5

Методи на безусловна оптимизација

5.1 Услови за оптималност

Да ја разгледаме задачата на безусловна оптимизација

$$\min_{x \in \mathbb{R}^n} f(x), \tag{5.1}$$

каде $f : \mathbb{R}^n \to \mathbb{R}$ е два пати непрекинато диференцијабилна функција.

Дефиниција 5.1. За точката $x^* \in \mathbb{R}^n$ велиме дека е

(i) *точка на глобален минимум* за f ако $f(x^*) \leq f(x)$ за сите $x \in \mathbb{R}^n$,

(ii) *точка на (слаб) локален минимум* за f ако постои околина $\mathcal{U}$ на x^* така да $f(x^*) \leq f(x)$ за $x \in \mathcal{U}$,

(iii) *точка на строг локален минимум* (или *точка на силен локален минимум*) за f ако постои околина $\mathcal{U}$ на x^* така да $f(x^*) \leq f(x)$ за $x \in \mathcal{U}$ при $x \neq x^*$,

(iv) *точка на изолиран локален минимум* за f во X ако постои околина $\mathcal{U}$ на x^* така да x^* е единствена точка на локален минимум во $\mathcal{U}$.

Забелешка 5.1. Секоја точка на изолиран локален минимум е точка на строг локален минимум, но обратното тврдење не секогаш е точно. На пример, функцијата

$$f(x) = x^4 \cos \frac{1}{x} + 2x^4,\ f(0) = 0$$

која е два пати непрекинато диференцијабилна, има строг локален минимум во точката $x^* = 0$ кој не е изолиран, бидејќи во околина на $x^* = 0$ постојат многу точки на строг локален минимум x_n, и тие може да се означат така да $x_n \to 0$ кога $n \to \infty$.

Теорема 5.1 (**Потребни услови за оптималност од прв ред**). *Ако x^* е точка на локален минимум на f од задачата (5.1) и f е непрекинато диференцијабилна на отворена околина на x^*, тогаш $\nabla f(x^*) = 0$.*

Доказ. (доказ во Nocedal, Wright, Теорема 2.2.) ■

Точката x^* која го задоволува условот $\nabla f(x^*) = 0$ се нарекува *стационарна точка.* Според Теорема 5.1, секоја точка на локален минимум е стационарна точка.

Теорема 5.2 (**Потребни услови за оптималност од втор ред**). *Ако x^* е точка на локален минимум на f од задачата (5.1) и $\nabla^2 f$ е непрекинат на отворена околина на x^*, тогаш $\nabla f(x^*) = 0$ и $\nabla^2 f(x^*)$ е позитивно полудефинитен.*

Доказ. (доказ во Nocedal, Wright, Теорема 2.3.) ■

Теорема 5.3 (**Доволни услови за оптималност од втор ред**). *Да претпоставиме дека $\nabla^2 f$ е непрекинат на отворена околина на x^* и дека $\nabla f(x^*) = 0$ и $\nabla^2 f(x^*)$ е позитивно дефинитен. Тогаш x^* е точка на строг локален минимум на f од задачата (5.1).*

Доказ. (доказ во Nocedal, Wright, Теорема 2.4.) ■

Забелешка 5.2. Доволните услови од втор ред не се и потребни, што значи дека може точката x^* да биде точка на строг локален минимум, а да не ги задоволува доволните услови. На пример, функцијата $f(x) = x^4$ има точка на строг локален минимум $x^* = 0$ во која Хесијанот се анулира, и затоа не е позитивно дефинитен.

Теорема 5.4. *Ако f е конвексна, тогаш секоја точка на локален минимум x^* е точка на глобален минимум на f од задачата (5.1). Ако уште f е и диференцијабилна, тогаш секоја стационарна точка x^* е точка на глобален минимум на f од задачата (5.1).*

Доказ. (доказ во Nocedal, Wright, Теорема 2.5.) ■

На овие резултати се темелат алгоритмите за безусловна оптимизација. Имено, алгоритмите за безусловна оптимизација бараат точка во која градиентот на f се анулира. При практична имплементација на алгоритмите, еден од најкористените критериуми за запирање е тој да вредноста $\|\nabla f(x_k)\|$ е помала од некоја однапред дадена точност, каде x_k е k-тото приближување генерирано од алгоритмот.

5.2 Општи поставки за методите на безусловна оптимизација

Обично методите на безусловна оптимизација не гарантираат дека ќе најдат глобално решение на задачата (5.1), туку решение кое е оптимално во некоја негова околина. Да го означиме ова локално решение со x^*.

Во општ случај, методите на безусловна оптимизација се итеративни методи, што значи дека почнувајќи од некоја *почетна точка* x_0, соодветниот алгоритам генерира низа од итерации $\{x_k\}$ според итеративната формула

$$x_{k+1} = x_k + \alpha_k p_k,\ k = 0, 1, 2, ..., \tag{5.2}$$

каде $p_k \in \mathbb{R}^n$ е *правецот на пребарување* и $\alpha_k > 0$ е *должината на чекорот* по правецот p_k. Итеративното правило обично се прави со цел да се добие монотона опаѓачка низа $\{f(x_k)\}$ од вредности на функцијата на целта. Еден доволен услов за тоа е да правецот p_k биде *опаѓачки правец*, односно да важи

$$p_k^T \nabla f(x_k) < 0. \tag{5.3}$$

Забелешка 5.3. Од Тејлоровата теорема, за произволни $p \in \mathbb{R}^n$ и $\varepsilon > 0$, имаме дека

$$f(x_k + \varepsilon p) = f(x_k) + \varepsilon p^T \nabla f(x_k) + \mathcal{O}(\varepsilon^2).$$

Затоа, ако ε е доволно мал, секој p за кој $p^T \nabla f(x_k) < 0$ предизвикува намалување кај f т.е. $f(x_k + \varepsilon p) < f(x_k)$.

Постојат и неколку главни барања кои практичните алгоритми треба да ги задоволат. Најнапред, алгоритмот мора да биде добро дефиниран и треба да конвергира кон x^* од секоја почетна точка x_0, така наречено барање на *глобална конвергенција*. Од друга страна, теоретскиот аспект на глобалната конвергенција е да се покаже дека точките на натрупување на низата $\{x_k\}$ се стационарни точки. При тоа, треба да се покаже дека

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{5.4}$$

Понекогаш методот не е глобално конвергентен во смисла на (5.4), но го задоволува послебиот резултат на глобална конвергенција

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0, \tag{5.5}$$

што значи дека постои подниза од норми на градиентот $\|f(x_{k_j})\|$ која конвергира кон нула, но не мора и целата низа.

Постојат теориски резултати и во врска со *локалната конвергенција*, и нивна цел е да се покаже дека постои околина на решението и таков избор на параметри на методот кои обезбедуваат конвергенција кон решението. Резултатите за локална конвергенција исто така се однесуваат и на брзината на конвергенција која главно може да биде линеарна, суперлинеарна или квадратна.

Дефиниција 5.2. Нека $\{x_k\}$ е низа во $\mathbb{R}^n$ која конвергира кон x^*.

(i) Конвергенцијата е *линеарна* ако постои константа $r \in (0,1)$ така да

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq r \text{ за сите доволно големи } k. \tag{5.6}$$

(ii) Конвергенцијата е *суперлинеарна* ако

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0. \tag{5.7}$$

(iii) Конвергенцијата е *квадратна* ако постои константа $M > 0$ така да

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq M \text{ за сите доволно големи } k. \tag{5.8}$$

5.3 Њутнов метод

Методите за решавање на задачата (5.1) обично се дизајнирани врз база на некоја *модел функција* (апроксимирачка функција) на функцијата на целта f. Најкористен е *квадратниот модел* кој во k-тата итерација ја апроксимира вредноста $f(x)$ со Тајлоров развој во околина на x_k, односно

$$f(x_k + p) \approx m_k(p) = f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T B_k p, \tag{5.9}$$

каде p е отстапувањето од x_k, и B_k е или Хесијанот $\nabla^2 f(x_k)$ или некоја негова апроксимација.

Њутновиот метод за правец на пребарување во k-тата итерација ја зема точката на минимум на $m_k(p)$, при што $B_k = \nabla^2 f(x_k)$. Имено, p_k се наоѓа како решение на линеарниот систем

$$\nabla^2 f(x_k) p = -\nabla f(x_k), \tag{5.10}$$

и *Њутновиот правец* p_k^N се дефинира како

$$p_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k), \tag{5.11}$$

а следната итерација се пресметува според итеративната формула

$$x_{k+1} = x_k + p_k^N, \tag{5.12}$$

што значи дека должината на чекорот во секоја итерација е $\alpha_k = 1$.

Забелешка 5.4. Њутновиот метод е добро дефиниран само во случај кога $\nabla^2 f(x_k)$ е позитивно дефинитна, затоа што само тогаш постои единствена точка на минимум на $m_k(p)$. Обично $\nabla^2 f(x_k)$ е позитивно дефинитна кога x_k е во некоја околина на x^*. Па, кога почетната точка x_0 е доволно блиску до x^*, Њутновиот метод конвергира кон решението на таков ефикасен начин така да останатите методи се обидуваат да го задржат ова локално однесување колку е тоа можно.

Забелешка 5.5. За Њутновиот правец (5.11) имаме

$$\nabla f(x_k)^T p_k^N = -p_k^N \nabla^2 f(x_k) p_k^N \leq -\sigma_k \|p_k^N\|^2,$$

за некој $\sigma_k > 0$, бидејќи $\nabla^2 f(x_k)$ е позитивно дефинитна. Тогаш, ако $\nabla f(x_k) \neq 0$, имаме дека $\nabla f(x_k)^T p_k^N < 0$ и според (5.3) заклучуваме дека Њутновиот правец е опаѓачки правец.

