részletesen kiírva

$$\begin{aligned} r_1 &= g_1(t) \\ r_2 &= g_2(t) \\ r_3 &= g_3(t) \end{aligned}$$

alakú. Ha $t \in D$, és így az érintővektor nem nulla, akkor lokálisan t kiküszöbölhető, és r_1, r_2, r_3 közül az egyikkel kifejezhető a másik kettő, azaz explicit előállításra térhetünk át. Megint az implicit függvény tétel szerint egy $F_1(r_1, r_2, r_3) = 0$, $F_2(r_1, r_2, r_3) = 0$ alakú egyenletrendszer, ahol $r_1, r_2, r_3 \in \mathbb{R}$ és F_1, F_2 valós értékű, folytonosan differenciálható függvények, valamint a

$$\begin{pmatrix} \frac{\partial F_1}{\partial r_1} & \frac{\partial F_1}{\partial r_2} & \frac{\partial F_1}{\partial r_3} \\ \frac{\partial F_2}{\partial r_1} & \frac{\partial F_2}{\partial r_2} & \frac{\partial F_2}{\partial r_2} \end{pmatrix}$$

mátrix rangja 2, lokálisan egy sima egydimenziós felületet, azaz sima görbét ad meg, mert lokálisan áttérhetünk explicit előállításra.

* **2.2.46. Felületi normális, érintősík, felületi görbék.** Ha $r : D \to \mathbb{R}^3$ egy kétdimenziós sima felület, $u \in D$, akkor a $\partial_1 r(u) \times \partial_2 r(u)$ vektort *felületi normális*nak, az

$$n(u) = \frac{\partial_1 r(u) \times \partial_2 r(u)}{\left|\partial_1 r(u) \times \partial_2 r(u)\right|}$$

vektort pedig *felületi normális egységvektor*nak fogjuk nevezni. Legtöbbször nem írjuk ki a változótól való függést. Világos, hogy $\partial_1 r$, $\partial_2 r$ és n egy (pozitív irányítású) bázist alkotnak. Az $r(u)$ ponton átmenő, $\langle p - r(u), n(u)\rangle = 0$, $p \in \mathbb{R}^3$ egyenletű síkot $r(u)$-beli (vagy $r(u)$-n átmenő) *érintősík*nak nevezzük. Ez az $r(u)$ pont körül elsőrendben közelíti a felületdarabot.

Egy $r \circ u$ alakban előállítható görbét, ahol $u : [a, b] \to D$ egy görbe, *felületi görbé*nek fogunk nevezni. Ha $u = (u_1, u_2)$ folytonosan differenciálható, akkor a felületi görbe is, és érintővektora

$$(r \circ u)' = u_1' \partial_1 r + u_2' \partial_2 r,$$

azaz egy, az $r\big(u(t)\big)$ pontból kiinduló, az érintősíkban haladó vektor. Felületi görbék például az úgynevezett paramétervonalak, amelyeknek egyenlete $u_1(t) = t$, u_2 konstans, illetve u_1 konstans, $u_2(t) = t$.

* **2.2.47. Approximáció.** Az approximáció feladata a következő. Legyen adott egy H halmazon értelmezett valós f függvény. Ismerjük $f(x)$ értékeit a H halmazon, vagy annak adott $x_1, x_2, \ldots, x_n$ pontjaiban. Keresendő egy m paraméteres $g(x, a_1, \ldots, a_m)$ függvényseregből az a függvény, amely f-et a „legjobban közelíti". Általában olyan függvények közül választjuk a közelítő függvényt, amelyek értékei könnyen számíthatók. Mivel számítógépen közvetlenül csak a négy alapművelet végezhető el, előnyben részesítjük a polinom és racionális törtfüggvény közelítéseket. Lehetséges az is, hogy érdemes H-t kisebb részekre osztani, és a részeken külön-külön keresni a közelítéseket. Egy adott halmazon való közelítést helyettesítéssel egy másik halmazon való közelítésre vezethetünk

vissza, például bármilyen korlátos zárt intervallumon való közelítés lineáris helyettesítéssel visszavezethető a $[0,1]$ vagy a $[-1,1]$ intervallumon való közelítésre. A lineáris helyettesítés előnye, hogy a polinomok polinomba, racionális törtfüggvények racionális törtfüggvénybe mennek át, és a fokszámok sem változnak.

Attól függően, hogy f és a közelítő függvény eltérését hogyan mérjük, különböző eseteket kapunk. Ha azt követeljük meg, hogy a közelítő függvény az $x_1, x_2, \dots, x_n$ pontokban megegyezzen f-el, kapjuk az *interpoláció* esetét. Ha azt követeljük meg, hogy az $x_1, x_2, \dots, x_n$ pontokban mért eltérések négyzetösszege minimális legyen, akkor a *legkisebb négyzetes közelítés*t kapjuk. Végül, ha az egész H-n vett eltérés szuprémumát minimalizáljuk, akkor az *egyenletesen legjobb közelítés*t kapjuk.

