Izbrana poglavja iz optimizacije - zapiski

Timotej Hrga, Janez Povh

Update on 17. januar 2023

Kazalo

Poglavje 1

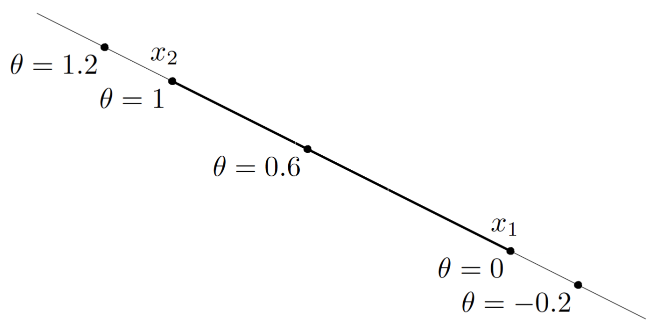
Sredstva iz konveksne analize

1.1 Konveksne množice

Definicija 1. *Množica $\mathcal{A} \subset \mathbb{R}^n$ je* **afina**, *če za vsaka $x, y \in \mathcal{A}$ in vsak $\lambda \in \mathbb{R}$ velja*

$$(1-\lambda)x + \lambda y \in \mathcal{A},$$

tj. vsaka premica skozi poljubni točki iz množice $\mathcal{A}$ leži v $\mathcal{A}$.



Slika 1.1: Parametrična oblika premice skozi točki x_1 in x_2 ima obliko $(1-\theta)x_1 + \theta x_2$, kjer $\theta \in \mathbb{R}$. Ko je $\theta \in [0,1]$, dobimo daljico s krajiščema x_1 in x_2.

Primer 1. *Množica rešitev sistema linearnih enačb $Ax = b$ je afina.*

Označimo $\mathcal{A} = \{x \mid Ax = b\}$. Naj bo $x, y \in \mathcal{A}$, tj. $Ax = b$ in $Ay = b$. Vzemimo poljuben $\lambda \in \mathbb{R}$. Potem velja

$$A\left((1-\lambda)x + \lambda y\right) = (1-\lambda)Ax + \lambda Ay = (1-\lambda)b + \lambda b = b.$$

Torej je $(1-\lambda)x + \lambda y \in \mathcal{A}$.

Vpeljimo nekatere oznake. Naj $\mathcal{S}^n$ označuje vektorski prostor simetričnih matrik reda n. Z

$$\mathcal{S}_+^n = \{X \in \mathcal{S}^n \mid X \succeq 0\}$$

označimo množico vseh *pozitivno semidefinitnih matrik* reda n, tj.

$$X \succeq 0 \quad \iff \quad z^T X z \geq 0 \text{ za vsak } z.$$

Definiramo, da je $A \succeq B$, če velja $A - B \succeq 0$.

Trditev 1 (Karakterizacija pozitivno semidefinitnih matrik)**.** *Naslednje trditve so ekvivalentne.*

(a) A je pozitivno semidefinitna.

(b) Vse lastne vrednosti matrike A so nenegativne.

(c) Obstaja taka matrika C, da velja $A = C^T C$.

Dokaz. $(a) \Rightarrow (b)$ Naj bo $A \succeq 0$. Po definiciji za vsak vektor x velja $x^T A x \geq 0$. Naj bo x normiran lastni vektor, ki pripada lastni vrednosti λ. Tedaj velja

$$x^T A x = \lambda x^T x = \lambda \|x\|^2 = \lambda \geq 0.$$

Torej so vse lastne vrednosti matrike A nenegativne.

$(b) \Rightarrow (c)$ Za simetrično matriko A obstaja spektralni (lastni) razcep $A = UDU^T$, pri čemer je U ortogonalna matrika in D diagonalna matrika z lastnimi vrednostmi matrike A po diagonali. Ker so po predpostavki vse lastne vrednosti nenegativne, obstaja taka matrika $\Lambda = \sqrt{D}$, kateri diagonalni elementi so koreni elementov matrike D. Od tod sledi

$$A = UDU^T = U\Lambda\Lambda U^T = C^T C,$$

če za matriko C vzamemo produkt ΛU^T.

$(c) \Rightarrow (a)$ Pozitivno definitnost preverimo po definiciji

$$x^T A x = x^T C^T C x = \|Cx\|_2^2 \geq 0. \qquad \square$$

Naj

$$\mathcal{S}^n_{++} = \{X \in \mathcal{S}^n \mid X \succ 0\}$$

označuje množico vseh *pozitivno definitnih matrik* reda n, tj.

$$X \succ 0 \quad \iff \quad z^T X z > 0 \text{ za vsak } z \neq 0$$

oziroma vse lastne vrednosti matrike X so pozitivne. Definiramo, da je $A \succ B$, če velja $A - B \succ 0$.

Naloga 1. *(VAJE) Dokaži, da je matrika $A = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$ pozitivno semidefinitna natanko tedaj, ko je $x, z \geq 0$ in $\det(A) = xz - y^2 \geq 0$.*

Rešitev: *Za lastni vrednosti λ_1 in λ_2 velja, da je $\det(A) = \lambda_1 \lambda_2 \geq 0$ in $\operatorname{sled}(A) = \lambda_1 + \lambda_2 \geq 0$. Od tod sledi, da je $\lambda_1, \lambda_2 \geq 0$.*

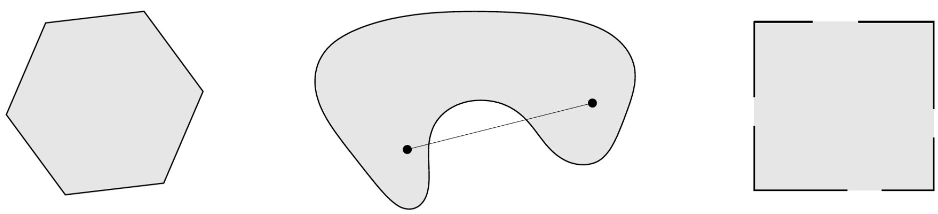
Definicija 2. *Množica $\mathcal{A} \subset \mathbb{R}^n$ je* **konveksna***, če za vsaka $x, y \in \mathcal{A}$ in za vsak $\lambda \in [0,1]$ velja*

$$(1-\lambda)x + \lambda y \in \mathcal{A},$$

tj. vsaka daljica s krajišči v množici $\mathcal{A}$ je v celoti vsebovana v $\mathcal{A}$ (glej Sliko 1.2).

Primer 2. *Vsaka afina množica je po definiciji konveksna.*

Definicija 3. *Točka x iz konveksne množice $\mathcal{A} \subset \mathbb{R}^n$ je* **ekstremna***, če iz $x = \lambda u + (1-\lambda)v$, za $u, v \in \mathcal{A}$ in $\lambda \in [0,1]$, sledi $x = u$ ali $x = v$.*



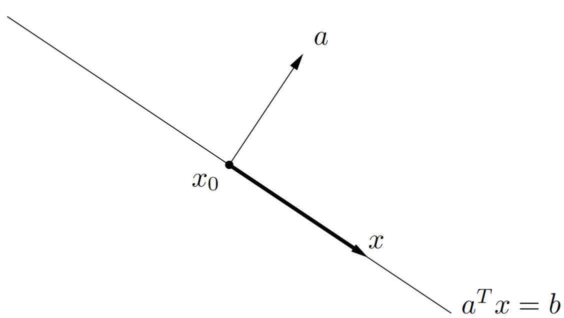
Slika 1.2: Primeri preprostih konveksnih in nekonveksnih množic. Prva množica je konveksna, drugi dve sta nekonveksni.

Primer 3. *Naslednje množice so konveksne.*

- *Za dani neničeln vektor a in število $b \in \mathbb{R}$ je* hiperravnina *z normalo a množica točk*

  $$\{x \mid a^T x = b\},$$

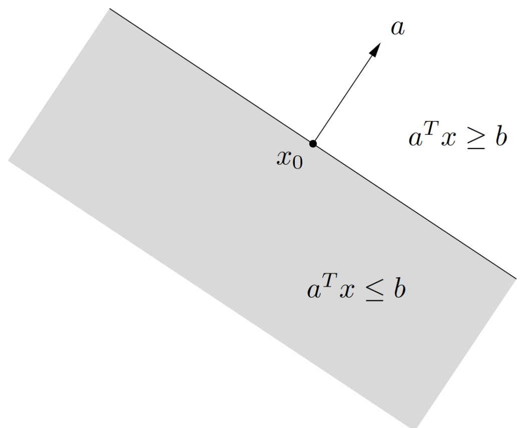
  tj. množica točk, ki zadošča linearni enačbi. V ravnini $\mathbb{R}^2$ je to premica, v prostoru $\mathbb{R}^3$ pa to določa ravnino. Hiperravnina je konveksna množica po Primeru 1.



- *Za dani neničeln vektor a in število $b \in \mathbb{R}$ je* polprostor *množica točk*

  $$\{x \mid a^T x \leq b\},$$

  tj. množica točk, ki zadošča linearni neenačbi.



Dokažimo konveksnost. Naj bo $\mathcal{A} = \{x \mid a^T x \leq b\}$. *Vzemimo* $x, y, \in \mathcal{A}$ *in* $\lambda \in [0,1]$. *Potem velja*

$$a^T((1-\lambda)x + \lambda y) = (1-\lambda)a^T x + \lambda a^T y \leq (1-\lambda)b + \lambda b = b.$$

Torej je $(1-\lambda)x + \lambda y \in \mathcal{A}$.

- *(VAJE)* Krogla *s središčem* x_c *in polmerom* $r > 0$ *je množica točk*

  $$\mathcal{B}(x_c, r) = \{x \mid \|x - x_c\| \leq r\}.$$

  Dokažimo konveksnost. Vzemimo $x, y \in \mathcal{B}(x_c, r)$ *in* $\lambda \in [0,1]$. *Potem velja* $\|x - x_c\| \leq r$ *in* $\|y - x_c\| \leq r$. *Sledi*

  $$\begin{aligned}\|(1-\lambda)x + \lambda y - x_c\| &= \|(1-\lambda)x + \lambda y - (1-\lambda+\lambda)x_c\| = \|(1-\lambda)(x - x_c) + \lambda(y - x_c)\| \\ &\leq (1-\lambda)\|x - x_c\| + \lambda\|y - x_c\| \leq (1-\lambda)r + \lambda r = r.\end{aligned}$$

  Torej je $(1-\lambda)x + \lambda y \in \mathcal{B}(x_c, r)$.

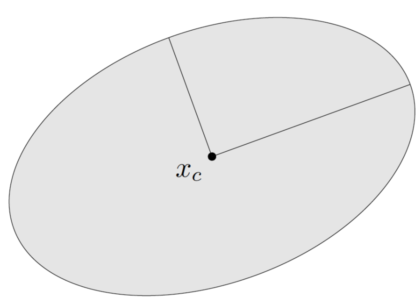
  Kroglo lahko zapišemo tudi kot množico točk

  $$\mathcal{B}(x_c, r) = \{x_c + u \mid \|u\| \leq r\}.$$

- *(VAJE) Naj* $\mathcal{S}^n$ *označuje vektorski prostor simetričnih matrik. Za dano pozitivno definitno matriko* $P \in \mathcal{S}^n$ *je* elipsoid *s središčem* x_c *množica točk*

  $$\mathcal{E}(x_c) = \{x \mid (x - x_c)^T P (x - x_c) \leq 1\}.$$

  Pri tem lastni vektorji matrike P^{-1} *določajo smeri polosi, koreni njenih lastnih vrednosti pa dolžine polosi. Na primer, če je* $P = \frac{1}{r^2} I$, *dobimo krog s polmerom* r.



Dokažimo konveksnost. Če definiramo normo $\|\cdot\|_P \colon \mathbb{R}^n \to \mathbb{R}$ *s predpisom*

$$\|x\|_P = \sqrt{x^T P x}, \tag{1.1}$$

lahko elipsoid zapišemo v obliki

$$\mathcal{E}(x_c) = \{x \mid \|x - x_c\|_P^2 \leq 1\}$$

oziroma

$$\mathcal{E}(x_c) = \{x \mid \|x - x_c\|_P \leq 1\}.$$

Naprej postopamo enako kot v dokazu konveksnosti za kroglo. Dokažimo še, da je s predpisom (1.1) *res definirana norma. Očitno je* $\|x\|_P \geq 0$ *in zaradi pozitivne definitnosti matrike* P *velja* $\|x\|_P = 0$ *natanko tedaj, ko je* $x = 0$. *Velja še*

$$\|\alpha x\|_P = \sqrt{(\alpha x)^T P (\alpha x)} = |\alpha| \sqrt{x^T P x} = |\alpha| \|x\|_P.$$

Kot zadnjo še moramo dokazati trikotniško neenakost. Ker je matrika P *pozitivno definitna, obstaja taka ortogonalna matrika* Q *in diagonalna matrika* D *s pozitivnimi elementi na diagonali, da velja* $P = QDQ^T$. *Torej lahko definiramo* $P^{\frac{1}{2}} = QD^{\frac{1}{2}}Q^T$, *kjer je* $D_{ii}^{\frac{1}{2}} = \sqrt{d_{ii}}$. *Od tod sledi*

$$\begin{aligned}
\|x + y\|_P^2 &= (x + y)^T P (x + y) = x^T P x + y^T P y + 2x^T P y \\
&= \|x\|_P^2 + \|y\|_P^2 + 2x^T P^{\frac{1}{2}} P^{\frac{1}{2}} y \\
&\overset{(1)}{\leq} \|x\|_P^2 + \|y\|_P^2 + 2\|P^{\frac{1}{2}} x\|_2 \|P^{\frac{1}{2}} y\|_2 \\
&= \|x\|_P^2 + \|y\|_P^2 + 2\|x\|_P \|y\|_P \\
&= \left(\|x\|_P + \|y\|_P\right)^2,
\end{aligned}$$

kjer smo pri (1) *uporabili Cauchy-Schwarzovo neenakost. Po korenjenju dobimo trikotniško neenakost.*

Definicija 4. *Naj bodo* $k \in \mathbb{N}$, $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ *in* $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. *Vektor* $\sum_{i=1}^{k} \lambda_i x_i$ *je* **konveksna kombinacija** *vektorjev* $x_1, x_2, \ldots, x_k$, *če velja*

$$\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0 \quad \textit{in} \quad \sum_{i=1}^{k} \lambda_i = 1.$$

Definicija 5. *Naj bo* $\mathcal{A} \subseteq \mathbb{R}^n$. **Konveksna ovojnica (ogrinjača)** $\mathcal{A}$, *ki jo označimo s* $\operatorname{conv}(\mathcal{A})$, *je množica vseh konveksnih kombinacij vektorjev iz* $\mathcal{A}$.

Definicija 6. *Množica $\mathcal{C} \in \mathbb{R}^n$ je* **konveksni stožec**, *če za vsaka $x, y \in \mathcal{A}$ in za vsaka skalarja $\lambda, \mu \geq 0$ velja*

$$\lambda x + \mu y \in \mathcal{A},$$

tj. množica je zaprta za nenegativne vsote. Pravimo, da je stožec $\mathcal{C}$ **končno generiran**, *če obstajajo generatorji $x_1, x_2, \ldots, x_k$, da velja*

$$\mathcal{C} = \{\lambda_1 x_1 + \ldots \lambda_k x_k \mid \lambda_1, \ldots, \lambda_k \geq 0\},$$

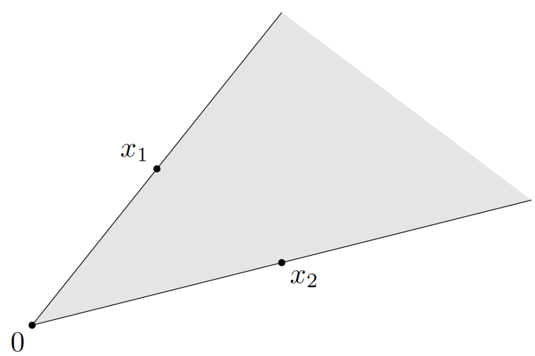
tj. vsaka element stožca lahko zapišemo kot **nenegativno kombinacijo** *generatorjev $x_1, x_2, \ldots, x_k$. Za dano množico $\mathcal{A} \subset \mathbb{R}^n$ imenujemo množico vseh končnih vsot $\{\sum_i \lambda_i x_i \mid \lambda_i \geq 0,\ x_i \in \mathcal{A}\}$* **stožčna ovojnica (ogrinjača)** *od $\mathcal{A}$. Oznaka:* $\operatorname{cone}(\mathcal{A})$.

Vsak konveksni stožec je po definiciji konveksna množica.

Primer 4. Nenegativni ortant

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$$

je konveksni stožec. Pri tem je neenakost mišljena po komponentah. Nenegativni ortant je končno generiran stožec z generatorji $e_1, e_2, \ldots, e_n$.



Slika 1.3: Končno generiran stožec z generatorjema x_1 in x_2.

Naloga 2. *(VAJE) Pokaži, da je za dano matriko A množica $\mathcal{C} = \{Ax \mid x \geq 0\}$ konveksni stožec.*

Rešitev: *Vzemimo $x_1, x_2 \in \mathcal{C}$ in $\lambda, \mu \geq 0$. Potem obstajata nenegativna vektorja u in v, tako da je $x_1 = Au$ in $x_2 = Av$. Sledi*

$$\lambda x_1 + \mu x_2 = \lambda A u + \mu A v = A\left(\lambda u + \mu v\right).$$

Ker je $\lambda u + \mu v \geq 0$, velja $\lambda x_1 + \mu x_2 \in \mathcal{C}$.

Primer 5. *Množica točk*

$$\mathcal{C} = \left\{(x,t) \in \mathbb{R}^{n+1} \mid \|x\| \leq t\right\}$$

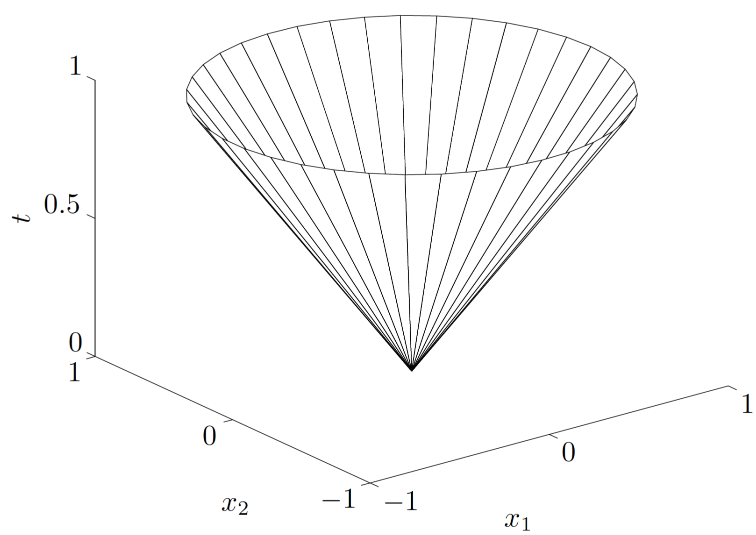
je konveksni stožec. Če za normo vzamemo Evklidsko ali 2-normo, dobimo Lorentzov stožec $\mathcal{L}^{n+1}$.

Vzemimo $(x_1, t_1), (x_2, t_2) \in \mathcal{C}$ *in* $\lambda, \mu \geq 0$. *Dokazujemo, da velja*

$$\lambda(x_1, t_1) + \mu(x_2, t_2) = (\lambda x_1 + \mu x_2, \lambda t_1 + \mu t_2) \in \mathcal{C}.$$

Po trikotniški neenakosti dobimo

$$\|\lambda x_1 + \mu x_2\| \leq \lambda\|x_1\| + \mu\|x_2\| \leq \lambda t_1 + \mu t_2.$$



Slika 1.4: Rob Lorentzovega stožca $\mathcal{L}^3 = \left\{(x_1, x_2, t) \mid \sqrt{x_1^2 + x_2^2} \leq t\right\}$.

Primer 6. *Množica pozitivno semidefinitnih matrik* $\mathcal{S}_+^n$ *je konveksni stožec.*

Vzemimo $X, Y \in \mathcal{S}_+^n$ *in* $\lambda, \mu \geq 0$. *Potem velja*

$$z^T(\lambda X + \mu Y)z = \lambda z^T X z + \mu z^T Y z \geq 0 \;\; \textit{za vsak } z.$$

Torej $\lambda X + \mu Y \in \mathcal{S}_+^n$.

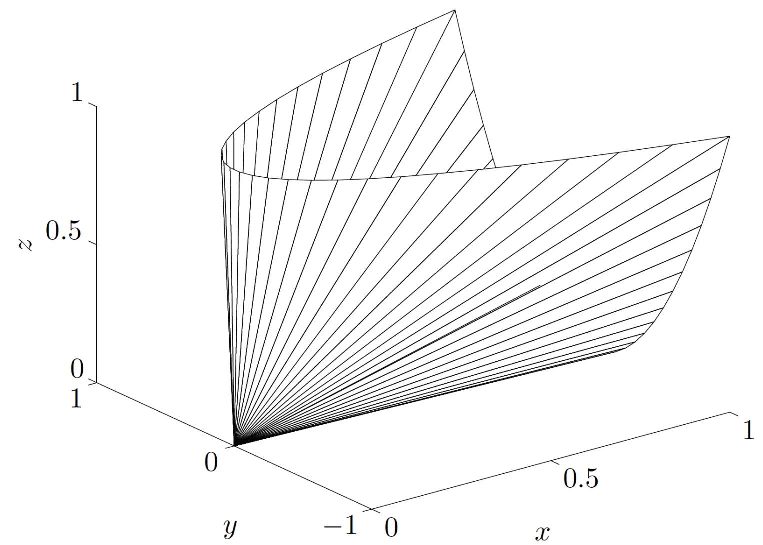
Naloga 3. *(VAJE) Naj bo* $C = \{(1,0),\ (1,1),\ (-1,-1),\ (0,0)\}$. *Določi pravilnost naslednjih izjav.*

(a) $\left(0, -\frac{1}{3}\right) \in \operatorname{conv}(C)$.

(b) $\left(0, \frac{1}{3}\right) \in \operatorname{conv}(C)$.

(c) $\left(0, \frac{1}{3}\right) \in \operatorname{cone}(C)$.

Rešitev: *Opazimo, da je konveksna ogrinjača množice* C *trikotnik z oglišči* $(1,0), (1,1)$ *in* $(-1,-1)$. *Stožčna ogrinjača pa je polprostor* $\{(x,y) \mid y \leq x\}$.



Slika 1.5: Rob stožca pozitivno semidefinitnih matrik $\mathcal{S}^2_+$. Matrika $A = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$ je pozitivno semidefinitna natanko tedaj, ko je $x, z \geq 0$ in $\det(A) = xz - y^2 \geq 0$.

(a) *Dokazujemo da* $\left(0, -\frac{1}{3}\right) \in \operatorname{conv}(C)$. *Iščemo nenegativne skalarje* λ_1, λ_2 *in* $\lambda_3 = 1 - \lambda_1 - \lambda_2$, *da velja*

$$\left(0, -\frac{1}{3}\right) = \lambda_1(1,0) + \lambda_2(1,1) + (1 - \lambda_1 - \lambda_2)(-1,-1).$$

To je sistem dveh linearnih enačb

$$2\lambda_1 + 2\lambda_2 = 1, \quad \lambda_1 + 2\lambda_2 = \frac{2}{3}$$

z nenegativno rešitvijo $\lambda_1 = \frac{1}{3}$, $\lambda_2 = \frac{1}{6}$ *in* $\lambda_3 = \frac{1}{2}$. *Torej* $\left(0, -\frac{1}{3}\right) \in \operatorname{conv}(C)$.

(b) *Dokazujemo da* $\left(0, \frac{1}{3}\right) \notin \operatorname{conv}(C)$. *Iščemo nenegativne skalarje* λ_1, λ_2 *in* $\lambda_3 = 1 - \lambda_1 - \lambda_2$, *da velja*

$$\left(0, \frac{1}{3}\right) = \lambda_1(1,0) + \lambda_2(1,1) + (1 - \lambda_1 - \lambda_2)(-1,-1).$$

Dobljen sistem

$$2\lambda_1 + 2\lambda_2 = 1, \quad \lambda_1 + 2\lambda_2 = \frac{4}{3}$$

nima nenegativne rešitve ($\lambda_1 = -\frac{1}{3}$), *zato* $\left(0, \frac{1}{3}\right) \notin \operatorname{conv}(C)$.

(c) *Dokazujemo, da* $\left(0, \frac{1}{3}\right) \notin \operatorname{cone}(C)$. *Iščemo nenegativna skalarja* λ_1 *in* λ_2 *da velja*

$$\left(0, \frac{1}{3}\right) = \lambda_1(-1,-1) + \lambda_2(1,1).$$

Dobljen sistem

$$-\lambda_1 + \lambda_2 = 0, \quad -\lambda_1 + \lambda_2 = \frac{1}{3}$$

nima rešitve, zato $\left(0, \frac{1}{3}\right) \notin \operatorname{cone}(C)$.

Definicija 7. *Točka x iz konveksnega stožca $\mathcal{A} \subset \mathbb{R}^n$ je* **ekstremni žarek**, *če iz $x = \lambda u + \mu v$, za $u, v \in \mathcal{A}$ in $\lambda, \mu \geq 0$, sledi $x = \hat{\lambda} u$ in $x = \hat{\mu} v$, za $\hat{\lambda}, \hat{\mu} \geq 0$.*

Opomba 1. *Lorentzov stožec in stožec pozitivno semidefinitnih matrik nista končno generirana. Ekstremni žarki za Lorentov stožec so točke oblike $(x_1, \ldots, x_{n-1}, \sqrt{\sum_{i=1}^{n-1} x_i^2})$. Teh točk pa je neskončno.*

V nadaljevanju se bomo pogosto sklicevali na notranje točkemnožic, zato tukaj ponovimo dfinicijo.

Definicija 8. *Točka $x \in C \subset \mathbb{R}^n$ je notranja, če obstaja $\varepsilon > 0$, da je krogla $\{y \mid \|x - y\| < \varepsilon\} \subset C$. Množico vseh notranjih točk od C označimo z $\operatorname{int}(C)$. Zaprtje množice C definiramo kot*

$$\mathrm{zaprtje(C)} := \mathbb{R}^{\mathrm{n}} \setminus \mathrm{int}(\mathbb{R}^{\mathrm{n}} \setminus \mathrm{C}).$$

Trditev 2. *Za stožec pozitivno semidefinitnih matrik velja*

$$\operatorname{int}\left(\mathcal{S}_+^n\right) = \mathcal{S}_{++}^n.$$

Dokaz. $\operatorname{int}\left(\mathcal{S}_+^n\right) \subseteq \mathcal{S}_{++}^n$:

Naj bo $A \in \operatorname{int}\left(\mathcal{S}_+^n\right)$. Potem obstaja tak $\varepsilon > 0$, da velja: $X \in \mathcal{S}^n \wedge \|A - X\| \leq \varepsilon \Rightarrow X \in \mathcal{S}_+^n$. Pri tem je $\| \cdot \|$ inducirana matrična norma. Vzemimo $X = A - \varepsilon I$. Ker je $\|A - X\| = \varepsilon \|I\| = \varepsilon$, velja $A - \varepsilon I \in \mathcal{S}_+^n$. Torej so vse lastne vrednosti $\lambda_i(A) - \varepsilon$ matrike $A - \varepsilon I$ nenegativne. Sledi $\lambda_i(A) \geq \varepsilon > 0$ in zato $A \in \mathcal{S}_{++}^n$.

$\mathcal{S}_{++}^n \subseteq \operatorname{int}\left(\mathcal{S}_+^n\right)$:

Naj bo $A \succ 0$ in označimo z $\lambda_{\min} > 0$ njeno najmanjšo lastno vrednost. Vzemimo kroglo $\mathcal{B}\left(A, \frac{1}{2}\lambda_{\min}\right) \subset \mathcal{S}^n$ in pokažimo, da je vsebovana v množici $\mathcal{S}_+^n$. Naj bo $Y \in \mathcal{B}\left(A, \frac{1}{2}\lambda_{\min}\right)$, tj. $\|A - Y\| \leq \frac{1}{2}\lambda_{\min}$. Potem za vsak x velja $x^T(A - Y)x \leq \frac{1}{2}\lambda_{\min}\|x\|^2$. Ker je $x^T A x \geq \lambda_{\min}\|x\|^2$, sledi

$$x^T Y x \geq x^T A x - \frac{1}{2}\lambda_{\min}\|x\|^2 \geq \frac{1}{2}\lambda_{\min}\|x\|^2 \geq 0.$$

Torej je $Y \in \mathcal{S}_+^n$. $\square$

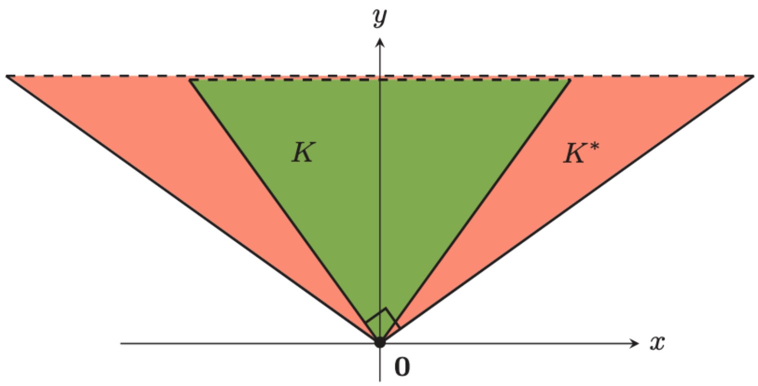
Definicija 9. *Naj bo $\mathcal{C}$ konveksni stožec.* **Dualni stožec** *je množica*

$$\mathcal{C}^* = \{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq 0 \ \text{ za vse } x \in \mathcal{C}\}.$$

Primer 7. *Dualni stožec linearnega podprostora $V \subseteq \mathbb{R}^n$ je njegov ortogonalni komplement*

$$V^\perp = \left\{y \mid v^T y = 0 \ \text{ za vsak } v \in V\right\}.$$

Lema 1. *Za vsak konveksni stožec $\mathcal{C}$ je dualni stožec $\mathcal{C}^*$ zaprt konveksni stožec.*



Slika 1.6: Primer konveksnega stožca K in dualnega stožca K^* v $\mathbb{R}^2$.

