ 1/15 \end{bmatrix}.$$

- deoarece $r_{12} = q_1^T A_2 = 2$ și $r_{22} = \|y_2\|_2 = 5$, rezultatul scris în forma matricială (16) este

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & -14/15 \\ 2/3 & 1/3 \\ 2/3 & 2/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} = QR.$$

- folosim termenul „clasică” pentru această versiune a metodei Gram–Schmidt, deoarece vom da o versiune actualizată, sau „modificată” la sfârșitul acestei secțiuni

Algoritmul 1 (Ortogonalizarea Gram–Schmidt clasică)

Fie $A_j, j = 1, \dots, n$ vectori liniar independenți.

for $j = 1, 2, \dots, n$

$y = A_j$

for $i = 1, 2, \dots, j - 1$

$r_{ij} = q_i^T A_j$

$y = y - r_{ij} q_i$

end

$r_{jj} = \|y\|_2$

$q_j = y / r_{jj}$

end

```

gramm_schmidt.m + 
1 function [Q, R] = gramm_schmidt(A)
2 [m, n] = size(A);
3
4 for j = 1:n
5     v = A(:, j);
6     for i = 1:j - 1
7         R(i, j) = Q(:, i)' * A(:, j);
8         v = v - R(i, j) * Q(:, i);
9     end
10    R(j, j) = norm(v);
11    Q(:, j) = v / R(j, j);
12 end
13 end
14
15
16 % A = [2 1;1 -1;2 1];
17 % [Q,R] = gramm_schmidt(A);
18 %disp(Q);
19 %disp(R);

```

Command Window

```

>> A=[1 -4;2 3;2 2];
>> [Q,R]=gramm_schmidt(A)

```

Q =

```

0.3333  -0.9333
0.6667   0.3333
0.6667   0.1333

```

R =

```

3.0000  2.0000
0      5.0000

```

- când metoda înregistrează un succes, se obișnuiește să se completeze matricea vectorilor ortogonali unitari pentru a da o bază completă a lui \mathbb{R}^m , și a obține astfel factorizarea QR „completă”
- acest lucru poate fi făcut, de exemplu, adăugând $m - n$ vectori suplimentari la A_j , astfel încât cei m vectori generează pe \mathbb{R}^m , și apoi aplicând metoda Gram–Schmidt
- în termenii bazei lui \mathbb{R}^m formată din q_1, \dots, q_m , vectorii inițiali pot fi exprimați astfel:

$$(A_1 | \dots | A_n) = (q_1 | \dots | q_m) \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & \cdots & r_{2n} & \\ \ddots & & \vdots & \\ & & & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}. \quad (17)$$

- această ecuație matricială reprezintă **factorizarea QR completă** a matricii $A = (A_1 | \dots | A_n)$, formată din vectorii inițiali
- observăm că dimensiunile matricilor din factorizarea QR completă sunt: A este de dimensiune $m \times n$, Q este pătratică de dimensiune $m \times m$, și matricea superior triangulară R are dimensiune $m \times n$, aceeași ca matricea A
- matricea Q din factorizarea QR completă ocupă un loc special în analiza numerică și îi vom da o definiție specială

Definiția 1

O matrice pătratică Q este **ortogonală** dacă $Q^T = Q^{-1}$.

- observăm că o matrice pătratică este ortogonală dacă și numai dacă are coloanele reprezentate prin vectori ortogonali unitari
- prin urmare, factorizarea QR completă este ecuația $A = QR$, unde Q este o matrice pătratică ortogonală și R este o matrice superior triangulară care are aceeași dimensiune cu A
- proprietatea de bază a unei matrici ortogonale este că păstrează norma euclidiană a unui vector

Lema 1

Dacă Q este o matrice ortogonală $m \times m$ și x este un vector m -dimensional, atunci $\|Qx\|_2 = \|x\|_2$.

