LABORATOR 6 - METODA GRADIENTILOR CONJUGAŢI. METODA CELOR MAI MICI PĂTRATE

1. Metoda gradienţilor conjugaţi

Metoda gradienţilor conjugaţi face parte din clasa metodelor de relaxare, ca urmare se aplică principiile generale studiate în cursul anterior. Fie deci sistemul $Ax = b$, unde A este simetrică şi pozitiv definită.

Definiţia 1. *Spunem că direcţiile p şi q sunt **direcţii conjugate** în raport cu matricea A dacă $\langle Ap, q\rangle = \langle p, Aq\rangle = 0$.*

Fie $v^{(0)}$ un vector de probă şi $r^{(0)} = Av^{(0)} - b$ vectorul rezidual corespunzător. La metoda gradienţilor conjugaţi pentru rezolvarea sistemului $Ax = b$, prima direcţie de relaxare se alege $p^{(1)} = -r^{(0)}$. În continuare avem:

$$t_{\min} = -\frac{\left\langle r^{(0)}, p^{(1)}\right\rangle}{\left\langle Ap^{(1)}, p^{(1)}\right\rangle} = \frac{\left\langle r^{(0)}, r^{(0)}\right\rangle}{\left\langle Ap^{(1)}, p^{(1)}\right\rangle}, \tag{1.1a}$$

$$v^{(1)} = v^{(0)} + q_1 p^{(1)}, \tag{1.1b}$$

unde

$$q_1 = t_{\min} = -\frac{\left\langle r^{(0)}, p^{(1)}\right\rangle}{\left\langle Ap^{(1)}, p^{(1)}\right\rangle}. \tag{1.2}$$

Valoarea minimă a funcţiei $F = F(v)$, când v parcurge dreapta ce trece prin $v^{(0)}$ şi are direcţia $p^{(1)} = -r^{(0)}$ este realizată în $v^{(1)}$. Fie

$$r^{(1)} = \operatorname{grad} F(v^{(1)}) = Av^{(1)} - b.$$

Din **Observaţia 1** din cursul anterior rezultă că

$$\left\langle r^{(1)}, p^{(1)}\right\rangle = -\left\langle r^{(1)}, r^{(0)}\right\rangle = 0. \tag{1.3}$$

Următoarea direcţie de relaxare $p^{(2)}$ se alege de forma $p^{(2)} = -r^{(1)} + c_1 p^{(1)}$ şi în plus să fie conjugată direcţiei $p^{(1)}$ în raport cu A.

Aşadar, avem:

$$0 = \left\langle Ap^{(2)}, p^{(1)}\right\rangle = \left\langle p^{(2)}, Ap^{(1)}\right\rangle = -\left\langle r^{(1)}, Ap^{(1)}\right\rangle + c_1\left\langle Ap^{(1)}, p^{(1)}\right\rangle,$$

de unde rezultă

$$c_1 = \frac{\left\langle r^{(1)}, Ap^{(1)}\right\rangle}{\left\langle Ap^{(1)}, p^{(1)}\right\rangle}. \tag{1.4}$$

Avem de asemenea

$$v^{(2)} = v^{(1)} + t_{\min} p^{(2)} = v^{(1)} - \frac{\left\langle r^{(1)}, p^{(2)}\right\rangle}{\left\langle Ap^{(2)}, p^{(2)}\right\rangle} p^{(2)}. \tag{1.5}$$

În general, pentru orice $k \geq 2$ obţinem

$$
\begin{aligned}
c_{k-1} &= \frac{\left\langle r^{(k-1)}, Ap^{(k-1)} \right\rangle}{\left\langle Ap^{(k-1)}, p^{(k-1)} \right\rangle}, && (1.6)\\
p^{(k)} &= -r^{(k-1)} + c_{k-1}p^{(k-1)} && (1.7)\\
q_k &= -\frac{\left\langle r^{(k-1)}, p^{(k)} \right\rangle}{\left\langle Ap^{(k)}, p^{(k)} \right\rangle}, && (1.8)\\
v^{(k)} &= v^{(k-1)} + q_k p^{(k)}. && (1.9)
\end{aligned}
$$

Metoda gradienţilor conjugaţi este definită de formulele (1.6)-(1.9). În continuare prezentăm unele simplificări şi proprietăţi suplimentare.

