DIDACTICA MATHEMATICA, Vol. 37 (2019) , pp. 29–43

METODA LUI NEWTON ÎN OPTIMIZAREA CONVEXĂ. APLICAŢII ÎN PYTHON

Luminiţa JARDA

Abstract. Within this paper we present from both a theoretical and an applied point of view Newton's approximation method in the general context of convex optimization.

MSC 2000. 54H25.
Key words. fixed point property, coincidence theorems, commuting operators.

1. FORMULAREA PROBLEMEI I DEFINIREA SOLUIEI

Contextul general n care vom lucra este urmtorul:

$$\begin{cases} f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \\ dom f \neq \emptyset \\ f \text{ este convex} \\ f \text{ este de clas } C^2 \text{ pe } dom f \end{cases}$$

Funcia f va fi numit **funcie obiectiv**.

Dorim s determinm soluiile optime ale urmtoarei probleme de optimizare:

$$(P) \quad \begin{cases} \min f(x) \\ x \in dom f \end{cases}$$

Pentru aceasta, vom presupune c problema are soluie, adic exist un punct de optim pentru aceasta, pe care l notm cu x^*. Un vector x^* din $dom f$ se numete **soluie pentru problema de optimizare (P)** dac $f(x^*) \leq f(x), \quad \forall x \in dom f$. **Valoarea minim a problemei (P)** o vom nota cu p^*, unde $p^* = f(x^*) = inf_x f(x)$.

Avnd n vedere c f este convex i difereniabil, aceast problem ar putea fi rezolvat punnd condiia ca gradientul funciei s se anuleze n x^*, ceea ce reprezint un sistem de n ecuaii n n variabile $(x_1, ..., x_n)$. Din pcate, doar n cteva cazuri speciale putem determina soluia acestui sistem n mod analitic.

Tocmai de aceea, n aceast lucrare nu studiem aceast abordare a soluiilor optime, ci vom calcula succesiv valorile unui ir de vectori $x^{(0)}, x^{(1)}, ... \in dom f$ astfel nct $f(x^{(k)}) \to p^*$, cnd $k \to \infty$. Condiia de oprire a algoritmului este ca $f(x^k) - p^* \leq \epsilon$, unde ϵ este o toleran specificat. Cu ajutorul unui astfel de algoritm vom putea obine o aproximaie a soluiei x^* cu precizia ϵ.

- **Punct de start**

Pentru metoda pe care o vom studia, cerem ca punctul de start, $x^{(0)}$, s satisfac urmtoarele condiii: $x^{(0)} \in dom f = \mathbb{R}^n$, iar mulimea sa de nivel,

$$S = \{x \in dom f \mid f(x) \leq f(x^{(0)})\},$$

s fie nchis. n particular, dac f este o funcie continu cu domeniul $dom f = \mathbb{R}^n$, toate mulimile ei de nivel sunt nchise. Prin urmare, pentru problema de minimizare considerat, cu $dom f = \mathbb{R}^n$, putem alege orice punct de start $x^{(0)} \in \mathbb{R}^n$.

- **Convexitate tare i consecine**

n cele ce urmeaz, presupunem c funcia f este tare convex pe S, ceea ce nseamn (pe lng faptul c f este convex), scris cu ajutorul unor inegaliti matriceale, c exist $m > 0$ astfel nct

$$mI \preceq \nabla^2 f(x), \forall x \in S. \tag{1}$$

Convexitatea tare are cteva consecine interesante. Una dintre ele este cea care urmeaz. Pentru $x, y \in S$, avem c

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \tfrac{1}{2}(y - x)^T \nabla^2 f(z)(y - x),$$

pentru un z aparinnd segmentului $[x, y]$. Din condiia de convexitate tare avem c ultimul termen din membrul drept trebuie s fie cel puin $\frac{m}{2}\|y - x\|_2^2$, aadar:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2}\|y - x\|_2^2, \forall x, y \in A. \tag{2}$$

Inegalitatea (2) poate fi folosit pentru a mrgini $f(x^*) - p^*$. Membrul drept al inegalitii (2) este o funcie convex cuadratic de y (pentru un x fixat). Punnd condiia ca gradientul n raport cu y s fie 0, gsim c $\tilde{y} = x - \frac{1}{m}\nabla f(x)$ minimizeaz membrul drept al inegalitii. Avem aadar c

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^T (y - x) + \tfrac{m}{2}\|y - x\|_2^2 \\ &\geq f(x) + \nabla f(x)^T (\tilde{y} - x) + \tfrac{m}{2}\|\tilde{y} - x\|_2^2 \\ &= f(x) - \tfrac{1}{2m}\|\nabla^2 f(x)\|_2^2. \end{aligned}$$

Acest lucru fiind valabil pentru orice $y \in S$, avem c

$$p^* \geq f(x) - \frac{1}{2m}\|\nabla^2 f(x)\|_2^2. \tag{3}$$

Ultima inegalitate ne spune c dac gradientul ntr-un punct este suficient de mic, atunci punctul este aproape de optim. Inegalitatea (3) poate fi interpretat i ca o condiie de suboptimalitate, generaliznd condiia de optimalitate (ca gradientul s fie nul):

$$\|\nabla f(x)\|_2 \leq (2m\epsilon)^{1/2} \Rightarrow f(x) - p^* \leq \epsilon. \tag{4}$$

