# Matematici Asistate de Calculator

Dr. Călin-Adrian POPA

## Cursul 4

27 Februarie 2019

# 3.4 Metode iterative

- eliminarea gaussiană este o secvenţă finită de operaţii care au ca rezultat o soluţie
- din acest motiv, eliminarea gaussiană este numită o metodă **directă** de rezolvare a sistemelor de ecuaţii liniare
- metodele directe, teoretic, dau soluţia exactă într-un număr finit de paşi
- metodele directe sunt în contrast cu metodele de găsire a soluţiilor descrise în Capitolul 2, care au o formă iterativă
- aşa-numitele metode **iterative** pot fi de asemenea aplicate pentru rezolvarea sistemelor de ecuaţii liniare
- asemenea iteraţiei de punct fix, metodele încep de la o valoare iniţială pe care o rafinează la fiecare pas, convergând către vectorul soluţie

# 3.4.1 Metoda lui Jacobi

- metoda lui Jacobi este o formă a iteraţiei de punct fix pentru un sistem de ecuaţii
- în IPF, primul pas este să rescriem ecuaţiile, pentru a găsi necunoscuta
- primul pas în metoda lui Jacobi este să facem acest lucru în următoarea formă standardizată: rezolvăm a $i$-a ecuaţie pentru a găsi a $i$-a necunoscută
- apoi, se iterează ca în iteraţia de punct fix, începând de la valoarea iniţială

**Exemplul 1**

- aplicaţi metoda lui Jacobi sistemului $3u + v = 5$, $u + 2v = 5$
- începem prin a rezolva prima ecuaţie pentru a găsi $u$ şi a doua ecuaţie pentru a găsi $v$
- vom folosi valoarea iniţială $(u_0, v_0) = (0, 0)$

# 3.4.1 Metoda lui Jacobi

- avem că

$$\begin{aligned} u &= \frac{5-v}{3} \\ v &= \frac{5-u}{2}. \end{aligned} \tag{1}$$

- cele două ecuaţii sunt iterate:

$$\begin{aligned} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_0}{3} \\ \frac{5-u_0}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-0}{3} \\ \frac{5-0}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ \frac{5}{2} \end{bmatrix} \\ \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_1}{3} \\ \frac{5-u_1}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-5/2}{3} \\ \frac{5-5/3}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ \frac{5}{3} \end{bmatrix} \\ \begin{bmatrix} u_3 \\ v_3 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_2}{3} \\ \frac{5-u_2}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-5/3}{3} \\ \frac{5-5/6}{2} \end{bmatrix} = \begin{bmatrix} \frac{10}{9} \\ \frac{25}{12} \end{bmatrix}. \end{aligned} \tag{2}$$

- paşii următori din metoda lui Jacobi arată o convergenţă către soluţie, care este $[1, 2]^T$

# 3.4.1 Metoda lui Jacobi

- acum să presupunem că ecuaţiile sunt date în ordine inversă

## Exemplul 2

- aplicaţi metoda lui Jacobi sistemului $u + 2v = 5, 3u + v = 5$
- rezolvăm prima ecuaţie pentru a găsi prima variabilă $u$ şi a doua ecuaţie pentru a găsi pe $v$
- începem cu

$$\begin{aligned} u &= 5 - 2v \\ v &= 5 - 3u. \end{aligned} \tag{3}$$

# 3.4.1 Metoda lui Jacobi

- cele două ecuaţii sunt iterate ca mai înainte, dar rezultatele sunt destul de diferite:

$$\begin{aligned}
\begin{bmatrix} u_0 \\ v_0 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} &= \begin{bmatrix} 5 - 2v_0 \\ 5 - 3u_0 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \\
\begin{bmatrix} u_2 \\ v_2 \end{bmatrix} &= \begin{bmatrix} 5 - 2v_1 \\ 5 - 3u_1 \end{bmatrix} = \begin{bmatrix} -5 \\ -10 \end{bmatrix} \\
\begin{bmatrix} u_3 \\ v_3 \end{bmatrix} &= \begin{bmatrix} 5 - 2v_2 \\ 5 - 3u_2 \end{bmatrix} = \begin{bmatrix} 25 \\ 20 \end{bmatrix}.
\end{aligned} \tag{4}$$

- în acest caz, metoda lui Jacobi eşuează, deoarece iteraţia diverge

- deoarece metoda lui Jacobi nu reuşeşte întotdeauna, este folositor să ştim condiţiile în care funcţionează
- o condiţie importantă este dată de următoarea definiţie:

# 3.4.1 Metoda lui Jacobi

## Definiţia 1

Matricea $n \times n$ $A = (a_{ij})$ este **strict diagonal dominantă** dacă, pentru orice $1 \leq i \leq n$, $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$. Cu alte cuvinte, fiecare intrare de pe diagonala principală domină rândul pe care se află în sensul în care este mai mare în valoare absolută decât suma valorilor absolute ale celorlalte intrări de pe rândul respectiv.

## Teorema 1

Dacă matricea $n \times n$ $A$ este strict diagonal dominantă, atunci (1) $A$ este o matrice nesingulară, şi (2) pentru orice vector $b$ şi orice valoare iniţială, metoda lui Jacobi aplicată ecuaţiei $Ax = b$ converge către soluţia unică.

- Teorema 1 spune că, dacă $A$ este strict diagonal dominantă, atunci metoda lui Jacobi aplicată ecuaţiei $Ax = b$ converge către o soluţie pentru orice valoare iniţială

# 3.4.1 Metoda lui Jacobi

- în Exemplul 1, matricea coeficienţilor este la început

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

  care este strict diagonal dominantă, deoarece $3 > 1$ şi $2 > 1$
- convergenţa este garantată în acest caz
- pe de altă parte, în Exemplul 2, metoda lui Jacobi este aplicată matricii

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

  care nu este diagonal dominantă, şi nu există o astfel de garanţie
- observăm că strict diagonal dominanţa este doar o condiţie suficientă
- metoda lui Jacobi s-ar putea totuşi să conveargă chiar şi în absenţa acestei condiţii

## Exemplul 3

- determinaţi dacă matricile

$$A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & -5 & 2 \\ 1 & 6 & 8 \end{bmatrix} \text{ şi } B = \begin{bmatrix} 3 & 2 & 6 \\ 1 & 8 & 1 \\ 9 & 2 & -2 \end{bmatrix}$$

  sunt strict diagonal dominante
- matricea $A$ este diagonal dominantă, deoarece $|3| > |1| + |-1|$, $|-5| > |2| + |2|$, şi $|8| > |1| + |6|$
- $B$ nu este diagonal dominantă, deoarece, de exemplu $|3| > |2| + |6|$ nu are loc
- totuşi, dacă primul şi al treilea rând ale lui $B$ sunt interschimbate, atunci $B$ este strict diagonal dominantă şi avem garanţia că metoda lui Jacobi va converge

# 3.4.1 Metoda lui Jacobi

- metoda lui Jacobi este o formă a iteraţiei de punct fix
- fie $D$ diagonala principală a lui $A$, $L$ triunghiul inferior al lui $A$ (intrările de sub diagonala principală), şi $U$ triunghiul superior al lui $A$ (intrările de deasupra diagonalei principale)
- atunci $A = L + D + U$, şi ecuaţia care trebuie rezolvată este $Lx + Dx + Ux = b$
- observăm că această utilizare a lui $L$ şi $U$ diferă de cea din factorizarea LU, deoarece toate intrările diagonale pentru aceşti $L$ şi $U$ sunt zero
- sistemul de ecuaţii $Ax = b$ poate fi rearanjat sub forma unei iteraţii de punct fix:

$$\begin{aligned} Ax &= b \\ (D + L + U)x &= b \\ Dx &= b - (L + U)x \\ x &= D^{-1}(b - (L + U)x). \end{aligned} \tag{5}$$

- deoarece $D$ este o matrice diagonală, inversa ei este matricea diagonală a inverselor intrărilor matricii $A$
- metoda lui Jacobi este doar iteraţia de punct fix din (5):

