Nemlineáris programozás

- Feltételmentes optimalizálás: egyenes menti keresés, a keresési tartomány fogalma, konvex függvények szélsőértékeinek tulajdonságai, nemdifferenciálható függvények minimumának megadása dichotóm kereséssel, differenciálható függvények minimumának megadása bináris kereséssel
- Többdimenziós feltételmentes minimumkeresés, a legmeredekebb irányok módszere
- Nemlineáris programok általános megoldása, belső és külső büntetőfüggvények alkalmazása, a büntetőfüggvények leggyakoribb megválasztása, a két módszer összehasonlítása

Emlékeztető: megengedett javítóirányok

- Adott a $\min f(\boldsymbol{x}) : \boldsymbol{x} \in X$ nemlineáris program, ahol $X = \big\{\boldsymbol{x} : g_i(\boldsymbol{x}) \leq 0, i \in \{1, \ldots, m\}\big\}$ és f és g_i folytonosan differenciálható az egész megengedett tartományon
- Felhasználjuk a tényt, hogy ha $\bar{\boldsymbol{x}}$ lokális minimum, akkor $\bar{\boldsymbol{x}}$-ben nincs $\boldsymbol{d}$ megengedett javítóirány, melyre $\exists\delta > 0$ hogy

$$\forall\lambda \in (0, \delta) : \quad \begin{aligned} & f(\bar{\boldsymbol{x}} + \lambda\boldsymbol{d}) < f(\bar{\boldsymbol{x}}) \\ & \bar{\boldsymbol{x}} + \lambda\boldsymbol{d} \in X \end{aligned}$$

- Csak szükséges feltétel, de például konvex program esetén elégséges is
- A nemlineáris programot megengedett javítóirányok mentén tett iteratív lépésekkel oldjuk meg

Emlékeztető: megengedett javítóirányok

- Keressünk pontot, melyben nincs megengedett javítóirány
- Legyen $J \subseteq I$ az $\bar{\boldsymbol{x}}$ pontban éles feltételek halmaza

$$\forall i \in J : g_i(\bar{\boldsymbol{x}}) = 0, \quad \forall i \in I \setminus J : g_i(\bar{\boldsymbol{x}}) < 0$$

- Például megengedett javítóirány a következő (ha létezik)

$$\boldsymbol{d} : \nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{d} < 0,\ \nabla g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} < 0 \quad \forall i \in J$$

- Ilyen $\boldsymbol{d}$ irány például az alábbi optimalizálási probléma megoldásával kereshető

$$\min \nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{d} :\ |\boldsymbol{d}| \leq 1,\ \nabla g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} < 0 \quad \forall i \in J$$

- Lineáris célfüggvény, lineáris/linearizálható feltételek
- Szukcesszív lineáris programozás

Emlékeztető: egyenes menti keresés

- Ha az optimum ~ 0, akkor jó eséllyel $\bar{\boldsymbol{x}}$ lokális minimum
  - patologikus esetekben a feltétel sajnos nem elégséges
- Ha az optimum negatív, akkor $\boldsymbol{d}$ megengedett javítóirány
- Keresés a $\bar{\boldsymbol{x}} + \lambda \boldsymbol{d} : \lambda > 0$ félegyenes mentén úgy, hogy
  - ne lépjünk ki a megengedett tartományból
  - a célfüggvény minimumát keressük $\boldsymbol{d}$ irányban
- Egyváltozós nemlineáris program, ahol a megengedett tartomány egy egyszerű intervallum (vagy $\mathbb{R}$)

$$\min f(\bar{\boldsymbol{x}} + \lambda \boldsymbol{d}) : \bar{\boldsymbol{x}} + \lambda \boldsymbol{d} \in X, \lambda \geq 0$$

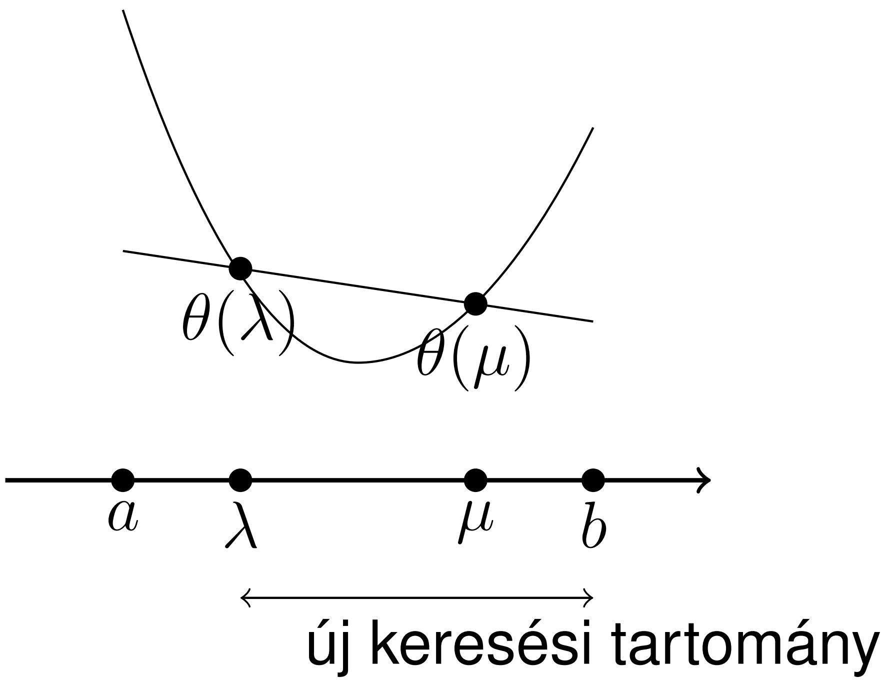
- Sok algoritmus működik a fenti séma szerint: találj egy javítóirányt majd egyenes menti keresés
- Hatékony egyenes menti kereső algoritmus szükséges