АЛГОРИТАМ 5.1 (Њутнов метод).

Чекор 0. Избери почетна точка x_0 доволно блиску до x^* и толеранција $\varepsilon > 0$. Стави $k = 0$.

Чекор 1. Ако $\|\nabla f(x_k)\| \leq \varepsilon$, тогаш СТОП. Инаку, пресметај го Њутновиот правец според $p_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$.

Чекор 2. Земи ја за следна итерација точката $x_{k+1} = x_k + p_k^N$, стави $k = k + 1$ и оди на Чекор 1.

Забелешка 5.6. Битно е да се забележи и дека Њутновиот метод наоѓа решение на стого конвексна квадратна функција $f(x) = x^T A x + b^T x$, каде A е позитивно дефинитна матрица, во само еден чекор. Имено, нека x^* е точката на минимум, тогаш $2Ax^* + b = 0$, па за првата итерација имаме

$$x_1 = x_0 - (2A)^{-1}(2Ax_0 + b) = -(2A)^{-1} b = x^*.$$

Теорема 5.5 (**Теорема за глобална и локална конвергенција на Њутновиот метод**). *Да претпоставиме дека f е два пати непрекинато диференцијабилна и дека Хесијанот $\nabla^2 f(x)$ е Липшиц непрекинат во околина на решението x^* во кое $\nabla f(x^*) = 0$ и $\nabla^2 f(x^*)$ е позитивно дефинитна. Да ја разгледаме итерацијата $x_{k+1} = x_k + p_k^N$, каде $p_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$ е Њутнов правец. Тогаш* *(i) ако почетната точка x_0 е доволно близу до x^*, тогаш низата од итерации $\{x_k\}$ конвергира кон x^*,*

(ii) брзината на конвергенција со која $\{x_k\}$ конвергира е квадратна, и

(iii) низата од норми на градиентот $\{\|\nabla f(x_k)\|\}$ конвергира квадратно кон нула.

Доказ. (доказ во Nocedal, Wright, Теорема 3.7.) ■

5.4 Методи на линиско пребарување

Еден *метод на линиско пребарување* во секоја итерација x_k, најнапред го избира правецот на пребарување p_k и потоа ја пресметува должината на чекорот α_k како решение на еднодимензионалната задача на минимизација

$$\min_{\alpha>0} f(x_k + \alpha p_k). \tag{5.13}$$

Тогаш, следната итерација се пресметува според формулата $x_{k+1} = x_k + \alpha_k p_k$ и целиот процес се повторува.

АЛГОРИТАМ 5.2 (Метод на линиско пребарување).

Чекор 0. Избери почетна точка x_0 и толеранција $\varepsilon > 0$. Стави $k = 0$.

Чекор 1. Ако $\|\nabla f(x_k)\| \leq \varepsilon$, тогаш СТОП. Инаку, пресметај го правецот на пребарување p_k.

Чекор 2. Реши ја еднодимензионалната задача (5.13) за да ја добиеш должината на чекорот α_k.

Чекор 3. Земи ја за следна итерација точката $x_{k+1} = x_k + \alpha_k p_k$, стави $k = k + 1$ и оди на Чекор 1.

5.4.1 Избор на правецот на пребарување

Во зависност од правецот на пребарување p_k, разликуваме повеќе методи на линиско пребарување.

Методот на најбрзо спуштање е метод на линиско пребарување кој во секој чекор се придвижува по должината на правецот $p_k = -\nabla f(x_k)$.

Методите кои го користат правецот $p_k = -B_k^{-1}\nabla f(x_k)$, каде B_k е или Хесијанот $\nabla^2 f(x_k)$ или некоја негова апроксимација познати се како *методи од Њутнов тип.* Ако $B_k = I$, единичната матрица, тогаш $p_k = -\nabla f(x_k)$ и се добива *методот на најбрзо спуштање.*

Забелешка 5.7. Кај методите од Њутнов тип важно е B_k да е позитивно дефинитна матрица што повлекува правецот на пребарување p_k да е добро дефиниран и опаѓачки, бидејќи тогаш $p_k^T \nabla f(x_k) = p_k^T B_k p_k < 0$.

Квази-Њутновите методи се методи од Њутнов тип кои користат апроксимација B_k на Хесијанот $\nabla^2 f(x_k)$ со особина да апроксимацијата B_{k+1} го задоволува следното својство познато како *секантна равенка*,

$$B_{k+1} s_k = y_k, \tag{5.14}$$

каде

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Понекогаш се поставуваат дополнителни барања за B_{k+1}, како симетричност и разликата на две последователни апроксимации B_k и B_{k+1} да има низок ранк.

Друга класа на правци на пребарување кои се користат кај *методите на конјугирани градиенти* го имаат следниот облик

$$p_k = -\nabla f(x_k) + \beta_k p_{k-1}, \tag{5.15}$$

каде $p_0 = -\nabla f(x_k)$ и β_k е скалар кој обезбедува p_k и p_{k-1} да бидат конјугирани правци[1].

5.4.2 Избор на должината на чекорот

По определување на правецот на пребарување p_k, методите на линиско пребарување треба да ја решат задачата (5.13) за да ја одредат должината на чекорот α_k. Решението на оваа задача на еднодимензионална миниминизација може да биде точно или приближно.

Едноставен пример за *точно линиско пребарување* е кога имаме квадратна функција на целта $f(x) = x^T Ax + b^T x$, каде A е симетрична позитивно дефинитна матрица и $b \in \mathbb{R}^n$. Тогаш точното решение на задачата (5.13) е

$$\alpha_k = -\frac{p_k^T (2Ax_k + b)}{2p_k^T A p_k}.$$

Попрактични се *техниките за приближно линиско пребарување*, кои наоѓаат приближно решение на задачата (5.13) проверувајќи низа од вредности кандидати за α се додека не бидат задоволени одредени услови.

[1] Два правца p и q се *конјугирани* во однос на симетрично позитивно дефинитна матрица A ако $p^T Aq = 0$.

Најкористен услов за приближно линиско пребарување е оној кој бара да должината на чекорот α_k доведе до *доволно намалување* на f, кое се мери со следното неравенство

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f(x_k)^T p_k, \tag{5.16}$$

за некоја константа $c_1 \in (0, 1)$. Неравенството (5.16) е познато како *условот на Армижо* (Armijo). Во пракса, c_1 се одбира да биде доволно мало, обично $c_1 = 10^{-4}$.

Лема 5.1. *Нека функцијата $f : \mathbb{R}^n \to \mathbb{R}$ е диференцијабилна на $\mathbb{R}^n$, градиентот ∇f е Липшиц непрекинат на $\mathbb{R}^n$ со Липшицова константа $L > 0$ и нека p_k е опаѓачки правец во x_k. Тогаш, за секој $\alpha \in (0, \overline{\alpha_k}]$ каде*

$$\overline{\alpha}_k = -\frac{2(1 - c_1) \nabla f(x_k)^T p_k}{L \|p_k\|^2} > 0, \tag{5.17}$$

е задоволено неравенството (5.16).

Доказ. (доказ во Измаилов, Лема 3, стр. 63) ■

Условот за доволно намалување (5.16) е задоволен за сите доволно мали α, па затоа користењето само на условот на Армижо не е доволно за да се направи разумен прогрес на алгоритмот. Затоа, условот на Армижо се комбинира со други процедури или услови.

Една процедура која се комбинира со условот на Армижо за да се одреди должината на чекорот α_k е *процедурата на константно намалување на чекорот* (*backtracking*) која почнувајќи од некоја почетна вредност $\alpha > 0$ ја намалува со мултипликатор (фактор) $\rho \in (0, 1)$ (или со интерполација) се додека не се задоволи условот на Армижо.

АЛГОРИТАМ 5.3 (Линиско пребарување за наоѓање на должината на чекорот со процедура на константно намалување на чекорот (backtracking)).

Чекор 0. Избери броеви $c_1, \rho \in (0, 1)$ (избери броеви $c_1, \underline{\rho}, \overline{\rho} \in (0, 1)$ такви да $\underline{\rho} < \overline{\rho}$) и избери почетна вредност $\alpha > 0$.

Чекор 1. Провери дали условот на Армижо (5.16) е задоволен.

Чекор 2. Ако (5.16) не е задоволен, тогаш стави $\alpha = \rho\alpha$ (избери нов α од $[\underline{\rho}\alpha, \overline{\rho}\alpha]$ со интерполација) и оди на Чекор 2, инаку заврши ја backtracking процедурата и за должина на чекорот зами ја вредноста $\alpha_k = \alpha$.

Пософистициран услов за корекција на Армижо условот се добива со воведување на *условот за закривеност*, кој бара α_k да го задоволува неравенството

$$\nabla f(x_k + \alpha p_k)^T p_k \geq c_2 \nabla f(x_k)^T p_k, \tag{5.18}$$

за некоја константа $c_2 \in (c_1, 1)$, каде c_1 е константата од (5.16). Во пракса $c_2 = 0.9$ кај Њутновиот и квази-Њутновиот метод и $c_2 = 0.1$ кај методот на нелинеарни коњугирани градиенти. Со условот на закривеност се избегнуваат многу малите должини на чекорот. Заедно условот на Армижо (5.16) и условот на закривеност (5.18) се познати како *Волфови услови* (Wolfe).

Лема 5.2. *Да претпоставиме дека $f : \mathbb{R}^n \to \mathbb{R}$ е непрекинато диференцијабилна. Нека p_k е опаѓачки правец во x_k, и нека f е ограничена од долу по должината на зракот $\{x_k + \alpha p_k \mid \alpha > 0\}$. Тогаш, ако $0 < c_1 < c_2 < 1$, постојат интервали на должини на чекорот кои ги задоволуваат Волфовите услови (5.16), (5.18).*

Доказ. (доказ во Nocedal, Wright, Лема 3.1. стр.40) ∎

Во пракса, процедурата на линиско пребарување базирана на Волфовите услови најмногу се користи кај квази-Њутновите методи.