Ha az f függvény értékeit csak az $x_1, x_2, \dots, x_n$ pontokban ismerjük, akkor az interpoláció a legkézenfekvőbb lehetőség. A legáltalánosabb esetben az

$$f(x_j) = g(x_j, a_1, \dots, a_m), \qquad j = 1, 2, \dots, n$$

egyenletrendszert kapjuk. Ezt az $a_1, \dots, a_m$ paraméterekre megoldva kapunk egy interpoláló függvényt. Például, ha $g_1, \dots, g_m$ adott és

$$g(x, a_1, \dots, a_m) = a_1 g_1(x) + a_2 g_2(x) + \dots + a_m g_m(x),$$

akkor lineáris egyenletrendszert kapunk. A legfontosabb az az eset, amikor f egy $[a,b]$ intervallumon értelmezett valós függvény, $x_1, x_2, \dots, x_n$ az $[a,b]$ pontjai, interpoláló függvényként pedig n-nél alacsonyabb fokú polinomot keresünk: ez a *Lagrange-interpoláció*. Mivel egy nem nulla legfeljebb n-ed fokú polinomnak legfeljebb n gyöke van, ha $c_0, \dots, c_n$ különböző, $d_0, \dots, d_n$ pedig tetszőleges elemei $\mathbb{K}$-nak, akkor legfeljebb egy olyan legfeljebb n-ed fokú f polinom létezik, amelyre $f(c_j) = d_j$, ha $j = 0, 1, \dots, n$, mert egyébként két ilyen polinom különbségének $n+1$ gyöke lenne. Mindig létezik is ilyen polinom, és az alábbi eljárással megkapható: legyen

$$l_j(x) = \frac{\prod_{i \neq j}(x - c_i)}{\prod_{i \neq j}(c_j - c_i)}$$

a j-edik *Lagrange interpolációs alappolinom* (erre $l_j(x_j) = 1$ és $l_j(x_i) = 0$, ha $i \neq j$), és legyen $f = \sum_{j=0}^{n} d_j l_j$.

Sokszor a közelítendő függvény mért értékei elég nagy hibát tartalmaznak. Ilyenkor nem célszerű azt kívánni, hogy a közelítő függvény az $x_1, x_2, \dots, x_n$ pontokban megegyezzen a mért értékekkel, mert akkor a hibát is tartalmazza. Elég azt megkövetelni, hogy a mért értékek közelében legyen. A *legkisebb négyzetek módszere* értelmében tehát az $a_1, \dots, a_m$ paramétereket úgy kell megválasztani, hogy ha $f(x_1), f(x_2), \dots, f(x_n)$ a mért értékek, akkor a

$$\Phi(a_1, \dots, a_m) = \sum_{i=1}^{n} w_i \big(f(x_i) - g(x_i, a_1, \dots, a_m)\big)^2$$

eltérés minimális legyen. Itt a w_i-k pozitív súlyok, amelyeket az $f(x_i)$ értékek szórásnégyzete reciprokának célszerű választani, mert minél pontatlanabb a mért érték, annál kisebb súllyal kívánjuk figyelembe venni. Ezt a minimumot a $\nabla\Phi(a_1, \dots, a_m) = 0$ egyenletrendszer megoldásával kereshetjük.

* **2.2.48. A Newton-módszer.** Ha $f \in \mathbb{R}^k \to \mathbb{R}^k$ és az $f(x) = 0$ egyenletet kívánjuk megoldani, természetes módon adódik a Newton-módszer: az egyenletet egy adott x_n érték birtokában az f függvényt lineáris részével közelítjük. Ekkor az

$$f(x_n) + f'(x_n)(x - x_n) \approx 0$$

egyenletet kapjuk, amiből a következő közelítésre az

$$f'(x_n)(x_{n+1} - x_n) = f'(x_n)\Delta x_n = -f(x_n)$$

egyenlet adódik, így x_{n+1} meghatározásához egy lineáris egyenletet kell megoldani, például Gauss-eliminációval. Az iteráció általában csak a gyök közvetlen közeléből indítva konvergens.

Gyakran használatos a módosított Newton-módszer is, ennél a derivált kiszámítására és invertálására csak egyszer van szükség, de a konvergencia lelassul:

$$\Delta x_n = -\big(f'(x_0)\big)^{-1} f(x_n).$$

Néhány lépés után újra számolva a deriváltat, a konvergencia gyorsítható.