Putem, de asemenea, s mrginim distana dintre un punct x i orice punct de optim, x^*. Pentru aceasta, folosim inegaliatatea (2) cu $y = x^*$:

$$p^* = f(x^*) \geq f(x) + \nabla f(x)^T(x^* - x) + \tfrac{m}{2}\|x^* - x\|_2^2$$
$$\geq f(x) - \|\nabla f(x)\|_2\|x^* - x\|_2 + \tfrac{m}{2}\|x^* - x\|_2^2,$$

unde pentru a doua inegalitate am folosit inegalitatea Cauchy-Buniakovski-Schwarz.

Din moment ce $p^* \leq f(x)$, trebuie s avem c

$$-\|\nabla f(x)\|_2\|x^* - x\|_2 + \tfrac{m}{2}\|x^* - x\|_2^2 \leq 0,$$

ceea ce nseamn c

$$\|x^* - x\|_2 \leq \tfrac{2}{m}\|\nabla f(x)\|_2.$$

O consecin a acestui rezultat este faptul c **punctul optimal x^* este *unic*.**

Inegalitatea (2) implic faptul c mulimile de nivel ale lui S sunt mrginite, deci n particular i S este mrginit. De aici rezult c i valoarea proprie maxim a funciei $\nabla^2 f(x)$, care este o funcie continu pe S, este mrginit superior pe S, ceea ce nseamn c exist o constant $M > 0$ astfel nct

$$\nabla^2 f(x) \preceq MI, \forall x \in S.$$

Aceast margine superioar pentru matricea Hessian implic faptul c pentru orice $x, y \in S$, avem c:

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \tfrac{M}{2}\|y - x\|_2^2,$$

relaie care este analoaga inegalitii (2). Minimizm fiecare membru dup y i obinem:

$$p^* \leq f(x) - \tfrac{1}{2M}\|\nabla^2 f(x)\|_2^2,$$

echivalenta relaiei (3).

Observaţia 1. *Este important de precizat c cele dou constante specifice convexitii tari, m i M, sunt rareori cunoscute n practic. Tocmai de aceea, condiia (4) nu poate fi folosit ca un criteriu practic de oprire. Ea este mai degrab un criteriu de oprire conceptual, care ne spune c pentru valori suficient de mici ale gradientului ne apropiem din ce n ce mai mult de valoarea optim. Prin urmare, dac terminm algoritmul cu $\|\nabla f(x)\|_2 \leq \eta$, unde η este ales suficient de mic, foarte probabil mai mic dect $(2m\epsilon)^{1/2}$, atunci vom avea i c $f(x^{(k)}) - p^* \leq \epsilon$ (foarte probabil).*

n cele ce urmeaz vom demonstra convergena metodei lui Newton, stabilind nite margini superioare pentru numrul de iteraii necesare pn cnd $f(x^{(k)}) - p^* \leq \epsilon$, aceste valori depinznd de constantele m i M. Rezultatele obinute sunt valoroase conceptual, deoarece ele ne arat c algoritmul converge, chiar dac marginea superioar a numrului de iteraii depinde de constante care sunt necunoscute.

2. SCURT ISTORIC. IDEEA METODEI LUI NEWTON

Metoda lui Newton pornete de la ideea de liniarizare. Mai exact, presupunnd c avem o funcie $g : \mathbb{R} \to \mathbb{R}$ derivabil, vom ncerca s rezolvm ecuaia

$$g(x) = 0.$$

Pornind cu un punct iniial, x_0, se consider aproximaia liniar a lui g n apropierea lui x_0:

$$g(x_0 + \triangle x) \approx g(x_0) + g'(x_0)\triangle x.$$

Prin urmare, rezolvnd noua ecuaie (una liniar), $g(x_0) + g'(x_0)\triangle x = 0$, vom obine o aproximaie a soluiei ecuaiei $g(x) = 0$. Considernd acest proces iterativ, avem c

$$x_{k+1} = x_k - g'(x_k)^{-1}g(x_k),\ k = 0, 1, ..$$

Newton a venit cu aceast metod n 1669. Este important de precizat c Newton a lucrat numai cu funcii polinomiale. Exemplul pe care a demonstrat metoda este urmtorul:

$$g(x) = x^3 - 2x - 5.$$

O prim aproximaie (punctul de start) a fost $x_0 = 2$, ecuaia devenind

$$g(2 + \triangle x) = (\triangle x)^3 + 6(\triangle x)^2 + 10\triangle x - 1 = 0.$$

Newton a neglijat termenii superiori i a rezolvat ecuaia liniar $10\triangle x - 1 = 0$. Noua aproximaie a soluiei a fost aadar $x = 2 + 0.1 = 2.1$. Apoi procesul se repet pornind de la noul punct. Convergena spre rdcin s-a dovedit a fi una foarte rapid.

Forma general ($x_{k+1} = x_k - g'(x_k)^{-1}g(x_k)$) a fost introdus ns n 1690 de ctre J. Raphson, motiv pentru care metoda se mai numete Newton-Raphson. Acesta nu a considerat funcia ca fiind neaprat polinomial, ci o funcie derivabil.