# 3.4.1 Metoda lui Jacobi

**Algoritmul 1 (Metoda lui Jacobi)**

$$x_0 = \text{vectorul iniţial}$$
$$x_{k+1} = D^{-1}(b-(L+U)x_k) \text{ for } k = 0,1,2,\ldots. \quad (6)$$

- pentru Exemplul 1,

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix},$$

iteraţia de punct fix (6) cu $x_k = \begin{bmatrix} u_k \\ v_k \end{bmatrix}$ este

$$\begin{aligned} \begin{bmatrix} u_{k+1} \\ v_{k+1} \end{bmatrix} &= D^{-1}(b-(L+U)x_k) \\ &= \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \left( \begin{bmatrix} 5 \\ 5 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_k \\ v_k \end{bmatrix} \right) \\ &= \begin{bmatrix} \frac{5-v_k}{3} \\ \frac{5-u_k}{2} \end{bmatrix}, \end{aligned}$$

care este aceeaşi cu versiunea originală

# 3.4.2 Metoda Gauss–Seidel şi SRS

- în strânsă legătură cu metoda lui Jacobi este o metodă iterativă numită metoda **Gauss–Seidel**
- singura diferenţă între metodele Gauss–Seidel şi Jacobi este aceea că în prima, cele mai recent actualizate valori ale necunoscutelor sunt utilizate la fiecare pas, chiar dacă actualizarea a avut loc în pasul curent
- întorcându-ne la Exemplul 1, vedem că metoda Gauss–Seidel arată după cum urmează:

$$
\begin{aligned}
\begin{bmatrix} u_0 \\ v_0 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_0}{3} \\ \frac{5-u_1}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-0}{3} \\ \frac{5-5/3}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ \frac{5}{3} \end{bmatrix} \\
\begin{bmatrix} u_2 \\ v_2 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_1}{3} \\ \frac{5-u_2}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-5/3}{3} \\ \frac{5-10/9}{2} \end{bmatrix} = \begin{bmatrix} \frac{10}{9} \\ \frac{35}{18} \end{bmatrix} \\
\begin{bmatrix} u_3 \\ v_3 \end{bmatrix} &= \begin{bmatrix} \frac{5-v_2}{3} \\ \frac{5-u_3}{2} \end{bmatrix} = \begin{bmatrix} \frac{5-35/18}{3} \\ \frac{5-55/54}{2} \end{bmatrix} = \begin{bmatrix} \frac{55}{54} \\ \frac{215}{108} \end{bmatrix}.
\end{aligned} \tag{7}
$$

# 3.4.2 Metoda Gauss–Seidel şi SRS

- observăm diferenţa dintre metodele Gauss–Seidel şi Jacobi: definiţia lui $v_1$ foloseşte $u_1$, nu $u_0$
- vedem convergenţa către soluţia $[1, 2]^T$, ca în metoda lui Jacobi, dar cu ceva mai mare precizie în acelaşi număr de paşi
- metoda Gauss–Seidel adesea converge mai repede decât metoda lui Jacobi, dacă este convergentă
- metoda Gauss–Seidel, la fel ca metoda lui Jacobi, converge către soluţie dacă matricea coeficienţilor este strict diagonal dominantă
- metoda Gauss–Seidel poate fi scrisă în formă matricială şi identificată ca o iteraţie de punct fix în care izolăm ecuaţia $(L + D + U)x = b$ în forma

$$(L + D)x_{k+1} = -Ux_k + b.$$

- observăm că folosirea intrărilor nou determinate ale lui $x_{k+1}$ este realizată prin includerea triunghiului inferior al lui $A$ în partea stângă a ecuaţiei
- rearanjând această ecuaţie, obţinem metoda Gauss–Seidel

## Algoritmul 2 (**Metoda Gauss–Seidel**)

$$\begin{aligned} x_0 &= \text{vectorul iniţial} \\ x_{k+1} &= D^{-1}(b - Ux_k - Lx_{k+1}) \textbf{ for } k = 0, 1, 2, \ldots. \end{aligned}$$

## Exemplul 4

- aplicaţi metoda Gauss–Seidel sistemului

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 4 & 1 \\ -1 & 2 & 5 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}.$$

- iteraţia Gauss–Seidel este

$$\begin{aligned} u_{k+1} &= \frac{4 - v_k + w_k}{3} \\ v_{k+1} &= \frac{1 - 2u_{k+1} - w_k}{4} \\ w_{k+1} &= \frac{1 + u_{k+1} - 2v_{k+1}}{5}. \end{aligned}$$

# 3.4.2 Metoda Gauss–Seidel şi SRS

- începând cu $x_0 = [u_0, v_0, w_0]^T = [0, 0, 0]^T$, calculăm

$$\begin{bmatrix} u_1 \\ v_1 \\ w_1 \end{bmatrix} = \begin{bmatrix} \frac{4-0+0}{3} \\ \frac{1-2(4/3)-0}{4} \\ \frac{1+(4/3)-2(-5/12)}{5} \end{bmatrix} = \begin{bmatrix} \frac{4}{3} \\ -\frac{5}{12} \\ \frac{19}{30} \end{bmatrix} \approx \begin{bmatrix} 1.3333 \\ -0.4167 \\ 0.6333 \end{bmatrix}$$

şi

$$\begin{bmatrix} u_2 \\ v_2 \\ w_2 \end{bmatrix} = \begin{bmatrix} \frac{101}{60} \\ -\frac{3}{4} \\ \frac{251}{300} \end{bmatrix} \approx \begin{bmatrix} 1.6833 \\ -0.7500 \\ 0.8367 \end{bmatrix}.$$

- sistemul este strict diagonal dominant, şi prin urmare iteraţia va converge către soluţia $[2, -1, 1]^T$

# 3.4.2 Metoda Gauss–Seidel şi SRS

- metoda numită a **supra-relaxărilor succesive (SRS)** preia direcţia dată de metoda Gauss–Seidel înspre soluţie şi „depăşeşte măsura” în încercarea de a accelera convergenţa
- fie $\omega$ un număr real, şi definim fiecare componentă a noii aproximări $x_{k+1}$ ca o medie ponderată de $\omega$ ori formula Gauss–Seidel şi de $1-\omega$ ori aproximarea curentă $x_k$
- numărul $\omega$ se numeşte **parametru de relaxare**, şi când $\omega > 1$ vorbim despre **supra-relaxare**

## Exemplul 5

- aplicaţi SRS cu $\omega = 1.25$ sistemului din Exemplul 4
- metoda supra-relaxărilor succesive ne dă

$$\begin{aligned} u_{k+1} &= (1-\omega)u_k + \omega\frac{4 - v_k + w_k}{3} \\ v_{k+1} &= (1-\omega)v_k + \omega\frac{1 - 2u_{k+1} - w_k}{4} \\ w_{k+1} &= (1-\omega)w_k + \omega\frac{1 + u_{k+1} - 2v_{k+1}}{5}. \end{aligned}$$

# 3.4.2 Metoda Gauss–Seidel şi SRS

- începând cu $[u_0, v_0, w_0]^T = [0, 0, 0]^T$, calculăm

$$\begin{bmatrix} u_1 \\ v_1 \\ w_1 \end{bmatrix} \approx \begin{bmatrix} 1.6667 \\ -0.7292 \\ 1.0312 \end{bmatrix}$$

şi

$$\begin{bmatrix} u_2 \\ v_2 \\ w_2 \end{bmatrix} \approx \begin{bmatrix} 1.9835 \\ -1.0672 \\ 1.0216 \end{bmatrix}.$$

- în acest exemplu, iteraţia SRS converge mai repede decât Jacobi şi Gauss–Seidel către soluţia $[2, -1, 1]^T$

- la fel ca în cazul metodelor Jacobi şi Gauss–Seidel, o deducere alternativă a SRS rezultă din tratarea sistemului ca o problemă de punct fix

# 3.4.2 Metoda Gauss–Seidel şi SRS

- problema $Ax = b$ poate fi scrisă $(L + D + U)x = b$, şi, când înmulţim cu $\omega$ şi rearanjăm, avem

$$\begin{aligned}(\omega L + \omega D + \omega U)x &= \omega b \\ (\omega L + D)x &= \omega b - \omega U x + (1-\omega)Dx \\ x &= (\omega L + D)^{-1}[(1-\omega)Dx - \omega U x] + \omega(D + \omega L)^{-1} b.\end{aligned}$$