* **2.2.49. Kvázi Newton-módszer.** A Newton-módszernél az $f'(x_n)\Delta x_n = -f(x_n)$ egyenletet kell megoldanunk minden lépésben. A kvázi Newton-módszernél x_{n+1} értékét az $A_n \Delta x_n = -\alpha_n f(x_n)$ egyenlet megoldásával nyerjük, ahol A_n az $f'(x_n)$ egy közelítése, az α_n pedig egy *relaxációs paraméter*. A közelítést nyerhetjük például differenciákkal közelítve a parciális deriváltakat. Ha minden irányban a differencia lépésköze konstansszor $|x_n - x|$ alatt marad, ahol x a gyök, akkor a konvergencia elméletileg kvadratikus, de a kerekítési hibák veszélyesek lehetnek. A *Broyden-módszer*nél $\alpha_n = 1$ minden n-re, $A_0 = f'(x_0)$, és A_{n+1}-et úgy próbáljuk meghatározni, hogy $A_{n+1}\Delta x_n = f(x_{n+1}) - f(x_n) = \Delta f(x_n)$ teljesüljön. Mivel ez az egyenlet alulhatározott, úgy választunk, hogy $A_{n+1} - A_n = \Delta A_n$ mátrixának normája (a szokásos bázisban) minimális legyen. Ez akkor teljesül, ha ΔA_n mátrixa a szokásos bázisban

$$\frac{\big[\Delta f(x_n) - A_n \Delta x_n\big]\big[\Delta x_n\big]'}{|\Delta x_n|^2}.$$

A konvergencia lineárisnál jobb, bár nem biztos, hogy $A_n \to f'(x)$. Lehetőség van mindjárt A_{n+1}^{-1} számítására: a szokásos bázisban $A_{n+1}^{-1} - A_n^{-1} = \Delta A_n^{-1}$ mátrixa

$$\frac{\big[\Delta x_n - A_n^{-1}\Delta f(x_n)\big]\big[\Delta x_n\big]'[A_n^{-1}]}{\big\langle \Delta x_n, A_n^{-1}\Delta f(x_n)\big\rangle}.$$

* **2.2.50. Feladat [10].** Alkalmazzuk a Newton-módszert és változatait az $e^{x_1^2 + x_2^2} - 1 = 0$, $e^{x_1^2 - x_2^2} - 1 = 0$ egyenletrendszer megoldására $x_1, x_2 = 0{,}1, 10, 20$ kezdőértékekkel.

* **2.2.51. Feladat [10].** Alkalmazzuk a Newton-módszert és változatait az $x_1+x_2-3=0$, $x_1^2+x_2^2-9=0$ egyenletrendszer megoldására $x_1=2$, $x_2=4$ kezdőértékekkel. Hova konvergál a Broyden-módszernél Q_n?

* **2.2.52. Kapcsolat minimumfeladatokkal.** Ismeretes, hogy egy valós értékű $\Phi \in \mathbb{R}^k \to \mathbb{R}$ függvény minimumának megkeresését visszavezethetjük a $\Phi'(x)=0$ egyenlet megoldására. Megfordítva, egy $f(x)=0$, $f \in \mathbb{R}^k \to \mathbb{R}^k$ operátoregyenlet megoldásait meghatározhatjuk úgy is, hogy megkeressük a $\Phi(x)=\|f(x)\|_2^2$ függvény minimumait. A $\|\ \|_2^2$ választás azért előnyös például $\|\ \|_2$ helyett, mert a $\|\ \|_2^2$ függvény differenciálható. Néha előnyösebb egy másik normát használni, például a $\Phi(x)=\sum_{i=1}^k w_i\big(f_i(x)\big)^2$ függvény minimumait keresni, ahol az f_i függvények az f függvény koordinátái. A pozitív w_i súlyok tetszés szerint választhatók, de alkalmas megválasztásuk a minimumfeladat megoldását megkönnyítheti.

* **2.2.53. Nelder–Mead-módszer minimumfeladatokra.** Ez egy valós értékű $\Phi \in \mathbb{R}^k \to \mathbb{R}$ függvény minimumának megkeresésére alkalmas heurisztikus módszer. Minden lépésben az $x_0, x_1, \dots, x_k \in \mathbb{R}^k$ pontokat módosítja, amelyek nincsenek egy hipersíkban, így egy úgynevezett szimplexet (kétdimenzióban háromszöget, három dimenzióban tetraédert, stb.) adnak meg, és amelyeket úgy indexelünk, hogy $\Phi(x_0) \le \Phi(x_1) \le \dots \le \Phi(k)$ teljesüljön. Ha adott $\varepsilon > 0$ tűrésre

$$\frac{1}{n}\sum_{j=0}^{k}\big(\Phi(x_j)-\overline{\Phi}\big)^2 < \varepsilon^2,$$

(ahol $\overline{\Phi}$ a $\Phi(x_j)$ értékek számtani közepe), akkor megállunk, az eredmény az x_j-k számtani közepe. Egyébként x_k-t tükrözzük a többi pont $\overline{x}_k$ számtani közepére, pontosabban legyen $y=\overline{x}_k+\alpha(\overline{x}_k-x_k)$; az α *tükrözési paraméter* rendszerint 1. Ha $\Phi(x_0) \le \Phi(y) \le \Phi(x_{k-1})$, akkor x_k-t kicseréljük y-nal. Ha még $\Phi(y) < \Phi(x_0)$ is teljesül, akkor kiszámítjuk $\Phi(z)$ értékét, ahol $z=\overline{x}_k+\beta(\overline{x}_k-x_k)$; β a *kiterjesztési paraméter*, rendszerint 2. Az x_k pontot $\Phi(z) < \Phi(y)$ esetén z-re, egyébként y-ra cseréljük.