La progresul acestei metode au contribuit i ali matematicieni celebri. Fourier a demonstrat n 1818 c metoda converge cuadratic n apropierea rdcinii, iar Cauchy (1829, 1847) a extins metoda pentru cazul multidimensional.

Datele istorice au fost preluate din lucrarea lui Polyak [3].

3. METODE DESCENDENTE. CUTARE BACKTRACKING

Pentru a rezolva problema de minimizare (P), ne propunem s determinm un ir minimizant $x^{(k)}$, unde $k = 0, 1, ...$, iar

$$x^{(k+1)} = x^{(k)} + t^{(k)}\triangle x^{(k)}.$$

Numrul iteraiei este dat de k, $\triangle x^{(k)}$ este un vector din $\mathbb{R}^n$ i se numete *pasul* sau *direcia de cutare* (chiar dac nu e nevoie s aib norma unitar), iar $t^{(k)} \geq 0$ ($t^{(k)} \in \mathbb{R}$) se numete *mrimea pasului* sau *lungimea pasului* la iteraia k (=0 cnd $x^{(k)}$ este punct de optim) .

O metod este **considerat metod descendent dac**

$$f(x^{(k+1)}) < f(x^{(k)}),$$

cu excepia situaiei n care $x^{(k)}$ este punct de optim. Asta implic faptul c pentru orice k avem c $x^{(k)} \in S$, unde S este mulimea de nivel iniial, iar n particular $x^{(k)} \in dom f$. Din convexitate tim c $\nabla(x^{(k)})^T(y - x^{(k)}) \geq 0$ implic $f(y) \geq f(x^{(k)})$, aadar direcia de cutare pentru o metod descendent trebuie s satisfac

$$\nabla(x^{(k)})^T(y - x^{(k)}) < 0.$$

DEFINIŢIA 1. *Numim* ***direcie descendent*** *a funciei f n punctul $x^{(k)}$ un vector care face un unghi optuz cu gradientul funciei n acel punct.*

ALGORITMUL 1. *Algoritm general pentru o metod descendent*
Se d un punct de start, $x \in domf$.
repeat
1. Determin o direcie descendent, $\triangle x$.
2. Alege mrimea pasului, $t > 0$.
3. Modific: $x = x + t\triangle x$.
until *criteriul de oprire este satisfcut.*

n ce privete criteriul de terminare, acesta poate fi verificat i imediat dup ce este determinat direcia descendent, $\triangle x$.

Cel de-al doilea pas ne spune unde de-a lungul semidreptei $\{x + t\triangle x \mid t \in \mathbb{R}_+\}$ va fi urmtoarea iteraie.

OBSERVAŢIA 2. *Mrimea pasului, t, poate fi determinat prin mai multe metode. Una dintre ele este o metod exact, analitic, cnd determinm* $t = argmin_{s\geq 0} f(x + s\triangle x)$. *Aceast metod se aplic atunci cnd costul (timp, memorie) aplicrii ei este sczut comparativ cu determinarea direciei descendente nsei.*

• Metoda Backtracking pentru determinarea mrimii pasului

Majoritatea metodelor de determinare a mrimii pasului folosite n practic sunt *inexacte*. Una dintre ele este o metod de tip *backtracking*. Acest algoritm de determinare a mrimii pasului utilizat n cadrul metodei lui Newton (la care ne vom referi ulterior) funcioneaz aproape la fel de eficient ca o metod *exact* n ceea ce privete convergena metodei. Algoritmul depinde de dou constante, α i β, cu $0 < \alpha < 0.5$ i $0 < \beta < 1$.

ALGORITMUL 2. *Backtracking pentru msura pasului*
Se d o direcie descendent $\triangle x$ *pentru* f *i un punct* $x \in domf$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$.
$t := 1$.
while $f(x + t\triangle x) > f(x) + \alpha t\nabla(x)^T\triangle x$ ***do***
$t := \beta t$.
end while

Aceast metod se numete *backtracking* deoarece ncepe cu un pas unitate i l reduce prin factorul β pn la ndeplinirea condiiei de oprire, $f(x + t\triangle x) \leq f(x) + \alpha t\nabla(x)^T\triangle x$. Din moment ce $\triangle x$ este o direcie descendent, avem c $\nabla(x)^T\triangle x < 0$, aadar pentru un t suficient de mic avem c

$$f(x + t\triangle x) \approx f(x) + t\nabla(x)^T\triangle x < f(x) + \alpha t\nabla(x)^T\triangle x,$$

ceea ce sugerez faptul c algoritmul se termin n cele din urm.

Condiia de backtracking este ilustrat n Figura 3.1.

