

# FELTÉTEL NÉLKÜLI OPTIMALIZÁLÁS ALGORITMUSAI

DR. NAGY TAMÁS  
egyetemi docens

Miskolci Egyetem  
Alkalmazott Matematikai Tanszék

„A bemutatott kutató munka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg”

„This research was carried out as part of the TAMOP-4.2.1.B-10/2/KONV-2010-0001 project with support by the European Union, co-financed by the European Social Fund.”

Miskolc, 2012

# Tartalomjegyzék

<b>1</b>	<b>Bevezetés</b>	<b>3</b>
<b>2</b>	<b>Newton típusú módszerek</b>	<b>3</b>
2.1	Newton módszer . . . . .	3
2.2	Levenberg-Marquardt módszer (módosított Newton módszer) . . . . .	8
2.3	Trust region módszer . . . . .	9
2.4	Kvázi Newton módszerek . . . . .	9
<b>3</b>	<b>Vonalmenti minimumkereső eljárások</b>	<b>15</b>
3.1	Egyváltozós minimumkereső eljárások . . . . .	16
3.1.1	Szimultán keresés . . . . .	17
3.1.2	Szekvenciális keresés . . . . .	18
3.1.3	Deriváltakat használó egyváltozós kereső eljárások . . . . .	26
3.2	Többváltozós vonalmenti kereső eljárások . . . . .	29
3.2.1	Deriváltakat nem használó többváltozós vonalmenti kereső eljárások . . . . .	29
3.2.2	Deriváltakat használó többváltozós vonalmenti kereső eljárások . . . . .	35

## 1. Bevezetés

Tekintsük a  $\min \{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}$  többváltozós feltétel nélküli minimalizálási feladatot. E feladat megoldására leggyakrabban az alábbi két módszer család valamelyikét alkalmazzuk. Az egyiknél a  $\nabla f(\mathbf{x}) = \mathbf{0}$  szükséges feltételt, mint egyenletrendszert valamilyen gyökkereső eljárással megoldjuk, a másiknál pedig a vonalmenti minimumkereső eljárásokat alkalmazzuk. A  $\nabla f(\mathbf{x}) = \mathbf{0}$  egyenletrendszer megoldására a Newton típusú módszereket ismertetjük. A vonalmenti kereső eljárások közül többet is bemutatunk, egyrészt azokat a módszereket, amelyek nem használják a deriváltakat ill. azokat, amelyeknél megköveteljük a minimalizálandó függvény differenciálhatóságát. Vegyük sorra ezeket a módszereket.

## 2. Newton típusú módszerek

### 2.1. Newton módszer

Először egyenletrendszer megoldására mutatjuk be a Newton módszert, majd utána optimalizálási feladat megoldására alkalmazzuk.

Legyenek  $f_1, f_2, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$  többváltozós differenciálható valós függvények. Tekintjük az alábbi  $m$  egyenletből álló egyenletrendszert:

$$\begin{aligned} f_1(\mathbf{x}) &= 0 \\ f_2(\mathbf{x}) &= 0 \\ &\dots \\ f_m(\mathbf{x}) &= 0 \end{aligned}$$

Közelítsük ezeket a függvényeket a többváltozós valós függvényekre megismert elsőfokú Taylor polinommal egy  $\mathbf{x}_k \in \mathbb{R}^n$  helyen, azaz

$$\begin{aligned} f_1(\mathbf{x}) &\approx f_1(\mathbf{x}_k) + \nabla f_1(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) \\ f_2(\mathbf{x}) &\approx f_2(\mathbf{x}_k) + \nabla f_2(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) \\ &\dots \\ f_m(\mathbf{x}) &\approx f_m(\mathbf{x}_k) + \nabla f_m(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) \end{aligned}$$

Az egyszerűbb írásmód végett térjünk át a mátrix-vektoros jelölésre. Legyen  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , az  $\mathbf{F}$  vektor elemei legyenek az  $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})$  függvények. Legyen továbbá a  $\mathbf{J}(\mathbf{x}_k)$   $m \times n$ -es mátrix olyan, amelynek  $i$ -edik sorvektora az  $i$ -edik függvény  $\mathbf{x}_k$  pontbeli gradiense, azaz  $\mathbf{J}(\mathbf{x}_k) = \nabla \mathbf{F}(\mathbf{x}_k)$ . Ezt a mátrixot Jacobi mátrixnak nevezzük. Ekkor a fenti formulák az alábbi mátrix-vektoros formában írhatók:

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_k) + \mathbf{J}(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k).$$

Válasszuk meg az  $\mathbf{x}_{k+1}$ -edik közelítést úgy, hogy a közelítő Taylor polinom zérus legyen, azaz

$$\mathbf{F}(\mathbf{x}_k) + \mathbf{J}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = \mathbf{0},$$

ebből pedig  $\mathbf{x}_{k+1}$ -et az alábbi lineáris egyenletrendszer megoldása szolgáltatja:

$$\mathbf{J}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\mathbf{F}(\mathbf{x}_k).$$

Amennyiben a  $\mathbf{J}(\mathbf{x}_k)$  mátrix négyzetes (az egyenletek száma megegyezik a változók számával) és létezik inverze, úgy az

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{J}(\mathbf{x}_k)^{-1}\mathbf{F}(\mathbf{x}_k).$$

formulát is használhatjuk a következő közelítés számítására.

A Newton módszer tehát az alábbiakból áll. Kiindulunk egy tetszőleges  $\mathbf{x}_1 \in \mathbb{R}^n$  vektorból és a

$$\begin{aligned} \mathbf{J}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) &= -\mathbf{F}(\mathbf{x}_k) \quad \text{ill.} \\ \mathbf{x}_{k+1} &= \mathbf{x}_k - \mathbf{J}(\mathbf{x}_k)^{-1}\mathbf{F}(\mathbf{x}_k) \end{aligned}$$

iterációs formulák valamelyikével meghatározzuk az  $\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_k, \dots$  sorozatot. Az eljárást addig folytatjuk, amíg két egymást követő közelítés távolsága egy pontossági tűrés ( $\varepsilon$ ) alá nem esik, azaz  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ . Természetesen más terminálási módszereket is alkalmazhatunk.

Most alkalmazzuk a Newton módszert a feltétel nélküli optimalizálásra, azaz keressük az  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , vagyis egy  $n$  változós valós függvény minimumát. Mint tudjuk a lokális minimum szükséges feltétele  $\nabla f(\mathbf{x}) = \mathbf{0}$ . Ez pedig nem más mint egy  $n$  változós  $n$  egyenletből álló egyenletrendszer. Ennek megoldására pedig rendelkezésünkre áll az előzőekben vázolt Newton módszer, egyszerűen  $\mathbf{F}(\mathbf{x})$  helyébe  $\nabla f(\mathbf{x})$ -et kell írunk. A Jacobi mátrix ebben az esetben olyan mátrix lesz, amelynek  $i$ -edik sorvektora nem más, mint az  $f(\mathbf{x})$  függvény gradiense  $i$ -edik elemének a gradiense, azaz  $\mathbf{H}(\mathbf{x}) = \nabla \nabla f(\mathbf{x}) = \nabla^2 f(\mathbf{x})$ . Ez a mátrix pedig nem más, mint az  $f(\mathbf{x})$  függvény  $\mathbf{H}(\mathbf{x})$  Hesse mátrixa.

Az  $f(\mathbf{x})$  függvény minimumának meghatározására szolgáló Newton módszer tehát az alábbiakból áll. Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  ( $k = 1$ ) vektorból és a

$$\mathbf{H}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\nabla f(\mathbf{x}_k).$$

lineáris egyenletrendszer megoldásából meghatározzuk a következő, azaz az  $\mathbf{x}_{k+1}$  közelítést. A  $\mathbf{H}(\mathbf{x})$  Hesse mátrix, mint tudjuk négyzetes, amennyiben viszont létezik az inverze, akkor az

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}(\mathbf{x}_k)^{-1}\nabla f(\mathbf{x}_k)$$

formula is használható. Mindkét formula a Newton módszert írja le, de célszerűnek tartjuk bevezetni a Newton I. és a Newton II. módszert, mert később szó lesz a kvázi Newton módszerekről és ezeket a módszereket a közölt két formulából vezetjük le. Továbbá célszerű bevezetni az

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$$

különbségvektort, amely tehát az ismeretlen  $\mathbf{x}_{k+1}$  közelítés és az előző  $\mathbf{x}_k$  közelítés különbsége. Ebben az esetben ennek a különbségvektornak a meghatározására az alábbi formulákat használjuk:

Newton I. módszer esetén

$$\mathbf{H}(\mathbf{x}_k)\mathbf{s}_k = -\nabla f(\mathbf{x}_k),$$

Newton II. módszer esetén

$$\mathbf{s}_k = -\mathbf{H}(\mathbf{x}_k)^{-1}\nabla f(\mathbf{x}_k).$$

Az  $\mathbf{s}_k$  különbségvektor ismeretében az

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$$

formulával kapjuk a következő közelítést.

Megjegyezzük, hogy az  $f(\mathbf{x})$  függvény minimumának meghatározására szolgáló Newton módszert levezethetjük a következőképpen is. Tekintsük az  $f(\mathbf{x})$  függvény másodfokú Taylor polinomját az  $\mathbf{x}_k$  pontban. Jelölje ezt  $q(\mathbf{x})$ , amely az alábbi

$$q(\mathbf{x}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_k)\mathbf{H}(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k).$$

Határozzuk meg  $q(\mathbf{x})$  minimumát. Mint tudjuk a minimum szükséges feltétele  $\nabla q(\mathbf{x}) = \mathbf{0}$  Elvégezve a gradiensképzést, kapjuk hogy

$$\nabla f(\mathbf{x}_k) + \mathbf{H}(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) = \mathbf{0},$$

amelyből a Newton módszerre korábban kapott formula adódik.

### Példa:

Határozzuk meg az

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + x_2 - 6x_1 - 8x_2 + 2$$

függvény minimumát Newton módszerrel az  $\mathbf{x}_1 = [0, 0]$  startvektorból kiindulva!

### Megoldás:

A gradiensvektor és a Hesse mátrix a következő:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 + 2x_2 - 6 \\ 2x_1 + 4x_2 - 8 \end{bmatrix}, \quad \mathbf{H}(\mathbf{x}) = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}.$$

Ezeknek az  $\mathbf{x}_1$  pontbeli értékeik a következők:

$$\nabla f(\mathbf{x}_1) = \begin{bmatrix} -6 \\ -8 \end{bmatrix}, \quad \mathbf{H}(\mathbf{x}_1) = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}.$$

Az  $\mathbf{s}_1 = [s_1, s_2]$  ismeretlen különbségvektort meghatározó  $\mathbf{H}(\mathbf{x}_1)\mathbf{s}_1 = -\nabla f(\mathbf{x}_1)$  lineáris egyenletrendszer vektor-mátrixos formában:

$$\begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = - \begin{bmatrix} -6 \\ -8 \end{bmatrix},$$

skaláris formában:

$$\begin{aligned} 2s_1 + 2s_2 &= 6 \\ 2s_1 + 4s_2 &= 8 \end{aligned}$$

A lineáris egyenletrendszer megoldását sokféleképpen meghatározhatjuk, például pivotálással vagy a két egyenletet kivonva egymásból. A megoldás:  $2s_2 = 2$ ,  $2s_1 = 6 - 2s_2$ , ezekből a keresett különbségvektor  $\mathbf{s}_1 = [2, 1]$ .

Az  $\mathbf{s}_1 = [s_1, s_2]$  különbségvektor meghatározására szolgáló lineáris egyenletrendszert a Hesse mátrix inverzének segítségével is megoldhatjuk, így egy mátrix és egy vektor szorzásával kapjuk a megoldást (Newton II. formula:  $\mathbf{s}_k = -\mathbf{H}(\mathbf{x}_k)^{-1} \nabla f(\mathbf{x}_k)$ ), de az invertálást el kell végezni. Az invertálást pivotálással is végezhetjük, de a  $2 \times 2$ -es mátrix inverze egyszerűen számolható, mégpedig úgy, hogy a főátlóban lévő elemeket felcseréljük, a mellékátlóban lévő elemeknek pedig az előjelét váltjuk ellenkezőre és minden elemet elosztunk a mátrix determinánsával. A Hesse mátrix inverze:

$$\mathbf{H}(\mathbf{x}_1)^{-1} = \frac{1}{4} \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}.$$

A megoldás

$$\mathbf{s}_1 = -\mathbf{H}(\mathbf{x}_1)^{-1} \nabla f(\mathbf{x}_1) = -\frac{1}{4} \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} -6 \\ -8 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Az  $\mathbf{s}_1$  különbségvektor ismeretében a következő közelítés:  $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{s}_1 = [0, 0] + [2, 1] = [2, 1]$ .

A példában azt tapasztaltuk, hogy egyetlen lépésben megkaptuk az optimális megoldást. Könnyen ellenőrizhető, hogy az optimumpont  $\bar{\mathbf{x}} = [2, 1]$ . Az optimális célfüggvény érték  $f_{\min} = -8$ . Ez mindig így van, ha a minimalizálandó függvény kvadratikus és konvex. Emlékezzünk a Newton módszer második levezetésére, amikor az  $f(\mathbf{x})$  függvényt annak másodfokú Taylor polinomjával helyettesítettük, ez pedig kvadratikus függvényeknél mindig önmagával egyezik meg.

### Példa:

Határozzuk meg az

$$f(x_1, x_2) = x_1^4 x_2^2 + 2x_1^2 x_2^2 + 17$$

függvény minimumát Newton módszerrel az  $\mathbf{x}_1 = [1, -1]$  startvektorból kiindulva!

### Megoldás:

A gradiensvektor és a Hesse mátrix a következő:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 4x_1^3 x_2^2 + 4x_1 x_2^2 \\ 2x_2 x_1^4 + 4x_2 x_1^2 \end{bmatrix}, \quad \mathbf{H}(\mathbf{x}) = \begin{bmatrix} 12x_1^2 x_2^2 + 4x_2^2 & 8x_2 x_1^3 + 8x_2 x_1 \\ 8x_2 x_1^3 + 8x_2 x_1 & 2x_1^4 + 4x_1^2 \end{bmatrix}.$$

Ezeknek az  $\mathbf{x}_1$  pontbeli értékeik a következők

$$\nabla f(\mathbf{x}_1) = \begin{bmatrix} 8 \\ -6 \end{bmatrix}, \quad \mathbf{H}(\mathbf{x}_1) = \begin{bmatrix} 16 & -16 \\ -16 & 6 \end{bmatrix}.$$

A megoldandó lineáris egyenletrendszer:  $\mathbf{H}(\mathbf{x}_k) \mathbf{s}_k = -\nabla f(\mathbf{x}_k)$ , amely az  $\mathbf{s}_1 = [s_1, s_2]$  ismeretlen különbségvektor két koordinátájára az alábbi:

$$\begin{aligned} 16s_1 - 16s_2 &= -8 \\ -16s_1 + 6s_2 &= 6 \end{aligned}$$

A megoldás, azaz az  $\mathbf{x}_1$  vektorhoz tartozó  $\mathbf{s}_1$  különbségvektor  $\mathbf{s}_1 = \left[-\frac{3}{10}, \frac{2}{10}\right]$ , vagy az inverz segítségével való egyenletrendszer megoldással:

$$\mathbf{s}_1 = -\mathbf{H}(\mathbf{x}_1)^{-1} \nabla f(\mathbf{x}_1) = -\frac{1}{-160} \begin{bmatrix} 6 & 16 \\ 16 & 16 \end{bmatrix} \begin{bmatrix} 4 \\ -3 \end{bmatrix} = \begin{bmatrix} -\frac{3}{10} \\ \frac{2}{10} \end{bmatrix}.$$

Az  $\mathbf{x}_2$  közelítés:  $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{s}_1 = [1, -1] + \left[-\frac{3}{10}, \frac{2}{10}\right] = \left[-\frac{7}{10}, \frac{8}{10}\right]$ .

Láthatjuk, hogy közelebb kerültünk az optimumhoz. További lépéseket nem végzünk, mivel a módszer lényegét bemutattuk.

**Példa:**

Határozzuk meg az

$$f(x_1, x_2) = -(x_1^2 + x_2^2 - 2x_1 - 4x_2 + 6)^{-1}$$

függvény minimumát Newton módszerrel az  $\mathbf{x}_1 = [0, 1]$  startvektorból kiindulva!

**Megoldás:**

A gradiensvektor és a Hesse mátrix a következő:

$$\begin{aligned} \nabla f(\mathbf{x}) &= \begin{bmatrix} \frac{2x_1-2}{(x_1^2+x_2^2-2x_1-4x_2+6)^2} \\ \frac{2x_2-4}{(x_1^2+x_2^2-2x_1-4x_2+6)^2} \end{bmatrix}, \\ \mathbf{H}(\mathbf{x}) &= \begin{bmatrix} \frac{2(x_1^2+x_2^2-2x_1-4x_2+6)-2(2x_1-2)^2}{(x_1^2+x_2^2-2x_1-4x_2+6)^3} & -\frac{2(2x_1-2)(2x_2-4)}{(x_1^2+x_2^2-2x_1-4x_2+6)^3} \\ -\frac{2(2x_1-2)(2x_2-4)}{(x_1^2+x_2^2-2x_1-4x_2+6)^3} & \frac{2(x_1^2+x_2^2-2x_1-4x_2+6)-2(2x_2-4)^2}{(x_1^2+x_2^2-2x_1-4x_2+6)^3} \end{bmatrix}. \end{aligned}$$

Az  $\mathbf{x}_1$  pontbeli értékeik pedig

$$\nabla f(\mathbf{x}_1) = \begin{bmatrix} -\frac{2}{9} \\ -\frac{2}{9} \end{bmatrix}, \quad \mathbf{H}(\mathbf{x}_1) = \begin{bmatrix} -\frac{2}{27} & -\frac{8}{27} \\ -\frac{8}{27} & -\frac{2}{27} \end{bmatrix}.$$

A megoldandó lineáris egyenletrendszer az  $\mathbf{s}_1$  vektor két koordinátájára

$$\begin{aligned} -\frac{2}{27}s_1 - \frac{8}{27}s_2 &= \frac{2}{9} \\ -\frac{8}{27}s_1 - \frac{2}{27}s_2 &= \frac{2}{9} \end{aligned}$$

A megoldás, azaz az  $\mathbf{x}_1$  vektorhoz tartozó  $\mathbf{s}_1$  különbségvektor  $\mathbf{s}_1 = \left[-\frac{3}{5}, -\frac{3}{5}\right]$ , vagy

$$\begin{aligned} \mathbf{s}_1 &= -\mathbf{H}(\mathbf{x}_1)^{-1}\nabla f(\mathbf{x}_1) = -\left(-\frac{2}{9}\right) \left(-\frac{2}{27} \begin{bmatrix} 1 & 4 \\ 4 & 1 \end{bmatrix}\right)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \\ &= -\left(-\frac{2}{9}\right)\left(-\frac{27}{2}\right)\frac{1}{15} \begin{bmatrix} 1 & -4 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{5} \\ -\frac{3}{5} \end{bmatrix}. \end{aligned}$$

amelyből az  $\mathbf{x}_2$  közelítés  $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{s}_1 = \mathbf{x}_1 + \mathbf{s}_1 = [0, 1] + \left[-\frac{3}{5}, -\frac{3}{5}\right] = \left[-\frac{3}{5}, \frac{2}{5}\right]$ .

Azt tapasztaljuk, hogy távolodunk az optimumponttól, a célfüggvényérték pedig nem csökken, hanem növekszik. Könnyen ellenőrizhető, hogy az optimumpont  $\bar{\mathbf{x}} = [1, 2]$ . Az  $\mathbf{x}_2$  új közelítéshez tartozó célfüggvény érték  $f(\mathbf{x}_2) = -\frac{25}{153} \approx -0.16$  nagyobb lett, mint a kezdőponthoz tartozó  $f(\mathbf{x}_1) = -\frac{1}{3} \approx -0.33$  célfüggvény érték, a minimumérték pedig  $f_{\min} = -\frac{1}{5} = 0.20$ .

A Newton módszer alkalmazásánál, mint e példa is mutatta, előfordulhat, hogy a Newton módszer nem konvergál az optimumponthoz. Ezen próbál segíteni az alábbi két módszer. Ezen módszerek mellett az is segít, ha a Newton módszert összekapcsoljuk egy vonalmenti optimalizálással. Ezt be is fogjuk mutatni a későbbiekben, ennek a példának a megoldásával.

**Megjegyzés:**

A három példamegoldásban az olvasónak bizonyára feltűntek azonos jelölések, mint például az  $\mathbf{x}_1, x_1, \mathbf{x}_2, x_2, \mathbf{s}_1, s_1$ . A félkövér betű mindig vektort jelöl, a normál betű pedig skalárt, általában vektorkoordinátát jelöl. A további példákban is ezt a jelölést alkalmazzuk, de ezen megjegyzés alapján a továbbiakban az olvasónak nem okozhat gondot a formulák jó értelmezése. Ha az olvasót ez mégis zavarja, akkor használhatja a vektorok megkülönböztetésére a felső indexet, ez hasznos lehet a nem nyomtatott, kézi írásnál.