- $\|Qx\|_2^2 = (Qx)^T Qx = x^T Q^T Qx = x^T x = \|x\|_2^2$
- produsul a două matrici ortogonale $m \times m$ este tot o matrice ortogonală
- factorizarea QR a unei matrici $m \times m$ folosind metoda Gram–Schmidt necesită aproximativ m^3 înmulțiri/împărțiri, de trei ori mai multe decât factorizarea LU, plus încă la fel de multe adunări

Exemplul 6

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$$

QR completă

$$y_1 = A = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$r_{11} = \|y_1\|_2 = \sqrt{1+2^2+2^2} = \sqrt{9} = 3$$

$$q_1 = \frac{y_1}{\|y_1\|_2} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

$$q_2 = \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix}$$

- adăugând un al treilea vector $A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, avem că

$$\begin{aligned} y_3 &= A_3 - q_1 q_1^T A_3 - q_2 q_2^T A_3 \\ &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{3}{3} \end{bmatrix} \frac{1}{3} - \begin{bmatrix} \frac{-14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix} \left(-\frac{14}{15}\right) = \frac{2}{225} \begin{bmatrix} 2 \\ 10 \\ -11 \end{bmatrix}, \end{aligned}$$

și $q_3 = y_3 / \|y_3\| = \begin{bmatrix} \frac{2}{15} \\ \frac{10}{15} \\ -\frac{11}{15} \end{bmatrix}$

- punând cele trei părți împreună, obținem factorizarea QR completă

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & -14/15 & 2/15 \\ 2/3 & 1/3 & 2/3 \\ 2/3 & 2/15 & -11/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} = QR.$$

- observăm că alegerea lui A_3 a fost arbitrară
- orică al treilea vector coloană liniar independent față de primii doi vectori coloană poate fi folosit
- comparăm acest rezultat cu factorizarea QR redusă din Exemplul 5

- există trei aplicații importante ale factorizării QR
- vom descrie două dintre ele aici; a treia este algoritmul QR pentru calculul valorilor proprii, introdus în Capitolul 10
- în primul rând, factorizarea QR poate fi folosită pentru a rezolva un sistem de n ecuații cu n necunoscute $Ax = b$
- factorizăm $A = QR$, și ecuația $Ax = b$ devine $QRx = b$ și $Rx = Q^T b$
- presupunând că A este nesingulară, intrările diagonale ale matricii superior triangulare R sunt nenule, astfel că R este nesingulară
- o substituție înapoi triangulară ne dă soluția x
- cum am menționat anterior, această abordare este de aproximativ trei ori mai costisitoare computațional față de abordarea LU
- a doua aplicație este pentru metoda celor mai mici pătrate
- fie A o matrice $m \times n$ cu $m \geq n$
- pentru a minimiza $\|Ax - b\|_2$, o rescriem sub forma $\|QRx - b\|_2 = \|Rx - Q^T b\|_2$ folosind Lema 1
- vectorul din norma euclidiană este

$$\begin{bmatrix} e_1 \\ \vdots \\ e_n \\ e_{n+1} \\ \vdots \\ e_m \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & \cdots & r_{2n} \\ \ddots & \ddots & \vdots \\ & & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} d_1 \\ \vdots \\ d_n \\ d_{n+1} \\ \vdots \\ d_m \end{bmatrix}, \quad (18)$$

unde $d = Q^T b$

- presupunem că $r_{ii} \neq 0$
- atunci partea superioară (e_1, \dots, e_n) a vectorului de eroare e poate fi făcută zero prin substituție înapoi
- alegerea valorilor x_i nu schimbă cu nimic partea inferioară a vectorului de eroare; evident, $(e_{n+1}, \dots, e_m) = (-d_{n+1}, \dots, -d_m)$
- prin urmare, soluția în sensul celor mai mici pătrate este minimizată folosind vectorul x obținut din substituția înapoi a părții superioare, și eroarea în sensul celor mai mici pătrate este $\|e\|_2^2 = d_{n+1}^2 + \dots + d_m^2$

Algoritm 2 (Cele mai mici pătrate folosind factorizarea QR)

Dându-se sistemul inconsistent $m \times n$ $Ax = b$, găsim factorizarea QR completă $A = QR$ și luăm

$$\begin{aligned} \hat{R} &= \text{submatricea } n \times n \text{ superioară a lui } R \\ \hat{d} &= \text{primele } n \text{ intrări superioare ale lui } d = Q^T b. \end{aligned}$$