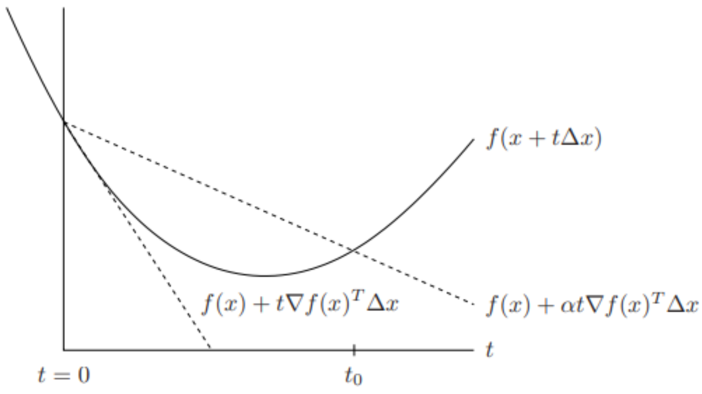

Figura. 3.1 – *Cutare backtracking.* Condiia de backtracking este ca f s rmn dedesubtul liniei punctate $f(x) + \alpha t \nabla(x)^T \triangle x$, adic $0 \leq t \leq t_0$.

4. METODA LUI NEWTON N OPTIMIZARE. EXEMPLE N $\mathbb{R}$

Ideea de optimizare presupune existena unei funcii obiectiv, pe care dorim fie s o minimizm, fie s o maximizm. Din punct de vedere teoretic, ajunge studiat un singur caz, pentru c maximizarea unei funcii f este echivalent cu minimizarea funciei $-f$. Pentru aceasta, pornim de la un punct oarecare al domeniului funciei i realizm modificri asupra lui dup o anumit regul. Procesul se ncheie fie cnd am ajuns suficient de aproape de valoarea optim, fie dup un numr maxim de iteraii prestabilit.

Pentru a nelege cum se aplic metoda lui Newton n optimizare, vom considera o funcie $f : \mathbb{R} \to \mathbb{R}$ polinomial. Fie ea

$$f(x) = -2x^3 + 8x^2 - 7x + 2.$$

Codul surs pentru implementarea metodei pe acest exemplu este scris n Python. Acest exemplu a fost inspirat din articolul Valentinei Alto [1]. Avem mai jos reprezentarea funciei pe intervalul [-1,1.3]. Am considerat un interval pe care funcia f este convex.

```
import numpy as np
import matplotlib.pyplot as plt
def f(x):
    return -2*x**3+8*x**2-7*x+2
x = np.linspace(-1, 1.3)
fig, ax = plt.subplots()
ax.plot(x, f(x), label='target')
ax.grid()
plt.legend()
```

Este important de precizat c funciile pe care se aplic metoda lui Newton trebuie s fie de dou ori derivabile, iar $f''(x) \neq 0$.

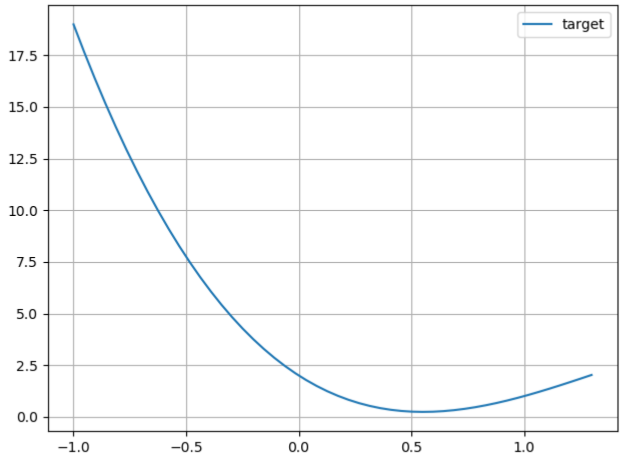

Figura. 4.2 – $f(x) = -2x^3 + 8x^2 - 7x + 2$ pe intervalul [-1,1.3]

Pentru a ndeplini i condiia cu privire la derivata a doua nenul, vom considera funcia f restrns la intervalul [-1,1.3] i pe acest interval vom cuta punctul de optim (n acest caz, de minim).

Metoda lui Newton n optimizare pornete de la ideea de a aproxima curba noastr cu funcia lui Taylor de ordin doi. Astfel, vom determina aceast aproximaie pentru un punct iniial $x = x_0$, dup care vom calcula miniimul acestei funcii. Punctul n care aproximaia cuadratic atinge valoarea minim va fi noul nostru x. Procesul continu iterativ, obinndu-se astfel o secven care se apropie tot mai mult de punctul de optim, x^*.

Mai nti calculm aproximaia lui Taylor n punctul iniial x_0:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2.$$

Pentru a determina minimul acestei funcii, punem condiia ca prima ei derivat s fie nul. Derivata e calculat n raport cu $(x - x_0)$, care reprezint ce trebuie adugat la x_0 pentru a obine minimul:

$$\frac{\partial[f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2]}{\partial(x-x_0)} = 0$$

$$f'(x_0) + f''(x_0)(x - x_0) = 0$$

$$x = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

Procesul fiind unul iterativ, putem generaliza formula:

$$x_{t+1} = x_t - \frac{f'(x_t)}{f''(x_t)}.$$

Pentru funcia dat mai sus va trebui s calculm aadar prima derivat, a doua derivat, la care alturm i aproximaia cuadratic:

```
def fprime(x):
    return -6*x**2+16*x-7
```

```
def fsecond(x):
    return -12*x + 16
def quadraticapprox(x, x0, f, fprime, fsecond):
    return f(x0)+fprime(x0)*(x-x0)+0.5*fsecond(x0)*(x-x0)**2
```