АЛГОРИТАМ 5.4 (Процедура на линиско пребарување за наоѓање на должината на чекорот базирана на Волфовите услови).

Чекор 0. Избери броеви $c_1, c_2 \in (0, 1)$ такви да $c_1 < c_2$. Стави $\underline{\alpha} = \overline{\alpha} = 0$ и избери почетна вредност $\alpha > 0$.

Чекор 1. Провери дали Волфовите услови (5.16), (5.18) се задоволени. Ако и двата услова се задоволени, тогаш оди на Чекор 6.

Чекор 2. Ако условот на Армижо (5.16) не е задоволен, тогаш стави $\overline{\alpha} = \alpha$ и оди на Чекор 5.

Чекор 3. Ако условот на закривеност (5.18) не е задоволен, тогаш стави $\underline{\alpha} = \alpha$ и оди на Чекор 4.

Чекор 4. Ако $\overline{\alpha} = 0$, тогаш избери нова вредност $\alpha > \underline{\alpha}$ (со екстраполација) и оди на Чекор 1.

Чекор 5. Избери нова вредност $\alpha \in (\underline{\alpha}, \overline{\alpha})$ (со интерполација) и оди на Чекор 1.

Чекор 6. Заврши ја процедурата и за должина на чекорот земи ја вредноста $\alpha_k = \alpha$.

5.4.3 Глобална конвергенција на методите на линиско пребарување

Влијанието на правецот на пребарување p_k на резултатите за глобална конвергенција на методите на линиско пребарување се разгледува преку аголот θ_k помеѓу p_k и правецот на најбрзо спуштање $-\nabla f(x_k)$, дефиниран со

$$\cos\theta_k = \frac{-\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \|p_k\|}. \tag{5.19}$$

Теорема 5.6. *Да ја разгледаме итерацијата од обликот (5.2), каде p_k е опаѓачки правец и α_k ги задоволува Волфовите услови (5.16), (5.18). Да претпоставиме дека f е ограничена од долу на $\mathbb{R}^n$ и дека f е непрекинато диференцијабилна на отворено множество $\mathcal{U}$ кое го содржи множеството $\Omega \overset{def}{=} \{x : f(x) \leq f(x_0)\}$, каде x_0 е почетната итерација. Претпоставуваме уште дека градиентот ∇f е Липшиц непрекинат на $\mathcal{U}$, односно, постои константа $L > 0$ така да*

$$\|\nabla_f(x) - \nabla f(\tilde{x})\| \leq L\|x - \tilde{x}\|, \quad \text{за сите } x, \tilde{x} \in \mathcal{U}. \tag{5.20}$$

Тогаш, важи условот на Зутенџик (Zoutendijk), односно

$$\sum_{k\geq 0} \cos^2\theta_k \|\nabla f(x_k)\|^2 < \infty. \tag{5.21}$$

Доказ. Од (5.18) и (5.2) имаме дека

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T p_k \geq (c_2 - 1)\nabla f(x_k)^T p_k,$$

додека пак Липшицовиот услов (5.20) повлекува дека

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T p_k \leq \alpha_k L \|p_k\|^2.$$

Со комбинирање на овие две релации, добиваме

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{\nabla f(x_k)^T p_k}{\|p_k\|^2}.$$

Со замена на ова неравенство во првиот Волфов услов (5.16), добиваме

$$f(x_{k+1}) \leq f(x_k) - c_1 \frac{1 - c_2}{L} \frac{(\nabla f(x_k)^T p_k)^2}{\|p_k\|^2}.$$

Од дефиницијата (5.19), може да ја запишеме оваа релација како

$$f(x_{k+1}) \leq f(x_k) - c\cos^2\theta_k \|\nabla f(x_k)\|^2,$$

каде $c = c_1(1-c_2)/L$. Со сумирање на овој израз по сите индекси помали или еднакви на k, добиваме

$$f(x_{k+1}) \leq f(x_0) - c\sum_{j=0}^{k}\cos^2\theta_j\|\nabla f(x_j)\|^2. \tag{5.22}$$

Бидејќи f е ограничена од долу, имаме дека $f(x_0)-f(x_{k+1})$ е помало од некоја позитивна константа, за сите k. Затоа со барање на граница во (5.22), добиваме

$$\sum_{k=0}^{\infty}\cos^2\theta_k\|\nabla f(x_k)\|^2 < \infty,$$

со што е завршен доказот. ∎

Забелешка 5.8. Условот на Зутенџик (Zoutendijk) (5.21) повлекува дека

$$\cos^2\theta_k\|\nabla f(x_k)\|^2 \to 0. \tag{5.23}$$

(i) Ако аголот θ_k е ограничен далеку од $\pi/2$, што значи, постои позитивна константа $\delta > 0$ така да $\cos\theta_k \geq \delta > 0$, за сите k, тогаш од (5.23) следи дека $\|\nabla f(x_k)\| \to 0$, односно методот на линиско пребарување е глобално конвергентен во смисла на (5.4).

(ii) За да се добие послабиот резултат за глобална конвергенција (5.5) доволно е да се покаже дека постои подниза $\{\cos\theta_{k_j}\}$ која е ограничена далеку од нула. Имено, ако претпоставиме дека (5.5) не важи, значи постои $\gamma > 0$ така да $\|\nabla f(x_k)\| \geq \gamma$ за сите доволно големи k. Со користење на последното неравенство и (5.23) заклучуваме дека $\cos\theta_k \to 0$, што е во контрадикција со постоењето на подниза $\{\cos\theta_{k_j}\}$ која е ограничена далеку од нула. И бидејќи за методот на најбрзо спуштање важи $\cos\theta_k = 1$, еден начин да се добие глобално конвергентен метод на линиско пребарување во смисла на (5.5), е да се одбере секоја m-та итерација да биде итерација на методот на најбрзо спуштање, каде p_k ќе биде опаѓачки правец, а должината на чекорот α_k се избира така да ги задоволува Волфовите услови.

5.5 Метод на најбрзо спуштање

Кога правецот на пребарување кај методите на линиско преебарување е антиградиентот, односно $p_k = -\nabla f(x_k)$, тогаш станува збор за *методот на најбрзо спуштање* или познат како *Кошиев метод*. Правецот на најбрзо спуштање $-\nabla f(x_k)$ е ортогонален на контурите на функцијата на целта f.

Забелешка 5.9. Според Тајлоровиот развој во Забелешка 5.3, функцијата на целта f ќе има најбрзо опаѓање ако единичниот правец p го минимизира изразот $p^T\nabla f(x_k)$. Бидејќи $p^T\nabla f(x_k) = \|p\|\|\nabla f(x_k)\|\cos\theta = \|\nabla f(x_k)\|\cos\theta$, каде θ е аголот меѓу p и $\nabla f(x_k)$, значи дека за да се достигне минимум треба $\cos\theta = -1$, односно $p = -\nabla f(x_k)/\|\nabla f(x_k)\|$.

АЛГОРИТАМ 5.5 (Метод на најбрзо спуштање).

Чекор 0. Избери почетна точка x_0 и толеранција $\varepsilon > 0$. Стави $k = 0$.

Чекор 1. Ако $\|\nabla f(x_k)\| \leq \varepsilon$, тогаш СТОП. Инаку, пресметај го правецот на пребарување $p_k = -\nabla f(x_k)$.

Чекор 2. Реши ја еднодимензионалната задача (5.13) за да ја добиеш должината на чекорот α_k.

Чекор 3. Земи ја за следна итерација точката $x_{k+1} = x_k + \alpha_k p_k$, стави $k = k + 1$ и оди на Чекор 1.

Забелешка 5.10. Според Забелешка 5.8, ако методот на најбрзо спуштање Алгоритам 5.5 при наоѓање на должината на чекорот α_k користи процедура на линиско пребарување базирана на Волфовите услови, тогаш $\cos\theta_k = 1 > 0$, односно аголот θ_k е ограничен далеку од $\pi/2$, па методот на најбрзо спуштање е глобално конвергентен во смисла на (5.4), т.е. $\|\nabla f(x_k)\| \to 0$, кога $k \to \infty$.

Теорема 5.7 (Теорема за глобална конвергенција на методот на најбрзо спуштање). *Нека f е непрекинато диференцијабилна функција. Тогаш, секоја точка на натрупување на низата $\{x_k\}$ генерирана од алгоритамот на методот на најбрзо спуштање Алгоритам 5.5 со точно линиско пребарување е стационарна точка.*

Доказ. (доказ во Sun, Yuan, Теорема 3.1.2. и во Freund) ■

Најчесто анализата на конвергенцијата (на локално ниво) кај методот на најбрзо спуштање се илуструра со случај на квадратна финкција на целта и точно линиско пребарување. Од претходно видовме дека, ако функцијата на целта е од обликот $f(x) = x^TAx + b^Tx$, каде A е симетрична позитивно дефинитна матрица и $b \in \mathbb{R}^n$, тогаш точното решение на еднодимензионалната задача на минимизација (5.13) е

$$\alpha_k = -\frac{p_k^T(2Ax_k + b)}{2p_k^TAp_k}.$$

Бидејќи $p_k = -\nabla f(x_k) = -2Ax_k + b$, за должината на чекорот имаме

$$\alpha_k = \frac{p_k^T p_k}{2p_k^T A p_k} = \frac{\nabla f(x_k)^T \nabla f(x_k)}{2\nabla f(x_k)^T A \nabla f(x_k)},$$

и секоја наредна итерација се пресметува по формулата

$$x_{k+1} = x_k - \Big(\frac{\nabla f(x_k)^T \nabla f(x_k)}{2\nabla f(x_k)^T A \nabla f(x_k)}\Big)\, \nabla f(x_k).$$

Од друга страна, единствената точка x^* на глобален минимиум на квадратната функцијата f е решението на равенката $2Ax + b = 0$.