**2.2. Levenberg-Marquardt módszer (módosított Newton módszer)**

Legyen adott az  $\mathbf{x}_k \in \mathbb{R}^n$  közelítő vektor és a  $\mathbf{H}(\mathbf{x}_k)$  Hesse mátrix. A Hesse mátrixot helyettesítsük egy  $\mathbf{B}$  szimmetrikus pozitív definit mátrix-szal, amelyet a következőképpen definiálunk

$$\mathbf{B} = \mathbf{H}(\mathbf{x}_k) + \varepsilon_k \mathbf{E},$$

ahol  $\varepsilon_k > 0$  paraméter, az  $\mathbf{E}$  pedig az egységmátrix. Ezzel a módosítással megnöveljük a Hesse mátrix főátlójában lévő elemek értékét. A Newton módszerben az  $\mathbf{s}_k$  meghatározása a

$$\mathbf{B}\mathbf{s}_k = (\mathbf{H}(\mathbf{x}_k) + \varepsilon_k \mathbf{E})\mathbf{s}_k = -\nabla f(\mathbf{x}_k)$$

lineáris egyenletrendszer megoldására fog módosulni. Az algoritmus során az  $\varepsilon_k$  paramétert módosítjuk, mégpedig egy arány függvényében. Ez az arány az  $\mathbf{x}_{k+1}$  és  $\mathbf{x}_k$  pontokbeli függvényértékek különbségének hányadosa, számlálóban a minimalizálandó függvény, nevezőben pedig annak kvadratikusan közelítő függvénye szerepel. Legyen ez az arány  $r_k$ , képletben

$$r_k = \frac{f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)}{q(\mathbf{x}_{k+1}) - q(\mathbf{x}_k)}.$$

**Levenberg és Marquardt az alábbi eljárást javasolta:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  vektorból és egy  $\varepsilon_k > 0$  paraméterből.

**Közbülső lépés:**

1. Kísérletet teszünk a  $\mathbf{B} = \mathbf{H}(\mathbf{x}_k) + \varepsilon_k \mathbf{E}$  mátrix Cholesky felbontására. Ha nem sikerül az  $\mathbf{L}\mathbf{L}^T$  felbontás, akkor az  $\varepsilon_k$  paraméter értékét négyszeresére növeljük, azaz:  $\varepsilon_k := 4\varepsilon_k$ . Addig folytatjuk a kísérletet a Cholesky felbontásra és vele párhuzamosan az  $\varepsilon_k$  értékének módosítását, amíg nem sikerül az  $\mathbf{L}\mathbf{L}^T$  felbontás.

2. Ha sikerül az  $\mathbf{L}\mathbf{L}^T$  felbontás, azaz a  $\mathbf{H}(\mathbf{x}_k) + \varepsilon_k \mathbf{E}$  mátrix pozitív definit, akkor megoldjuk az  $\mathbf{L}\mathbf{L}^T \mathbf{s}_k = -\nabla f(\mathbf{x}_k)$  egyenletrendszert  $\mathbf{s}_k$ -ra.

3. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ .

4. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást a következő sorral.

5. Kiszámítjuk az  $r_k$  arányt.

Ha  $0 < r_k < 0.25$ , akkor legyen  $\varepsilon_{k+1} = 4\varepsilon_k$ .

Ha  $r_k > 0.75$ , akkor legyen  $\varepsilon_{k+1} = \frac{1}{2}\varepsilon_k$ .

Ha  $0.25 \leq r_k \leq 0.75$ , akkor legyen  $\varepsilon_{k+1} = \varepsilon_k$ .

Ha  $r_k \leq 0$ , akkor legyen  $\varepsilon_{k+1} = 4\varepsilon_k$  és az  $\mathbf{x}_{k+1}$  értékét visszaállítjuk  $\mathbf{x}_k$ -ra, azaz  $\mathbf{x}_{k+1} := \mathbf{x}_k$ .

6. Folytatjuk az eljárást,  $k := k + 1$ .



### 2.3. Trust region módszer

Minden  $\mathbf{x}_k$  közelítéshez definiálunk egy megbízhatósági tartomány (trust region) a következőképpen:

$$\Omega_k = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_k\| \leq \Delta_k\},$$

ahol  $\Delta_k > 0$  trust region paraméter. Ebben a tartományban keressük az  $\mathbf{x}_{k+1}$  közelítést, mégpedig a  $q(\mathbf{x})$  kvadratikus közelítő függvény minimumaként. Itt az algoritmus során a  $\Delta_k$  paramétert módosítjuk, hasonlóan a módosított Newton módszernél látottakhoz.

#### Trust region módszer algoritmus:

**Induló lépés** ( $k = 1$ ): Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  vektorból és egy  $\Delta_k > 0$  paraméterből.

#### Közbülső lépés:

1. Megoldjuk a  $\min \{q(\mathbf{x}) : \mathbf{x} \in \Omega_k\}$  optimalizálási feladatot, legyen az optimális megoldás  $\mathbf{x}_{k+1}$ .
2. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást a következő sorral.
3. Kiszámítjuk az  $r_k$  arányt.
4. Ha  $0 < r_k < 0.25$ , akkor legyen  $\Delta_{k+1} = \|\mathbf{x}_{k+1} - \mathbf{x}_k\| / 4$ .  
Ha  $r_k > 0.75$  és  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| = \Delta_k$  (az  $\mathbf{x}_{k+1}$  a trust region határán van), akkor legyen  $\Delta_{k+1} = 2\Delta_k$ .  
Ha a fenti két eset egyike sem áll fenn, akkor legyen  $\Delta_{k+1} = \Delta_k$ .  
Ha  $r_k \leq 0$ , akkor legyen  $\Delta_{k+1} = \|\mathbf{x}_{k+1} - \mathbf{x}_k\| / 4$  és az  $\mathbf{x}_{k+1}$  értékét visszaállítjuk  $\mathbf{x}_k$ -ra, azaz  $\mathbf{x}_{k+1} := \mathbf{x}_k$ .
5. Folytatjuk az eljárást,  $k := k + 1$ .

### 2.4. Kvázi Newton módszerek

A Newton módszernek **egyenletrendszer** megoldására szolgáló változata esetén szükségünk van az adott pontban a Jacobi mátrixra ill. **optimalizálás** esetén a Hesse mátrixra. A kvázi Newton módszereknél ezeket az elsőrendű ill. a másodrendű deriváltakat nem számítjuk ki, hanem helyette egy egyszerű képlettel számítjuk ki azok közelítését. Tekintsük az  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$  egyenletrendszert, ahol  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Ha az  $\mathbf{F}$  függvényt az  $\mathbf{x}_k \in \mathbb{R}^n$  pontbeli elsőfokú Taylor polinommal közelítjük, akkor az  $\mathbf{x}_{k+1}$  pontban a függvényérték a közelítése

$$\mathbf{F}(\mathbf{x}_{k+1}) \approx \mathbf{F}(\mathbf{x}_k) + \mathbf{J}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k).$$

Mivel a mi feladatunk alapvetően egy  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  függvény minimumának a meghatározása, ezért a kvázi Newton módszerhez szükséges levezetéseket csak optimalizálás esetére végezzük el. A szükséges feltétel szerint az optimalizálásban a  $\nabla f(\mathbf{x}) = \mathbf{0}$  egyenletrendszert kell megoldani, tehát a fenti összefüggés az alábbiakra módosul

$$\nabla f(\mathbf{x}_{k+1}) \approx \nabla f(\mathbf{x}_k) + \mathbf{H}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k).$$

Vezessük be az  $\mathbf{y}_k$  jelölést a gradiensek különbségére, az  $\mathbf{s}_k$  jelölést pedig az  $\mathbf{x}$  vektorok különbségére, azaz legyen

$$\begin{aligned} \mathbf{s}_k &= \mathbf{x}_{k+1} - \mathbf{x}_k \\ \mathbf{y}_k &= \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k) \end{aligned}$$

ekkor a fenti összefüggés az alábbi egyszerű alakban írható

$$\mathbf{y}_k \approx \mathbf{H}(\mathbf{x}_k)\mathbf{s}_k.$$

Ha most a  $\mathbf{H}(\mathbf{x}_k)$  Hesse mátrixot egy  $\mathbf{B}_k$  mátrix-szal helyettesítjük, akkor

$$\mathbf{y}_k \approx \mathbf{B}_k\mathbf{s}_k.$$

Válasszuk meg a Hesse mátrix következő helyettesítését, a  $\mathbf{B}_{k+1}$  mátrixot úgy, hogy egyenlőség álljon fenn az előző formulában, azaz

$$\mathbf{y}_k = \mathbf{B}_{k+1}\mathbf{s}_k.$$

Ezt az összefüggést kvázi Newton feltételnek vagy más elnevezéssel szelő egyenletnek is szokás nevezni.

A  $\mathbf{B}_{k+1}$  mátrixot sokféleképpen meg lehet határozni. Szokásos eljárás a  $\mathbf{B}_{k+1}$  meghatározására az, hogy a  $\mathbf{B}_k$  mátrixhoz egy vagy több diadikus mátrixot adunk hozzá.

### Egy diadikus mátrix alkalmazása (egyrangú formula)

Legyen  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ,  $\alpha \in \mathbb{R}$  és módosítsuk a  $\mathbf{B}$  mátrixot az alábbiak szerint

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \alpha\mathbf{u}\mathbf{v}^T.$$

Ennek ki kell elégíteni a kvázi Newton feltételt, azaz

$$\mathbf{y}_k = \mathbf{B}_{k+1}\mathbf{s}_k = (\mathbf{B}_k + \alpha\mathbf{u}\mathbf{v}^T)\mathbf{s}_k = \mathbf{B}_k\mathbf{s}_k + (\alpha\mathbf{v}^T\mathbf{s}_k)\mathbf{u}.$$

Ebből az egyenletből  $\mathbf{u} = \mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k$ ,  $\alpha = 1/(\mathbf{v}^T\mathbf{s}_k)$  adódik,  $\mathbf{v}$  pedig egy alkalmasan választott vektor. Szokásos a  $\mathbf{v} = \mathbf{u}$ , vagy a  $\mathbf{v} = \mathbf{s}_k$  választás, amelyből az új  $\mathbf{B}$  mátrix

$$\begin{aligned} \mathbf{B}_{k+1} &= \mathbf{B}_k + \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^T}{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^T\mathbf{s}_k} \quad \text{ill.} \\ \mathbf{B}_{k+1} &= \mathbf{B}_k + \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{s}_k}. \end{aligned}$$

### Két diadikus mátrix alkalmazása (kétrangú formula)

Az egyszerűség kedvéért az egyes diádokban most ugyanazokat a vektorokat használjuk, legyen  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ,  $\alpha, \beta \in \mathbb{R}$  és módosítsuk a  $\mathbf{B}$  mátrixot az alábbiak szerint

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \alpha\mathbf{u}\mathbf{u}^T + \beta\mathbf{v}\mathbf{v}^T.$$

Ekkor

$$\mathbf{y}_k = \mathbf{B}_{k+1}\mathbf{s}_k = (\mathbf{B}_k + \alpha\mathbf{u}\mathbf{u}^T + \beta\mathbf{v}\mathbf{v}^T)\mathbf{s}_k = \mathbf{B}_k\mathbf{s}_k + (\alpha\mathbf{u}^T\mathbf{s}_k)\mathbf{u} + (\beta\mathbf{v}^T\mathbf{s}_k)\mathbf{v}.$$

Ebből az egyenletből  $\mathbf{u} = \mathbf{y}_k$ ,  $\mathbf{v} = \mathbf{B}_k\mathbf{s}_k$ ,  $\alpha = 1/(\mathbf{u}^T\mathbf{s}_k)$ ,  $\beta = -1/(\mathbf{v}^T\mathbf{s}_k)$  adódik, amelyből az új  $\mathbf{B}$  mátrix

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{y}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k} - \frac{(\mathbf{B}_k\mathbf{s}_k)(\mathbf{B}_k\mathbf{s}_k)^T}{(\mathbf{B}_k\mathbf{s}_k)^T\mathbf{s}_k}.$$

Most további eljárásokat ismertetünk. Míg az előzőekben a Hesse mátrixot, most a Hesse mátrix inverzét helyettesítjük. A kiinduló formula az előzőekből ismert

$$\nabla f(\mathbf{x}_{k+1}) \approx \nabla f(\mathbf{x}_k) + \mathbf{H}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k).$$

A Hesse mátrix inverzét használva az ismert jelöléseket figyelembe véve az alábbi formulát kapjuk

$$\mathbf{s}_k \approx \mathbf{H}(\mathbf{x}_k)^{-1} \mathbf{y}_k.$$

Ha most a  $\mathbf{H}(\mathbf{x}_k)^{-1}$  mátrixot egy  $\mathbf{D}_k$  mátrix-szal helyettesítjük, akkor

$$\mathbf{s}_k \approx \mathbf{D}_k \mathbf{y}_k.$$

Válasszuk meg a  $\mathbf{D}_{k+1}$  mátrixot úgy, hogy

$$\mathbf{s}_k = \mathbf{D}_{k+1} \mathbf{y}_k.$$

Ezt az összefüggést szintén kvázi Newton feltételnek szokás nevezni.

A  $\mathbf{D}_{k+1}$  mátrix meghatározását a  $\mathbf{B}_{k+1}$  mátrix meghatározásához hasonlóan végezzük. A levezetést az olvasóra bízunk, itt csupán a formulákat közöljük.

### Egy diadikus mátrix alkalmazása (egyrangú formula)

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \frac{(\mathbf{s}_k - \mathbf{D}_k \mathbf{y}_k)(\mathbf{s}_k - \mathbf{D}_k \mathbf{y}_k)^T}{(\mathbf{s}_k - \mathbf{D}_k \mathbf{y}_k)^T \mathbf{y}_k}.$$

### Két diadikus mátrix alkalmazása (kétrangú formula)

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{(\mathbf{D}_k \mathbf{y}_k)(\mathbf{D}_k \mathbf{y}_k)^T}{(\mathbf{D}_k \mathbf{y}_k)^T \mathbf{y}_k}.$$

Érdemes megjegyezni a formulák közötti duális kapcsolatot, miszerint a  $\mathbf{D}$  és  $\mathbf{B}$ , valamint az  $\mathbf{s}$  és  $\mathbf{y}$  betűk cseréjével egyik formulából adódik a másik formula.

Összefoglalóan, a sok lehetséges kvázi Newton módszert használó eljárások közül egy-egy eljárást ismertetünk. A Hesse mátrixot helyettesítő eljárások közül a leggyakrabban használt eljárás a BFGS eljárás, míg a Hesse mátrix inverzét helyettesítő eljárások közül a leggyakrabban használt eljárás a DFP eljárás.

### BFGS (Broyden-Fletcher-Goldfarb-Shanno) eljárás

**Induló lépés** ( $k = 1$ ): Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  vektorból és egy szimmetrikus pozitív definit  $\mathbf{B}_k$  mátrixból.

#### Közbülső lépés:

1. Megoldjuk a

$$\mathbf{B}_k \mathbf{s}_k = -\nabla f(\mathbf{x}_k)$$

lineáris egyenletrendszer  $\mathbf{s}_k$  különbségvektorra.

2. Meghatározzuk a következő közelítést:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k.$$

3. Az alábbi formulákkal kiszámítjuk az  $\mathbf{y}_k$  vektort és a Hesse mátrixot helyettesítő mátrix következő,  $\mathbf{B}_{k+1}$  közelítő mátrixát:

$$\begin{aligned}\mathbf{y}_k &= \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k), \\ \mathbf{B}_{k+1} &= \mathbf{B}_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{(\mathbf{B}_k \mathbf{s}_k)(\mathbf{B}_k \mathbf{s}_k)^T}{(\mathbf{B}_k \mathbf{s}_k)^T \mathbf{s}_k}.\end{aligned}$$

Az eljárást addig folytatjuk, amíg  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ .

A fenti eljárást kidolgozóiról Broyden, Fletcher, Goldfarb, Shanno kutatókról nevezték el.

Megjegyzés: Az egyrangú formula esetén, ha a  $\mathbf{v} = \mathbf{s}_k$  választást használjuk, akkor a módszert Broyden módszernek nevezzük.

### DFP (Davidon-Fletcher-Powell) eljárás

**Induló lépés** ( $k = 1$ ): Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  vektorból és egy szimmetrikus pozitív definit  $\mathbf{D}_k$  mátrixból.

#### Közbülső lépés:

1. Kiszámítjuk az  $\mathbf{s}_k$  különbségvektort a következő formulával

$$\mathbf{s}_k = -\mathbf{D}_k \nabla f(\mathbf{x}_k).$$

2. Meghatározzuk a következő közelítést:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k.$$

3. Az alábbi formulákkal kiszámítjuk az  $\mathbf{y}_k$  vektort és a Hesse mátrix inverzét helyettesítő mátrix következő,  $\mathbf{D}_{k+1}$  közelítő mátrixát:

$$\begin{aligned}\mathbf{y}_k &= \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k), \\ \mathbf{D}_{k+1} &= \mathbf{D}_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{(\mathbf{D}_k \mathbf{y}_k)(\mathbf{D}_k \mathbf{y}_k)^T}{(\mathbf{D}_k \mathbf{y}_k)^T \mathbf{y}_k}.\end{aligned}$$

Az eljárást addig folytatjuk, amíg  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ .

A fenti eljárást kidolgozóiról Davidon, Fletcher, Powell kutatókról nevezték el.

Fentebb a kétrangú formulát használtuk  $\mathbf{D}_{k+1}$  számítására, értelemszerűen alkalmazható az egyrangú formula is.

#### Példa:

Oldjuk meg az alábbi optimalizálási feladatot Broyden-Fletcher-Goldfarb-Shanno módszerrel!

$$f(x_1, x_2) = x_1^2 + \frac{1}{2}x_2^2 + 7 \rightarrow \min!$$

Legyen az  $\mathbf{x}_1$  startvektor és a Hesse mátrixot helyettesítő  $\mathbf{B}_1$  startmátrix az alábbi:

$$\mathbf{x}_1 = (1, -1), \quad \mathbf{B}_1 = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}.$$

#### Megoldás:

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ x_2 \end{bmatrix}.$$

1. lépés:

$$\mathbf{x}_1 = (1, -1), \quad \nabla f(\mathbf{x}_1) = (2, -1)$$

A megoldandó egyenletrendszer ( $\mathbf{B}_1 \mathbf{s}_1 = -\nabla f(\mathbf{x}_1)$ ):

$$\begin{aligned} 2s_1 - s_2 &= -2 \\ -s_1 + 4s_2 &= 1 \end{aligned}$$

Az egyenletrendszer megoldása:  $s_1 = -1$ ,  $s_2 = 0$ , így  $\mathbf{s}_1 = (-1, 0)$ .

2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{s}_1 = (1, -1) - (-1, 0) = (0, -1)$$

Most előkészítjük a  $\mathbf{B}_2$  mátrix számítását.

$$\nabla f(\mathbf{x}_2) = (0, -1)$$

$$\mathbf{y}_1 = \nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1) = (0, -1) - (2, -1) = (-2, 0)$$

$\mathbf{B}_1 \mathbf{s}_1 = (-2, 1)$ , ezt számolás nélkül is tudjuk az egyenletrendszerből, ugyanis  $\mathbf{B}_1 \mathbf{s}_1 = -\nabla f(\mathbf{x}_1)$ .

$$\mathbf{y}_1 \mathbf{y}_1^T = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \begin{bmatrix} -2 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{y}_1^T \mathbf{s}_1 = \begin{bmatrix} -2 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = 2$$

$$(\mathbf{B}_1 \mathbf{s}_1)(\mathbf{B}_1 \mathbf{s}_1)^T = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \begin{bmatrix} -2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}$$

$$(\mathbf{B}_1 \mathbf{s}_1)^T \mathbf{s}_1 = \begin{bmatrix} -2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = 2$$

$$\mathbf{B}_2 = \mathbf{B}_1 + \frac{\mathbf{y}_1 \mathbf{y}_1^T}{\mathbf{y}_1^T \mathbf{s}_1} - \frac{(\mathbf{B}_1 \mathbf{s}_1)(\mathbf{B}_1 \mathbf{s}_1)^T}{(\mathbf{B}_1 \mathbf{s}_1)^T \mathbf{s}_1} = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & \frac{7}{2} \end{bmatrix}$$

3. lépés:

$$\mathbf{x}_2 = (0, -1), \quad \nabla f(\mathbf{x}_2) = (0, -1)$$

A megoldandó egyenletrendszer ( $\mathbf{B}_2 \mathbf{s}_2 = -\nabla f(\mathbf{x}_2)$ ):

$$\begin{aligned} 2s_1 - s_2 &= 0 \\ -s_1 + 4s_2 &= 1 \end{aligned}$$

Az egyenletrendszer megoldása:  $s_1 = \frac{1}{7}$ ,  $s_2 = \frac{2}{7}$ , így  $\mathbf{s}_2 = (\frac{1}{7}, \frac{2}{7})$ .

4. lépés:

$$\mathbf{x}_3 = \mathbf{x}_2 + \mathbf{s}_2 = (0, -1) - (\frac{1}{7}, \frac{2}{7}) = (\frac{1}{7}, -\frac{9}{7})$$

A számítást nem folytatjuk tovább, hiszen a BFGS módszer lépéseit bemutattuk.

**Feladat:**

A gyakorlás végett számítsa ki a  $\mathbf{B}_3$  mátrixot!