**Algoritmul 3 (Metoda supra-relaxărilor succesive (SRS))**

$$\begin{aligned}x_0 &= \text{vectorul iniţial} \\ x_{k+1} &= (\omega L + D)^{-1}[(1-\omega)Dx_k - \omega U x_k] + \omega(D + \omega L)^{-1} b \ \textbf{for}\ k = 0, 1, 2, \ldots\end{aligned}$$

- SRS cu $\omega = 1$ este exact Gauss–Seidel
- parametrul $\omega$ poate de asemenea să fie mai mic decât 1, dând astfel naştere unei metode numită metoda sub-relaxărilor succesive

# 3.5 Metode pentru matrici simetrice şi pozitiv definite

- matricile simetrice au o poziţie privilegiată în analiza sistemelor liniare datorită structurii lor speciale şi pentru că au doar aproximativ jumătate din numărul de intrări independente al unor matrici generale
- aceasta ridică întrebarea dacă nu cumva factorizări ca factorizarea LU pot fi realizate la jumătate din complexitatea computaţională, şi folosind doar jumătate din locaţiile de memorie
- pentru matricile simetrice şi pozitiv definite, acest scop poate fi atins de către factorizarea Cholesky
- matricile simetrice şi pozitiv definite permit de asemenea o abordare destul de diferită pentru rezolvarea ecuaţiei $Ax = b$, care nu depinde de o factorizare matricială
- această nouă abordare, numită metoda gradienţilor conjugaţi, este utilă mai ales pentru matrici mari
- în debutul secţiunii, definim conceptul de pozitiv definire pentru o matrice simetrică
- apoi, arătăm că orice matrice simetrică şi pozitiv definită $A$ poate fi factorizată sub forma $A = R^T R$ pentru o matrice superior triangulară $R$, fapt ce reprezintă factorizarea Cholesky
- încheiem secţiunea cu algoritmul gradienţilor conjugaţi

# 3.5.1 Matrici simetrice şi pozitiv definite

## Definiţia 2

Matricea $n \times n$ $A$ este **simetrică** dacă $A^T = A$. Matricea $A$ este **pozitiv definită** dacă $x^T Ax > 0$ pentru orice vector $x \neq 0$.

## Exemplul 6

- arătaţi că matricea $A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$ este simetrică şi pozitiv definită
- evident, $A$ este simetrică
- pentru a arăta că este pozitiv definită, aplicăm definiţia:

$$\begin{aligned} x^T Ax &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2x_1^2 + 4x_1x_2 + 5x_2^2 \\ &= 2(x_1 + x_2)^2 + 3x_2^2. \end{aligned}$$

- această expresie este întotdeauna nenegativă, şi nu poate fi zero decât dacă $x_2 = 0$ şi $x_1 + x_2 = 0$, care împreună implică $x = 0$

## Exemplul 7

- arătaţi că matricea simetrică $A = \begin{bmatrix} 2 & 4 \\ 4 & 5 \end{bmatrix}$ nu este pozitiv definită
- calculăm $x^T Ax$ prin completarea pătratului:

$$\begin{aligned} x^T Ax &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2x_1^2 + 8x_1x_2 + 5x_2^2 \\ &= 2(x_1^2 + 4x_1x_2) + 5x_2^2 \\ &= 2(x_1 + 2x_2)^2 - 8x_2^2 + 5x_2^2 \\ &= 2(x_1 + 2x_2)^2 - 3x_2^2. \end{aligned}$$

- luând $x_1 = -2$ şi $x_2 = 1$, de exemplu, rezultatul va fi mai mic decât zero, ceea ce contrazice definiţia unei matrici pozitiv definite

- observăm că o matrice simetrică şi pozitiv definită trebuie să fie nesingulară, deoarece este imposibil ca un vector nenul $x$ să satisfacă $Ax = 0$

# 3.5 Metode pentru matrici simetrice şi pozitiv definite

- există încă trei fapte importante care au loc pentru această clasă de matrici

**Teorema 2**

Presupunem că $A$ este o matrice simetrică $n \times n$ cu intrări reale. Atunci valorile proprii ale lui $A$ sunt numere reale, şi mulţimea vectorilor proprii unitari ai lui $A$ este o mulţime ortonormală $\{w_1, \ldots, w_n\}$ care formează o bază a lui $\mathbb{R}^n$.

**Faptul 1**

Dacă matricea $n \times n$ $A$ este simetrică, atunci $A$ este pozitiv definită dacă şi numai dacă toate valorile proprii ale sale sunt pozitive.

- Teorema 2 ne spune că mulţimea vectorilor proprii este ortonormală şi formează o bază a lui $\mathbb{R}^n$
- dacă $A$ este pozitiv definită şi $Av = \lambda v$ pentru un vector nenul $v$, atunci $0 < v^T Av = v^T(\lambda v) = \lambda ||v||_2^2$, deci $\lambda > 0$

# 3.5 Metode pentru matrici simetrice şi pozitiv definite

- pe de altă parte, dacă toate valorile proprii ale lui $A$ sunt pozitive, atunci putem scrie orice vector nenul $x = c_1 v_1 + \ldots + c_n v_n$, unde $v_i$ sunt vectori unitari ortonormali şi nu toţi $c_i$ sunt zero
- atunci $x^T Ax = (c_1 v_1 + \ldots + c_n v_n)^T (\lambda_1 c_1 v_1 + \ldots + \lambda_n c_n v_n) = \lambda_1 c_1^2 + \ldots + \lambda_n c_n^2 > 0$, deci $A$ este pozitiv definită
- valorile proprii ale lui $A$ din Exemplul 6 sunt 6 şi 1
- valorile proprii ale lui $A$ din Exemplul 7 sunt aproximativ 7.77 şi $-0.77$

**Faptul 2**

Dacă $A$ este o matrice $n \times n$ simetrică şi pozitiv definită şi $X$ este o matrice $n \times m$ de rang maxim, cu $n \geq m$, atunci $X^T AX$ este o matrice $m \times m$ simetrică şi pozitiv definită.

- matricea este simetrică deoarece $(X^T AX)^T = X^T AX$
- considerăm un $m$-vector nenul $v$
- observăm că $v^T (X^T AX) v = (Xv)^T A(Xv) \geq 0$, cu egalitate doar dacă $Xv = 0$, datorită faptului că $A$ este pozitiv definită
- deoarece $X$ are rang maxim, coloanele ei sunt liniar independente, astfel că $Xv = 0$ implică $v = 0$

# 3.5 Metode pentru matrici simetrice şi pozitiv definite

**Definiţia 3**

O submatrice **principală** a matricii pătratice $A$ este o submatrice pătratică ale cărei intrări diagonale sunt intrări diagonale ale lui $A$.

**Faptul 3**

Orice submatrice principală a unei matrici simetrice şi pozitiv definite este de asemenea simetrică şi pozitiv definită.