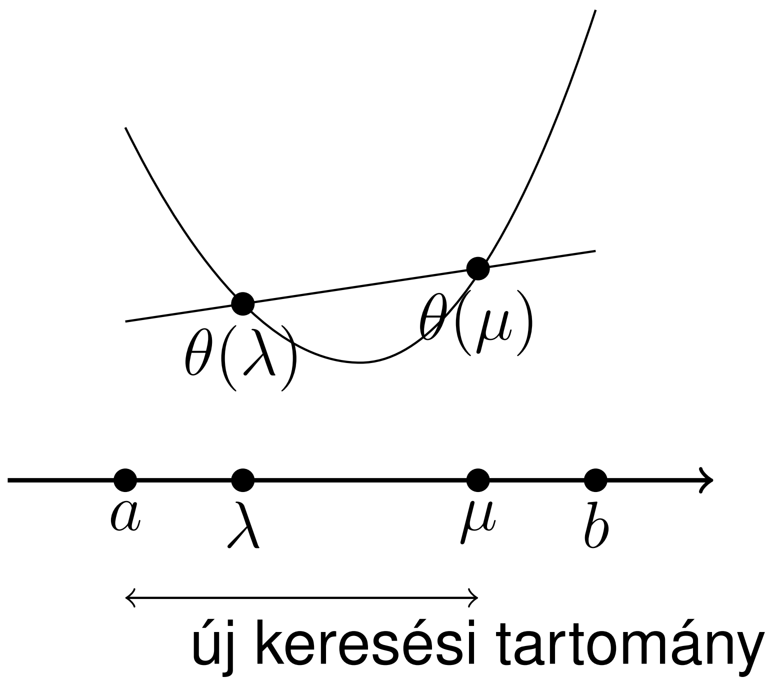
Egyenes menti keresés

- Keressük a $\min \theta(\lambda) : \lambda \in [a, b]$ **konvex** program optimális megoldását
  - csak konvex függvényekkel foglalkozunk
  - de az ötletek nemkonvex függvényekre is általánosíthatók
- A $[a, b]$ intervallum neve **keresési intervallum**
- Az algoritmusok ötlete, hogy a keresési intervallum bizonyos részeiről belátjuk, hogy nem tartalmazzák a minimumot, így ezeket elhagyjuk
- Ehhez felhasználjuk a konvex függvények tulajdonságait
- Először egy nemdifferenciálható konvex függvényeken is működő algoritmust mutatunk

Dichotóm keresés

- **Tétel:** legyen θ az $[a,b]$ keresési intervallumon konvex $\mathbb{R} \mapsto \mathbb{R}$ függvény, és legyen $\lambda, \mu \in [a,b]$ úgy, hogy $\lambda < \mu$. Ekkor
  - ha $\theta(\lambda) > \theta(\mu)$, akkor $\forall z \in [a, \lambda) : \theta(z) \geq \theta(\mu)$
  - ha $\theta(\lambda) < \theta(\mu)$, akkor $\forall z \in (\mu, b] : \theta(z) \geq \theta(\lambda)$





Dichotóm keresés

- **Biz.:** tegyük fel először, hogy $\theta(\lambda) > \theta(\mu)$ és legyen $z \in [a, \lambda)$
- Tegyük fel, hogy $\theta(z) < \theta(\mu)$, és lássuk be, hogy így ellentmondásra jutunk
- Mivel $z < \lambda < \mu$, ezért λ felírható z és μ konvex kombinációjaként
- A konvexitás miatt $\theta(\lambda) \leq \max\{\theta(z), \theta(\mu)\} = \theta(\mu)$, ami ellentmond a feltevésnek, hogy $\theta(\lambda) > \theta(\mu)$
- A másik eset hasonlóan látható be $\square$
- A tétel felhasználható arra, hogy két mintavétel után a keresési tartomány egy részét elhagyjuk

Dichotóm keresés

- Például ha valamely $\lambda < \mu$ pontokra $\theta(\lambda) > \theta(\mu)$, akkor az $[a, \lambda)$ pontokban már biztos nem lehet θ minimuma
- Az új keresési tartomány tehát $[\lambda, b]$
- Ha ellenben $\theta(\lambda) < \theta(\mu)$ adódik, akkor a $(\mu, b]$ intervallum elhagyható, és az új keresési tartomány $[a, \mu]$
- Fennállhat még $\theta(\lambda) = \theta(\mu)$, de erre a gyakorlatban nagyon kicsi az esély (például a véges pontosságú számábrázolás miatt)
- Az új keresési tartomány nagysága vagy $b - \lambda$, vagy $\mu - a$
- Azt szeretnénk, ha az új keresési tartomány nagysága $\max\{b - \lambda, \mu - a\}$ a legkisebb legyen
- Ehhez a $\lambda = \mu = \frac{1}{2}(a + b)$ választás lenne a legjobb
- Nekünk viszont két, egymástól különböző λ és μ pont kell