**Példa:**

Oldjuk meg az alábbi optimalizálási feladatot Davidon-Fletcher-Powell módszerrel!

$$f(x_1, x_2) = x_1^2 + 2x_2^2 - 9 \rightarrow \min!$$

Legyen az  $\mathbf{x}_1$  startvektor és a Hesse mátrixot helyettesítő  $\mathbf{D}_1$  startmátrix az alábbi:

$$\mathbf{x}_1 = (-1, 1), \quad \mathbf{D}_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

**Megoldás:**

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix}.$$

1. lépés:

$$\mathbf{x}_1 = (-1, 1), \quad \nabla f(\mathbf{x}_1) = (-2, 4),$$

$$\mathbf{s}_1 = -\mathbf{D}_1 \nabla f(\mathbf{x}_1) = - \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ -10 \end{bmatrix} \text{ vagy más jelöléssel } \mathbf{s}_1 = (6, -10)$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{s}_1 = (-1, 1) + (6, -10) = (5, -9)$$

Most előkészítjük a  $\mathbf{D}_2$  mátrix számítását.

$$\nabla f(\mathbf{x}_2) = (10, -36)$$

$$\mathbf{y}_1 = \nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1) = (10, -36) - (-2, 4) = (12, -40)$$

$$\mathbf{D}_1 \mathbf{y}_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 12 \\ -40 \end{bmatrix} = (52, -92)$$

$$\mathbf{s}_1 \mathbf{s}_1^T = \begin{bmatrix} 6 \\ -10 \end{bmatrix} \begin{bmatrix} 6 & -10 \end{bmatrix} = \begin{bmatrix} 36 & -60 \\ -60 & 100 \end{bmatrix}$$

$$\mathbf{s}_1^T \mathbf{y}_1 = \begin{bmatrix} 6 & -10 \end{bmatrix} \begin{bmatrix} 12 \\ -40 \end{bmatrix} = 472$$

$$(\mathbf{D}_1 \mathbf{y}_1)(\mathbf{D}_1 \mathbf{y}_1)^T = \begin{bmatrix} 52 \\ -92 \end{bmatrix} \begin{bmatrix} 52 & -92 \end{bmatrix} = \begin{bmatrix} 2704 & -4784 \\ -4784 & 8464 \end{bmatrix}$$

$$(\mathbf{D}_1 \mathbf{y}_1)^T \mathbf{y}_1 = \begin{bmatrix} 52 & -92 \end{bmatrix} \begin{bmatrix} 12 \\ -40 \end{bmatrix} = 4304$$

$$\begin{aligned} \mathbf{D}_2 &= \mathbf{D}_1 + \frac{\mathbf{s}_1 \mathbf{s}_1^T}{\mathbf{s}_1^T \mathbf{y}_1} - \frac{(\mathbf{D}_1 \mathbf{y}_1)(\mathbf{D}_1 \mathbf{y}_1)^T}{(\mathbf{D}_1 \mathbf{y}_1)^T \mathbf{y}_1} = \\ &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} + \frac{1}{472} \begin{bmatrix} 36 & -60 \\ -60 & 100 \end{bmatrix} - \frac{1}{4304} \begin{bmatrix} 2704 & -4784 \\ -4784 & 8464 \end{bmatrix} = \\ &= \begin{bmatrix} \frac{14221}{31742} & -\frac{495}{31742} \\ -\frac{495}{31742} & \frac{7787}{31742} \end{bmatrix} = \begin{bmatrix} 0.44802 & -0.015594 \\ -0.015594 & 0.24532 \end{bmatrix} \end{aligned}$$

3. lépés:

$$\mathbf{x}_2 = (5, -9), \quad \nabla f(\mathbf{x}_2) = (10, -36)$$

$$\mathbf{s}_2 = -\mathbf{D}_2 \nabla f(\mathbf{x}_2) = - \begin{bmatrix} 0.44802 & -0.015594 \\ -0.015594 & 0.24532 \end{bmatrix} \begin{bmatrix} 10 \\ -36 \end{bmatrix} = \begin{bmatrix} -5.0416 \\ 8.9875 \end{bmatrix} \text{ vagy más}$$

jelöléssel  $\mathbf{s}_2 = (-5.0416, 8.9875)$

4. lépés:

$$\mathbf{x}_3 = \mathbf{x}_2 + \mathbf{s}_2 = (5, -9) + (-5.0416, 8.9875) = (-0.0416, -0.0125)$$

A számítást nem folytatjuk tovább, hiszen a DFP módszer lépéseit bemutattuk.

**Megjegyzés:**

További formulák is származtathatók a

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{(1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u})}, \quad 1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u} \neq 0.$$

Sherman-Morrison-Woodbury formula segítségével. A BFGS eljárásban szereplő  $\mathbf{B}$  mátrix inverzét használhatjuk a DFP eljáráshoz  $\mathbf{D}$  mátrixként, ekkor a  $\mathbf{D}$  mátrixra vonatkozó formula:

$$\mathbf{D}_{k+1}^{BFGS} = \mathbf{D}_k + \left(1 + \frac{(\mathbf{D}_k\mathbf{y}_k)^T\mathbf{y}_k}{\mathbf{s}_k^T\mathbf{y}_k}\right) \frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k} - \frac{\mathbf{s}_k(\mathbf{D}_k\mathbf{y}_k)^T + (\mathbf{D}_k\mathbf{y}_k)\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k}.$$

A DFP eljárásban szereplő  $\mathbf{D}$  mátrix inverzét használhatjuk a BFGS eljáráshoz  $\mathbf{B}$  mátrixként, ekkor a  $\mathbf{B}$  mátrixra vonatkozó formula:

$$\mathbf{B}_{k+1}^{DFP} = \mathbf{B}_k + \left(1 + \frac{(\mathbf{B}_k\mathbf{s}_k)^T\mathbf{s}_k}{\mathbf{y}_k^T\mathbf{s}_k}\right) \frac{\mathbf{y}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k} - \frac{\mathbf{y}_k(\mathbf{B}_k\mathbf{s}_k)^T + (\mathbf{B}_k\mathbf{s}_k)\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k}.$$

Hasonlóan érvényes a dualitás ebben az esetben is.

### 3. Vonalmenti minimumkereső eljárások

A módszerek lényege a következőképpen foglalható össze.

**Induló lépés** ( $k = 1$ ): Kiindulunk egy  $\mathbf{x}_k \in \mathbb{R}^n$  pontból és választunk egy alkalmas  $\mathbf{d}_k \in \mathbb{R}^n$  kereső irányvektort.

**Közbülső lépés:** A  $\mathbf{d}_k$  irányban megkeressük a függvény minimumát, azaz a

$$\min \{f(\mathbf{x}_k + \lambda\mathbf{d}_k) : \lambda \geq 0\}$$

minimalizálási feladatot kell megoldani. Ez a feladat egy egyváltozós minimalizálási probléma. Legyen a  $\varphi(\lambda) = f(\mathbf{x}_k + \lambda\mathbf{d}_k)$  egyváltozós függvény minimuma  $\lambda_k$ . Ekkor a következő közelítés  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k\mathbf{d}_k$ .

A módszerek abban fognak különbözni, hogy a kereső irányt milyen módon határozzuk meg. Mielőtt azonban ezekre a módszerekre rátérnénk bemutatjuk az egyváltozós minimalizálási technikákat, hisz a vonalmenti minimumkereső eljárások magját az egyváltozós optimalizálási probléma adja.

### 3.1. Egyváltozós minimumkereső eljárások

Tekintsük a  $\min \{f(x) : x \in \mathbb{R}\}$  egyváltozós minimalizálási feladatot. Első lépésként határozzunk meg egy olyan  $[a, b]$  intervallumot, amely tartalmazza a minimumpontot. Ezt az intervallumot szokás **bizonytalansági intervallumnak** is tekinteni, hiszen csak azt tudjuk, hogy a minimumpont itt keresendő, de a helyéről nincs információnk. A következőkben bemutatott módszerek lényege abban áll, hogy a bizonytalansági intervallumot lépésről lépésre szűkítjük. Ehhez nem teszünk mást, mint az intervallumot részekre osztjuk (általában két belső pont felvételével) és az osztópontokban a függvényértéket kiszámítjuk, a függvényértékek nagysága alapján fogjuk az új (szűkebb) bizonytalansági intervallumot meghatározni. Ahhoz, hogy az eljárás konvergens legyen a függvénynek bizonyos konvexitási tulajdonsággal kell rendelkeznie. Megjegyezzük, hogy sok esetben anélkül is alkalmazzuk az alábbi eljárásokat, hogy meggyőződnénk a konvexitásról, ekkor azonban nincs garancia az eljárások konvergenciájára. Az alábbi tétel adja meg a szűkítés lehetőségét.

#### TÉTEL

Legyen  $f : \mathbb{R} \rightarrow \mathbb{R}$  szigorúan kvázikonvex függvény az  $[a, b]$  intervallumon. Legyen  $c, d \in (a, b)$  olyan, hogy  $c < d$ .

- (a) Ha  $f(c) > f(d)$ , akkor  $f(x) \geq f(d)$  minden  $x \in [a, c)$  esetén, így az új bizonytalansági intervallum  $[c, b]$ .
- (b) Ha  $f(c) \leq f(d)$ , akkor  $f(x) \geq f(c)$  minden  $x \in (d, b]$  esetén, így az új bizonytalansági intervallum  $[a, d]$ .

#### Bizonyítás

(a) Indirekte tegyük fel, hogy  $f(x) < f(d)$  minden  $x \in [a, c)$  esetén. Mivel  $c = \lambda x + (1 - \lambda)d$ ,  $\lambda \in (0, 1)$  és a függvény szigorúan kvázikonvex, így

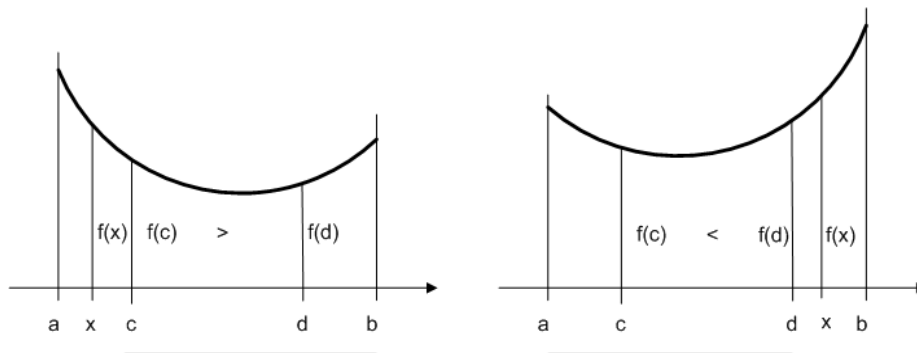
$$f(c) < \max \{f(x), f(d)\} = f(d).$$

Ez pedig ellentmond annak, hogy  $f(c) > f(d)$ .

(b) Hasonlóan bizonyítható (i)-hez.

**Q.e.d.**

A tétel és a bizonyítás jobb megértését az alábbi ábrával segítjük elő:



Az alábbi alfejezetekben háromféle keresési módszercsaládot mutatunk be: a szimultán keresést, a szekvenciális keresést és a függvény deriváltjait is használó keresést.



### 3.1.1. Szimultán keresés

**Uniform keresés** A szimultán keresési módszerek közül a legelterjedtebb módszer az uniform keresés.

Osszuk be az  $[a, b]$  intervallumot  $n$  egyenlő részre.

Az így kapott  $a = x_0, x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n = b$  pontokban számítsuk ki a függvényértékeket, majd határozzuk meg ezek minimumát. Tegyük fel, hogy a minimum az  $x_i$  pontban éretik el. Az előző tétel alapján mondhatjuk, hogy sem az  $x_{i-1}$  pont előtt, sem az  $x_{i+1}$  pont után nem lehet a minimumpont, így az új bizonytalansági intervallum  $[x_{i-1}, x_{i+1}]$ . Ha azt akarjuk, hogy az új bizonytalansági intervallum kicsi legyen, akkor elég nagy számú függvény-kiértékelést kell végezni, ezért gyakran használják azt a módszert, hogy először kevesebb részre osztják az intervallumot, aztán később finomítják a beosztást.

Az eljárást addig folytatjuk, amíg olyan  $[a_k, b_k]$  bizonytalansági intervallumhoz nem érünk, amelynek hossza ( $L_k = b_k - a_k$ ) kisebb, mint egy előre megadott tűrés ( $\varepsilon$ ) kétszerese, azaz ha valamely  $k$ -ra  $L_k < 2\varepsilon$ . Ha a minimumpontot a megállásnál kapott  $[a_k, b_k]$  intervallum középpontjára választjuk, akkor  $\varepsilon$ -nál kisebb hibát követünk el a minimumpont közelítésében.

**Példa:**

Oldjuk meg az  $f(x) = x^2 - 5x + 10 \rightarrow \min!$  egyváltozós optimalizálási feladatot uniform keresés módszerrel, ha a bizonytalansági intervallum  $[a, b] = [1, 5]$ ,  $\varepsilon = 0.5$ .

**Megoldás:**

1. lépés:

Az első közelítésbeli bizonytalansági intervallum  $[a_1, b_1] = [1, 5]$ , melynek hossza  $L_1 = 4$ . Osszuk be az  $[a_1, b_1]$  intervallumot  $n = 5$  egyenlő részre. Ekkor a lépésköz  $h = L_1/n = 0.8$ , az  $x_i$  értékek pedig  $x_i = a_1 + i \cdot h$  ( $i = 0, 1, \dots, n$ ).

Az így kapott  $a_1 = x_0, x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n = b_1$  pontokat és a hozzájuk tartozó a függvényértékeket az alábbi táblázatban közöljük.

$i$	0	1	2	3	4	5
$x_i$	1	1.8	2.6	3.4	4.2	5
$f(x_i)$	6.0	4.24	3.76	4.56	6.64	10.0

A függvényértékek minimuma 3.76 és ezt az  $x_2$  helyen veszi fel. Az új bizonytalansági intervallum tehát a minimumpontot közvetlenül megelőző és követő intervallumok uniója, azaz  $[1.8, 3.4]$ .

2. lépés:

A második közelítésbeli bizonytalansági intervallum  $[a_2, b_2] = [1.8, 3.4]$ , melynek hossza  $L_2 = 1.6$ . Osszuk be az  $[a_2, b_2]$  intervallumot  $n = 4$  egyenlő részre. Ekkor a lépésköz  $h = L_2/n = 0.4$ . Az így kapott  $a_2 = x_0, x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n = b_2$  pontokat és a hozzájuk tartozó a függvényértékeket az alábbi táblázatban közöljük.

$i$	0	1	2	3	4
$x_i$	1.8	2.2	2.6	3.0	3.4
$f(x_i)$	4.24	3.84	3.76	4.0	4.56

A függvényértékek minimuma 3.76 és ezt az  $x_2$  helyen veszi fel. Az új bizonytalansági intervallum tehát a minimumpontot közvetlenül megelőző és követő intervallumok uniója, azaz  $[2.2, 3.0]$ .

3. lépés:

A harmadik közelítésbeli bizonytalansági intervallum  $[a_3, b_3] = [2.2, 3.0]$ , melynek hossza  $L_3 = 0.8$ . Itt megállunk, mivel  $L_3 < 2\varepsilon$ . A minimumpont közelítésére az intervallum felezőpontját választjuk, így  $x_{\min} \approx 2.6$ . Megjegyezzük, hogy jobb eredmény eléréséhez általában kisebb  $\varepsilon$  értéket szokás megadni. Az osztópontok száma is általában nagyobb a gyakorlatban.

### 3.1.2. Szekvenciális keresés

A szekvenciális keresések lényege az, hogy az  $[a, b]$  intervallum belsejében felvesszünk két pontot, jelölje ezeket  $c, d$  ( $c < d$ ). A közölt tétel értelmében

ha  $f(c) > f(d)$ , akkor az új bizonytalansági intervallum  $[c, b]$ ,

ha pedig  $f(c) \leq f(d)$ , akkor az új bizonytalansági intervallum  $[a, d]$ .

Mivel nem tudjuk, hogy melyik eset fog előfordulni, így a  $c$  és  $d$  pontokra az optimális választás az, amelynél az új bizonytalansági intervallumok hossza megegyezik. Az alábbi eljárások mindegyikében ez az optimális elv érvényesül. Az eljárások különbözősége a két pont konkrét megválasztásában van.

Az eljárások ismertetése előtt bevezetjük a következő jelöléseket. Jelölje az eredeti  $[a, b]$  intervallumot  $[a_1, b_1]$ . A további intervallumokat pedig jelölje  $[a_k, b_k]$ . Az  $[a_k, b_k]$  intervallumban választott két pontot jelölje  $c_k, d_k$ . Továbbá jelölje az intervallumok hosszát  $L_k$ , ahol  $L_k = b_k - a_k$ .

A szekvenciális eljárások közül hármat ismertetünk, nevezetesen a Dichotomous keresést, a Golden section (arany metszés) keresést és a Fibonacci keresést.

**Dichotomous keresés** Legegyszerűbben úgy tudjuk biztosítani az új bizonytalansági intervallumok hosszának megegyezését, ha az intervallum középpontjától balra és jobbra azonos távolságra választjuk meg a  $c$  és a  $d$  pontokat. Jelölje a középponttól való távolságot, egy alkalmasan választott  $\delta > 0$  szám.

Ekkor az alábbi formulák adódnak ( $k = 1, 2, 3, \dots$ ):

$$\begin{aligned} c_k &= \frac{a_k + b_k}{2} - \delta = a_k + \frac{L_k}{2} - \delta, \\ d_k &= \frac{a_k + b_k}{2} + \delta = a_k + \frac{L_k}{2} + \delta. \end{aligned}$$

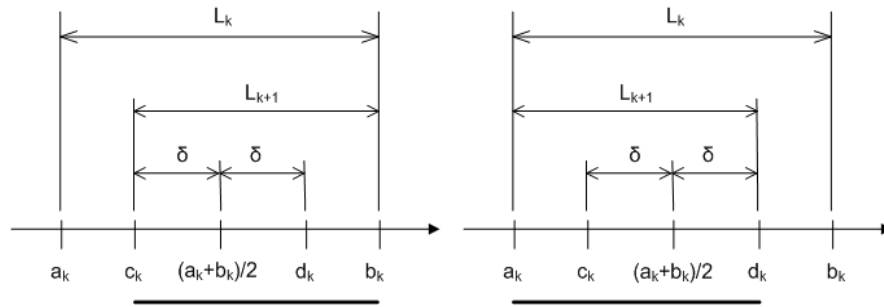
Az új bizonytalansági intervallum számítása:

Ha  $f(c_k) > f(d_k)$ , akkor az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [c_k, b_k]$ ,

ha  $f(c_k) \leq f(d_k)$ , akkor az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [a_k, d_k]$ .

Nyilvánvaló, hogy mindkét esetben  $L_{k+1} = \frac{L_k}{2} + \delta$ .

Az alábbi ábra szemlélteti a fent leírtakat:



Az eljárást addig folytatjuk, amíg valamely  $k$ -ra  $L_k < 2\varepsilon$ , ahol  $\varepsilon > 0$  a pontossági előírás. Ha a minimumpontot a megállásnál kapott  $[a_k, b_k]$  intervallum középpontjára választjuk, akkor  $\varepsilon$ -nál kisebb hibát követünk el a minimumpont közelítésében.

A következőkben egy összefüggést adunk az  $L_{k+1}$  és az első intervallumbeli  $L_1$  között. Azt tudjuk, hogy  $L_{k+1} = \frac{L_k}{2} + \delta$ , ezt felhasználva:

$$\begin{aligned}
 L_2 &= \frac{L_1}{2} + \delta \\
 L_3 &= \frac{L_2}{2} + \delta = \frac{\frac{L_1}{2} + \delta}{2} + \delta = \frac{L_1}{2^2} + \frac{\delta}{2} + \delta \\
 L_4 &= \frac{L_3}{2} + \delta = \frac{\frac{L_1}{2^2} + \frac{\delta}{2} + \delta}{2} + \delta = \frac{L_1}{2^3} + \frac{\delta}{2^2} + \frac{\delta}{2} + \delta \\
 &\dots \\
 L_{k+1} &= \frac{L_1}{2^k} + \frac{\delta}{2^{k-1}} + \frac{\delta}{2^{k-2}} + \dots + \frac{\delta}{2} + \delta = \\
 &= \frac{L_1}{2^k} + \delta \left( \frac{\left(\frac{1}{2}\right)^k - 1}{\frac{1}{2} - 1} \right) = \frac{L_1}{2^k} + 2\delta (2^k - 1)
 \end{aligned}$$

Az utolsó összefüggésnél felhasználtuk azt, hogy ott egy  $\frac{1}{2}$  hányadosú mértani sorozat  $k$  darab tagjának összege szerepel. A képlet segítségével előre megállapíthatjuk a lépésszámot egy adott  $\varepsilon$  pontossági előíráshoz. A  $\frac{L_1}{2^{k-1}} + 2\delta (2^{k-1} - 1) < 2\varepsilon$  egyenlőtlenségből kellene  $k$ -t kifejezni, amely nem bonyolult, ha a  $\tilde{k} = 2^{k-1}$  helyettesítéssel élünk, mert  $\tilde{k}$ -re egy másodfokú egyenlőtlenség adódik, utána pedig a  $k$  is egyszerűen meghatározható.

#### Példa:

Oldjuk meg az  $f(x) = x^2 - 5x + 10 \rightarrow \min!$  egyváltozós optimalizálási feladatot dichotomous módszerrel, ha a biztonalansági intervallum  $[a, b] = [1, 5]$ ,  $\delta = 0.2$ ,  $\varepsilon = 0.5$ .

#### Megoldás:

1. lépés:

Az első közelítésbeli bizonytalansági intervallum  $[a_1, b_1] = [1, 5]$ , melynek hossza  $L_1 = 4$ . A két közbülső pont és a hozzájuk tartozó függvényértékek:

$$\begin{aligned}
 c_1 &= a_1 + \frac{L_1}{2} - \delta = 2.8, & f(c_1) &= 3.84, \\
 d_1 &= a_1 + \frac{L_1}{2} + \delta = 3.2, & f(d_1) &= 4.24
 \end{aligned}$$

2. lépés:

Mivel  $f(c_1) < f(d_1)$ , így az új bizonytalansági intervallum  $[a_2, b_2] = [a_1, d_1] = [1, 3.2]$ , melynek hossza  $L_2 = 2.2$ . A két közbülső pont és a hozzájuk tartozó függvényértékek:

$$\begin{aligned} c_2 &= 1.9, & f(c_2) &= 4.11, \\ d_2 &= 2.3, & f(d_2) &= 3.79 \end{aligned}$$

3. lépés:

Mivel  $f(c_2) > f(d_2)$ , így az új bizonytalansági intervallum  $[a_3, b_3] = [c_2, b_2] = [1.9, 3.2]$ , melynek hossza  $L_3 = 1.3$ . A két közbülső pont és a hozzájuk tartozó függvényértékek:

$$\begin{aligned} c_3 &= 2.35, & f(c_3) &= 3.7725, \\ d_3 &= 2.75, & f(d_3) &= 3.8125 \end{aligned}$$

4. lépés:

Mivel  $f(c_3) < f(d_3)$ , így az új bizonytalansági intervallum  $[a_4, b_4] = [a_3, d_3] = [1.9, 2.75]$ , melynek hossza  $L_4 = 0.85$ . Itt megállunk, mivel  $L_4 < 2\varepsilon$ . A minimumpont közelítése  $x_{\min} = \frac{a_4+b_4}{2} \approx 2.325$ . Általában kisebb  $\varepsilon$  értéket szokás megadni, de ezzel is elég közel kerültünk a pontos minimumhelyhez, az  $\bar{x} = 2.5$  értékhez.