Ha $\Phi(y) > \Phi(x_{k-1})$, akkor a szimplexet összehúzzuk: legyen $w=x_k+\gamma(\overline{x}_k-x_k)$, ahol a γ *összehúzási paraméter* rendszerint $1/2$. Ha $\Phi(w) < \Phi(x_k)$, az x_k pontot kicseréljük w-vel; egyébként az x_0 kivételével az összes x_j pontot az $x_0+\delta(x_j-x_0)$ ponttal helyettesítjük, ahol a δ *redukciós paraméter* rendszerint $1/2$.

* **2.2.54. Az iránymenti csökkentés módszere.** Lényege, hogy a $\Phi \in \mathbb{R}^k \to \mathbb{R}$ függvény minimumhelyét úgy keressük, hogy minden x_n közelítéshez meghatározunk egy e_n irányt, amerre a függvény értéke tovább csökken, és a közelítést ebbe az irányba mozdítjuk el: $\Delta x_n = \alpha_n e_n$. Ha egy bázis vektorait ($\pm$ előjellel) választjuk e_n-nek úgy, hogy az utolsó bázisvektor után újrakezdjük az elsővel, akkor a *ciklikus koordinátánkénti csökkentés módszerét* kapjuk. Ha a deriválható Φ egy nyílt részhalmazon van értelmezve, és minden lépésben $e_n=-\nabla\Phi(x_n)$, a legmeredekebb csökkenés iránya, akkor a *gradiens módszert* kapjuk. Ha $e_n=-\nabla\Phi(x_n)+\beta_{n-1}e_{n-1}$ alkalmas β_{n-1} konstanssal, akkor a

*konjugált gradiens módszer*t kapjuk: az egyik változatnál $\beta_{-1} = \beta_0 = 0$, $e_{-1} = 0$ és

$$\beta_n = \frac{\left|\nabla\Phi(x_n)\right|^2}{\left|\nabla\Phi(x_{n-1})\right|^2}, \quad \text{ha} \quad n > 0,$$

míg a másik változatnál $\beta_{-1} = \beta_0 = 0$, $e_{-1} = 0$ és

$$\beta_n = \frac{\left\langle \nabla\Phi(x_n), \nabla\Phi(x_n) - \nabla\Phi(x_{n-1})\right\rangle}{\left|\nabla\Phi(x_{n-1})\right|^2}, \quad \text{ha} \quad n > 0.$$

Az e_n kijelölése után az elmozdulás mértékének meghatározásához az egyváltozós $g(t) = \Phi(x_n + te_n)$ függvényt kell minimalizálnunk. (Ha sikerül, a kapott pontban a gradiens merőleges lesz e_n-re. Ez az észrevétel adja a konjugált gradiens módszerek alapját: a $-\nabla\Phi(x_n)$ irányba minimalizálva, utána tudnánk még tovább csökkenteni az e_{n-1} irányba; a konjugált gradiens módszernél ezt próbáljuk „megelőlegezni".) A $t \mapsto g(t)$ függvényt csak közelítőleg fogjuk minimalizálni. Ez történhet a következőképpen. A $t_0 = 0$, t_1 és t_2 pontokban meghatározzuk a g függvény értékét, majd Lagrange-interpolációval meghatározzuk azt a $p(t)$ másodfokú polinomot, amely ezekben a pontokban megegyezik $g(t)$-vel. A t_1 értékének jó megválasztásához az előző lépések adhatnak útmutatást, úgy választjuk, hogy $g(t_1) < g(t_0)$ teljesüljön; t_2 választása függhet $g(t_1)$-től is. A gradiens és a konjugált gradiens módszereknél a másodfokú polinomot inkább úgy választjuk, hogy a $t_0 = 0$ helyen értéke $\Phi(x_n)$, deriváltja $\partial_{e_n}\Phi(x_n)$ legyen, így t_2-re nincs szükség. A t^* közelítést úgy kapjuk, hogy vesszük p minimumhelyét. (Arra is gondolnunk kell, hogy a másodfokú polinomban a másodfokú tag együtthatója nempozitív is lehet.) Bár valamelyik pontot elhagyva, helyette a t^* pontot tekintve, ez a lépés megismételhető, általában nem érdemes g minimumhelyét túl pontosan meghatározni, hanem t^* meghatározása után $\alpha_n = t^*$ választással áttérhetünk az $x_{n+1} = x_n + \alpha_n e_n$ pontra. Az alulrelaxálás, azaz az adott lépésben optimális t^*-nál kisebb α_n választása általában növeli a módszer hatékonyságát, mert a cikkcakk pálya kisimul. Az alulrelaxálást például α_n olyan választásával biztosíthatjuk, amelyre az