Deoarece $r^{(k-1)}$ este ortogonal pe direcţia $p^{(k-1)}$ rezultă

$$\left\langle r^{(k-1)}, p^{(k)} \right\rangle = \left\langle r^{(k-1)}, -r^{(k-1)} + c_{k-1}p^{(k-1)} \right\rangle = -\left\langle r^{(k-1)}, r^{(k-1)} \right\rangle$$

şi deci

$$q_k = \frac{\left\langle r^{(k-1)}, r^{(k-1)} \right\rangle}{\left\langle Ap^{(k)}, p^{(k)} \right\rangle} > 0. \tag{1.10}$$

Pe de altă parte, din (1.9) avem

$$r^{(k)} = Av^{(k)} - b = Av^{(k-1)} - b + q_k Ap^{(k)} = r^{(k-1)} + q_k Ap^{(k)}.$$

Obţinem deci următoarea relaţie de recurenţă:

$$r^{(k)} = r^{(k-1)} + q_k Ap^{(k)}. \tag{1.11}$$

Observăm că

$$\left\langle r^{(k)}, r^{(k-1)} \right\rangle = 0. \tag{1.12}$$

Într-adevăr, din (1.11) rezultă

$$\left\langle r^{(k)}, r^{(k-1)} \right\rangle = \left\langle r^{(k-1)}, r^{(k-1)} \right\rangle + q_k \left\langle r^{(k-1)}, Ap^{(k)} \right\rangle. \tag{1.13}$$

Pe de altă parte, tinând seama de (1.7) şi (1.10) şi de faptul că $p^{(k)}$ şi $p^{(k-1)}$ sunt A-conjugate, avem

(1.14)

$$q_k \left\langle r^{(k-1)}, Ap^{(k)} \right\rangle = \left\langle r^{(k-1)}, r^{(k-1)} \right\rangle \frac{\left\langle r^{(k-1)}, Ap^{(k)} \right\rangle}{\left\langle Ap^{(k)}, -r^{(k-1)} + c_{k-1}p^{(k-1)} \right\rangle} = -\left\langle r^{(k-1)}, r^{(k-1)} \right\rangle.$$

Din (1.13) şi (1.14) rezultă (1.12).

Deoarece $Ap^{(k)}$ oricum trebuie calculat, rezultă că vectorul rezidual $r^{(k)}$ se va calcula din relaţia de recurenţă (1.11) şi nu prin înlocuirea directă a lui $v^{(k)}$ în expresia $Av = b$.

În continuare vom stabili o altă formulă pentru coeficientul c_{k-1}. Din (1.11) şi (1.12) rezultă

$$\left\langle r^{(k-1)}, Ap^{(k-1)} \right\rangle = \left\langle r^{(k-1)}, \frac{1}{q_{k-1}}(r^{(k-1)} - r^{(k-2)}) \right\rangle = \frac{1}{q_{k-1}} \left\langle r^{(k-1)}, r^{(k-1)} \right\rangle.$$

Ţinând seama de (1.6) şi (1.10) obţinem

$$c_{k-1} = \frac{\left\langle r^{(k-1)}, Ap^{(k-1)} \right\rangle}{\left\langle Ap^{(k-1)}, p^{(k-1)} \right\rangle} = \frac{\left\langle r^{(k-1)}, r^{(k-1)} \right\rangle}{\left\langle r^{(k-2)}, r^{(k-2)} \right\rangle}.$$

Are loc următorul rezultat:

Teorema 1. *La metoda gradienţilor conjugaţi direcţiile de relaxare* $p^{(k)}, k = 1, 2, ...$ *sunt conjugate două câte două în raport cu matricea* A*, iar vectorii reziduali* $r^{(k)}, k = 0, 1, ...$ *sunt ortogonali doi câte doi.*

Demonstraţie. Demonstraţia se face prin inducţie după k. Pentru $k = 1$ avem $\left\langle r^{(1)}, r^{(0)} \right\rangle = 0$ din (1.3), iar pentru prima afirmaţie nu avem ce arăta. Ipoteza inductivă este:

$$\left\langle r^{(i)}, r^{(j)} \right\rangle = 0, \text{ pentru } i \neq j, 0 \leq i, j \leq k \tag{1.15}$$

$$\left\langle p^{(i)}, Ap^{(j)} \right\rangle = 0, \text{ pentru } i \neq j, 1 \leq i, j \leq k. \tag{1.16}$$