Acum putem defini metoda lui Newton, folosind o precizie epsilon pentru a ne apropia de optim:

```
def newton(x0, fprime, fsecond, maxiter=100, eps=0.0001):
    x=x0
        for i in range(maxiter):
            xnew=x-(fprime(x)/fsecond(x))
            if np.abs(xnew-x¡eps):
                return xnew
            x = xnew
    return x
```

Realiznd un grafic al funciei noastre, precum i pentru aproximaia cuadratic n x^* obinut prin metoda lui Newton, putem observa c minimul lor coincide aproximativ.

```
xstar=newton(0, fprime, fsecond)
fig, ax = plt.subplots()
ax.plot(x, f(x), label='target')
ax.grid()
ax.plot(x, quadraticapprox(x, xstar , f, fprime, fsecond),
color='red', label='quadratic approximation')
ax.set_ylim([-2,3])
ax.axhline(y=0, color='k')
ax.axvline(x=0, color='k')
ax.axvline(x = xstar, color='green')
plt.legend()
```

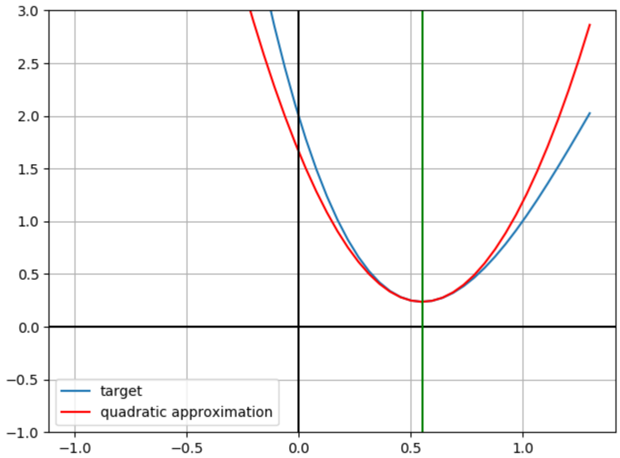

Figura. 4.3 – Funcia $f(x) = -2x^3 + 8x^2 - 7x + 2$ este ilustrat cu albastru, iar aproximaia cuadratic n x^* determinat cu metoda lui Newton este ilustrat cu rou. Cu verde este ilustrat verticala ce trece prin x^*.

OBSERVAŢIA 3. *n cazul de fa, timpul de rulare al algoritmului nu a fost lung, pentru c am pornit din apropierea soluiei* ($x_0 = 0$). *Metoda lui Newton s-a dovedit ns a fi extrem de eficient i pentru situaii mai puin avantajoase ale alegerii punctlui de start. Cu toate acestea, ea poate fi aplicat numai n condiiile precizate mai sus, adic n cazul funciilor care se comport bine, cum este funcia noastr polinomial.*

- **Aplicaia 2**

Fie funcia $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = 7x^4 + e^x - cos(x),$$

o funcie convex pe $\mathbb{R}$, nepolinomial de data aceasta. Pentru reprezentare vom folosi ns numai intervalul $[-1, 1]$, unde se gsete de fapt i punctul de minim.

Codul pentru implementarea metodei lui Newton pe acest exemplu este similar cu cel folosit anterior, cu urmtoarele diferene pentru definirea funciilor (f i primele dou derivate ale sale), precum i a intervalului de reprezentare grafic:

```
def f(x):
    return 7*x**4+np.exp(x)-np.cos(x)
x = np.linspace(-1, 1)

def fprime(x):
    return 28*x**3+np.exp(x)+np.sin(x)
def fsecond(x):
    return 84*x**2 + np.exp(x)+np.cos(x)
```

n continuare ne vom referi la metoda lui Newton pe $\mathbb{R}^n$.

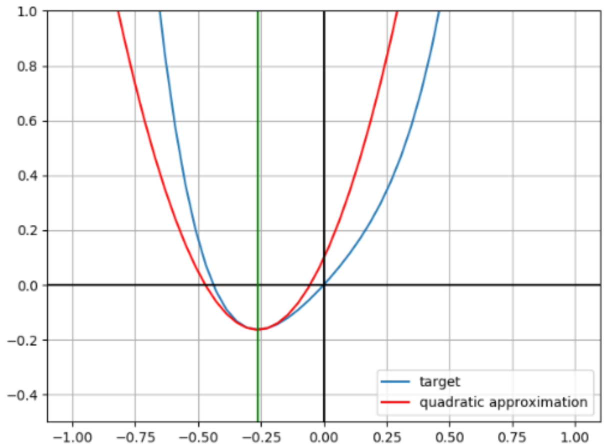

Figura. 4.4 – Funcia obiectiv, $f(x) = 7x^4 + e^x - cos(x)$, este ilustrat cu albastru, iar aproximaia cuadratic n x^* determinat cu metoda lui Newton este ilustrat cu rou. Cu verde este ilustrat verticala ce trece prin x^*. Se observ c minimele coincid aproximativ.