Rezolvăm $\hat{R}\hat{x} = \hat{d}$ pentru a găsi soluția în sensul celor mai mici pătrate \bar{x} .

$$A_3 \cdot \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} = \frac{1}{3}$$

$$A_3 q_2^T = -\frac{14}{15}$$

Cele_mai_mici_patrate_cu_fQR.m

```

1 x=(2*(0:10)/5)';
2 y=1+x+x.^2+x.^3+x.^4+x.^5+x.^6+x.^7;
3 A=[x.^0 x.^1 x.^2 x.^3 x.^4 x.^5 x.^6 x.^7];
4 [Q,R]=qr(A);
5 b=Q'*y;
6 c=R(1:8,1:8)\b(1:8)
7
8
9
10
11

```

Command Window

>> Cele_mai_mici_patrate_cu_fQR

c =

```

1.0000
1.0000
1.0000
1.0000
1.0000
1.0000
1.0000
1.0000
1.0000

```

Exemplul 7

- folosiți factorizarea QR completă pentru a rezolva problema de tip cele mai mici pătrate
- $$\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$$
- trebuie să rezolvăm $Rx = Q^T b$, sau

$$\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 5 & 10 & 10 \\ -14 & 5 & 2 \\ 2 & 10 & -11 \end{bmatrix} \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \\ 3 \end{bmatrix}.$$

- eroarea în sensul celor mai mici pătrate va fi $\|e\|_2 = \|(0, 0, 3)\|_2 = 3$
- egalizând părțile superioare, obținem

$$\begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \end{bmatrix},$$

a cărui soluție este $\bar{x}_1 = 3.8$, $\bar{x}_2 = 1.8$

```
>> A = [1 -4; 2 3; 2 2];
>> b = [-3; 15; 9];
>> [Q, R] = qr(A);
>> y = Q' * b;
>> c = R(1:2, :) \ y(1:2)

c =
3.8000
1.8000
```

5.4. Cele mai mici pătrate neliniare

- soluția în sensul celor mai mici pătrate a sistemului liniar de ecuații $Ax = b$ minimizează norma euclidiană a rezidualului $\|Ax - b\|_2$
- am învățat două metode de a găsi soluția x , una bazată pe ecuațiile normale și alta bazată pe factorizarea QR
- niciuna dintre cele două metode nu poate fi aplicată dacă ecuațiile sunt neliniare
- în această secțiune, vom dezvolta metoda Gauss–Newton de rezolvare a problemelor neliniare de tip cele mai mici pătrate
- în plus față de ilustrarea utilizării acestei metode pentru a rezolva probleme de tip intersecție de cercuri, vom aplica metoda Gauss–Newton pentru a interpola datele folosind modele cu coeficienți neliniari

5.4.1. Metoda Gauss–Newton

- considerăm sistemul de m ecuații cu n necunoscute

$$\begin{aligned} r_1(x_1, \dots, x_n) &= 0 \\ &\vdots \\ r_m(x_1, \dots, x_n) &= 0. \end{aligned} \tag{19}$$

- suma păratelor erorilor este reprezentată de funcția

$$E(x_1, \dots, x_n) = \frac{1}{2}(r_1^2 + \dots + r_m^2) = \frac{1}{2}r^T r,$$

unde $r = [r_1, \dots, r_m]^T$

- constantă $1/2$ a fost inclusă în definiție pentru a simplifica formulele deduse ulterior
- pentru a minimiza E , facem gradientul $F(x) = \nabla E(x)$ egal zero:

$$0 = F(x) = \nabla E(x) = \nabla \left(\frac{1}{2}r(x)^T r(x) \right) = r(x)^T D r(x). \tag{20}$$

- observăm că am folosit regula produsului scalar pentru gradient
- reamintim metoda lui Newton pentru mai multe variabile, și o aplicăm funcției $F(x)^T = (r^T D r)^T = (D r)^T r$ privită ca vector coloană