- de exemplu, dacă

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

este simetrică şi pozitiv definită, atunci la fel este şi

$$\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}.$$

# 3.5.2 Factorizarea Cholesky

- pentru a demonstra ideea principală, începem cu cazul $2 \times 2$
- toate problemele importante apar în acest caz, extensia la cazul general fiind imediată
- considerăm matricea simetrică şi pozitiv definită

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

- din Faptul 3 despre matricile simetrice şi pozitiv definite, ştim că $a > 0$
- în plus, ştim că determinantul $ac - b^2$ al lui $A$ este pozitiv, deoarece determinantul este produsul valorilor proprii, care sunt toate pozitive, conform Faptului 1
- scriind $A = R^T R$ unde $R$ este o matrice superior triangulară, implică forma

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \sqrt{a} & 0 \\ u & v \end{bmatrix} \begin{bmatrix} \sqrt{a} & u \\ 0 & v \end{bmatrix} = \begin{bmatrix} a & u\sqrt{a} \\ u\sqrt{a} & u^2 + v^2 \end{bmatrix},$$

şi vrem să verificăm dacă acest lucru este posibil
- comparând partea dreaptă şi partea stângă, obţinem identităţile $u = b/\sqrt{a}$ şi $v^2 = c - u^2$
- observăm că $v^2 = c - (b/\sqrt{a})^2 = c - b^2/a > 0$, care rezultă din faptul că determinantul este pozitiv

# 3.5.2 Factorizarea Cholesky

- deducem că $v$ poate fi definit ca un număr real, şi deci factorizarea Cholesky

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \sqrt{a} & 0 \\ \frac{b}{\sqrt{a}} & \sqrt{c - b^2/a} \end{bmatrix} \begin{bmatrix} \sqrt{a} & \frac{b}{\sqrt{a}} \\ 0 & \sqrt{c - b^2/a} \end{bmatrix} = R^T R$$

  există pentru matrici $2 \times 2$ simetrice şi pozitiv definite
- factorizarea Cholesky nu este unică; puteam la fel de bine să alegem $v$ ca fiind rădăcina pătrată a lui $c - b^2/a$ cu semnul minus
- următorul rezultat garantează că aceeaşi idee funcţionează şi pentru cazul $n \times n$

**Teorema 3 (Teorema factorizării Cholesky)**

Dacă $A$ este o matrice $n \times n$ simetrică şi pozitiv definită, atunci există o matrice $n \times n$ superior triangulară $R$ astfel încât $A = R^T R$.

- construim pe $R$ prin inducţie după $n$
- cazul $n = 2$ a fost făcut mai sus

# 3.5.2 Factorizarea Cholesky

- considerăm că $A$ este partiţionată ca

$$A = \left[\begin{array}{c|c} a & b^T \\ \hline b & C \end{array}\right]$$

  unde $b$ este un $(n-1)$-vector şi $C$ este o submatrice $(n-1) \times (n-1)$
- vom folosi înmulţirea pe blocuri pentru a simplifica argumentul
- luăm $u = b/\sqrt{a}$, ca în cazul $2 \times 2$
- luând $A_1 = C - uu^T$ şi definind matricea inversabilă

$$S = \left[\begin{array}{c|c} \sqrt{a} & u^T \\ \hline 0 & \\ \vdots & I \\ 0 & \end{array}\right]$$

# 3.5.2 Factorizarea Cholesky

- avem că

$$S^T \left[\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right] S = \left[\begin{array}{c|ccc} \sqrt{a} & 0 & \dots & 0 \\ \hline u & & I & \end{array}\right] \left[\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} \sqrt{a} & u^T & & \\ \hline 0 & & & \\ \vdots & & I & \\ 0 & & & \end{array}\right]$$

$$= \left[\begin{array}{c|ccc} a & b^T & & \\ \hline b & uu^T + A_1 & & \end{array}\right] = A.$$

- observăm că $A_1$ este simetrică şi pozitiv definită
- aceasta rezultă din faptul că

$$\left[\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right] = (S^T)^{-1} A S^{-1}$$

  este simetrică şi pozitiv definită din Faptul 2, şi la fel este şi submatricea principală $(n-1) \times (n-1)$ $A_1$ din Faptul 3
- din ipoteza de inducţie, $A_1 = V^T V$, unde $V$ este superior triangulară

# 3.5.2 Factorizarea Cholesky

- în final, definim matricea superior triangulară

$$R = \left[\begin{array}{c|c} \sqrt{a} & u^T \\ \hline 0 & \\ \vdots & V \\ 0 & \end{array}\right]$$

şi verificăm că

$$R^T R = \left[\begin{array}{c|ccc} \sqrt{a} & 0 & \dots & 0 \\ \hline u & & V^T & \end{array}\right] \left[\begin{array}{c|c} \sqrt{a} & u^T \\ \hline 0 & \\ \vdots & V \\ 0 & \end{array}\right] = \left[\begin{array}{c|c} a & b^T \\ \hline b & uu^T + V^T V \end{array}\right] = A,$$

încheind astfel demonstraţia

- construcţia demonstraţiei se poate face explicit, în ceea ce a devenit algoritmul standard pentru factorizarea Cholesky
- matricea $R$ este construită din exterior înspre interior
- mai întâi, găsim $r_{11} = \sqrt{a_{11}}$ şi luăm restul primului rând al lui $R$ ca fiind $u^T = b^T / r_{11}$

# 3.5.2 Factorizarea Cholesky

- atunci $uu^T$ este scăzut din submatricea principală $(n-1) \times (n-1)$ inferioară, şi aceiaşi paşi sunt repetaţi pentru a umple al doilea rând al lui $R$
- aceşti paşi sunt continuaţi până când toate rândurile lui $R$ sunt determinate
- potrivit teoremei, noua submatrice principală este pozitiv definită la fiecare pas al construcţiei, deci, din Faptul 3, rezultă că intrarea din colţul din stânga sus este pozitivă, şi extragerea rădăcinii pătrate este validă
- această abordare poate fi pusă direct în următorul algoritm
- folosim notaţia cu „două puncte” pentru a specifica submatricile respective

# 3.5.2 Factorizarea Cholesky

**Algoritmul 4 (Factorizarea Cholesky)**

**for** $k = 1, 2, \ldots, n$
  **if** $A_{kk} < 0$, **stop**, **end**
  $R_{kk} = \sqrt{A_{kk}}$
  $u^T = \frac{1}{R_{kk}} A_{k,k+1:n}$
  $R_{k,k+1:n} = u^T$
  $A_{k+1:n,k+1:n} = A_{k+1:n,k+1:n} - uu^T$
**end**

- matricea $R$ rezultată este superior triangulară şi satisface $A = R^T R$

**Exemplul 8**

- găsiţi factorizarea Cholesky a lui $\begin{bmatrix} 4 & -2 & 2 \\ -2 & 2 & -4 \\ 2 & -4 & 11 \end{bmatrix}$

# 3.5.2 Factorizarea Cholesky

- primul rând al lui $R$ este $R_{11} = \sqrt{a_{11}} = 2$, urmat de $R_{1,2:3} = [-2, 2]/R_{11} = [-1, 1]$:

$$R = \left[\begin{array}{c|cc} 2 & -1 & 1 \\ \hline & & \\ & & \end{array}\right].$$

- scăzând produsul $uu^T = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix}$ din submatricea principală $2 \times 2$ inferioară $A_{2:3,2:3}$ a lui $A$, obţinem

$$\left[\begin{array}{c|cc} & & \\ \hline & 2 & -4 \\ & -4 & 11 \end{array}\right] - \left[\begin{array}{c|cc} & & \\ \hline & 1 & -1 \\ & -1 & 1 \end{array}\right] = \left[\begin{array}{c|cc} & & \\ \hline & 1 & -3 \\ & -3 & 10 \end{array}\right].$$

- acum, repetăm aceiaşi paşi pentru submatricea $2 \times 2$ şi găsim $R_{22} = 1$ şi $R_{23} = -3/1 = -3$:

$$R = \left[\begin{array}{c|c|c} 2 & -1 & 1 \\ \hline & 1 & -3 \\ \hline & & \end{array}\right].$$

# 3.5.2 Factorizarea Cholesky

- submatricea principală $1 \times 1$ inferioară a lui $A$ este $10 - (-3)(-3) = 1$, deci $R_{33} = \sqrt{1}$
- factorul Cholesky al lui $A$ este

$$R = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix}.$$

- rezolvarea sistemului $Ax = b$ pentru matricea simetrică şi pozitiv definită $A$ urmează aceeaşi idee ca şi factorizarea LU
- acum că $A = R^T R$ este un produs a două matrici triangulare, trebuie să rezolvăm sistemul inferior triangular $R^T c = b$ şi sistemul superior triangular $Rx = c$ pentru a determina soluţia $x$

# 3.5.3 Metoda gradienţilor conjugaţi

- metoda gradienţilor conjugaţi se bazează pe generalizarea ideii cunoscute de produs scalar
- produsul scalar euclidian $(v, w) = v^T w$ este simetric şi liniar în intrările $v$ şi $w$, deoarece $(v, w) = (w, v)$ şi $(\alpha v + \beta w, u) = \alpha(v, u) + \beta(w, u)$, pentru orice scalari $\alpha$ şi $\beta$
- produsul scalar euclidian este de asemenea şi pozitiv definit, prin aceea că $(v, v) > 0$ dacă $v \neq 0$

**Definiţia 4**

Fie $A$ o matrice $n \times n$ simetrică şi pozitiv definită. Pentru doi $n$-vectori $v$ şi $w$, definim $A$-**produsul scalar** prin $(v, w)_A = v^T A w$. Vectorii $v$ şi $w$ sunt $A$-conjugaţi dacă $(v, w)_A = 0$.