$$\frac{\Phi(x_n + \alpha_n e_n) - \Phi(x_n)}{\alpha_n}$$

differenciahányados és a $\partial_{e_n}\Phi(x_n)$ irány menti derivált hányadosa σ és $1 - \sigma$ közé esik valamely $0 < \sigma < 1/2$ értékre. (Rendszerint $10^{-5} \le \sigma \le 10^{-1}$.) Az alsó korlát kiküszöböli a túl hosszú, a felső pedig a túl rövid lépést. A túl rövid lépés kiküszöbölésére azt is megkövetelhetjük, hogy a $\partial_{e_n}\Phi(x_n + \alpha_n e_n)$ és $\partial_{e_n}\Phi(x_n)$ irány menti deriváltak hányadosának abszolút értéke, vagy legalább a hányados kisebb legyen, mint valamely γ érték, ahol $\sigma < \gamma < 1$. (Rendszerint $0{,}1 \le \gamma \le 0{,}5$.) Az alkalmas α_n értéket a $t^*\varrho^k$, $k \in \mathbb{Z}$ értékek között keressük, $k = 0$-val kezdve, ahol $0 < \varrho < 1$.

A konvergencia általában elég lassú, és semmilyen garancia nincs arra, hogy bármelyik módszer egy abszolút minimumhoz konvergál, bár konvergenciahalmaza általában bővebb, mint a Newton-módszeré. Ezért szokás a közelítés kezdetén az iránymenti csökkentések módszerét alkalmazni, majd áttérni a gyorsabb Newton-módszerre.

* **2.2.55. Példa.** Az előző pont jelöléseivel, tekintsük azt a tipikus esetet, amikor egy x^* lokális minimumhelyen a második deriváltra, amely szimmetrikus bilineáris forma, a vele képzett kvadratikus forma pozitív definit. Ekkor a Taylor-formula szerint

$$\Phi(x) \approx \frac{1}{2}\Phi''(x^*)(x-x^*, x-x^*) + \Phi(x^*).$$

Vizsgáljuk a különböző módszereket a

$$\varphi(x) = \frac{1}{2}\Phi''(x^*)(x-x^*, x-x^*) + \Phi(x^*)$$

függvényre. Mivel bármely B szimmetrikus bilineáris formára rögzített y-ra $x \mapsto B(x,y)$ lineáris funkcionál, előáll $\langle x, Ay\rangle$ alakban. Mivel valós térben B a másik változójában is lineáris, A lineáris operátor. Mivel B szimmetrikus, $\langle x, Ay\rangle = \langle y, Ax\rangle = \langle Ax, y\rangle = \langle x, A^*y\rangle$, amiből A önadjungált. Jelölje A a $\Phi''(x^*)$-hoz tartozó önadjungált operátort, amely nyilván pozitív definit is. Ezzel

$$\begin{aligned}\varphi(x) &= \frac{1}{2}\big\langle A(x-x^*), x-x^*\big\rangle + \Phi(x^*) = \frac{1}{2}\langle Ax, x\rangle - \langle Ax^*, x\rangle + \frac{1}{2}\langle Ax^*, x^*\rangle + \Phi(x^*)\\ &= \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c,\end{aligned}$$

ahol $b = Ax^*$ és $c = \frac{1}{2}\langle Ax^*, x^*\rangle + \Phi(x^*)$. A minimalizálással az $Ax = b$ lineáris egyenletrendszert oldjuk meg.

Bevezetve az $r = Ax - b$ jelölést, $A(x - x^*) = r$, és bármely e irányra a $g(t) = \varphi(x + te)$ függvényre

$$g(t) - g(0) = t\langle Ax, e\rangle - t\langle b, e\rangle = t\langle Ax - b, e\rangle = t\langle r, e\rangle.$$

Innen átosztva, $\partial_e\varphi(x) = \langle Ax - b, e\rangle = \langle r, e\rangle$ és így $\nabla\varphi(x) = Ax - b = r$. Hasonlóan kiszámolható, hogy $g'(t) = 0$ akkor teljesül, ha $t = -\langle r, e\rangle/\langle Ae, e\rangle$. Legyen $x = x_n$, $e = e_n$, $r_n = Ax_n - b$ választással α_n ez a t, azaz $\alpha_n = -\langle r_n, e_n\rangle/\langle Ae_n, e_n\rangle$. A hiba méréséhez tekintsük az új $(x, y) \mapsto \langle Ax, y\rangle = \langle x, y\rangle_A$ belső szorzatot és a belőle származó $\|x\|_A$ normát. Ezzel