Dokaz. Vzemimo $y, z \in \mathcal{C}^*$ in $\lambda, \mu \geq 0$. Potem za vse $x \in \mathcal{C}$ velja

$$\langle \lambda y + \mu z, x \rangle = \lambda \langle y, x \rangle + \mu \langle z, x \rangle \geq 0.$$

Torej $\lambda y + \mu z \in \mathcal{C}^*$. Dokažimo še zaprtost. Za vsak x je množica $\{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq 0\}$ zaprt polprostor. Torej je $\mathcal{C}^*$ kot presek zaprtih množic zaprt. □

Lema 2. *S predpisom*

$$\langle A, B \rangle = \operatorname{sled}(A^T B) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

je definiran skalarni produkt v prostoru matrik $\mathbb{R}^{m \times n}$. Normo porojeno s tem skalarnim produktom imenujemo Frobeniusova norma

$$\|A\|_F = \sqrt{\sum_{i,j=1}^{n} a_{ij}^2}.$$

Dokaz. Preverimo lastnosti. $\langle A, A \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \geq 0$ in velja, da je $\langle A, A \rangle = 0$ natanko tedaj, ko je A ničelna matrika. Simetričnost velja $\langle A, B \rangle = \langle B, A \rangle$, saj je sled tranponirane matrike enaka sledi prvotne matrike. Linearnost pa velja zaradi

$$\begin{aligned}
\langle \alpha A + \beta B, C \rangle &= \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\alpha a_{ij} + \beta b_{ij}\right) c_{ij} \\
&= \alpha \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} c_{ij} + \beta \sum_{i=1}^{m} \sum_{j=1}^{n} b_{ij} c_{ij} \\
&= \alpha \langle A, C \rangle + \beta \langle B, C \rangle.
\end{aligned}$$

□

Lema 3. *Za $X, Y \succeq 0$ velja*

$$\langle X, Y\rangle = 0 \iff XY = 0.$$

Dokaz. ($\Rightarrow$) Naj bo $\langle X, Y\rangle = 0$. Ker sta X in Y pozitivno semidefinitni matriki, obstajata po Trditivi 1 taki matriki U in V, da velja $X = U^T U$ in $Y = V^T V$. Tedaj je

$$0 = \langle X, Y\rangle = \langle U^T U, V^T V\rangle = \mathrm{sled}(U^T U V^T V) = \mathrm{sled}(U V^T V U^T) = \langle V U^T, V U^T\rangle = \|V U^T\|_F^2.$$

Pri tem smo uporabili lastnost $\mathrm{sled}(AB) = \mathrm{sled}(BA)$. Tore je $VU^T = 0$ oziroma $UV^T = 0$. Od tod sledi

$$XY = U^T \underbrace{UV^T}_{0} V = 0.$$

($\Leftarrow$)

$$\langle X, Y\rangle = \mathrm{sled}(XY) = \mathrm{sled}(0) = 0. \qquad \square$$

Definicija 10. *Pravimo, da je stožec $\mathcal{C}$* **sebidualen***, če velja $\mathcal{C}^* = \mathcal{C}$.*

Primer 8. *Stožci $\mathbb{R}^n_+$, $\mathcal{L}^{n+1}$ in $\mathcal{S}^n_+$ so sebidualni.*

- $\left(\mathbb{R}^n_+\right)^* = \mathbb{R}^n_+$

  Naj bo $y \in \mathbb{R}^n_+$. Ker za vse $x \in \mathbb{R}^n_+$ velja $\langle y, x\rangle = \sum_{i=1}^{n} x_i y_i \geq 0$, smo pokazali inkluzijo $\mathbb{R}^n_+ \subseteq \left(\mathbb{R}^n_+\right)^*$. Za dokaz obratne inkluzije vzemimo $y \in \left(\mathbb{R}^n_+\right)^*$, tj. velja $\langle y, x\rangle \geq 0$ za vse $x \in \mathbb{R}^n_+$. Vzemimo $x = e_i$. Sledi $\langle y, x\rangle = \langle y, e_i\rangle = y_i \geq 0$. Torej je $y \geq 0$.

- $(\mathcal{L}^{n+1})^* = \mathcal{L}^{n+1}$

  Naj bo $(x_1, t_1) \in \mathcal{L}^{n+1}$. Ker za vsak $(x_2, t_2) \in \mathcal{L}^{n+1}$ po Cauchy-Schwarzovi neenakosti velja

  $$\langle (x_1, t_1), (x_2, t_2)\rangle = \langle x_1, x_2\rangle + t_1 t_2 \geq -\|x_1\|\|x_2\| + t_1 t_2 \geq -t_1 t_2 + t_1 t_2 = 0,$$

  smo pokazali inkluzijo $\mathcal{L}^{n+1} \subseteq (\mathcal{L}^{n+1})^*$. Za dokaz obratne inkluzije bomo pokazali, da velja

  $$(x_1, t_1) \notin \mathcal{L}^{n+1} \Longrightarrow (x_1, t_1) \notin \left(\mathcal{L}^{n+1}\right)^*.$$

  Naj bo torej $(x_1, t_1) \notin \mathcal{L}^{n+1}$, tj. $\|x_1\| > t_1$. Vzemimo $(x_2, t_2) = (-x_1, \|x_1\|)$, ki očitno pripada stožcu $\mathcal{L}^{n+1}$. Potem velja

  $$\langle (x_1, t_1), (x_2, t_2)\rangle = \langle x_1, x_2\rangle + t_1 t_2 = -\|x_1\|^2 + t_1\|x_1\| < -\|x_1\|^2 + \|x_1\|^2 = 0.$$

  Torej $(x_1, t_1) \notin (\mathcal{L}^{n+1})^*$.

- $\left(\mathcal{S}_+^n\right)^* = \mathcal{S}_+^n$

  Naj bo $Y \succeq 0$. Tedaj obstaja taka ortogonalna matrika Q in diagonalna matrika D z nenegativnimi elementi na diagonali, da velja $Y = QDQ^T$. Potem za vsako matriko $X \in \mathcal{S}_+^n$ velja

$$\langle X, Y\rangle = \operatorname{sled}(XY) = \operatorname{sled}(XQDQ^T) \overset{(1)}{=} \operatorname{sled}(Q^TXQD) = \sum_{i=1}^{n} \left(Q^TXQ\right)_{ii} d_{ii} \overset{(2)}{\geq} 0.$$

  Pri tem smo pri (1) smo upoštevali ciklično lastnost sledi in pri (2) da je matrika $Q^TXQ \in \mathcal{S}_+^n$, če je $X \in \mathcal{S}_+^n$. Vsi diagonalni elementi pozitivno semidefinitne matrike pa so nenegativni. Torej velja inkluzija $\mathcal{S}_+^n \subseteq \left(\mathcal{S}_+^n\right)^*$. Za dokaz obratne inkluzije vzemimo $Y \in \left(\mathcal{S}_+^n\right)^*$, tj. $\langle X, Y\rangle \geq 0$ za vsako matriko $X \in \mathcal{S}_+^n$. Za poljuben $x \in \mathbb{R}^n$ je matrika $X = xx^T$ očitno pozitivno semidefinitna. Potem velja

$$x^TYx = \operatorname{sled}(x^TYx) = \operatorname{sled}(xx^TY) = \operatorname{sled}(XY) = \langle X, Y\rangle \geq 0.$$

  Torej je $Y \in \mathcal{S}_+^n$.

Naloga 4. *(VAJE) Poišči dualna stožca naslednjih stožcev.*

(a) $K_1 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 = 0\}$

(b) $K_2 = \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_1| \leq x_2\}$

Rešitev:

(a) K_1 je linearni podpostor, tj. premica s smernim vektorjem $(-1, 1)$. Dualni stožec je orgonalni komplement $K_1^* = K_1^\perp = \{(x_1, x_2) \in \mathbb{R}^2 \mid -x_1 + x_2 = 0\}$.

(b) K_2 je Lorentzov stožec $\mathcal{L}^2$, ki je sebidualen. Zato je $K_2^* = K_2$.

Naloga 5. *(VAJE) Za dano matriko $A \in \mathbb{R}^{m\times n}$ poišči dualni stožec stožca $\mathcal{C} = \{Ax \mid x \geq 0\}$.*

Rešitev: *Po definiciji dualnega stožca velja*

$$\begin{aligned}\mathcal{C}^* &= \{y \mid \langle y, z\rangle \geq 0 \text{ \textit{za vsak} } z \in \mathcal{C}\}\\ &= \{y \mid \langle y, Ax\rangle \geq 0,\ x \geq 0\}\\ &= \{y \mid \langle A^Ty, x\rangle \geq 0,\ x \geq 0\}.\end{aligned}$$

Ker je nenegativni ortant sebidualen, dobimo

$$\mathcal{C}^* = \{y \mid A^Ty \geq 0\}.$$

Definicija 11. *Za dane matrike $A \in \mathcal{S}^m$, $C \in \mathcal{S}^n$, $B \in \mathbb{R}^{m\times n}$ je podana bločno simetrična matrika X oblike*

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}.$$

Denimo, da je A nesingularna. Matriki

$$S = C - B^T A^{-1} B$$

pravimo **Schurov komplement** *matrike A v X.*

Trditev 3. *Naj bo $A \in \mathcal{S}^n_{++}$, $C \in \mathcal{S}^n$ in $B \in \mathbb{R}^{m\times n}$. Za matriko $X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ in Schurov komplement $S = C - B^T A^{-1} B$ velja:*

(a) $\det(X) = \det(A)\det(S)$.

(b) $X \succeq 0$ *natanko tedaj, ko je* $S \succeq 0$.

(b) $X \succ 0$ *natanko tedaj, ko je* $S \succ 0$.

Dokaz. Ker je matrika A pozitivno definitna, lahko definiramo nesingularno matriko

$$P = \begin{bmatrix} I_m & -A^{-1}B \\ 0 & I_n \end{bmatrix}.$$

Tako je

$$P^T X P = P^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} P = \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix}.$$

Ker je $\det(P) = 1$, velja točka (a) zaradi multiplikativnosti determinante in iz dejstva, da je determinanta bločne matrike enaka produktu determinant posameznih blokov. Točki (b) in (c) pa sledita iz dejstva, da je bločno diagonalna matrika pozitivno semidefinitna (definitna) natanko takrat, ko so vsi njeni diagonalni bloki pozitivno semidefinitni (definitni). Dodatno smo še upoštevali, da je za nesingularno matriko P matrika $P^T X P \in \mathcal{S}^n_+$ $(\mathcal{S}^n_{++})$ natako tedaj, ko je $X \in \mathcal{S}^n_+$ $(\mathcal{S}^n_{++})$. □

V nadaljevanju si bomo ogledali operacije, ki ohranjajo konveksnost množic.

Trditev 4. *Naj bo $\mathcal{I} \neq \emptyset$ neprazna množica indeksov in naj bo za vsak $i \in \mathcal{I}$ množica $\mathcal{A}_i$ konveksna. Potem je tudi presek*

$$\bigcap_{i\in\mathcal{I}} \mathcal{A}_i = \{x \in \mathbb{R}^n \mid \forall i \in \mathcal{I} : x \in \mathcal{A}_i\}$$

konveksna množica.

Dokaz. Vzemimo $x, y \in \bigcap\limits_{i \in \mathcal{I}} \mathcal{A}_i$ in $\lambda \in [0,1]$. Potem sta $x, y \in \mathcal{A}_i$ za vsak $i \in \mathcal{I}$. Zaradi konveksnosti množic $\mathcal{A}_i$ potem velja $(1-\lambda)x + \lambda y \in \mathcal{A}_i$ za vsak $i \in \mathcal{I}$. Po definiciji preseka sledi, da je $(1-\lambda)x + \lambda y \in \bigcap\limits_{i \in \mathcal{I}} \mathcal{A}_i$. $\square$

Primer 9. *Stožec pozitivno semidefinitnih matrik lahko zapišemo kot*

$$\mathcal{S}_+^n = \bigcap_{z \neq 0} \{X \in \mathcal{S}^n \mid z^T X z \geq 0\}.$$

Za vsak $z \neq 0$ je $f(X) = z^T X z$ linearna funkcija v X, saj velja

$$f(X) = z^T X z = \mathrm{sled}(z^T X z) = \mathrm{sled}(X z z^T) = \langle X, z z^T \rangle.$$

Zato so množice

$$\{X \in \mathcal{S}^n \mid z^T X z \geq 0\}$$

polprostori v $\mathcal{S}^n$. Torej je $\mathcal{S}_+^n$ presek neskončno mnogo polprostorov in je zato konveksna množica.

Naloga 6. *(VAJE) Dokaži, da je množica*

$$C = \{X \in \mathcal{S}^n \mid X \succeq 0,\ X_{ii} = 1,\ i = 1, \ldots, n\}$$

konveksna.

Rešitev: *Množica C je presek stožca pozitivno semidefinitnih matrik in afine množice, definirane z linearnimi enačbami $X_{ii} = 1$, $i = 1, \ldots, n$. V prostoru matrik $\mathbb{R}^{m \times n}$ ima linearna enačba obliko $\langle A, X \rangle = b$, kjer je $A \in \mathbb{R}^{m \times n}$ in $b \in \mathbb{R}$.*

Naloga 7. *(VAJE) Dokaži, da je množica*

$$\left\{a \in \mathbb{R}^k \mid p(0) = 1,\ |p(t)| \leq 1 \text{ za } \alpha \leq t \leq \beta \right\},$$

kjer je $p(t) = a_1 + a_2 t + \ldots + a_k t^{k-1}$, konveksna.

Rešitev: *Množica je presek neskončno mnogo polprostorov*

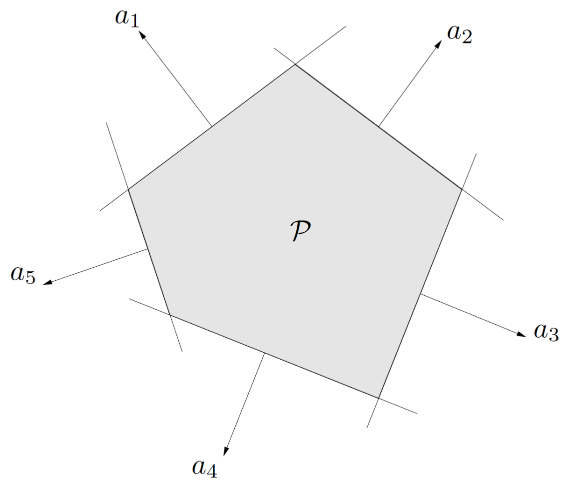
$$\left\{a \in \mathbb{R}^k \mid -1 \leq a_1 + a_2 t + \ldots + a_k t^{k-1} \leq 1\right\},$$

parametriziranih s $t \in [\alpha, \beta]$, in hiperravnine $\{a \in \mathbb{R}^k \mid a_1 = 0\}$.

Definicija 12. *Za dane $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^m$ in $d \in \mathbb{R}^p$ je* **konveksni polieder** *množica*

$$\{x \in \mathbb{R}^n \mid Ax \leq b,\ Cx = d\},$$

tj. množica rešitev sistema (končno mnogo) linearnih enačb in neenačb. Omejenemu poliedru pravimo politop.



Slika 1.7: Polieder $\mathcal{P}$ je presek petih polprostorov.

Ker je konveksni polieder po definiciji presek končno mnogo polprostorov in hiperravnin, je po Trditvi 4 konveksna množica.

Naloga 8. *(VAJE) Katere izmed množic S_1, S_2 in S_3 so poliedri? Zapiši jih v obliki*

$$\{x \mid Ax \leq b,\ Cx = d\}.$$

(a) $S_1 = \{y_1 a_1 + y_2 a_2 \mid -1 \leq y_1 \leq 1,\ -1 \leq y_2 \leq 1\}$*, kjer sta* $a_1, a_2 \in \mathbb{R}^2$ *linearno neodvisna vektorja.*

(b) $S_2 = \left\{x \in \mathbb{R}^n \mid x \geq 0,\ e^T x = 1,\ \sum_{i=1}^n a_i x_i = b_1,\ \sum_{i=1}^n a_i^2 x_i = b_2\right\}$*, kjer je* $a_1, \ldots, a_n \in \mathbb{R}$ *in* $b_1, b_2 \in \mathbb{R}$*.*

(c) $S_3 = \left\{x \in \mathbb{R}^n \mid x \geq 0,\ x^T y \leq 1 \ \text{ za vsak } y, \text{ za katerega velja } \|y\|_2 = 1\right\}$*.*

Rešitev:

(a) S_1 *je paralelogram z oglišči* $a_1 + a_2$, $a_1 - a_2$, $-a_1 + a_2$ *in* $-a_1 - a_2$*. Torej gre za polieder, ki je presek 4 polravnin. Določimo enačbo tiste polravnine, ki je določena s premico skozi točki* $a_1 + a_2$ *in* $-a_1 + a_2$*. Enačba polravnine je oblike* $c^T x \leq c^T x_0$*, kjer je* c *normalni vektor in* x_0 *točka, ki leži na premici. V našem primeru je* $x_0 = a_2$*. Normalo* c *lahko zapišemo kot linearno kombinacijo vektorjev* a_1 *in* a_2*, tj.*

$$c = \alpha_1 a_1 + \alpha_2 a_2.$$

Ker sta vektorja a_1 *in* c *pravokotna, dobimo*

$$0 = c^T a_1 = \alpha_1 \|a_1\|^2 + \alpha_2 a_1^T a_2.$$

Od tod sledi $\alpha_1 = -\alpha_2 \frac{a_1^T a_2}{\|a_1\|^2}$*. Torej je*

$$c = a_2 - \frac{a_1^T a_2}{\|a_1\|^2} a_1.$$

Podobno določimo preostale 3 polravnine.

(b) *S_2 je polieder, ki je definiran kot presek stožca $x \geq 0$ in afinega podprostora, določenega s 3 linearnimi enačbami.*

(c) *S_3 je presek stožca $x \geq 0$ in enotske krogle $\{x \mid \|x\|_2 \leq 1\}$. Velja namreč*

$$x^T y \leq 1,\ \|y\|_2 = 1 \iff \|x\|_2 \leq 1.$$

($\Leftarrow$) *Denimo, da velja $\|x\|_2 \leq 1$. Potem po Cauchy-Schwarzovi neenakosti velja*

$$x^T y \leq \|x\|_2 \|y\|_2 \leq 1$$

za vsak $\|y\|_2 = 1$.

($\Rightarrow$) *Obratno, naj bo x neničeln vektor, za katerega velja $x^T y \leq 1$ za vsak $\|y\|_2 = 1$. Vzemimo $y = \frac{x}{\|x\|_2}$, ki ima normo 1. Potem je*

$$x^T y = \frac{x^T x}{\|x\|_2} = \|x\|_2.$$

Torej velja $\|x\|_2 \leq 1$.

Ker se krogle ne da zapisati kot sistem končno mnogo linearnih enačb in neenačb, S_3 ni polieder.

Definicija 13. *Preslikava $f \colon \mathcal{U} \to \mathcal{V}$ med vektorskima prostoroma $\mathcal{U}$ in $\mathcal{V}$ je* **afina**, *če velja*

$$f((1-\lambda)x + \lambda y) = (1-\lambda) f(x) + \lambda f(y)$$

za vsaka $x, y \in \mathcal{U}$ in $\lambda \in \mathbb{R}$.

Trditev 5. *Vsaka afina preslikava $f \colon \mathbb{R}^n \to \mathbb{R}^m$ je oblike $f(x) = Ax + b$ za neko matriko $A \in \mathbb{R}^{m \times n}$ in vektor $b \in \mathbb{R}^m$.*

Dokaz. Ker je f afina, velja

$$f(\lambda x) = f(\lambda x + (1-\lambda)0) = \lambda f(x) + (1-\lambda) f(0) \tag{1.2}$$

in

$$\begin{aligned} f(x+y) = f\left(\frac{1}{2}2x + \frac{1}{2}2y\right) &= \frac{1}{2} f(2x) + \frac{1}{2} f(2y) \\ &= \frac{1}{2}(2f(x) - f(0)) + \frac{1}{2}(2f(y) - f(0)) = f(x) + f(y) - f(0). \end{aligned}$$

Definirajmo preslikavo $L \colon \mathbb{R}^n \to \mathbb{R}^m$ s predpisom $L(x) = f(x) - f(0)$. Ta preslikava je linearna, saj velja:

$$L(x+y) = f(x+y) - f(0) = f(x) + f(y) - 2f(0) = L(x) + L(y)$$

in

$$L(\lambda x) = f(\lambda x) - f(0) = \lambda f(x) + (1-\lambda) f(0) - f(0) = \lambda(f(x) - f(0)) = \lambda L(x).$$

Ker je L linearna, jo lahko zapišemo kot $L(x) = Ax$, za neko matriko $A \in \mathbb{R}^{m \times n}$. Če definiramo še $b = f(0)$, dobimo, da je $f(x) = Ax + b$. $\square$

Trditev 6. *Naj bo f afina preslikava. Potem je*

1. *slika $f(\mathcal{C})$ konveksne množice $\mathcal{C}$ konveksna:*

$$\mathcal{C} \text{ konveksna } \implies f(\mathcal{C}) = \{f(x) \mid x \in \mathcal{C}\} \text{ konveksna}$$

2. *praslika $f^{-1}(\mathcal{C})$ konveksne množice $\mathcal{C}$ konveksna:*

$$\mathcal{C} \text{ konveksna } \implies f^{-1}(\mathcal{C}) = \{x \mid f(x) \in \mathcal{C}\} \text{ konveksna}$$

Dokaz. 1. Vzemimo $y_1, y_2 \in f(\mathcal{C})$ in $\lambda \in [0,1]$. Potem obstajata $x_1, x_2 \in \mathcal{C}$, da je $y_1 = f(x_1)$ in $y_2 = f(x_2)$. Velja

$$(1-\lambda)y_1 + \lambda y_2 = (1-\lambda)f(x_1) + \lambda f(x_2) = f\left((1-\lambda)x_1 + \lambda x_2\right).$$

Zaradi konveksnosti množice $\mathcal{C}$ velja, da je $(1-\lambda)y_1 + \lambda y_2 \in f(\mathcal{C})$.

2. Vzemimo $x_1, x_2 \in f^{-1}(\mathcal{C})$ in $\lambda \in [0,1]$. Potem velja

$$f((1-\lambda)x_1 + \lambda x_2) = (1-\lambda)f(x_1) + \lambda f(x_2) \in \mathcal{C},$$

ker je množica $\mathcal{C}$ konveksna in $f(x_1), f(x_2) \in \mathcal{C}$. □

Primer 10. *Ker so raztegi, translacije, projekcije, rotacije in zrcaljenja afine preslikave, ohranjajo konveksnost množic.*

Naloga 9. *(VAJE) Dokaži, da je za $P \succ 0$ elipsoid*

$$\mathcal{E}(x_c) = \{x \mid (x-x_c)^T P (x-x_c) \leq 1\}$$

slika $f(\mathcal{B})$ enotske krogle $\mathcal{B}$, kjer je $f(u) = P^{-\frac{1}{2}}u + x_c$.

Rešitev: *Velja*

$$f(\mathcal{B}) = \{f(u) \mid \|u\| \leq 1\} = \left\{P^{-\frac{1}{2}}u + x_c \,|\, \|u\| \leq 1\right\}.$$

To pa je ravno množica točk x, ki zadošča neenačbi

$$(x-x_c)^T P (x-x_c) = u^T P^{-\frac{1}{2}} P P^{-\frac{1}{2}} u = \|u\|^2 \leq 1.$$

Ker je krogla konveksna množica in preslikava f afina, smo dokazali, da je elipsoid konveksna množica.

Naloga 10. *(VAJE) Dokaži, da je za dano matriko $P \in \mathcal{S}_+^n$ in vektor $c \in \mathbb{R}^n$ množica*

$$\mathcal{H} = \left\{x \mid x^T P x \leq (c^T x)^2,\ c^T x \geq 0\right\}$$

konveksna.

Rešitev: *Definirajmo afino preslikavo $f\colon \mathbb{R}^n \to \mathbb{R}^{n+1}$ s predpisom $f(x) = \left(P^{\frac{1}{2}}x, c^T x\right)$. Potem je množica $\mathcal{H}$ praslika Lorentzovega stožca*

$$\left\{(z,t) \mid z^T z \leq t^2,\ t \geq 0\right\}.$$

Zato je po Trditvi 6 konveksna.

Naloga 11. *(VAJE) Naj bodo dane simetrične matrike $A_1, \ldots, A_m$ in B. Dokaži, da je množica rešitev* linearne matrične neenakosti

$$\{x \in \mathbb{R}^m \mid x_1 A_1 + \ldots x_m A_m \preceq B\} \tag{1.3}$$

konveksna.

Rešitev: *Definirajmo afino preslikavo $f\colon \mathbb{R}^m \to \mathcal{S}^n$ s predpisom $f(x) = B - \sum_{i=1}^m x_i A_i$. Potem velja*

$$f^{-1}(\mathcal{S}_+^n) = \left\{x \in \mathbb{R}^m \;\middle|\; B - \sum_{i=1}^m x_i A_i \succeq 0\right\}.$$

Ker je stožec pozitivno semidefinitnih matrik konveksna množica, je množica (1.3) *po Trditvi 6 konveksna.*

1.2 Konveksne funkcije

Definicija 14. *Funkcija $f\colon \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ je* **konveksna**, *če je njena domena $\mathcal{D}$ konveksna množica in velja*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \tag{1.4}$$

za vsaka $x, y \in \mathcal{D}$ in vsak $\lambda \in [0,1]$. Funkcija f je **konkavna**, *če je $-f$ konveksna. Funkcija f je* **strogo konveksna**, *če je njena domena $\mathcal{D}$ konveksna množica in velja*

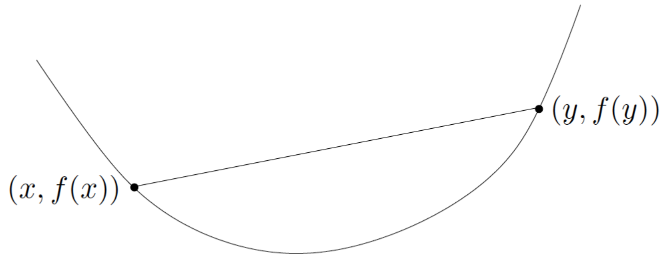
$$f((1-\lambda)x + \lambda y) < (1-\lambda)f(x) + \lambda f(y)$$

za vsaka $x, y \in \mathcal{D}$, $x \neq y$ in vsak $\lambda \in (0,1)$.

Geometrijsko lahko definicijo konveksne funkcije povemo takole: Funkcija je konveksna natanko tedaj, ko zveznica katerih koli dveh točk na njenem grafu leži nad grafom (ali na njem). Glej sliko 1.8.

Primer 11.

- *Vsaka afina funkcija je po definiciji konveksna (in konkavna).*



Slika 1.8: Graf konveksne funkcije.

- *Dane so funkcije ene spremenljivke:* e^{ax} *je konveksna na* $\mathbb{R}$ *za vsak* $a \in \mathbb{R}$, $-\ln x$ *in* $x \ln x$ *sta konveksni na* $\mathbb{R}_{++}$. x^a *je konveksna na* $\mathbb{R}_{++}$ *za* $a \geq 1$ *ali* $a \leq 0$, *in konkavna za* $a \in [0, 1]$. *Konveksnost pokažemo tako, da preverimo, da je drugi odvod nenegativen.*

- *Za dani vektorski prostor* $\mathcal{U}$ *je vsaka norma* $\|\cdot\|: \mathcal{U} \to \mathbb{R}$ *konveksna funkcija. Lastnost* (1.4) *sledi iz trikotniške neenakosti.*

  (VAJE) Na primer, funkcija $f: \mathbb{R}^{m\times n} \to \mathbb{R}$ *s predpisom* $f(X) = \|X\|_2 = \sigma_{\max}(X) = \sqrt{\lambda_{\max}(X^T X)}$ *je konveksna. Pri tem* $\lambda_{\max}(X)$ *in* $\sigma_{\max}(X)$ *označujeta največjo lastno in največjo singularno vrednost matrike* X.