Теорема 5.8. *Нека функцијата на целта од задачата (5.1) е квадратна т.е. $f(x) = x^T Ax + b^T x$, каде A е симетрична позитивно дефинитна матрица и $b \in \mathbb{R}^n$ и нека x^* е нејзиното оптимално решение. Тогаш, низата $\{x_k\}$ генерирана од алгоритамот на методот на најбрзо спуштање Алгоритам 5.5 со точно линиско пребарување конвергира кон x^* и при тоа брзината на конвергенција е линеарно, односно важи следната граница*

$$\frac{\|x_{k+1} - x^*\|_A^2}{\|x_k - x^*\|_A^2} \le \Big(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\Big)^2, \tag{5.24}$$

каде $0 < \lambda_1 \le \cdots \le \lambda_n$ се сопствените вредности на матрицата A.

Доказ. (доказ во Sun, Yuan, Теорема 3.1.5. и во Freund) ∎

Забелешка 5.11. Од дефинираноста на нормата $\|\cdot\|_A$ како $\|x\|_A^2 = x^T Ax$ за квадратната функција $f(x) = x^T Ax + b^T x$ имаме дека $f(x) - f(x^*) = \|x - x^*\|_A^2$. Затоа како последица од (5.24) заклучуваме дека $f(x_k)$ конвергира линеарно кон минимумот $f(x^*)$. Специјално кога сопствените вредности на A се сите еднакви, конвергенцијата се постигнува е во една итерација.

И кога функцијата на целта има општ облик, исто така може да се постигне линеарна конвергенција на методот на најбрзо спуштање.

Теорема 5.9 (Теорема за локална конвергенција на методот на најбрзо спуштање). *Нека функцијата на целта $f : \mathbb{R}^n \to \mathbb{R}$ од задачата (5.1) е два пати непрекинато диференцијабилна, и нека низата $\{x_k\}$ генерирана од алгоритамот на методот на најбрзо спуштање Алгоритам 5.5 со точно линиско пребарување конвергира кон x^* каде $\nabla^2 f(x^*)$ е позитивно дефинитна. Тогаш, конвергенцијата е линеарна и важи*

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \le \Big(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\Big)^2, \tag{5.25}$$

каде $0 < \lambda_1 \le \cdots \le \lambda_n$ се сопствените вредности на $\nabla^2 f(x^*)$.

Доказ. (доказ во Sun, Yuan, Теорема 3.1.7.) ∎

5.6 Њутнов метод со линиско пребарување

Њутновиот метод е локален метод, што значи дека кога почетната точка е далеку од решението не се знае дали $\nabla^2 f(x_k)$ е позитивно дефинитна, и следствено дали p_k^N е опаѓачки правец, па конвергенцијата не е загарантирана. Од друга страна, методите на линиско пребарување може да обезбедат глобална конвергенција (Забелешка 5.8), па затоа добро е да се комбинира Њутновиот метод со линиското пребарување.

АЛГОРИТАМ 5.6 (Њутнов метод со линиско пребарување).

Чекор 0. Избери почетна точка x_0 и толеранција $\varepsilon > 0$. Стави $k = 0$.

Чекор 1. Ако $\|\nabla f(x_k)\| \leq \varepsilon$, тогаш СТОП. Инаку, пресметај го Њутновиот правец на пребарување $p_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$.

Чекор 2. Реши ја еднодимензионалната задача (5.13) за да ја добиеш должината на чекорот α_k.

Чекор 3. Земи ја за следна итерација точката $x_{k+1} = x_k + \alpha_k p_k^N$, стави $k = k + 1$ и оди на Чекор 1.

Овде ги изложуваме теоремите за глобална и локална конвергенција на Њутновиот метод со точно линиско пребарување.

Теорема 5.10 (Теорема за глобална конвергенција на Њутновиот метод со линиско пребарување). *Да претпоставиме дека f е два пати непрекинато диференцијабилна и нека за секоја почетна точка x_0, постои константа $m > 0$ така да*

$$u^T \nabla^2 f(x) u \geq m\|u\|^2, \text{ за сите } u \in \mathbb{R}^n, x \in \Omega(x_0), \tag{5.26}$$

каде $\Omega(x_0) = \{x \mid f(x) \leq f(x_0)\}$. Нека $\{x_k\}$ е низа генерирана со Алгоритам 5.6. Тогаш,

(i) ако $\{x_k\}$ е конечна, тогаш последната точка е оптимално решение на задачата (5.1),

(ii) ако $\{x_k\}$ е бесконечна, тогаш таа конвергира кон единственото решение на задачата (5.1).

Доказ. (доказ во Sun, Yuan, Теорема 3.2.4, и кај Vujcic, Asic, Milicic, Теорема 5.4.1.) ∎

Теорема 5.11 (Теорема за локална конвергенција на Њутновиот метод со линиско пребарување). *Нека f е два пати непрекинато диференцијабилна функција. Нека $\{x_k\}$ е низа генерирана со Алгоритам 5.6 која конвергира кон x^* така да $\nabla f(x^*) = 0$ и $\nabla^2 f(x^*)$ е позитивно дефинитна. Тогаш,*

(i) низата $\{x_k\}$ суперлинеарно конвергира,

(ii) ако уште и Хесијанот $\nabla^2 f(x)$ е Липшиц непрекинат во околина на решението x^*, тогаш низата $\{x_k\}$ квадратно конвергира.

Доказ. (доказ во Vujcic, Asic, Milicic, Теорема 5.4.2.) ∎

5.7 Методи на конјугирани градиенти

Главна карактеристика на методите на конјугирани градиенти е таа дека правците кои тие ги генерираат се конјугирани. Тие го пресметуваат новиот правец p_k користејќи го само претходниот правец p_{k-1}, односно за $\beta_k > 0$ правецот p_k се пресметува според

$$p_k = -\nabla f(x_k) + \beta_k p_{k-1}.$$

Најнапред го воведуваме поимот за коњугирани правци и го излагаме методот на коњугирани правци.

Дефиниција 5.3. Нека A е симетрична позитивно дефинитна матрица од ред n. Ненултите вектори $p_0, p_1, ..., p_{m-1} \in \mathbb{R}^n$, $m \leq n$ се *конјугирани* (или A-*конјугирани*), ако

$$p_i^T A p_j = 0, \text{ за } i \neq j.$$

Ако $A = I$ тогаш конјугираните вектори се нарекуваат *ортогонални.*

Забелешка 5.12. Конјугираните вектори $p_0, p_1, ..., p_{m-1}$ се линеарно независни.

АЛГОРИТАМ 5.7 (Метод на коњугирани правци).

Чекор 0. Избери почетна точка x_0 и толеранција $\varepsilon > 0$. Пресметај го $\nabla f(x_0)$ и ако $\nabla f(x_0) \neq 0$ тогаш избери почетен опаѓачки правец p_0 така да $p_0^T \nabla f(x_0) < 0$, во спротивно СТОП. Стави $k = 0$.

Чекор 1. Ако $\|\nabla f(x_k)\| \leq \varepsilon$, тогаш СТОП.

Чекор 2. Реши ја еднодимензионалната задача (5.13) за да ја добиеш должината на чекорот α_k.

Чекор 3. Земи ја за следна итерација точката $x_{k+1} = x_k + \alpha_k p_k$.

Чекор 4. Пресметај го правецот p_{k+1} така да е конјугиран со сите дотогашни правци p_j, $j = 0, 1, ..., k$.

Чекор 5. Стави $k = k + 1$ и оди на Чекор 1.

Теорема 5.12. *Нека функцијата на целта во задачата (5.1) е строго конвексна квадратна функција т.е. $f(x) = x^T Ax + b^T x$, каде A е симетрична позитивно дефинитна матрица. Тогаш, за секоја почетна точка x_0 низата од итерации $\{x_k\}$ генерирана со методот на конјугирани правци Алгоритам 5.7, конвергира кон точката на минимум во најмногу n чекори.*

Доказ. (доказ во Vujcic, Asic, Milicic, Теорема 5.5.1, во Sun, Yuan, Теорема 4.1.3, и кај Nocedal, Wright, Теорема 5.1) ∎

Сите методи на конјугирани правци се ослонуваат на резултатот од последната теорема. Имено, коњугираноста заедно со точното линиско пербарување доведува до квадратна терминација на методот.

Методите на конјугирани градиенти се состојат од постапка за пресметување на конјугираните правци од обликот $p_k = -\nabla f(x_k) + \beta_k p_{k-1}$, каде $\beta_k > 0$. Постојат повеќе формули за пресметување на β_k, овде излагаме некои од нив.

- Fletcher - Reeves (FR) формула

$$\beta_k = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_{k-1})^T \nabla f(x_{k-1})} \tag{5.27}$$

- Polak - Ribière - Polyak (PRP) формула

$$\beta_k = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\nabla f(x_{k-1})^T \nabla f(x_{k-1})} \tag{5.28}$$

- Hestenes - Stiefel (HS) (или Crowder - Wolfe (CW)) формула

$$\beta_k = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{p_{k-1}^T (\nabla f(x_k) - \nabla f(x_{k-1}))} \tag{5.29}$$

- Dixon формула

$$\beta_k = -\frac{\nabla f(x_k)^T \nabla f(x_k)}{p_{k-1}^T \nabla f(x_{k-1})} \tag{5.30}$$

- Dai - Yuan (DY) формула

$$\beta_k = \frac{\nabla f(x_k)^T \nabla f(x_k)}{p_{k-1}^T(\nabla f(x_k) - \nabla f(x_{k-1}))} \tag{5.31}$$

- Daniel формула

$$\beta_k = \frac{p_{k-1}^T \nabla^2 f(x_{k-1}) \nabla f(x_k)}{p_{k-1}^T \nabla^2 f(x_{k-1}) p_{k-1}} \tag{5.32}$$

Сите овие формули се еквивалентни при минимизирање на строго конвексна квадратна финкција, односно генерираат еднакви низи од правци на пребарување. Додека, во поопшт случај на функција на целта и при приближно линиско пребарување, нивното однесување значително се разликува.