**Golden section (aranymetszés) keresés** A másik, szintén nem bonyolult megoldás arra, hogy az új intervallumok mindkét esetben azonos hosszúságúak legyenek a következő. Válasszuk meg a  $c, d$  pontokat úgy, hogy a  $[c, b]$  és az  $[a, d]$  intervallum hossza az intervallum  $\phi$ -szerese ( $1/2 < \phi < 1$ ) legyen, azaz

$$\begin{aligned} b - c &= \phi L \\ d - a &= \phi L \end{aligned}$$

Ekkor az alábbi képletek határozzák meg az osztópontokat az egyes lépésekben

$$\begin{aligned} c_k &= b_k - \phi L_k = a_k + (1 - \phi)L_k, \\ d_k &= a_k + \phi L_k. \end{aligned}$$

Természetesen használhatjuk az alábbi képleteket is

$$\begin{aligned} c_k &= a_k + (1 - \phi)L_k = b_k - \phi L_k, \\ d_k &= b_k - (1 - \phi)L_k. \end{aligned}$$

Nyilvánvaló, hogy mindkét esetben  $L_{k+1} = \phi L_k$ .

Az új bizonytalansági intervallum számítása:

Ha  $f(c_k) > f(d_k)$ , akkor az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [c_k, b_k]$ ,

ha  $f(c_k) \leq f(d_k)$ , akkor az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [a_k, d_k]$ .

A  $\phi$  értéke az  $(1/2, 1)$  intervallumban szabadon megválasztható. Mivel az eljárás legidőigényesebb része a függvényérték kiszámítása, ezért célszerű a  $\phi$  szorzót úgy megválasztani, hogy az új bizonytalansági intervallum egyik belső pontja megegyezzen az előző bizonytalansági intervallum egyik belső pontjával, azaz ha előírjuk, hogy az  $[a_{k+1}, b_{k+1}] = [c_k, b_k]$  intervallumban

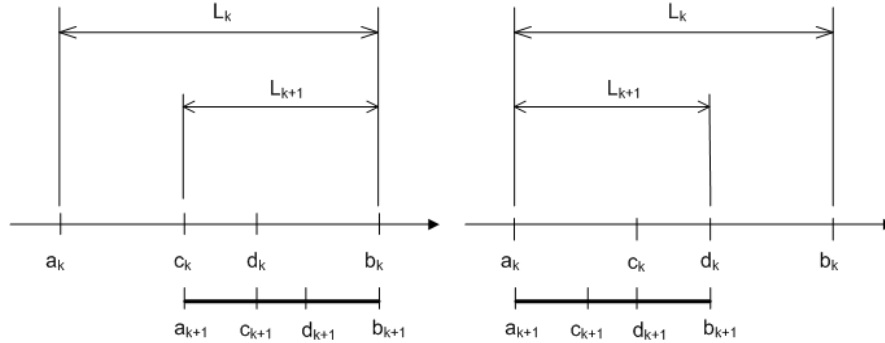
$$c_{k+1} = d_k,$$

ill. az  $[a_{k+1}, b_{k+1}] = [a_k, d_k]$  intervallumban pedig

$$d_{k+1} = c_k$$

teljesüljön, akkor ezekben a pontokban nem kell újra függvényt számolni.

Az alábbi ábra szemlélteti a fent leírtakat:



Most vizsgáljuk meg, hogy milyen  $\phi$  szorzó esetén lehet a fenti jó tulajdonságot biztosítani.

Az  $[a_{k+1}, b_{k+1}] = [c_k, b_k]$  új bizonytalansági intervallum esetén

$$\begin{aligned} c_{k+1} &= b_{k+1} - \phi L_{k+1} = b_k - \phi^2 L_k \\ d_k &= b_k - (1 - \phi)L_k, \end{aligned}$$

amelyek akkor egyeznek meg, ha  $\phi^2 = 1 - \phi$ .

Az  $[a_{k+1}, b_{k+1}] = [a_k, d_k]$  új bizonytalansági intervallum esetén

$$\begin{aligned} d_{k+1} &= a_{k+1} + \phi L_{k+1} = a_k + \phi^2 L_k \\ c_k &= a_k + (1 - \phi)L_k, \end{aligned}$$

amelyek akkor egyeznek meg, ha  $\phi^2 = 1 - \phi$ . Tehát mindkét esetben a  $\phi$  szorzóra ugyanaz a

$$\phi^2 + \phi - 1 = 0$$

másodfokú egyenlet adódik, amelynek megoldása  $\phi = \frac{\sqrt{5}-1}{2} \approx 0.618$ . Azt az eljárást, amelynél a  $\phi$  értékét ilyen módon választjuk, golden section vagy magyarul aranymetszés módszernek nevezzük. Ennél a módszernél tehát az új bizonytalansági intervallum hossza közelítőleg 62 %-a az előző bizonytalansági intervallum hosszának.

A golden section módszernél ( $\phi = 0,618$ ) az új bizonytalansági intervallum és abban lévő két pont számítása a következő:

Ha  $f(c_k) > f(d_k)$ , akkor az új bizonytalansági intervallum

$$\begin{aligned} [a_{k+1}, b_{k+1}] &= [c_k, b_k] \\ c_{k+1} &= d_k \\ d_{k+1} &= a_{k+1} + \phi L_{k+1}, \end{aligned}$$

ha  $f(c_k) \leq f(d_k)$ , akkor az új bizonytalansági intervallum

$$\begin{aligned} [a_{k+1}, b_{k+1}] &= [a_k, d_k] \\ c_{k+1} &= a_{k+1} + (1 - \phi)L_{k+1} \\ d_{k+1} &= c_k. \end{aligned}$$

Az eljárást addig folytatjuk, amíg valamely  $k$ -ra  $L_k < 2\varepsilon$ , ahol  $\varepsilon > 0$  a pontossági előírás. Ha a minimumpontot a megállásnál kapott  $[a_k, b_k]$  intervallum középpontjára választjuk, akkor  $\varepsilon$ -nál kisebb hibát követünk el a minimumpont közelítésében.

Azt tudjuk, hogy az intervallumok redukciós aránya  $\phi$ , azaz  $L_{k+1} = \phi L_k$ . A redukciós arányt  $L_{k+1}$  és az első intervallumbeli  $L_1$  között az alábbiak szerint számolhatjuk:

$$\begin{aligned} L_2 &= \phi L_1 \\ L_3 &= \phi L_2 = \phi\phi L_1 = \phi^2 L_1 \\ L_4 &= \phi L_3 = \phi\phi^2 L_1 = \phi^3 L_1 \\ &\dots \\ L_{k+1} &= \phi^k L_1 \end{aligned}$$

Az  $L_{k+1} = \phi^k L_1$  formula segítségével előre megállapíthatjuk a lépésszámot egy adott  $\varepsilon$  a pontossági előíráshoz. Ha a minimumpontot az  $[a_k, b_k]$  intervallum középpontjára választjuk, akkor  $L_k < 2\varepsilon$  összefüggésből  $k$ -ra a  $\phi^{k-1} L_1 < 2\varepsilon$  adódik. Ebből az egyenlőtlenségből  $k > 1 + \log_{\phi} \frac{2\varepsilon}{L_1}$ , így a  $k$ -t az  $(1 + \log_{\phi} \frac{2\varepsilon}{L_1})$  mennyiséghez a számegyenesen jobbra eső legközelebbi egészre kell választani. Ezzel elkerülhetjük, hogy minden lépésnél megvizsgáljuk az intervallum hosszát.

#### Példa:

Oldjuk meg az  $f(x) = x^2 - 5x + 10 \rightarrow \min!$  egyváltozós optimalizálási feladatot arany-metszés (golden section) módszerrel, ha a bizonytalansági intervallum  $[a, b] = [1, 5]$ ,  $\varepsilon = 0.5$ .

#### Megoldás:

1. lépés:

Az első közelítésbeli bizonytalansági intervallum  $[a_1, b_1] = [1, 5]$ , melynek hossza  $L_1 = 4$ . A két közbülső pont és a hozzájuk tartozó függvényértékek:

$$\begin{aligned} c_1 &= a_1 + 0.382 L_1 = 2.528, & f(c_1) &= 3.75, \\ d_1 &= a_1 + 0.618 L_1 = 3.472, & f(d_1) &= 4.69 \end{aligned}$$

2. lépés:

Mivel  $f(c_1) < f(d_1)$ , így az új bizonytalansági intervallum  $[a_2, b_2] = [a_1, d_1] = [1, 3.472]$ , melynek hossza  $L_2 = 2.472$ . A két közbülső pont és a hozzájuk tartozó függvényértékek az alábbiak. Ne feledjük, hogy most a  $d_2$  pont az előző intervallumbeli  $c_1$  értékkel azonos.

$$\begin{aligned} c_2 &= a_2 + 0.382 L_2 = 1.944, & f(c_2) &= 4.06, \\ d_2 &= c_1 = 2.528, & f(d_2) &= f(c_1) = 3.75 \end{aligned}$$

3. lépés:

Mivel  $f(c_2) > f(d_2)$ , így az új bizonytalansági intervallum  $[a_3, b_3] = [c_2, b_2] = [1.944, 3.472]$ , melynek hossza  $L_3 = 1.528$ . A két közbülső pont és a hozzájuk tartozó függvényértékek az alábbiak. Ez esetben a  $c_3$  pont lesz az előző intervallumbeli  $d_2$  értékkel azonos.

$$\begin{aligned} c_3 &= d_2 = 2.528, & f(c_3) &= f(d_2) = 3.75, \\ d_3 &= a_3 + 0.618 L_3 = 2.888, & f(d_3) &= 3.90 \end{aligned}$$

4. lépés:

Mivel  $f(c_3) < f(d_3)$ , így az új bizonytalansági intervallum  $[a_4, b_4] = [a_3, d_3] = [1.944, 2.888]$ , melynek hossza  $L_4 = 0.944$ . Itt megállunk, mivel  $L_4 < 2\varepsilon$ . A minimumpont közelítése  $x_{\min} = \frac{a_4+b_4}{2} \approx 2.416$ . Most is elég közel kerültünk a pontos minimumhelyhez, az  $\bar{x} = 2.5$  értékhez.

**Fibonacci keresés** A golden section módszernél láttuk, hogy a bizonytalansági intervallumok hossza minden lépésben azonos arányban csökken, azaz  $L_{k+1} = \phi L_k$ . A Fibonacci módszernél a redukciós arány lépésről lépésre változik, de megmarad a golden section módszernek az a jó tulajdonsága, hogy az első intervallumban kettő, a többi intervallumban pedig csak egy függvény-kiértékelésre van szükség. Az eljárás alapját a Fibonacci számok képezik, amelyeket az alábbi rekurzív formula definiál:

$$\begin{aligned} F_0 &= F_1 = 1, \\ F_k &= F_{k-1} + F_{k-2} \quad k = 2, 3, \dots \end{aligned}$$

A sorozatból az első 10 Fibonacci számot közöljük: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...

A módszer elején kiválasztunk egy  $n$  természetes számot, ennek választására később visszatérünk. A Fibonacci módszernél az  $n$  és a Fibonacci számok segítségével az alábbi formulákkal számítjuk ki az  $[a_k, b_k]$  intervallum belsejében a  $c_k, d_k$  pontokat:

$$\begin{aligned} c_k &= a_k + \frac{F_{n-k-1}}{F_{n-k+1}} L_k, \\ d_k &= a_k + \frac{F_{n-k}}{F_{n-k+1}} L_k. \end{aligned}$$

A képletek könnyen megjegyezhetők, mivel mindkét belső pont számítása azonos elveken nyugszik, a baloldali végponthoz hozzá kell adni a bizonytalansági intervallum hosszának valahányszorosát és ez a szorzó két Fibonacci szám hányadosa. A hányadosokban a nevező mindkét esetben azonos, a számláló pedig  $d_k$  esetén a nevezőt közvetlenül megelőző Fibonacci szám,  $c_k$  esetén pedig a nevezőt kettővel megelőző Fibonacci szám. A  $k = 1$  esetben a nevező  $F_n$ , a többi  $k$  esetén pedig mindig eggyel kisebb indexűek a képletekben szereplő Fibonacci számok.

Az alábbiakban három állítást közlünk, amelyek a Fibonacci módszer fő tulajdonságait mutatják.

### 1. ÁLLÍTÁS

Akár a baloldali, akár a jobboldali intervallum lesz az új bizonytalansági intervallum, mindkét esetben

$$L_{k+1} = \frac{F_{n-k}}{F_{n-k+1}} L_k,$$

és az intervallumok hosszára igaz a Fibonacci számok képzéséhez formailag hasonló összefüggés

$$L_k = L_{k+1} + L_{k+2}, \quad k = 1, 2, \dots, n-3.$$

A bizonytalansági intervallumok hosszának csökkenése minden lépésben a Fibonacci számok hányadosával arányos.

## 2. ÁLLÍTÁS

Akár a baloldali, akár a jobboldali intervallum lesz az új bizonytalansági intervallum, mindkét esetben az új bizonytalansági intervallumbeli egyik pont megegyezik az előző bizonytalansági intervallum egyik belső pontjával, azaz ha az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [c_k, b_k]$ , akkor

$$c_{k+1} = d_k,$$

ha pedig az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [a_k, d_k]$ , akkor

$$d_{k+1} = c_k.$$

## 3. ÁLLÍTÁS

A  $k = n - 1$  esetben

$$c_{n-1} = d_{n-1}.$$

Ez az állítás azt fejezi ki, hogy egy bizonyos lépés után az algoritmus megáll, mivel a két belső pont megegyezik.

Az állítások egyszerű számolással bizonyíthatóak, amelyet az olvasóra bízunk.

Összefoglalva, a Fibonacci módszernél az új bizonytalansági intervallum és abban lévő két pont számítása a következő:

Ha  $f(c_k) > f(d_k)$ , akkor az új bizonytalansági intervallum

$$\begin{aligned} [a_{k+1}, b_{k+1}] &= [c_k, b_k] \\ c_{k+1} &= d_k \\ d_{k+1} &= a_{k+1} + \frac{F_{n-k}}{F_{n-k+1}} L_{k+1}, \end{aligned}$$

ha  $f(c_k) \leq f(d_k)$ , akkor az új bizonytalansági intervallum

$$\begin{aligned} [a_{k+1}, b_{k+1}] &= [a_k, d_k] \\ c_{k+1} &= a_{k+1} + \frac{F_{n-k-1}}{F_{n-k+1}} L_{k+1} \\ d_{k+1} &= c_k. \end{aligned}$$

Az eljárást  $k = n - 1$  lépésig folytatjuk és a 3. állításbeli  $c_{n-1} = d_{n-1}$  közös értékre választjuk a minimumpontot. Ekkor a hiba, amit elkövethetünk legfeljebb  $L_{n-1}/2$ . Könnyen ellenőrizhető, hogy a közös érték nem más, mint a  $k = n - 1$  intervallum középpontja.

Most visszatérünk az  $n$  értékének megválasztására. Az első intervallum esetén kettő, a 2. állítás miatt pedig a további intervallumokban csak egy függvény-kiértékelés szükséges. Ha az utolsó, azaz a  $k = n - 1$ -edik intervallumban is kiszámítjuk a függvény minimumát, akkor a függvény-kiértékelések száma  $n$ , tehát az előre nem definiált  $n$  a függvény-kiértékelések számát jelenti. Ha előírjuk, hogy a választott minimumpontnak legfeljebb  $\varepsilon > 0$  legyen a hibája, azaz

$$\frac{L_{n-1}}{2} < \varepsilon,$$



akkor az  $n$  értéke ennek a formulának a segítségével számítható. A képletben  $L_{n-1}$  helyére használható formulát kell írunk, amelyet az 1. állítás segítségével az alábbiak szerint végzünk el. Először számítsuk ki a redukciós arányt  $L_{k+1}$  és az első intervallumbeli  $L_1$  között, amelyet az alábbiak szerint végezhetünk

$$\begin{aligned} L_2 &= \frac{F_{n-1}}{F_n} L_1 \\ L_3 &= \frac{F_{n-2}}{F_{n-1}} L_2 = \frac{F_{n-2}}{F_{n-1}} \frac{F_{n-1}}{F_n} L_1 = \frac{F_{n-2}}{F_n} L_1 \\ L_4 &= \frac{F_{n-3}}{F_{n-2}} L_3 = \frac{F_{n-3}}{F_{n-2}} \frac{F_{n-2}}{F_n} L_1 = \frac{F_{n-3}}{F_n} L_1 \\ &\dots \\ L_{k+1} &= \frac{F_{n-k}}{F_n} L_1 \end{aligned}$$

Ezt felhasználva,  $k = n - 2$  helyettesítéssel

$$L_{n-1} = \frac{F_2}{F_n} L_1 = \frac{2}{F_n} L_1.$$

A hibára vonatkozó képlet használható formában az alábbiak szerint írható

$$\frac{L_1}{F_n} < \varepsilon,$$

amelyből az adott kezdeti intervallumhossz ( $L_1$ ) és egy előre megadott pontossági tűrés ( $\varepsilon$ ) ismertében  $n$  értékét úgy kell megválasztani, hogy

$$F_n > \frac{L_1}{\varepsilon}.$$

#### Példa:

Oldjuk meg az  $f(x) = x^2 - 5x + 10 \rightarrow \min!$  egyváltozós optimalizálási feladatot Fibonacci módszerrel, ha a bizonytalansági intervallum  $[a, b] = [1, 5]$ .

#### Megoldás:

Válasszuk meg az  $n$  értékét  $n = 4$ -re. Ennek javasolt megválasztására a példa végén visszatérünk. Az algoritmushoz szükséges Fibonacci számok:  $F_0 = F_1 = 1$ ,  $F_2 = 2$ ,  $F_3 = 3$ ,  $F_4 = 5$ .

1. lépés:

Az első közelítésbeli bizonytalansági intervallum  $[a_1, b_1] = [1, 5]$ , melynek hossza  $L_1 = 4$ . A két közbülső pont és a hozzájuk tartozó függvényértékek:

$$\begin{aligned} c_1 &= a_1 + \frac{F_2}{F_4} L_1 = 1 + \frac{2}{5} \cdot 4 = 2.6, & f(c_1) &= 3.76, \\ d_1 &= a_1 + \frac{F_3}{F_4} L_1 = 1 + \frac{3}{5} \cdot 4 = 3.4, & f(d_1) &= 4.56 \end{aligned}$$

Ahogy korábban említettük, az első intervallumbeli közbülső pontok számításánál a nevezőben mindig  $F_n$  áll, a számlálóban pedig attól kettővel ill. eggyel kisebb indexű Fibonacci szám áll. A további lépésekben a Fibonacci számok indexe mindig eggyel csökken.

2. lépés:

Mivel  $f(c_1) < f(d_1)$ , így az új bizonytalansági intervallum  $[a_2, b_2] = [a_1, d_1] = [1, 3.4]$ , melynek hossza  $L_2 = 2.4$ . A két közbülső pont és a hozzájuk tartozó függvényértékek az alábbiak. Ne feledjük, hogy most a  $d_2$  pont az előző intervallumbeli  $c_1$  értékkel azonos.

$$\begin{aligned} c_2 &= a_2 + \frac{F_1}{F_3} L_2 = 1.8, & f(c_2) &= 4.24, \\ d_2 &= c_1 = 2.6, & f(d_2) &= f(c_1) = 3.76 \end{aligned}$$

3. lépés:

Mivel  $f(c_2) > f(d_2)$ , így az új bizonytalansági intervallum  $[a_3, b_3] = [c_2, b_2] = [1.8, 3.4]$ , melynek hossza  $L_3 = 1.6$ . A két közbülső pont és a hozzájuk tartozó függvényértékek az alábbiak. Ez esetben a  $c_3$  pont lesz az előző intervallumbeli  $d_2$  értékkel azonos.

$$\begin{aligned} c_3 &= d_2 = 2.6, & f(c_3) &= f(d_2) = 3.76, \\ d_3 &= a_3 + \frac{F_1}{F_2} L_3 = 2.6, & f(d_3) &= 3.76 \end{aligned}$$

4. lépés:

Mivel  $c_3 = d_3$ , így az algoritmust be kell fejezni. A minimumpont közelítése  $x_{\min} \approx 2.6$ , vagyis a közbülső pontok közös értéke.

Ha az  $n$  értékét önkényesen választjuk meg, akkor - mint ahogy a példában tapasztaltuk - nem tudjuk befolyásolni a pontosságot. Amennyiben azt szeretnénk, hogy a pontossági tűrés például  $\varepsilon = 0.05$  legyen, akkor az  $n$  értékét olyanra kell választani, hogy fennálljon az

$$F_n > \frac{L_1}{\varepsilon}.$$

összefüggés, ez pedig azt jelenti példánkban, hogy  $F_n > \frac{4}{0.05} = 80$ . A Fibonacci számokból adódik, hogy  $n = 10$ , mivel  $F_{10} = 87$ .

A példa során azt is ellenőrizhetjük, hogy az intervallumok hosszára a Fibonacci számokhoz hasonló összefüggés érvényes,  $L_1 = L_2 + L_3$ .