5. PASUL LUI NEWTON

DEFINIŢIA 2. *Pentru un $x \in dom f = \mathbb{R}^n$, definim pasul lui Newton(direcia lui Newton) ca fiind vectorul:*

$$\triangle x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x). \tag{5}$$

OBSERVAŢIA 4. *Se poate observa c pentru $n = 1$, pasul lui Newton este cel folosit n exemplul anterior, adic $-\frac{f'(x)}{f''(x)}$. De asemenea, este important de precizat c n exemplul de mai sus s-a luat un pas unitar (msura pasului $t = 1$). Pe $\mathbb{R}^n$ vom lucra i cu pai care nu sunt unitari, iar paii unitari vor corespunde fazei pure din aplicarea metodei lui Newton.*

Avnd n vedere c matricea Hessian, $\nabla^2 f(x)$, este pozitiv semidefinit n fiecare punct la domeniului de definiie (f este convex), este uor de verificat c direcia(pasul) lui Newton este una *descendent*:

$$\nabla f(x)^T \triangle x_{nt} = -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0,$$

pentru orice x cu excepia situaiei n care x este optim ($\nabla f(x) = 0$).

- **Minimizator pentru a doua aproximaie a lui Taylor**

La fel ca n exemplul pe $\mathbb{R}$, pasul lui Newton reprezint ce trebuie adugat la a doua aproximaie a lui Taylor pentru f ca aceasta(aproximaia) s ia valoarea minim.

- **Soluie pentru condiia liniar de optimalitate**

Condiia de optimalitate pentru un punct x^* este ca $\nabla f(x^*) = 0$. Liniariznd aceast condiie n apropierea lui x, avem c

$$\nabla f(x + v) \approx \nabla f(x) + \nabla^2 f(x) v = 0.$$

Aceast ecuaie liniar are soluia $v = \triangle x_{nt}$. Prin urmare, pasul lui Newton este ceea ce trebuie adugat la x pentru ca s aib loc condiia liniarizat de optimalitate.

• **Invarina afin**

O caracteristic foarte important a pasului lui Newton este faptul c este **afin invariant**. Cu alte cuvinte, pasul lui Newton este independent de schimbri afine de coordonate.

Demonstraţie. Fie $T \in \mathbb{R}^{n\times n}$ o matrice nesingular i definim $g(y) = f(Ty)$. Atunci, avem c:

$$\nabla g(y) = T^T \nabla f(x), \text{ iar } \nabla^2 g(y) = T^T \nabla^2 f(x) T,$$

unde $x = Ty$. Calculm pasul lui Newton pentru g n y:

$$\triangle y_{nt} = -(T^T \nabla^2 f(x) T)^{-1} (T^T \nabla f(x))$$

$$\triangle y_{nt} = -T^{-1} \nabla^2 f(x)^{-1} \nabla f(x)$$

$$\triangle y_{nt} = T^{-1} \triangle x_{nt}.$$

Prin urmare, pasul lui Newton pentru f i pentru g sunt legate prin aceeai transformare liniar, iar:

$$x + \triangle x_{nt} = T(y + \triangle y_{nt}).$$

□

6. DECREMENTUL LUI NEWTON

DEFINIŢIA 3. *Numim decrementul lui Newton n punctul x urmtoarea cantitate:*

$$\lambda(x) = (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}. \tag{6}$$

Decrementul lui Newton este foarte important n studiul convergenei metodei lui Newton. El este folosit, de asemenea, ca i criteriu de oprire al algoritmului.

Fie $\tilde{f}$ a doua aproximaie a lui Taylor pentru f. Calculm urmtoarea cantitate:

$$f(x) - inf_y \tilde{f}(y) = f(x) - \tilde{f}(x + \triangle x_{nt}) = \frac{1}{2}\lambda^2(x)$$

Prin urmare, $\lambda^2/2$ este un estimator pentru $f(x) - p^*$, estimator bazat pe aproximaia cuadratic a lui f n x.

• **Cutare backtracking**

Decrementul lui Newton apare i n cutarea de tip backtracking a msurii pasului lui Newton, ntruct

$$\nabla f(x)^T \triangle x_{nt} = -\lambda(x)^2.$$

Aceasta este constanta folosit n cutarea de tip backtracking i ea poate fi interpretat ca derivata dup direcie a lui f n x dup direcia pasului lui Newton:

$$-\lambda(x)^2 = \nabla f(x)^T \triangle x_{nt} = \frac{\partial}{\partial t} f(x + \triangle x_{nt} t)|_{t=0}.$$

OBSERVAŢIA 5. *Decrementul lui Newton este, la fel ca pasul lui Newton, afin invariant.*

7. ALGORITMUL METODEI LUI NEWTON

Algoritmul 3. *Algoritmul metodei lui Newton*
Se d un punct de start, $x \in \mathbf{dom} f$ *i o toleran,* $\epsilon > 0$.
Repet
1. Calculeaz pasul i decrementul lui Newton:
$\triangle x_{nt} := -\nabla^2 f(x)^{-1} \nabla f(x)$; $\lambda^2 = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$.
2. Criteriu de oprire. Oprete-te dac $\lambda^2/2 \leq \epsilon$.
3. Msura pasului lui Newton. Se alege msura pasului, t, *prin metoda backtracking.*
4. $x := x + t\triangle x_{nt}$.