АЛГОРИТАМ 5.8 (Метод на коњугирани градиенти).

Чекор 0. Избери почетна точка x_0 и толеранција $\varepsilon > 0$.

Чекор 1. Ако $\|\nabla f(x_0)\| \leq \varepsilon$, тогаш СТОП. Инаку, стави $p_0 = -\nabla f(x_0)$ и $k = 0$.

Чекор 2. Реши ја еднодимензионалната задача (5.13) за да ја добиеш должината на чекорот α_k.

Чекор 3. Земи ја за следна итерација точката $x_{k+1} = x_k + \alpha_k p_k$.

Чекор 4. Ако $\|\nabla f(x_{k+1})\| \leq \varepsilon$, тогаш СТОП.

Чекор 5. Ако $k + 1 \in \{n, 2n, 3n, ...\}$, тогаш стави $p_{k+1} = -\nabla f(x_{k+1})$, инаку стави $p_{k+1} = -\nabla f(x_{k+1}) + \beta_{k+1} p_k$, каде β_{k+1} се пресметува според некоја од формулите (5.27)-(5.32).

Чекор 6. Стави $k = k + 1$ и оди на Чекор 2.

Забелешка 5.13. Главна карактеристика на изложениот алгоритам на коњугирани градиенти Алгоритам 5.8, е таа да после n итерации процедурата на барање на правци на пребарување се рестартира при што повторно како на почетокот од алгоритмот правецот на најбрзо спуштање се зема за правец на пребарување. На тој начин алгоритмот периодично се освежува.

Формулата *PRP* има особина да доведе до автоматско рестартирање, во смисла дека кога алгоритмот напредува споро и $\nabla f(x_{k+1}) \approx \nabla f(x_k)$, тогаш според *PRP* формулата $\beta_{k+1} \approx 0$ што повлекува да $p_{k+1} \approx -\nabla f(x_{k+1})$.

Повеќето од формулите за β_k не гарантираат дека p_k е опаѓачки правец. За да се обезбеди секој правец да биде опаѓачки обично се применува рестартирање секој пат кога не е задоволен условот за опаѓачки правец, односно ако $p_{k+1}^T \nabla f(x_{k+1}) > 0$, тогаш се зема $p_{k+1} = -\nabla f(x_{k+1})$.

Теорема 5.13. *Нека функцијата на целта во задачата (5.1) е строго конвексна квадратна функција т.е. $f(x) = x^T Ax + b^T x$, каде A е симетрична позитивно дефинитна матрица. Тогаш, за секоја почетна точка x_0 низата од итерации $\{x_k\}$ генерирана со PRP методот на конјугирани градиенти конвергира кон точката на минимум во најмногу n чекори и при тоа генерираните правци се A-конјугирани.*

Доказ. (доказ во Vujcic, Asic, Milicic, Теорема 5.5.2 за PRP методот, забелешка: во Sun, Yuan, Теорема 4.2.1 го дава соодветното тврдење за FR методот) ∎

Теорема 5.14 (Теорема за глобална конвергенција на PRP методот). *Нека f е два пати непрекинато диференцијабилна функција и нека за секоја почетна точка x_0, постои константа $m > 0$ така да*

$$u^T \nabla^2 f(x) u \geq m\|u\|^2, \text{ за сите } u \in \mathbb{R}^n, x \in \Omega(x_0), \tag{5.33}$$

каде $\Omega(x_0) = \{x \mid f(x) \leq f(x_0)\}$ е конвексно и ограничено. Нека $\{x_k\}$ е низа генерирана од PRP методот со точно линиско пребарување, тогаш таа конвергира кон единствената точка на минимим x^* на f.

Доказ. (доказ во Sun, Yuan, Теорема 4.3.3, кај Vujcic, Asic, Milicic, Теоремите 5.5.3, 5.5.4, 5.5.5, забелешка: во Sun, Yuan, Теорема 4.3.1 (FR)-точно, Теорема 4.3.5 (FR)-приближно, кај Nocedal, Wright, Теорема 5.8 (FR)-приближно) ∎

Забелешка 5.14. Доколку линиското пребарување е приближно и базирано на Волфовите услови, при дискусија на глобалната конвергенција може да го искористиме резултатот на Зутенџик, Теорема 5.6. Имено, кај алгоритмите со рестарт, ако ги означиме со $k_1, k_2, ...$ итерациите кога се применува рестартирањето, односно кога за правец на пребарување се зема $p_k = -\nabla f(x_k)$, $k = k_1, k_2, ...$, од условот на Зутенџик (5.21), следи дека

$$\sum_{k=k_1,k_2,...} \|\nabla f(x_k)\|^2 < \infty.$$

Ако дозволиме по најмногу $\overline{n}$ итерации помеѓу две рестартирања, тогаш низата $\{k_j\}$ ќе биде бесконечна и ќе следи дека $\lim_{j\to\infty} \|\nabla f(x_{k_j})\| = 0$, од каде се добива дека важи послабиот резултат за глобална конвергенција

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0.$$

Забелешка 5.15. Поголемиот дел од теориските резултати за брзината на конвергенција на методите на конјугирани градиенти птетпоставуваат дека линиското пребарување е точно. Имено, за методот на конјугирани градиенти со рестартирање може да се покаже дека под одредени услови n-чекорно квадратно конвергира т.е. $\|x_{k+n} - x^*\| = \mathcal{O}(\|x_k - x^*\|^2)$ или дека има n-чекорно суперквадратна конвергенција т.е $\|x_{k+n} - x^*\| = o(\|x_k - x^*\|^2)$.

(За две бесконечни низи од ненегативни реални броеви $\{x_k\}$ и $\{y_k\}$, означуваме со $x_k = \mathcal{O}(y_k)$, ако постои позитивна константа $C > 0$ така да $|x_k| \le C|y_k|$ за сите доволно големи k и означуваме со $x_k = o(y_k)$, ако $\lim_{k\to\infty} \frac{x_k}{y_k} = 0$.)

5.8 Квази-Њутнови методи

Квази-Њутновите методи користат правец на пребарување од облик $p_k = -B_k^{-1}\nabla f(x_k)$, каде B_k е апроксимација на Хесијанот $\nabla^2 f(x_k)$, која во секоја итерација се обновува со матрицата B_{k+1} која ја задоволува *секантната равенка* или *квази-Њутнова равенка* (5.35).

Имено, Тејлоровата теорема тврди дека

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p\, dt,$$

за секој $p \in \mathbb{R}^n$ и некој $t \in (0,1)$, од каде имаме

$$\nabla f(x+p) = \nabla f(x) + \nabla^2 f(x+p)p + \int_0^1 [\nabla^2 f(x+tp) - \nabla^2 f(x+p)]p\, dt,$$

за секој $p \in \mathbb{R}^n$ и некој $t \in (0,1)$. Бидејќи $\nabla^2 f$ е непрекинат, големината на последниот интеграл е $o(\|p\|)$. Ако ставиме $x = x_k$ и $p = x_{k+1} - x_k$, добиваме

$$\nabla f(x_{k+1}) = \nabla f(x_k) + \nabla^2 f(x_{k+1})(x_{k+1} - x_k) + o(\|x_{k+1} - x_k\|).$$

Кога x_k и x_{k+1} лежат во околина близу до решението x^*, каде $\nabla^2 f$ е позитивно дефинитен, последното разложување евентуално е доминирано од изразот $\nabla^2 f(x_{k+1})(x_{k+1} - x_k)$, па може да запишеме

$$\nabla^2 f(x_{k+1})(x_{k+1} - x_k) \approx \nabla f(x_{k+1}) - \nabla f(x_k). \tag{5.34}$$

Тогаш новата апроксимација B_{k+1} на Хесијанот се избира да го имитира својството (5.34) на вистинскиот Хесијан $\nabla^2 f(x_{k+1})$, односно да ја задоволи *секантната равенка*,

$$B_{k+1}s_k = y_k, \tag{5.35}$$

каде

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Во пракса многу често се користи апроксимацијата H_k на инверзниот Хесијан која во секоја итерација се обновува со H_{k+1} која ја задоволува *секантната равенка*,

$$s_k = H_{k+1} y_k, \tag{5.36}$$

и тогаш правецот на пребарување сепресметува според $p_k = -H_k \nabla f(x_k)$.

Ако се користи апроксимацијата на Хесијанот B_k важно е таа да биде позитивно дефинитна. Па, ако ја помножиме секантната равенка (5.35) со s_k од лево, ќе добиеме $s_k^T B_{k+1} s_k = s_k^T y_k$. Значи дека ако

$$s_k^T y_k > 0 \tag{5.37}$$

тогаш матрицата B_{k+1} е позитивно дефинитна. Неравенството (5.37) се нарекува *услов за закривеност*.

АЛГОРИТАМ 5.9 (Квази-Њутнов метод).

Чекор 0. Избери почетна точка x_0, толеранција $\varepsilon > 0$ и некоја симетрична позитивно дефинитна матрица B_0 за почетна апроксимација на Хесијанот, често $B_0 = I$ (односно некоја симетрична позитивно дефинитна матрица H_0 за почетна апроксимација на инверзниот Хесијан, често $H_0 = I$). Стави $k = 0$.

Чекор 1. Ако $\|\nabla f(x_k)\| \leq \varepsilon$, тогаш СТОП. Инаку, пресметај го правецот на пребарување според формулата $p_k = -B_k^{-1} \nabla f(x_k)$ (односно според формулата $p_k = -H_k \nabla f(x_k)$).