Dichotóm keresés

- A megoldás, hogy az ideális $\frac{1}{2}(a+b)$ pont körül szimmetrikusan ϵ távolságra mérünk

$$\lambda = \tfrac{1}{2}(a+b) - \epsilon, \qquad \mu = \tfrac{1}{2}(a+b) + \epsilon$$

- Az ϵ konstanst úgy választjuk meg, hogy
  - a két mérési pont λ és μ már kellően távol legyen ahhoz, hogy $\theta(\lambda)$ és $\theta(\mu)$ számottevően különbözzön
  - de a kapott új intervallum $\frac{1}{2}(a+b)+\epsilon$ hossza ne nőjön meg túlzottan az optimális $\frac{1}{2}(a+b)$ értékről
- Miután meghatároztuk λ és μ értékét, kiértékeljük a függvényt ezekben a pontokban, és $\theta(\lambda)$ és $\theta(\mu)$ értéke szerint új keresési tartományt kapunk
- A keresés vége, ha keresési tartomány már elég kicsi

Dichotóm keresés

- Konvex θ függvény minimalizálása egy $[a, b]$ keresési tartományban
- Inicializálás: válasszunk $\epsilon > 0$ konstanst és egy $l > 0$ keresési pontosságot. Legyen $a_1 = a$, $b_1 = b$, a kezdeti keresési tartomány $[a_1, b_1]$ és legyen $k = 1$

1. Ha $b_k - a_k < l$, akkor kiszállás: a minimum pont az $[a_k, b_k]$ tartományban van

- Ellenkező esetben állítsuk be λ_k és μ_k értékét az alábbiak szerint

$$\lambda_k = \frac{1}{2}(a_k + b_k) - \epsilon, \qquad \mu_k = \frac{1}{2}(a_k + b_k) + \epsilon$$

2. Ha $\theta(\lambda_k) < \theta(\mu_k)$, akkor $a_{k+1} = a_k$ és $b_{k+1} = \mu_k$, ellenkező esetben $a_{k+1} = \lambda_k$ és $b_{k+1} = b_k$

- $k = k + 1$ és vissza az 1. lépésre

Dichotóm keresés: példa

- A keresési tartomány hossza a k-adik lépésben

$$b_k - a_k = \frac{1}{2^k}(b-a) + 2\epsilon\left(1 - \frac{1}{2^k}\right)$$

- Ebből megadható az adott l pontosság eléréséhez szükséges lépésszám
- Például a $\theta(\lambda) = \lambda^2$ függvény minimuma a $[-5, 15]$ tartományon $\epsilon = 10^{-2}$ és $l = 2\epsilon$ választás mellett

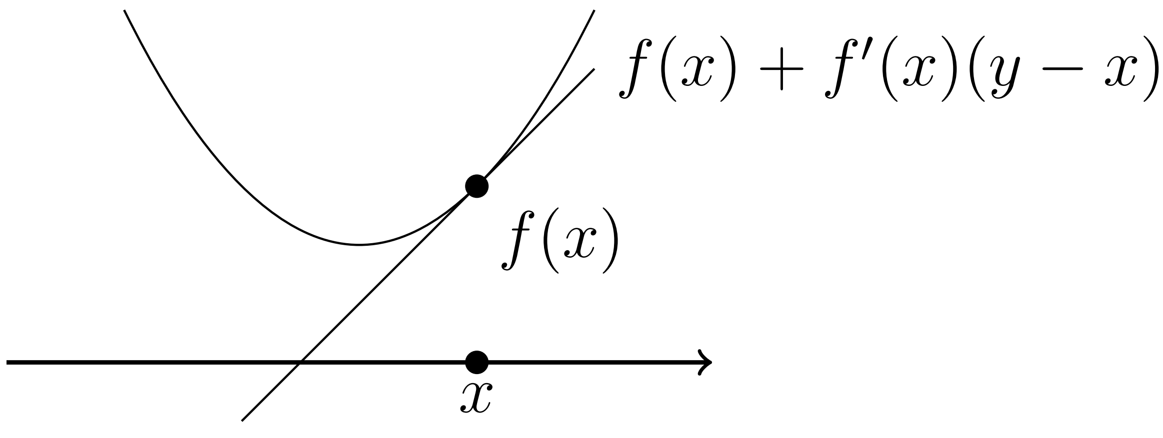
k	a	b	λ	μ	$\theta(\lambda)$	$\theta(\mu)$	$b-a$
1	-5.0000	15.0000	4.9900	5.0100	24.9001	25.1001	20.0000
2	-5.0000	5.0100	-0.0050	0.0150	0.0000	0.0002	10.0100
3	-5.0000	0.0150	-2.5025	-2.4825	6.2625	6.1628	5.0150
4	-2.5025	0.0150	-1.2538	-1.2338	1.5719	1.5221	2.5175
$\vdots$	...	...	...	...	...	...	...
18	-0.0100	0.0101	-0.0100	0.0100	0.0001	0.0001	0.0201

Differenciálható függvény minimalizálása

- Ha a minimalizálandó függvény nemcsak konvex, de differenciálható is, akkor az egyenes menti keresés egyszerűbb
- **Tétel:** egy folytonosan differenciálható f függvény akkor és csak akkor konvex egy X tartományon, ha

$$\forall \boldsymbol{x}, \boldsymbol{y} \in X : \quad f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) \tag{1}$$