Metoda descris mai sus este o metod general descendent n care am utilizat pasul lui Newton ca i direcie de cutare. Singura diferen este c se verific criteriul de oprire imediat dup calcularea decrementului lui Newton, iar nu dup updatarea lui x.

Observaţia 6. *Metoda lui Newton de mai sus este numit i metoda amortizat a lui Newton sau metoda pzit, pentru a fi deosebit de metoda pur a lui Newton, care folosete mereu un pas unitar, adic* $t = 1$.

8. ANALIZA CONVERGENEI

• **Presupuneri**

Presupunem, ca mai devreme, c f este de dou ori difereniabil i tare convex, cu constanta $m > 0$, ceea ce nseamn c $mI \preceq \nabla^2 f(x)$ pentru orice x din S. De asemenea, am vzut c tare convexitatea implic i existena unei constante $M > 0$ astfel nct $\nabla^2 f(x) \preceq MI$ pentru orice $x \in S$.

Pe lng aceste condiii, adugm i faptul ca matricea Hessian a lui f s fie Lipschitz-continu pe S, cu constanta L:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2, \forall x, y \in S. \tag{7}$$

La modul general, coeficientul L ne indic ct de bine poate fi aproximat f de o funcie cuadratic. Astfel, ne putem atepta ca pentru o funcie a crei aproximaie cuadratic variaz puin, metoda lui Newton va funciona bine.

• **Ideea demonstraiei convergenei**

Vom demonstra c exist dou constante γ i η corespunztoare, fiecare dintre ele, uneia din cele dou faze ale desfurrii algoritmului. Cele dou constante ndeplinesc urmtoarele constrngeri: $0 < \eta < m^2/L$ i $\gamma > 0$.

n ceea ce privete fazele algoritmului, prima dintre ele mai este numit i faza *amortizat* a metodei lui Newton, deoarece poate alege i pai care s nu fie ntregi(msura pasului, t, s nu fie 1), iar cea de-a doua faz este faza *pur* a metodei lui Newton, cnd sunt luai numai pai ntregi($t = 1$). Vom vedea c faza a doua mai este numit i *faza de convergen cuadratic*, iar odat ajuni n aceast faz, convergena este foarte rapid.

Cele dou faze sunt descrise matematic dup cum urmeaz:

1. **Faza amortizat.** Dac $\|\nabla f(x^{(k)})\|_2 \geq \eta$, atunci

$$f(x^{(k+1)}) - f(x^{(k)}) \leq -\gamma. \tag{8}$$

2. **Faza pur.** Dac $\|\nabla f(x^{(k)})\|_2 < \eta$, atunci cutarea de tip backtracking ia msura pasului $t^{(k)} = 1$ i

$$\frac{L}{2m^2}\|\nabla f(x^{(k+1)})\|_2 \leq (\frac{L}{2m^2}\|\nabla f(x^{(k)})\|_2)^2. \tag{9}$$

Observaţia 7. *Analiznd implicaiile pentru a doua faz, observm c dac la iteraia k avem $|\ \nabla f(x^{(k)})\|_2 < \eta$, atunci, din moment ce $\eta \leq m^2/L$, avem c*

$$|\ \nabla f(x^{(k+1)})\|_2 < \tfrac{L}{2m^2}\eta^2 < \eta/2 < \eta.$$

Prin urmare, condiia pentru a doua faz este satisfcut i pentru iteraia $k+1$. Asta nseamn c odat ce are loc a doua condiie, (9), ea va avea loc pentru toate iteraiile viitoare, adic pentru orice $l \geq k$ msura pasului luat va fi $t = 1$ i

$$\frac{L}{2m^2}\|\nabla f(x^{(l+1)})\|_2 \leq (\frac{L}{2m^2}\|\nabla f(x^{(l)})\|_2)^2. \tag{10}$$

Aplicnd recursiv relaia (10), obinem c pentru orice $l \geq k$

$$\tfrac{L}{2m^2}\|\nabla f(x^{(l)})\|_2 \leq (\tfrac{L}{2m^2}\|\nabla f(x^{(k)})\|_2)^{2^{l-k}} \leq (\tfrac{1}{2})^{2^{l-k}},$$

acest rezultat implicnd faptul c

$$f(x^{(l)}) - p^* \leq \tfrac{1}{2m}\|\nabla f(x^{(l)})\|_2^2 \leq \tfrac{2m^3}{L^2}(\tfrac{1}{2})^{2^{l-k+1}}.$$

Ultima relaie ne spune c odat satisfcut a doua condiie, convergena este extrem de rapid. Mai mult, putem spune c dup un numr suficient de mare de iteraii, numrul de cifre corecte se dubleaz la fiecare iteraie. Acest fenomen poart numele de *convergen cuadratic.*

n continuare, vom estima convergena total a metodei, lund n considerare cele dou etape menionate mai sus.