Trebuie să arătăm că

$$\left\langle r^{(k+1)}, r^{(j)} \right\rangle = 0, \text{ pentru } 0 \leq j \leq k \tag{1.17}$$

$$\left\langle p^{(k+1)}, Ap^{(j)} \right\rangle = 0, \text{ pentru } 1 \leq j \leq k. \tag{1.18}$$

Dacă $j = k$, atunci $\left\langle p^{(k+1)}, Ap^{(k)} \right\rangle = \left\langle Ap^{(k+1)}, p^{(k)} \right\rangle = 0$ deoarece $p^{(k+1)}$ şi $p^{(k)}$ sunt A-conjugate. Fie $1 \leq j < k$; atunci folosind (1.16)

$$\left\langle p^{(k+1)}, Ap^{(j)} \right\rangle = \left\langle -r^{(k)} + c_{k-1} p^{(k)}, Ap^{(j)} \right\rangle = \left\langle -r^{(k)}, Ap^{(j)} \right\rangle.$$

Dar $Ap^{(j)} = \dfrac{1}{q_j}(r^{(j)} - r^{(j-1)})$ şi din (1.15) obţinem $\left\langle p^{(k+1)}, Ap^{(j)} \right\rangle = 0$. Am demonstrat (1.18).

Pentru $j = k$, (1.17) este adevărată din (1.12). Fie $0 \leq j < k$. Din (1.11) şi (1.15) rezultă

$$\begin{aligned} \left\langle r^{(k+1)}, r^{(j)} \right\rangle &= \left\langle r^{(k)} + q_{k+1} Ap^{(k+1)}, r^{(j)} \right\rangle = q_{k+1} \left\langle Ap^{(k+1)}, r^{(j)} \right\rangle \\ &= q_{k+1} \left\langle Ap^{(k+1)}, -p^{(j+1)} + c_j p^{(j)} \right\rangle. \end{aligned}$$

Folosind relaţia (1.18) şi faptul că A este simetrică, rezultă $\left\langle r^{(k+1)}, r^{(j)} \right\rangle = 0$ şi cu aceasta teorema este demonstrată. □

Din **Teorema 1** rezultă că vectorii reziduali $r^{(k)}$ sunt ortogonali doi câte doi şi deci liniar independenţi dacă sunt nenuli. Cum nu pot exista decât cel mult n vectori liniar independenţi, rezultă că teoretic ar trebui să obţinem soluţia exactă în cel mult n paşi. În practică nu se realizează acest lucru din cauza erorilor de calcul care intervin şi care fac ca vectorii $r^{(k)}$ să nu formeze un sistem ortogonal. În concluzie, se continuă calculul lui $r^{(k)}$ până devine foarte mic sau nul, deoarece valoarea funcţiei pătratice $F(v)$ se micşorează la fiecare pas.

Metoda gradienţilor conjugaţi se dovedeşte foarte utilă pentru sistemele în care matricea A are multe zerouri şi fiecare ecuaţie are o anumită regularitate internă. Astfel de sisteme apar în procesul de discretizare a problemei la limită a ecuaţiilor cu derivate parţiale de tip eliptic.

Algoritmul MATLAB pentru metoda gradienţilor conjugaţi

```
clc
% Introducere A,b,n
```

```
x=zeros(n,1);
r=A*x-b;
for i=1:n
i
    if (i==1)
        p=-r
    else
        c=(r'*A*p)/(p'*A*p)
        p=-r+c*p
    end
    q=(r'*r)/(p'*A*p)
    x=x+q*p
    r=r+q*A*p
end
disp('Solutia aproximativa a sistemului prin metoda gradienţilor conjugaţi')
x
fprintf('\neste obtinuta la pasul %g.\n',i);
disp('Verificare')
A*x
```

2. Metoda celor mai mici pătrate

În procesul de prelucrare a datelor, apar sisteme de ecuaţii supradimensionate şi subdimensionate. Abordăm pentru început problema sistemelor supradimensionate. Fie sistemul:

$$\left\{\begin{array}{l} a_{11}x_1 + ... + a_{1n}x_n = b_1 \\ ... \\ a_{m1}x_1 + ... + a_{mn}x_n = b_m \end{array}\right. , m > n. \tag{2.1}$$

Evident, un asemenea sistem este posibil să nu aibă soluţie. Fie

$$r = Ax - b = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \tag{2.2}$$

vectorul rezidual, unde

$$r_i = a_{i1}x_1 + ... + a_{in}x_n, i = 1, ..., m.$$

Considerăm funcţia pătratică:

$$f(x) = f(x_1, ..., x_n) = \langle r, r\rangle = r_1^2 + r_2^2 + ... + r_m^2. \tag{2.3}$$

Definiţia 2. *Se numeşte **soluţie în sensul celor mai mici pătrate** a sistemului (2.1) acel vector x^* pentru care funcţia (2.3) are valoarea minimă.*

Dacă $\min_{x\in\mathbb{R}^n} f(x) = f(x^*) = 0$, atunci $r_i(x^*) = 0$, pentru orice $i = 1, ..., m$. Rezultă că sistemul (2.1) este compatibil şi $x = x^*$ este soluţia sa exactă.