- Egy $\boldsymbol{x}$ pontba húzott érintő alsó becslést f értékére X-en



Differenciálható függvény minimalizálása

- **Biz.:** először tegyük fel, hogy f konvex, majd belátjuk, hogy az (1) feltétel teljesül
- f konvex: az $f(\boldsymbol{x})$ és $f(\boldsymbol{y})$ közötti vonalszakasz a függvényt felülről becsli $\boldsymbol{x}$ és $\boldsymbol{y}$ között

$$\forall\lambda \in [0,1] : f(\lambda\boldsymbol{y} + (1-\lambda)\boldsymbol{x}) \leq \lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x})$$

- Használjuk fel, hogy $\mu a + (1-\mu)b = b + \mu(a-b)$

$$f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) \leq f(\boldsymbol{x}) + \lambda(f(\boldsymbol{y}) - f(\boldsymbol{x}))$$

- $f(\boldsymbol{x})$-et a bal oldalra rendezve és osztva λ-val

$$\frac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda} \leq f(\boldsymbol{y}) - f(\boldsymbol{x})$$

Differenciálható függvény minimalizálása

- Innen $\lambda \to 0$ közelítéssel a kívánt egyenlőtlenséget kapjuk

$$\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) \le f(\boldsymbol{y}) - f(\boldsymbol{x})$$

ugyanis f függvény $(\boldsymbol{y}-\boldsymbol{x})$ irány menti deriváltja

$$\lim_{\lambda\to 0}\frac{f(\boldsymbol{x}+\lambda(\boldsymbol{y}-\boldsymbol{x}))-f(\boldsymbol{x})}{\lambda} = \nabla^T f(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})$$

- Megfordítva: lássuk be, hogy f konvexitása következik az (1) feltételből
- Legyen $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$ az X halmaz két tetszőleges pontja és legyen $\boldsymbol{x}$ a két pont bármely konvex kombinációja

$$\boldsymbol{x} = \lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2, \quad \lambda \in [0,1]$$

Differenciálható függvény minimalizálása

- Írjuk fel (1) feltételt arra az esetre, ha $\boldsymbol{y} = \boldsymbol{x}_1$

$$f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{x}_1 - \boldsymbol{x}) \tag{2}$$

- Írjuk fel a feltételt arra az esetre is, ha $\boldsymbol{y} = \boldsymbol{x}_2$

$$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{x}_2 - \boldsymbol{x}) \tag{3}$$

- Szorozzuk be (2)-t λ-val és (3)-at $1 - \lambda$-val, majd a két egyenlőtlenséget összeadva

$$\begin{aligned} \lambda f(\boldsymbol{x}_1) &\geq \lambda f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^T (\boldsymbol{x}_1 - \boldsymbol{x}) \\ (1-\lambda) f(\boldsymbol{x}_2) &\geq (1-\lambda) f(\boldsymbol{x}) + (1-\lambda) \nabla f(\boldsymbol{x})^T (\boldsymbol{x}_2 - \boldsymbol{x}) \\ \hline \end{aligned}$$

$$\lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 - \boldsymbol{x})$$

Differenciálható függvény minimalizálása

- Mivel $\boldsymbol{x}$ az $\boldsymbol{x}_1$ és $\boldsymbol{x}_2$ pontok konvex kombinációja
  - egyrészt $f(\boldsymbol{x}) = f(\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2)$ helyettesíthető a jobb oldal első tagjában
  - másrészt a gradiens mögötti zárójeles tagra

$$\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 - \boldsymbol{x} =$$
$$\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 - (\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) = 0$$

- Így pont a konvexitás feltételét kapjuk

$$\lambda f(\boldsymbol{x}_1) + (1-\lambda)f(\boldsymbol{x}_2) \geq f(\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \qquad \square$$

- A konvex programozás egyik legfontosabb tétele

Differenciálható függvény minimalizálása

- **Következmény:** legyen f egy folytonosan differenciálható konvex függvény és legyen $\bar{\boldsymbol{x}} \in \mathbb{R}^n$. Ekkor $f(\bar{\boldsymbol{x}})$ akkor és csak akkor globális minimuma f-nek, ha $\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{0}$
- **Biz.:** korábban láttuk, hogy $\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{0}$ szükséges ahhoz, hogy $f(\bar{\boldsymbol{x}})$ minimum legyen
- Azonban általános esetben $\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{0}$ nem elégséges, mert inflexiós pontokban, illetve nyeregpontokban is teljesül
- Konvex függvényekre azonban a feltétel elégséges
- Tegyük fel, hogy $\nabla f(\bar{\boldsymbol{x}}) = \boldsymbol{0}$ és alkalmazzuk az iménti tételt $\bar{\boldsymbol{x}}$ pontra és egy tetszőlegesen választott $\boldsymbol{y}$ pontra

$$f(\boldsymbol{y}) \geq f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{y} - \bar{\boldsymbol{x}}) = f(\bar{\boldsymbol{x}}) + 0$$

- Tehát $f(\bar{\boldsymbol{x}})$ globális minimum $\square$

Differenciálható függvény minimalizálása

- Egyváltozós esetben a tételek lényegesen egyszerűsödnek
- A $\theta(\lambda) : \mathbb{R} \mapsto \mathbb{R}$ függvény akkor és csak akkor konvex, ha

$$\forall \mu, \nu : \quad \theta(\mu) \geq \theta(\nu) + \theta'(\nu)(\mu - \nu) \tag{4}$$