- *(VAJE) Funkcija* $f: \mathbb{R}^n \to \mathbb{R}$ *definirana s predpisom* $f(x) = \max_i x_i$ *je konveksna, saj za* $\lambda \in [0, 1]$ *velja*

$$\begin{aligned} f((1-\lambda)x + \lambda y) &= \max_i((1-\lambda)x_i + \lambda y_i) \\ &\leq (1-\lambda)\max_i x_i + \lambda \max_i y_i \\ &= (1-\lambda)f(x) + \lambda f(y). \end{aligned}$$

Naloga 12. *(VAJE) Dokaži, da rang matrike* $r: \mathbb{R}^{n\times n} \to \mathbb{R}$ *ni konveksna funkcija.*

Rešitev: *Trditev bomo dokazali s protiprimerom. Vzemimo matriki*

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{in} \quad Y = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Potem je za vsak $\lambda \in (0, 1)$

$$r((1-\lambda)X + \lambda Y) = 2 > 1 = (1-\lambda)r(X) + \lambda r(Y).$$

V nadaljevanju bomo razvili orodja, s katerimi bomo lažje preverjali konveksnost funkcij.

Trditev 7. *Funkcija* $f: \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ *je konveksna natanko tedaj, ko je funkcija* $g: \mathbb{R} \to \mathbb{R}$ *definirana s predpisom*

$$g(t) = f(x + tv)$$

konveksna na domeni $\{t \mid x + tv \in \mathcal{D}\}$ *za vsak* $x \in \mathcal{D}$ *in vsak* $v \in \mathbb{R}^n$.

Dokaz. $(\Rightarrow)$ Vzemimo $a, b \in \{t \mid x + tv \in \mathcal{D}\}$ in $\lambda \in [0,1]$. Potem zaradi konveksnosti funkcije f velja

$$\begin{aligned} g((1-\lambda)a + \lambda b) &= f\left(x + ((1-\lambda)a + \lambda b)\, v\right) = f\left((1-\lambda)(x+av) + \lambda(x+bv)\right) \\ &\leq (1-\lambda)f(x+av) + \lambda f(x+bv) = (1-\lambda)g(a) + \lambda g(b). \end{aligned}$$

$(\Leftarrow)$ Naj bosta $x, y \in \mathcal{D}$ in $\lambda \in [0,1]$. Vzemimo $v = y - x$. Potem zaradi konveksnost funkcije g velja

$$\begin{aligned} f((1-\lambda)x + \lambda y) &= f(x + \lambda(y-x)) = g(\lambda) = g((1-\lambda)\cdot 0 + \lambda \cdot 1) \\ &\leq (1-\lambda)g(0) + \lambda g(1) = (1-\lambda)f(x) + \lambda f(y). \end{aligned}$$

□

Opomba 2. *Podobno dokažemo, da je zvezno odvedljiva funkcija f strogo konveksna natanko tedaj, ko je funkcija $g\colon \mathbb{R} \to \mathbb{R}$ definirana s predpisom*

$$g(t) = f(x + tv)$$

strogo konveksna.

Trditev pove, da je funkcija konveksna natanko tedaj, ko je njena zožitev na poljubno premico konveksna funkcija ene spremenljivke.

Primer 12. *Dokažimo, da je funkcija $f\colon \mathcal{S}^n_{++} \to \mathbb{R}$ definirana s predpisom*

$$f(X) = -\ln\det(X)$$

strogo konveksna. Funkciji f pravimo **pregradna ali barierna funkcija** *pozitivno semidefinitnega stožca $\mathcal{S}^n_+$.*

Naj bo $X \succ 0$ in $V \in \mathcal{S}^n$. Po Trditvi 7 je dovolj preveriti strogo konveksnost funkcije $g(t) = -\ln\det(X + tV)$. Ker je $X \succ 0$, obstaja $X^{\frac{1}{2}}$. Potem velja

$$\begin{aligned} g(t) &= -\ln\det(X + tV) = -\ln\det\Big(X^{\frac{1}{2}}(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})X^{\frac{1}{2}}\Big) \\ &= -\ln\Big(\det\Big(X^{\frac{1}{2}}\Big)\det\Big(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}\Big)\det\Big(X^{\frac{1}{2}}\Big)\Big) \\ &= -\ln\det(X) - \ln\det\Big(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}\Big) \\ &= -\ln\det(X) - \sum_{i=1}^{n}\ln(1 + t\lambda_i), \end{aligned}$$

kjer so λ_i lastne vrednosti matrike $X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$. Pri tem smo upoštevali, da je determinanta enaka produktu lastnih vrednosti in da so lastne vrednosti matrike $I + A$ enake $1 + \lambda_i(A)$. Ker pa velja

$$g'(t) = -\sum_{i=1}^{n}\frac{\lambda_i}{1 + t\lambda_i} \quad \textit{in} \quad g''(t) = \sum_{i=1}^{n}\frac{\lambda_i^2}{(1 + t\lambda_i)^2}$$

in je drugi odvod pozitiven, je funkcija g strogo konveksna.

Definicija 15. *Naj bo $f : \mathbb{R}^n \to \mathbb{R}$ zvezno odvedljiva funkcija.* **Gradient** *funkcije f izračun v točki x je (stolpični) vektor*

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T.$$

Za funkcijo $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ definiramo $\nabla f(X)$ kot matriko, za katero velja

$$\nabla f(X)_{ij} = \frac{\partial f(X)}{\partial x_{ij}}.$$

Smerni odvod funkcije f v smeri vektorja $v \in \mathbb{R}^n$ je definiran kot

$$\nabla_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}.$$

Opomba 3. *Za smerni odvod odvedljive funkcije f v točki x v smeri v velja*

$$\nabla_v f(x) = \nabla f(x)^T v.$$

Trditev 8 (***karakterizacija konveksnosti s 1. odvodom***). *Naj bo $\mathcal{D}$ odprta konveksna množica v $\mathbb{R}^n$ in $f : \mathcal{D} \to \mathbb{R}$ zvezno odvedljiva funkcija na $\mathcal{D}$. Potem je f konveksna na $\mathcal{D}$ natanko tedaj, ko za vse $x, y \in \mathcal{D}$ velja:*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x). \tag{1.5}$$

Dokaz. ($\Rightarrow$) Naj bo f konveksna. Potem po definiciji za vsaka $x, y \in \mathcal{D}$ in vsak $\lambda \in [0, 1]$ velja

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda) f(x) + \lambda f(y).$$

Če preuredimo, dobimo

$$f(x + \lambda(y - x)) - f(x) \leq \lambda(f(y) - f(x))$$

oziroma

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

Ko pošljemo $\lambda \downarrow 0$, dobimo na desni strani neenakosti smerni odvod funkcije f v smeri vektorja $y - x$. Od tod sledi

$$f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

($\Leftarrow$) Naj velja neenakost (1.5). Vzemimo $x, y \in \mathcal{D}$ in parameter $\lambda \in [0, 1]$. Naj bo

$$z = (1 - \lambda)x + \lambda y.$$

Potem velja

$$f(x) \geq f(z) + \nabla f(z)^T (x - z)$$

in

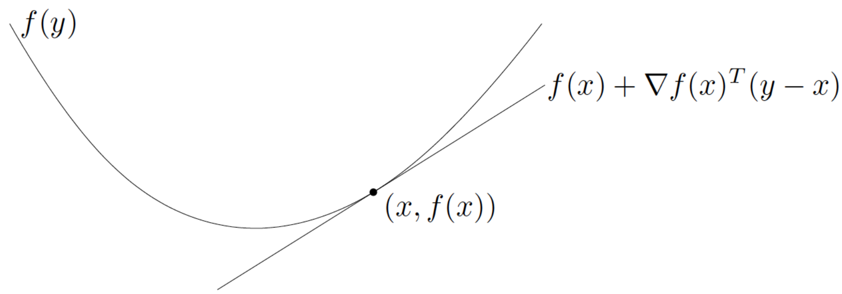
$$f(y) \geq f(z) + \nabla f(z)^T (y - z).$$

Če prvo neenakost pomnožimo z $1-\lambda$, drugo z λ in ju seštejemo, dobimo

$$
\begin{aligned}
(1-\lambda)f(x)+\lambda f(y) &\geq f(z)+\nabla f(z)^T((1-\lambda)x+\lambda y-z)\\
(1-\lambda)f(x)+\lambda f(y) &\geq f(z)\\
(1-\lambda)f(x)+\lambda f(y) &\geq f((1-\lambda)x+\lambda y).
\end{aligned}
$$

Torej je f koveksna. □

Geometrijsko lahko trditev povemo takole: Gladka funkcija je konveksna natanko tedaj, ko tangentna hiperravnina na graf funkcije v kateri koli točki leži pod grafom funkcije (ali se ga dotika). Glej sliko 1.9.



Slika 1.9: Za zvezno odvedljive funkcije je konveksnost ekvivalentna pogoju $f(y) \geq f(x) + \nabla f(x)^T(y-x)$ za vsaka x, y.

Opomba 4. *Zvezno odvedljiva funkcija je strogo konveksna na odprti konveksni množici $\mathcal{D}$ natanko tedaj, ko za vse $x, y \in \mathcal{D}, x \neq y$, velja*

$$f(y) > f(x) + \nabla f(x)^T(y-x).$$

Trditev 9 (*karakterizacija konveksnosti z 2. odvodom*). *Naj bo $\mathcal{D}$ odprta konveksna množica v $\mathbb{R}^n$ in $f\colon \mathcal{D} \to \mathbb{R}$ dvakrat zvezno odvedljiva funkcija na $\mathcal{D}$. Potem je f konveksna na $\mathcal{D}$ natanko tedaj, ko je njena Hessejeva matrika*

$$\nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right]_{i,j=1}^n$$

pozitivno semidefinitna za vse $x \in \mathcal{D}$.

Dokaz. Po Trditvi 7 velja, da je funkcija f konveksna natako tedaj, ko je funkcija $g(t) = f(x+tv)$ konveksna za vsak $x \in \mathcal{D}$ in vsak v. Za odvode funkcije g velja

$$g'(t) = v^T \nabla f(x+tv) \quad \text{in} \quad g''(t) = v^T \nabla^2 f(x+tv) v.$$

Iz analize vemo, da je funkcija ene spremenljivke konveksna na nekem območju natanko tedaj, ko je njen drugi odvod tam nenegativen. Torej je $g''(t) \geq 0$ za poljubna x in v, za katera velja $x+tv \in \mathcal{D}$ natanko tedaj, ko je Hessejeva matrika pozitivno semidefinitna na $\mathcal{D}$. □

Opomba 5. *Za strogo konveksne funkcije velja trditev samo v eno smer, tj. če je Hessejeva matrika funkcije f pozitivno definitna, potem je funkcija f strogo konveksna. To vidimo iz Taylorjevefa razvoja funkcije f. Velja*

$$f(y) = f(x) + \nabla f(y)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + t(y - x))(y - x)$$

za nek $t \in [0, 1]$. Če je torej Hessejeva matrika pozitivna definitna, dobimo neenakost

$$f(y) > f(x) + \nabla f(y)^T (y - x),$$

kar potrjuje strogo konveksnost funkcije f. Obrat v splošnem ne velja. Na primer, $f(x) = x^4$ je strogo konveksna, vendar $f''(0) = 0$.

Lema 4. *Za dane $A \in \mathcal{S}^n$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ ter funkciji $f, g \colon \mathbb{R}^n \to \mathbb{R}$ podani s predpisoma*

$$f(x) = b^T x + c = \sum_{i=1}^{n} b_i x_i + c$$

in

$$g(x) = x^T A x + b^T x + c = \sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c = \sum_{i=1}^{n} \left(a_{ii} x_i^2 + b_i x_i + \sum_{j \neq i} a_{ij} x_i x_j \right) + c$$

velja

$$\nabla f(x) = b, \quad \nabla g(x) = 2Ax + b \quad \textit{in} \quad \nabla^2 g(x) = 2A.$$

Dokaz. Parcialni odvodi so enaki

$$\begin{aligned}
\frac{\partial f}{\partial x_k}(x) &= b_k, \\
\frac{\partial g}{\partial x_k}(x) &= 2a_{kk} x_k + \sum_{j \neq k} a_{kj} x_j + \sum_{i \neq k} a_{ik} x_i + b_k \\
&= \sum_{j=1}^{n} a_{kj} x_j + \sum_{j=1}^{n} a_{jk} x_j + b_k = 2 \sum_{j=1}^{n} a_{kj} x_j + b_k, \\
\frac{\partial^2 g}{\partial x_k x_l}(x) &= 2a_{kl}.
\end{aligned}$$

□

Naloga 13. *(VAJE) Izračunaj gradienta naslednjih funkcij.*

1. *Za dano matriko $A \in \mathbb{R}^{m \times n}$ definiramo funkcijo $f \colon \mathbb{R}^{m \times n} \to \mathbb{R}$ s predpisom*

$$f(X) = \langle A, X \rangle = \operatorname{sled}(A^T X).$$

Potem je $\nabla f(X) = A$.

2. *Dana je funkcija $f\colon \mathcal{S}^n_{++} \to \mathbb{R}$ s predpisom $f(X) = \ln\det(X)$. Potem velja $\nabla f(X) = X^{-1}$.*

Rešitev:

1. *Za $f(X) = \langle A, X\rangle = \mathrm{sled}(A^T X) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_{ij}$ velja*
$$\frac{\partial f(X)}{\partial x_{ij}} = a_{ij}.$$
Od tod sledi, da je $\nabla f(X) = A$.

2. *Po verižnem pravilu za odvajanje dobimo*
$$\nabla \ln\det(X) = \frac{1}{\det(X)}\nabla\det(X).$$
Gradient determinante bomo izračunali s pomočjo razvoja determinante po i-ti vrstici.
$$\frac{\partial}{\partial x_{ij}}\left(\sum_j x_{ij}C_{ij}\right) = C_{ij}.$$
Pri tem smo z C_{ij} označili kofaktor elementa x_{ij}. Od tod sledi $\nabla\det(X) = C$, pri čemer je C matrika kofaktorjev. Po formuli za inverz matrike X
$$X^{-1} = \frac{1}{\det(X)}C^T,$$
dobimo, da velja $\nabla \ln\det(X) = X^{-1}$.

Primer 13. *Kvadratična funkcija*
$$f(x) = \|Ax-b\|_2^2 = x^TA^TAx - 2\left(A^Tb\right)^T x + b^Tb.$$
je konveksna, saj po Lemi 4 velja, da je $\nabla^2 f(x) = 2A^TA \succeq 0$ za vsak x. Splošneje, za $P \in \mathcal{S}^n$ je kvadratična funkcija
$$f(x) = \frac{1}{2}x^TPx + q^Tx + r$$
konveksna natanko tedaj, ko velja $P \succeq 0$.

Naloga 14. *(VAJE) Dokaži, da je funkcija $f\colon \mathbb{R}\times\mathbb{R}_{++} \to \mathbb{R}$ definirana s predpisom*
$$f(x,y) = \frac{x^2}{y}$$
konveksna.

Rešitev: *Za Hessejevo matriko velja*
$$\nabla^2 f(x,y) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3}\end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & \frac{2x^2}{y^3}\end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x\end{bmatrix}\begin{bmatrix} y \\ -x\end{bmatrix}^T.$$
Ker so matrike ranga 1 po definiciji pozitivno semidefinitne, je f konveksna funkcija.

Naloga 15. *(VAJE) Dokaži, da je funkcija $f\colon \mathbb{R}^n \to \mathbb{R}$ definirana s predpisom*

$$f(x) = \ln\left(\sum_{k=1}^{n} e^{x_k}\right)$$

konveksna.

Rešitev: *Parcialni odvodi 2. reda so enaki*

$$\frac{\partial^2 f}{\partial x_k x_l}(x) = \frac{-e^{x_k}e^{x_l}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2} \quad \text{in} \quad \frac{\partial^2 f}{\partial x_k^2}(x) = \frac{e^{x_k}}{\sum_{k=1}^{n} e^{x_k}} - \frac{-e^{x_k}e^{x_k}}{\left(\sum_{k=1}^{n} e^{x_k}\right)^2}.$$

Če označimo $z_k = e^{x_k}$, lahko zapišemo Hessejevo matriko v kompaktni obliki kot

$$\nabla^2 f(x) = \frac{1}{e^T z}\operatorname{Diag}(z) - \frac{1}{(e^T z)^2} z z^T.$$

Pozitivno semidefinitnost Hesssejeve matrike preverimo po definicijo. Naj bo $v \in \mathbb{R}^n$ poljuben vektor. Potem velja

$$\begin{aligned} v^T \nabla^2 f(x) v &= \frac{1}{e^T z}\sum_{k=1}^{n} z_k v_k^2 - \frac{1}{(e^T z)^2}\left(\sum_{k=1}^{n} z_k v_k\right)^2 \\ &= \frac{\left(\sum_{k=1}^{n} z_k v_k^2\right)\left(\sum_{k=1}^{n} z_k\right) - \left(\sum_{k=1}^{n} z_k v_k\right)^2}{(e^T z)^2}. \end{aligned}$$

Če vzamemo vektorja $(\sqrt{z_1}, \ldots, \sqrt{z_n})$ in $(\sqrt{z_1}v_1, \ldots, \sqrt{z_n}v_n)$, potem po Cauchy-Schwarzovi neenakosti velja

$$\left(\sum_{k=1}^{n} z_k v_k\right)^2 = \left(\sum_{k=1}^{n} \sqrt{z_k}\sqrt{z_k} v_k\right)^2 \leq \left(\sum_{k=1}^{n} z_k v_k^2\right)\left(\sum_{k=1}^{n} z_k\right).$$

Zato je $v^T \nabla^2 f(x) v \geq 0$.

Naloga 16. *(VAJE) Katere od spodnjih funkcij so konveksne oziroma konkavne?*

(a) Funkcija $f_1\colon (-1,1) \to \mathbb{R}$ je definirana s predpisom $f_1(x) = \frac{1}{1-x^2}$.

(b) Funkcija $f_2\colon \mathbb{R}^2 \to \mathbb{R}$ je definirana s predpisom $f_2(x_1,x_2) = x_1 x_2$.

(c) Funkcija $f_3\colon \mathbb{R}^2_{++} \to \mathbb{R}$ je definirana s predpisom $f_3(x_1,x_2) = -\frac{1}{x_1 x_2}$.

(d) Funkcija $f_4\colon \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}$ je definirana s predpisom $f_4(x_1,x_2) = \frac{x_1^2}{x_2}$.

Rešitev: *Konveksnost oziroma konkavnost bomo določili s pomočjo drugega odvoda oziroma Hessejevih matrik.*

(a) *Za drugi odvod velja* $f''(x) = \frac{6x^2+2}{(1-x^2)^3} > 0$. *Zato je funkcija* f_1 *strogo konveksna.*

(b) *Hessejeva matrika* $\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *ni niti pozitivno semidefinitna niti negativno semidefinitna. Zato funkcija* f_2 *ni konveksna niti konkavna.*

(c) *Hessejeva matrika* $\nabla^2 f(x) = -\frac{1}{x_1x_2} \begin{bmatrix} \frac{2}{x_1^2} & \frac{1}{x_1x_2} \\ \frac{1}{x_1x_2} & \frac{2}{x_2^2} \end{bmatrix}$ *je negativno semidefinitna, saj velja*

$$\operatorname{sled}(\nabla^2 f(x)) = -\frac{1}{x_1x_2}\left(\frac{2}{x_1^2} + \frac{2}{x_2^2}\right) < 0 \quad \textit{in} \quad \det\big(\nabla^2 f(x)\big) = \frac{3}{x_1^4x_2^4} > 0.$$

Torej je funkcija f_3 *strogo konkavna.*

(d) *Za Hessejevo matriko velja*

$$\nabla^2 f(x) = \begin{bmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{bmatrix} = \frac{2}{x_2}\begin{bmatrix} 1 \\ -\frac{x_1}{x_2} \end{bmatrix}\begin{bmatrix} 1 & -\frac{x_1}{x_2} \end{bmatrix} \succeq 0.$$

Torej je funkcija f_4 *konveksna.*

Povezavo med konveksnimi množicami in konveksnimi funkcijami povzemata naslednja lema in trditev.

Lema 5. *Naj bo funkcija* $f\colon \mathbb{R}^n \to \mathbb{R}$ *konveksna. Potem je množica*

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid f(x) \leq 0\}$$

konveksna.

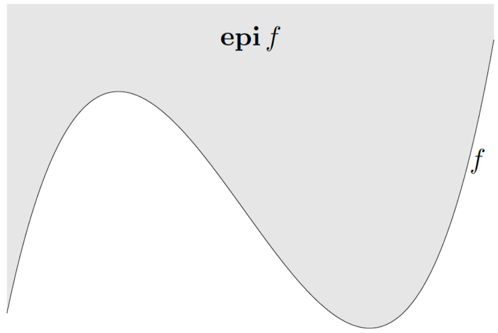
Dokaz. Vzemimo $x, y \in \mathcal{F}$ in $\lambda \in [0,1]$. Potem je $(1-\lambda)x + \lambda y \in \mathcal{F}$, saj velja

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \leq 0.$$ □

Definicija 16. Epigraf *funkcije* $f\colon \mathbb{R}^n \to \mathbb{R}$ *je množica točk, ki ležijo na grafu funkcije* f *ali nad njim:*

$$\operatorname{epi}(f) = \left\{(x,t) \in \mathbb{R}^{n+1} \mid f(x) \leq t\right\}.$$

Trditev 10. *Funkcija* $f\colon \mathbb{R}^n \to \mathbb{R}$ *je konveksna natanko tedaj, ko je* $\operatorname{epi}(f)$ *konveksna množica.*



Slika 1.10: Epigraf funkcije f.

Dokaz. Naj bo f konveksna funkcija. Vzemimo $(x_1, t_1), (x_2, t_2) \in \operatorname{epi}(f)$ in $\lambda \in [0, 1]$. Dokažimo, da $(1-\lambda)(x_1, t_1) + \lambda(x_2, t_2) \in \operatorname{epi}(f)$:

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \leq (1-\lambda)t_1 + \lambda t_2.$$

Obratno, denimo, da je epigraf konveksna množica. Vzemimo $x_1, x_2 \in \mathbb{R}^n$ in $\lambda \in [0, 1]$. Točki $(x_1, f(x_1))$ in $(x_2, f(x_2))$ pripadata epigrafu funkcije f. Zaradi konveksnosti množice $\operatorname{epi}(f)$ je potem

$$(1-\lambda)(x_1, f(x_1)) + \lambda(x_2, f(x_2)) = ((1-\lambda)x_1 + \lambda x_2, (1-\lambda)f(x_1) + \lambda f(x_2)) \in \operatorname{epi}(f).$$

Torej velja

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2)$$

in zato je funkcija f konveksna. □

Primer 14. *Epigraf 2-norme je Lorenzov stožec $\mathcal{L}^{n+1}$ (glej Primer 5).*

Naloga 17. *(VAJE) Dokaži, da je množica*

$$E = \left\{(x, y) \in \mathbb{R}^2_{++} \mid xy \geq 1\right\}$$

konveksna.

Rešitev: *Množica E je epigraf funkcije $f\colon \mathbb{R}_{++} \to \mathbb{R}$ definirane s predpisom $f(x) = \frac{1}{x}$. Ker je f konveksna funkcija $\left(f''(x) = \frac{2}{x^3} > 0\right)$, je množica E po Trditvi 10 konveksna.*

V nadaljevanju si bomo ogledali operacije, ki ohranjajo konveksnost funkcij.

Trditev 11. *1. Naj bosta $f_1, f_2\colon \mathbb{R}^n \to \mathbb{R}$ konveksni funkciji in $\alpha_1, \alpha_2 \geq 0$. Potem je nenegativna kombinacija*

$$\alpha_1 f_1 + \alpha_2 f_2$$

funkcij f_1 in f_2 konveksna funkcija.

2. *Naj bo $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ in funkcija $f\colon \mathbb{R}^m \to \mathbb{R}$ konveksna. Potem je kompozitum*
$$f(Ax+b)$$
funkcije f in afine funkcije konveksna funkcija.

3. *Naj bo $g\colon \mathbb{R}^n \to \mathbb{R}$ konveksna funkcija in $f\colon \mathbb{R} \to \mathbb{R}$ konveksna in nepadajoča. Potem je kompozitum*
$$f \circ g$$
konveksna funkcija.

4. *Naj bodo $f_1, \dots, f_m$ konveksne funkcije. Potem je*
$$f(x) = \max\{f_1(x), \dots, f_m(x)\}$$
konveksna funkcija.

5. *Naj bo $\mathcal{A}$ dana množica in naj bo $f\colon \mathbb{R}^n \times \mathcal{A} \to \mathbb{R}$ konveksna funkcija v spremenljivki x za vsak $y \in \mathcal{A}$. Potem je*
$$g(x) = \sup_{y\in\mathcal{A}} f(x,y)$$
konveksna funkcija.

6. *Naj bo $\mathcal{C}$ konveksna množica in naj bo $f\colon \mathbb{R}^n \times \mathcal{C} \to \mathbb{R}$ konveksna funkcija. Potem je*
$$g(x) = \inf_{y\in\mathcal{C}} f(x,y)$$
konveksna funkcija.

Dokaz. Konveksnost funkcij preverimo po definiciji. Vzemimo $x, y \in \mathbb{R}^n$ in $\lambda \in [0,1]$. Potem velja

1.
$$\begin{aligned}(\alpha_1 f_1 + \alpha_2 f_2)\left((1-\lambda)x + \lambda y\right) &= \alpha_1 f_1\left((1-\lambda)x+\lambda y\right) + \alpha_2 f_2\left((1-\lambda)x+\lambda y\right)\\ &\le \alpha_1(1-\lambda)f_1(x) + \alpha_1\lambda f_1(y) + \alpha_2(1-\lambda)f_2(x) + \alpha_2\lambda f_2(y)\\ &= (1-\lambda)(\alpha_1 f_1+\alpha_2 f_2)(x) + \lambda(\alpha_1 f_1+\alpha_2 f_2)(y)\end{aligned}$$

2.
$$\begin{aligned}f\left(A((1-\lambda)x+\lambda y)+b\right) &= f((1-\lambda)(Ax+b)+\lambda(Ay+b))\\ &\le (1-\lambda)f(Ax+b)+\lambda f(Ay+b)\end{aligned}$$

3.
$$\begin{aligned}(f\circ g)((1-\lambda)x+\lambda y) &= f\left(g((1-\lambda)x+\lambda y)\right)\\ &\overset{(1)}{\le} f((1-\lambda)g(x)+\lambda g(y))\\ &\overset{(2)}{\le} (1-\lambda)(f(g(x))+\lambda f(g(y))\\ &= (1-\lambda)(f\circ g)(x)+\lambda(f\circ g)(y)\end{aligned}$$

Pri (1) smo upoštevali konveksnost funkcije g in da je funkcija f nepadajoča, pri (2) pa konveksnost funkcije f.

4.
$$\begin{aligned} f((1-\lambda)x+\lambda y) &= f_i((1-\lambda)x+\lambda y) \\ &\leq (1-\lambda)f_i(x)+\lambda f_i(y) \\ &\leq (1-\lambda)f(x)+\lambda f(y) \end{aligned}$$

5. Naj bosta $x_1, x_2 \in \mathbb{R}^n$. Po definiciji funkcije g velja
$$g((1-\lambda)x_1+\lambda x_2) = \sup_{y\in\mathcal{A}} f((1-\lambda)x_1+\lambda x_2, y).$$
Po definiciji supremuma potem za vsak $\varepsilon > 0$ obstaja y_ε, da velja
$$\sup_{y\in\mathcal{A}} f((1-\lambda)x_1+\lambda x_2, y) \leq f((1-\lambda)x_1+\lambda x_2, y_\varepsilon)+\varepsilon.$$
Zaradi konveksnosti funkcije f sledi
$$\begin{aligned} g((1-\lambda)x_1+\lambda x_2) &= \sup_{y\in\mathcal{A}} f((1-\lambda)x_1+\lambda x_2, y) \\ &\leq f((1-\lambda)x_1+\lambda x_2, y_\varepsilon)+\varepsilon \\ &\leq (1-\lambda)f(x_1,y_\varepsilon)+\lambda f(x_2,y_\varepsilon)+\varepsilon \\ &\leq (1-\lambda)\sup_{y\in\mathcal{A}} f(x_1,y)+\lambda\sup_{y\in\mathcal{A}} f(x_2,y)+\varepsilon \\ &= (1-\lambda)g(x_1)+\lambda g(x_2)+\varepsilon. \end{aligned}$$
Ker to velja za vsak $\varepsilon > 0$, dobimo želeno neenakost.