### 3.1.3. Deriváltakat használó egyváltozós kereső eljárások

**Bisection módszer** Tegyük fel, hogy az egyváltozós függvény, amelynek minimumát keressük, differenciálható konvex függvény. Induljunk ki egy olyan  $[a, b]$  bizonytalansági intervallumból, amely tartalmazza a minimumpontot. A módszer lényege, hogy minden lépésben az intervallum belsejében választunk egy pontot (optimális választás a felezőpont) és a pontbeli derivált ( $f'$ ) értékétől függően választjuk ki a baloldali vagy a jobboldali intervallumot.

**A Bisection módszer algoritmus:**

**Induló lépésben** legyen  $[a, b] = [a_1, b_1]$ , ( $k = 1$ ).

**Közbülső lépés:** Adott az  $[a_k, b_k]$  intervallum. Meghatározzuk a felezőpontot, jelölje ezt  $c_k$ .

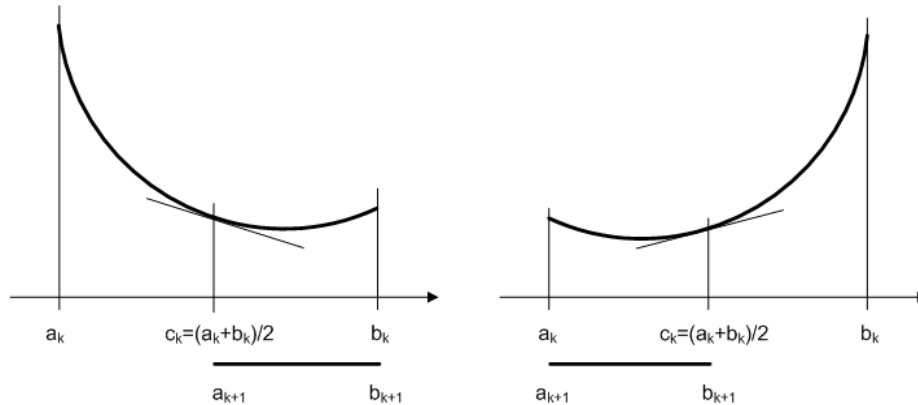
Ha  $f'(c_k) = 0$ , akkor a függvény konvexitása miatt a  $c_k$  pont a minimumpont.

Ha  $f'(c_k) < 0$ , akkor a függvény konvexitása miatt az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [c_k, b_k]$ .

Ha  $f'(c_k) > 0$ , akkor a függvény konvexitása miatt az új bizonytalansági intervallum  $[a_{k+1}, b_{k+1}] = [a_k, c_k]$ .

Az eljárást addig folytatjuk, amíg valamely  $k$ -ra  $L_k = b_k - a_k < 2\varepsilon$ , ahol  $\varepsilon > 0$  a pontossági előírás. Ekkor, ha a minimumpontot a megállásnál kapott  $[a_k, b_k]$  intervallum középpontjára választjuk, akkor  $\varepsilon$ -nál kisebb hibát követünk el a minimumpont közelítésében.

Az algoritmus lépései az alábbi ábrán is követhetők:



Ha az intervallum közepére választjuk a pontot, akkor  $L_{k+1} = \frac{L_k}{2}$ , azaz feleződik a bizonytalansági intervallum. Ezért ezt a módszert intervallumfelező eljárásnak is nevezzük.

#### Példa:

Oldjuk meg az  $f(x) = x^3 - 9x + 7 \rightarrow \min!$  egyváltozós optimalizálási feladatot Bisection módszerrel, ha a bizonytalansági intervallum  $[a, b] = [1, 3]$ ,  $\varepsilon = 0.005$ .

#### Megoldás:

Előkészületként határozzuk meg a függvény első deriváltját:

$$f'(x) = 3x^2 - 9$$

#### 1. lépés:

Az első közelítésbeli bizonytalansági intervallum  $[a_1, b_1] = [1, 5]$ , melynek hossza  $L_1 = 2$ , középpontja (felezőpontja)  $c_1 = 2$ .

#### 2. lépés:

Mivel  $f'(2) = 3 > 0$ , így az új bizonytalansági intervallum  $[a_2, b_2] = [a_1, c_1] = [1, 2]$ , melynek hossza  $L_2 = 1$ , felezőpontja  $c_2 = 1.5$ .

#### 3. lépés:

Mivel  $f'(1.5) = -2.25 < 0$ , így az új bizonytalansági intervallum  $[a_3, b_3] = [c_2, b_2] = [1.5, 2]$ , melynek hossza  $L_3 = 0.5$ , felezőpontja  $c_3 = 1.75$ .

#### 4. lépés:

Mivel  $f'(1.75) = 0.1875 > 0$ , így az új bizonytalansági intervallum  $[a_4, b_4] = [a_3, c_3] = [1.5, 1.75]$ , melynek hossza  $L_4 = 0.25$ , felezőpontja  $c_4 = 1.625$ . Itt megállunk, mivel az algoritmus fő lépéseit bemutattuk. Egyébként addig folytatjuk az eljárást, amíg el nem érjük azt a lépést, amelyben  $L_k < 2\varepsilon = 0.01$ , és ekkor  $x_{\min} = c_k$  lesz a minimumpont  $\varepsilon$  pontosságú közelítése.

Megfigyelhető, hogy a bizonytalansági intervallumok hossza minden lépésben a felére csökken.

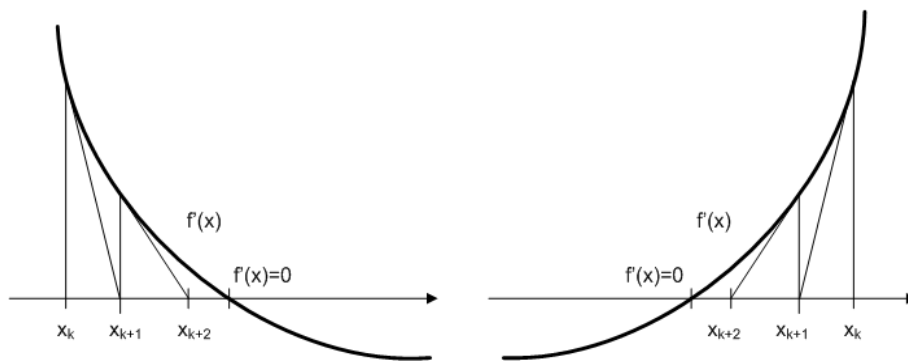
**Newton módszer** Tegyük fel, hogy az egyváltozós függvény, amelynek minimumát keressük, kétszer differenciálható függvény. Induljunk ki egy  $x_1$  pontból és alkalmazzuk az  $f'(x) = 0$  szükséges feltételre, mint stacionárius egyenlet megoldására a Newton iterációs eljárást, amely a következő

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \quad k = 1, 2, \dots$$

A Newton módszerről jelen tananyag elején részletesen beszéltünk, igaz, hogy a többváltozós esetet tárgyaltuk. A fenti formulát egyszerűen kapjuk a többváltozós esetből, hiszen a gradiens vektor helyett az első deriváltat, a Hesse mátrix helyett pedig a második deriváltat kell szerepeltetni.

Az eljárást addig folytatjuk, amíg valamely  $k$ -ra  $|x_{k+1} - x_k| < \varepsilon$ , ahol  $\varepsilon > 0$  a pontossági előírás. Ekkor az  $x_{k+1}$  pont legfeljebb  $\varepsilon$  számmal tér el a minimumponttól.

Az algoritmus lépései az alábbi ábrán is követhetők:



Az egyváltozós esetben az  $f'(x)$  deriváltfüggvénynek az  $x_k$  pontjában érintőt húzunk és ahol ez metszi a vízszintes tengelyt az lesz az új  $x_{k+1}$  közelítés. Tehát minden lépésben az  $f'(x)$  deriváltfüggvényt az érintőegyenessel közelítjük. A Newton módszert szokás érintőmódszernek is nevezni.

#### Példa:

Oldjuk meg az  $f(x) = x^3 - 9x + 7 \rightarrow \min!$  egyváltozós optimalizálási feladatot Newton módszerrel, ha a kiinduló közelítő megoldás  $x_1 = 3$ ,  $\varepsilon = 0.005$ .

#### Megoldás:

Előkészületként határozzuk meg a függvény első és második deriváltját

$$f'(x) = 3x^2 - 9, \quad f''(x) = 6x$$

2. közelítés:

$$x_2 = x_1 - \frac{3x_1^2 - 9}{6x_1} = 2$$

Az utolsó két közelítés távolsága:  $|x_2 - x_1| = 1$ . Hasonló lépésekkel nyerjük a további közelítéseket, amelyek rendre az alábbiak:

3. közelítés:  $x_3 = 1.7500$ ,  $|x_3 - x_2| = 0.25$
4. közelítés:  $x_4 = 1.7321$ ,  $|x_4 - x_3| = 0.0179$
5. közelítés:  $x_5 = 1.7321$ ,  $|x_5 - x_4| = 0$

Láthatjuk a gyors konvergenciát, már a 4. közelítésben megkaptuk a  $x_{\min} = \sqrt{3}$  pontos értéket négy tizedes pontossággal.

### 3.2. Többváltozós vonalmenti kereső eljárások

Ebben a fejezetben részletesen ismertetjük a többváltozós vonalmenti eljárásokat, az első alfejezetben a deriváltakat nem használó, a második alfejezetben pedig a deriváltakat is használó módszereket mutatjuk be.

#### 3.2.1. Deriváltakat nem használó többváltozós vonalmenti kereső eljárások

**Ciklikus koordináta módszer** A ciklikus koordináta módszerben a kereső irányok az egységvektorok.

**Ciklikus koordináta módszer algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy  $\mathbf{x}_k \in \mathbb{R}^n$  pontból.

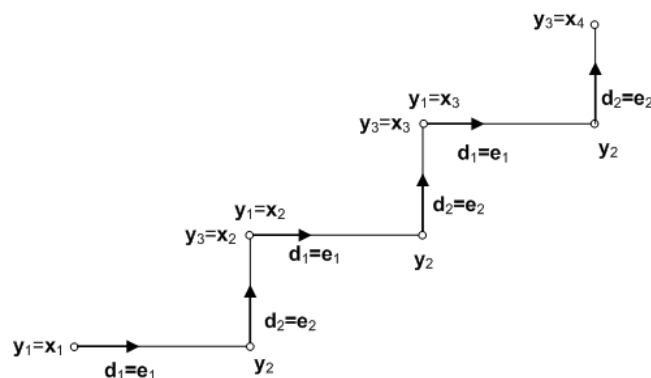
**Közbülső lépés:**

1. Legyen  $\mathbf{y}_1 = \mathbf{x}_k$ . Az  $\mathbf{y}_1$  pontból kiindulva az  $\mathbf{e}_1$  egységvektor irányában megkeressük a  $\varphi(\lambda) = f(\mathbf{y}_1 + \lambda \mathbf{e}_1)$  függvény minimumát, legyen ez az  $\mathbf{y}_2$  pont, majd az  $\mathbf{y}_2$  pontból kiindulva az  $\mathbf{e}_2$  egységvektor irányában megkeressük a  $\varphi(\lambda) = f(\mathbf{y}_2 + \lambda \mathbf{e}_2)$  függvény minimumát, legyen ez az  $\mathbf{y}_3$  pont, ezt folytatva, azaz **ciklikusan** változtatva a kereső egységvektorokat, utolsóként az  $\mathbf{e}_n$  egységvektor irányában történő keresés után kapjuk az  $\mathbf{y}_{n+1}$  pontot.

2. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{y}_{n+1}$ .

3. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást,  $k := k + 1$ .

Az algoritmus lépései az alábbi ábrán is követhetők:



**Példa:**

Oldjuk meg az alábbi optimalizálási feladatot ciklikus koordináta módszerrel! Legyen az  $\mathbf{x}_1 = (0, 0)$  a startvektor és legyen  $\varepsilon = 0.005$  a pontossági tűrés.

$$f(x_1, x_2) = x_1^2 - x_1 x_2 + 2x_2^2 - 2x_1 - 3x_2 + 7 \rightarrow \min!$$

**Megoldás:**

1. lépés:

$$\mathbf{x}_1 = (0, 0)$$

$$\mathbf{y}_1 = \mathbf{x}_1 = (0, 0), \quad \mathbf{d}_1 = \mathbf{e}_1 = (1, 0), \quad \mathbf{y} = \mathbf{y}_1 + \lambda \mathbf{d}_1 = (\lambda, 0)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(\lambda, 0) = \lambda^2 - \lambda \cdot 0 + 2 \cdot 0^2 - 2\lambda - 3 \cdot 0 + 7 = \lambda^2 - 2\lambda + 7 \rightarrow \min!$$

$$\lambda_1 = 1$$

$$\mathbf{y}_2 = \mathbf{y}_1 + \lambda_1 \mathbf{d}_1 = (\lambda_1, 0) = (1, 0), \quad \mathbf{d}_2 = \mathbf{e}_2 = (0, 1), \quad \mathbf{y} = \mathbf{y}_2 + \lambda \mathbf{d}_2 = (1, \lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1, \lambda) = 2\lambda^2 - 4\lambda + 6 \rightarrow \min!$$

$$\lambda_2 = 1$$

$$\mathbf{y}_3 = \mathbf{y}_2 + \lambda_2 \mathbf{d}_2 = (1, 1)$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{y}_3 = (1, 1)$$

$\|\mathbf{x}_2 - \mathbf{x}_1\| = \sqrt{2} > \varepsilon$ , tehát folytatni kell az eljárást.

$$\mathbf{y}_1 = \mathbf{x}_2 = (1, 1), \quad \mathbf{d}_1 = \mathbf{e}_1 = (1, 0), \quad \mathbf{y} = \mathbf{y}_1 + \lambda \mathbf{d}_1 = (1 + \lambda, 1)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 + \lambda, 1) = (\lambda + 1)^2 - 3\lambda + 3 \rightarrow \min!$$

$$\lambda_1 = \frac{1}{2}$$

$$\mathbf{y}_2 = \mathbf{y}_1 + \lambda_1 \mathbf{d}_1 = \left(\frac{3}{2}, 1\right), \quad \mathbf{d}_2 = \mathbf{e}_2 = (0, 1), \quad \mathbf{y} = \mathbf{y}_2 + \lambda \mathbf{d}_2 = \left(\frac{3}{2}, 1 + \lambda\right)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f\left(\frac{3}{2}, 1 + \lambda\right) = 2(\lambda + 1)^2 - \frac{9}{2}\lambda + \frac{7}{4} \rightarrow \min!$$

$$\lambda_2 = \frac{1}{8}$$

$$\mathbf{y}_3 = \mathbf{y}_2 + \lambda_2 \mathbf{d}_2 = \left(\frac{3}{2}, \frac{9}{8}\right)$$

3. lépés:

$$\mathbf{x}_3 = \mathbf{y}_3 = \left(\frac{3}{2}, \frac{9}{8}\right)$$

$\|\mathbf{x}_3 - \mathbf{x}_2\| = 0.177 > \varepsilon$ , tehát folytatni kell az eljárást.

Mivel az algoritmus fő lépéseit bemutattuk, ezért az eljárást befejezzük. Megjegyezzük, hogy az optimális megoldás  $\bar{\mathbf{x}} = \left[\frac{11}{7}, \frac{8}{7}\right] = (1.57, 1.14)$ , amelyhez már az  $\mathbf{x}_3 = \left(\frac{3}{2}, \frac{9}{8}\right) = (1.5, 1.125)$  közelítés elég közel van.

**Hooke-Jeeves módszer** A Hooke-Jeeves a módszer hasonló a ciklikus koordináta módszerhez, mivel itt is az egységvektorokat használjuk kereső irányoknak. A különbség az, hogy az egységvektorok irányába történő keresési ciklus után elvégzünk egy  $\mathbf{x}_{k+1} - \mathbf{x}_k$  irányban történő keresést is.

**Hooke-Jeeves módszer algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy  $\mathbf{x}_k \in \mathbb{R}^n$  pontból.

**Közbülső lépés:**

1. Legyen  $\mathbf{y}_1 = \mathbf{x}_k$ . Az  $\mathbf{y}_1$  pontból kiindulva az  $\mathbf{e}_1$  egységvektor irányában megkeressük a  $\varphi(\lambda) = f(\mathbf{y}_1 + \lambda \mathbf{e}_1)$  függvény minimumát, legyen ez az  $\mathbf{y}_2$  pont, majd az  $\mathbf{y}_2$  pontból kiindulva az  $\mathbf{e}_2$  egységvektor irányában megkeressük a  $\varphi(\lambda) = f(\mathbf{y}_2 + \lambda \mathbf{e}_2)$  függvény minimumát, legyen ez az  $\mathbf{y}_3$  pont, ezt folytatva, azaz ciklikusan változtatva a kereső egységvektorokat, utolsóként az  $\mathbf{e}_n$  egységvektor irányában történő keresés után kapjuk az  $\mathbf{y}_{n+1}$  pontot.

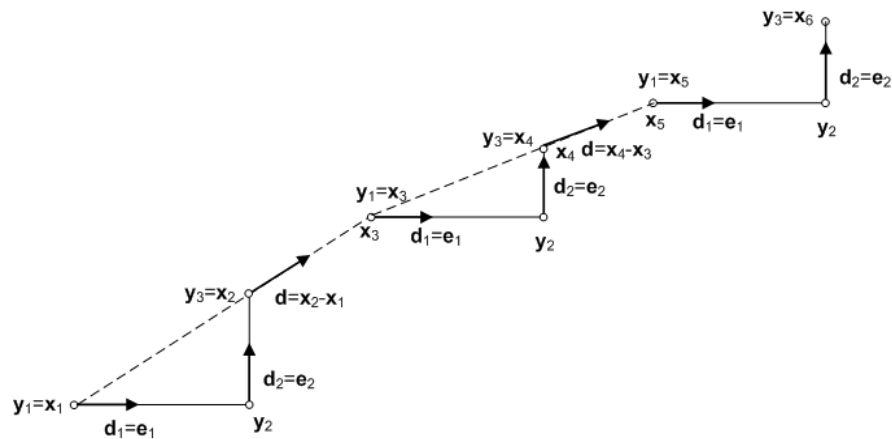
2. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{y}_{n+1}$ .

3. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást a következő sorral.

4. Legyen  $\mathbf{d} = \mathbf{x}_{k+1} - \mathbf{x}_k$  és az  $\mathbf{x}_{k+1}$  pontból kiindulva a  $\mathbf{d}$  irányban megkeressük a  $\varphi(\lambda) = f(\mathbf{x}_{k+1} + \lambda \mathbf{d})$  függvény minimumát, legyen ez az  $\mathbf{x}_{k+2}$  pont.

5. Folytatjuk az eljárást,  $k := k + 1$ .

Az algoritmus lépései az alábbi ábrán is követhetők:

**Példa:**

Oldjuk meg az alábbi optimalizálási feladatot Hooke-Jeeves módszerrel! Legyen az  $\mathbf{x}_1 = (0, 0)$  a startvektor és legyen  $\varepsilon = 0.005$  a pontossági tűrés.

$$f(x_1, x_2) = x_1^2 - x_1 x_2 + 2x_2^2 - 2x_1 - 3x_2 + 7 \rightarrow \min!$$

**Megoldás:**

1. lépés:

$$\mathbf{x}_1 = (0, 0)$$

$$\mathbf{y}_1 = \mathbf{x}_1 = (0, 0), \quad \mathbf{d}_1 = \mathbf{e}_1 = (1, 0), \quad \mathbf{y} = \mathbf{y}_1 + \lambda \mathbf{d}_1 = (\lambda, 0)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(\lambda, 0) = \lambda^2 - \lambda(0) + 2(0)^2 - 2\lambda - 3(0) = \lambda^2 - 2\lambda + 7 \rightarrow \min!$$

$$\lambda_1 = 1$$

$$\mathbf{y}_2 = \mathbf{y}_1 + \lambda_1 \mathbf{d}_1 = (\lambda_1, 0) = (1, 0), \quad \mathbf{d}_2 = \mathbf{e}_2 = (0, 1), \quad \mathbf{y} = \mathbf{y}_2 + \lambda \mathbf{d}_2 = (1, \lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1, \lambda) = 2\lambda^2 - 4\lambda + 6 \rightarrow \min!$$

$$\lambda_2 = 1$$

$$\mathbf{y}_3 = \mathbf{y}_2 + \lambda_2 \mathbf{d}_2 = (1, 1)$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{y}_3 = (1, 1)$$

$$\|\mathbf{x}_2 - \mathbf{x}_1\| = \sqrt{2} > \varepsilon, \text{ tehát folytatjuk az eljárást.}$$

3. lépés:

Itt térünk el a ciklikus koordináták módszertől, most az irányvektor legyen az előző két közelítő  $\mathbf{x}$  vektor különbsége, azaz  $\mathbf{d} = \mathbf{x}_2 - \mathbf{x}_1 = (1, 1) - (0, 0) = (1, 1)$

$$\mathbf{x}_2 = (1, 1), \quad \mathbf{d} = (1, 1), \quad \mathbf{x} = \mathbf{x}_2 + \lambda \mathbf{d} = (1 + \lambda, 1 + \lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 + \lambda, 1 + \lambda) = 2(\lambda + 1)^2 - 5\lambda + 2 \rightarrow \min!$$

$$\lambda = \frac{1}{4}$$

$$\mathbf{x}_3 = \mathbf{x}_2 + \lambda \mathbf{d} = \left(\frac{5}{4}, \frac{5}{4}\right)$$

4. lépés:

$\mathbf{x}_3 = \left(\frac{5}{4}, \frac{5}{4}\right)$  közelítésből kiindulva újra elvégzünk egy ciklust az egységvektorokkal.

$$\mathbf{y}_1 = \mathbf{x}_3 = \left(\frac{5}{4}, \frac{5}{4}\right), \quad \mathbf{d}_1 = \mathbf{e}_1 = (1, 0), \quad \mathbf{y} = \mathbf{y}_1 + \lambda \mathbf{d}_1 = \left(\frac{5}{4} + \lambda, \frac{5}{4}\right)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f\left(\frac{5}{4} + \lambda, \frac{5}{4}\right) = \left(\lambda + \frac{5}{4}\right)^2 - \frac{13}{4}\lambda + \frac{37}{16} \rightarrow \min!$$

$$\lambda_1 = \frac{3}{8}$$

$$\mathbf{y}_2 = \mathbf{y}_1 + \lambda_1 \mathbf{d}_1 = \left(\frac{13}{8}, \frac{5}{4}\right), \quad \mathbf{d}_2 = \mathbf{e}_2 = (0, 1), \quad \mathbf{y} = \mathbf{y}_2 + \lambda \mathbf{d}_2 = \left(\frac{13}{8}, \frac{5}{4} + \lambda\right)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f\left(\frac{13}{8}, \frac{5}{4} + \lambda\right) = 2\left(\lambda + \frac{5}{4}\right)^2 - \frac{37}{8}\lambda + \frac{39}{64} \rightarrow \min!$$

$$\lambda_2 = -\frac{3}{32}$$

$$\mathbf{y}_3 = \mathbf{y}_2 + \lambda_2 \mathbf{d}_2 = \left(\frac{13}{8}, \frac{37}{32}\right)$$

5. lépés:  $\mathbf{x}_4 = \mathbf{y}_3 = \left(\frac{13}{8}, \frac{37}{32}\right)$

$$\|\mathbf{x}_4 - \mathbf{x}_3\| = 0.53 > \varepsilon, \text{ tehát folytatni kell az eljárást.}$$

Mivel az algoritmus fő lépéseit bemutattuk, ezért az eljárást befejezzük. A folytatásban a  $\mathbf{d} = \mathbf{x}_4 - \mathbf{x}_3$  irányvektorral végzünk egy egyváltozós minimalizálást, majd újra egy ciklust végzünk az egységvektorokkal, stb. Megjegyezzük, hogy az optimális megoldás  $\bar{\mathbf{x}} = \left[\frac{11}{7}, \frac{8}{7}\right] = (1.57, 1.14)$ , amelyhez már az  $\mathbf{x}_4 = (1.625, 1.156)$  közelítés elég közel van. Közelebb jutottunk, mint a ciklikus koordináták módszerrel, igaz, hogy eggyel több egyváltozós minimalizálást végeztünk.