- observăm că noul produs scalar moşteneşte proprietăţile de simetrie, liniaritate şi pozitiv definire de la matricea $A$
- deoarece $A$ este simetrică, la fel este şi $A$-produsul scalar: $(v, w)_A = v^T A w = (v^T A w)^T = w^T A v = (w, v)_A$

# 3.5.3 Metoda gradienţilor conjugaţi

- $A$-produsul scalar este de asemenea liniar, iar pozitiv definirea rezultă din faptul că dacă $A$ este pozitiv definită, atunci

$$(v, v)_A = v^T A v > 0, \text{ dacă } v \neq 0$$

- în sens strict, metoda gradienţilor conjugaţi este o metodă directă, şi ajunge la soluţia $x$ a sistemului simetric şi pozitiv definit $Ax = b$ cu următoarea buclă finită:

**Algoritmul 5 (Metoda gradienţilor conjugaţi)**

$x_0 =$ vectorul iniţial
$d_0 = r_0 = b - Ax_0$
**for** $k = 0, 1, 2, \ldots, n-1$
  **if** $r_k = 0$, **stop**, **end**
  $\alpha_k = \frac{r_k^T r_k}{d_k^T A d_k}$
  $x_{k+1} = x_k + \alpha_k d_k$
  $r_{k+1} = r_k - \alpha_k A d_k$
  $\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$
  $d_{k+1} = r_{k+1} + \beta_k d_k$
**end**

# 3.5.3 Metoda gradienţilor conjugaţi

- dăm acum o descriere informală a acestei iteraţii, care va fi urmată de demonstraţia faptelor necesare în Teorema 4
- iteraţia gradienţilor conjugaţi actualizează trei vectori diferiţi la fiecare pas
- vectorul $x_k$ este aproximarea soluţiei la pasul $k$
- vectorul $r_k$ reprezintă rezidualul aproximării $x_k$
- acest lucru este clar pentru $r_0$ din definiţie, şi, în timpul iteraţiei, observăm că

$$\begin{aligned} Ax_{k+1} + r_{k+1} &= A(x_k + \alpha_k d_k) + r_k - \alpha_k A d_k \\ &= Ax_k + r_k, \end{aligned}$$

  şi astfel prin inducţie $r_k = b - Ax_k$, pentru orice $k$
- în final, vectorul $d_k$ reprezintă noua direcţie de căutare folosită pentru a actualiza aproximarea $x_k$ în versiunea îmbunătăţită $x_{k+1}$
- metoda reuşeşte deoarece fiecare rezidual este definit astfel încât să fie ortogonal pe toţi rezidualii anteriori
- dacă acest lucru poate fi făcut, metoda va rămâne fără direcţii ortogonale în care să caute şi va trebui să ajungă la un rezidual egal cu zero şi la o soluţie corectă în cel mult $n$ paşi

# 3.5.3 Metoda gradienţilor conjugaţi

- cheia pentru a realiza ortogonalitatea între reziduali se dovedeşte a fi alegerea direcţiilor de căutare $d_k$ ca fiind conjugate două câte două
- conceptul de conjugare generalizează conceptul de ortogonalitate, dând şi numele algoritmului
- acum vom explica alegerile lui $\alpha_k$ şi $\beta_k$
- direcţiile $d_k$ sunt alese din subspaţiul liniar generat de rezidualii anteriori, după cum se poate vedea prin inducţie din ultima linie a buclei algoritmului
- pentru a asigura faptul că următorul rezidual este ortogonal pe toţi rezidualii anteriori, $\alpha_k$ este ales tocmai pentru ca noul rezidual $r_{k+1}$ să fie ortogonal pe direcţia $d_k$:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k d_k \\ b - Ax_{k+1} &= b - Ax_k - \alpha_k Ad_k \\ r_{k+1} &= r_k - \alpha_k Ad_k \\ 0 = d_k^T r_{k+1} &= d_k^T r_k - \alpha_k d_k^T Ad_k \\ \alpha_k &= \frac{d_k^T r_k}{d_k^T Ad_k}. \end{aligned}$$

# 3.5.3 Metoda gradienţilor conjugaţi

- aceasta nu este exact forma lui $\alpha_k$ din algoritm, dar observăm că, deoarece $d_{k-1}$ este ortogonal pe $r_k$, avem

$$\begin{aligned} d_k - r_k &= \beta_{k-1} d_{k-1} \\ r_k^T d_k - r_k^T r_k &= 0, \end{aligned}$$

  ceea ce justifică rescrierea $r_k^T d_k = r_k^T r_k$
- în al doilea rând, coeficientul $\beta_k$ este ales pentru a asigura $A$-conjugarea direcţiilor $d_k$ două câte două:

$$\begin{aligned} d_{k+1} &= r_{k+1} + \beta_k d_k \\ 0 = d_k^T A d_{k+1} &= d_k^T A r_{k+1} + \beta_k d_k^T A d_k \\ \beta_k &= -\frac{d_k^T A r_{k+1}}{d_k^T A d_k}. \end{aligned}$$

- expresia pentru $\beta_k$ poate fi rescrisă în forma mai simplă folosită în algoritm, după cum se arată în ecuaţia (9) de mai jos
- Teorema 4 de mai jos verifică faptul că toţi rezidualii $r_k$ produşi de către iteraţia gradienţilor conjugaţi sunt ortogonali unii pe alţii

# 3.5.3 Metoda gradienţilor conjugaţi

- deoarece ei sunt vectori $n$-dimensionali, cel mult $n$ din rezidualii $r_k$ pot fi ortogonali doi câte doi, deci ori $r_n$ ori un $r_k$ anterior trebuie să fie zero, rezolvând astfel $Ax = b$
- prin urmare, după cel mult $n$ paşi, metoda gradienţilor conjugaţi ajunge la o soluţie
- în teorie, metoda este una directă, şi nu una iterativă
- înainte de a ajunge la teorema care garantează succesul metodei gradienţilor conjugaţi, este instructiv să urmărim un exemplu în aritmetică exactă

**Exemplul 9**

- rezolvaţi $\begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$ folosind metoda gradienţilor conjugaţi

# 3.5.3 Metoda gradienţilor conjugaţi

- urmând algoritmul de mai sus, avem că

$$
\begin{aligned}
x_0 &= \begin{bmatrix}0\\0\end{bmatrix}, \; r_0 = d_0 = \begin{bmatrix}6\\3\end{bmatrix} \\
\alpha_0 &= \frac{\begin{bmatrix}6\\3\end{bmatrix}^T \begin{bmatrix}6\\3\end{bmatrix}}{\begin{bmatrix}6\\3\end{bmatrix}^T \begin{bmatrix}2 & 2\\2 & 5\end{bmatrix} \begin{bmatrix}6\\3\end{bmatrix}} = \frac{45}{6\cdot 18 + 3\cdot 27} = \frac{5}{21} \\
x_1 &= \begin{bmatrix}0\\0\end{bmatrix} + \frac{5}{21}\begin{bmatrix}6\\3\end{bmatrix} = \begin{bmatrix}10/7\\5/7\end{bmatrix} \\
r_1 &= \begin{bmatrix}6\\3\end{bmatrix} - \frac{5}{21}\begin{bmatrix}18\\27\end{bmatrix} = \begin{bmatrix}12/7\\-24/7\end{bmatrix} \\
\beta_0 &= \frac{r_1^T r_1}{r_0^T r_0} = \frac{144\cdot 5/49}{36+9} = \frac{16}{49} \\
d_1 &= \begin{bmatrix}12/7\\-24/7\end{bmatrix} + \frac{16}{49}\begin{bmatrix}6\\3\end{bmatrix} = \begin{bmatrix}180/49\\-120/49\end{bmatrix} \\
\alpha_1 &= \frac{\begin{bmatrix}12/7\\-24/7\end{bmatrix}^T \begin{bmatrix}12/7\\-24/7\end{bmatrix}}{\begin{bmatrix}180/49\\-120/49\end{bmatrix}^T \begin{bmatrix}2 & 2\\2 & 5\end{bmatrix} \begin{bmatrix}180/49\\-120/49\end{bmatrix}} = \frac{7}{10} \\
x_2 &= \begin{bmatrix}10/7\\5/7\end{bmatrix} + \frac{7}{10}\begin{bmatrix}180/49\\-120/49\end{bmatrix} = \begin{bmatrix}4\\-1\end{bmatrix} \\
r_2 &= \begin{bmatrix}12/7\\-24/7\end{bmatrix} - \frac{7}{10}\begin{bmatrix}2 & 2\\2 & 5\end{bmatrix}\begin{bmatrix}180/49\\-120/49\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.
\end{aligned}
$$