6. Naj bosta $x_1, x_2 \in \mathbb{R}^n$. Naj bo $\varepsilon > 0$. Po definiciji infimuma obstajata taka $y_1, y_2 \in \mathcal{C}$, da za $i = 1, 2$ velja
$$f(x_i, y_i) \leq \inf_{y\in\mathcal{C}} f(x_i, y)+\frac{\varepsilon}{2}.$$
Potem zaradi konveksnosti množice $\mathcal{C}$ in konveksnosti funkcije f sledi
$$\begin{aligned} g((1-\lambda)x_1+\lambda x_2) &= \inf_{y\in\mathcal{A}} f((1-\lambda)x_1+\lambda x_2, y) \\ &\leq f((1-\lambda)x_1+\lambda x_2, (1-\lambda)y_1+\lambda y_2) \\ &\leq (1-\lambda)f(x_1,y_1)+\lambda f(x_2,y_2) \\ &\leq (1-\lambda)g(x_1)+\lambda g(x_2)+\varepsilon. \end{aligned}$$
Ker to velja za vsak $\varepsilon > 0$, dobimo želeno neenakost. □

Opomba 6. *Produkt in kompozitum konveksnih funkcij v splošnem nista konveksni funkciji. Na primer, za konveksni funkciji $f(x) = -x$ in $g(x) = x^2$ velja*

$$(fg)(x) = -x^3 \quad \text{in} \quad (f \circ g)(x) = -x^2.$$

Posledica 1. *Naj bo $\mathcal{A}$ dana množica in naj bo $f\colon \mathbb{R}^n \times \mathcal{A} \to \mathbb{R}$ konkavna funkcija v spremenljivki x za vsak $y \in \mathcal{A}$. Potem je*

$$g(x) = \inf_{y \in \mathcal{A}} f(x, y)$$

konkavna funkcija.

Dokaz. Funkcijo g lahko zapišemo kot

$$g(x) = \inf_{y \in \mathcal{A}} f(x, y) = -\sup_{y \in \mathcal{A}}((-f(x, y))$$

Ker je $-f$ konveksna v spremenljivki x, velja po 5. točki Trditve 11, da je funkcija $\sup_{y \in \mathcal{A}}((-f(x, y))$ konveksna. Torej je g konkavna. ☐

Primer 15. *(VAJE)*

- *Za dano matriko $A \in \mathbb{R}^{m \times n}$ in vektor $b \in \mathbb{R}^m$ je funkcija $f\colon \mathbb{R}^n \to \mathbb{R}$ definirana s predpisom*

  $$f(x) = \|Ax + b\|$$

  konveksna. Vsaka norma je konveksna funkcija, f pa je kompozitum norme in afine funkcije. Zato je po 2. točki Trditve 11 konveksna funkcija.

- *Dana je matrika $A \in \mathbb{R}^{m \times n}$ in vektor $b \in \mathbb{R}^m$.* **Pregradna ali barierna funkcija** *poliedra $\mathcal{P}$, ki je podan s sistemom linearih neenačb $Ax \leq b$ je funkcija $f\colon \operatorname{int}(\mathcal{P}) \to \mathbb{R}$ definirana s predpisom*

  $$f(x) = -\sum_{i=1}^{m} \ln\left(b_i - a_i^T x\right).$$

  Funkcija f je kompozitum konveksne in afine funkcije, zato je po 2. točki Trditve 11 konveksna funkcija.

- *Za dano matriko $A \in \mathbb{R}^{m \times n}$ in vektor $b \in \mathbb{R}^m$ je odsekoma linearna funkcija $f\colon \mathbb{R}^n \to \mathbb{R}$ definirana s predpisom*

  $$f(x) = \max_{i=1,\ldots,m} \left(a_i^T x + b_i\right)$$

  po 4. točki Trditve 11 konveksna funkcija.

- *Funkcija* $g\colon \mathcal{S}^n \to \mathbb{R}$ *definirana s predpisom*

$$g(X) = \sup_{\|y\|_2=1} y^T X y$$

je konveksna. Najprej dokažimo, da za vsak $X \in \mathcal{S}^n$ *velja*

$$g(X) = \lambda_{\max}(X),$$

kjer $\lambda_{\max}(X)$ *označuje največjo lastno vrednost matrike* X*. Ker je* X *simetrična matrika, obstaja ortonormirana baza iz lastnih vektorjev* $\{v_1, \dots, v_n\}$*. Pripadajoče lastne vrednosti označimo z* λ_i*,* $i = 1, \dots, n$*. Razvoj poljubnega vektorja* y *po tej bazi je potem enak* $y = \sum_{i=1}^{n} \alpha_i v_i$*, kjer je* $\sum_{i=1}^{n} \alpha_i^2 = 1$ *zaradi pogoja* $\|y\|_2 = 1$*. Od tod dobimo*

$$Xy = \sum_{i=1}^{n} \alpha_i X v_i = \sum_{i=1}^{n} \alpha_i \lambda_i v_i$$

in

$$y^T X y = \sum_{i=1}^{n} \alpha_i^2 \lambda_i \le \lambda_{\max}(X) \sum_{i=1}^{n} \alpha_i^2 = \lambda_{\max}(X).$$

Enakost je dosežena, če za y *vzamemo lastni vektor, ki pripada največji lastni vrednosti.*

Opazimo, da je funkcija $f(X, y) = y^T X y$ *za vsak enotski vektor* y *linearna funkcija v spremenljivki* X *(torej konveksna), saj velja*

$$y^T X y = \mathrm{sled}(y^T X y) = \mathrm{sled}(X y y^T) = \langle X, y y^T \rangle.$$

Zato je po 5. točki Trditve 11 funkcija g *konveksna.*

Podobno pokažemo, da je tudi $\lambda_{\min}(X)$ *konveksna funkcija.*

- *Dana je konveksna množica* $\mathcal{C} \subseteq \mathbb{R}^n$*. Funkcija* $g\colon \mathbb{R}^n \to \mathbb{R}$*, ki je definirana s predpisom*

$$g(x) = \inf_{y \in \mathcal{C}} \|x - y\|$$

in določa razdaljo točke x *do množice* $\mathcal{C}$*, je konveksna.*

Funkcija

$$f(x, y) = \|x - y\| = \left\| \begin{bmatrix} I & -I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right\|$$

je namreč po 2. točki Trditve 11 konveksna, saj je kompozitum norme in afine funkcije. Po 6. točki iste trditve je potem g *konveksna.*

Poglavje 2

Konveksni optimizacijski problemi

Definicija 17. **Splošni optimizacijski problem** *v standardni obliki je problem*

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p. \end{aligned} \tag{2.1}$$

Problem sprašuje po tisti spremenljivki x, ki zadošča vsem omejitvam $f_i(x) \leq 0$, $i = 1, \ldots, m$ in $h_j(x) = 0$, $j = 1, \ldots, p$, ter minimizira funkcijo $f_0(x)$. Funkciji f_0 pravimo **kriterijska (tudi namenska funkcija)***, neznanki $x \in \mathbb{R}^n$ pa* **optimizacijska spremenljivka***. Element $x \in \mathbb{R}^n$ je* **dopustna rešitev** *optimizacijskega problema* (2.1)*, če je v domeni $\mathcal{D} = \bigcap_{i=0}^m \mathcal{D}_{f_i} \cap \bigcap_{j=1}^p \mathcal{D}_{h_j}$ (implicitne omejitve) ter zadošča vsem neenakostim $f_i(x) \leq 0$, $i = 1, \ldots, m$ in enakostim $h_j(x) = 0$, $j = 1, \ldots, p$ (eksplicitne omejitve). Pravimo, da je problem* **nedopusten***, če je množica dopustnih rešitev prazna, sicer pa je* **dopusten***. Problemu* (2.1) *brez eksplicitnih omejitev $(m = p = 0)$ pravimo* **nevezan optimizacijski problem***.*

Primer 16. *Optimizacijski problem*

$$\text{minimiziraj} \quad f_0(x) = -\sum_{i=1}^k \ln\left(b_i - a_i^T x\right)$$

je nevezan optimizacijski problem z implicitnimi omejitvami $a_i^T x < b_i$, $i = 1, \ldots, k$.

Naj

$$\mathcal{F} = \{x \in \mathcal{D} \mid f_i(x) \leq 0,\ i = 1, \ldots, m,\ h_j(x) = 0,\ j = 1, \ldots, p\}$$

označuje **množico dopustnih rešitev** *problema* (2.1). **Optimalna vrednost problema** (2.1) *je definirana kot*

$$p^* = \inf\left\{f_0(x) \mid x \in \mathcal{F}\right\}.$$

Če je problem nedopusten, potem definiramo $p^* = \infty$. V primeru, ko je problem navzdol neomejen pa velja $p^* = -\infty$. Dopustni rešitvi $x^* \in \mathcal{F}$ problema (2.1) *pravimo* **optimalna rešitev ali globalni minimum***, če velja*

$$f_0(x^*) = p^* \quad \textit{oziroma} \quad f(x) \leq f(y) \ \textit{ za vsak } y \in \mathcal{F}.$$

Točka $x \in \mathcal{F}$ je **lokalni minimum**, *če*

$$\exists R > 0 \;\; \forall y \in \mathcal{F} : \; \|x - y\| \leq R \implies f(x) \leq f(y).$$

Vsak globalni ekstrem je tudi lokalni ekstrem, obratno pa v splošnem seveda ne velja.

Opomba 7. *Vsako maksimizacijsko nalogo preprosto prevedemo na minimizacijsko nalogo, saj velja*

$$\sup\{f_0(x) \mid x \in \mathcal{F}\} = -\inf\{-f_0(x) \mid x \in \mathcal{F}\}.$$

Primer 17. *Vzemimo $n = 1$, $m = p = 0$, tj. nevezan optimizacijski problem v eni spremenljivki.*

- *Za $f_0(x) = \frac{1}{x}$ z domeno $\mathcal{D} = \mathbb{R}_{++}$ velja $p^* = 0$, vendar optimalna rešitev x^* ne obstaja.*
- *Za $f_0(x) = -\ln x$ z domeno $\mathcal{D} = \mathbb{R}_{++}$ velja $p^* = -\infty$.*
- *Za $f_0(x) = x \ln x$ z domeno $\mathcal{D} = \mathbb{R}_{++}$ velja $p^* = -\frac{1}{e}$ in $x^* = \frac{1}{e}$.*
- *Za $f_0(x) = x^3 - 3x$ z domeno $\mathcal{D} = \mathbb{R}$ velja $p^* = -\infty$. Pri $x = 1$ je lokalni minimum.*

Definicija 18. Konveksni optimizacijski problem *v standardni obliki je problem*

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_j^T x = b_j, \quad j = 1, \ldots, p, \end{aligned}$$

kjer so $f_0, f_1, \ldots, f_m \colon \mathbb{R}^n \to \mathbb{R}$ konveksne funkcije in $h_j(x) = a_j^T x - b_j$, $j = 1, \ldots, p$, afine. V kompaktni obliki ga zapišemo kot

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b. \end{aligned} \tag{2.2}$$

Za konveksni optimizacijski problem (2.2) ***torej velja, da minimiziramo (maksimiziramo) konveksno (konkavno) funkcijo po konveksni množici***. Da bo množica dopustnih rešitev

$$\mathcal{F} = \{x \in \mathcal{D} \mid f_i(x) \leq 0, \; i = 1, \ldots, m, \; Ax = b\}$$

optimizacijskega problema (2.2) konveksna, smo morali zahtevati, da so funkcije, ki določajo neenakosti, konveksne in funkcije, ki določajo enakosti, afine. Potem je množica $\mathcal{F}$ po Lemi 5 presek m konveksnih množic ter p hiperravnin in je zato konveksna.

Naloga 18. *(VAJE) Določi, kateri optimizacijski problemi so konveksni.*

(a)

$$\begin{aligned} \text{minimiziraj} \quad & x_1^2 + x_2^2 \\ \text{pri pogojih} \quad & \frac{x_1}{1+x_2^2} \leq 0 \\ & (x_1 + x_2)^2 = 0 \end{aligned}$$

(b)

$$\begin{aligned} \text{minimiziraj} \quad & x_1^2 + x_2^2 \\ \text{pri pogojih} \quad & x_1 \leq 0 \\ & x_1 + x_2 = 0 \end{aligned}$$

Rešitev:

(a) Ni konveksni optimizacijski problem. Funkcija $h_1(x) = (x_1 + x_2)^2$ *ni afina. Še več, funkcija* $f_1(x) = \frac{x_1}{1+x_2^2}$ *ni konveksna. Njena Hessejeva matrika je*

$$\nabla^2 f_1(x) = \begin{bmatrix} 0 & -\frac{2x_2}{(1+x_2^2)^2} \\ -\frac{2x_2}{(1+x_2^2)^2} & \frac{8x_1x_2^2}{(1+x_2^2)^3} - \frac{2x_1}{(1+x_2^2)^2} \end{bmatrix}$$

z determinanto $-\frac{4x_2^2}{(1+x_2^2)^4}$*. Torej sta lastni vrednosti različno predznačeni.*

(b) Je konveksni optimizacijski problem, saj sta $f_0(x) = x_1^2 + x_2^2$ *in* $f_1(x) = x_1$ *konveksni ter* $h_1(x) = x_1 + x_2$ *afina.*

Fundamentalna lastnost konveksnih optimizacijskih problemov je ta, da je vsak lokalni minimum tudi globalni.

Izrek 1. *Dan je konveksni optimizacijski problem* (2.2)*. Potem je vsak lokalni minimum tudi globalni.*

Dokaz. Predpostavimo, da je element $x \in \mathcal{F}$ lokalni minimum, tj. x je dopusten in obstaja tak $R > 0$, da velja

$$f_0(x) = \inf\{f_0(z) \mid z \in \mathcal{F},\ \|z - x\| \leq R\}.$$

Denimo, da x ni globalni minimum, tj. obstaja dopusten $y \in \mathcal{F}$, da velja $f_0(y) < f_0(x)$. Torej $\|y-x\| > R$, saj bi sicer veljalo $f_0(x) \leq f_0(y)$. Do protislovja bomo prišli tako, da bomo skonstruirali dopustno rešitev $z \in \mathcal{F}$ z lastnostjo $\|z - x\| \leq R$, za katero velja $f_0(z) < f_0(x)$. Vzemimo

$$z = (1-\lambda)x + \lambda y, \quad \lambda = \frac{R}{2\|y - x\|}$$

Ker je z konveksna kombinacija elementov $x, y \in \mathcal{F}$ in je $\mathcal{F}$ konveksna množica, sledi, da je z dopusten. Velja tudi

$$\|z - x\| = \lambda \|y - x\| = \frac{R}{2} < R.$$

Zaradi konveksnosti funkcije f_0 potem sledi

$$\begin{aligned} f_0(z) &= f_0((1-\lambda)x + \lambda y) \\ &\leq (1-\lambda) f_0(x) + \lambda f_0(y) \\ &< (1-\lambda) f_0(x) + \lambda f_0(x) \\ &= f_0(x). \end{aligned}$$

Torej velja $f_0(z) < f_0(x)$ in zato x ni lokalni minimum. Prišli smo do protislovja. □

Izrek 2. *Dan je konveksni optimizacijski problem* (2.2). *Denimo, da je kriterijska funkcija f_0 strogo konveksna na dopustni množici $\mathcal{F}$. Če obstaja optimalna rešitev, potem je enolična.*

Dokaz. Denimo, da obstajata dve optimalni rešitvi x in y. To pomeni, da za vsak $z \in \mathcal{F}$ velja

$$f(x) = f(y) \leq f(z).$$

Vzemimo $z = \frac{1}{2}(x+y)$. Zaradi konveksnosti množice $\mathcal{F}$ velja $z \in \mathcal{F}$. Zaradi strogo kovenksnosti funkcije f_0 pa dobimo

$$f(z) = f\left(\frac{1}{2}x + \frac{1}{2}y\right) < \frac{1}{2}f(x) + \frac{1}{2}f(y) = f(x).$$

Prišli smo do protislovja. □

Izrek 3. *Dan je konveksni optimizacijski problem* (2.2). *Denimo, da je kriterijska funkcija f_0 zvezno odvedljiva. Potem je x optimalna rešitev natanko tedaj, ko je x dopusten in velja*

$$\nabla f_0(x)^T (y - x) \geq 0 \quad \text{za vsak } y \in \mathcal{F}. \tag{2.3}$$

Dokaz. ($\Rightarrow$) Naj bo x optimalna rešitev. Torej je x dopusten. Predpostavimo, da pogoj (2.3) ne velja, tj. obstaja $y \in \mathcal{F}$, da je

$$\nabla f_0(x)^T (y - x) < 0.$$

Vzemimo $z(t) = (1-t)x + ty$, kjer je $t \in [0,1]$. Ker je $z(t)$ konveksna kombinacija točk x in y, je $z(t)$ dopusten. Ker je

$$\frac{d}{dt} f_0(z(t))\Big|_{t=0} = \nabla f_0(z(t))^T z'(t)\Big|_{t=0} = \nabla f_0(x)^T (y - x) < 0,$$

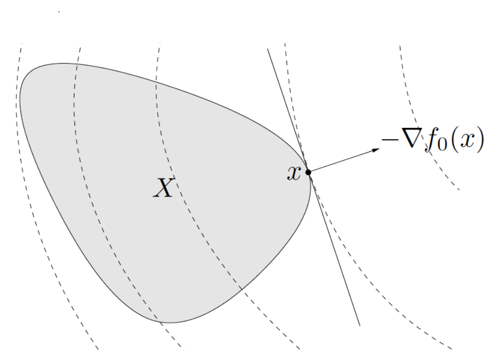
dobimo, da za dovolj majhne t velja $f_0(z(t)) < f_0(x)$. Torej x ni optimalna rešitev.

($\Leftarrow$) Ker je kriterijska funkcija f_0 konveksna in zvezno odvedljiva, velja po Trditvi 8

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T (y - x)$$

za vsak $y \in \mathcal{F}$. Ker je izpolnjen pogoj (2.3), sledi $f_0(y) \geq f_0(x)$. Torej je x globalni minimum. □

Geometrijsko lahko pogoj optimalnosti povemo takole: Če je $\nabla f_0(x) \neq 0$, potem $-\nabla f_0(x)$ določa podporno hiperravnino na množico dopustnih rešitev v točki x. *Podporna hiperravnina* za konveksno množico $\mathcal{C}$ v točki x_0 je ravnina z enačbo $a^T x = a^T x_0$, za katero velja $a^T x \leq a^T x_0$ za vsak $x \in C$. Konveksna množica ima lahko več kot eno podporno hiperravnino v neki točki.



Slika 2.1: Geometrijska interpretacija pogojev optimalnosti (2.3).

Kasneje se bomo pri teoriji dualnosti bolj podrobno posvetili pogojem, ki določajo optimalnost rešitev. Tu si oglejmo nekaj preprostih zgledov.

Naloga 19. *(VAJE) Dokaži, da je $x^* = \left(1, \frac{1}{2}, -1\right)$ optimalna rešitev optimizacijskega problema*

$$\begin{aligned} \text{minimiziraj} \quad & \frac{1}{2} x^T P x + q^T x + r \\ \text{pri pogojih} \quad & -1 \leq x_i \leq 1, \quad i = 1, 2, 3. \end{aligned}$$

Pri tem je

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \quad q = \begin{bmatrix} -22 \\ -14.5 \\ 13 \end{bmatrix}, \quad r = 1.$$

Rešitev: *Najprej preverimo, da gre za konveksni optimizacijski problem. Ker so vsi glavni minorji nenegativni ($\det(A_{II}) \geq 0$ za vsako podmatriko A_{II}, kjer je $I \subseteq \{1, \ldots, n\}$), je P res pozitivno semidefinitna matrika.*

Optimalnost rešitve bomo dokazali tako, da bomo za x^* preverili pogoj (2.3). Gradient kriterijske funkcije je v točki x^* enak

$$\nabla f_0(x^*) = P x^* + q = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}.$$

Od tod dobimo, da je pogoj optimalnosti enak

$$0 \leq \nabla f_0(x^*)^T (y - x) = -1(y_1 - 1) + 2(y_2 + 1) = -y_1 + 2y_2 + 3,$$

ki pa je očitno izpolnjen za vsak y, ki zadošča pogojem $-1 \leq y_i \leq 1$.

Posledica 2 (Pogoj optimalnosti za nevezan optimizacijski problem)**.** *Dan je nevezan optimizacijski problem*

$$\text{minimiziraj} \quad f_0(x),$$

kjer je f_0 zvezno odvedljiva konveksna funkcija. Potem je x^* optimalna rešitev natanko tedaj, ko velja

$$\nabla f_0(x) = 0.$$

Dokaz. Po Izreku 3 je x^* je optimalna rešitev natanko tedaj, ko velja

$$\nabla f_0(x^*)^T(y - x^*) \geq 0$$

za vsak dopusten y, tj. $y \in \mathbb{R}^n$. Vzemimo $y = x^* - \nabla f_0(x^*)$. Potem sledi

$$\|\nabla f_0(x^*)\|_2^2 \leq 0$$

oziroma $\nabla f_0(x^*) = 0$. □

Posledica pove, da za konveksni optimizacijski problem enačba $\nabla f_0(x) = 0$ podaja potrebni in zadostni pogoj za nastop ekstrema.

Definicija 19. *Naj bo A realna simetrična matrika reda n in ranga $k \leq n$. Označimo z $\lambda_1, \ldots, \lambda_k$ in $q_1, \ldots, q_k$ k neničelnih lastnih vrednosti in pripadajočih lastnih vektorjev matrike A. Tako lahko zapišemo*

$$A = Q\Lambda Q^T = \sum_{i=1}^{k} \lambda_i q_i q_i^T.$$

Pseudoinverz *matrike A je definiran kot*

$$A^+ = \sum_{i=1}^{k} \frac{1}{\lambda_i} q_i q_i^T.$$

Primer 18 (Nevezana kvadratična optimizacija)**.** *Dan je optimizacijski problem*

$$\text{minimiziraj} \quad \frac{1}{2}x^T P x + q^T x.$$

Da bo kriterijska funkcija konveksna, zahtevamo, da je $P \succeq 0$. Potrebni in zadostni pogoj za nastop ekstrema x je

$$\nabla f_0(x) = Px + q = 0.$$

Ločimo primere.

- *Če $q \notin \operatorname{im}(P)$, potem optimalna rešitev ne obstaja. V tem primeru je f_0 neomejena navzdol.*
- *če je $P \succ 0$ (f_0 je v tem primeru strogo konveksna), potem obstaja enolična optimalna rešitev $x^* = -P^{-1}q$.*

- *Če je P singularna (vsaj ena lastna vrednost enaka 0) in $q \in \operatorname{im} P$, potem je množica optimalnih rešitev enaka $-P^{+}q + \ker P$. Pri tem P^{+} označuje psevdoinverz matrike P.*

Naloga 20. *(VAJE) Poišči optimalno rešitev in optimalno vrednost optimizacijskega problema*

$$\text{minimiziraj} \quad x^T P x + q^T x + r,$$

kjer je

$$P = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad q = \begin{bmatrix} 2 \\ -2 \end{bmatrix} \quad \textit{in} \quad r = -1.$$

Rešitev: *Lastni vrednosti matrike P sta 0 in 2. Zato je Hessejeva matrika*

$$\nabla^2 f_0(x) = 2P \succeq 0.$$

Torej je kriterijska funkcija konveksna. Potrebni in zadostni pogoj za nastop ekstrema x^* je

$$\nabla f_0(x) = 2Px^* + q = 0.$$

Ker je P singularna in $q \in \operatorname{im} P$, je množica optimalnih rešitev enaka $x^* = -\frac{1}{2}P^{+}q + \ker P$. Matriko P lahko z uporabo lastnega razcepa zapišemo kot

$$P = 2\left(\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right)\left(\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \end{bmatrix}\right).$$

Od tod dobimo, da je pseudoinverz enak

$$P^{+} = \frac{1}{2}\left(\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right)\left(\frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \end{bmatrix}\right) = \frac{1}{4}P.$$

Sledi, da je

$$x^* = -\frac{1}{2}P^{+}q + \ker P = -\frac{1}{8}Pq + \ker P = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + \begin{bmatrix} t \\ t \end{bmatrix}.$$

Torej je množica optimalnih rešitev enaka $\{(x_1, x_2) \in \mathbb{R}^n \mid x_2 = x_1 + 1\}$. Če to vstavimo v kriterijsko funkcijo, dobimo, da je optimalna vrednost enaka -2.

Posledica 3 (Izreka 3)**.** *Dan je konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & Ax = b, \end{aligned}$$

kjer je f_0 zvezno odvedljiva konveksna funkcija in $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$. Element x^* je optimalna rešitev natanko tedaj, ko obstaja $\nu \in \mathbb{R}^p$, da velja

$$Ax^* = b \quad \textit{in} \quad \nabla f_0(x^*) + A^T \nu = 0.$$

Dokaz. Po Izreku 3 je x^* je optimalna rešitev natanko tedaj, ko je x dopusten in velja

$$\nabla f_0(x^*)^T(y - x^*) \geq 0$$

za vsak y, ki zadošča $Ay = b$. Ker sta x in y dopustna, razlika $z = y - x$ leži v jedru matrike A, tj. $Az = 0$. Torej velja $\nabla f_0(x^*)^T z \geq 0$ za vsak $z \in \ker A$. Element $-z$ tudi pripada jedru, zato velja $-\nabla f_0(x^*)^T z \geq 0$. Dobimo, da je $\nabla f_0(x^*)^T z = 0$ za vsak $z \in \ker A$. Od tod sledi, da

$$\nabla f_0(x^*) \in (\ker A)^{\perp} = \operatorname{im}\left(A^T\right).$$

Torej obstaja $\nu \in \mathbb{R}^p$, da je

$$\nabla f_0(x) = -A^T \nu.$$

□

Primer 19 (Kvadratična optimizacija z linearnimi enakostmi)**.** *Za $P \succeq 0$ je dan optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & \frac{1}{2}x^T P x + q^T x \\ \text{pri pogojih} \quad & Ax = b, \end{aligned}$$

kjer je $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, $q \in \mathbb{R}^n$. Predpostavimo, da je rank $r(A) = p < n$. Element $x^+ \in \mathbb{R}^n$ je optimalna rešitev natanko tedaj, ko obstaja $\nu^* \in \mathbb{R}^p$, da velja

$$Ax^* = b \quad \textit{in} \quad Px^* + q + A^T \nu^* = 0.$$

Torej optimalno rešitev dobimo preko reševanja sistema $n + p$ linearnih enačb

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \tag{2.4}$$

za $n + p$ neznank.

Ločimo primere.

- *Če je matrika sistema obrnljiva, potem obstaja enolična optimalna rešitev x^*.*
- *Če je matrika singularna in je sistem rešljiv, potem je vsaka rešitev optimalna.*
- *V primeru da pa sistem ni rešljiv, je kvadratični program bodisi nedopusten bodisi neomejen.*

Glede singularnosti matrike sistema (2.4) *lahko pokažemo naslednje ekvivalentne trditve:*

(a) *Matrika* $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$ *je nesingularna.*

(b) $Ax = 0, x \neq 0 \Longrightarrow x^T P x > 0.$

(c) $P + A^T A \succ 0.$

Dokaz: $(a) \Rightarrow (b)$ *Denimo, da velja* $Ax = 0$, $x \neq 0$ *in* $x^T P x = 0$. *Ker je* $P \succeq 0$, *velja* $Px = 0$. *Za vektor neničeln vektor* $[x, 0]^T$ *potem velja*

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = 0$$

Torej je matrika $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$ *singularna, kar je protislovje.*

$(b) \Rightarrow (a)$ *Denimo, da je matrika* $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$ *singularna, tj. obstaja neničeln vektor* $[x, z]^T$, *da velja*

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = 0.$$

Torej je $Px + A^T z = 0$ *in* $Ax = 0$. *Če prvo enačbo z levo pomnožimo z* x^T *in upoštevamo* $Ax = 0$, *dobimo*

$$x^T P x + x^T A^T z = x^T P x + (Ax)^T z = x^T P x = 0,$$

kar je protislovje.

$(b) \Rightarrow (c)$ *Če je* $x \neq 0$ *in* $Ax = 0$, *potem velja*

$$x^T (P + A^T A) x = x^T P x > 0.$$

Če pa velja $x \neq 0$ *in* $Ax \neq 0$, *dobimo*

$$x^T (P + A^T A) x = \underbrace{x^T P x}_{\geq 0} + \underbrace{\|Ax\|_2^2}_{>0} > 0.$$

V obeh primerih dobimo, da je matrika $P + A^T A$ *pozitivno definitna.*

$(c) \Rightarrow (b)$ *Denimo, da velja* $Ax = 0$, $x \neq 0$ *in* $x^T P x = 0$. *Potem je*

$$x^T (P + A^T A) x = 0.$$

Torej matrika $P + A^T A$ *ne more biti pozitivno definitna.*

Iz točke (c) *vidimo, da je* $P \succ 0$ *zadostni pogoj za obrnljivost matrike* $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$.

Opomba: *O pomenu spremenljivke* ν^* *bomo izvedeli v poglavju o dualnosti.*

Primer 20. *(VAJE) Poišči minimum funkcije* $f(x, y, z) = 2x^2 + y^2 + \frac{z^2}{2}$ *pri pogojih* $x + y + z = 10$ *in* $x - y = 5$.