În general, sistemul (2.1) nu este compatibil şi $\min_{x\in\mathbb{R}^n} f(x) = f(x^*) > 0$, iar $x = x^*$ este un substitut pentru soluţia sistemului şi anume soluţia în sensul celor mai mici pătrate.

Funcţia f se poate pune sub forma

$$
\begin{aligned}
(2.4) \qquad f(x) &= \langle r,r\rangle = \langle Ax-b, Ax-b\rangle = \langle Ax, Ax\rangle - 2\langle Ax, b\rangle + \langle b,b\rangle \\
&= \left\langle A^T Ax, x\right\rangle - 2\left\langle A^T b, x\right\rangle + \langle b,b\rangle .
\end{aligned}
$$

Teorema 2. *Dacă $rang(A) = n$, atunci sistemul (2.1) admite o unică soluţie în sensul celor mai mici pătrate şi aceasta este soluţia (unică) a sistemului*

$$
A^T Ax = A^T b. \tag{2.5}
$$

Sistemul (2.5) se numeşte ***sistemul normal al lui Gauss.***

Demonstraţie. Punctele de extrem ale funcţiei f se găsesc printre punctele critice, adică $\operatorname{grad} f(x_0) = 0$ dacă x_0 este punct de extrem, echivalent $A^T Ax_0 - A^T b = 0$. Cum $rang(A) = rang(A^T) = rang(A^T A) = n$, obţinem că sistemul (2.5) admite o unică soluţie x^*, care este punct critic pentru f.

Matricea $B = A^T A$ este simetrică şi semipozitiv definită. Dacă presupunem că $\langle Bx, x\rangle = 0$, obţinem $\langle Ax, Ax\rangle = 0$, deci $Ax = 0$. Cum $rang(A) = n < m$, avem $x = 0$, deci B este pozitiv definită. Deoarece $d^2 f(x) = 2\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} dx_i dx_j$, pentru orice x, obţinem că diferenţiala de ordinul doi este pozitiv definită, deci unicul punct critic x^* este punct de minim pentru f. $\square$

Aşadar, în ipoteza $rang(A) = n$, soluţia sistemului (2.1), în sensul celor mai mici pătrate, este unică şi este soluţia sistemului (2.5). Acest sistem este simetric şi pozitiv definit, aşadar se rezolvă prin metoda Cholesky sau una din metodele de relaxare.

Observaţia 1. *Soluţia sistemului (2.5) este $x^* = (A^T A)^{-1} A^T b$. Matricea $P = (A^T A)^{-1} A^T$ se numeşte* ***pseudoinversa*** *matricii (dreptunghiulare) A. Dacă A este pătratică, obţinem $P = A^{-1}(A^T)^{-1} A^T = A^{-1}$, deci noţiunea de pseudoinversă generalizează noţiunea de matrice inversă pentru matrici dreptunghiulare.*

Exemplul 1. ***Dreapta de regresie***

Să presupunem că vrem să găsim o dreaptă $y = mx + n$ care să treacă prin punctele

$$
M_1(0,0), M_2(2,1), M_3(5,3), M_4(8,5), M_5(10,6).
$$

Se obţine astfel sistemul

$$
\begin{cases}
0\cdot m + n = 0 \\
2\cdot m + n = 1 \\
5\cdot m + n = 3 \\
8\cdot m + n = 5 \\
10\cdot m + n = 6
\end{cases} . \tag{2.6}
$$

Sistemul (2.6) este supradimensionat şi incompatibil. Avem:

$$
A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 5 & 1 \\ 8 & 1 \\ 10 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 5 \\ 6 \end{bmatrix}, B = A^T A = \begin{bmatrix} 193 & 25 \\ 25 & 5 \end{bmatrix}, A^T b = \begin{bmatrix} 117 \\ 15 \end{bmatrix}.
$$