**Rosenbrock módszer** A Rosenbrock módszer hasonlóan indul mint a Hooke-Jeeves módszer, azonban a második ciklustól kezdve nem az egységvektorokat használjuk kereső iránynak, hanem olyan  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$  vektorokat, amelyek (mint ahogy az egységvektorok is), lineárisan függetlenek és ortogonálisak (merőlegesek) egymásra. Az első irányvektort, a  $\mathbf{d}_1$  vektort  $\mathbf{d}_1 = \mathbf{x}_{k+1} - \mathbf{x}_k$ -re választjuk, a többit pedig a Gram-Schmidt féle ortogonalizációs eljárással határozzuk meg. A Gram-Schmidt féle ortogonalizációs eljárás ismertetésére itt nem térünk ki.

#### Rosenbrock módszer algoritmus:

**Induló lépés** ( $k = 1$ ): Kiindulunk egy  $\mathbf{x}_k \in \mathbb{R}^n$  pontból és a kezdeti kereső vektorokat az egységvektorokra választjuk, azaz  $\mathbf{d}_i = \mathbf{e}_i$ ,  $i = 1, 2, \dots, n$ .

#### Közbülső lépés:

1. Legyen  $\mathbf{y}_1 = \mathbf{x}_k$ . Az  $\mathbf{y}_1$  pontból kiindulva a  $\mathbf{d}_1$  irányban megkeressük a  $\varphi(\lambda) = f(\mathbf{y}_1 + \lambda \mathbf{d}_1)$  függvény minimumát, legyen ez az  $\mathbf{y}_2$  pont, majd az  $\mathbf{y}_2$  pontból kiindulva a  $\mathbf{d}_2$  irányban megkeressük a  $\varphi(\lambda) = f(\mathbf{y}_2 + \lambda \mathbf{d}_2)$  függvény minimumát, legyen ez az  $\mathbf{y}_3$  pont, ezt folytatva, azaz ciklikusan változtatva a kereső irányvektorokat, utolsóként a  $\mathbf{d}_n$  irányban történő keresés után kapjuk az  $\mathbf{y}_{n+1}$  pontot.

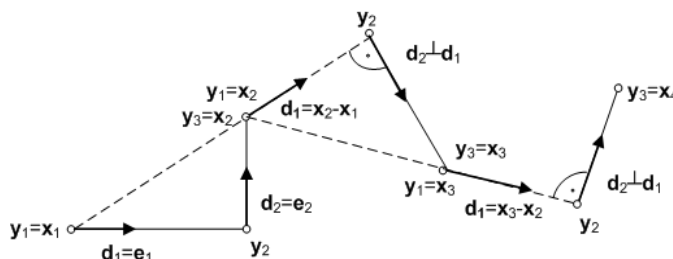
2. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{y}_{n+1}$ .

3. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást a következő sorral.

4. Meghatározzuk az új  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$  lineárisan független, egymásra merőleges irányvektorokat a Gram-Schmidt féle ortogonalizációs eljárással úgy, hogy  $\mathbf{d}_1 = \mathbf{x}_{k+1} - \mathbf{x}_k$  legyen.

5. Folytatjuk az eljárást,  $k := k + 1$ .

Az algoritmus lépései az alábbi ábrán is követhetők:



#### Példa:

Oldjuk meg az alábbi optimalizálási feladatot Rosenbrock módszerrel! Legyen az  $\mathbf{x}_1 = (0, 0)$  a startvektor és legyen  $\varepsilon = 0.005$  a pontossági tűrés.

$$f(x_1, x_2) = x_1^2 - x_1 x_2 + 2x_2^2 - 2x_1 - 3x_2 + 7 \rightarrow \min!$$

#### Megoldás:

1. lépés:

$$\mathbf{x}_1 = (0, 0)$$

$$\mathbf{y}_1 = \mathbf{x}_1 = (0, 0), \quad \mathbf{d}_1 = (1, 0), \quad \mathbf{y} = \mathbf{y}_1 + \lambda \mathbf{d}_1 = (\lambda, 0)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(\lambda, 0) = \lambda^2 - \lambda(0) + 2(0)^2 - 2\lambda - 3(0) = \lambda^2 - 2\lambda + 7 \rightarrow \min!$$

$$\lambda_1 = 1$$

$$\mathbf{y}_2 = \mathbf{y}_1 + \lambda_1 \mathbf{d}_1 = (\lambda_1, 0) = (1, 0), \quad \mathbf{d}_2 = (0, 1), \quad \mathbf{y} = \mathbf{y}_2 + \lambda \mathbf{d}_2 = (1, \lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1, \lambda) = 2\lambda^2 - 4\lambda + 6 \rightarrow \min!$$

$$\lambda_2 = 1$$

$$\mathbf{y}_3 = \mathbf{y}_2 + \lambda_2 \mathbf{d}_2 = (1, 1)$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{y}_3 = (1, 1)$$

$$\|\mathbf{x}_2 - \mathbf{x}_1\| = \sqrt{2} > \varepsilon, \text{ tehát folytatni kell az eljárást.}$$

3. lépés:

Itt térünk el a ciklikus koordináták ill. a Hooke-Jeeves módszertől, most képezzünk két irányvektort, amelyek egymásra merőlegesek, az első irányvektor legyen az előző két közelítő  $\mathbf{x}$  vektor különbsége, azaz  $\mathbf{d}_1 = \mathbf{x}_2 - \mathbf{x}_1 = (1, 1) - (0, 0) = (1, 1)$ , a másik irányvektor ( $\mathbf{d}_2$ ) pedig erre merőleges legyen. A  $\mathbf{d}_2$  vektor akkor merőleges a  $\mathbf{d}_1$  vektorra, ha  $\mathbf{d}_1 \mathbf{d}_2 = 0$ . Két ilyen vektor is van, az  $(1, -1)$  és a  $(-1, 1)$  vektorok. Válasszuk a  $\mathbf{d}_2 = (1, -1)$  vektort. Most a  $\mathbf{d}_1$  és  $\mathbf{d}_2$  két vektorok irányában végzünk egymás után egy-egy egyváltozós optimalizálást.

$$\mathbf{y}_1 = \mathbf{x}_2 = (1, 1), \quad \mathbf{d}_1 = (1, 1), \quad \mathbf{x} = \mathbf{x}_2 + \lambda \mathbf{d}_1 = (1 + \lambda, 1 + \lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 + \lambda, 1 + \lambda) = 2(\lambda + 1)^2 - 5\lambda + 2 \rightarrow \min!$$

$$\lambda_1 = \frac{1}{4}$$

$$\mathbf{y}_2 = \mathbf{y}_1 + \lambda_1 \mathbf{d}_1 = \left(\frac{5}{4}, \frac{5}{4}\right), \quad \mathbf{d}_2 = (1, -1), \quad \mathbf{y} = \mathbf{y}_2 + \lambda \mathbf{d}_2 = \left(\frac{5}{4} + \lambda, \frac{5}{4} - \lambda\right)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f\left(\frac{5}{4} + \lambda, \frac{5}{4} - \lambda\right) = \lambda + \left(\lambda + \frac{5}{4}\right)^2 + 2\left(\frac{5}{4} - \lambda\right)^2 - \left(\lambda + \frac{5}{4}\right)\left(\frac{5}{4} - \lambda\right) + \frac{3}{4} \rightarrow \min!$$

$$\lambda_2 = \frac{3}{16}$$

$$\mathbf{y}_3 = \mathbf{y}_2 + \lambda_2 \mathbf{d}_2 = \left(\frac{23}{16}, \frac{17}{16}\right)$$

4. lépés:

$$\mathbf{x}_3 = \mathbf{y}_3 = \left(\frac{23}{16}, \frac{17}{16}\right)$$

$$\|\mathbf{x}_3 - \mathbf{x}_2\| = 0.44 > \varepsilon, \text{ tehát folytatni kell az eljárást.}$$

Mivel az algoritmus fő lépéseit bemutattuk, ezért az eljárást befejezzük. A folytatáshoz meghatározzuk a két, egymásra merőleges irányvektorokat:  $\mathbf{d}_1 = \mathbf{x}_3 - \mathbf{x}_2 = \left(\frac{7}{16}, \frac{1}{16}\right)$ ,  $\mathbf{d}_2 = \left(\frac{1}{16}, -\frac{7}{16}\right)$  és ezekkel elvégzünk egy-egy egyváltozós minimalizálást, majd újra meghatározzuk a két, egymásra merőleges irányvektort, stb.

Megjegyezzük, hogy a merőleges irányvektorok előállításához nem volt szükség a Gram-Schmidt féle ortogonalizációs eljárásra, mivel kétváltozós volt a feladat.

### 3.2.2. Deriváltakat használó többváltozós vonalmenti kereső eljárások

**Gradiens módszer (Cauchy módszer, 1847)** Mint tudjuk, a gradiens vektor a függvény legnagyobb növekedésének irányába mutat. Legkézenfekvőbb tehát a  $\mathbf{d}$  kereső irányra a  $\mathbf{d} = -\nabla f(\mathbf{x})$  választás, hisz ebben az irányban várható a függvény legnagyobb csökkenése. Szokás ezt a módszert a legnagyobb csökkenés módszerének is nevezni. A módszer lényege az alábbiakban foglalható össze.

**Gradiens módszer algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy  $\mathbf{x}_k \in \mathbb{R}^n$  pontból.

**Közbülső lépés:**

1. Az irányvektor meghatározása:  $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$ .
2. Megoldjuk a  $\min \{f(\mathbf{x}_k + \lambda \mathbf{d}_k) : \lambda \geq 0\}$  minimalizálási feladatot  $\lambda$ -ra, legyen az optimális megoldás  $\lambda_k$ .
3. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$ .
4. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást,  $k := k + 1$ .

**Példa:**

Oldjuk meg az alábbi optimalizálási feladatot gradiens módszerrel. Legyen az  $\mathbf{x}_1 = (0, 0)$  a startvektor és legyen  $\varepsilon = 0.005$  a pontossági tűrés.

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2 - x_1 + x_2 + 5 \rightarrow \min!$$

**Megoldás:**

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 2x_2 - 1 \\ 2x_1 + 4x_2 + 1 \end{bmatrix}.$$

1. lépés:  $\mathbf{x}_1 = (0, 0)$

$$\nabla f(\mathbf{x}_1) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mathbf{d}_1 = -\nabla f(\mathbf{x}_1) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\mathbf{x} = \mathbf{x}_1 + \lambda \mathbf{d}_1 = (\lambda, -\lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(\lambda, -\lambda) = \lambda^2 - 2\lambda + 5 \rightarrow \min!$$

$$\lambda_1 = 1$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1 \mathbf{d}_1 = (\lambda_1, -\lambda_1) = (1, -1), \quad \mathbf{d}_2 = -\nabla f(\mathbf{x}_2) = (1, 1), \quad \mathbf{x} = \mathbf{x}_2 + \lambda \mathbf{d}_2 = (1 + \lambda, -1 + \lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 + \lambda, -1 + \lambda) = 5\lambda^2 - 2\lambda + 4 \rightarrow \min!$$

$$\lambda_2 = \frac{1}{5}$$

3. lépés:

$$\mathbf{x}_3 = \mathbf{x}_2 + \lambda_2 \mathbf{d}_2 = \left(\frac{6}{5}, -\frac{4}{5}\right), \quad \mathbf{d}_3 = -\nabla f(\mathbf{x}_3) = \left(\frac{1}{5}, -\frac{1}{5}\right), \quad \mathbf{x} = \mathbf{x}_3 + \lambda \mathbf{d}_3 = \left(\frac{6}{5} + \frac{1}{5}\lambda, -\frac{4}{5} - \frac{1}{5}\lambda\right)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f\left(\frac{6}{5} + \frac{1}{5}\lambda, -\frac{4}{5} - \frac{1}{5}\lambda\right) = \frac{1}{25}\lambda^2 - \frac{2}{25}\lambda + \frac{19}{5} \rightarrow \min!$$

$$\lambda_3 = 1$$

4. lépés:

$$\mathbf{x}_4 = \mathbf{x}_3 + \lambda_3 \mathbf{d}_3 = \left(\frac{7}{5}, -1\right), \quad \mathbf{d}_4 = -\nabla f(\mathbf{x}_4) = \left(\frac{1}{5}, \frac{1}{5}\right), \quad \mathbf{x} = \mathbf{x}_4 + \lambda \mathbf{d}_4 = \left(\frac{7}{5} + \frac{1}{5}\lambda, -1 + \frac{1}{5}\lambda\right)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f\left(\frac{7}{5} + \frac{1}{5}\lambda, -1 + \frac{1}{5}\lambda\right) = \frac{1}{5}\lambda^2 - \frac{2}{25}\lambda + \frac{94}{25} \rightarrow \min!$$

$$\lambda_4 = \frac{1}{5}$$

5. lépés:

$$\mathbf{x}_5 = \mathbf{x}_4 + \lambda_4 \mathbf{d}_4 = \left(\frac{36}{25}, -\frac{24}{25}\right), \quad \mathbf{d}_5 = -\nabla f(\mathbf{x}_4) = \left(\frac{1}{25}, -\frac{1}{25}\right), \quad \mathbf{x} = \mathbf{x}_5 + \lambda \mathbf{d}_5 = \left(\frac{36}{25} + \frac{1}{25}\lambda, -\frac{24}{25} - \frac{1}{25}\lambda\right)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f\left(\frac{36}{25} + \frac{1}{25}\lambda, -\frac{24}{25} - \frac{1}{25}\lambda\right) = \frac{1}{625}\lambda^2 - \frac{2}{625}\lambda + \frac{469}{125} \rightarrow \min!$$

$$\lambda_5 = 1$$

6. lépés:

$$\mathbf{x}_6 = \mathbf{x}_5 + \lambda_5 \mathbf{d}_5 = \left(\frac{37}{25}, -1\right) = (1.48, -1).$$

Mivel az algoritmus fő lépéseit bemutattuk, ezért az eljárást befejezzük. Megjegyezzük, hogy az 5. lépésben már elég közel kerültünk a minimumhelyhez, amely  $\bar{\mathbf{x}} = (1.5, -1)$ . Szokás a két utolsó közelítés távolsága helyett azt a megállási (terminálási) feltételt is használni, amelynél a  $\|\nabla f(\mathbf{x}_k)\|$  elegendően közel van a zérushoz.

**Newton típusú módszerek** Az előzőekben megismertük a Newton módszert és a kvázi Newton módszereket. Mindegyiknél egy  $\mathbf{s}_k$  különbségvektort határoztunk meg és ezt közvetlenül hozzáadtuk az  $\mathbf{x}_k$  vektorhoz, úgy kaptuk meg a következő közelítést, azaz  $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ . A most tárgyalásra kerülő módszerekben is egy lineáris egyenletrendszer kell megoldani, de a kapott vektort csupán egy iránynak használjuk és ebben az irányban elvégzünk egy vonalmenti minimalizálást. A következő közelítést ennek a vonalmenti minimalizálásnak az eredménye adja. Továbbra is megtartjuk a Newton típusú módszereknél használatos jelöléseket, tehát az  $\mathbf{x}$  vektorok különbségére az  $\mathbf{s}$  vektort, a gradiens vektorok különbségére pedig az  $\mathbf{y}$  vektort használjuk, azaz  $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$  és  $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$ . Az irányvektorra pedig a szokásos  $\mathbf{d}$  jelölést használjuk. A következőkben a Newton módszernek ill. a két kvázi Newton módszernek a vonalmenti minimalizálást is tartalmazó algoritmusát ismertetjük.

**Newton módszer I. algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy  $\mathbf{x}_k \in \mathbb{R}^n$  pontból.

**Közbülső lépés:**

1. Irányvektor meghatározása: Megoldjuk a  $\mathbf{H}(\mathbf{x}_k)\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$  lineáris egyenletrendszert  $\mathbf{d}_k$ -ra.

2. Megoldjuk a  $\min\{f(\mathbf{x}_k + \lambda \mathbf{d}_k)\}$  minimalizálási feladatot  $\lambda$ -ra, legyen az optimális megoldás  $\lambda_k$ .

3. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$ .

4. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást,  $k := k + 1$ .

**Newton módszer II. algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy  $\mathbf{x}_k \in \mathbb{R}^n$  pontból.

**Közbülső lépés:**

1. Irányvektor meghatározása:  $\mathbf{d}_k = -\mathbf{H}(\mathbf{x}_k)^{-1} \nabla f(\mathbf{x}_k)$ .
2. Megoldjuk a  $\min \{f(\mathbf{x}_k + \lambda \mathbf{d}_k)\}$  minimalizálási feladatot  $\lambda$ -ra, legyen az optimális megoldás  $\lambda_k$ .
3. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$ .
4. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást,  $k := k + 1$ .

Ismételten megjegyezzük, hogy a Newton I. és a Newton II. algoritmus mindegyikében a  $\mathbf{d}$  irányvektort ugyanannak az egyenletrendszernek a megoldásaként kapjuk. A Newton I. módszerben az egyenletrendszert valamilyen módszerrel oldjuk meg, a Newton II. módszerben pedig a megoldást az együtthatómátrix (Hesse mátrix) inverze segítségével határozzuk meg.

**Példa:**

Határozzuk meg az

$$f(x_1, x_2) = -(x_1^2 + x_2^2 - 2x_1 - 4x_2 + 6)^{-1}$$

függvény minimumát Newton módszerrel az  $\mathbf{x}_1 = [0, 1]$  startvektorból kiindulva, de használjunk vonalmenti optimalizálást!

**Megoldás:**

Emlékeztetjük az olvasót, hogy ezt a feladatot a Newton módszer c. fejezetben megpróbáltuk megoldani a Newton módszerrel és azt tapasztaltuk, hogy a módszer nem konvergált. Most ugyanezt a feladatot próbáljuk megoldani a Newton módszer és a vonalmenti optimalizálás összekapcsolásával. A lineáris egyenletrendszer megoldásaként kapott  $\mathbf{s}_1 = [-\frac{3}{5}, -\frac{3}{5}]$  vektort tehát most nem adjuk hozzá az  $\mathbf{x}_1$  vektorhoz, hanem a  $\mathbf{s}_1$  vektort csak egy irányvektorként használjuk és ebben az irányban elvégzünk egy egyváltozós optimalizálást. Célszerű a  $\mathbf{d}_1$  irányvektort egész számokkal felírni, mert így az egyváltozós optimalizálási feladat megoldása egyszerűbbé válik, tehát legyen  $\mathbf{d}_1 = [-1, -1]$ . Az  $\mathbf{x} = \mathbf{x}_1 + \lambda \mathbf{d}_1 = [-\lambda, 1 - \lambda]$  vektort behelyettesítve a célfüggvénybe az egyváltozós optimalizálási feladatot

$$\varphi(\lambda) = -\frac{1}{(-\lambda)^2 + (1 - \lambda)^2 - 2(-\lambda) - 4(1 - \lambda) + 6} \rightarrow \min!$$

A megoldás  $\lambda$ -ra  $\lambda_1 = -1$ , így  $\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1 \mathbf{d}_1 = [0, 1] + (-1)[-1, -1] = [1, 2]$ . Tovább nem folytatjuk a megoldást, mert az  $\mathbf{x}_2$  pontban a gradiens eltűnik. A  $\nabla f(\mathbf{x}_2) = \mathbf{0}$  pedig azt jelenti, hogy a minimum szükséges feltétele teljesül az  $\mathbf{x}_2$  pontban.