- deoarece $r_2 = b - Ax_2 = 0$, soluţia este $x_2 = [4, -1]^T$

# 3.5.3 Metoda gradienţilor conjugaţi

## Teorema 4

Fie $A$ o matrice $n \times n$ simetrică şi pozitiv definită şi fie $b \neq 0$ un vector. În metoda gradienţilor conjugaţi, presupunem că $r_k \neq 0$ pentru $k < n$ (dacă $r_k = 0$, ecuaţia este rezolvată). Atunci, pentru orice $1 \leq k \leq n$,

(a) următoarele trei subspaţii liniare ale lui $\mathbb{R}^n$ sunt egale:

$$\langle x_1, \ldots, x_k \rangle = \langle r_0, \ldots, r_{k-1} \rangle = \langle d_0, \ldots, d_{k-1} \rangle,$$

(b) rezidualii $r_k$ sunt ortogonali doi câte doi: $r_k^T r_j = 0$ pentru $j < k$,

(c) direcţiile $d_k$ sunt $A$-conjugate două câte două: $d_k^T A d_j = 0$ pentru $j < k$.

- (a) pentru $k = 1$, observăm că $\langle x_1 \rangle = \langle d_0 \rangle = \langle r_0 \rangle$, deoarece $x_0 = 0$
- prin definiţie $x_k = x_{k-1} + \alpha_{k-1} d_{k-1}$
- aceasta implică prin inducţie faptul că $\langle x_1, \ldots, x_k \rangle = \langle d_0, \ldots, d_{k-1} \rangle$
- un argument similar folosind $d_k = r_k + \beta_{k-1} d_{k-1}$ arată că $\langle r_0, \ldots, r_{k-1} \rangle$ este egal cu $\langle d_0, \ldots, d_{k-1} \rangle$
- pentru (b) şi (c), folosim inducţia
- dacă $k = 0$ nu există nimic de demonstrat

# 3.5.3 Metoda gradienţilor conjugaţi

- presupunem că (b) şi (c) au loc pentru $k$, şi vom demonstra (b) şi (c) pentru $k+1$
- înmulţim definiţia lui $r_{k+1}$ cu $r_j^T$ la stânga:

$$r_j^T r_{k+1} = r_j^T r_k - \frac{r_k^T r_k}{d_k^T A d_k} r_j^T A d_k. \tag{8}$$

- dacă $j \leq k-1$, atunci $r_j^T r_k = 0$ din ipoteza de inducţie (b)
- deoarece $r_j$ poate fi exprimat ca o combinaţie de $d_0, \ldots, d_j$, avem $r_j^T A d_k = 0$ din ipoteza de inducţie (c), şi deci (b) are loc
- pe de altă parte, dacă $j = k$, atunci $r_k^T r_{k+1} = 0$ rezultă din nou din (8), deoarece $d_k^T A d_k = r_k^T A d_k + \beta_{k-1} d_{k-1}^T A d_k = r_k^T A d_k$, folosind ipoteza de inducţie (c)
- aceasta demonstrează (b)
- acum că $r_k^T r_{k+1} = 0$, ecuaţia (8) cu $j = k+1$ ne spune că

$$\frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} = -\frac{r_{k+1}^T A d_k}{d_k^T A d_k}. \tag{9}$$

# 3.5.3 Metoda gradienţilor conjugaţi

- aceasta, împreună cu înmulţirea definiţiei lui $d_{k+1}$ la stânga cu $d_j^T A$, ne dau

$$d_j^T A d_{k+1} = d_j^T A r_{k+1} - \frac{r_{k+1}^T A d_k}{d_k^T A d_k} d_j^T A d_k. \tag{10}$$

- dacă $j = k$, atunci $d_k^T A d_{k+1} = 0$ din (10), folosind simetria lui $A$
- dacă $j \leq k-1$, atunci $Ad_j = (r_j - r_{j+1})/\alpha_j$ (din definiţia lui $r_{k+1}$) este ortogonal pe $r_{k+1}$, arătând că primul termen din partea dreaptă a lui (10) este zero, şi al doilea termen este zero din ipoteza de inducţie, ceea ce completează argumentul pentru (c)
- în Exemplul 9, observăm că $r_1$ este ortogonal pe $r_0$, după cum ne garantează Teorema 4
- acest fapt este cheia succesului metodei gradienţilor conjugaţi: fiecare nou rezidual $r_i$ este ortogonal pe toţi rezidualii $r_i$ anteriori
- dacă unul dintre $r_i$ se dovedeşte a fi zero, atunci $Ax_i = b$ şi $x_i$ este soluţia
- dacă nu, după $n$ paşi ai buclei, $r_n$ este ortogonal pe spaţiul liniar generat de cei $n$ vectori ortogonali doi câte doi $r_0, \ldots, r_{n-1}$, care trebuie să fie egal cu $\mathbb{R}^n$
- deci $r_n$ trebuie să fie vectorul zero, şi $Ax_n = b$

# 3.6 Sisteme de ecuaţii neliniare

- Capitolul 2 prezintă metode de rezolvare a unei ecuaţii într-o necunoscută, de obicei neliniară
- în acest capitol, am studiat metode de găsire a soluţiilor pentru sisteme de ecuaţii, dar am impus condiţia ca ecuaţiile să fie liniare
- combinarea dintre neliniaritate şi „mai mult decât o ecuaţie” ridică gradul de dificultate în mod considerabil
- această secţiune descrie metoda lui Newton şi variante ale ei pentru găsirea soluţiei sistemelor de ecuaţii neliniare

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- metoda lui Newton pentru o singură variabilă

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

  oferă ideea principală pentru metoda lui Newton pentru mai multe variabile
- ambele sunt deduse pornind de la aproximarea liniară oferită de dezvoltarea în serie Taylor
- de exemplu, fie

$$\begin{aligned} f_1(u, v, w) &= 0 \\ f_2(u, v, w) &= 0 \\ f_3(u, v, w) &= 0 \end{aligned} \tag{11}$$

  trei ecuaţii neliniare în trei necunoscute $u$, $v$, $w$
- definim funcţia cu valori vectoriale $F(u, v, w) = (f_1, f_2, f_3)$, şi notăm problema (11) prin $F(x) = 0$, unde $x = (u, v, w)$

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- analogul derivatei $f'$ din cazul unei singure variabile este **matricea jacobiană** definită prin

$$DF(x) = \begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} & \frac{\partial f_1}{\partial w} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} & \frac{\partial f_2}{\partial w} \\ \frac{\partial f_3}{\partial u} & \frac{\partial f_3}{\partial v} & \frac{\partial f_3}{\partial w} \end{bmatrix}.$$