Najprej preverimo konveksnost funkcije f. *Hessejeva matrika*

$$\nabla^2 f(x) = P = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

je očitno pozitivno semidefinitna. Optimalno rešitev dobimo preko reševanja sistema linearnih enačb

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}. \tag{2.5}$$

V našem primeru dobimo

$$\begin{bmatrix} 4 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 10 \\ 5 \end{bmatrix}.$$

Ker je $P \succ 0$, je matrika $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$ obrnljiva. Pri reševanju sistem linearnih enačb (2.5) *izkoristimo strukturo (ker je P diagonalna). Iz prve enačbe eliminiramo x^*:*

$$x^* = P^{-1}(-q - A^T\nu^*)$$

Če to vstavimo v drugo enačbo, dobimo $(q = 0)$

$$AP^{-1}A^T\nu^* = -AP^{-1}q - b = -b.$$

Optimalna rešitev je $\nu^* = (-7, -9)$ s pripadajočim $x^* = (4, -1, 7)$.

Posledica 4 (Izreka 3)**.** *Dan je konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & x \geq 0, \end{aligned}$$

kjer je f_0 zvezno odvedljiva konveksna funkcija. Element $x^* \in \mathbb{R}^n$ je optimalna rešitev natanko tedaj, ko velja

$$x^* \geq 0, \quad \nabla f_0(x^*) \geq 0 \quad \textit{in} \quad x_i^* \nabla f_0(x^*)_i = 0 \;\; \textit{za vsak} \;\; i = 1, \ldots, n.$$

Dokaz. Po Izreku 3 je x^* je optimalna rešitev natanko tedaj, ko je $x^* \geq 0$ in velja

$$\nabla f_0(x^*)^T(y - x^*) \geq 0 \tag{2.6}$$

za vsak $y \geq 0$. Funkcija $y \mapsto \nabla f_0(x^*)^T(y - x^*)$ je afina v y. Da bo za vsak $y \geq 0$ navzdol omejena, mora veljati $\nabla f_0(x^*) \geq 0$. Če sedaj vstavimo $y = 0$ v neenačbo (2.6), dobimo $-\nabla f_0(x^*)^T x^* \geq 0$. Ker sta vektorja $\nabla f_0(x^*)$ in x^* nenegativna (torej je njun skalarni produt nenegativen), sledi, da je

$$\nabla f_0(x^*)^T x^* = \sum_{i=1}^{n} x_i^* \nabla f_0(x^*)_i = 0.$$

Ker je vsota nenegativnih števil enaka nič natanko tedaj, ko je vsak člen enak nič, dobimo pogoj

$$x_i^* \nabla f_0(x^*)_i = 0 \;\; \text{za vsak} \;\; i = 1, \ldots, n. \qquad \square$$

V nadaljevanju si oglejmo dva pristopa, ki omogočata pretvorbo optimizacijskega problema v drugega, ki mu je ekvivalenten. Pravimo, da sta optimizacijska problema *ekvivalentna*, če rešitev enega problema lahko enostavno dobimo iz rešitve drugega (in obratno).

Uvedba dopolnilnih spremenljivk za linearne neenakosti. Z uvedbo dopolnilnih spremenljivk lahko pretvorimo vsako neenakost v enakost. Pri konveksnem optimizacijskem problemu zahtevamo, da so funkcije, ki določajo omejitve enakosti, afine. Zato si oglejmo to pretvorbo za linearne neenakosti. Optimizacijski problem

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & a_i^T x \leq b_i, \quad i = 1, \ldots, m \end{aligned}$$

je ekvivalenten problemu

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & a_i^T x + s_i = b_i, \quad i = 1, \ldots, m \\ & s_i \geq 0, \quad i = 1, \ldots, m. \end{aligned}$$

Epigraf oblika. Epigraf oblika konveksnega optimizacijskega problema (2.2) je

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & f_0(x) \leq t \\ & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ & Ax = b. \end{aligned} \tag{2.7}$$

Optimizacijski spremenljivki sta x in t. To še je zmeraj konveksni optimizacijski problem, saj je kriterijska funkcija linearna, funkcija $(x, t) \mapsto f_0(x) - t$, ki določa omejitev neekanosti pa konveksna (vsota konveksne in afine funkcije). Očitno je optimalna vrednost problema (2.7) dosežena pri tistem (x^*, t^*), za katerega velja $t^* = f_0(x^*)$.

2.1 Družine konveksnih optimizacijskih problemov

2.1.1 Linearno programiranje

Če je kriterijska funkcija linearna in so funkcije, ki določajo pogoje enakosti in neenakosti afine, potem optimizacijskemu problemu rečemo ***linearni program (LP)***. V standardni obliki ga zapišemo kot

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & Ax = b \\ & x \geq 0, \end{aligned}$$

kjer je $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$ in $b \in \mathbb{R}^p$. Množica dopustnih rešitev $\mathcal{F}$ je presek končno mnogo hiperravnin in polprostorov. Torej pri linearnem programiranju minimiziramo linearno funkcijo po poliedru. Do optimalne rešitve lahko pridemo grafično, če se znotraj množice $\mathcal{F}$ pomikamo v smeri vektorja $-c$ tako dolgo, da še ostanemo dopustni.

Naloga 21. *(VAJE) Z grafično metodo poišči optimalni rešitvi naslednjih linearnih programov.*

(a)

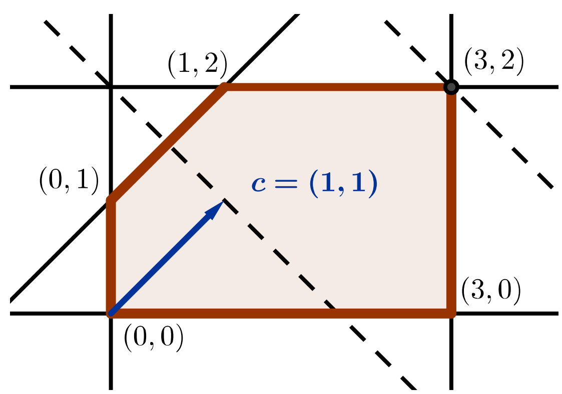
$$
\begin{aligned}
\text{maksimiziraj} \quad & x_1 + x_2 \\
\text{pri pogojih} \quad & -x_1 + x_2 \leq 1 \\
& x_1 \leq 3 \\
& x_2 \leq 2 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

(b)

$$
\begin{aligned}
\text{maksimiziraj} \quad & x_1 \\
\text{pri pogojih} \quad & x_1 - x_2 \leq 1 \\
& -x_1 + x_2 \leq 2 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

Rešitev: *Narišimo množici dopustnih rešitev.*

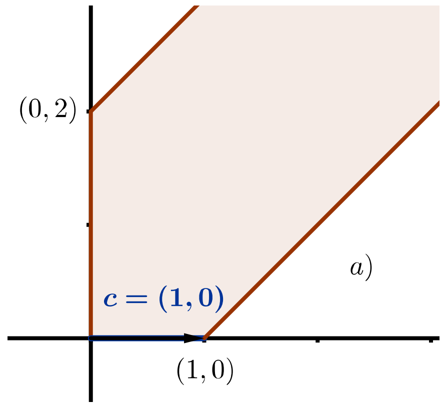
(a) Izmed vseh dopustnih rešitev x iščemo tisto, ki maksimizira kriterijsko funkcijo. Vsak vektor, ki leži na pravokotnici na vektor $(1,1)$, da z njim enak skalarni produkt. Optimalno rešitev dobimo tako, da transliramo pravokotnico v smeri vektorja $(1,1)$ tako dolgo, da še ostanemo znotraj množice dopustnih rešitev. Na sliki vidimo, da je optimalna rešitev $(3,2)$ z vrednostjo 5.



Slika 2.2: Množica dopustnih rešitev iz primera (a).

(b) Premikamo se v smeri vektorja $(1,0)$ tako dolgo, da še ostanemo dopustni. Ker lahko vrednost kriterijske funkcije poljubno povečamo, je problem neomejen.

Opomba 8. *Pomembna lastnost linearnih programov je ta, da če ima naloga optimalno rešitev, tedaj je ta vedno dosežena v (vsaj) enem oglišču poliedra.*



Slika 2.3: Množica dopustnih rešitev iz primera (b).

V primeru, da je podan splošni linearni program

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & Gx \leq h \\ & Ax = b, \end{aligned}$$

ga z uvedbo dopolnilnih spremenljivk s_i pretvorimo na standardno obliko. Dobimo

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & Gx + s = h \\ & Ax = b \\ & s \geq 0. \end{aligned}$$

V drugem koraku prosto spremenljivko x zapišemo kot razliko dveh nenegativnih spremenljivk x^+ in x^-, tj. $x = x^+ - x^-$. Od tod dobimo

$$\begin{aligned} \text{minimiziraj} \quad & c^T x^+ - c^T x^- \\ \text{pri pogojih} \quad & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & x^+, x^-, s \geq 0 \end{aligned}$$

oziroma (če pišemo $z = [x^+, x^-, s]^T$)

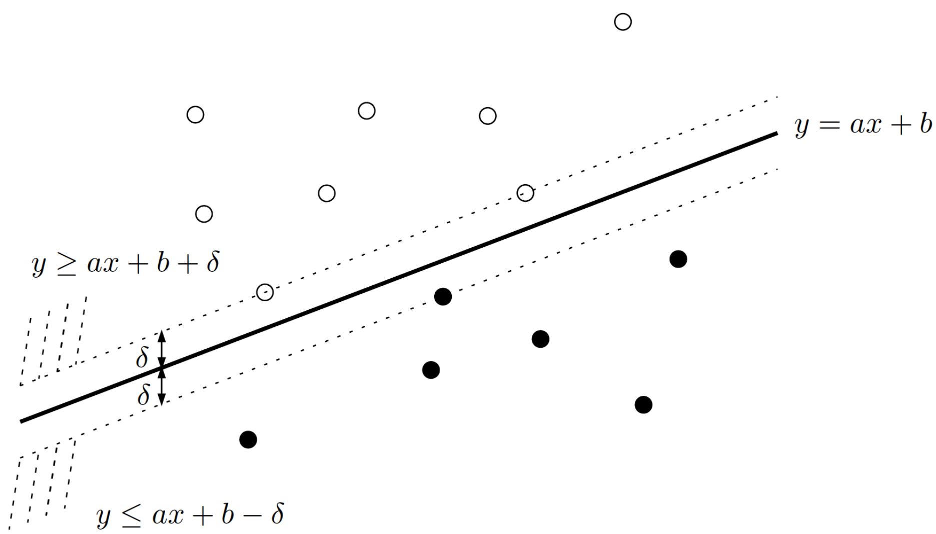
$$\begin{aligned} \text{minimiziraj} \quad & [c, -c, 0]^T z \\ \text{pri pogojih} \quad & \begin{bmatrix} G & -G & I \\ A & -A & 0 \end{bmatrix} z = \begin{bmatrix} h \\ b \end{bmatrix} \\ & z \geq 0, \end{aligned}$$

kar je standardna oblika linearnega programa.

Naloga 22 (Separacija točk). *(VAJE) V ravnini imamo podanih m belih točk $p_1, p_2, \dots, p_m$ in n črnih točk $q_1, q_2, \dots, q_n$. Zanima nas, ali obstaja premica, ki separira obe skupini točk, tj. vse bele točke ležijo strogo na eni strani in vse črne točke strogo na drugi strani. Ločimo tri primere. Najprej preverimo, ali obstaja navpična premica, ki loči točke. To je enostavna naloga, katere rešitev ne potrebuje znanja linearnega programiranja. Druga možnost je, da iščemo poševno premico oblike $y = ax + b$, tako da so vse bele točke nad njo in vse črne točke pod njo. Taka premica obstaja, če najdemo skalarja a in b, da velja*

$$\begin{aligned} y(p_i) &> ax(p_i) + b, \quad i = 1, \dots, m \\ y(q_j) &< ax(q_j) + b, \quad j = 1, \dots, n. \end{aligned} \tag{2.8}$$

Stroge neenakosti niso dovoljene pri konveksni optimizaciji. Zato uvedemo dopolnilno spremenljivko δ, ki meri vrzel oziroma razmik med levo in desno stranjo vsake neenakosti (glej Sliko 2.4). Ker želimo



Slika 2.4: Separacija belih in črnih točk s premico.

maksimizirati ta razmik, dobimo linearni program

$$\begin{aligned} \text{maksimiziraj} \quad & \delta \\ \text{pri pogojih} \quad & y(p_i) \geq ax(p_i) + b + \delta, \quad i = 1, \dots, m \\ & y(q_j) \leq ax(q_j) + b - \delta, \quad j = 1, \dots, n \end{aligned}$$

za optimizacijske spremenljivke a, b in δ. Optimalna vrednost δ je pozitivna natanko takrat, ko ima sistem (2.8) *strogih neenakosti rešitev, tj. ko lahko s poševno premico separiramo bele (nad premico) in črne točke (pod premico). Pri tretji možnosti, ko se vlogi belih in črnih točk zamenjata, postopamo enako.*

Z enakim pristopom lahko ločimo tudi dve skupini točk v $\mathbb{R}^n$. V tem primeru za separacijo uporabimo hiperravnino. Še več, da se testirati ali dve skupini točk lahko separiramo s funkcijo oblike $f(x) = a_1\varphi_1(x) + \ldots + a_k\varphi_k(x)$, kjer so $\varphi_1, \varphi_2, \dots, \varphi_k$ dane (nelinearne) funkcije in $a_1, \dots, a_k$ skalarji. Vsak tak problem lahko modeliramo z linearnim programom.

Primer 21 (Minimizacija odsekoma afine funkcije)**.** *Dan je optimizacijski problem*

$$\text{minimiziraj} \quad \max_{i=1,\dots,m} \left(a_i^T x + b_i\right).$$

Iz 3. točke Primera 15 vemo, da je kriterijska funkcija konveksna. Torej gre za konveksni optimizacijski problem. Opazimo tudi, da kriterijska funkcija ni odvedljiva v točkah, kjer se grafi funkcij $x \mapsto a_i^T x + b_i$ *sekajo. Izkaže se, da lahko problem prevedemo na linearni program. Najprej zapišimo problem v epigraf obliki*

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & \max_{i=1,\dots,m} \left(a_i^T x + b_i\right) \le t. \end{aligned}$$

Ta pa je ekvivalenten linearnemu programu

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & a_i^T x + b_i \le t, \quad i = 1, \dots, m. \end{aligned}$$

Primer 22 (Center Čebišova)**.** *(VAJE) Center Čebišova za dani polieder* $\mathcal{P} = \{x \mid a_i^T x \le b_i,\ i = 1, \dots, m\}$ *je središče* x_c *največje krogle*

$$\mathcal{B} = \{x_c + u \mid \|u\|_2 \le r\},$$

ki jo lahko včrtamo poliedru $\mathcal{P}$*.*

Velja, da $a_i^T x \le b_i$ *za vsak* $x \in \mathcal{B}$ *natanko tedaj, ko je* $a_i^T(x_c + u) \le b_i$ *za vsak* $i = 1, \dots, m$ *in za vsak* $\|u\|_2 \le r$*. Oziroma*

$$\sup\left\{a_i^T(x_c + u) \mid \|u\|_2 \le r\right\} \le b_i \quad i = 1, \dots, m. \tag{2.9}$$

Levo stran neenačbe (2.9) *je določena z rešitvijo problema*

$$\begin{aligned} \text{maksimiziraj} \quad & a_i^T u \\ \text{pri pogojih} \quad & \|u\|_2 \le r. \end{aligned}$$

Ker maksimiziramo linearno funkcijo po krogli, je optimalna rešitev očitno enaka $u^* = r\frac{a_i}{\|a_i\|_2}$*. Optimalna vrednost programa je* $r\|a_i\|_2$*.*

Če to vstavimo v (2.9) *in glede na to, da iščemo največji polmer* r*, lahko formuliramo problem centra Čebišova kot konveksni optimizacijski problem*

$$\begin{aligned} \text{maksimiziraj} \quad & r \\ \text{pri pogojih} \quad & a_i^T x_c + r\|a_i\|_2 \le b_i, \quad i = 1, \dots, m. \end{aligned}$$

Opazimo, da so omejitve linearne v spremenljivkah x_c *in* r*. Torej gre za linearni program.*

Opomba 9. *Zraven algoritmov, ki jih bomo spoznali v 6. poglavju, lahko rešujemo linearne programe tudi s pomočjo* ***simpleksne metode****.*

Naloga 23 (Celoštevilsko linearno programiranje)**.** *Dan je neusmerjen graf $G = (V, E)$ z $n = |V|$ vozlišči. Pravimo, da je množica vozlišč $I \subseteq V$* stabilna ali neodvisna, *če nobeni dve vozlišči iz množice I nista sosednji. Iščemo stabilno množico z največjo močjo. Formuliraj nalogo kot optimizacijski problem.*

Rešitev: *Vsakemu vozlišču $v \in V$ priredimo spremenljivko x_v, ki lahko zavzame vrednost 0 ali 1. Če je $x_v = 1$, vozlišče v pripada stabilni množici I, sicer pa ne. Če sta vozlišči x_u in x_v povezani, lahko kvečjemu eno pripada stabilni množici I. To lahko zapišemo kot $x_u + x_v \leq 1$. Ker iščemo največjo stabilno množico, lahko problem formuliramo kot*

$$
\begin{aligned}
\text{maksimiziraj} \quad & \sum_{v \in V} x_v \\
\text{pri pogojih} \quad & x_u + x_v \leq 1 \qquad \textit{za vsako povezavo } \{u, v\} \in E \\
& x_v \in \{0, 1\}.
\end{aligned}
$$

Vidimo, da so prisotne omejitve, ki zahtevajo, da so spremenljivke celoštevilske ali celo binarne. Gre za celoštevilski linearni program. Problemi takega tipa so v splošnem zelo težki za reševanje.

2.1.2 Kvadratično programiranje

Če minimiziramo konveksno kvadratično funkcijo in so funkcije, ki določajo pogoje enakosti in neenakosti afine, potem optimizacijskemu problemu rečemo ***kvadratični program (QP)***. V standardni obliki ga zapišemo kot

$$
\begin{aligned}
\text{minimiziraj} \quad & x^T P x + q^T x \\
\text{pri pogojih} \quad & Ax = b \\
& x \geq 0,
\end{aligned}
$$

Pri tem so $P \in \mathcal{S}_+^n$, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$ in $b \in \mathbb{R}^p$. Torej pri kvadratičnem programiranju minimiziramo kvadratično funkcijo po poliedru.

Primer 23 (Problem najmanjših kvadratov)**.** *Poseben primer kvadratičnega programiranja je problem najmanjših kvadratov. Za dano matriko $A \in \mathbb{R}^{m \times n}$ polnega ranga in vektor $b \in \mathbb{R}^m$ rešujemo problem*

$$\text{minimiziraj} \quad \|Ax - b\|_2^2.$$

Kriterijsko funkcijo lahko namreč zapišemo kot

$$f_0(x) = \|Ax - b\|_2^2 = x^T A^T A x - 2 \left(A^T b\right)^T x + b^T b.$$

Po Lemi 4 je f_0 konveksna funkcija, saj je $\nabla^2 f(x) = 2A^T A \succeq 0$ za vsak x.

Ker gre za nevezan optimizacijski problem, je po Posledici 2 x^* optimalna rešitev natanko tedaj, ko reši enačbo $\nabla f_0(x^*) = 2A^T A x - 2A^T b = 0$. Matrika AA^T je zaradi polnega ranga matrike A obrnljiva. Denimo namreč, da velja $AA^T x = 0$. Potem velja

$$x^T A A^T x = \|A^T x\|_2^2 = 0.$$

Od tod sledi, $A^T x = 0$. Ker pa je A polnega ranga, velja $x = 0$. Torej je matrika AA^T nesingularna. Tako dobimo analitično rešitev problema

$$x^* = \left(A^T A\right)^{-1} A^T b = A^+ b,$$

kjer je A^+ psevdo inverz pravokotne matrike A.

V primeru, da pa matrika A ni polnega ranga, moramo rešiti sistem linearnih enačb $A^T Ax = A^T b$. Ta sistem ima vedno rešitev. Denimo, da imamo za A naslednji singularni (SVD) razcep

$$A = U\Sigma V^T = [U_1\ U_2] \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}.$$

Potem je

$$A^T A = V\Sigma^2 V^T = [V_1\ V_2] \begin{bmatrix} \hat{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}.$$

Sistem $A^T Ax = A^T b$ je torej v bločni matrični obliki enak

$$[V_1\ V_2] \begin{bmatrix} \hat{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} x = A^T Ax = V\Sigma^2 V^T = [V_1\ V_2] \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} b.$$

Če uvedemo $y = V^T x$ in obe strani pomnožimo z V^T, dobimo

$$\begin{bmatrix} \hat{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix} y = \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} U^T b.$$

Če označimo $y = [y_1, y_2]$, potem dobimo rešitev tega sistema: $y_1 = \hat{\Sigma}^{-1} U_1^T b$, y_2 poljuben. Torej je $x = Vy = V_1 \hat{\Sigma}^{-1} U_1^T b + V_2 y_2$, kjer y_2 poljuben. V bločnem zapisu dobimo

$$x = V \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T b = A^+ b.$$

Naloga 24. *(VAJE) Za dano matriko A polnega ranga in vektor b poišči optimalno rešitev kvadratičnega programa*

$$\begin{aligned} \text{minimiziraj} \quad & \|x\|_2^2 \\ \text{pri pogojih} \quad & Ax = b. \end{aligned}$$

Rešitev: *Geometrijsko problem sprašuje po vektorju iz afine podmnožice $\{x \mid Ax = b\}$, ki je najbližje izhodišču. Po Posledici 3 je x optimalna rešitev natanko tedaj, ko obstaja tak ν, da velja*

$$2x + A^T \nu = 0 \quad \textit{in} \quad Ax = b.$$

V matrični obliki je to sistem linearnih enačb

$$\begin{bmatrix} 2I & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}.$$

Ker je $I \succ 0$ (glej Primer 19), je matrika $\begin{bmatrix} 2I & A^T \\ A & 0 \end{bmatrix}$ nesingularna. Če iz prve enačbe izrazimo x, dobimo $x = -\frac{1}{2}A^T\nu$. Z vstavljanjem v drugo, dobimo, da ν reši enačbo $-\frac{1}{2}AA^T\nu = b$. Od tod dobimo optimalno rešitev problema

$$x^* = A^T \left(AA^T\right)^{-1} b.$$

Primer 24. *Vzemimo problem najmanjših kvadratov in 2-normo nadomestimo z 1-normo ali ∞-normo:*

$$\text{minimiziraj} \quad \|Ax - b\|_p,$$

kjer je $p = 1$ ali $p = \infty$.

Za razliko od $\|Ax-b\|_2^2$ funkciji $\|Ax-b\|_1$ in $\|Ax-b\|_\infty$ nista odvedljivi (tudi, če ju kvadriramo), zato moramo postopati drugače. Izkaže se, da lahko oba problema pretvorimo v linearni program.

$(p = 1)$ *Pišimo $y_i = |a_i^T x - b_i|$. Potem je*

$$\|Ax - b\|_1 = \sum_{i=1}^{m} |a_i^T x - b_i| = \sum_{i=1}^{m} y_i.$$

Nekonveksni pogoj $y_i = |a_i^T x - b_i|$ zamenjamo z $|a_i^T x - b_i| \leq y_i$. Tako dobimo linearni program

$$\begin{aligned} \text{minimiziraj} \quad & e^T y \\ \text{pri pogojih} \quad & a_i^T x - b_i \leq y_i, \quad i = 1, \dots, m \\ & a_i^T x - b_i \geq -y_i, \quad i = 1, \dots, m. \end{aligned}$$

Kriterijska funkcija je neodvisna od x, zato pri fiksnem x res doseže minimum pri pogoju $y_i = |a_i^T x - b_i|$. Torej sta problema ekvivalentna.

$(p = \infty)$ *Zapišimo problem v epigraf obliki*

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & \max_{i=1,\dots,m} \left|a_i^T x - b_i\right| \leq t. \end{aligned}$$

Ta pa je ekvivalenten linearnemu programu

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & a_i^T x + b_i \leq t, \quad i = 1, \dots, m \\ & a_i^T x + b_i \geq -t, \quad i = 1, \dots, m. \end{aligned}$$

Metoda najmanjših kvadratov se na primer uporablja za izravnanje podatkov (aproksimacija podatkov z danim polinomom). Napram 2-normi je 1-norma manj občutljiva na osamelce in je zato velikokrat boljša izbira.

Primer 25. *Naj bodo* $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \in \mathbb{R}^2$*. Poiščimo premico* $y = \alpha x + \beta$*, ki se tem točkam čim bolje prilega v smeri* y*-osi. Iščemo torej* α, β*, da bo vsota* $\sum_i |y_i - \alpha x_i - \beta|^2$ *minimalna. To je problem najmanjših kvadratov iz Primera 23, če vamemo*

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \ b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Če niso vse točke enake med sabo, potem je matrika A *polnega ranga in velja, da je optimalna rešitev*

$$[\alpha, \ \beta] = (A^T A)^{-1} A^T b,$$

oz.

$$\alpha = \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

$$\beta = \frac{n \sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2}.$$

Naloga 25 (Projekcija na afini podprostor)**.** *(VAJE) Za dano matriko* $A \in \mathbb{R}^{m \times n}$ *polnega ranga in vektor* $b \in \mathbb{R}^m$ *naj bo* $\mathcal{P} = \{x \mid Ax = b\}$ *množica rešitev sistema linearnih enačb* $Ax = b$*. Za dan vektor* $z \notin \mathcal{P}$ *iščemo* $\Pi(z) \in \mathcal{P}$*, ki je pravokotna projekcija vektorja* z *na podprostor* P*. Formuliraj nalogo kot optimizacijski problem in pokaži, da velja*

$$\Pi(z) = z - A^T (A A^T)^{-1} (Az - b).$$

Rešitev: *Iščemo pravokotno projekcijo vektorja* z *na afini podprostor* $\mathcal{P}$*. Izmed vseh elementov* $x \in \mathcal{P}$ *iščemo tistega, ki je najbližje vektorju* z*, tj.*

$$\begin{aligned} \text{minimiziraj} \quad & \frac{1}{2}\|x - z\|_2^2 \\ \text{pri pogojih} \quad & Ax = b. \end{aligned}$$

Kriterijsko funkcijo zapišemo kot

$$f_0(x) = \frac{1}{2}\|x - z\|_2^2 = \frac{1}{2}(x^T - z^T)(x - z) = \frac{1}{2}x^T x - z^T x + \frac{1}{2} z^T z.$$

Po Posledici 3 je x *optimalna rešitev natanko tedaj, ko obstaja tak* ν*, da velja*

$$Ax = b \quad \text{in} \quad \nabla f_0(x) + A^T \nu = 0.$$

Iz česar dobimo sistem enačb $x - z + A^T \nu = 0$*,* $Ax = b$*. Če iz prve enačbe izrazimo* x *in vstavimo v drugo enačbo, dobimo* $\nu = (AA^T)^{-1}(Az - b)$*. Matrika* AA^T *je zaradi polnega ranga matrike* A *obrnljiva. Končno iz prve enačbe dobimo*

$$\Pi(z) = x = z - A^T \nu = z - A^T (AA^T)^{-1} (Az - b).$$

2.1.3 Kvadratično programiranje s kvadratičnimi omejitvami

Če linearne neenakosti v kvadratičnem programu zamenjamo s kvadratičnimi, dobimo primer ***kvadratičnega programa s kvadratičnimi omejitvami (QCQP)***. V standardni obliki ga zapišemo kot

$$\begin{aligned} \text{minimiziraj} \quad & x^T P_0 x + q_0^T x \\ \text{pri pogojih} \quad & x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\ & Ax = b. \end{aligned}$$

Pri tem so $P_i \in \mathcal{S}_+^n$, $q_i \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$ in $b \in \mathbb{R}^p$. Zaradi pozitivne semidefinitnosti matrik P_i so kriterijska funkcija in fukcije, ki določajo neenakosti, konveksne. Če velja $P_1, \ldots, P_m \succ 0$, potem je množica dopustnih rešitev presek m elipsoidov in afine podmnožice.