$$\begin{aligned}\|x_{n+1} - x^*\|_A^2 &= \|x_n - x^* + \alpha_n e_n\|_A^2 = \|x_n - x*\|_A^2 + 2\alpha_n\langle x_n - x^*, e_n\rangle + \alpha_n^2\|e_n\|_A^2\\ &= \Big(1 - \frac{\langle r_n, e_n\rangle^2}{\langle Ae_n, e_n\rangle\langle A^{-1}r_n, r_n\rangle}\Big)\|x_n - x^*\|_A^2,\end{aligned}$$

mivel $r_n = Ax_n - b$ jelöléssel $\|x_n - x^*\|_A^2 = \langle A^{-1}r_n, r_n\rangle$. A gradiens-módszernél

$$\begin{aligned}\|x_{n+1} - x^*\|_A^2 &= \|x_n - x^* + \alpha_n e_n\|_A^2\\ &= \Big(1 - \frac{\|r_n\|^4}{\langle Ar_n, r_n\rangle\langle A^{-1}r_n, r_n\rangle}\Big)\|x_n - x^*\|_A^2,\end{aligned}$$

azaz lineáris konvergenciát várhatunk.

Megmutatjuk, hogy a konjugált gradiens módszer két változata ebben a példában egybeesik. Az jellemzi őket, hogy $\langle e_n, e_{n-1}\rangle_A = \langle -r_{n-1} + \beta_{n-1}e_{n-1}, e_{n-1}\rangle_A = 0$. Ehhez az kell, hogy $\beta_{n-1} = \langle Ar_n, e_{n-1}\rangle / \langle Ae_{n-1}, e_{n-1}\rangle$ teljesüljön. Indukcióval megmutatjuk, hogy a fenti α_n és β_{n-1} választással, ha $r_0, r_1, \dots, r_n$ nem nullák, akkor $e_0, e_1, \dots, e_n$ sem nullák és ugyanazt az alteret feszítik ki, továbbá $\beta_0, \beta_1, \dots, \beta_{n-1}$ és $\alpha_0, \alpha_1, \dots, \alpha_n$ értelmezve vannak, az utóbbiak nem nullák,

$$\alpha_n = -\frac{\|r_n\|^2}{\|e_n\|_A^2} \quad \text{és} \quad \beta_{n-1} = \frac{\langle r_n, r_n - r_{n-1}\rangle}{\|r_{n-1}\|^2} = \frac{\|r_n\|^2}{\|r_{n-1}\|^2},$$

valamint $\langle e_j, e_n\rangle_A = 0$ és $\langle e_j, r_n\rangle = 0$, ha $j < n$. Az $n = 0$ esetben az állítás teljesül ($\beta_{-1} = \beta_0 = 0$, $e_{-1} = 0$ miatt $e_0 = -r_0$). Tegyük fel, hogy $r_{n+1} \neq 0$. Mivel az e_n irányban minimalizáltunk, $e_n \perp r_{n+1}$, azaz $\langle e_n, r_{n+1}\rangle = 0$. A β_n definíciója szerint $\langle e_{n+1}, e_n\rangle_A = 0$. (Nyilván β_n értelmezve van.) Mivel $e_{n+1} = -r_{n+1} + \beta_n e_n$, az r_{n+1} kombinálható e_{n+1}-ből és e_n-ből, és e_{n+1} kombinálható r_{n+1}-ből és e_n-ből. Az $x_{j+1} = x_j + \alpha_j e_j$ összefüggésből $r_{j+1} = r_j + \alpha_j Ae_j$, amiből $j < n$ esetén egyrészt

$$\langle e_j, r_{n+1}\rangle = \langle e_j, r_n\rangle + \alpha_n \langle e_j, Ae_n\rangle = 0,$$

másrészt

$$\langle Ae_j, e_{n+1}\rangle = \langle Ae_j, r_{n+1}\rangle + \beta_n \langle Ae_j, e_n\rangle = \langle Ae_j, r_{n+1}\rangle = 0,$$

mivel $\alpha_j \neq 0$, így Ae_j benne van az $r_0, \dots, r_n$, vagyis az $e_0, \dots, e_n$ által kifeszített altérben. Ha $e_{n+1} = 0$ lenne, akkor $r_{n+1} = \beta_n e_n$ teljesülne, ami ellentmond annak, hogy r_{n+1} nem nulla, és merőleges e_n-re. Így α_{n+1} értelmezve van és e_{n+1} definíciójából

$$\alpha_{n+1} = -\frac{\langle e_{n+1}, r_{n+1}\rangle}{\|e_{n+1}\|_A^2} = \frac{\langle r_{n+1} - \beta_n e_n, r_{n+1}\rangle}{\|e_{n+1}\|_A^2} = \frac{\|r_{n+1}\|^2}{\|e_{n+1}\|_A^2} \neq 0.$$