- A minimumpont feltétele pedig egyszerűen $\theta'(\lambda) = 0$
- A két tétel segítségével differenciálható f függvény minimuma gyorsabban kereshető
- A dichotóm keresés esetében a keresési tartomány megfelezéséhez minden iterációban két, eltérő pontban kellett kiértékelni a függvényt
- Ha azonban θ differenciálható, akkor elég egy pontban kiértékelni a deriváltját θ'-t
- Egyszerű bináris kereséssel megoldható a feladat

Differenciálható függvény minimalizálása

- Keressük a $\min \theta(\lambda) : \lambda \in [a, b]$ **konvex** program optimális megoldását, ahol θ **differenciálható** $[a, b]$ halmazon
- Tegyük fel, hogy valamely $\lambda \in [a, b]$ pontban ismerjük a derivált értékét, $\theta'(\lambda)$-t. Három eset lehetséges
  - $\theta'(\lambda) = 0$: ekkor a fentiek miatt θ a λ pontban éri el a minimumát
  - $\theta'(\lambda) > 0$: ekkor minden $\mu > \lambda$ esetén $\theta'(\lambda)(\mu - \lambda) > 0$, ezért (4) miatt $\theta(\mu) > \theta(\lambda)$ adódik. Következésképp a minimumpont nem lehet a $[\lambda, b]$ tartományban, így a keresési tartomány a $[a, \lambda]$ intervallumra szűkíthető
  - $\theta'(\lambda) < 0$: pont megfordítva, ebben az esetben $\mu < \lambda$ pontokra kapjuk a $\theta(\mu) > \theta(\lambda)$ becslést. Ugyanis $\theta'(\lambda)(\mu - \lambda) > 0$, mert $\theta'(\lambda) < 0$ és $\mu - \lambda < 0$. Az új keresési tartomány $[\lambda, b]$

Differenciálható függvény minimalizálása

- A λ pontot úgy kell kiválasztanunk, hogy az új keresési tartomány $\max\{\lambda - a, b - \lambda\}$ mérete a legrosszabb esetet feltételezve is a minimális legyen
- Könnyű látni, hogy ez az $[a, b]$ intervallum felezőpontjában következik be, vagyis $\lambda = \frac{1}{2}(a + b)$
- Egyszerű bináris keresés:
  - Inicializálás: válasszunk $l > 0$ keresési pontosságot. Legyen $a_1 = a$, $b_1 = b$, és legyen $k = 1$

1. legyen $\lambda_k = \frac{1}{2}(a_k + b_k)$ és tekintsük $\theta'(\lambda_k)$ értékét
2. ha $\theta'(\lambda) = 0$ vagy $b_k - a_k < l$, akkor kiszállás: a minimum pont az $[a_k, b_k]$ tartományban van
3. ha $\theta'(\lambda) > 0$, akkor $a_{k+1} = a_k$ és $b_{k+1} = \lambda_k$, különben $a_{k+1} = \lambda_k$ és $b_{k+1} = b_k$, és vissza az 1. lépésre

Differenciálható függvény minimalizálása

- A k-adik lépésben a keresési tartomány nagysága $\frac{1}{2^k}(b-a)$
- Ebből becsülhető az adott l keresési pontosság eléréséhez szükséges lépésszám
- Keressük a $\theta(\lambda) = \lambda^2 + 2\lambda$ függvény minimumát a $[-5, 15]$ tartományon, az $l = 2 \cdot 10^{-2}$ választás mellett

k	a	b	λ	$\theta'(\lambda)$	$\theta(\lambda)$	$b-a$
1	-5.0000	15.0000	5.0000	12.0000	35.0000	20.0000
2	-5.0000	5.0000	0.0000	2.0000	0.0000	10.0000
3	-5.0000	0.0000	-2.5000	-3.0000	1.2500	5.0000
$\vdots$	...	...	...	...	...	...
11	-1.0156	-0.9961	-1.0059	0.0118	-1.0000	0.0195

- A $\theta'(\lambda) = 2\lambda + 2 = 0$ egyenletet kielégítő pontot keressük, de $\theta'(\lambda) = 0$ sokszor nem oldható meg zárt alakban

Többdimenziós szélsőérték-keresés

- A gyakorlatban sokszor kell $\mathbb{R}^n \mapsto \mathbb{R}$ függvények minimumát keresni egy adott tartományon
- Tehát adott a $\min f(\boldsymbol{x}) : \boldsymbol{x} \in X$ nemlineáris program, és legyen most $X = \mathbb{R}^n$
- Tegyük fel, hogy f folytonosan differenciálható és konvex
- A minimum abban az $\bar{\boldsymbol{x}}$ pontban áll elő, melyre $\nabla f(\bar{\boldsymbol{x}}) = \mathbf{0}$
- Ez az egyenlet bonyolult f függvényekre analitikusan nem oldható meg, ezért numerikus módszereket kell bevetnünk
- Konvex program minimumának keresésére van módszerünk: a megengedett irányok módszere
- Itt most jelentősen egyszerűsödik, mert minden irány megengedett irány ($X = \mathbb{R}^n$)
- Csak arról kell gondoskodnunk, hogy javítóirány is legyen