- regula produsului matrice/vector poate fi aplicată pentru a da

$$DF(x)^T = D((Dr)^T r) = (Dr)^T \cdot Dr + \sum_{i=1}^m r_i Dc_i,$$

unde c_i este a i -a coloană a lui Dr

- observăm că $Dc_i = H_{r_i}$, matricea derivelor parțiale secunde, sau **hessiană** lui r_i :

$$H_{r_i} = \begin{bmatrix} \frac{\partial^2 r_i}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 r_i}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 r_i}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 r_i}{\partial x_n \partial x_n} \end{bmatrix}.$$

- aplicarea metodei lui Newton poate fi simplificată prin renunțarea la anumiți termeni
- renunțând la suma de m termeni de mai sus, obținem următorul algoritm:

Algoritmul 3 (Metoda Gauss–Newton)

Pentru a minimiza $r_1(x)^2 + \cdots + r_m(x)^2$.

Luăm x^0 = vectorul inițial.

for $k = 0, 1, 2, \dots$

$$A = Dr(x^k) \quad (21)$$

$$A^T A v^k = -A^T r(x^k)$$

$$x^{k+1} = x^k + v^k \quad (22)$$

end

observăm că fiecare pas al metodei Gauss–Newton este o reminiscență a ecuațiilor normale, în care matricea coeficientilor a fost înlocuită cu Dr

metoda Gauss–Newton poate fi rezolvată pentru a găsi o rădăcină a gradientului erorii pătratice

deși gradientul trebuie să fie zero într-un punct de minim, invers nu este adevărat, deci este posibil ca metoda să conveargă către un punct de maxim sau către un punct să

trebuie să fim atenți când interpretăm rezultatul algoritmului

```

1 function x = Problema_10(x0, r, Dr, k)
2
3 x = x0;
4
5 for i = 0 : k
6   A = Dr(x);
7
8   v = (A' * A) \ (-A' * r(x));
9
10  x = x + v;
11 end
12
13
```

- două cercuri care se intersectează, se pot intersecta în unul sau două puncte, dacă nu cumva cercurile coincid
- trei cercuri din plan, însă, nu au puncte comune de intersecție
- într-un astfel de caz, căutăm punctul din plan care este cel mai aproape de fi punctul de intersecție în sensul celor mai mici pătrate
- pentru trei cercuri, aceasta presupune trei ecuații neliniare cu două necunoscute x, y
- Exemplul 8 arată cum metoda Gauss–Newton rezolvă această problemă neliniară de tip cele mai mici pătrate
- Exemplul 9 definește cel mai bun punct într-un alt mod: punctul unic de intersecție al celor 3 cercuri se obține permitând ca razele lor să fie schimbată cu o valoare comună K
- aceasta ne dă o problemă cu trei ecuații cu trei necunoscute x, y, K , nu o problemă de tip cele mai mici pătrate, care poate fi rezolvată folosind metoda lui Newton pentru mai multe variabile
- în fine, Exemplul 10 adaugă un al patrulea cerc

- soluția celor patru ecuații cu trei necunoscute x, y, K este din nou o problemă de tip cele mai mici pătrate care poate fi rezolvată folosind metoda Gauss-Newton

Exemplul 8

- considerăm trei cercuri în plan care au centrele $(x_1, y_1) = (-1, 0)$, $(x_2, y_2) = (1, 1/2)$, $(x_3, y_3) = (1, -1/2)$ și, respectiv, razele $R_1 = 1$, $R_2 = 1/2$, $R_3 = 1/2$
- folosiți metoda Gauss-Newton pentru a găsi punctul pentru care suma pătratelor distanțelor până la cele trei cercuri este minimizată
- cercurile sunt prezentate în Figura 6(a)