n *faza amortizat*, observm c f descrete la fiecare iteraie cu cel puin γ. Aadar, numrul maxim de iteraii pentru aceast etap va fi maxim

$$\tfrac{f(x^{(0)})-p^*}{\gamma}.$$

Cu privire la etapa de convergen cuadratic, n *faza pur*, avem c

$$\tfrac{2m^3}{L^2}(\tfrac{1}{2})^{2^n} < \epsilon \Rightarrow (\tfrac{1}{2})^{2^n} < \epsilon/\epsilon_0,$$

unde cu ϵ_0 am notat pe $\frac{2m^3}{L^2}$ i cu n numrul de iteraii

$$\Rightarrow 2^n = log_{\frac{1}{2}}(\epsilon/\epsilon_0) = \tfrac{log_2(\epsilon/\epsilon_0)}{log_2\frac{1}{2}} = log_2(\epsilon_0/\epsilon).$$

Avem aadar c numrul de iteraii pentru a doua faz nu poate depi $log_2 log_2(\epsilon_0/\epsilon)$.
Prin urmare, numrul maxim de iteraii pn cnd $f(x) - p^* \leq \epsilon$ este mrginit superior de

$$\frac{f(x^{(0)})-p^*}{\gamma} + log_2 log_2(\epsilon_0/\epsilon).$$

Observaţia 8. *Pentru cel de-al doilea termen, $log_2 log_2(\epsilon_0/\epsilon)$, corespunztor etapei de convergen cuadratic, acesta crete extrem de ncet pentru un ϵ dat. Astfel, acesta poate fi considerat o constant din motive practice. Se observ c pentru o constant egal cu 6 (6 iteraii n faza de convergen cuadratic) se obine o precizie foarte bun, $\epsilon \approx 5 * 10^{-20}\epsilon_0$. Prin urmare, o margine superior pentru numrul de iteraii n metoda lui Newton pentru a obine o aproximaie foarte bun a soluiei este*

$$\frac{f(x^{(0)})-p^*}{\gamma} + 6.$$

Din analiza celor doua faze ale metodei lui Newton reies urmtoarele valori pentru γ i η (n funcie de constantele m, M i L):

$$\gamma = \alpha\beta\eta^2 \frac{m}{M^2}, \quad \eta = min\{1, 3(1-2\alpha)\}\frac{m^2}{L}.$$

Observaţia 9. *Putem spune, deci, c pentru o aproximare foarte bun a soluiei, numrul maxim de iteraii necesare este mrginit superior de*

$$6 + \frac{1}{\alpha\beta\eta^2\frac{m}{M^2}}(f(x^{(0)}) - p^*) =$$

$$6 + \frac{M^2L^2/m^5}{\alpha\beta\ min\{1, 9(1-2\alpha)^2\}}(f(x^{(0)}) - p^*).$$

Ideile de baz privind optimizarea convex, metodele descendente, algoritmul, pasul, decrementul lui Newton, precum i analiza convergenei metodei au fost preluate din lucrarea lui Boyd i Vandenberghe [2], unde se gsete i demonstraia complet privind convergena metodei.

9. AVANTAJE I DEZAVANTAJE

Teoria i practica au relevat cteva avantaje nsemnate pentru metoda lui Newton:

- Este afin invariant, dup cum pasul i decrementul lui Newton sunt.
- Convergena n metoda lui Newton este rapid n general i cuadratic n apropierea lui x^*. Odat ce s-a trecut n etapa de convergen cuadratic, e nevoie de aproximativ 6 iteraii pentru a obine o soluie foarte precis.
- Metoda lui Newton se scaleaz bine n funcie de dimensiunea problemei. Performana ei pe probleme n $\mathbb{R}^{10000}$ este asemntoare cu cea n $\mathbb{R}^{10}$. Ele difer printr-o cretere mic a numrului de pai cerui.
- Performana acestei metode nu depinde de parametrii algoritmului (pentru diferite alegeri ale parametrilor α i β din cadrul cutrii backtracking numrul de iteraii pentru a obine o soluie foarte bun variaz nesemnificativ).

Principalul dezavantaj al metodei lui Netwton este costul formrii i memorrii matricei Hessiene, mpreun cu costul calculrii pasului lui Newton. Cu toate acestea, exist multe cazuri n care se poate exploata structura problemei i reduce substanial costul calculrii pasului lui Newton.

OBSERVAŢIA 10. *n ceea ce privete efortul computaional, amintim i o alt alternativ. Exist o familie de algoritmi numii metode cvasi-Newton, care necesit mai puin efort de calcul pentru a determina direcia de cutare, dar au aceeai convergen rapid n apropierea lui* x^*.

BIBLIOGRAFIE

[1] Valentina ALTO: *Optimization algorithms: the Newton Method*, https://medium.com/swlh/optimization-algorithms-the-newton-method-4bc6728fb3b6.
[2] Stephan BOYD, Lieven VANDENBERGHE : *Convex optimization*, Cambridge University Press, United Kingdom, 2004.
[3] B.T. POLYAK: *Newtons method and its use in optimization*, https://www.researchgate.net/publication/221989049_Newton's_method_and_its_use_in_optimization.

Faculty of Mathematics and Computer Science
"Babeş-Bolyai" University
Str. Kogălniceanu, no. 1
400084 Cluj-Napoca, Romania