Iz strukture programov je očitna hierarhija

$$LP \subseteq QP \subseteq QCQP.$$

2.1.4 Semidefinitno programiranje

Semidefinitni program (SDP) je problem, pri katerem iščemo minimum linearne funkcije pri afinih pogojih in dodatni zahtevi, da so vse nastopajoče spremenljivke vzete iz pozitivno semidefinitne matrike. Naj bodo matrike $C, A_1, \ldots, A_m \in \mathcal{S}^n$ in vektor $b \in \mathbb{R}^m$. Semidefinitni program v standardni obliki je optimizacijski problem

$$\begin{aligned} \text{minimiziraj} \quad & \langle C, X \rangle \\ \text{pri pogojih} \quad & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0. \end{aligned} \tag{2.10}$$

Primer 26. *Vzemimo $n = 3$ in $m = 2$. Naj bo*

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 0 \\ 3 & 0 & 7 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 7 \\ 1 & 7 & 5 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 2 & 8 \\ 2 & 6 & 0 \\ 8 & 0 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 11 \\ 19 \end{bmatrix}.$$

Optimizacijska spremenljivka je v tem primeru 3×3 simetrična matrika

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{bmatrix}.$$

Semidefinitni program lahko potem zapišemo kot

$$\begin{aligned} \text{minimiziraj} \quad & x_{11} + 4x_{12} + 6x_{13} + 9x_{22} + 7x_{33} \\ \text{pri pogojih} \quad & x_{11} + 2x_{13} + 3x_{22} + 14x_{13} + 5x_{33} = 11 \\ & 4x_{12} + 16x_{13} + 6x_{22} + 4x_{33} = 19 \\ & \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{bmatrix} \succeq 0. \end{aligned}$$

Semidefinitne programe bomo srečevali tudi v obliki

$$\begin{aligned} \text{maksimiziraj} \quad & b^T y \\ \text{pri pogojih} \quad & C - \sum_{i=1}^{m} y_i A_i \succeq 0, \end{aligned} \tag{2.11}$$

tj. maksimiziramo linearno funkcijo po dopustni množici, ki je določena z ***linearno matrično neenakostjo*** $\sum_{i=1}^{m} y_i A_i \preceq C$. Pri tem je y prosta spremenljivka v $\mathbb{R}^m$.

Primer 27. *Za podatke iz Primera 26 je semidefinitni program oblike* (2.11) *enak*

$$\begin{aligned} \text{maksimiziraj} \quad & 11y_1 + 19y_2 \\ \text{pri pogojih} \quad & \begin{bmatrix} 1-y_1 & 2-2y_2 & 3-y_1-8y_2 \\ 2-2y_2 & 9-3y_1-6y_2 & -7y_1 \\ 3-y_1-8y_2 & -7y_1 & 7-5y_1-4y_2 \end{bmatrix} \succeq 0. \end{aligned}$$

Izkaže se, da je problem (2.11) ***dual*** prvotnega problema (glej Poglavje 3). Množica dopustnih rešitev semidefinitnega programa je konveksna. V problemu (2.10) je to presek afinega podprostora in stožca pozitivno semidefinitnih matrik, v problemu (2.11) pa je konveksna po Nalogi 11. Semidefinitni program je tako posplošitev linearnega programa, saj stožec $\mathbb{R}^n_+$ $(x \geq 0)$ nadomestimo s stožcem $\mathcal{S}^n_+$ $(X \succeq 0)$.

Naloga 26. *Zapiši linearni program*

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & Ax \leq b \end{aligned}$$

kot semidefinitni program.

Rešitev: *Omejitev* $Ax \leq b$ *je ekvivalentna pogoju, da je diagonalna matrika* $\mathrm{Diag}(b-Ax)$ *pozitivno semidefinitna. Zato lahko problem zapišemo kot semidefinitni program*

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & \mathrm{Diag}(b - Ax) \succeq 0. \end{aligned}$$

Še več, velja namreč hierarhija

$$LP \subseteq QP \subseteq QCQP \subseteq SDP.$$

Dokažimo zadnjo inkluzijo, tj. vsak QCQP problem lahko zapišemo kot semidefinitni program. To bomo pokazali za poseben primer. Denimo, da je dan konveksni kvadratični program s kvadratičnimi neenakostmi (brez omejitev enakosti)

$$\begin{aligned} \text{minimiziraj} \quad & x^T P_0 x + q_0^T x \\ \text{pri pogojih} \quad & x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m. \end{aligned}$$

Najprej zapišimo problem v epigraf obliki

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & x^T P_0 x + q_0^T x \leq t \\ & x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m. \end{aligned}$$

Ker so matrike P_i pozitivno semidefinitne, po Trditvi 1 obstajajo take matrike Q_i, $i = 0, 1, \ldots, m$, da velja $P_i = Q_i^T Q_i$. Z uporabo Schurovega komplementa lahko potem neenakosti

$$x^T Q_0^T Q_0 x + q_0^T x \leq t \quad \text{in} \quad x^T Q_i^T Q_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m$$

zapišemo kot pozitivno semidefinitni omejitve

$$\begin{bmatrix} I & Q_0 x \\ x^T Q_0^T & t - q_0^T x \end{bmatrix} \succeq 0 \quad \text{in} \quad \begin{bmatrix} I & Q_i x \\ x^T Q_i^T & -r_i - q_i^T x \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m.$$

Če to zložimo v bločno diagonalno matriko, dobimo problem

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & \begin{bmatrix} I & Q_0 x & & & & & \\ x^T Q_0^T & t - q_0^T x & & & & & \\ & & I & Q_1 x & & & \\ & & x^T Q_1^T & -r_1 - q_1^T x & & & \\ & & & & \ddots & & \\ & & & & & I & Q_1 x \\ & & & & & x^T Q_1^T & -r_1 - q_1^T x \end{bmatrix} \succeq 0. \end{aligned}$$

Opazimo, da so koeficienti neznane matrike afine funkcije spremenljivk $x \in \mathbb{R}^n$ in $t \in \mathbb{R}$. Torej gre za semidefinitni program oblike (2.11).

Naloga 27 (minimizacija največje lastne vrednosti matrike). *(VAJE) Za dane simetrične matrike $A_i \in \mathcal{S}^n$, $i = 1, \ldots, m$, je podan optimizacijski problem*

$$\text{minimiziraj} \quad \lambda_{\max}\left(\mathcal{A}(x)\right),$$

kjer je preslikava $\mathcal{A}\colon \mathbb{R}^m \to \mathcal{S}^n$ definirana s predpisom $A(x) = A_0 + x_1 A_1 + \ldots + x_m A_m$. Dokaži, da je problem konveksen in ga prevedi na semidefinitni program.

Rešitev: *Gre za nevezan optimizacijski problem. Dokažimo konveksnost kriterijske funkcije. Opazimo, da je preslikava $\mathcal{A}$ afina. Po Primeru 15 je funkcija $f(X) = \lambda_{\max}(X)$, ki dani simetrični matriki priredi njeno največjo lastno vrednost, konveksna. Zato je po 2. točki Trditve 11 kompozitum $f \circ \mathcal{A}$ konveksna funkcija.*

Zapišimo problem v epigraf obliki

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & \lambda_{\max}\left(\mathcal{A}(x)\right) \leq t. \end{aligned}$$

Omejitev $\lambda_{\max}(\mathcal{A}(x)) \leq t$ je ekvivalentna omejitvam $\lambda_i(\mathcal{A}(x)) \leq t$ za vsak $i = 1, \ldots, n$ oziroma $\lambda_i(tI - \mathcal{A}(x)) \geq 0$ za vsak $i = 1, \ldots, n$. Po definiciji to pomeni, da je matrika $tI - \mathcal{A}(x)$ pozitivno semidefinitna. Torej je prvotni problem ekvivalenten semidefinitnemu programu

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & tI - \mathcal{A}(x) \succeq 0 \end{aligned}$$

v spremenljivkah $x \in \mathbb{R}^m$ in $t \in \mathbb{R}$.

Naloga 28 (minimizacija največje singularne vrednosti matrike). *(VAJE) Za dane $A_i \in \mathbb{R}^{p \times q}$, $i = 1, \ldots, m$, je podan optimizacijski problem*

$$\text{minimiziraj} \quad \|\mathcal{A}(x)\|_2 = \sigma_{\max}(\mathcal{A}(x)) = \sqrt{\lambda_{\max}\left(\mathcal{A}(x)^T \mathcal{A}(x)\right)},$$

kjer je preslikava $\mathcal{A} \colon \mathbb{R}^m \to \mathbb{R}^{p \times q}$ definirana s predpisom $A(x) = A_0 + x_1 A_1 + \ldots + x_m A_m$. Dokaži, da je problem konveksen in ga prevedi na semidefinitni program.

Rešitev: *Ker je kriterijska funkcija kompozitum norme in afine preslikave, je po 2. točki Trditve 11 konveksna funkcija.*

Zapišimo problem v epigraf obliki

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & \|\mathcal{A}(x)\|_2 \leq t. \end{aligned}$$

Omejitev $\|\mathcal{A}(x)\|_2 \leq t$ je po kvadriranju enaka pogoju $\lambda_{\max}\big(\mathcal{A}(x)^T \mathcal{A}(x)\big) \leq t^2$ oziroma

$$\mathcal{A}(x)^T \mathcal{A}(x) \preceq t^2 I, \quad t \geq 0.$$

To pa s pomočjo Schurovega komplementa lahko zapišemo kot $\begin{bmatrix} tI & \mathcal{A}(x) \\ \mathcal{A}(x)^T & tI \end{bmatrix} \succeq 0$. Torej je prvotni problem ekvivalenten semidefinitnemu programu

$$\begin{aligned} \text{minimiziraj} \quad & t \\ \text{pri pogojih} \quad & \begin{bmatrix} tI & \mathcal{A}(x) \\ \mathcal{A}(x)^T & tI \end{bmatrix} \succeq 0 \end{aligned}$$

v spremenljivkah $x \in \mathbb{R}^m$ in $t \in \mathbb{R}$.

Reševanje konveksnih optimizacijskih problemov. Konveksne optimizacijske probleme se v splošnem rešuje z ***metodami notranjih točk***, ki jih bomo spoznali v 6. poglavju.

Definicija 20. *Polinom $p \in \mathbb{R}[\bar{x}]$ je mogoče zapisati kot* vsoto kvadratov*, če obstajajo polinomi $q_1, \ldots, q_k$, da velja $p = \sum_i q_i^2$.*

Izrek 4. *Polynom $p \in \mathbb{R}[\bar{x}]$ v n spremenljivkah in stopnje $2d$ je vsota kvadratov natanko tedaj, ko obstaja pozitivno semidefinitna matrika Q, da velja*

$$p(\bar{x}) = X^T Q X,$$

kjer je X stolpec, ki vsebuje vse polinome stopnje $\leq d = \deg(p)/2$, tj.

$$X = [1, x_1, \ldots, x_n, x_1 x_2, \ldots, x_n^d]^T.$$

Dokaz. ($\Rightarrow$) Denimo, da je polinom p vsota kvadratov, tj. obstajajo taki polinomi $q_1, \ldots, q_k$, da velja

$$p(x) = \sum_{i=1}^{k} q_i^2(x).$$

Vsak polinom q_i lahko razvijemo po bazi $q_i(x) = q_i^T X$. Od tod dobimo

$$p(x) = \sum_{i=1}^{k} q_i^2(x) = \sum_{i=1}^{k} (a_i^T X)^2 = \sum_{i=1}^{k} X^T a_i a_i^T X = X^T \left(\sum_{i=1}^{k} a_i a_i^T \right) X.$$

Pri tem je $Q = \sum\limits_{i=1}^{k} a_i a_i^T$ pozitivno semidefinitna matrika.

($\Leftarrow$) Naj obstaja $Q \succeq 0$, da velja $p(\bar{x}) = X^T Q X$. Če izračunamo razcep Choleskega za matriko $Q = V^T V$ in z v_i^T označimo i-to vrstico matrike V, lahko zapišemo

$$p(\bar{x}) = X^T Q X = X^T V^T V X = (VX)^T (VX) = \|VX\|_2^2.$$

Tore je p vsota kvadratov polinomov $q_i(x) = v_i^T X$. □

Primer 28. *Obravnavajmo $p(x_1, x_2) = 2x_1^4 + 2x_1^3 x_2 - x_1^2 x_2^2 + 5x_2^4$. Ker je p 4-forma, velja, da je p vsota kvadratov natanko tedaj, ko je nenegativna. Velja tudi, da lahko v vsoti kvadratov nastopajo le monomi $x_1^2, x_2^2, x_1 x_2$.*

Polynom p ima torej razcep na vsoto kavdratov natanko tedaj, ko obstaja pozitivno semidefinitna matrika Q, da velja $p(x_1, x_2) = X^T Q X$, za $X = (x_1^2, x_2^2, x_1 x_2)^T$. Matrika Q mora zadoščati linearnim omejitvam. Ker imamo v p monom x_1^4 s koefiicentom 2, mora veljati $q_{1,1} = 2$ in podobno $q_{2,2} = 5$. Monom $x_1^3 x_2$ ima koeficient 2 in je lahko dobljen kot $x_1^2 \cdot x_1 x_2$, torej mora veljati $q_{1,3} + q_{3,1} = 2$ ozoiroma $q_{1,3} = q_{3,1} = 1$. Monom $x_1^2 x_2^2$ lahko dobimo kot produkt x_1^2 in x_2^2 ali kot kvadrat od $x_1 x_2$, torej $q_{1,2} + q_{2,1} + q_{3,3} = -1$. Monom $x_1 x_2^3$ ne nastopa v p, torej $q_{2,3} + q_{3,2} = 0$. Ker je Q simetrična: $q_{2,3} = q_{3,2} = 0$. Te enačbe lahko rešimo na pamet. Če postavimo $q_{3,3} = 5$, dobimo $q_{1,2} = q_{2,1} = -3$. Matrika Q je s tem v celoti določena:

$$Q = \begin{bmatrix} 2 & -3 & 1 \\ -3 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \end{bmatrix}^T \cdot \begin{bmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

Torej imamo naslednji razcep za p na vsote kvadratov:

$$p(x_1, x_2) = \frac{1}{2} \left((2x_1^2 - 3x_2^2 + x_1 x_2)^2 + (x_2^2 + 3x_1 x_2)^2 \right).$$

Poglavje 3

Teorija dualnosti

3.1 Dualni problem

Dan je splošni (ne nujno konveksni) optimizacijski problem v standardni obliki

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_j(x) = 0, \quad j = 1, \dots, p, \end{aligned} \tag{3.1}$$

kjer je $x \in \mathbb{R}^n$ optimizacijska spremenljivka. Predpostavimo, da je domena $\mathcal{D} = \bigcap_{i=0}^{m} \mathcal{D}_{f_i} \cap \bigcap_{j=1}^{p} \mathcal{D}_{h_j}$ neprazna in naj p^* označuje optimalno vrednost problema. Osnovna ideja Lagrangeove dualnosti je, da vzamemo funkcije, ki določajo omejitve problema (3.1) in jih z uteženo kombinacijo prenesemo v kriterijsko funkcijo.

Definicija 21. **Lagrangeova funkcija** $L\colon \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ *optimizacijska problema* (3.1) *je definirana s predpisom*

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x).$$

Števila λ_i, $i = 1, \dots, m$ in ν_j, $j = 1, \dots, p$ se imenujejo Lagrangeovi multiplikatorji (množitelji), ki pripadajo omejitvam neenakosti in enakosti. Vektorjema λ in ν pravimo dualni spremenljivki.

Definicija 22. **Dualna funkcija** $g\colon \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, *pripadajoča optimizacijskemu problemu* (3.1), *je definirana s predpisom*

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x) \right).$$

Ko je Lagrangeova funkcija navzdol neomejena po x, je vrednost dualne funkcije $-\infty$. Zato vektor (λ, ν) pripada $\mathcal{D}_g$, če $g(\lambda, \nu) > -\infty$. Četudi prvotni problem (3.1) ni konveksen, je dualna funkcija vedno konkavna.

Lema 6. *Dualna funkcija je konkavna funkcija.*

Dokaz. Opazimo, da je za vsak x Lagrangeova funkcija afina v spremenljivkah λ in ν. Ker je dualna funkcija g infimum družine afinih funkcij (torej konkavnih) po točkah, sledi po točki 5 trditve 11, da je konkavna funkcija. □

Posledica 5. *Definicijsko območje $\mathcal{D}_g$ je konveksna množica.*

Dokaz. Naj bosta (λ_1, ν_1) in (λ_2, ν_2) iz $\mathcal{D}_g$ ter $\mu \in [0, 1]$. Potem je zaradi konkavnosti g

$$g((1-\mu)(\lambda_1, \nu_1) + \mu(\lambda_2, \nu_2)) \geq (1-\mu)g(\lambda_1, \nu_1) + \mu g(\lambda_2, \nu_2) > -\infty.$$ □

Vrednost dualne funkcije nam da spodnjo mejo za optimalno vrednost p^* problema (3.1).

Trditev 12. *Za vsak $\nu \in \mathbb{R}^p$ in $\lambda \in \mathbb{R}^m_+$ velja*

$$g(\lambda, \nu) \leq p^*.$$

Dokaz. Naj bo $\overline{x}$ dopustna rešitev problema (3.1) in $\lambda \geq 0$. Potem velja

$$\sum_{i=1}^{m} \lambda_i f_i(\overline{x}) + \sum_{j=1}^{p} \nu_j h_j(\overline{x}) \leq 0,$$

saj je vsak člen v prvi vsoti nepozitiven in vsak člen v drugi vsoti enak 0. Od tod sledi

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\overline{x}, \lambda, \nu) = f_0(\overline{x}) + \sum_{i=1}^{m} \lambda_i f_i(\overline{x}) + \sum_{j=1}^{p} \nu_j h_j(\overline{x}) \leq f_0(\overline{x}).$$

Ker velja $g(\lambda, \nu) \leq f_0(\overline{x})$ za poljubno dopustno rešitev $\overline{x}$, dobimo želeno neenakost. □

Zgornja trditev pove, da dualna funkcija podaja parametrizirano spodnjo mejo za p^* v odvisnosti od $\lambda \geq 0$ in poljubnega ν. Naravno vprašanje je, kako poiskati čim boljšo spodnjo mejo za prvotni problem (3.1).

Definicija 23. Dualni problem *prirejen problemu* (3.1) *je optimizacijski problem oblike*

$$\begin{aligned} \text{maksimiziraj} \quad & g(\lambda, \nu) \\ \text{pri pogojih} \quad & \lambda \geq 0. \end{aligned} \tag{3.2}$$

Prvotnemu problemu (3.1) *rečemo* **primarni problem***. Dualni spremenljivki λ in ν sta* **dualno dopustni***, če $\lambda \geq 0$ in $(\lambda, \nu) \in \mathcal{D}_g$ $(g(\lambda, \nu) > -\infty)$. Optimalno vrednost dualnega problema označimo z d^*.*

Velikokrat se zapis dualnega problema poenostavi, če implicitne omejitve $(\lambda, \nu) \in \mathcal{D}_g$ zapišemo eksplicitno. Četudi je primarni problem nekonveksen, je dualni problem vedno konveksni, ker maksimiziramo konkavno funkcijo po konveksni podmnožici. Razliki $f_0(x) - g(\lambda, \nu)$, ki je po Trditvi 12 za primarno dopusten x in dualno dopusten (λ, ν) vedno nenegativna, pravimo ***dualni razmik***. Če torej najdemo primarno dopusten x in dualno dopusten (λ, ν), za katera velja $f_0(x) = g(\lambda, \nu)$, imamo certifikat za optimalnost teh rešitev.

Primer 29. *(VAJE) Linearni program v standardni obliki je problem*

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & Ax = b \\ & x \geq 0, \end{aligned}$$

kjer je $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$ in $b \in \mathbb{R}^p$. Pri tem so $f_i(x) = -x_i$, $i = 1, \ldots, n$. Sestavimo Lagrangeovo funkcijo

$$L(x, \lambda, \nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b) = c^T x - \lambda^T x + \nu^T (Ax - b) = -b^T \nu + (c - \lambda + A^T \nu)^T x. \quad (3.3)$$

Ker je L afina v x, dobimo dualno funkcijo

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu, & \text{če } \ c - \lambda + A^T \nu = 0 \\ -\infty, & \text{sicer} \end{cases},$$

saj je linearna funkcija (3.3) v spremenljivki x navzdol omejena le, če je vektor koeficientov pred x identično enak 0. Od tod dobimo dualni problem

$$\begin{aligned} \text{maksimiziraj} \quad & -b^T \nu \\ \text{pri pogojih} \quad & c - \lambda + A^T \nu = 0 \\ & \lambda \geq 0 \end{aligned}$$

oziroma

$$\begin{aligned} \text{maksimiziraj} \quad & -b^T \nu \\ \text{pri pogojih} \quad & c + A^T \nu \geq 0. \end{aligned}$$

Primer 30. *(VAJE) Za $A \in \mathbb{R}^{p \times n}$ je dan optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & \|x\|^2 \\ \text{pri pogojih} \quad & Ax = b. \end{aligned}$$

Problem je brez omejitev neenakosti. Lagrangeeva funkcija je

$$L(x, \nu) = x^T x + \nu^T (Ax - b).$$

Ker je L konveksna kvadratična funkcija v x, poiščemo x, ki minimizira L, iz pogoja optimalnosti

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0.$$

Sledi, da je $x = -\frac{1}{2} A^T \nu$. Od tod dobimo dualno funkcijo

$$g(\nu) = L\left(-\frac{1}{2} A^T \nu, \nu\right) = -\frac{1}{4} \nu^T A A^T \nu - b^T \nu,$$

ki je očitno konkavna kvadratična funkcija z domeno $\mathbb{R}^p$. Dobimo, da je dualni problem

$$\text{maksimiziraj} \quad -\frac{1}{4} \nu^T A A^T \nu - b^T \nu.$$

3.2 Šibka in krepka dualnost

Optimalna vrednost d^* dualnega problema je po definiciji najboljša spodnja meja za p^*. Od tod dobimo pomembno neenakost.

Posledica 6 (Šibka dualnost)**.** *Naj bo p^* optimalna vrednost primarnega problema in d^* optimalna vrednost dualnega problema. Potem velja*

$$d^* \leq p^*.$$

Neenakost velja tudi v primeru, ko primarni problem ni konveksni in ko vrednosti p^* in d^* nista končni. Na primer, če je primarni problem navzdol neomejen ($p^* = -\infty$), potem je $d^* = -\infty$, kar pomeni, da je dualni problem nedopusten. In obratno, če je dualni problem navzgor neomejen ($d^* = \infty$), je primarni problem nedopusten.

Razlika $p^* - d^*$, ki je vedno nenegativna, se imenuje ***optimalni dualni razmik***. Če velja enakost

$$p^* = d^*,$$

tj. optimalni dualni razmik je enak nič, potem pravimo, da velja ***krepka dualnost***. Krepka dualnost ne velja za splošne optimizacijske probleme. Če pa je problem (3.1) konveksni in so dodatno izponjene zahteve po regularnosti, potem pa krepka dualnost velja. Eden izmed takih pogojev regularnosti je Slaterjev pogoj.

Definicija 24. *Dan je konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{aligned} \tag{3.4}$$

pri čemer so $f_0, \ldots, f_m$ konveksne funkcije. Pravimo, da velja **Slaterjev pogoj***, če je problem* (3.4) strogo dopusten*, tj. če obstaja notranja točka množice dopustnih rešitev:*

$$\exists x \in \operatorname{rel}\operatorname{int}(\mathcal{D}) : \quad f_i(x) < 0, \; i = 1, \ldots, m, \quad Ax = b. \tag{3.5}$$

Izrek 5. *[o krepki dualnosti] Za konveksni optimizacijski problem, ki izpolnjuje Slaterjev pogoj* (3.5)*, velja krepka dualnost. Dodatno, če velja $p^* > -\infty$ (primarni problem ni neomejen oz. dualni problem ni nedopusten), potem je optimalna vrednost dualnega problema dosežena, tj. obstaja dualno dopusten (λ, ν), da je $p^* = d^* = g(\lambda, \nu)$.*

Dokaz. TODO □

Posledica 7. *Če primarni in dualni problem izpolnjujeta Slaterjev pogoj, potem velja krepka dualnost in vrednost $p^* = d^*$ je dosežena za oba problema.*

Opomba 10. *Pri Slaterjevem pogoju* (3.5) *je dovolj zahtevati strogo neenakost za vse funkcije f_i, ki niso afine. Za vse afine funkcije je dovolj preveriti dopustnost notranje točke.*

Primer 31. *(VAJE) Vzemimo optimizacijski problem iz Primera 30. Ker so omejitve afine, je Slaterjev pogoj izpolnjen, če je program dopusten. Pokazali smo, da velja*

$$x = -\frac{1}{2}A^T\nu, \tag{3.6}$$

pri čemer je ν optimalna rešitev dualnega problema

$$\text{maksimiziraj} \quad -\frac{1}{4}\nu^T AA^T\nu - b^T\nu.$$

Ker je kriterijska funkcija konkavna in ni dodatnih omejitev (gre za problema iz nevezane optimizacije), dobimo optimalno rešitev iz pogoja

$$\nabla g(\nu) = -\frac{1}{2}AA^T\nu - b = 0.$$

Od tod sledi $\nu^* = -2(AA^T)^{-1}b$. Če to uporabimo v enačbi 3.6, dobimo optimalno rešitev

$$x^* = A^T(AA^T)^{-1}b.$$

Optimalni vrednosti primarnega in dualnega problema sta enaki $p^* = d^* = b^T(AA^T)^{-1}b$. Torej velja krepka dualnost.

Primer 32. *Obravnavajmo linearni program*

$$\begin{aligned} \text{minimiziraj} \quad & x_1 + 2x_2 \\ \text{pri pogojih} \quad & x_1 + x_2 = 1, \\ & 2x_1 + 2x_2 = 3 \end{aligned} \tag{3.7}$$

Njegov dual je

$$\begin{aligned} \text{maksimiziraj} \quad & -\lambda_1 - 3\lambda_2 \\ \text{pri pogojih} \quad & \lambda_1 + 2\lambda_2 = -1, \\ & \lambda_1 + 2\lambda_2 = -2 \end{aligned} \tag{3.8}$$

Očitno sta oba linearna programa nedopustna, torej je $p^* = \infty, q^* = -\infty$.

Primer 33. *(VAJE) Dan je linearni program*

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & Ax \leq b, \end{aligned} \tag{3.9}$$

kjer je $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ in $b \in \mathbb{R}^m$. Dualna funkcija je

$$g(\lambda) = \inf_x \left(\left(c + A^T\lambda\right)^T x - b^T\lambda\right) = \begin{cases} -b^T\lambda, & \text{če} \;\; c + A^T\lambda = 0 \\ -\infty, & \text{sicer} \end{cases}.$$

Iz česar sledi, da je dualni problem

$$\begin{aligned} \text{maksimiziraj} \quad & -b^T\lambda \\ \text{pri pogojih} \quad & A^T\lambda + c = 0 \\ & \lambda \geq 0. \end{aligned} \tag{3.10}$$

Ker so vse funkcije afine, je Slaterjev pogoj izpolnjen, če obstaja primarno dopusten x za (3.9) *ali dualno dopusten λ za* (3.10). *V obeh primerih velja krepka dualnost. Torej za linearne programe vedno velja $p^* = d^*$, razen, ko sta oba problema nedopustna ($p^* = \infty$, $d^* = -\infty$).*

Naloga 29. *(VAJE) Dan je linearni program*

$$\begin{aligned} \text{minimiziraj} \quad & x \\ \text{pri pogojih} \quad & \begin{bmatrix} 0 \\ 1 \end{bmatrix} x \leq \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \end{aligned}$$

Zapiši dualni linearni program in pokaži, da velja $p^* = \infty$ in $d^* = -\infty$.

Dualni problem je po (3.10) *enak*

$$\begin{aligned} \text{maksimiziraj} \quad & -\lambda_1 + \lambda_2 \\ \text{pri pogojih} \quad & \lambda_2 = -1 \\ & \lambda_1, \lambda_2 \geq 0. \end{aligned}$$

Očitno sta oba problema nedopustna, zato velja $p^* = \infty$ in $d^* = -\infty$.

Primer 34. *(VAJE) Dan je konveksni kvadratični program*

$$\begin{aligned} \text{minimiziraj} \quad & x^T P x \\ \text{pri pogojih} \quad & Ax \leq b, \end{aligned}$$

kjer je $P \in \mathcal{S}^n_{++}$, $A \in \mathbb{R}^{m\times n}$ in $b \in \mathbb{R}^m$. Dualna funkcija je

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b) \right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T\lambda.$$

Pri tem smo poiskali x, ki minimizira L, iz pogoja

$$\nabla_x L(x, \lambda) = 2Px + A^T\lambda = 0$$

oziroma $x = -\frac{1}{2}P^{-1}A^T\lambda$. Od tod sledi, da je dualni problem

$$\begin{aligned} \text{maksimiziraj} \quad & -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T\lambda \\ \text{pri pogojih} \quad & \lambda \geq 0. \end{aligned}$$

Ker je dualni problem vedno dopusten, krepka dualnost v tem primeru vedno velja.

Naslednji protiprimer potrjuje, da zgolj konveksnost ni dovolj, da bi veljala krepka dualnost.