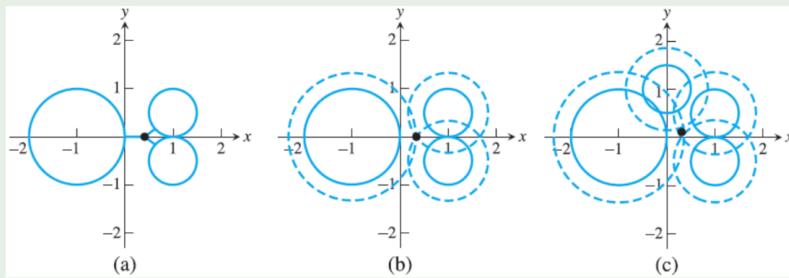


Figura 6: Puncte de intersecție a trei cercuri. (a) Punctul de intersecție în sensul celor mai mici pătrate, găsit folosind metoda Gauss-Newton. (b) Mărind razele cu o valoare comună ne dă un tip diferit de punct de intersecție folosind metoda lui Newton pentru mai multe variabile. (c) Cele patru cercuri din Exemplul 10 cu punctul de tip cele mai mici pătrate găsit folosind metoda Gauss-Newton.

- punctul (x, y) care trebuie găsit, minimizează suma pătratelor erorilor reziduale:

$$\begin{aligned}r_1(x, y) &= \sqrt{(x - x_1)^2 + (y - y_1)^2} - R_1 \\r_2(x, y) &= \sqrt{(x - x_2)^2 + (y - y_2)^2} - R_2 \\r_3(x, y) &= \sqrt{(x - x_3)^2 + (y - y_3)^2} - R_3.\end{aligned}$$

- aceasta rezultă din faptul că distanța de la un punct (x, y) la un cerc cu centru (x_i, y_i) și raza R_i este $|\sqrt{(x - x_i)^2 + (y - y_i)^2} - R_i|$
- jacobianul lui $r(x, y)$ este

$$Dr(x, y) = \begin{bmatrix} \frac{x - x_1}{S_1} & \frac{y - y_1}{S_1} \\ \frac{x - x_2}{S_2} & \frac{y - y_2}{S_2} \\ \frac{x - x_3}{S_3} & \frac{y - y_3}{S_3} \end{bmatrix},$$

unde $S_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$ pentru $i = 1, 2, 3$

- iterația Gauss-Newton cu vectorul inițial $(x^0, y^0) = (0, 0)$ converge la $(\bar{x}, \bar{y}) = (0.412891, 0)$ cu şase zecimale exacte după şapte pași

- o altă problemă pentru trei cercuri ne dă un tip diferit de răspuns
- în loc să căutăm puncte care seamănă cel mai mult cu puncte de intersecție, putem expanda (sau contracta) razele cercurilor cu o valoare comună până când cercurile au un punct de intersecție comun
- aceasta este echivalent cu rezolvarea sistemului

$$\begin{aligned}r_1(x, y, K) &= \sqrt{(x - x_1)^2 + (y - y_1)^2} - (R_1 + K) = 0 \\r_2(x, y, K) &= \sqrt{(x - x_2)^2 + (y - y_2)^2} - (R_2 + K) = 0 \\r_3(x, y, K) &= \sqrt{(x - x_3)^2 + (y - y_3)^2} - (R_3 + K) = 0. \quad (23)\end{aligned}$$

- punctul (x, y) identificat în acest mod este în general diferit de soluția de tip cele mai mici pătrate din Exemplul 8

Exemplul 9

- rezolvăți sistemul (23) pentru a găsi (x, y, K) , folosind cercurile din Exemplul 8
- sistemul constă din trei ecuații neliniare cu trei necunoscute, ceea ce înseamnă că poate fi rezolvat folosind metoda lui Newton pentru mai multe variabile
- jacobianul este

$$Dr(x, y, K) = \begin{bmatrix} \frac{x-x_1}{S_1} & \frac{y-y_1}{S_1} & -1 \\ \frac{x-x_2}{S_2} & \frac{y-y_2}{S_2} & -1 \\ \frac{x-x_3}{S_3} & \frac{y-y_3}{S_3} & -1 \end{bmatrix}.$$

- metoda lui Newton ne dă soluția $(x, y, K) = (1/3, 0, 1/3)$ în trei pași
- punctul de intersecție $(1/3, 0)$ și cele trei cercuri cu razele expandate cu $K = 1/3$ apar în Figura 6(b)
- Exemplele 8 și 9 prezintă două puncte de vedere diferite asupra sensului „punctului de intersecție” al unui grup de cercuri
- Exemplul 10 combină cele două abordări diferite