- dezvoltarea în serie Taylor pentru funcţii vectoriale în jurul lui $x_0$ este

$$F(x) = F(x_0) + DF(x_0) \cdot (x - x_0) + O((x - x_0)^2).$$

- scriem $f(x) = O(g(x))$ dacă şi numai dacă există $M > 0$ astfel încât $|f(x)| \leq M|g(x)|$
- de exemplu, dezvoltarea liniară a lui $F(u, v) = (e^{u+v}, \sin u)$ în jurul lui $x_0 = (0, 0)$ este

$$\begin{aligned} F(x) &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} e^0 & e^0 \\ \cos 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + O(x^2) \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} u + v \\ u \end{bmatrix} + O(x^2). \end{aligned}$$

- metoda lui Newton se bazează pe o aproximare liniară, ignorând termenii $O(x^2)$

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- ca în cazul unei singure variabile, fie $x = r$ rădăcina, şi fie $x_0$ aproximarea curentă
- atunci

$$0 = F(r) \approx F(x_0) + DF(x_0) \cdot (r - x_0),$$

sau

$$-DF(x_0)^{-1} F(x_0) \approx r - x_0. \tag{12}$$

- prin urmare, o mai bună aproximare a rădăcinii rezultă din rezolvarea ecuaţiei (12) pentru a găsi $r$

**Algoritmul 6 (Metoda lui Newton pentru mai multe variabile)**

$$\begin{aligned} x_0 &= \text{vectorul iniţial} \\ x_{k+1} &= x_k - (DF(x_k))^{-1} F(x_k) \textbf{ for } k = 0, 1, 2, \ldots. \end{aligned}$$

- deoarece calculul inverselor este dificil din punct de vedere computaţional, vom folosi un truc pentru a evita acest calcul
- la fiecare pas, în loc să urmăm definiţia anterioară literal, luăm $x_{k+1} = x_k - s$, unde $s$ este soluţia ecuaţiei $DF(x_k)s = F(x_k)$

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- acum, doar eliminarea gaussiană este necesară pentru a realiza un pas al metodei, în loc de calculul inversei, care necesită de trei ori mai multe operaţii
- prin urmare, pasul de iterare pentru metoda lui Newton pentru mai multe variabile este

$$\begin{cases} DF(x_k)s = -F(x_k) \\ x_{k+1} = x_k + s. \end{cases} \tag{13}$$

## Exemplul 10

- folosiţi metoda lui Newton cu vectorul iniţial $(1, 2)$ pentru a găsi o soluţie a sistemului

$$\begin{aligned} v - u^3 &= 0 \\ u^2 + v^2 - 1 &= 0. \end{aligned}$$

# 3.6.1 Metoda lui Newton pentru mai multe variabile

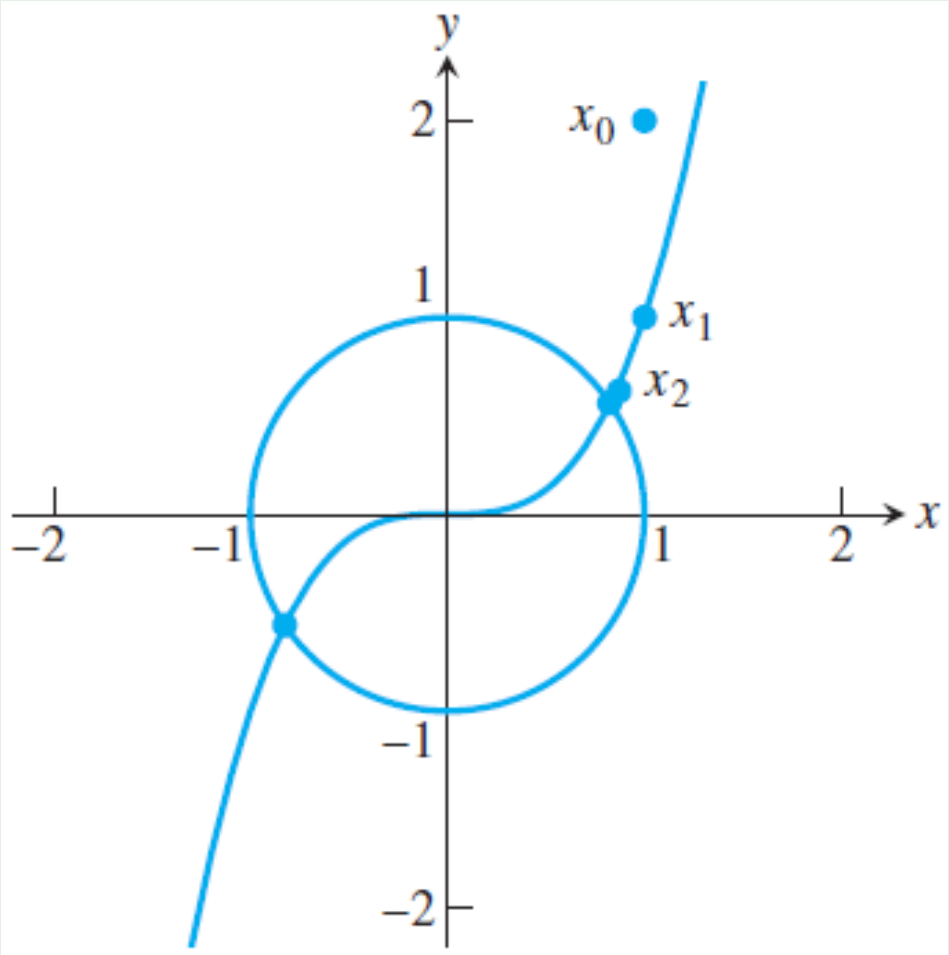


Figura 1: **Metoda lui Newton pentru Exemplul 10.** Cele două rădăcini sunt punctele de pe cerc. Metoda lui Newton produce punctele care converg către soluţia aproximativă $(0.8260, 0.5636)$.

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- Figura 1 prezintă mulţimile pe care $f_1(u, v) = v - u^3$ şi $f_2(u, v) = u^2 + v^2 - 1$ sunt zero şi cele două puncte de intersecţie ale lor, care sunt soluţiile sistemului de ecuaţii
- matricea jacobiană este

$$DF(u, v) = \begin{bmatrix} -3u^2 & 1 \\ 2u & 2v \end{bmatrix}.$$

- folosind punctul iniţial $x_0 = (1, 2)$, în primul pas trebuie să rezolvăm ecuaţia matricială (13):

$$\begin{bmatrix} -3 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = - \begin{bmatrix} 1 \\ 4 \end{bmatrix}.$$

- soluţia este $s = (0, -1)$, astfel că prima iteraţie produce $x_1 = x_0 + s = (1, 1)$

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- al doilea pas necesită rezolvarea ecuaţiei

$$\begin{bmatrix} -3 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = - \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

- soluţia este $s = (-1/8, -3/8)$ şi $x_2 = x_1 + s = (7/8, 5/8)$
- ambele iteraţii sunt prezentate în Figura 1
- următorii paşi sunt detaliaţi în tabelul de mai jos:

| pasul | $u$ | $v$ |
|---|---|---|
| 0 | 1.00000000000000 | 2.00000000000000 |
| 1 | 1.00000000000000 | 1.00000000000000 |
| 2 | 0.87500000000000 | 0.62500000000000 |
| 3 | 0.82903634826712 | 0.56434911242604 |
| 4 | 0.82604010817065 | 0.56361977350284 |
| 5 | 0.82603135773241 | 0.56362416213163 |
| 6 | 0.82603135765419 | 0.56362416216126 |
| 7 | 0.82603135765419 | 0.56362416216126 |

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- familiara dublare a numărului de zecimale exacte la fiecare pas, caracteristică pătratic convergenţei, este evidentă în acest tabel
- simetria ecuaţiilor arată că dacă $(u, v)$ este o soluţie, atunci şi $(-u, -v)$ este o soluţie, după cum se poate observa în Figura 1
- a doua soluţie poate fi de asemenea găsită prin aplicarea metodei lui Newton cu un punct iniţial aflat în vecinătatea acesteia

## Exemplul 11

- folosiţi metoda lui Newton pentru a găsi soluţiile sistemului

$$\begin{aligned} f_1(u, v) &= 6u^3 + uv - 3v^3 - 4 = 0 \\ f_2(u, v) &= u^2 - 18uv^2 + 16v^3 + 1 = 0. \end{aligned}$$