Чекор 2. Пресметај ја должината на чекорот α_k како решение на еднодимензионалната задача на минимизација (5.13), точно или приближно со помош на некоја процедура за линиско пребарување, обично со процедурата на линиско пребарување базирана на Волфовите услови.

Чекор 3. Земи ја за следна итерација точката $x_{k+1} = x_k + \alpha_k p_k$.

Чекор 4. Пресметај ја B_{k+1} (односно H_{k+1}) според некоја формула за обновување така да биде задоволена секантната равенка (5.35) (односно (5.36)), стави $k = k + 1$ и оди на Чекор 1.

Забелешка 5.16. Ако матриците B_k се позитивно дефинитни со рамномерно ограничен коефициент на условеност, односно, постои позитивна константа $M > 0$ така да $\|B_k\| \|B_k^{-1}\| \leq M$ за сите k, тогаш имаме дека $\cos\theta_k \geq \frac{1}{M}$, односно аголот θ_k е ограничен далеку од $\pi/2$. Па, според Забелешка 5.8, ако квази-Њутновиот метод Алгоритам 5.9 при наоѓање на должината на чекорот α_k користи процедура на линиско пребарување базирана на Волфовите услови, тогаш тој е глобално конвергентен во смисла на (5.4), т.е. $\|\nabla f(x_k)\| \to 0$, кога $k \to \infty$.

5.8.1 SR1, DFP и BFGS обновувачки формули

Една едноставна обновувачка формула за апроксимацијата на Хесијанот B_{k+1} е *симетричната-ранк-еден* (SR1) формула, дефинирана со

$$\text{(SR1)} \quad B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}, \tag{5.38}$$

која не гарантира дека обновената матрица ќе биде позитивно дефинитна, но ја зачувува симетријата на матрицата и B_{k+1} се разликува од B_k за ранк-1 матрица. Соодветната обновувачка формула за апроксимацијата H_{k+1} на инверзниот Хесијан H_{k+1} е дадена со

$$\text{(SR1)} \quad H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k}. \tag{5.39}$$

Теорема 5.15. *Нека функцијата на целта во задачата (5.1) е строго конвексна квадратна функција т.е. $f(x) = x^T Ax + b^T x$, каде A е симетрична позитивно дефинитна матрица. Тогаш, за секоја почетна точка x_0 и секоја почетна симетрична матрица H_0, итерациите $\{x_k\}$ генерирани со SR1 методот со обновувачка формула (5.39) конвергираат кон точката на минимум во најмногу n чекори, кога $(s_k - H_k y_k)^T y_k \neq 0$ за секој k. Ако после n чекори правците на пребарување p_i се линеарно независни, тогаш $H_n = A^{-1}$.*

Доказ. (доказ во Sun, Yuan, Теорема 5.1.2., во Nocedal, Wright, Теорема 8.1)∎

Друг добар аспект на SR1 формулата, е тој што генерираните матрици се добри апроксимации на Хесијанот.

При практична имплементација на SR1 методот, ако $(s_k - H_k y_k)^T y_k = 0$ или ако изразот $(s_k - H_k y_k)^T y_k$ е многу блиску до нулата, се поставува $H_{k+1} = H_k$ и така се избегнува нумеричката нестабилност и се спречува методот од неуспех.

Квази-Њутново обновување за апроксимацијата на Хесијанот B_{k+1} кое ја задржува позитивната дефинитност кај матриците, е *DFP обновувачката формула* дефинирана со

$$\text{(DFP)} \quad B_{k+1} = (I - \rho_k y_k s_k^T) B_k (I - \rho_k s_k y_k^T) + \rho_k y_k y_k^T, \tag{5.40}$$

каде

$$\rho_k = \frac{1}{y_k^T s_k}. \tag{5.41}$$

Оваа формула оригинално била предложена од Дејвидон (Davidon) во 1959, и проучувана, имплементирана и популаризирана од Флечер и Пауел (Fletcher and Powell). Соодветната *DFP обновувачка формула* за апроксимацијата на инверзниот Хесијан H_{k+1} е дефинирана со

$$\text{(DFP)} \quad H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}. \tag{5.42}$$

Главни карактеристики на DFP обновувањето се тие да разликата меѓу матриците B_k и B_{k+1} е ранк-2 матрица и се задржува симетријата на матриците. Исто така со DFP обновувањето се генерираат позитивно дефинитни матрици секогаш кога почетната апроксимација B_0 е позитивно дефинитна и условот на закривеност (5.37) е задоволен.

Теорема 5.16. *Нека функцијата на целта во задачата (5.1) е строго конвексна квадратна функција т.е. $f(x) = x^T Ax + b^T x$, каде A е симетрична позитивно дефинитна матрица. Тогаш, за секоја почетна точка x_0 и секоја почетна симетрична и позитивно дефинитна матрица H_0, правците на пребарување генерирани со DFP методот со обновувачка формула (5.42) ги поседуваат следните својства:*

(i) Секантната равенка е задоволена за сите претходни правци на пребарување т.е. $H_{i+1} y_j = s_j$, $j = 0, 1, ..., i$.

(ii) Ако $H_0 = I$, тогаш правците на пребарување се коњугирани т.е. $s_i^T A s_j = 0$ за $j = 0, 1, ..., i-1$.

(iii) Методот завршува во најмногу n чекори, и ако заврши точно во n-тиот чекор, тогаш важи $H_n = A^{-1}$.

Доказ. (доказ во Sun, Yuan, Теорема 5.1.7., во Nocedal, Wright, Теорема 8.4) ∎

Најпопуларен квази-Њутнов метод е *BFGS методот*, именуван според неговите откривачи Бројден, Флечер, Голдфарб и Шано (Broyden, Fletcher, Goldfarb and Shanno), и од 1970 овој метод генерално се смета за најефективен. Соодветната *BFGS обновувачка формула* за апроксимацијата на Хесијанот B_{k+1} е дефинирана со

$$\text{(BFGS)} \quad B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}. \tag{5.43}$$

Соодветната *BFGS обновувачка формула* за апроксимацијата H_{k+1} на инверзниот Хесијан е дефинирана со

$$\text{(BFGS)} \quad H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T, \tag{5.44}$$

каде
$$\rho_k = \frac{1}{y_k^T s_k}. \tag{5.45}$$

Основните својства на BFGS обновувањето се тие да разликата меѓу матриците B_k и B_{k+1} е ранк-2 матрица и обновувањето ја задржува симетријата на матриците. Но, една од најпосакуваните особини на BFGS обновувањето е генерирањето на позитивно дефинитни матрици секогаш кога почетната апроксимација B_0 е позитивно дефинитна и условот на закривеност (5.37) е задоволен.

Забелешка 5.17. Условот на закривеност (5.37) ќе биде задоволен ако линиското пребарување е базирано на Волфовите услови (5.16), (5.18). Кога пак f е строго конвексна функција, условот за закривеност (5.37) е задоволено за било кои два вектори x_k и x_{k+1}.

Теорема 5.17. *Нека функцијата на целта во задачата (5.1) е строго конвексна квадратна функција т.е. $f(x) = x^T Ax + b^T x$, каде A е симетрична позитивно дефинитна матрица. Тогаш, за секоја почетна точка x_0 и секоја почетна симетрична и позитивно дефинитна матрица B_0, правците на пребарување генерирани со BFGS методот со обновувачка формула (5.43) ги поседуваат следните својства:*

(i) Секантната равенка е задоволена за сите претходни правци на пребарување т.е. $B_{i+1}s_j = y_j$, $j = 0, 1, ..., i$.

(ii) Ако $H_0 = I$, тогаш правците на пребарување се коњугирани т.е. $s_i^T As_j = 0$ за $j = 0, 1, ..., i - 1$.

(iii) Методот завршува во најмногу n чекори, и ако заврши точно во n-тиот чекор, тогаш важи $B_n = A$.

Доказ. (во Nocedal, Wright, Теорема 8.4) ∎

Иако обновувачката формула DFP е прилично ефективна, таа брзо била заменета со BFGS формулата. Една од причините зошто BFGS обновувањето е попосакувано е таа да и кога се користи многу слаб критериум за линиско пребарување (да речеме $c_2 = 0.9$ во Волфовите услови), DFP методот често дава лоши резултати, што не е случај со BFGS методот. Но, кога параметарот c_2 е мал (да речеме $c_2 \leq 0.1$) тогаш не постои голема разлика меѓу повеќето обновувачки формули.

Интересно е да се напомене дека DFP и BFGS обновувачките формули се дуали една на друга, во смисла дека едната се добива од другата со замената $s \leftrightarrow y$, $B \leftrightarrow H$.

5.8.2 Анализа на конвергенцијата кај квази-Њутновите методи

Обновувачките формули за Хесијанот кај квази-Њутновите методи, ја прават анализата на конвергенцијата покомплексна отколку истата кај методот на најбрзо спуштање и Њутновиот метод. Овде ќе ја разгледаме глобалната и локалната конвергенција на BFGS методот. Потребно е воведување на одредени претпоставки.

Претпоставка 5.1.

(1) Функцијата на целта f е два пати непрекинато диференцијабилна.

(2) Функцијата на целта f е рамномерно конвексна т.е. постојат позитивни константи m и M така да за сите $x \in \Omega = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ кое е конвексно множество, важи

$$m\|u\|^2 \leq u^T \nabla^2 f(x) u \leq M\|u\|^2$$

за секој $u \in \mathbb{R}^n$.

Забелешка 5.18. Од Претпоставка 5.1 (2) следи дека Хесијанот $\nabla^2 f(x)$ е позитивно дефинитна на Ω, и f има единствен минимум на Ω.