Exemplul 10

- considerăm patru cercuri cu centrele $(-1, 0), (1, 1/2), (1, -1/2), (0, 1)$ și, respectiv, razele $1, 1/2, 1/2, 1/2$
- găsiți punctul (x, y) și constanta K pentru care suma pătratelor distanțelor de la punct la cele patru cercuri cu razele mărite cu valoarea K (deci, cu razele respectiv $1 + K, 1/2 + K, 1/2 + K, 1/2 + K$) este minimizată
- aceasta este o simplă combinație a celor două exemple anterioare
- sunt patru ecuații cu trei necunoscute x, y, K

- rezidualul în sensul celor mai mici pătrate este asemănător cu (23), dar cu patru termeni, iar jacobianul este

$$Dr(x, y, K) = \begin{bmatrix} \frac{x-x_1}{S_1} & \frac{y-y_1}{S_1} & -1 \\ \frac{x-x_2}{S_2} & \frac{y-y_2}{S_2} & -1 \\ \frac{x-x_3}{S_3} & \frac{y-y_3}{S_3} & -1 \\ \frac{x-x_4}{S_4} & \frac{y-y_4}{S_4} & -1 \end{bmatrix}.$$

- metoda Gauss–Newton ne dă soluția $(\bar{x}, \bar{y}) = (0.311385, 0.112268)$ cu $\bar{K} = 0.367164$, prezentată în Figura 6(c)
- analogul Exemplului 10 pentru sfere în trei dimensiuni reprezintă fundamentalul matematic al Global Positioning System (GPS)

5.4.2. Metode cu parametri neliniari

- o aplicație importantă a metodei Gauss–Newton este de a interpola modele cu coeficienți neliniari
- fie $(t_1, y_1), \dots, (t_m, y_m)$ punctele de interpolare și $y = f_c(x)$ funcția care urmează să fie interpolată, unde $c = [c_1, \dots, c_p]^T$ este un set de parametri care va fi ales pentru a minimiza suma pătratelor rezidualilor

$$\begin{aligned} r_1(c) &= f_c(t_1) - y_1 \\ &\vdots \\ r_m(c) &= f_c(t_m) - y_m. \end{aligned}$$

- cazul particular (19) apare destul de des pentru a fi tratat separat aici
- dacă parametrii c_1, \dots, c_p apar în model liniar, atunci aceste ecuații liniare cu necunoscutele c_i , și ecuațiile normale sau soluția folosind factorizarea QR, ne dă alegerea optimă a parametrilor c
- dacă parametrii c_i apar neliniar în model, același tratament rezultă într-un sistem de ecuații care este neliniar în necunoscutele c_i
- de exemplu, interpolând modelul $y = c_1 t^{c_2}$ pentru punctele (t_i, y_i) , obținem ecuațiile neliniare

$$\begin{aligned} y_1 &= c_1 t_1^{c_2} \\ y_2 &= c_1 t_2^{c_2} \\ &\vdots \\ y_m &= c_1 t_m^{c_2}. \end{aligned}$$

- deoarece c_2 apare neliniar în acest model, sistemul de ecuații nu poate fi pus în formă matricială
- în Secțiunea 5.2, am tratat această dificultate schimbând problema: am „liniarizat modelul” luând logaritmul ambelor părți ale modelului și am minimizat eroarea în aceste coordonate transformate folosind cele mai mici pătrate
- în cazurile în care coordonatele transformate sunt într-adevăr coordonatele potrivite în care să minimizăm eroarea, acest lucru este adekvat
- pentru a rezolva problema de tip cele mai mici pătrate inițială, însă, apelăm la metoda Gauss–Newton
- aceasta este folosită pentru a minimiza funcția de eroare E ca funcție de vectorul de parametri c
- matricea Dr este matricea derivatelor parțiale ale erorilor r_i în funcție de parametrii c_j , care sunt