**Példa:**

Határozzuk meg az

$$f(x_1, x_2) = x_1^4 x_2^2 + 2x_1^2 x_2^2 + 17$$

függvény minimumát Newton módszerrel az  $\mathbf{x}_1 = [1, -1]$  startvektorból kiindulva, de használjunk vonalmenti optimalizálást!

**Megoldás:**

Emlékeztetjük az olvasót, hogy ezt a feladatot a Newton módszer c. fejezetben elkezdjük megoldani Newton módszerrel, most a pedig a Newton módszert összekapcsoljuk a vonalmenti optimumkereséssel.

A gradiensvektor és a Hesse mátrix a következő:

$$\nabla f(\mathbf{x}) = \left[ \begin{array}{c} 4x_1^3x_2^2 + 4x_1x_2^2 \\ 2x_2x_1^4 + 4x_2x_1^2 \end{array} \right], \quad \mathbf{H}(\mathbf{x}) = \left[ \begin{array}{cc} 12x_1^2x_2^2 + 4x_2^2 & 8x_2x_1^3 + 8x_2x_1 \\ 8x_2x_1^3 + 8x_2x_1 & 2x_1^4 + 4x_1^2 \end{array} \right].$$

Ezeknek az  $\mathbf{x}_1$  pontbeli értékeik a következők:

$$\nabla f(\mathbf{x}_1) = \left[ \begin{array}{c} 8 \\ -6 \end{array} \right], \quad \mathbf{H}(\mathbf{x}_1) = \begin{array}{cc} 16 & -16 \\ -16 & 6 \end{array}.$$

A megoldandó lineáris egyenletrendszer az **irányvektorra**:  $\mathbf{H}(\mathbf{x}_k)\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$ , amely a  $\mathbf{d}_1 = [d_1, d_2]$  ismeretlen irányvektor két koordinátájára az alábbi:

$$\begin{aligned} 16d_1 - 16d_2 &= -8 \\ -16d_1 + 6d_2 &= 6 \end{aligned}$$

A megoldás, azaz az  $\mathbf{x}_1$  vektorhoz tartozó  $\mathbf{d}_1$  irányvektor  $\mathbf{d}_1 = \left[-\frac{3}{10}, \frac{2}{10}\right]$ , vagy az inverz segítségével való egyenletrendszer megoldás:

$$\mathbf{d}_1 = -\mathbf{H}(\mathbf{x}_1)^{-1}\nabla f(\mathbf{x}_1) = -\frac{1}{-160} \begin{bmatrix} 6 & 16 \\ 16 & 16 \end{bmatrix} \begin{bmatrix} 4 \\ -3 \end{bmatrix} = \begin{bmatrix} -\frac{3}{10} \\ \frac{2}{10} \end{bmatrix}.$$

Az  $\mathbf{x}_2$  közelítést egy vonalmenti optimalizálással határozzuk meg. Célszerű az iránynak egész számot választani az egyszerűbb számolás miatt, így a  $\mathbf{d}_1 = [-3, 2]$  irányvektorral dolgozunk.

A  $\mathbf{d}_1$  irányban egy tetszőleges pont  $\mathbf{x} = \mathbf{x}_1 + \lambda\mathbf{d}_1 = [1, -1] + \lambda[-3, 2] = [1 - 3\lambda, -1 + 2\lambda]$ .

Az egyváltozós optimalizálási feladat:

$$\varphi(\lambda) = f(1 - 3\lambda, -1 + 2\lambda) = (1 - 3\lambda)^4(-1 + 2\lambda)^2 + 2(1 - 3\lambda)^2(-1 + 2\lambda)^2 \rightarrow \min!$$

Egy hatodfokú polinomnak a minimumát kell meghatározni, erre használhatjuk az egyváltozós minimumkereső módszerek valamelyikét. De most speciális ez a polinom, mert egyszerűen szorzattá alakítható:

$$\varphi(\lambda) = (1 - 3\lambda)^2(-1 + 2\lambda)^2((1 - 3\lambda)^2 + 2)$$

Minden tényező pozitív, így a függvény legkisebb értéke zérus lehet, amelyet a  $\lambda_1^{(1)} = \frac{1}{3}$  és a  $\lambda_1^{(2)} = \frac{1}{2}$  értékeknél vesz fel, így az  $\mathbf{x}_2$  közelítésre az alábbiak adódnak:  $\mathbf{x}_2^{(1)} = \left[0, -\frac{1}{3}\right]$  és az  $\mathbf{x}_2^{(2)} = \left[-\frac{1}{2}, 0\right]$ . Mindkettő optimumpont. A célfüggvény tüzetesebb vizsgálatával megállapítható, hogy minden olyan  $(x_1, x_2)$  vektor optimális megoldás, amelynél az  $x_1x_2$  szorzat zérus.

**Kvázi Newton típusú módszerek (BFGS és DFP módszerek)** Az elv hasonló a Newton módszernél látottakhoz, csak itt a kvázi Newton módszerekkel határozzuk meg a  $d_k$  irányvektort és ezt nem adjuk hozzá közvetlenül az  $\mathbf{x}_k$  vektorhoz, hanem a  $d_k$  vektort csupán egy iránynak használjuk, amely irányban elvágunk egy vonalmenti minimalizálást.

**BFGS (Broyden-Fletcher-Goldfarb-Shanno) eljárás algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  vektorból és egy szimmetrikus pozitív definit  $\mathbf{B}_k$  mátrixból.

**Közbülső lépés:**

1. Irányvektor meghatározása: Megoldjuk a  $\mathbf{B}_k \mathbf{d}_k = -\nabla f(\mathbf{x}_k)$  lineáris egyenletrendszert  $\mathbf{d}_k$ -ra.
2. Megoldjuk a  $\min \{f(\mathbf{x}_k + \lambda \mathbf{d}_k) : \lambda \geq 0\}$  minimalizálási feladatot  $\lambda$ -ra, legyen az optimális megoldás  $\lambda_k$ .
3. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$ .
4. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást a következő sossal.
5. Meghatározzuk az alábbiakat

$$\begin{aligned} \mathbf{s}_k &= \mathbf{x}_{k+1} - \mathbf{x}_k & \text{vagy} & & \mathbf{s}_k &= \lambda_k \mathbf{d}_k, \\ \mathbf{y}_k &= \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k), \\ \mathbf{B}_{k+1} &= \mathbf{B}_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{(\mathbf{B}_k \mathbf{s}_k)(\mathbf{B}_k \mathbf{s}_k)^T}{(\mathbf{B}_k \mathbf{s}_k)^T \mathbf{s}_k}. \end{aligned}$$

6. Folytatjuk az eljárást,  $k := k + 1$ .

**DFP (Davidon-Fletcher-Powell) eljárás algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  vektorból és egy szimmetrikus pozitív definit  $\mathbf{D}_k$  mátrixból.

**Közbülső lépés:**

1. Irányvektor meghatározása:  $\mathbf{d}_k = -\mathbf{D}_k \nabla f(\mathbf{x}_k)$ .
2. Megoldjuk a  $\min \{f(\mathbf{x}_k + \lambda \mathbf{d}_k) : \lambda \geq 0\}$  minimalizálási feladatot  $\lambda$ -ra, legyen az optimális megoldás  $\lambda_k$ .
3. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$ .
4. Megállunk, ha  $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon$ , egyébként folytatjuk az eljárást a következő sossal.
5. Meghatározzuk az alábbiakat

$$\begin{aligned} \mathbf{s}_k &= \mathbf{x}_{k+1} - \mathbf{x}_k & \text{vagy} & & \mathbf{s}_k &= \lambda_k \mathbf{d}_k, \\ \mathbf{y}_k &= \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k), \\ \mathbf{D}_{k+1} &= \mathbf{D}_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{(\mathbf{D}_k \mathbf{y}_k)(\mathbf{D}_k \mathbf{y}_k)^T}{(\mathbf{D}_k \mathbf{y}_k)^T \mathbf{y}_k}. \end{aligned}$$

6. Folytatjuk az eljárást,  $k := k + 1$ .

Emlékeztetjük az olvasót, hogy a BFGS módszerben a  $\mathbf{B}_k$  mátrix a Hesse mátrixot helyettesíti, míg a DFP módszerben a  $\mathbf{D}_k$  mátrix a Hesse mátrix inverzét helyettesíti.

A következőkben az egyik leggyakrabban használt módszerre, a DFP módszerre vonatkozóan két tételt ismertetünk.

**TÉTEL**

Legyen  $\mathbf{x}_1 \in \mathbb{R}^n$  tetszőleges vektor és  $\mathbf{D}_1$  szimmetrikus pozitív definit mátrix. A DFP módszer alkalmazása esetén, ha  $\nabla f(\mathbf{x}_k) \neq \mathbf{0}$ ,  $k = 1, 2, \dots$ , akkor igazak az alábbi állítások

- i)  $\mathbf{d}_1, \mathbf{d}_2, \dots$  kereső vektorok javító (csökkenő) irányvektorok,
- ii)  $\mathbf{D}_1, \mathbf{D}_2, \dots$  mátrixok szimmetrikus pozitív definit mátrixok.

Mielőtt a másik tételre rátérnénk definiáljuk a konjugált vektorok fogalmát.

### Konjugált vektorok fogalma

Legyen  $\mathbf{C}$  egy  $n \times n$ -es szimmetrikus mátrix. A  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}^n$  vektorokat  $\mathbf{C}$ -konjugáltaknak vagy egyszerűen konjugáltaknak nevezzük, ha lineárisan függetlenek és  $\mathbf{v}_i^T \mathbf{C} \mathbf{v}_j = 0$  minden  $i \neq j$  esetén.

Megjegyzés: Ha  $\mathbf{C} = \mathbf{E}$ , akkor ortogonális vektorokról beszélünk.

### TÉTEL

Legyen a minimalizálandó függvény kvadratikus, azaz  $f(\mathbf{x}) = \mathbf{c}\mathbf{x} + \frac{1}{2}\mathbf{x}^T \mathbf{H}\mathbf{x}$  és tegyük fel, hogy  $\mathbf{H}$  szimmetrikus pozitív definit mátrix. Alkalmazzuk ennek a feladatnak a megoldására a DFP módszert, amelyben a kiindulásnál legyen  $\mathbf{x}_1 \in \mathbb{R}^n$  tetszőleges vektor és  $\mathbf{D}_1$  szimmetrikus pozitív definit mátrix. Végezzünk el  $n$  darab iterációt. Ha minden  $k = 1, 2, \dots, n$  esetén  $\nabla f(\mathbf{x}_k) \neq \mathbf{0}$ , akkor igazak az alábbi állítások

- i)  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$  kereső irányok  $\mathbf{H}$ -konjugáltak,
- ii)  $\mathbf{D}_{n+1} = \mathbf{H}^{-1}$ ,
- iii)  $\mathbf{x}_{k+1}$  az optimális megoldás.

### Példa:

Oldjuk meg az alábbi optimalizálási feladatot Broyden-Fletcher-Goldfarb-Shanno módszerrel!

$$f(x_1, x_2) = x_1^2 + x_2^2 - 3x_1 + 6 \rightarrow \min!$$

Legyen az  $\mathbf{x}_1$  startvektor és a Hesse mátrixot helyettesítő  $\mathbf{B}_1$  startmátrix az alábbi:

$$\mathbf{x}_1 = (2, 1), \quad \mathbf{B}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

### Megoldás:

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - 3 \\ 2x_2 \end{bmatrix}.$$

1. lépés:

$$\mathbf{x}_1 = (2, 1), \quad \nabla f(\mathbf{x}_1) = (1, 2)$$

Az  $\mathbf{x}_1$  vektorhoz tartozó irányvektor meghatározásához megoldandó egyenletrendszer ( $\mathbf{B}_1 \mathbf{d}_1 = -\nabla f(\mathbf{x}_1)$ ):

$$\begin{aligned} 2d_1 + d_2 &= -1 \\ d_1 + d_2 &= -2 \end{aligned}$$

Az egyenletrendszer megoldása:  $d_1 = 1$ ,  $d_2 = -3$ , így a keresett irányvektor  $\mathbf{d}_1 = (1, -3)$ .

$$\mathbf{x} = \mathbf{x}_1 + \lambda \mathbf{d}_1 = (2 + \lambda, 1 - 3\lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(2 + \lambda, 1 - 3\lambda) = (2 + \lambda)^2 + (1 - 3\lambda)^2 - 3(2 + \lambda) + 6 \rightarrow \min!$$

$$\lambda_1 = \frac{1}{4}$$



2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1 \mathbf{d}_1 = (2, 1) + \frac{1}{4}(1, -3) = \left(\frac{9}{4}, \frac{1}{4}\right)$$

Most előkészítjük a  $\mathbf{B}_2$  mátrix számítását.

$$\nabla f(\mathbf{x}_2) = \left(\frac{3}{2}, \frac{1}{2}\right)$$

$$\mathbf{s}_1 = \mathbf{x}_2 - \mathbf{x}_1 = \lambda_1 \mathbf{d}_1 = \left(\frac{1}{4}, -\frac{3}{4}\right)$$

$$\mathbf{y}_1 = \nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1) = \left(\frac{3}{2}, \frac{1}{2}\right) - (1, 2) = \left(\frac{1}{2}, -\frac{3}{2}\right)$$

$$\mathbf{B}_1 \mathbf{s}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ -\frac{3}{4} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} -1 \\ -2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} \\ -\frac{2}{4} \end{bmatrix}$$

Megjegyzés: Itt nem igaz, hogy  $\mathbf{B}_k \mathbf{s}_k = -\nabla f(\mathbf{x}_k)$ , csak akkor igaz, ha  $\lambda_k = 1$ .

Javasoljuk az olvasónak, hogy a törtértékű mátrixok és vektorok közötti műveleteknél a fenti kiemeléssel éljen, mert az egész számokkal való műveletvégzés kisebb hibalehetőséget rejt. A mátrix-vektor művelet végén az osztásról ne feledkezzen meg, amikor további számolást végzünk az eredménnyel, akkor nem is érdemes elvégezni az osztást.

$$\mathbf{y}_1 \mathbf{y}_1^T = \frac{1}{2} \begin{bmatrix} 1 \\ -3 \end{bmatrix} \begin{bmatrix} 1 & -3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & -\frac{3}{4} \\ -\frac{3}{4} & \frac{9}{4} \end{bmatrix}$$

$$\mathbf{y}_1^T \mathbf{s}_1 = \frac{1}{2} \begin{bmatrix} 1 & -3 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ -\frac{3}{4} \end{bmatrix} = \frac{1}{8} 10 = \frac{5}{4}$$

$$(\mathbf{B}_1 \mathbf{s}_1)(\mathbf{B}_1 \mathbf{s}_1)^T = \frac{1}{4} \begin{bmatrix} -1 \\ -2 \end{bmatrix} \begin{bmatrix} -1 & -2 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{16} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{4} \end{bmatrix}$$

$$(\mathbf{B}_1 \mathbf{s}_1)^T \mathbf{s}_1 = \frac{1}{4} \begin{bmatrix} -1 & -2 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ -\frac{3}{4} \end{bmatrix} = \frac{1}{16} 5 = \frac{5}{16}$$

$$\begin{aligned} \mathbf{B}_2 &= \mathbf{B}_1 + \frac{\mathbf{y}_1 \mathbf{y}_1^T}{\mathbf{y}_1^T \mathbf{s}_1} - \frac{(\mathbf{B}_1 \mathbf{s}_1)(\mathbf{B}_1 \mathbf{s}_1)^T}{(\mathbf{B}_1 \mathbf{s}_1)^T \mathbf{s}_1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \frac{1}{\frac{5}{4}} \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix} - \frac{1}{\frac{5}{16}} \frac{1}{16} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \frac{4}{5} & -\frac{12}{5} \\ -\frac{12}{5} & \frac{36}{5} \end{bmatrix} - \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

3. lépés:

$$\mathbf{x}_2 = \left(\frac{9}{4}, \frac{1}{4}\right), \quad \nabla f(\mathbf{x}_2) = \left(\frac{3}{2}, \frac{1}{2}\right)$$

Az  $\mathbf{x}_2$  vektorhoz tartozó irányvektor meghatározásához megoldandó egyenletrendszer ( $\mathbf{B}_2 \mathbf{d}_2 = -\nabla f(\mathbf{x}_2)$ ):

$$\begin{aligned} 2d_1 &= -\frac{3}{2} \\ 2d_2 &= -\frac{1}{2} \end{aligned}$$

Az egyenletrendszer megoldása:  $d_1 = -\frac{3}{4}$ ,  $d_2 = -\frac{1}{4}$ , így a keresett irányvektor  $\mathbf{d}_2 = \left(-\frac{3}{4}, -\frac{1}{4}\right)$

A számítást nem folytatjuk tovább, hiszen a BFGS módszer lépéseit bemutattuk.

**Feladat:**

A gyakorlás kedvéért végezze el az egyváltozós optimalizálást, határozza meg az  $\mathbf{x}_3$  vektort és a  $\mathbf{B}_3$  mátrixot.

**Példa:**

Oldjuk meg az alábbi optimalizálási feladatot Davidon-Fletcher-Powell módszerrel!

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2 - x_1 + x_2 + 5 \rightarrow \min!$$

Legyen az  $\mathbf{x}_1$  startvektor és a Hesse mátrixot helyettesítő  $\mathbf{D}_1$  startmátrix az alábbi:

$$\mathbf{x}_1 = (0, 0), \quad \mathbf{D}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

**Megoldás:**

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 2x_2 - 1 \\ 2x_1 + 4x_2 + 1 \end{bmatrix}.$$

1. lépés:

$$\mathbf{x}_1 = (0, 0), \quad \nabla f(\mathbf{x}_1) = (-1, 1)$$

Az  $\mathbf{x}_1$  ponthoz tartozó irányvektor:  $\mathbf{d}_1 = -\mathbf{D}_1 \nabla f(\mathbf{x}_1) = (1, -1)$

$$\mathbf{x} = \mathbf{x}_1 + \lambda \mathbf{d}_1 = (\lambda, -\lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(\lambda, -\lambda) = \lambda^2 - 2\lambda + 5 \rightarrow \min!$$

$$\lambda_1 = 1$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1 \mathbf{d}_1 = (\lambda_1, -\lambda_1) = (1, -1)$$

Most előkészítjük a  $\mathbf{D}_2$  mátrix számítását.

$$\nabla f(\mathbf{x}_2) = (-1, -1)$$

$$\mathbf{s}_1 = \mathbf{x}_2 - \mathbf{x}_1 = \lambda_1 \mathbf{d}_1 = (1, -1)$$

$$\mathbf{y}_1 = \nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1) = (-1, -1) - (-1, 1) = (0, -2)$$

$$\mathbf{D}_1 \mathbf{y}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$\mathbf{s}_1 \mathbf{s}_1^T = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$\mathbf{s}_1^T \mathbf{y}_1 = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} = 2$$

$$(\mathbf{D}_1 \mathbf{y}_1)(\mathbf{D}_1 \mathbf{y}_1)^T = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \begin{bmatrix} 0 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

$$(\mathbf{D}_1 \mathbf{y}_1)^T \mathbf{y}_1 = \begin{bmatrix} 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} = 4$$

$$\mathbf{D}_2 = \mathbf{D}_1 + \frac{\mathbf{s}_1 \mathbf{s}_1^T}{\mathbf{s}_1^T \mathbf{y}_1} - \frac{(\mathbf{D}_1 \mathbf{y}_1)(\mathbf{D}_1 \mathbf{y}_1)^T}{(\mathbf{D}_1 \mathbf{y}_1)^T \mathbf{y}_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Az  $\mathbf{x}_2$  ponthoz tartozó irányvektor:  $\mathbf{d}_2 = -\mathbf{D}_2 \nabla f(\mathbf{x}_2) = (1, 0)$

$$\mathbf{x} = \mathbf{x}_2 + \lambda \mathbf{d}_2 = (1 + \lambda, -1)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 + \lambda, -1) = \lambda^2 - \lambda + 4 \rightarrow \min!$$

$$\lambda_2 = \frac{1}{2}$$

3. lépés:

$$\mathbf{x}_3 = \mathbf{x}_2 + \lambda_2 \mathbf{d}_2 = (1 + \lambda_2, -1) = \left(\frac{3}{2}, -1\right)$$

Most előkészítjük a  $\mathbf{D}_3$  mátrix számítását.

$$\nabla f(\mathbf{x}_3) = (0, 0).$$

Mivel az  $\mathbf{x}_3$  közelítéshez tartozó gradiensvektor zérus, így befejezzük az eljárást. Az  $\mathbf{x}_3 = (1.5, -1)$  pontban teljesül a minimum szükséges feltétele és a függvény konvex, így megkaptuk az optimális megoldást.

Mivel a minimalizálandó függvényünk kvadratikus és Hesse mátrixa pozitív definit, folytassuk tovább a számítást, hogy az algoritmushoz közölt tételek állításainak helyességét ellenőrizhessük.

$$\mathbf{s}_2 = \mathbf{x}_3 - \mathbf{x}_2 = \lambda_2 \mathbf{d}_2 = \left(\frac{1}{2}, 0\right)$$

$$\mathbf{y}_2 = \nabla f(\mathbf{x}_3) - \nabla f(\mathbf{x}_2) = (1, 1)$$

$$\mathbf{D}_2 \mathbf{y}_2 = \frac{1}{2} \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{s}_2 \mathbf{s}_2^T = \frac{1}{2} \frac{1}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{s}_2^T \mathbf{y}_2 = \frac{1}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2}$$

$$(\mathbf{D}_2 \mathbf{y}_2)(\mathbf{D}_2 \mathbf{y}_2)^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$(\mathbf{D}_2 \mathbf{y}_2)^T \mathbf{y}_2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1$$

$$\mathbf{D}_3 = \mathbf{D}_2 + \frac{\mathbf{s}_2 \mathbf{s}_2^T}{\mathbf{s}_2 \mathbf{y}_2^T} - \frac{(\mathbf{D}_2 \mathbf{y}_2)(\mathbf{D}_2 \mathbf{y}_2)^T}{(\mathbf{D}_2 \mathbf{y}_2)^T \mathbf{y}_2} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Tovább már nem számolunk. A célfüggvény Hesse mátrixa

$$\mathbf{H} = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix},$$

és ellenőrizhetjük a fenti tételek egyik állításának helyességét, miszerint a  $\mathbf{D}_3$  mátrix a Hesse mátrix inverze.