Ecuaţiile normale ale lui Gauss sunt:

$$\begin{cases} 193m + 25n = 117 \\ 25m + 5n = 15 \end{cases}.$$

Soluţia exactă este $m = \dfrac{21}{34}, n = -\dfrac{3}{34}$*, iar dreapta de regresie este* $y = \dfrac{21}{34}x - \dfrac{3}{34}$*. Această dreaptă nu trece prin punctele* M_i*, dar este acea dreaptă din plan care trece (într-un anume sens) cel mai aproape de aceste puncte.*

Sistemele subdimensionate apar în probleme legate de ajustarea datelor. Să presupunem că măsurând n cantităţi $x_1, x_2, ..., x_n$ găsim valorile $l_1, l_2, ..., l_n$. Pe de altă parte, necunoscutele $x_1, x_2, ..., x_n$ satisfac anumite ecuaţii, şi anume:

$$\begin{cases} a_{11}x_1 + ... + a_{1n}x_n = b_1 \\ ... \\ a_{m1}x_1 + ... + a_{mn}x_n = b_m \end{cases}, m < n. \tag{2.7}$$

Datorită lipsei de acurateţe a măsurătorilor se pune condiţia ca sumele pătratelor corecţiilor $\sum_{i=1}^{n}(l_i - x_i)^2$ să fie minimă. Se obţine astfel o problemă de extrem cu legături.

Exemplul 2. *Să presupunem că măsurând unghiurile unui triunghi* x_1, x_2, x_3 *găsim valorile* l_1, l_2, l_3*. Evident avem legătura* $x_1 + x_2 + x_3 = \pi$*. Impunem condiţia ca abaterile datorate impreciziei măsurătorilor să fie cât mai mici, deci* $(x_1 - l_1)^2 + (x_2 - l_2)^2 + (x_3 - l_3)^2$ *să fie minimă.*

Revenind la cazul general, se pune problema să determinăm n necunoscute $x_1, x_2, ..., x_n$ care satisfac legăturile (2.7) şi care minimizează funcţia:

$$f(x_1, ..., x_n) = (x_1 - l_1)^2 + ... + (x_n - l_n)^2.$$

Considerăm funcţia auxiliară

$$\begin{aligned} \Phi(x_1, ..., x_n, \lambda_1, ..., \lambda_m) \quad = \quad & (x_1 - l_1)^2 + ... + (x_n - l_n)^2 \\ & +\lambda_1(a_{11}x_1 + ... + a_{1n}x_n - b_1) \\ & +... + \lambda_m(a_{m1}x_1 + ... + a_{mn}x_n - b_m). \end{aligned}$$

Punctele critice ale funcţiei Φ sunt date de sistemul

$$\begin{cases} 2(x_1 - l_1) + (a_{11}\lambda_1 + ... + a_{m1}\lambda_m) = 0 \\ ... \\ 2(x_n - l_n) + (a_{1n}\lambda_1 + ... + a_{mn}\lambda_m) = 0 \\ a_{11}x_1 + ... + a_{1n}x_n - b_1 = 0 \\ ... \\ a_{m1}x_1 + ... + a_{mn}x_n - b_m = 0 \end{cases}. \tag{2.8}$$

Dacă notăm $l = [l_1, ..., l_m]^T$ şi $\lambda = [\lambda_1, ..., \lambda_m]^T$, atunci sistemul 2.8 se scrie sub forma matriceală:

$$x = \frac{1}{2}A^T\lambda + l \tag{2.9}$$

$$Ax = b \tag{2.10}$$

Înlocuind (2.9) în (2.10) obţinem sistemul:

$$\frac{1}{2}AA^T\lambda = b - Al. \tag{2.11}$$

Dacă $rang(A) = m$, atunci $rang(A^T A) = m$ şi sistemul (2.11) are soluţie unică. Mai mult, matricea AA^T este pătratică, simetrică şi pozitiv definită, deci rezolvarea sistemului (2.11) se poate face cu metoda Cholesky sau cu una din metodele de relaxare. Rezolvând sistemul (2.11) obţinem multiplicatorii Lagrange $\lambda_1, ..., \lambda_m$, iar din relaţia $x = \frac{1}{2}A^T\lambda + l$ determinăm punctul de minim condiţionat al funcţiei f (cum $d^2\Phi = 2(dx_1^2 + ... + dx_n^2) > 0$ rezultă că avem într-adevăr un punct de minim).