Legmeredekebb irányok módszere

- Valamely $\bar{\boldsymbol{x}}$ pontban egy $\boldsymbol{d}$ irány javítóirány akkor és csak akkor, ha $\nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{d} < 0$
- Keressük azt az egység abszolút értékű $\boldsymbol{d}$ irányt, amely mentén a legmeredekebb a célfüggvény
- **Tétel:**

$$\operatorname*{argmin}_{\boldsymbol{d}:\|\boldsymbol{d}\|=1} \nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{d} = \frac{-\nabla f(\bar{\boldsymbol{x}})}{\|\nabla f(\bar{\boldsymbol{x}})\|}$$

- **Biz.:** a Cauchy–Schwarz egyenlőtlenségből, illetve $\|\boldsymbol{d}\| = 1$ miatt

$$\nabla f(\bar{\boldsymbol{x}})^T \boldsymbol{d} \geq -\|\nabla f(\bar{\boldsymbol{x}})\| \|\boldsymbol{d}\| \geq -\|\nabla f(\bar{\boldsymbol{x}})\|$$

- Egyenlőség akkor és csak akkor, ha $\boldsymbol{d} = \frac{-\nabla f(\bar{\boldsymbol{x}})}{\|\nabla f(\bar{\boldsymbol{x}})\|}$ $\square$

Legmeredekebb irányok módszere

- A tétel értelmében a $\boldsymbol{d} = \frac{-\nabla f(\bar{\boldsymbol{x}})}{\|\nabla f(\bar{\boldsymbol{x}})\|}$ a legmeredekebb megengedett javítóirány
- Vagyis ebben az irányban egységnyit haladva csökken a leginkább a célfüggvény értéke
- A lépésköz meghatározásához keresés a $\frac{-\nabla f(\bar{\boldsymbol{x}})}{\|\nabla f(\bar{\boldsymbol{x}})\|}$ irány mentén
- Célszerű a bináris keresést használni

Legmeredekebb irányok módszere

- Oldjuk meg a $\min f(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n$ feltételmentes nemlineáris programot, ahol f folytonosan differenciálható konvex függvény
- Inicializálás: válasszunk $\epsilon > 0$ pontosság határt, $\boldsymbol{x}_1$ kezdeti pontot és legyen $k = 1$

1 Ha $\|\nabla f(\boldsymbol{x}_k)\| < \epsilon$, akkor kiszállás, $\boldsymbol{x}_k$ minimumpont

2 Egyébként $\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$

3 Egyenes menti keresés: oldjuk meg a

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

feltételmentes egyenes menti keresési feladatot, és legyen az optimális megoldás λ_k

4 Legyen $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$ és ugrás vissza az 1. lépésre

Legmeredekebb irányok módszere

- Egyszerű és könnyen implementálható módszer
- Elméletben véges lépésben konvergál, a gyakorlatban azonban kevéssé használható
- Az algoritmus az optimum közelében egyre rövidebb, a legutóbbi lépésre majdnem merőleges irányban egyre kisebb lépéseket tesz meg: cikk-cakk
- Ez az összes elsőrendű módszerre érvényes, amelyet tárgyaltunk
- A jelenség kiküszöböléséhez általában másodrendű módszereket használnak
- Newton módszer, konjugált gradiens módszer, stb.
- Itt nem tárgyaljuk

Büntetőfüggvények módszere

- Az egy- és többváltozós feltételmentes optimalizálási algoritmusok segítségével megoldhatók feltételeket is tartalmazó nemlineáris programok
- Az ötlet az, hogy a feltételeket „beépítjük” a célfüggvénybe büntetőfüggvényeken keresztül: a megváltoztatott feltételmentes optimalizálási feladat a feltételes program optimumába konvergál
- A feltételek megsértése erősen növeli a célfüggvény értékét, így a minimalizálás során a megoldó algoritmusok törekednek majd a feltételek kielégítésére
    - külső büntetőfüggvények: a megengedett tartományból való kilépést büntetjük, így nem megengedett megoldások sorozatán keresztül jutunk az optimumba
    - belső büntetőfüggvények: a generált pontok nem léphetnek ki a megengedett tartományból

Külső büntetőfüggvények

- Tekintsük az alábbi egyszerű formájú nemlineáris programokat

$$\min f(\boldsymbol{x}) : h(\boldsymbol{x}) = 0$$

  ahol f és h két $\mathbb{R}^n \mapsto \mathbb{R}$ függvény és $\boldsymbol{x} \in \mathbb{R}^n$

- Helyettesítsük a fenti feltételes optimalizálási feladatot egy feltételmentes feladattal

$$\min f(\boldsymbol{x}) + \mu h^2(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n$$

  ahol $\mu > 0$ valamilyen nagy skalár

- Intuitívan látható, hogy az optimumban a $\mu h^2(\boldsymbol{x})$ **külső büntetőfüggvény** értéke közel lesz zéróhoz, ellenkező esetben ez a tag erősen megnöveli a célfüggvény értékét
- Elég nagy μ értéket választva általában biztosítható, hogy a két feladat optimuma ugyanoda essen