**Feladat:**

Ellenőrizze le a tételben megfogalmazott további állítások helyességét is!

A Davidon-Fletcher-Powell módszer az egyik leggyakrabban alkalmazott eljárás, ezért az egyes lépések eredményeit a könnyebb tájékozódás kedvéért egy táblázatban összefoglalva is közöljük.

$k$	$\mathbf{x}_k$	$\nabla f(\mathbf{x}_k)$	$\mathbf{D}_k$	$\mathbf{d}_k$	$\mathbf{x}(\lambda)$	$\lambda_k$	$s_k$	$\mathbf{y}_k$
1	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} \lambda \\ -\lambda \end{bmatrix}$	1	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -2 \end{bmatrix}$
2	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} -1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 + \lambda \\ -1 \end{bmatrix}$	$\frac{1}{2}$	$\begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
3	$\begin{bmatrix} \frac{3}{2} \\ -1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$					

**Példa:**

Oldjuk meg az alábbi optimalizálási feladatot Broyden-Fletcher-Goldfarb-Shanno módszerrel!

$$f(x_1, x_2) = x_1^2 - x_1x_2 + x_2^2 \rightarrow \min!$$

Legyen az  $\mathbf{x}_1$  startvektor és a Hesse mátrixot helyettesítő  $\mathbf{B}_1$  startmátrix az alábbi:

$$\mathbf{x}_1 = (1, -1), \quad \mathbf{B}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

**Megoldás:**

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 \end{bmatrix}.$$

1. lépés:

$$\mathbf{x}_1 = (1, -1), \quad \nabla f(\mathbf{x}_1) = (3, -3)$$

Az  $\mathbf{x}_1$  vektorhoz tartozó irányvektor meghatározásához megoldandó egyenletrendszer ( $\mathbf{B}_1\mathbf{d}_1 = -\nabla f(\mathbf{x}_1)$ ):

$$\begin{aligned} 2d_1 + d_2 &= -3 \\ d_1 + d_2 &= 3 \end{aligned}$$

Az egyenletrendszer megoldása:  $d_1 = -6$ ,  $d_2 = 9$ , így a keresett irányvektor  $\mathbf{d}_1 = (-6, 9)$ . Mivel irányvektorról van szó, a hossza nem játszik szerepet, ezért javasoljuk, hogy az egyváltozós optimalizálási feladatban a fele olyan hosszú  $\mathbf{d}_1 = (-2, 3)$  irányvektort használjunk, így kisebb számokkal kell dolgozni.

$$\mathbf{x} = \mathbf{x}_1 + \lambda\mathbf{d}_1 = (1 - 2\lambda, -1 + 3\lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 - 2\lambda, -1 + 3\lambda) = (1 - 2\lambda)^2 - (1 - 2\lambda)(-1 + 3\lambda) + (-1 + 3\lambda)^2$$

$$\lambda_1 = \frac{15}{38}$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1\mathbf{d}_1 = (1, -1) + \frac{15}{38}(-2, 3) = \left(\frac{8}{38}, \frac{7}{38}\right)$$

Most előkészítjük a  $\mathbf{B}_2$  mátrix számítását.

$$\nabla f(\mathbf{x}_2) = \left(\frac{9}{38}, \frac{6}{38}\right)$$

$$\mathbf{s}_1 = \mathbf{x}_2 - \mathbf{x}_1 = \lambda_1\mathbf{d}_1 = \frac{15}{38}(-2, 3)$$

$$\mathbf{y}_1 = \nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1) = \left(\frac{9}{38}, \frac{6}{38}\right) - (3, -3) = \frac{15}{38}(-7, 8)$$

$$\mathbf{B}_1\mathbf{s}_1 = \frac{15}{38} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \frac{15}{38} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\mathbf{y}_1\mathbf{y}_1^T = \frac{15}{38} \frac{15}{38} \begin{bmatrix} -7 \\ 8 \end{bmatrix} \begin{bmatrix} -7 & 8 \end{bmatrix} = \frac{15}{38} \frac{15}{38} \begin{bmatrix} 49 & -56 \\ -56 & 64 \end{bmatrix}$$

$$\mathbf{y}_1^T\mathbf{s}_1 = \frac{15}{38} \frac{15}{38} \begin{bmatrix} -7 & 8 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \frac{15}{38} \frac{15}{38} 38 = \frac{15}{38} 15$$

$$(\mathbf{B}_1\mathbf{s}_1)(\mathbf{B}_1\mathbf{s}_1)^T = \frac{15}{38} \frac{15}{38} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} = \frac{15}{38} \frac{15}{38} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$(\mathbf{B}_1 \mathbf{s}_1)^T \mathbf{s}_1 = \frac{15}{38} \frac{15}{38} \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \frac{15}{38} \frac{15}{38} 5$$

$$\begin{aligned} \mathbf{B}_2 &= \mathbf{B}_1 + \frac{\mathbf{y}_1 \mathbf{y}_1^T}{\mathbf{y}_1^T \mathbf{s}_1} - \frac{(\mathbf{B}_1 \mathbf{s}_1)(\mathbf{B}_1 \mathbf{s}_1)^T}{(\mathbf{B}_1 \mathbf{s}_1)^T \mathbf{s}_1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \frac{\frac{15}{38} \frac{15}{38}}{\frac{15}{38} \frac{15}{38} 5} \begin{bmatrix} 49 & -56 \\ -56 & 64 \end{bmatrix} - \frac{\frac{15}{38} \frac{15}{38}}{\frac{15}{38} \frac{15}{38} 5} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \frac{49}{38} & -\frac{56}{38} \\ -\frac{56}{38} & \frac{64}{38} \end{bmatrix} - \begin{bmatrix} \frac{1}{5} & -\frac{1}{5} \\ -\frac{1}{5} & \frac{1}{5} \end{bmatrix} = \frac{1}{190} \begin{bmatrix} 587 & -52 \\ -52 & 472 \end{bmatrix} \end{aligned}$$

3. lépésben először az  $\mathbf{x}_2$  vektorhoz tartozó  $\mathbf{d}_2$  irányvektort kell meghatározni a  $\mathbf{B}_2 \mathbf{d}_2 = -\nabla f(\mathbf{x}_2)$  egyenletrendszer megoldásával, majd az iránymenti optimalizálás következik. A számítást nem folytatjuk tovább, hiszen a BFGS módszer lépéseit ezzel is bemutattuk.

#### Feladat:

A gyakorlás kedvéért határozza meg az irányvektort, végezze el az egyváltozós optimalizálást, majd határozza meg az  $\mathbf{x}_3$  vektort és a  $\mathbf{B}_3$  mátrixot.

#### Példa:

Oldjuk meg az előző optimalizálási feladatot Davidon-Fletcher-Powell módszerrel!

$$f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 \rightarrow \min!$$

Legyen az  $\mathbf{x}_1$  startvektor és a Hesse mátrixot helyettesítő  $\mathbf{D}_1$  startmátrix az alábbi:

$$\mathbf{x}_1 = (1, -1), \quad \mathbf{D}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

#### Megoldás:

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 \end{bmatrix}.$$

1. lépés:

$$\mathbf{x}_1 = (1, -1), \quad \nabla f(\mathbf{x}_1) = (3, -3)$$

$$\text{Az } \mathbf{x}_1 \text{ ponthoz tartozó irányvektor: } \mathbf{d}_1 = -\mathbf{D}_1 \nabla f(\mathbf{x}_1) = - \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = (-3, 0)$$

Az egyszerűség kedvéért most is használhatjuk a rövidebb irányvektort, legyen ez  $\mathbf{d}_1 = (-1, 0)$ .

$$\mathbf{x} = \mathbf{x}_1 + \lambda \mathbf{d}_1 = (1, -1) + \lambda(-1, 0) = (1 - \lambda, -1)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 - \lambda, -1) = (1 - \lambda)^2 - (1 - \lambda)(-1) + (-1)^2 \rightarrow \min!$$

$$\lambda_1 = \frac{3}{2}$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1 \mathbf{d}_1 = \frac{1}{2}(-1, -2)$$

Most előkészítjük a  $\mathbf{D}_2$  mátrix számítását.

$$\nabla f(\mathbf{x}_2) = \frac{3}{2}(0, -1)$$

$$\mathbf{s}_1 = \mathbf{x}_2 - \mathbf{x}_1 = \lambda_1 \mathbf{d}_1 = \frac{3}{2}(-1, 0)$$

$$\mathbf{y}_1 = \nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1) = (0, -\frac{3}{2}) - (3, -3) = \frac{3}{2}(-2, 1)$$

$$\begin{aligned}\mathbf{D}_1 \mathbf{y}_1 &= \frac{3}{2} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \frac{3}{2} \begin{bmatrix} -3 \\ -1 \end{bmatrix} \\ \mathbf{s}_1 \mathbf{s}_1^T &= \frac{3}{2} \frac{3}{2} \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \end{bmatrix} = \frac{3}{2} \frac{3}{2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ \mathbf{s}_1^T \mathbf{y}_1 &= \frac{3}{2} \frac{3}{2} \begin{bmatrix} -1 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \frac{3}{2} \frac{3}{2} 2 \\ (\mathbf{D}_1 \mathbf{y}_1)(\mathbf{D}_1 \mathbf{y}_1)^T &= \frac{3}{2} \frac{3}{2} \begin{bmatrix} -3 \\ -1 \end{bmatrix} \begin{bmatrix} -3 & -1 \end{bmatrix} = \frac{3}{2} \frac{3}{2} \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \\ (\mathbf{D}_1 \mathbf{y}_1)^T \mathbf{y}_1 &= \frac{3}{2} \frac{3}{2} \begin{bmatrix} -3 & -1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \frac{3}{2} \frac{3}{2} 5\end{aligned}$$

$$\begin{aligned}\mathbf{D}_2 &= \mathbf{D}_1 + \frac{\mathbf{s}_1 \mathbf{s}_1^T}{\mathbf{s}_1^T \mathbf{y}_1} - \frac{(\mathbf{D}_1 \mathbf{y}_1)(\mathbf{D}_1 \mathbf{y}_1)^T}{(\mathbf{D}_1 \mathbf{y}_1)^T \mathbf{y}_1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \frac{\frac{3}{2} \frac{3}{2}}{\frac{3}{2} \frac{3}{2} 2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \frac{\frac{3}{2} \frac{3}{2}}{\frac{3}{2} \frac{3}{2} 5} \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \frac{9}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{1}{5} \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 17 & 4 \\ 4 & 8 \end{bmatrix}\end{aligned}$$

3. lépés:

$$\mathbf{x}_2 = \frac{1}{2}(-1, -2), \quad \nabla f(\mathbf{x}_2) = \frac{3}{2}(0, -1)$$

$$\text{Az } \mathbf{x}_2 \text{ ponthoz tartozó irányvektor: } \mathbf{d}_2 = -\mathbf{D}_2 \nabla f(\mathbf{x}_2) = -\frac{1}{10} \frac{3}{2} \begin{bmatrix} 17 & 4 \\ 4 & 8 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} =$$

$$\frac{3}{10} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Az egyszerűség kedvéért most is használhatjuk a rövidebb irányvektort, legyen ez  $\mathbf{d}_2 = (1, 2)$ .

$$\mathbf{x} = \mathbf{x}_2 + \lambda \mathbf{d}_2 = \frac{1}{2}(-1, -2) + \lambda(1, 2) = (\lambda - \frac{1}{2}, 2\lambda - 1)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(\lambda - \frac{1}{2}, 2\lambda - 1) = (\lambda - \frac{1}{2})^2 - (\lambda - \frac{1}{2})(2\lambda - 1) + (2\lambda - 1)^2 \rightarrow \min!$$

$$\lambda_2 = \frac{1}{2}$$

4. lépés:

$$\mathbf{x}_3 = \mathbf{x}_2 + \lambda_2 \mathbf{d}_2 = (0, 0)$$

Most előkészítjük a  $\mathbf{D}_3$  mátrix számítását.

$$\nabla f(\mathbf{x}_3) = (0, 0).$$

Mivel az  $\mathbf{x}_3$  közelítéshez tartozó gradiensvektor zérus, így befejezzük az eljárást. Az  $\mathbf{x}_3 = (0, 0)$  pontban teljesül a minimum szükséges feltétele és a függvény konvex, így megkaptuk az optimális megoldást.

**Konjugált gradiens módszer (Fletcher, Reeves)** Az irány megválasztására eddig sok módszert láttunk, a gradiens módszerben  $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$ , a DFP módszerben  $\mathbf{d}_k = -\mathbf{D}_k \nabla f(\mathbf{x}_k)$ . A DFP módszer úgy is felfogható, hogy nem a legnagyobb csökkenés irányába mutató vektort használjuk irányvektornak, hanem annak egy mátrix-szal szorzott transzformáltját. A konjugált gradiens módszerben sem a gradiens vektor  $(-1)$ -szeresét használjuk irányvektornak, hanem ahhoz egy vektort adunk hozzá, amely az előző lépésben használt irányvektor irányába mutat, képletben

$$\mathbf{d}_{k+1} = -\nabla f(\mathbf{x}_{k+1}) + \alpha_k \mathbf{d}_k.$$

Több módszer ismert az  $\alpha_k$  szorzóra, Fletcher és Reeves az alábbi szorzót javasolta

$$\alpha_k = \frac{\|\nabla f(\mathbf{x}_{k+1})\|^2}{\|\nabla f(\mathbf{x}_k)\|^2}.$$

**Fletcher-Reeves módszer algoritmus:**

**Induló lépés** ( $k = 1$ ): Kiindulunk egy tetszőleges  $\mathbf{x}_k \in \mathbb{R}^n$  vektorból és a  $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$  irányvektorból.

**Közbülső lépés:**

1. Megállunk, ha  $\|\nabla f(\mathbf{x}_k)\| < \varepsilon$ , egyébként folytatjuk az eljárást a következő sorral.
2. Megoldjuk a  $\min \{f(\mathbf{x}_k + \lambda \mathbf{d}_k) : \lambda \geq 0\}$  minimalizálási feladatot  $\lambda$ -ra, legyen az optimális megoldás  $\lambda_k$ .
3. Meghatározzuk a következő közelítést:  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$ .
4. Irányvektor meghatározása:  $\mathbf{d}_{k+1} = -\nabla f(\mathbf{x}_{k+1}) + \frac{\|\nabla f(\mathbf{x}_{k+1})\|^2}{\|\nabla f(\mathbf{x}_k)\|^2} \mathbf{d}_k$ .
5. Folytatjuk az eljárást,  $k := k + 1$ .

**TÉTEL**

Legyen a minimalizálandó függvény kvadratikus, azaz  $f(\mathbf{x}) = \mathbf{c}\mathbf{x} + \frac{1}{2}\mathbf{x}^T \mathbf{H}\mathbf{x}$ . Alkalmazzuk ennek a feladatnak a megoldására a Fletcher-Reeves módszert, amelyben a kiinduláskor legyen  $\mathbf{x}_1 \in \mathbb{R}^n$  tetszőleges vektor és  $\mathbf{d}_1 = -\nabla f(\mathbf{x}_1)$ . Végezzünk el  $n$  darab iterációt. Ha minden  $k = 1, 2, \dots, n$  esetén  $\nabla f(\mathbf{x}_k) \neq \mathbf{0}$ , akkor igazak az alábbi állítások

- i)  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$  kereső irányok  $\mathbf{H}$ -konjugáltak,
- ii)  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$  kereső irányok javító (csökkenő) irányok,
- iii)  $\alpha_k = \frac{\|\nabla f(\mathbf{x}_{k+1})\|^2}{\|\nabla f(\mathbf{x}_k)\|^2} = \frac{\mathbf{d}_k^T \mathbf{H} \nabla f(\mathbf{x}_{k+1})}{\mathbf{d}_k^T \mathbf{H} \mathbf{d}_k}$ ,  $k = 1, 2, \dots, n$ .

**Példa:**

Oldjuk meg az alábbi optimalizálási feladatot Fletcher-Reeves módszerrel. Legyen az  $\mathbf{x}_1 = (0, 0)$  a startvektor és legyen  $\varepsilon = 0.005$  a pontossági tűrés.

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 2x_2^2 - x_1 + x_2 + 5 \rightarrow \min!$$

**Megoldás:**

Előkészületként határozzuk meg a célfüggvény gradiensét

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 2x_2 - 1 \\ 2x_1 + 4x_2 + 1 \end{bmatrix}$$

1. lépés:

Az első lépés mindig megegyezik a gradiens módszerrel, mivel ekkor még nincs előző irányvektorunk.

$$\mathbf{x}_1 = (0, 0), \quad \mathbf{d}_1 = -\nabla f(\mathbf{x}_1) = (1, -1), \quad \mathbf{x} = \mathbf{x}_1 + \lambda \mathbf{d}_1 = (\lambda, -\lambda)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(\lambda, -\lambda) = \lambda^2 - 2\lambda + 5 \rightarrow \min!$$

$$\lambda_1 = 1$$

2. lépés:

$$\mathbf{x}_2 = \mathbf{x}_1 + \lambda_1 \mathbf{d}_1 = (\lambda_1, -\lambda_1) = (1, -1), \quad \nabla f(\mathbf{x}_2) = (-1, -1)$$

Ettől a lépéstől térünk el a gradiens módszertől, mert az irányvektor meghatározása kissé bonyolultabb.

Ehhez számítsuk ki a két utolsó közelítéshez tartozó gradiensek hosszának négyzetét:

$$\|\nabla f(\mathbf{x}_1)\|^2 = 2, \quad \|\nabla f(\mathbf{x}_2)\|^2 = 2$$

Az új irányvektor

$$\mathbf{d}_2 = -\nabla f(\mathbf{x}_2) + \frac{\|\nabla f(\mathbf{x}_2)\|^2}{\|\nabla f(\mathbf{x}_1)\|^2} \mathbf{d}_1 = -(-1, -1) + \frac{2}{2}(1, -1) = (2, 0)$$

$$\mathbf{x} = \mathbf{x}_2 + \lambda \mathbf{d}_2 = (1 + 2\lambda, -1)$$

Az egyváltozós optimalizálási feladat és megoldása:

$$\varphi(\lambda) = f(1 + 2\lambda, -1) = 4\lambda^2 - 2\lambda + 4 \rightarrow \min!$$

$$\lambda_2 = \frac{1}{4}$$

3. lépés:

$$\mathbf{x}_3 = \mathbf{x}_2 + \lambda_2 \mathbf{d}_2 = \left(\frac{3}{2}, -1\right), \quad \mathbf{d}_3 = -\nabla f(\mathbf{x}_3) = (0, 0).$$

Mivel az  $\mathbf{x}_3$  közelítéshez tartozó gradiensvektor zérus, így befejezzük az eljárást. Az  $\mathbf{x}_3 = (1.5, -1)$  pontban teljesül a minimum szükséges feltétele és a függvény konvex, így megkaptuk az optimális megoldást.

Mivel a minimalizálandó függvényünk kvadratikus, így érvényesek a Fletcher-Reeves módszerre a fentebb közölt tétel állításai.

#### **Feladat:**

Ellenőrizze le a tételben megfogalmazott állítások helyességét.

Végül az egyváltozós optimalizálási feladat megoldásáról kell szólnunk. A példákban úgy oldottuk meg az egyváltozós minimalizálást, hogy a  $\varphi'(\lambda) = 0$  stacionárius egyenletet megoldottuk és ellenőriztük a minimum  $\varphi''(\lambda) > 0$  elégséges feltételét. A példákban a  $\varphi(\lambda)$  függvény differenciálható volt. Amennyiben nem differenciálható a függvény vagy a stacionárius egyenlet túl bonyolult, akkor a korábbi fejezetben tárgyalt egyváltozós minimumkereső eljárások valamelyikével kell elvégezni az egyváltozós minimalizálást (leginkább a Golden section módszert használják). Ezekben a módszerekben szükségünk van egy bizonytalansági intervallumra és a függvényre kvázikonvexitást írtunk elő. Az  $[a, b]$  bizonytalansági intervallum baloldali végpontját 0-ra választjuk, hiszen az adott  $\mathbf{x}_k$  ponttól kell a  $\varphi(\lambda) = f(\mathbf{x}_k + \lambda \mathbf{d}_k)$  függvény minimumát meghatározni. A jobboldali végpont megválasztása már körülményesebb, választhatjuk például 1-re. Nem biztos azonban, hogy a  $\varphi(\lambda)$  függvény a  $[0, 1]$  bizonytalansági intervallumban kvázikonvex. Nagyobb az esélyünk erre, ha kisebb intervallumot választunk. Ha a függvény menetéről van valami ismeretünk, akkor könnyebb a döntés. Amennyiben a függvény kvázikonvex és a bizonytalansági intervallum belsejében lévő pontra adódik a minimum, akkor biztosak lehetünk abban, hogy az adott irányban megtaláltuk a minimumpontot. Ha a minimumpont valamelyik végpontra adódik, akkor nem lehetünk biztosak afelől, hogy a minimumpontot találtuk meg, mert az intervallumon kívül is lehet a minimumpont. Ha a jobboldali végpont adódott, akkor az új kezdő bizonytalansági intervallumra az  $[1, 2]$ -t választjuk. Ha a baloldali végpont adódott, akkor az új kezdő bizonytalansági intervallumra a  $[-1, 0]$ -t választjuk. Ebben a intervallumban is elvégezzük az egyváltozós optimalizálást és szükség esetén további új kezdő intervallumot adunk meg.