Primer 35. *(VAJE) Dan je konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & e^{-x} \\ \text{pri pogojih} \quad & \frac{x^2}{y} \leq 0 \end{aligned}$$

za optimizacijski spremenljivki x in $y > 0$. Zaradi omejitve $\frac{x^2}{y} \leq 0$ sledi, da je $x = 0$. Tako dobimo $p^* = 1$. Sestavimo Lagrangeovo funkcijo

$$L(x, \lambda) = e^{-x} + \lambda \frac{x^2}{y}.$$

Dualna funkcija je tako definirana s predpisom

$$g(\lambda) = \inf_{x,y>0} \left(e^{-x} + \lambda \frac{x^2}{y} \right) = \begin{cases} 0, & \text{če } \lambda \geq 0 \\ -\infty, & \text{sicer.} \end{cases}$$

Dualni problem potem lahko zapišemo kot

$$\begin{aligned} \text{maksimiziraj} \quad & 0 \\ \text{pri pogojih} \quad & \lambda \geq 0. \end{aligned}$$

Optimalna vrednost dualnega problema je očitno enaka $d^* = 0$ in optimalni dualni razmik je $p^* - d^* = 1$, kar pomeni, da krepka dualnost v tem primeru ne velja. Množica dopustnih rešitev ima namreč prazno notranjost, saj je $x = 0$ za vsako dopustno rešitev (x, y). Torej Slaterjev pogoj v tem primeru ni izpolnjen.

Z naslednjim primerom bomo pokazali, da krepka dualnost ne velja za splošne optimizacijske probleme.

Primer 36. *Za dano simetrično matriko $W \in \mathcal{S}^n$ rešujemo optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & x^T W x \\ \text{pri pogojih} \quad & x_i^2 = 1, \quad i = 1, \ldots, n. \end{aligned} \tag{3.11}$$

Gre za nekonveksni problem, saj funkcije, ki določajo omejitve enakosti, niso afine. Množica dopustnih rešitev vsebuje 2^n diskretnih vektorjev ($x_i = \pm 1$). Torej število dopustnih rešitev raste eksponentno z velikostjo problema n. Gre za NP-težki optimizacijski problem, za katerega ni znan učinkovit algoritem, ki bi rešil problem v polinomskem času.

Sestavimo Lagrangeovo funkcijo

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T \left(W + \mathrm{Diag}(\nu) \right) x - e^T \nu.$$

Pri tem e *označuje vektor samih enic in* $\operatorname{Diag}(\nu)$ *diagonalno matriko z vektorjem* ν *na diagonali. Dualna funkcija je potem*

$$g(\nu) = \inf_x \left(x^T \left(W + \operatorname{Diag}(\nu) \right) x - e^T \nu \right) = \begin{cases} -e^T \nu, & \text{če } \; W + \operatorname{Diag}(\nu) \succeq 0 \\ -\infty, & \text{sicer} \end{cases}.$$

Od tod dobimo, da dualni problem lahko zapišemo kot semidefinitni program

$$\begin{aligned} \text{maksimiziraj} \quad & -e^T \nu \\ \text{pri pogojih} \quad & W + \operatorname{Diag}(\nu) \succeq 0. \end{aligned}$$

Če na primer vzamemo dualno dopusten $\nu = -\lambda_{\min}(W) e$, *dobimo spodnjo mejo* $n\lambda_{\min}(W) \le p^*$. *Krepka dualnost ne more veljati, saj bi sicer imeli metodo, ki rešuje problem* (3.11).

Z naslednjim zgledom pa bomo pokazali, da krepka dulanost lahko velja, četudi primarni problem ni konveksni.

Primer 37. *(VAJE) Za dano simetrično matriko* $C \in \mathcal{S}^n$ *rešujemo optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & x^T C x \\ \text{pri pogojih} \quad & \|x\|^2 = 1. \end{aligned} \tag{3.12}$$

Očitno gre za nekonveksni optimizacijski problem, saj v splošnem matrika C *ni pozitivno semidefinitna in omejitev enakosti ni afina. Kljub temu minimum obstaja, saj minimiziramo zvezno funkcijo na kompaktni množici. Po 4. točki Primera 15 že vemo, da je optimalna vrednost problema enaka* $\lambda_{\min}(C)$.

Poiščimo sedaj dual problema (3.12). *Dualna funkcija je*

$$g(\nu) = \inf_x \left(x^T C x + \nu(1 - \|x\|^2) \right) = \inf_x \left(x^T (C - \nu I) x + \nu \right) = \begin{cases} \nu, & \text{če } \; C - \nu I \succeq 0 \\ -\infty, & \text{sicer} \end{cases}.$$

Dualni problem lahko zapišemo kot semidefinitni program

$$\begin{aligned} \text{maksimiziraj} \quad & \nu \\ \text{pri pogojih} \quad & C - \nu I \succeq 0. \end{aligned}$$

Ker je matrika $C - \nu I$ *pozitivno semidefinitna, so vse njene lastne vrednosti* $\lambda_i(C) - \nu$ *nenegativne. Od tod dobimo, da velja neenakost* $\lambda_i(C) \ge \nu$ *za vsak* i. *Največji* ν, *ki izpolnjuje tak pogoj, pa je ravno* $\lambda_{\min}(C)$. *Torej je* $d^* = \lambda_{\min}(C)$ *in krepka dualnost velja.*

3.3 Pogoji optimalnosti

Vzemimo sedaj splošni optimizacijski problem (3.1) in predpostavimo, da velja krepka dualnost, ter da sta x^* in (λ^*, ν^*) primarno-dualni par optimalnih rešitev. Tedaj velja

$$\begin{aligned} f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{j=1}^p \nu_j^* h_j(x) \right) \\ &\overset{(1)}{\le} f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{j=1}^p \nu_j^* h_j(x^*) \overset{(2)}{\le} f_0(x^*). \end{aligned}$$

Od tod sledi, da obe neenakosti morata biti izpolnjeni kot enakosti. Iz prve enakosti (1) dobimo, da

$$x^* \text{ minimizira funkcijo } L(x, \lambda^*, \nu^*) = f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{j=1}^{p} \nu_j^* h_j(x).$$

To pove, kako izračunati optimalno rešitev primarnega problema, če imamo dualno optimalno rešitev (λ^*, ν^*).

Iz druge enakosti (2) pa dobimo $\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$. Ker je vsota nepozitivnih števil enaka nič natanko takrat, ko je vsak člen enak nič, dobimo pogoj ***komplementarnosti***

$$\lambda_i^* f_i(x^*) = 0 \text{ za vsak } i = 1, \ldots, m.$$

Če je torej $\lambda_i^* > 0$, potem je $f_i(x^*) = 0$, kar pomeni, da je v optimalni rešitvi i-ta omejitev neenakosti izpolnjena kot enakost. In obratno, če velja $f_i(x^*) < 0$ (pravimo, da neenakost $f_i(x^*) \leq 0$ ni aktivna), potem mora biti $\lambda_i^* = 0$. Ničelni Lagrangeov multiplikator λ_i^* tako nakazuje, da i-ta nenenakost najbrž ni aktivna v optimalni rešitvi x^*.

V nadaljevanju bomo predpostavili, da so funkcije $f_0, \ldots, f_m, h_1, \ldots, h_p$ zvezno odvedljive. Ker x^* minimizira funkcijo $L(x, \lambda^*, \nu^*)$, mora veljati

$$\nabla_x L(x, \lambda^*, \nu^*) = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(x) = 0.$$

Tako smo dobili, da za primarno-dualni par optimalnih rešitev x^* in (λ^*, ν^*) veljajo

Karush-Kuhn-Tuckerjevi (KKT) pogoji:

- ***primarna dopustnost***: $f_i(x^*) \leq 0,\ i = 1, \ldots, m$ in $h_j(x^*) = 0,\ j = 1, \ldots, p$
- ***dualna dopustnost***: $\lambda^* \geq 0$
- ***komplementarnost***: $\lambda_i^* f_i(x^*) = 0,\ i = 1, \ldots, m$
- ***stacionarnost***: $\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(x) = 0$

Povzemimo rezultat v trditvi.

Trditev 13. *Naj bo dan splošni optimizacijski problem* (3.1)*, pri katerem so f_i, h_j zvezno odvedljive in za katerega velja velja krepka dualnost. Potem par primarno-dualnih optimalnih rešitev x^* in (λ^*, ν^*) zadošča KKT pogojem.*

Trditev pove, da so za splošni optimizacijski problem KKT pogoji potrebni za optimalnost rešitev.

Posledica 8. *Naj bo x^* optimalna rešitev konveksnega problema, ki izpolnjuje Slaterjev pogoj. Naj bo (λ^*, ν^*) optimalna rešitev dualnega problema. Potem (x^*, λ^*, ν^*) zadošča KKT pogojem.*

Izkaže pa se, da so v primeru konveksnega optimizacijskega problema KKT pogoji tudi zadostni za optimalnost rešitev.

Trditev 14. *Naj bo dan konveksni optimizacijski problem* (3.4)*. Naj bodo $\overline{x}$, $\overline{\lambda}$ in $\overline{\nu}$ vektorji, ki zadoščajo KKT pogojem:*

$$\begin{aligned} f_i(\overline{x}) &\leq 0, \quad i = 1, \ldots, m \\ h_j(\overline{x}) &= 0, \quad j = 1, \ldots, p \\ \overline{\lambda}_i &\geq 0, \quad i = 1, \ldots, m \\ \overline{\lambda}_i f_i(\overline{x}) &= 0, \quad i = 1, \ldots, m \\ \nabla f_0(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda}_i \nabla f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\nu}_j \nabla h_j(\overline{x}) &= 0. \end{aligned}$$

Potem je $\overline{x}$ optimalna rešitev problema (3.4)*, $(\overline{\lambda}, \overline{\nu})$ optimalna rešitev dualnega problema in optimalni dualni razmik je enak nič.*

Dokaz. Prvi trije pogoji zagotavljajo, da sta $\overline{x}$ in $(\overline{\lambda}, \overline{\nu})$ primarno-dualni dopustni rešitvi. Zaradi pogoja $\overline{\lambda} \geq 0$ je Lagrangeova funkcija $L(x, \overline{\lambda}, \overline{\nu})$ konveksna funkcija v spremenljivlki x (saj je nenegativna vsota konveksnih funkcij, ki ji prištejemo afino funkcijo, še vedno konveksna funkcija (ref TODO)). Pogoj stacionarnosti pravi, da je gradient te funkcije po spremenljivki x enak nič, kar pomeni, da $\overline{x}$ minimizira $L(x, \overline{\lambda}, \overline{\nu})$ (ref TODO). Iz tega sledi

$$g(\overline{\lambda}, \overline{\nu}) = L(\overline{x}, \overline{\lambda}, \overline{\nu}) = f_0(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda}_i f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\nu}_j h_j(\overline{x}) = f_0(\overline{x}), \tag{3.13}$$

kjer smo pri zadnji enakosti upoštevali dopustnost $\overline{x}$ in pogoj komplementarnosti. Enakost (3.13) pove, da je dualni razmik enak 0, zato je $\overline{x}$ in $(\overline{\lambda}, \overline{\nu})$ par primarno-dualnih optimalnih rešitev. □

Izrek 6. *Naj bo dan konveksni optimizacijski problem* (3.4)*, pri čemer so kriterijska funkcija f_0 ter funkcije f_i odvedljive in predpostavimo, da velja krepka dualnost. Potem so KKT pogoji potrebni in zadostni pogoji za optimalnost, tj. x^* je optimalna rešitev problema* (3.4) *natanko takrat, ko obstaja tak (λ^*, ν^*), da (x^*, λ^*, ν^*) zadošča KKT pogojem. Pri tem je (λ^*, ν^*) optimalna rešitev dualnega problema.*

Dokaz. Uporabimo Trditvi 13 in 14. □

Za konveksni optimizacijski problem lahko Izrek (6) formuliramo kot naslednji ekvivalentni trditvi:

- x^* in (λ^*, ν^*) zadoščata KKT pogojem.
- x^* in (λ^*, ν^*) sta optimalni rešitvi primarnega in dualnega problema, ter krepka dualnost velja.

V primeru, ko je izpolnjen Slaterjev pogoj, pa dobimo poenostavljeni ekvivalentni trditvi:

- x^* in (λ^*, ν^*) zadoščata KKT pogojem.
- x^* in (λ^*, ν^*) sta optimalni rešitvi primarnega in dualnega problema.

Opomba 11. *Z naslednjim primerom bomo pokazali, da v primeru, ko krepka dualnost ne velja, sta lahko x^* in (λ^*, ν^*) optimalni rešitvi primarnega in dualnega problema, četudi ne zadoščata KKT pogojem.*

Primer 38. *(VAJE) Vzemimo optimizacijski problem iz Primera 35. Pokazali smo, da je $x = 0$ in $p^* = 1$, ter da je poljuben $\lambda \geq 0$ optimalna rešitev dualnega problema z optimalno vrednostjo $d^* = 0$. Torej krepka dualnost ne velja.*

Zapišimo KKT pogoje

$$x = 0, \quad \lambda \geq 0, \quad \lambda x^2 = 0, \quad -e^{-x} + 2\lambda x = 0.$$

Če vstavimo $x = 0$ v enačbo $-e^{-x} + 2\lambda x = 0$, dobimo $-1 = 0$, kar je protislovje. Torej nelinearni sistem, ki ga dobimo iz KKT pogojev, nima rešitve.

Naloga 30. *(VAJE) S pomočjo KKT pogojev poišči optimalno rešitev problema*

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & x^T Q x \leq 1, \end{aligned}$$

kjer je $Q \succ 0$ in $c \in \mathbb{R}^n$ neničeln vektor. Funkcija $f(x) = x^T Q x - 1$ je zaradi pozitivne definitnosti matrike Q konveksna. Torej gre za konveksni optimizacijski problem. Množica dopustnih rešitev je elipsoid v $\mathbb{R}^n$. Sestavimo Lagrangeovo funkcijo

$$L(x, \lambda) = c^T x + \lambda(x^T Q x - 1).$$

Par (x, λ) je optimalen, če izpolnjuje KKT pogoje:

$$x^T Q x \leq 1, \quad \lambda \geq 0, \quad \lambda(x^T Q x - 1) = 0, \quad \nabla_x L(x, \lambda) = c + 2\lambda Q x = 0.$$

Iz pogoja stacionarnosti dobimo $x = -\frac{1}{2\lambda} Q^{-1} c$. Ker je $\lambda \neq 0$ (saj $c \neq 0$), mora zaradi komplementarnosti veljati $x^T Q x = 1$. Če vstavimo x, dobimo

$$\frac{1}{4\lambda^2} c^T Q^{-1} c = 1.$$

Iz česar sledi, da je $\lambda = \frac{1}{2}\sqrt{c^T Q^{-1} c} \geq 0$. Končno dobimo

$$x = -\frac{Q^{-1} c}{\sqrt{c^T Q^{-1} c}} \quad \textit{in} \quad p^* = c^T x = -\sqrt{c^T Q^{-1} c}.$$

Primer 39 (Konveksni kvadratični program z linearnimi omejitvami)**.** *Dan je optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & \frac{1}{2} x^T P x + q^T x \\ \text{pri pogojih} \quad & Ax = b, \end{aligned}$$

kjer je $P \succeq 0$. Očitno je Slaterjev pogoj izpolnjen, zato lahko optimalni rešitvi primarnega in dualnega programa izračunamo iz KKT sistema. KKT pogoja sta

$$Ax^* = b \quad \textit{in} \quad Px^* + q + A^T \nu^* = 0.$$

V matrični obliki ju lahko zapišemo kot sistem linearnih enačb

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

Če je matrika sistema nesingularna, potem obstaja enolična primarno-dualno optimalna rešitev (x^*, ν^*). V primeru, ko je matrika sistema singularna in je sistem rešljiv, je vsak par rešitev (x^*, ν^*) optimalen. Če pa zgornji sistem ni rešljiv, je optimizacijski problem neomejen navzdol ali nedopusten. Glej Primer 19.

Pri izpeljavi KKT pogojev smo pokazali, da v primeru ko velja krepka dualnost in obstaja par primarno-dualnih optimalnih rešitev x^* in (λ^*, ν^*), potem

$$x^* \text{ minimizira Lagrangevo funkcijo } L(x, \lambda^*, \nu^*) = f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{j=1}^{p} \nu_j^* h_j(x). \tag{3.14}$$

Če je optimizijski problem konveksni, potem je funkcija $L(x, \lambda^*, \nu^*)$ konveksna in rešitev problema (3.14) izračunamo iz pogoja $\nabla_x L(x, \lambda^*, \nu^*) = 0$. Če je dobljena rešitev primarno dopustna, je tudi optimalna, sicer optimalna vrednost primarnega problema ni dosežena. Iz (3.14) tako dobimo metodo, ki izračuna optimalno rešitev primarnega problema, če poznamo optimalno rešitev dualnega problema. To je v praksi pomembno, saj lahko struktura dualnega problema omogoča lažje reševanje prvotnega problema.

Primer 40. *(VAJE) Za dano matriko $A \in \mathbb{R}^{m \times n}$ in vektor $b \in \mathbb{R}^m$ je podan konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & \sum_{i=1}^{n} x_i \ln x_i \\ \text{pri pogojih} \quad & Ax \leq b \\ & e^T x = 1. \end{aligned}$$

z domeno $\mathbb{R}^n_{++}$. Pri tem smo e označili vektor samih enic. Naj a_i označuje i-ti stolpec matrike A. Sestavimo Lagrangeovo funkcijo

$$\begin{aligned} L(x, \lambda, \nu) &= \sum_{i=1}^{n} x_i \ln x_i + \lambda^T (Ax - b) + \nu(e^T x - 1) \\ &= \sum_{i=1}^{n} x_i \ln x_i + \left(A^T \lambda + \nu e\right)^T x - b^T \lambda - \nu \\ &= \sum_{i=1}^{n} x_i \ln x_i + \sum_{i=1}^{n} (a_i^T \lambda + \nu) x_i - b^T \lambda - \nu. \end{aligned}$$

Za določitev dualne funkcije $g(\lambda, \nu) = \inf_{x>0} L(x, \lambda, \nu)$ izračunamo parcialne odvode

$$\frac{\partial L}{\partial x_i} = \ln x_i + 1 + a_i^T \lambda + \nu = 0.$$

Od tod dobimo, da je $x_i = e^{-\left(1+\nu+a_i^T\lambda\right)} > 0$. *Dualni problem je potem*

$$\begin{aligned} \text{maksimiziraj} \quad & -\nu - b^T\lambda - e^{-\nu-1}\sum_{i=1}^n e^{-a_i^T\lambda} \\ \text{pri pogojih} \quad & \lambda \geq 0. \end{aligned}$$

Denimo, da primarni problem izpolnjuje Slaterjev pogoj, tj. obstaja $x > 0$, *za katerega velja* $Ax \leq b$ *in* $e^T x = 1$. *Potem vemo, da velja krepka dualnost in obstaja optimalna rešitev* (λ^*, ν^*). *Predpostavimo, da smo rešili dualni problem. Potem je*

$$L(x, \lambda^*, \nu^*) = \sum_{i=1}^n x_i \ln x_i + \lambda^{*T}(Ax - b) + \nu^*(e^T x - 1)$$

navzdol omejena strogo konveksna funkcija. Torej obstaja enolični globalni minimum, ki je dosežen v

$$x_i^* = e^{-\left(1+\nu^*+a_i^T\lambda^*\right)}, \quad i = 1, \ldots, n.$$

Če je dobljen x^* *primarno dopusten, je tudi optimalen, sicer optimalna vrednost problema ni dosežena.*

Primer 41. *Za dan vektor* $a \in \mathbb{R}^n$ *in skalar* $b \in \mathbb{R}$ *je podan konveksni optimizacijski problema*

$$\begin{aligned} \text{minimiziraj} \quad & \sum_{i=1}^n f_i(x_i) \\ \text{pri pogojih} \quad & a^T x = b. \end{aligned}$$

Pri tem so $f_i \colon \mathbb{R} \to \mathbb{R}$ *strogo konveksne in odvedljive funkcije. Iz Lagrangeove funkcije*

$$L(x, \nu) = \sum_{i=1}^n f_i(x_i) + \nu\left(a^T x - b\right) = -b\nu + \sum_{i=1}^n \left(f_i(x_i) + \nu a_i x_i\right)$$

dobimo dualno funkcijo

$$g(\nu) = \inf_x \left(-b\nu + \sum_{i=1}^n \left(f_i(x_i) + \nu a_i x_i\right)\right) = -b\nu + \sum_{i=1}^n \inf_{x_i}\left(f_i(x_i) + \nu a_i x_i\right) = -b\nu + \sum_{i=1}^n f_i^*(\nu a_i).$$

Pri tem smo označili

$$f^*(y) = \inf_x \left(f(x) + yx\right).$$

Dualni problem je potem

$$\text{maksimiziraj} \quad -b\nu + \sum_{i=1}^n f_i^*(\nu a_i),$$

kjer je $\nu \in \mathbb{R}$. *Torej iščemo maksimum konkavne funkcije ene spremenljivke. Ker so* f_i *strogo konveksne, je Lagrangeova funkcija tudi strogo konveksna in optimalno rešitev primarnega problema dobimo iz*

$$\nabla_x L(x, \nu^*) = 0 \quad \textit{oz.} \quad f_i'(x_i) = -\nu^* a_i.$$

3.4 Dualnost pri semidefinitnem programiranju

Za dane simetrične matrike $C, A_1, \dots, A_m \in \mathcal{S}^n$ in vektor $b \in \mathbb{R}^m$ je podan primarni semidefinitni program (SDP) v standardni obliki

$$\begin{aligned} \text{minimiziraj} \quad & \langle C, X\rangle \\ \text{pri pogojih} \quad & \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m \\ & X \succeq 0 \quad (-X \preceq 0). \end{aligned} \tag{3.15}$$

Pri tem je s predpisom $\langle A, B\rangle = \text{sled}(AB)$ definiran skalarni produkt v prostoru simetričnih matrik. Naj p^* označuje optimalno vrednost problema (3.15). Omejitev neenakosti je mišljena glede na stožec pozitivno semidefinitnih matrik, tj. $X \in \mathcal{S}^n_+$. Ker ni oblike $f_i(X) \leq 0$ za končno mnogo funkcij f_i, je potrebno teorijo dualnosti razviti posebej.

Lagrangeova funkcija $L\colon \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n \to \mathbb{R}$ optimizacijskega problema (3.15) je definirana s predpisom

$$L(X, y, Z) = \langle C, X\rangle + \sum_{i=1}^m y_i \left(b_i - \langle A_i, X\rangle\right) - \langle Z, X\rangle.$$

Dualna funkcija je potem

$$\begin{aligned} g(y, Z) &= \inf_X \left(\langle C, X\rangle + \sum_{i=1}^m y_i \left(b_i - \langle A_i, X\rangle\right) - \langle Z, X\rangle\right) \\ &= \inf_X \left(b^T y + \left\langle C - \sum_{i=1}^m y_i A_i - Z, X\right\rangle\right) \\ &= \begin{cases} b^T y, & \textit{če } \; C - \sum\limits_{i=1}^m y_i A_i - Z = 0 \\ -\infty, & \textit{sicer} \end{cases}. \end{aligned}$$

Trditev 15. *Za vsak $Z \succeq 0$ velja*

$$g(y, Z) \leq p^*.$$

Dokaz. Naj bo $\overline{X}$ dopustna rešitev semidefinitnega programa (3.15) in $Z \succeq 0$. Po Lemi TODO? velja, da je skalarni produkt matrik $\overline{X}$ in Z nenegativen. Zato velja

$$\sum_{i=1}^m y_i \Bigg(\underbrace{b_i - \langle A_i, \overline{X}\rangle}_{=0}\Bigg) - \underbrace{\langle Z, \overline{X}\rangle}_{\geq 0} \leq 0.$$

Od tod sledi

$$g(y, Z) = \inf_X L(X, y, Z) \leq L(\overline{X}, y, Z) = \langle C, \overline{X}\rangle + \sum_{i=1}^m y_i \left(b_i - \langle A_i, \overline{X}\rangle\right) - \langle Z, \overline{X}\rangle \leq \langle C, \overline{X}\rangle.$$

Ker velja $g(y, Z) \leq \langle C, \overline{X}\rangle$ za poljubno dopustno rešitev $\overline{X}$, dobimo želeno neenakost. $\square$

Opomba 12. *Če bi v konveksnem optimizacijskem problemu* (3.15) *namesto stožca pozitivno semidefinitnih matrik nastopal splošni konveksni stožec $\mathcal{C}$, bi morali v Trditvi 15 za dualno spremenljivko Z zahtevati, da pripada dualnemu stožcu $\mathcal{C}^*$.*

Dualni problem prirejen problemu (3.15) je tako semidefinitni program oblike

$$\begin{aligned} \text{maksimiziraj} \quad & b^T y \\ \text{pri pogojih} \quad & C - \sum_{i=1}^{m} y_i A_i = Z \\ & Z \succeq 0 \end{aligned}$$

oziroma

$$\begin{aligned} \text{maksimiziraj} \quad & b^T y \\ \text{pri pogojih} \quad & \sum_{i=1}^{m} y_i A_i \preceq C. \end{aligned} \tag{3.16}$$

Posledica 9 (Šibka dualnost za SDP)**.** *Naj bosta p^* in d^* optimalni vrednosti primarnega in dualnega semidefinitnega programa. Potem velja*

$$d^* \leq p^*.$$

Definirajmo operator $\mathcal{A}\colon \mathcal{S}^n \to \mathbb{R}^m$ s predpisom

$$\mathcal{A}(X) = \begin{bmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{bmatrix}.$$

Njemu prirejen adjungiran operator $\mathcal{A}^T\colon \mathbb{R}^m \to \mathcal{S}^n$ je potem definiran z zahtevo, da velja

$$\langle \mathcal{A}(X), y \rangle = \left\langle X, \mathcal{A}^T(y) \right\rangle$$

za vsak $X \in \mathcal{S}^n$ in za vsak $y \in \mathbb{R}^m$. Od tod dobimo

$$\left\langle X, \mathcal{A}^T(y) \right\rangle = \langle \mathcal{A}(X), y \rangle = \sum_{i=1}^{m} y_i \langle A_i, X \rangle = \left\langle X, \sum_{i=1}^{m} y_i A_i \right\rangle.$$

Iz česar sledi

$$\mathcal{A}^T(y) = \sum_{i=1}^{m} y_i A_i.$$

S pomočjo operatorjev $\mathcal{A}$ in $\mathcal{A}^T$ lahko primarno-dualni par semidefinitnih programov (3.15) in (3.16) zapišemo v kompaktni obliki kot

$$\begin{aligned} \text{minimiziraj} \quad & \langle C, X \rangle \\ \text{pri pogojih} \quad & \mathcal{A}(X) = b \\ & X \succeq 0 \end{aligned} \tag{3.17}$$

in

$$\begin{aligned} \text{maksimiziraj} \quad & b^T y \\ \text{pri pogojih} \quad & \mathcal{A}^T(y) \preceq C. \end{aligned} \tag{3.18}$$

Opomba 13. *Čeprav vemo, da je dualni problem po definiciji konveksen, lahko to preverimo tudi direktno. V optimizacijskem problemu* (3.18) *maksimiziramo konkavno funkcijo po konveksni množici (glej Nalogo 11). Torej je problem res konveksen.*

Definicija 25 (Slaterjev pogoj za SDP)**.** *Pravimo, da je primarni semidefinitni program* (3.17) *strogo dopusten, če obstaja pozitivno definitna matrika* ($X \succ 0$)*, ki zadošča omejitvam* $\mathcal{A}(X) = b$*. Dualni semidefinitni program* (3.18) *je strogo dopusten, če obstaja* $y \in \mathbb{R}^m$*, da velja* $\mathcal{A}^T(y) \prec C$*.*

Izrek 7 (Krepka dualnost za SDP)**.** *Naj bo dan semidefinitni program, ki izpolnjuje Slaterjev pogoj. Potem velja krepka dualnost in optimalna vrednost dualnega problema je dosežena, če* $p^* > -\infty$*.*

Dokaz. Dokažemo podobno kot Izrek 5. □

Primer 42. *Dan je semidefinitni program*

$$\begin{aligned} \text{minimiziraj} \quad & \left\langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \right\rangle \\ \text{pri pogojih} \quad & \left\langle \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, X \right\rangle = 1, \quad \left\langle \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, X \right\rangle = 0, \quad X \succeq 0 \end{aligned}$$

in njegov dual

$$\begin{aligned} \text{maksimiziraj} \quad & y_1 \\ \text{pri pogojih} \quad & \begin{bmatrix} 0 & -y_1 & 0 \\ -y_1 & -y_2 & 0 \\ 0 & 0 & -y_1 + 1 \end{bmatrix} \succeq 0. \end{aligned}$$

Po Lemi TODO? velja, da množica dopustnih rešitev primarnega problema vsebuje pozitivno semidefinitne matrike oblike

$$X = \begin{bmatrix} x_1 & 0 & x_2 \\ 0 & 0 & 0 \\ x_2 & 0 & 1 \end{bmatrix}$$

in da je množica dopustnih rešitev dualnega problema enaka

$$\left\{ y \in \mathbb{R}^2 \mid y_1 = 0,\ y_2 \leq 0 \right\}.$$

Torej je $p^* = 1$ *in* $d^* = 0$*. Krepka dualnost v tem primeru ne velja. Niti primarni niti dualni program ne izpolnjuje Slaterjevega pogoja.*

Primer 43. *(VAJE) Dan je par primarno-dualnih semidefinitnih programov*

$$\begin{aligned} \text{minimiziraj} \quad & \left\langle \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, X \right\rangle \\ \text{pri pogojih} \quad & \left\langle \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, X \right\rangle = 2, \quad X \succeq 0, \end{aligned} \qquad \begin{aligned} \text{maksimiziraj} \quad & 2y \\ \text{pri pogojih} \quad & \begin{bmatrix} 1 & -y \\ -y & 0 \end{bmatrix} \succeq 0. \end{aligned}$$

Množica dopustnih rešitev primarnega programa vsebuje matrike oblike

$$X = \begin{bmatrix} x_1 & 1 \\ 1 & x_2 \end{bmatrix}.$$

Zaradi semidefinitne omejitve morajo biti vsi diagonalni elementi in determinanta nenegativni. Zaradi tega velja $x_1, x_2 \geq 0$ in $x_1 \geq \frac{1}{x_2}$. Ko gre $x_2 \to \infty$, se vrednost x_1 približuje 0. Zato optimalna vrednost programa $p^* = 0$ ni dosežena. Da bo matrika $\begin{bmatrix} 1 & -y \\ -y & 0 \end{bmatrix}$ dualno dopustna, mora njena determinanta biti nenegativna. Dobimo, da je $y = 0$. Torej je $d^* = 0$. Krepka dualnost velja, ampak ker optimalna vrednost primarnega programa ni dosežena, dualni problem ne more izpolnjevati Slaterjeva pogoja.