Mivel $x_{n+1} = x_n + \alpha_n e_n$-ből $r_{n+1} = r_n + \alpha_n Ae_n$, azt is kapjuk, hogy

$$\begin{aligned}
\beta_n &= \frac{\langle e_{n+1}, r_{n+1}\rangle}{\langle Ae_n, e_n\rangle} = \frac{\langle r_{n+1} - r_n, r_{n+1}\rangle}{\alpha_n \langle Ae_n, e_n\rangle} \\
&= \frac{\langle r_{n+1} - r_n, r_{n+1}\rangle}{\langle r_{n+1} - r_n, e_n\rangle} = -\frac{\langle r_{n+1} - r_n, r_{n+1}\rangle}{\langle r_n, e_n\rangle} \\
&= -\frac{\langle r_{n+1} - r_n, r_{n+1}\rangle}{\langle r_n, -r_n + \beta_{n-1}e_{n-1}\rangle} = \frac{\langle r_{n+1} - r_n, r_{n+1}\rangle}{\|r_n\|^2} = \frac{\|r_{n+1}\|^2}{\|r_n\|^2};
\end{aligned}$$

az utolsó egyenlőség azért teljesül, mert r_n benne van az $e_0, \dots, e_n$ által generált altérben.

Vegyük észre hogy a konjugált gradiens módszer véges sok lépésben konvergál, mert $e_0, \dots, e_n$ ortogonálisak az $(x, y) \mapsto \langle x, y\rangle_A$ belső szorzatra, így $r_j = 0$ kell legyen valamely $j \leq k$-ra, hiszen $\mathbb{R}^k$-ban vagyunk.

* **2.2.56. Kvázi Newton-módszerek minimalizálásra.** Mint már említettük, minimalizáláskor a lokális minimumhely közelében érdemes áttérni egy Newton-típusú módszerre. Ez azt jelenti, hogy Φ minimuma helyett $f = \nabla\Phi$ zérushelyét keressük. Bármelyik módszer használható, de javítások is lehetségesek. A gyakran használt Broyden-módszert például úgy szokás megváltoztatni, hogy az iteráció során A_n szimmetrikus legyen, mert $\Phi''(x_n)$ szimmetrikus bilineáris forma. Az egyik szokásos választásnál $\Delta A_n = A_{n+1} - A_n$ mátrixa a szokásos bázisban

$$\frac{\big[\Delta f(x_n) - A_n\Delta x_n\big]\big[\Delta f(x_n) - A_n\Delta x_n\big]'}{\big\langle \Delta f(x_n) - A_n\Delta x_n, \Delta x_n\big\rangle}.$$

Nem tanácsos tovább folytatni az iterációt, ha a $\Delta f(x_n) - A_n\Delta x_n$ és Δx_n által bezárt szög koszinusza túl kicsi lesz, mondjuk 10^{-8} alá megy. Egyébként erre a formulára $A_{n+1}^{-1} - A_n^{-1} = \Delta A_n^{-1}$ mátrixa a szokásos bázisban

$$\frac{\big[\Delta x_n - A_n^{-1}\Delta f(x_n)\big]\big[\Delta x_n - A_n^{-1}\Delta f(x_n)\big]'}{\big\langle \Delta x_n - A_n^{-1}\Delta f(x_n), \Delta f(x_n)\big\rangle}.$$

Még elterjedtebb az a változat, amelyben ΔA_n mátrixa a szokásos bázisban

$$\frac{\big[\Delta f(x_n)\big]\big[\Delta f(x_n)\big]'}{\big\langle \Delta f(x_n), \Delta x_n\big\rangle} - \frac{\big[A_n\Delta x_n\big]\big[A_n\Delta x_n\big]'}{\big\langle \Delta x_n, A_n\Delta x_n\big\rangle}.$$

Erre a formulára ΔA_n^{-1} mátrixa a szokásos bázisban

$$\begin{aligned}&\frac{\big\langle \Delta x_n, \Delta f(x_n)\big\rangle + \big\langle \Delta f(x_n), A_n^{-1}\Delta f(x_n)\big\rangle}{\big\langle \Delta x_n, \Delta f(x_n)\big\rangle^2}\big[\Delta x_n\big]\big[\Delta x_n\big]' \\ &- \frac{\big[A_n^{-1}\Delta f(x_n)\big]\big[\Delta x_n\big]' + \big[\Delta x_n\big]\big[\Delta f(x_n)\big]'\big[A_n^{-1}\big]}{\big\langle \Delta x_n, \Delta f(x_n)\big\rangle}.\end{aligned}$$

* **2.2.57. Feladat [10].** A $\Phi(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ függvény minimalizálásán hasonlítsunk össze különböző módszereket.

* **2.2.58. Feladat [10].** A $\Phi(x, y) = (x^2+y-11)^2+(y^2+x-7)$ függvénynek keressük meg a négy abszolút minimumát és az egyetlen lokális maximumát, összehasonlítva a különböző módszereket.