Теорема 5.18 (Теорема за глобална конвергенција на BFGS методот). *Нека B_0 е симетрична позитивно дефинитна почетна матрица, и нека x_0 е почетна итерација за која е исполнета Претпоставка 5.1. Тогаш, низата $\{x_k\}$ генерирана од BFGS методот со линиско пребарување базирано на Волфовите услови, конвергира кон точката на минимум x^* на f.*

Доказ. (доказ во Nocedal, Wright, Теорема 8.5, во Sun, Yuan, Теорема 5.3.6.) ∎

Теорема 5.19. *Нека важи Претпоставка 5.1. Да ја разгледаме низата од итерации $x_{k+1} = x_k - B_k^{-1}\nabla f(x_k)$, каде $\{B_k\}$ е низа од симетрични и позитивно дефинитни матрици. Нека $\{x_k\}$ конвергира кон x^* така да $\nabla f(x^*) = 0$ и $\nabla^2 f(x^*)$ е позитивно дефинитна. Тогаш, $\{x_k\}$ конвергира суперлинеарно кон x^* ако и само ако*

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla^2 f(x^*))p_k\|}{\|p_k\|} = 0. \tag{5.46}$$

Доказ. (доказ во Nocedal, Wright, Теорема 3.6, во Sun, Yuan, Теорема 5.4.6) ∎

Теорема 5.20. *Нека важи Претпоставка 5.1. Да ја разгледаме низата од итерации* $x_{k+1} = x_k - \alpha_k B_k^{-1} \nabla f(x_k)$, *каде* $\{B_k\}$ *е низа од симетрични и позитивно дефинитни матрици и* α_k *ги задоволува Волфовите услови со* $c_1 \leq \frac{1}{2}$. *Ако низата* $\{x_k\}$ *конвергира кон* x^* *така да* $\nabla f(x^*) = 0$ *и* $\nabla^2 f(x^*)$ *е позитивно дефинитна и ако за правецот на пребарување важи (5.46), тогаш* $\alpha_k \to 1$ *и* $\{x_k\}$ *конвергира суперлинеарно кон* x^*.

Доказ. (доказ во Sun, Yuan, Теорема 5.4.8, во Nocedal, Wright, Теорема 3.5) ∎

Претпоставка 5.2. Хесијанот $\nabla^2 f(x)$ е Липшиц непрекинат во x^*, односно важи

$$\|\nabla^2 f(x) - \nabla^2 f(x^*)\| \leq L\|x - x^*\|,$$

за сите x во околина на x^*, каде L е позитивна константа.

Теорема 5.21 (Теорема за локална конвергенција на BFGS методот). *Нека* f *е два пати непрекинато диференцијабилна. Нека низата* $\{x_k\}$ *генерирана од BFGS методот конвергира кон точката на минимум* x^* *во која важи Претпоставка 5.2, и нека важи условот*

$$\sum_{k=1}^{\infty} \|x_k - x^*\| < \infty. \tag{5.47}$$

Тогаш, $\{x_k\}$ *конвергира суперлинеарно кон* x^*.

Доказ. (доказ во Nocedal, Wright, Теорема 8.6, и Sun, Yuan, Теорема 5.4.16.)∎

5.9 Методи без пресметување на изводи

Во овој дел се изложени некои методи за оптимизација на еднодимезинални и повеќедимензионални функции кои користат само вредности на функцијата на целта.

5.9.1 Метод на златен пресек и метод на Фибоначи

Нека $f : \mathbb{R} \to \mathbb{R}$ е непрекината функција на интервалот $[a, b]$. Овие методи се состојат во формирање на низа од интервали $[a_k, b_k]$ кои ја содржат оптималната точка x^*. Но, најнапред излагаме алгоритам за одредување на почетен интервал кој ја содржи точката x^*.

АЛГОРИТАМ 5.10 (Одредување на интервал на решението).

Чекор 0. Избери почетна точка $x_0 \in \mathbb{R}$ и правец на пребарување $h \in \mathbb{R}$. Стави $k = 0$.

Чекор 1. Ако $f(x_0 + h) < f(x_0)$ стави $x_1 = x_0 + h$, $k = 1$ и оди на Чекор 2. Во спротивно, сави $h = -h$ и ако $f(x_0 + h) < f(x_0)$ стави $x_1 = x_0 + h$, $k = 1$ и оди на Чекор 2. Ако $f(x_0+h) \geq f(x_0)$ стави $h = h/2$ и повтори го Чекор 1.

Чекор 2. Стави $h = 2h$ и пресметај $x_{k+1} = x_k + h$.

Чекор 3. Ако $f(x_{k+1}) < f(x_k)$ стави $k = k+1$ и оди на Чекор 2. Ако $f(x_{k+1}) \geq f(x_k)$ сопри го пребарувањето и за интервал кој го содржи решението земи го интервалот $[x_{k-1}, x_{k+1}]$.

Методот на златен пресек и методот на Фибоначи подлежат на една иста шема и се разликуваат само во изборот на параметарот λ. Методот на златен пресек ја користо вредноста $\lambda = \bar{\lambda} = \frac{\sqrt{5}-1}{2} \approx 0.618$, додека методот на Фибоначи ја користи вредноста $\lambda = \lambda_{n,k} = \frac{F_{n-k}}{F_{n-k+1}}$, каде $F_0 = F_1 = 1$, $F_{k+1} = F_k + F_{k-1}$, $k = 1, 2, ...$ е низата од Фибоначиеви броеви.

АЛГОРИТАМ 5.11 (Шема на методот на златен пресек и методот на Фибоначи).

Чекор 0. Избери почетен интервал $[a_0, b_0]$ кој го содржи решението, толеранција $\varepsilon > 0$ и вредност за параметарот $1/2 < \lambda < 1$. Стави $k = 0$.

Чекор 1. Пресметај $\Delta_k = b_k - a_k$. Ако $\Delta_k \leq \varepsilon$, тогаш оди на Чекор 4, инаку оди на Чекор 2.

Чекор 2. Пресметај ги $y_k = a_k + (1 - \lambda)\Delta_k$ и $z_k = a_k + \lambda\Delta_k$.

Чекор 3. Ако $f(y_k) \leq f(z_k)$, тогаш стави $a_{k+1} = a_k$, $b_{k+1} = z_k$, во спротивно стави $a_{k+1} = y_k$, $b_{k+1} = b_k$. Стави $k = k + 1$ и оди на Чекор 1.

Чекор 4. За оптимално решение земи ја вредноста a_k или b_k во која функцијата на целта има помала вредност.

Забелешка 5.19. Името на методот на златен пресек доаѓа од дефиницијата за златен пресек, имено точката $y \in [a, b]$ е *златен пресек* на интервалот $[a, b]$ ако важи

$$\frac{b-a}{b-y} = \frac{b-y}{y-a}.$$

Се покажува дека за $\lambda = \bar{\lambda} = \frac{\sqrt{5}-1}{2}$ точките y_k и z_k од Алгоритам 5.11 се точки на златен пресек на интервалот $[a_k, b_k]$. Исто така, се покажува дека секој од интервалите $[a_k, b_k]$ ја содржи оптималната точка x^*, ако почетниот интервал $[a_0, b_0]$ ја содржи, и при тоа $\Delta_k \to 0$ кога $k \to \infty$, односно за однапред дадена толеранција $\varepsilon > 0$ методот на златен пресек Алгоритам 5.11 стопира после конечно многу чекори.

Забелешка 5.20. За низата Фибоначиеви броеви $F_0 = F_1 = 1$, $F_{k+1} = F_k + F_{k-1}$, $k = 1, 2, ...$ се покажува дека важи *формулата на Бинет*, таа е

$$F_k = \frac{1}{\sqrt{5}}\Big(\Big(\frac{1+\sqrt{5}}{2}\Big)^{k+1} - \Big(\frac{1-\sqrt{5}}{2}\Big)^{k+1}\Big). \tag{5.48}$$

Исто така се покажува дека методот на Фибоначи Алгоритам 5.11 генерира низа од интервали $[a_k, b_k]$ кои ја содржат оптималната точка x^*, ако почетниот интервал $[a_0, b_0]$ ја содржи, и при тоа за однапред дадена толеранција $\varepsilon > 0$ алгоритмот ќе сопре после конечно многу чекори, ако n се одбере така да $\frac{\Delta_0}{F_n} \le \varepsilon$, односно $F_n \ge \frac{\Delta_0}{\varepsilon} = \frac{b_0 - a_0}{\varepsilon}$.

5.9.2 Метод на квадратна апроксимација

Нека $f : \mathbb{R} \to \mathbb{R}$ е непрекината функција на интервалот $[a, b]$ кој ја содржи точката на минимум x^*. Идејата на методот на квадратна апроксимација е да функцијата f се апроксимира доволно прецизно со полином од втор степен на интервалот $[a, b]$, и потоа точката ма минимум на полиномот да се земе за приближна вредност на x^*. Овој метод често се применува како краен чекор на методот на златен пресек откако е веќе најден доволно мал интервал $[a_k, b_k]$ кој ја содржи точката на минимум x^*.

Дефиниција 5.4. Тројката броеви $x_1 < x_2 < x_3$ се нарекува *погодна тројка броеви* ако $f(x_2) \le \min\{f(x_1), f(x_3)\}$ и $f(x_2) < \max\{f(x_1), f(x_3)\}$.

Забелешка 5.21. Последниот услов во Дефиниција 5.4 повлекува дека точките $(x_i, f(x_i))$, $i = 1, 2, 3$ не лежат на права паралелна со апцисната оска. Да забележиме дека ако x_1, x_2, x_3 е погодна тројка броеви, тогаш интервалот $[x_1, x_3]$ ја содржи оптималната точка x^*.