Külső büntetőfüggvények

- Egyenlőtlenség feltételt tartalmazó nemlineáris programok esetében másfajta külső büntetőfüggvény kell

$$\min f(\boldsymbol{x}) : g(\boldsymbol{x}) \leq 0$$

- Ekkor ugyanis $\mu g^2(\boldsymbol{x})$ mind a $g(\boldsymbol{x}) < 0$ és a $g(\boldsymbol{x}) > 0$ megoldásokat büntetné, holott nekünk csak a másodikat kell
- Helyettesítsünk az alábbi feltételmentes feladattal

$$\min f(\boldsymbol{x}) + \mu \max\{0, g(\boldsymbol{x})\} : \boldsymbol{x} \in \mathbb{R}^n$$

- Értelemszerűen csak akkor jelenik meg plusz tag a célfüggvényben, ha $g(\boldsymbol{x}) > 0$, ha $g(\boldsymbol{x}) < 0$, akkor $\max\{0, g(\boldsymbol{x})\} = 0$ és nincs büntetés
- Még jobb a $\mu(\max\{0, g(\boldsymbol{x})\})^2$ büntetőfv.: differenciálható

Külső büntetőfüggvények

- Általános esetben a nemlineáris program több egyenlőség és egyenlőtlenség típusú feltételt is tartalmaz

$$\begin{aligned} \min \ & f(\boldsymbol{x}) & \\ \text{s.t.} \ & g_i(\boldsymbol{x}) \le 0 & i \in \{1, \ldots, m\} \\ & h_i(\boldsymbol{x}) = 0 & i \in \{1, \ldots, n\} \end{aligned}$$

- Ekkor a külső büntetőfüggvény formája

$$\alpha(\boldsymbol{x}) = \sum_{i=1}^{m} \Phi(g_i(\boldsymbol{x})) + \sum_{i=1}^{n} \Psi(h_i(\boldsymbol{x}))$$

$$\Phi(y) = (\max\{0, y\})^p, \qquad \Psi(y) = |y|^p$$

valamilyen $p > 0$ egész számra

Külső büntetőfüggvények

- Ekkor az alábbi feltételmentes nemlineáris programot kell megoldani

$$\min f(\boldsymbol{x}) + \mu \left(\sum_{i=1}^{m} (\max\{0, g_i(\boldsymbol{x})\})^p + \sum_{i=1}^{n} |h_i(\boldsymbol{x})|^p \right) :$$

$$\boldsymbol{x} \in \mathbb{R}^n$$

- Megoldható a korábban ismertetett módszerekkel
- Tetszőleges, akár nem megengedett kezdőpontból is indítható
- Hátránya, hogy csak az optimum közelében ér a megengedett tartományba
- A jó működéshez nagy μ kell: numerikus instabilitás

Külső büntetőfüggvények: példa

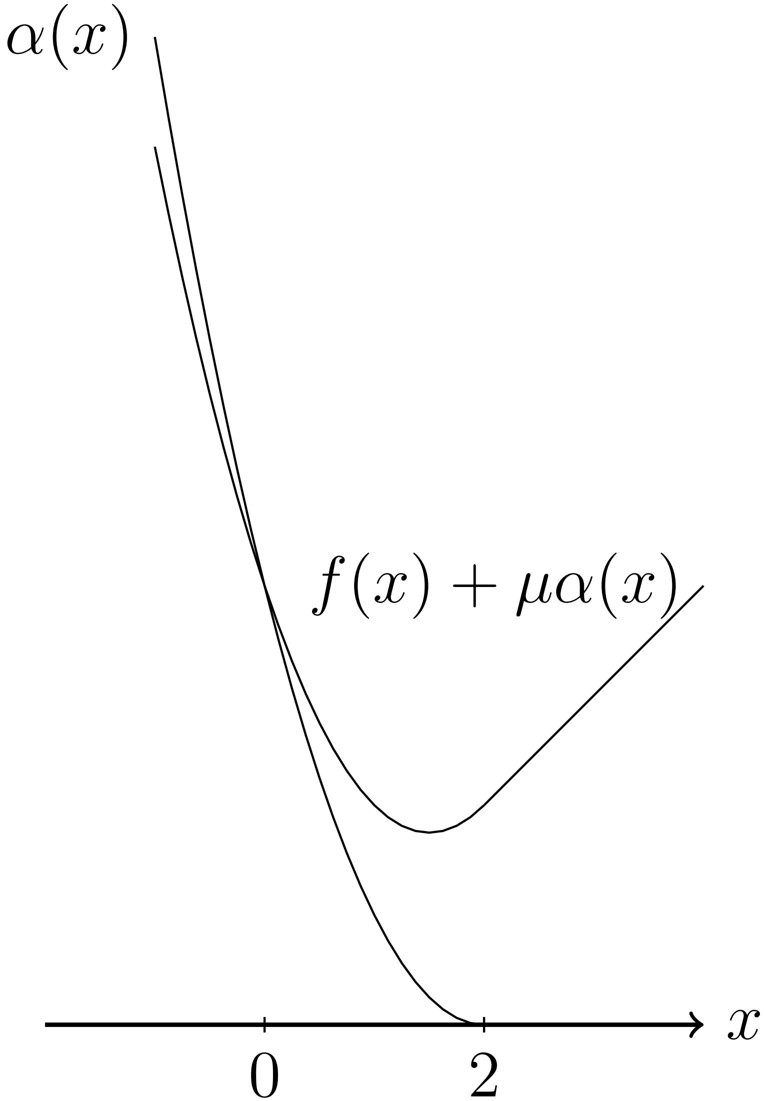
- Tekintsük a feltételes (nem)lineáris programot

$$\begin{aligned} \min \;& x \\ \text{s.t.} \;& -x+2 \leq 0 \end{aligned}$$

- Az optimum a $\bar{x}=2$ pontban $f(\bar{x})=2$
- Oldjuk meg külső büntetőfüggvény használatával
- Írjuk fel az $\alpha(x) = (\max\{0, g(x)\})^2$ külső büntetőfüggvény segítségével feltételmentes nemlineáris programként

$$\alpha(x) = \begin{cases} 0 & \text{ha } x \geq 2 \\ (-x+2)^2 & \text{ha } x < 2 \end{cases}$$