$$(Dr)_{ij} = \frac{\partial r_i}{\partial c_j} = f_{c_j}(t_i).$$

- cu această informație, metoda Gauss–Newton (21) poate fi implementată

Exemplul 11

- folosiți metoda Gauss–Newton pentru a interpola datele privind numărul total de mașini existente în lume din Exemplul 1 folosind un model exponentional (neliniarizat)
- găsirea celei mai bune interpolări de tip cele mai mici pătrate pentru date folosind un model exponentional înseamnă găsirea celor c_1, c_2 care minimizează REMP-ul pentru erorile $r_i = c_1 e^{c_2 t_i} - y_i$, $i = 1, \dots, m$
- folosind liniarizarea modelului din secțiunea anterioară, am minimizat REMP-ul pentru erorile modelului logaritmic $\ln y_i - (\ln c_1 + c_2 t_i)$

- valorile parametrilor c_i care minimizează REMP-ul în cele două sensuri diferite sunt diferite în general
- pentru a calcula cel mai bun model în sensul celor mai mici pătrate folosind metoda Gauss–Newton, definim

$$r = \begin{bmatrix} c_1 e^{c_2 t_1} - y_1 \\ \vdots \\ c_1 e^{c_2 t_m} - y_m \end{bmatrix},$$

și luăm derivatele în funcție de parametrii c_1 și c_2 pentru a obține

$$Dr = \begin{bmatrix} e^{c_2 t_1} & c_1 t_1 e^{c_2 t_1} \\ \vdots & \vdots \\ e^{c_2 t_m} & c_1 t_m e^{c_2 t_m} \end{bmatrix}.$$

- acest model este interpolat pentru numărul total de mașini existente în lume, unde t este măsurat în ani scurți din 1970, și mașinile în milioane
- cinci pași din metoda Gauss–Newton (21) pornind de la valoarea inițială $(c_1, c_2) = (50, 0.1)$ ne dău $(c_1, c_2) \approx (58.51, 0.05772)$ cu patru zecimale exacte
- cel mai bun model exponentional în sensul celor mai mici pătrate pentru aceste date este

$$y = 58.51 e^{0.05772t}. \quad (24)$$

- valoarea REMP-ului este 7.68, ceea ce reprezintă eroarea medie de modelare, în sensul celor mai mici pătrate, de 7.68 milioane de mașini (a se vedea Figura 7)
- cel mai bun model (24) poate fi comparat cu cel mai bun model exponentional liniarizat

$$y = 54.03 e^{0.06152t}$$

calculat în Exemplul 1

- acesta a fost obținut din ecuațiile normale aplicate modelului liniarizat $\ln y = \ln c_1 + c_2 t$
- valoarea REMP-ului pentru erorile r_i ale modelului liniarizat este 9.56, mai mare decât REMP-ul din (24), după cum era de așteptat

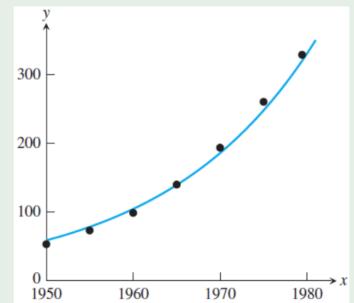


Figura 7: Interpolarea exponentională a numărului total de mașini existente în lume, fără a folosi liniarizarea. Cel mai bun model în sensul celor mai mici pătrate este $y = 58.51 e^{0.05772t}$.

- totuși, modelul liniarizat minimizează REMP-ul pentru erorile $\ln y_i - (\ln c_1 + c_2 t_i)$, dând o valoare de 0.0357, mai mică decât valoarea corespunzătoare 0.0568 pentru modelul (24), tot după cum era de așteptat
- fiecare dintre modele este interpolarea optimă în spațiul de date corespunzător
- morala este că există algoritmi computaționali pentru rezolvarea oricărei dintre probleme
- minimizarea valorilor r_i este problema de tip cele mai mici pătrate standard, dar utilizatorul trebuie să decidă pe baza contextului datelor dacă este mai potrivit să minimizeze erori sau erori logaritmice

5.4.3. Metoda Levenberg - Marquardt

- metoda Levenberg-Marquardt poate fi considerată ca o combinație între metoda Gauss-Newton și metoda gradientului, care va fi introdusă pentru probleme generale de optimizare în Capitolul 11
- algoritmul este o simplă modificare a metodei Gauss-Newton

Algoritmul 4 (Metoda Levenberg-Marquardt)

Pentru a minimiza $r_1(x)^2 + \dots + r_m(x)^2$.