- observăm că $(u, v) = (1, 1)$ este o soluţie
- mai există însă altele două

# 3.6.1 Metoda lui Newton pentru mai multe variabile

- matricea jacobiană este

$$DF(u,v) = \begin{bmatrix} 18u^2 + v & u - 9v^2 \\ 2u - 18v^2 & -36uv + 48v^2 \end{bmatrix}.$$

- care soluţie este găsită de metoda lui Newton depinde de punctul iniţial, exact ca în cazul unidimensional
- folosind punctul iniţial $(u_0, v_0) = (2, 2)$, iterarea formulei anterioare ne dă următorul tabel:

# 3.6.1 Metoda lui Newton pentru mai multe variabile

| pasul | $u$ | $v$ |
|---|---|---|
| 0 | 2.00000000000000 | 2.00000000000000 |
| 1 | 1.37258064516129 | 1.34032258064516 |
| 2 | 1.07838681200443 | 1.05380123264984 |
| 3 | 1.00534968896520 | 1.00269261871539 |
| 4 | 1.00003367866506 | 1.00002243772010 |
| 5 | 1.00000000111957 | 1.00000000057894 |
| 6 | 1.00000000000000 | 1.00000000000000 |
| 7 | 1.00000000000000 | 1.00000000000000 |

- alţi vectori iniţiali conduc la alte două rădăcini, care sunt aproximativ $(0.865939, 0.462168)$ şi $(0.886809, -0.294007)$

- metoda lui Newton este o alegere bună dacă jacobianul poate fi calculat
- dacă nu, cea mai bună alternativă este metoda lui Broyden, subiectul subsecţiunii următoare

# 3.6.2 Metoda lui Broyden

- metoda lui Newton pentru rezolvarea unei ecuaţii într-o necunoscută necesită cunoaşterea derivatei
- dezvoltarea acestei metode în Capitolul 2 a fost urmată de discuţia despre metoda secantei, folositoare pentru cazul în care derivata nu este disponibilă sau este prea greu de evaluat
- acum că avem o versiune a metodei lui Newton pentru sisteme de ecuaţii neliniare $F(x) = 0$, ne confruntăm cu aceeaşi întrebare: Ce se întâmplă dacă matricea jacobiană $DF$ nu este disponibilă?
- deşi nu există o extensie simplă a metodei lui Newton la o metodă a secantei pentru sisteme, Broyden a sugerat o metodă care se apropie cel mai mult de acest deziderat
- să presupunem că $A_i$ este cea mai bună aproximare a matricii jacobiene disponibilă la pasul $i$, şi că a fost folosită pentru a da naştere următoarei aproximări astfel:

$$x_{i+1} = x_i - A_i^{-1} F(x_i). \tag{14}$$

# 3.6.2 Metoda lui Broyden

- pentru a actualiza $A_i$ la $A_{i+1}$ pentru pasul următor, am dori să respectăm aspectul de derivată al jacobianul $DF$, care trebuie să satisfacă
$$A_{i+1}\delta_{i+1} = \Delta_{i+1}, \tag{15}$$
unde $\delta_{i+1} = x_{i+1} - x_i$ şi $\Delta_{i+1} = F(x_{i+1}) - F(x_i)$
- pe de altă parte, pentru complementul ortogonal al lui $\delta_{i+1}$, nu avem alte informaţii noi
- prin urmare, cerem ca
$$A_{i+1}w = A_i w \tag{16}$$
pentru orice $w$ care satisface $\delta_{i+1}^T w = 0$
- o matrice care satisface atât (15) cât şi (16) este
$$A_{i+1} = A_i + \frac{(\Delta_{i+1} - A_i\delta_{i+1})\delta_{i+1}^T}{\delta_{i+1}^T\delta_{i+1}}. \tag{17}$$
- metoda lui Broyden foloseşte pasul din metoda lui Newton (14) pentru a găsi aproximarea următoare, şi actualizează aproximarea jacobianului ca în (17)
- rezumând, algoritmul începe cu un vector iniţial $x_0$ şi o aproximare iniţială a jacobianului $A_0$, care poate fi aleasă matricea identitate

# 3.6.2 Metoda lui Broyden

**Algoritmul 7 (Metoda lui Broyden I)**

$x_0$= vectorul iniţial
$A_0$= matricea iniţială
**for** $i = 0, 1, 2, \ldots$

$$x_{i+1} = x_i - A_i^{-1} F(x_i)$$

$$A_{i+1} = A_i + \frac{(\Delta_{i+1} - A_i \delta_{i+1}) \delta_{i+1}^T}{\delta_{i+1}^T \delta_{i+1}}.$$

**end**
unde $\delta_{i+1} = x_{i+1} - x_i$ şi $\Delta_{i+1} = F(x_{i+1}) - F(x_i)$.

- observăm că pasul de tip Newton se realizează prin rezolvarea ecuaţiei $A_i \delta_{i+1} = F(x_i)$, la fel ca în metoda lui Newton
- tot ca metoda lui Newton, metoda lui Broyden nu oferă garanţia convergenţei către o soluţie
- o a doua abordare a metodei lui Broyden evită pasul relativ dificil din punct de vedere computaţional $A_i \delta_{i+1} = F(x_i)$
- deoarece oricum doar aproximăm derivata $DF$ în timpul iteraţiei, mai bine aproximăm inversa lui $DF$, cea de care avem nevoie în pasul Newton

# 3.6.2 Metoda lui Broyden

- refacem deducerea metodei lui Broyden din punctul de vedere al matricii $B_i = A_i^{-1}$
- ne-ar plăcea să avem

$$\delta_{i+1} = B_{i+1}\Delta_{i+1}, \tag{18}$$

unde $\delta_{i+1} = x_{i+1} - x_i$ şi $\Delta_{i+1} = F(x_{i+1}) - F(x_i)$, şi, pentru orice $w$ care satisface $\delta_{i+1}^T w = 0$, să avem $A_{i+1}w = A_i w$, sau

$$B_{i+1}A_i w = w. \tag{19}$$

- o matrice care satisface atât (18) cât şi (19) este

$$B_{i+1} = B_i + \frac{(\delta_{i+1} - B_i\Delta_{i+1})\delta_{i+1}^T B_i}{\delta_{i+1}^T B_i \Delta_{i+1}}. \tag{20}$$

- noua versiune a iteraţiei, care nu necesită rezolvarea vreunei ecuaţii, este

$$x_{i+1} = x_i - B_i F(x_i). \tag{21}$$

- algoritmul rezultat se numeşte metoda lui Broyden II

# 3.6.2 Metoda lui Broyden

**Algoritmul 8 (Metoda lui Broyden II)**

$x_0$= vectorul iniţial
$B_0$= matricea iniţială
**for** $i = 0, 1, 2, \ldots$

$$x_{i+1} = x_i - B_i F(x_i)$$

$$B_{i+1} = B_i + \frac{(\delta_{i+1} - B_i \Delta_{i+1}) \delta_{i+1}^T B_i}{\delta_{i+1}^T B_i \Delta_{i+1}}$$

**end**
unde $\delta_{i+1} = x_{i+1} - x_i$ şi $\Delta_{i+1} = F(x_{i+1}) - F(x_i)$.

- la început, un vector iniţial $x_0$ şi o aproximare iniţială pentru $B_0$ sunt necesare
- dacă este imposibil să se calculeze derivate, alegerea $B_0 = I$ poate fi folosită

# 3.6.2 Metoda lui Broyden

- un dezavantaj al metodei lui Broyden II este că aproximările jacobianului, necesare pentru anumite aplicaţii, nu sunt uşor disponibile
- matricea $B_i$ este o aproximare a inversei matricii jacobiene
- metoda lui Broyden I, pe de altă parte, actualizează pe $A_i$, care reprezintă o aproximare a jacobianului
- ambele versiuni ale metodei lui Broyden au o convergenţă superliniară, fiind uşor mai lente decât convergenţa pătratică a metodei lui Newton
- dacă o formulă a jacobianului este disponibilă, are loc o accelerare a convergenţei dacă se foloseşte inversa lui $DF(x_0)$ pe post de matrice iniţială $B_0$

Vă mulţumesc!