* **2.2.59. Feladat [11].** A

$$\Phi(x_1, x_2, \dots, x_n) = \sum_{j=1}^{n-1}\big((1 - x_j)^2 + 100\varepsilon_j(x_{j+1} - x_j^2)^2\big)$$

függvény — ahol $\varepsilon_1, \dots \varepsilon_{n-1} \in \,]0, 1]$ szabad paraméterek — minimalizálásán hasonlítsunk össze különböző módszereket. (A függvénynek egy globális minimuma van, de még ha minden ε_j egy, akkor is már $4 \le n \le 7$ esetén van egy lokális minimuma a $(-1, 1, 1, \dots, 1)$ pont közelében.)

* **2.2.60. Algebrai egyenletek megoldása.** A Newton-módszer minden további nélkül alkalmazható komplex változós, komplex értékű f függvényre is. Hogy a konvergenciát biztosabbá tegyük, kezdetben alkalmazhatjuk a gradiens módszert az $|f|$ függvényre, majd a Newton-módszert. Mivel a Taylor-formulából $f(x+h) \approx f(x) + f'(x)h$, az abszolút érték akkor nő a leggyorsabban, ha $f(x)$ és $f'(x)h$ iránya megegyezik, azaz ha $f(x) = r(h)f'(x)h$ valamely $r(h) \geq 0$-ra. Innen az $|f|$ gradiense $f(x)/f'(x)$ irányú, ha $f(x) \neq 0$ és $f'(x) \neq 0$. Ha $r(h)h = f(x)/f'(x)$, $r(h) > 0$, akkor $\big|f(x+h)\big| - \big|f(x)\big| \approx \big|f'(x)\big||h|$, ahonnan átosztva $|h| \to 0$ határátmenettel $\big|\nabla|f|(x)\big| = \big|f'(x)\big|$. Összegezve

$$\nabla|f|(x) = \frac{f(x)\big|f'(x)\big|^2}{f'(x)\big|f(x)\big|}.$$

Így az $x_{n+1} = x_n - tf(x_n)/f'(x_n)$ alakban kell keresnünk a következő közelítést, ahol t pozitív valós szám. Eljárhatunk úgy, hogy először $t = 1$-el próbálkozunk, ami a Newton-módszernek felel meg, majd a gradiens-módszernél leírtak szerint változtatjuk t-t. Az iteráció csak akkor szakad meg, ha valamelyik lépésben $f'(x_n) = 0$.

Polinomokra alkalmazva ezt az eljárást, meghatározhatjuk a polinom gyökeit. Megjegyezzük, hogy még ha a polinom valós együtthatós, akkor is ajánlatos egy véletlen komplex kezdőértékből indítani az iterációt, mert így nagyon kicsi a valószínűsége, hogy megszakad, és újra kell indítani. Ezzel az eljárással megkaphatjuk a polinom egy zérushelyét. A polinomot elosztva a gyöktényezővel, eggyel alacsonyabb fokú polinomot kapunk, amit ugyanúgy kezelhetünk. A számítási hibák felhalmozódása miatt a talált közelítő értékek pontosságát érdemes az eredeti polinom felhasználásával végzett iterációval növelni.

2.3. Integrálszámítás

2.3.1. Többváltozós függvények integrálja. Egy $\mathbb{R}^m$-beli *m-dimenziós tégla* alatt egydimenziós $T_j = [a_j, b_j]$, $-\infty < a_j < b_j < +\infty$ korlátos, zárt, pozitív hosszúságú intervallumok

$$T = T_1 \times T_2 \times \cdots \times T_m$$

Descartes-szorzatát értjük. A T tégla $|T|$ *mérték*e az oldalhosszak szorzata, $\prod_{j=1}^m (b_j - a_j)$. A T tégla egy *felosztás*án tégláknak egy olyan $T_1, T_2, \dots T_k$ véges sorozatát érjük, amelyek egyesítése T, és amelyeknek páronként nincs közös belső pontjuk.

Egy P *pontozott felosztás* alatt egy felosztást értünk, amelynek minden T_j résztéglájához meg van adva egy $t_j \in T_j$ pont, azaz $P = \big\{(T_j, t_j) : t_j \in T_j, j = 1, 2, \dots, k\big\}$. Legyen $\delta : T \to \mathbb{R}^+$ egy tetszőleges függvény. Azt mondjuk, hogy a T egy (T_j, t_j) pontozott résztéglája δ-finom, ha $T_j \subset \mathbb{U}_{\delta(t_j)}(t_j)$; a P pontozott felosztás *δ-finom*, ha a (T_j, t_j) résztéglák mind δ-finomak, ha $j = 1, 2, \dots, k$. Legyen $f : T \to \mathbb{K}^n$ egy függvény. A T tégla P pontozott felosztásához hozzárendeljük az

$$s(f, P) = \sum_{j=1}^{k} |T_j| f(t_j)$$