Luăm x^0 = vectorul inițial, λ = constantă.

for $k = 0, 1, 2, \dots$

$$A = Dr(x^k)$$

$$(A^T A + \lambda \text{diag}(A^T A))v^k = -A^T r(x^k)$$

$$x^{k+1} = x^k + v^k$$

end

- cazul $\lambda = 0$ este identic cu metoda Gauss-Newton
- creșterea parametrului λ accentuează efectul diagonalei matricii $A^T A$, și, în general, face ca metoda să conveargă pornind de la o mulțime mai mare de vectori inițiali x^0 decât metoda Gauss-Newton

```
1 function x = Problema_12(x0, r, Dr, lambda, k)
2
3     x = x0;
4
5     for i = 0 : k
6         A = Dr(x);
7         v = (A' * a + lambda * diag(diag(A' * A))) \ (-A' * r(x));
8
9         x = x + v;
10    end
```

Exemplul 12

- folosiți metoda Levenberg-Marquardt pentru a interpola modelul $y = c_1 e^{-c_2(t - c_3)^2}$ pentru punctele $(t_i, y_i) = \{(1, 3), (2, 5), (2, 7), (3, 5), (4, 1)\}$
- trebuie să găsim acei c_1, c_2, c_3 care minimizează REMP-ul vectorului de eroare

$$r = \begin{bmatrix} c_1 e^{-c_2(t_1 - c_3)^2} - y_1 \\ \vdots \\ c_1 e^{-c_2(t_5 - c_3)^2} - y_5 \end{bmatrix}.$$

- derivata lui r evaluată în cele cinci puncte este matricea 5×3

$$Dr = \begin{bmatrix} e^{-c_2(t_1 - c_3)^2} & -c_1(t_1 - c_3)^2 e^{-c_2(t_1 - c_3)^2} & 2c_1c_2(t_1 - c_3)e^{-c_2(t_1 - c_3)^2} \\ \vdots & \vdots & \vdots \\ e^{-c_2(t_5 - c_3)^2} & -c_1(t_5 - c_3)^2 e^{-c_2(t_5 - c_3)^2} & 2c_1c_2(t_5 - c_3)e^{-c_2(t_5 - c_3)^2} \end{bmatrix}$$

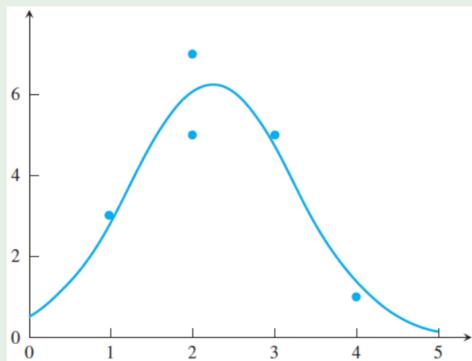


Figura 8: Modelul de interpolare din Exemplul 12. Metoda Levenberg–Marquardt este folosită pentru a găsi cel mai bun model în sensul celor mai mici pătrate $y = 6.301e^{-0.5088(t-2.249)^2}$, care este reprezentat grafic sub forma curbei continue.

- metoda Levenberg–Marquardt cu vectorul inițial $(c_1, c_2, c_3) = (1, 1, 1)$ și λ fixat la valoarea 50 converge către cel mai bun model în sensul celor mai mici pătrate $y = 6.301e^{-0.5088(t-2.249)^2}$
- cel mai bun model este reprezentat grafic împreună cu punctele de interpolare în Figura 8
- metoda Gauss–Newton corespunzătoare diverge la infinit pornind de la acest vector inițial