Külső büntetőfüggvények: példa



- Az $f(x)+\mu\alpha(x)$ konvex, így minimumpontja kereshető
- A derivált, ha $x<2$:

$$(f(x)+\mu\alpha(x))' = 1+2\mu(x-2) = 0$$

- Ebből a minimumpont

$$\bar{x} = 2 - \frac{1}{2\mu}$$

- Közelíti az eredeti probléma optimumát, ahogy $\mu \to \infty$
- Általában is igaz

Külső büntetőfüggvények: példa

- Oldjuk meg a feltételes nemlineáris programot

$$\min x_1^2 + x_2^2 : x_1 + x_2 - 1 = 0$$

- Az optimális megoldás $\begin{bmatrix} x_1 & x_2 \end{bmatrix}^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}^T$, az optimum $\frac{1}{2}$
- Alkalmazzuk az $\alpha(y) = |y|^2$ külső büntetőfüggvényt

$$\min x_1^2 + x_2^2 + \mu(x_1 + x_2 - 1)^2 : x_1, x_2 \in \mathbb{R}$$

- $\forall \mu \geq 0$ esetén konvex, így a gradiensből adódik a minimum

$$2x_1 + 2\mu(x_1 + x_2 - 1) = 0$$
$$2x_2 + 2\mu(x_1 + x_2 - 1) = 0$$

- A két egyenletből $x_1 = x_2 = \frac{1}{2 + \frac{1}{\mu}}$, $\mu \to \infty$ esetén optimális

Belső büntetőfüggvények

- A külső büntetőfüggvények módszere nem megengedett pontok sorozatán keresztül konvergál az optimumba
- A belső büntetőfüggvények módszerénél a generált pontok nem léphetnek ki a megengedett tartományból
- Adva van a nemlineáris program egyenlőtlenség feltételekkel

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0 \qquad\qquad i \in \{1, \ldots, m\}
\end{aligned}
$$

- Ehelyett oldjuk meg a feltételmentes optimalizálási problémát

$$
\min f(\boldsymbol{x}) + \mu\beta(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n
$$

Belső büntetőfüggvények

- A β belső büntetőfüggvény a megengedett tartományon valamilyen nemnegatív értéket vesz fel, míg a megengedett tartomány határához közelítve tart a végtelenbe

$$\beta(\boldsymbol{x}) = \sum_{i=1}^{m} \Phi(g_i(\boldsymbol{x}))$$

ahol a Φ függvénynek az alábbi tulajdonságai vannak

$$\Phi(y) \geq 0 \text{ ha } y < 0, \qquad \lim_{y \to 0^-} \Phi(y) = \infty$$

- A belső büntetőfüggvény megtiltja a megengedett tartomány határának átlépését
- Egyenlőségfeltételek nehezen kezelhetőek ilyen módon

Belső büntetőfüggvények

- Tipikus belső büntetőfüggvények

$$\beta(\boldsymbol{x}) = \sum_{i=1}^{m} \frac{-1}{g_i(\boldsymbol{x})}, \qquad \beta(\boldsymbol{x}) = -\sum_{i=1}^{m} \ln\left(\min\left\{1, -g_i(\boldsymbol{x})\right\}\right)$$

- A kiegészített célfüggvény feltételmentes optimalizálással minimalizálható, és a folyamat során sosem lépünk ki a megengedett tartományból
- Hátránya, hogy az optimum pontban minden feltétel jelentkezik a célfüggvényben, nemcsak az aktívak
- Továbbá feltétlenül megengedett pontból kell indítani, különben a büntetőfüggvény „rossz oldalára” kerülünk
- Akkor működik jól, ha μ értéke kicsi, ez azonban rosszul kondícionáltsághoz és numerikus instabilitáshoz vezethet

Belső büntetőfüggvények: példa

- Tekintsük a nemlineáris programot

$$\begin{aligned} \min \;& x \\ \text{s.t.} \;& -x+1 \leq 0 \end{aligned}$$

- Az optimum 1, amit a célfüggvény az $\bar{x}=1$ pontban vesz fel
- Legyen a belső büntetőfüggvény a következő

$$\beta(x) = \frac{-1}{-x+1} = \frac{1}{x-1} \qquad x \neq 1$$

- A kapott feltételmentes optimalizálási feladat

$$\min x + \frac{\mu}{x-1} : x \in \mathbb{R}$$

Belső büntetőfüggvények: példa

- A $\min x + \frac{\mu}{x-1} : x \in \mathbb{R}$ feladat csak akkor ad értelmes megoldást, ha $x > 1$, erről külön kell gondoskodnunk
- A célfüggvény konvex, minimuma a $\frac{d}{dx}(f(x) + \mu\beta(x)) = 0$ egyenletből $\bar{x} = 1 + \sqrt{\mu}$, optimális ha $\mu \to 0$