Изборот на погодна тројка броеви меѓу четири броја $x_1 < x_2 < x_3 < x_4$ за кои важи $f(x_2) \le \min\{f(x_1), f(x_4)\}$ и $f(x_3) \le \min\{f(x_1), f(x_4)\}$ се прави на следниот начин: Ако $f(x_2) < f(x_3)$, тогаш погодната тројка броеви е x_1, x_2, x_3 и ако $f(x_2) > f(x_3)$, тогаш погодната тројка броеви е x_2, x_3, x_4. Ако $f(x_2) = f(x_3)$, тогаш $x^* \in [x_2, x_3]$ и секоја од тројките x_1, x_2, x_3 и x_2, x_3, x_4 е погодна тројка броеви. Во овој случај препорачливо е да се одбере онаа тројка која одговара на пократкиот од интервалите $[x_1, x_3]$ и $[x_2, x_4]$. Или пак, да се одбере точка $x_5 \in [x_2, x_3]$ и тогаш тројката x_2, x_5, x_3 е погодна тројка броеви.

Како што беше погоре кажано, функцијата на цалта f се апроксимира со полином од втор степен $P_2(x) = ax^2 + bx + c$ и точката на минимум на полиномот на интервалот на кој се изведува оптимизацијата се зема за приближна вредност на решението x^* во тој чекор. Нека x_1, x_2, x_3 е погодна тројка броеви, тогаш апроксимациониот полином на интервалот $[x_1, x_3]$ се наоѓа со интерполација, односно треба да бидат задоволени условите $P_2(x_i) = f(x_i),\ i = 1, 2, 3$. Ако означиме со $f_i = f(x_i),\ i = 1, 2, 3$, тогаш коефициентите a, b, c ги добиваме со решавање на системот

$$ax_i^2 + bx_i + c = f_i,\ i = 1, 2, 3,$$

и за следно приближување на решението x^* го ја земаме точката на минимум на апроксимациониот полином, односно

$$\overline{x} = -\frac{b}{2a} = \frac{1}{2} \cdot \frac{(x_1^2 - x_2^2)(f_2 - f_3) - (x_2^2 - x_3^2)(f_1 - f_2)}{(x_1 - x_2)(f_2 - f_3) - (x_2 - x_3)(f_1 - f_2)}. \tag{5.49}$$

Може да се покаже дека $|\overline{x} - x_2| \leq \frac{1}{2} \max\{x_2 - x_1, x_3 - x_2\}$, од каде следи дека новото приближување $\overline{x} \in [x_1, x_3]$, што е предуслов за формирање на следна погодна тројка броеви која одговара на помал интервал, подинтервал од $[x_1, x_3]$.

АЛГОРИТАМ 5.12 (Метод на квадратна апроксимација).

Чекор 0. Избери почетна погодна тројка броеви x_1, x_2, x_3 и толеранција $\varepsilon > 0$.

Чекор 1. Пресметај го приближувањето $\overline{x}$ според формулата (5.49).

Чекор 2. Ако $|\overline{x} - x_2| \leq \varepsilon$, тогаш оди на Чекор 4, инаку оди на Чекор 3.

Чекор 3. Избери нова погодна тројка броеви меѓу броевите x_1, x_2, x_3 и $\overline{x}$ која одговара на помал интервал, изврши преозначување на броевите (повторно со $x_1 < x_2 < x_3$) и оди на Чекор 1.

Чекор 4. За оптимално решение земи ја вредноста $\overline{x}$.

Забелешка 5.22. Изборот на почетна погодна тројка броеви x_1, x_2, x_3 се изведува слично како изборот на почетен интервал кој го содржи решението Алгоритам 5.10 со доплнителни чекори за одредување на точка од интервалот која со крајните точки ќе формираат погодна тројка броеви.

Понатаму, доколу именителот во (5.49) е нула, тогаш за следно приближување се зема $\overline{x} = x_2$ и Алгоритам 5.12 запира при што за оптимално решение се зема x_2.

Забелешка 5.23. Се покажува дека низата приближувања генерирана од Алгоритам 5.12 конвергира кон решението x^* со брзина на конвергенција од ред 1.32, што значи дека постои константа $\overline{M} > 0$ така да

$$|\overline{x}_{k+1} - x^*| \leq \overline{M}\ |\overline{x}_k - x^*|^{1.32} \text{ за сите доволно големи } k,$$

каде $\overline{x}_k$ е k-тото приближување генерирано со алгоритамот.

5.9.3 Метод на Хук и Џивс

Нека $f : \mathbb{R}^n \to \mathbb{R}$ е повеќедимензионална функција на целта. Методот на Хук и Џивс користи комбинирање на правците на пребарување, почнува со координатни правци, па се префрла на правец по должината на разликата од последните две итерации, и повторно од почеток.

Да ги означиме со $e_1, e_2, ..., e_n$ се покоординатните правци, односно векторот e_j има единица на j-тото место и останатите се нули, $j = 1, 2, ..., n$.

АЛГОРИТАМ 5.13 (Метод на Хук и Џивс).

Чекор 0. Избери почетна точка x_0 и толеранција $\varepsilon > 0$. Стави $y_1 = x_0$, $j = 1$ и $k = 0$.

Чекор 1. Реши ја задачата на еднодимензионална мимизација $\min_{\lambda>0} f(y_j + \lambda e_j)$ за да ја добиеш должината на покоординатниот чекор λ_j.

Чекор 2. Стави $y_{j+1} = y_j + \lambda_j e_j$. Ако $j < n$ тогаш стави $j = j+1$ и оди на Чекор 1. Ако $j = n$ тогаш за следна итерација земи ја точката $x_{k+1} = y_{n+1}$ и оди на Чекор 3.

Чекор 3. Ако $\|x_{k+1} - x_k\| \leq \varepsilon$ тогаш СТОП, инаку оди на Чекор 4.

Чекор 4. Стави $p_k = x_{k+1} - x_k$ и реши ја задачата на еднодимензионална мимизација $\min_{\alpha>0} f(x_{k+1} + \alpha p_k)$ за да ја добиеш должината на чекорот α_k.

Чекор 5. Стави $y_1 = x_{k+1} + \alpha_k p_k$, $j = 1$, $k = k + 1$ и оди на Чекор 1.

Забелешка 5.24. Конвергенцијата на методот на Хук и Џивс се постигнува при релативно слаби услови, доволно е да се претпостави дека f е диференцијабилна и ограничена од долу функција на целта.

5.9.4 Метод на Пауел и Зангвил

Нека $f : \mathbb{R}^n \to \mathbb{R}$ е повеќедимензионална функција на целта. Методот на Пауел и Зангвил користи множество од покоординатни правци на пребарување кое во секоја итерација се менува со цел да доведе до опаѓање на функцијата на целта.

АЛГОРИТАМ 5.14 (Метод на Пауел и Зангвил).

Чекор 0. Избери почетна точка $x_{0,0}$, почетни покоординатни единични вектори $p_{1,0}, p_{2,0}, ..., p_{n,0}$ и толеранција $\varepsilon > 0$. Стави $\delta_0 = 1$ и $k = 0$.

Чекор 1. За $i = 1, 2, ..., n$ реши ја задачата на еднодимензионална мимизација $\min_{\alpha>0} f(x_{i-1,k} + \alpha p_{i,k})$ за да ја добиеш должината на покоординатниот чекор $\alpha_{i,k}$ и стави $x_{i,k} = x_{i-1,k} + \alpha_{i,k} p_{i,k}$.

Чекор 2. Стави $\lambda_k = \|x_{n,k} - x_{0,k}\|$. Ако $\lambda_k \leq \varepsilon$ тогаш СТОП, инаку оди на Чекор 3.

Чекор 3. Стави $p_{n+1,k} = (x_{n,k} - x_{0,k})/\lambda_k$. Реши ја задачата на еднодимензионална мимизација $\min_{\alpha>0} f(x_{n,k} + \alpha p_{n+1,k})$ за да ја добиеш должината на чекорот $\alpha_{n+1,k}$.

Чекор 4. Стави $x_{k+1,0} = x_{n,k} + \alpha_{n+1,k} p_{n+1,k}$. Ако $\|x_{0,k+1} - x_{0,k}\| \leq \varepsilon$ тогаш СТОП, инаку оди на Чекор 5.

Чекор 5. Одреди го m така да $\alpha_{m,k} = \max\{\alpha_{i,k} \mid i = 1, 2, ..., n\}$.

(i) Ако $\alpha_{m,k}\delta_k/\lambda_k \geq \varepsilon$ тогаш стави $p_{i,k+1} = p_{i,k}$ за $i \neq m$, $p_{m,k+1} = p_{n+1,k}$, $\delta_{k+1} = \alpha_{m,k}\delta_k/\lambda_k$ и оди на Чекор 6.

(ii) Ако $\alpha_{m,k}\delta_k/\lambda_k < \varepsilon$ тогаш стави $p_{i,k+1} = p_{i,k}$, $i = 1, 2, ..., n$, $\delta_{k+1} = \delta_k$ и оди на Чекор 6.

Чекор 6. Стави $k = k + 1$ и оди на Чекор 1.

Забелешка 5.25. На местото од почетните покоординатни вектори $p_{1,0}, p_{2,0}, ..., p_{n,0}$ може да се земе било кое множество од n линеарно независни нормализирани вектори. Чекор 4 од Алгоритам 5.14 обезбедува да новите правци бидат повторно независни.

Забелешка 5.26. Се покажува дека низите $\{x_{0,k}\}$, $\{x_{1,k}\}$, ..., $\{x_{n,k}\}$ генерирани со методот на Пауел и Зангвил Алгоритам 5.14 конвергираат кон оптималната точка, ако f е строго конвексна непрекинато диференцијабилна функција на целта.

Меѓутоа резултатите од нумеричките тестирања покажуваат дека брзината на конвергенција на сите координатни методи е приближно n пати поспора од брзината на конвергенција на методот на најбрзо спуштање.