Poglavje 4

Nevezana konveksna optimizacija

V tem poglavju si bomo ogledali, kako poiščemo minimum gladke konveksne funkcije. Podan je nevezan optimizacijski problem oblike

$$\text{minimiziraj} \quad f(x), \tag{4.1}$$

pri čemer je $f\colon \mathcal{D}_f \subset \mathbb{R}^n \to \mathbb{R}$ dvakrat zvezno odvedljiva konveksna funkcija. Predpostavimo, da optimalna rešitev x^* obstaja. Optimalno vrednost označimo z $p^* = f(x^*) = \inf_x f(x)$.

Ker je funkcija f odvedljiva in konveksna, je

$$\nabla f(x^*) = 0 \tag{4.2}$$

potreben in zadosten pogoj za nastop ekstrema x^*. Torej je reševanje nevezanega optimizacijskega problema (4.1) enako reševanju (v splošnem) nelinearnega sistema (4.2) n enačb z n neznankami $x_1, \ldots, x_n$. Če je f kvadratna funkcija oblike $f(x) = \frac{1}{2}x^T Px + q^T x$, se problem prevede na reševanje sistema linearnih enačb $Px^* + q = 0$. Sicer je pa glavna ideja vseh metod ta, da skonstruiramo zaporedje približkov $x^{(k)} \in \mathcal{D}_f$, $k = 0, 1, \ldots$, ki konvergira proti optimalni rešitvi x^*, tj.

$$f\left(x^{(k)}\right) \to p^*, \quad \text{ko gre } \ k \to \infty.$$

Iskanje optimalne rešitve lahko tako interpretiramo kot reševanje pogoja optimalnosti (4.2) z iterativno metodo. Ko za dano toleranco $\varepsilon > 0$ in začetni približek $x^{(0)}$ dosežemo zaustavitveni pogoj $f\left(x^{(k)}\right) - p^* \leq \varepsilon$, se algoritem ustavi in kot približek za optimalno rešitev vrne $x^{(k)}$.

Metode v grobem razdelimo na metode 1. reda (*gradientna metoda*), ki za izračun približkov uporabljajo gradient, medtem ko pri metodah 2. reda (*Newtonova metoda*) še dodatno izkoristimo informacijo o Hessejevi matriki.

Primer 44. *1. Za dano matriko $A \in \mathbb{R}^{m\times n}$ in vektor $b \in \mathbb{R}^m$ je podan konveksni optimizacijski problem*

$$\text{minimiziraj} \quad f(x) = \ln\left(\sum_{i=1}^m e^{a_i^T x + b_i}\right).$$

Pogoj optimalnosti je

$$\nabla f(x^*) = \frac{1}{\sum_{j=1}^m e^{a_j^T x^* + b_j}} \sum_{i=1}^m e^{a_i^T x^* + b_i} a_i = 0,$$

ki v splošnem nima analitične rešitve. Zato moramo uporabiti interativno metodo. Ker je $\mathcal{D}_f = \mathbb{R}^n$*, lahko za* $x^{(0)}$ *izberemo poljubno točko.*

2. *Za dano matriko* $A \in \mathbb{R}^{m\times n}$ *in vektor* $b \in \mathbb{R}^m$ *je podan konveksni optimizacijski problem*

$$\text{minimiziraj} \quad f(x) = -\sum_{i=1}^{m} \ln\left(b_i - a_i^T x\right).$$

Domena kriterijske funkcije f *je odprta množica*

$$\mathcal{D}_f = \left\{x \in \mathbb{R}^n \mid a_i^T x < b_i,\ i = 1, \ldots, m\right\},$$

zato začetni približek $x^{(0)}$ *mora ležati v notranjosti poliedra, podanega z neenačbami* $Ax \leq b$*. Funkciji* f *pravimo* pregradna ali barierna funkcija za sistem neenačb $Ax \leq b$.

3. *Za dane matrike* $F_i \in \mathcal{S}^n$*,* $i = 0, 1, \ldots, m$ *je podan konveksni optimizacijski problem*

$$\text{minimiziraj} \quad f(x) = -\ln\det(F(x)).$$

Pri tem je $F\colon \mathbb{R}^m \to \mathcal{S}^n$ *afina funkcija oblike*

$$F(x) = F_0 + x_1F_1 + \ldots + x_mF_m = F_0 + \sum_{i=1}^{m} x_iF_i.$$

Domena kriterijske funkcije f *je odprta množica*

$$\mathcal{D}_f = \{x \in \mathbb{R}^m \mid F(x) \succ 0\}.$$

Funkciji f *pravimo* pregradna ali barierna funkcija za linearno matrično neenakost $F(x) \succeq 0$.

4.1 Krepka konveksnost

Tekom poglavja bomo privzeli, da je kriterijska funkcija krepko konveksna.

Definicija 26. *Pravimo, da je funkcija* $f\colon \mathbb{R}^n \to \mathbb{R}$ krepko konveksna*, če obstaja* $m > 0$*, da velja*

$$\nabla^2 f(x) \succeq mI$$

za vsak $x \in \mathbb{R}^n$*.*

Opomba 14. *Če je funkcija krepko konveksna, potem sledi, da je tudi strogo konveksna. Namreč za* $m > 0$ *pogoj* $\nabla^2 f(x) \succeq mI$ *implicira* $\nabla^2 f(x) \succ 0$*.*

Lema 7. *Naj bo $f\colon \mathbb{R}^n \to \mathbb{R}$ krepko konveksna funkcija. Potem velja*

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{m}{2}\|y-x\|_2^2 \tag{4.3}$$

za vsaka x in y iz $\mathbb{R}^n$.

Dokaz. Po Taylorjevi formuli za poljubna x in y velja

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

kjer leži z na daljici $[x, y]$. Zaradi krepke konveksnosti lahko zadnji člen na desni strani neenakosti ocenimo navzdol:

$$\frac{1}{2}(y-x)^T\nabla^2 f(z)(y-x) \geq \frac{m}{2}\|y-x\|_2^2.$$

Dobimo želeno neenakost. □

Ko je $m = 0$, dobimo neenakost (1.5), ki karakterizira konveksne funkcije. Spodnja meja je afina funkcija. Če pa je funkcija krepko konveksna ($m > 0$), dobimo spodnjo mejo, ki je kvadratna funkcija spremenljivke y.

Lema 8. *Naj bo $f\colon \mathbb{R}^n \to \mathbb{R}$ krepko konveksna funkcija. Potem velja neenakost*

$$f(x) - p^* \leq \frac{1}{2m}\left\|\nabla f\left(x\right)\right\|_2^2, \tag{4.4}$$

ki poda oceno, kako blizu optimalne rešitve smo.

Dokaz. Ker je f krepko konveksna, velja neenakost (4.3). Desna stran je konveksna kvadratna funkcija v spremenkjivki y. Minimum $\widehat{y}$ je dosežen, kjer je gradient enak nič. Dobimo

$$\nabla f(x) + m(\widehat{y} - x) = 0$$

oziroma

$$\widehat{y} = x - \frac{1}{m}\nabla f(x).$$

Če to vstavimo v kvadratno funkcijo na desni strani neenačbe, dobimo oceno

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

Ker to velja za vsak y, dobimo

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

□

V praksi ponavadi konstante m ne poznamo. Kljub temu ocena (4.4) pokaže, da je v primeru majhne norme gradienta $\nabla f\left(x^{(k)}\right)$ vrednost $f\left(x^{(k)}\right)$ blizu p^*. Zato se za dano toleranco $\varepsilon > 0$ pogoj $\left\|\nabla f\left(x^{(k)}\right)\right\|_2^2 < \varepsilon$ uporablja kot zaustavitveni kriterij v algoritmih.

4.2 Gradientna metoda

Splošni nastavek je, da se iterativno približujemo minimumu. Naj bo $x^{(k)}$ tekoči približek. Izberemo vektor $\Delta x^{(k)} \in \mathbb{R}^n$ in v tej smeri poiščemo naslednji približek oblike

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)},$$

da je $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$. Torej na vsakem koraku želimo zmanjšati vrednost funkcije f. Vektorju Δx in številu $t \geq 0$ pravimo *smer* in *premik*.

Zaradi zahteve $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$ in konveksnosti funkcije f velja

$$f\left(x^{(k)}\right) + t \nabla f\left(x^{(k)}\right)^T \Delta x^{(k)} \leq f\left(x^{(k)} + t \Delta x^{(k)}\right) = f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right).$$

Od tod dobimo pogoj

$$\nabla f\left(x^{(k)}\right)^T \Delta x^{(k)} < 0.$$

Smeri Δx, ki zadošča temo pogoju, pravimo *smer spusta.* Torej je kot med negativnim gradientom (smer največjega padanja funkcije) in smerjo spusta ostri.

Zapišimo splošni algoritem za iskanje minimuma funkcije f.

Algorithm 4.2.1: Splošna metoda spusta.

---

vhodni podatki: začetni približek $x^{(0)} \in \mathcal{D}_f$

for $k = 0, 1, \ldots,$ **do**
- Izberi smer spusta $\Delta x^{(k)}$.
- Določi premik $t^{(k)} > 0$.
- Izračunaj novi približek $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$.

Pogosto se kot zaustavitveni pogoj uporablja $\|\nabla f(x)\|_2 \leq \varepsilon$, kjer je $\varepsilon > 0$ dana toleranca. Ta izbira je smiselna, saj iterativno rešujemo sistem $\nabla f(x) = 0$ in je dodatno motivirana zaradi neenakosti (4.4). Ko bo norma gradienta dovolj majhna, bo vrednost $f\left(x^{(k)}\right)$ blizu p^*.

Pogledati še moramo:

1. kako določimo premik t,
2. kako izberemo smer Δx.

Izbira premika. Pri določanju premika t imamo opravka s funkcijo ene spremenljivke

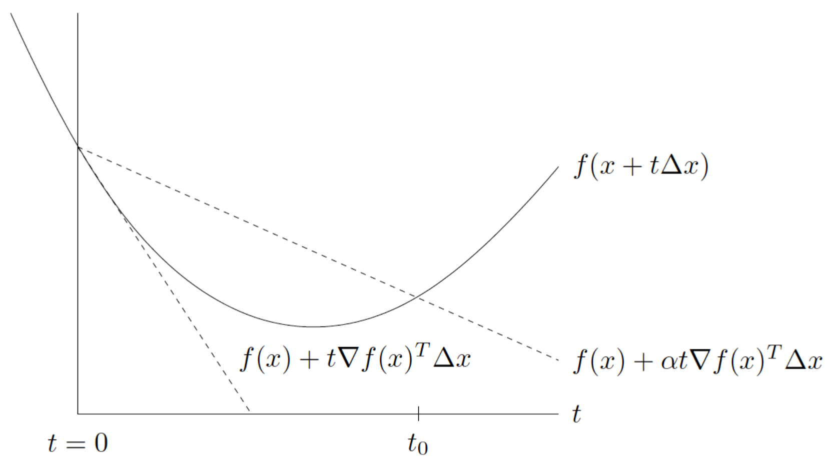
$$g(t) = f\left(x^{(k)} + t \Delta x^{(k)}\right)$$

Iščemo tak $t^{(k)} \geq 0$, da je $g\left(t^{(k)}\right) < g(0)$. Za določitev premika se v praksi uporabljata dva pristopa.

- *Metoda največjega spusta (exact line search)*: poiščemo $t^{(k)}$, kjer funkcija g doseže svoj minimum. Za to analitično rešimo nelinearno enačbo $g'\left(t^{(k)}\right) = 0$ ali pa uporabimo kakšno metodo za računanje minimuma funkcije ene spremenljivke (iskanje ničle odvoda), npr. bisekcijo ali tangentno metodo.

- *Metoda diskretnega spusta (backtracking line search)*: Najprej izberemo dva parametra α in β, za katera velja $\alpha \in (0, 0.5)$ in $\beta \in (0, 1)$. Na sliki 4.1 vidimo, da je premica z enačbo $y = f(x) + t\nabla f(x)^T \Delta x$ tangenta na graf funkcije $g(t) = f(x + t\Delta x)$ v točki $t = 0$. Zmanjšajmo naklon te premice s faktorjem α. Dobimo premico z enačbo $y = f(x) + \alpha t \nabla f(x)^T \Delta x$, ki je na sliki 4.1 poleg tangente označena s črtkano črto. Ideja je, da za premik želimo najti takšen t, da bo graf funkcije g pod premico z enačbo $y = f(x) + \alpha t \nabla f(x)^T \Delta x$. To naredimo tako, da začnemo s premikom $t = 1$ in ga zmanjšujemo s faktorjem β toliko časa, dokler ni izpolnjen pogoj

$$f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x. \tag{4.5}$$



Slika 4.1: Metoda diskretnega spusta. Na sliki je graf funkcije g (zožitev funkcije f v smeri Δx). Spodnja črtkana črta prikazuje tangento na graf funkcije g v točki $t = 0$, zgornja črta pa premico z naklonom, ki je za faktor α manjši. Pogoj (4.5) je izpolnjen, ko graf funkcije g leži pod zgornjo črtkano črto, tj. za vse $0 \le t \le t_0$.

Algorithm 4.2.2: Metoda diskretnega spusta.

---

vhodni podatki: smer spusta Δx, približek $x \in \mathcal{D}_f$, $\alpha \in (0, 0.5)$ in $\beta \in (0, 1)$

while $f(x + t\Delta x) \ge f(x) + \alpha t \nabla f(x)^T \Delta x$ **do**

  $t \leftarrow \beta t$

Tipični vrednosti za parametra α in β sta $\alpha \in (0.01, 0.3)$ in $\beta = 0.5$.

Izbira smeri. Naravna izbira za smer je negativni gradient $\Delta x = -\nabla f(x)$, saj funkcija f v tej smeri najhitreje pada. Negativni gradient je res smer spusta, saj velja

$$\nabla f(x)^T \Delta x = -\|\nabla f(x)\|_2^2 < 0,$$

razen, če je x že optimalen ($\nabla f(x) = 0$). Tako dobimo *gradientno metodo*.

Algorithm 4.2.3: Gradientna metoda.

vhodni podatki: začetni približek $x^{(0)} \in \mathcal{D}_f$

for $k = 0, 1, \ldots,$ **do**
- $\Delta x^{(k)} = -\nabla f\left(x^{(k)}\right)$.
- Določi premik $t^{(k)} > 0$ (največji spust ali diskretni spust).
- Izračunaj novi približek $x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)}$.

Analiza konvergence gradientne metode. Definirajmo množico $S = \{x \in \mathcal{D}_f \mid f(x) \leq f(x_0)\}$. Za dani začetni približek x_0 iščemo minimum funkcije f, zato se lahko omejimo na območje S.

Lema 9. *Naj bo $S = \{x \in \mathcal{D}_f \mid f(x) \leq f(x_0)\}$. Denimo, da je f krepko konveksna funkcija na S. Potem je množica S kompaktna.*

Dokaz. Ker velja $S = f^{-1}\left((-\infty, f(x_0)]\right)$, je S zaprta množica. Pokažimo še omejenost. Ker je funkcija f krepko konveksna na S, velja (4.3). Če v to neenakost vstavimo $x = x^*$ in upoštevamo, da je x^* globalni minimum ($\nabla f(x^*) = 0$), dobimo, da za vsak $y \in S$ velja

$$f(y) \geq f(x^*) + \frac{m}{2}\|y - x^*\|_2^2.$$

Po definiciji množice S velja

$$f(x_0) \geq f(y) \geq f(x^*) + \frac{m}{2}\|y - x^*\|_2^2.$$

Od tod dobimo oceno

$$\|y - x^*\|_2^2 \leq \frac{2}{m}\left(f(x_0) - f(x^*)\right).$$

Torej je množica S tudi omejena in posledično kompaktna. □

Posledica 10. *Naj bo funkcija f krepko konveksna na množici $S = \{x \in \mathcal{D}_f \mid f(x) \leq f(x_0)\}$, tj. obstaja $m > 0$, da velja $\nabla^2 f(x) \succeq mI$. Potem obstaja konstanta $M > 0$, da velja*

$$\nabla^2 f(x) \preceq MI$$

za vsak $x \in S$. Torej so lastne vrednosti Hessejeve matrike tudi navzgor omejene.

Dokaz. Največja lastna vrednost matrike $\nabla^2 f(x)$ je zvezna funkcija od x. Na zaprti in omejeni, torej kompaktni množici doseže svoj maksimum. Zato obstaja število $M > 0$, da velja $\lambda_{\max}\left(\nabla^2 f(x)\right) \leq M$ oziroma $\nabla^2 f(x) \preceq MI$. □

Podobno kot smo iz pogoja $\nabla^2 f(x) \succeq mI$ izpeljali neenakost (4.3), lahko pokažemo, da iz $\nabla^2 f(x) \preceq MI$ sledi

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{M}{2}\|y - x\|_2^2 \tag{4.6}$$

za vsaka x in y iz S.

Izrek 8. *Naj bo f krepko konveksna funkcija, za katero velja $mI \preceq \nabla^2 f(x) \preceq MI$ za vsak $x \in \mathcal{D}_f$. Potem za gradientno metodo, ki uporablja metodo največjega spusta (exact line search) velja*

$$f\left(x^{(k+1)}\right) - p^* \leq \left(1 - \frac{m}{M}\right)\left(f\left(x^{(k)}\right) - p^*\right). \tag{4.7}$$

Torej ima gradientna metoda linearno konvegenco. *Dodatno, za $c = 1 - \frac{m}{M}$ je število potrebnih korakov, da zadostimo pogoju $f\left(x^{(k)}\right) - p^* \leq \varepsilon$ omejeno z*

$$k \leq \frac{\ln\left(\frac{f\left(x^0\right) - p^*}{\varepsilon}\right)}{\ln\left(\frac{1}{c}\right)}.$$

Dokaz. Zaradi lažjega zapisa vpeljimo oznako $x^+ = x - t\nabla f(x)$ za $x^{(k+1)} = x^{(k)} - t^{(k)}\nabla f\left(x^{(k)}\right)$. Definirajmo funkcijo $\overline{f}(t) = f(x - t\nabla f(x))$. Če vstavimo v neenakost (4.6) $y = x - t\nabla f(x)$, dobimo kvadratično zgornjo mejo za $\overline{f}$:

$$\overline{f}(t) \leq f(x) - t\|\nabla f(x)\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x)\|_2^2.$$

Obe strani zgornje neenakosti minimizirajmo po t. Na levi strani dobimo $\overline{f}(t_{exact})$, pri čemer je premik t_{exact} dobljen po metodi največjega spusta. Na desni strani je minimum kvadratne funkcije dosežen pri $t = \frac{1}{M}$ z vrednostjo $f(x) - \frac{1}{2M}\|\nabla f(x)\|_2^2$. Od tod dobimo zvezo

$$f(x^+) = \overline{f}(t_{exact}) \leq f(x) - \frac{1}{2M}\|\nabla f(x)\|_2^2.$$

Če na obeh straneh te neenakosti odštejemo p^*, dobimo

$$f(x^+) - p^* \leq f(x) - p^* - \frac{1}{2M}\|\nabla f(x)\|_2^2. \tag{4.8}$$

Zaradi neenakosti (4.3) velja $\|\nabla f(x)\|_2^2 \geq 2m\left(f(x) - p^*\right)$. Če to uporabimo v (4.8), dobimo

$$f(x^+) - p^* \leq \left(1 - \frac{m}{M}\right)\left(f\left(x\right) - p^*\right),$$

kar pa je neenakost (4.7). Če to neenakost rekurzivno uporabimo, sledi

$$f\left(x^{(k)}\right) - p^* \leq c^k\left(f\left(x^{(0)}\right) - p^*\right), \tag{4.9}$$

kjer je $c = 1 - \frac{m}{M} < 1$. To dokazuje, da $f\left(x^{(k)}\right)$ konvergira proti p^*, ko gre $k \to \infty$. Neenakost $f\left(x^{(k)}\right) - p^* \leq \varepsilon$ dobimo po kvečjemu

$$k \leq \frac{\ln\left(\frac{f\left(x^0\right) - p^*}{\varepsilon}\right)}{\ln\left(\frac{1}{c}\right)}.$$

iteracij. $\square$

Števili m in M sta povezani s pogojenostnim številom Hessejeve matrike. Spomnimo se, da je *pogojenostno število* obrnljive matrike A definirano kot

$$\kappa(A) = \|A\|_2 \cdot \|A^{-1}\|_2 = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

oziroma

$$\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)},$$

če je matrika simetrična in pozitivno definitna.

Za krepko konveksno funkcijo f tako dobimo oceno za pogojenostno število Hessejeve matrike:

$$\kappa\left(\nabla^2 f(x)\right) = \frac{\lambda_{\max}\left(\nabla^2 f(x)\right)}{\lambda_{\min}\left(\nabla^2 f(x)\right)} \leq \frac{M}{m}.$$

V naslednjem primeru bomo videli, kako veliko pogojenostno število Hessejeve matrike vpliva na počasno konvergenco gradientne metode.

Primer 45. *Za dani parameter $\gamma > 0$ je podan optimizacijski problem*

$$\text{minimiziraj} \quad f(x) = \frac{1}{2}\left(x_1^2 + \gamma x_2^2\right).$$

Za iskanje minimuma uporabimo gradientno metodo, ki za določitev premika uporablja metodo največjega spusta. Za začetni približek izberemo $x^{(0)} = (\gamma, 1)$. V tem primeru lahko vsak naslednji približek določimo analitično:

$$x_1^{(k)} = \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k, \quad x_2^{(k)} = \left(-\frac{\gamma-1}{\gamma+1}\right)^k. \tag{4.10}$$

Pri tem je

$$f\left(x^{(k)}\right) = \left(\frac{\gamma-1}{\gamma+1}\right)^{2k} f\left(x^{(0)}\right).$$

Potek algoritma pri vrednosti parametra $\gamma = 10$ je prikazan na Sliki 4.2. Približki, dobljeni po gradientni metodi skonvergirajo proti optimalni rešitvi $x^* = (0,0)$. Če bi nivojnice bile krožnice ($\gamma = 1$), bi gradientna metodo skonvergirala v enem koraku. Bolj ko so sploščene, več korakov potrebuje gradientna metoda. Izkaže se, da je to povezano s pogojenostnim številom Hessejeve matrike. Večji je $\kappa\left(\nabla^2 f(x)\right)$, več iteracij potrebuje gradientna metoda, da skonvergira.

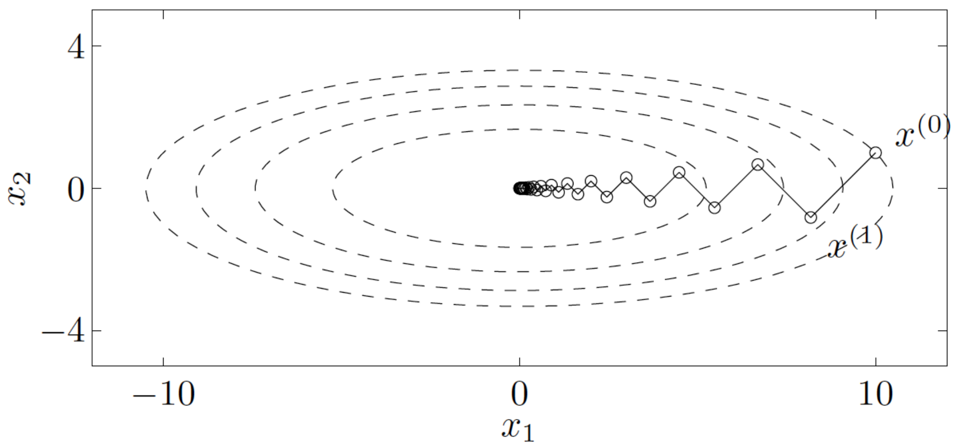
Naloga 31. *(VAJE) Preveri, da veljajo formule* (4.10).

Rešitev: *Gradient funkcije $f(x) = \frac{1}{2}\left(x_1^2 + \gamma x_2^2\right)$ v točki $x^{(k)}$ je enak*

$$\nabla f\left(x^{(k)}\right) = \begin{bmatrix} x_1^{(k)} \\ \gamma x_2^{(k)} \end{bmatrix}.$$

Ker je začetni približek enak $x^{(0)} = (\gamma, 1)$, dobimo po gradientni metodi

$$x^{(1)} = \begin{bmatrix} (1-t)\gamma \\ 1-\gamma t \end{bmatrix}.$$



Slika 4.2: Na sliki so nivojnice funkcije $f(x) = \frac{1}{2}(x_1^2 + 10x_2^2)$ in približki, dobljeni po gradienti metodi, ki uporablja metodo največjega spusta.

Ker uporabljamo metodo največjega spusta, iščemo tak t, ki minimizira funkcijo

$$g(t) = \frac{1}{2}(1-t)^2\gamma^2 + \frac{1}{2}\gamma(1-\gamma t)^2.$$

Ker je

$$g'(t)\gamma^2(t(1+\gamma)-2),$$

dobimo, da je minimum dosežen pri $t_{opt} = \frac{2}{1+\gamma}$. od tod sledi

$$1 - t_{opt} = \frac{\gamma-1}{\gamma+1}, \quad 1 - \gamma t_{opt} = -\frac{\gamma-1}{\gamma+1}.$$

Če to ponovimo za vsak približek, dobimo formulo (4.10).

Primer 46. *Za $n = 100$ in $m = 500$ je podan konveksni optimizacijski problem oblike*

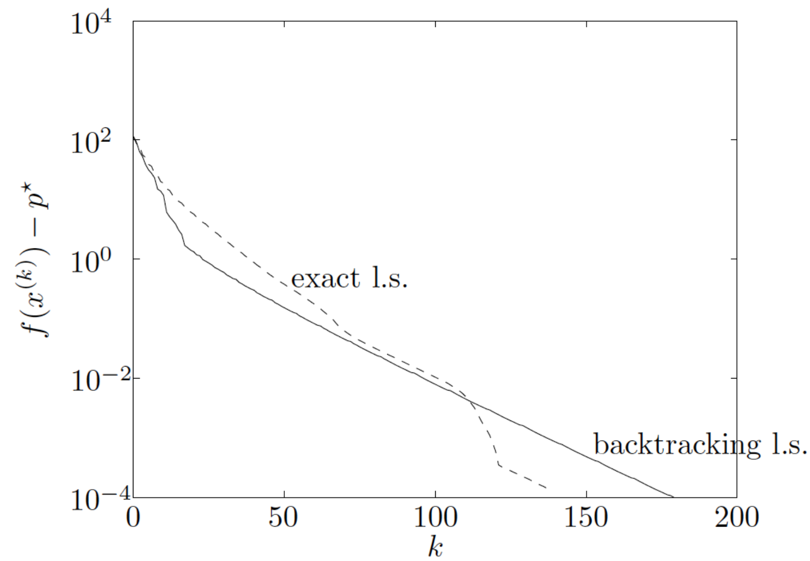
$$\text{minimiziraj} \quad c^T x - \sum_{i=1}^{m} \ln\big(b_i - a_i^T x\big). \tag{4.11}$$

Obnašanje gradientne metodo, ki uporablja metodi največjega spusta in diskretnega spusta $(\alpha = 0.1, \beta = 0.5)$ *je prikazano na Sliki 4.3.*

4.3 Newtonova metoda

Newtonova metoda za iskanje minimuma funkcije je aplikacija Newtonove metode za reševanje sistema nelinearnih enačb. Sistem, ki ga rešujemo je $\nabla f(x) = 0$.

Metodo lahko izpeljemo preko Taylorjeve vrste ali z linearizacijo pogoja optimalnosti.



Slika 4.3: Na sliki je prikazano obnašanje gradientne metode, ki uporablja metodi največjega spusta in diskretnega spusta ($\alpha = 0.1, \beta = 0.5$) za optimizacijski problem (4.11). Skala na y osi je logaritemska. Na sliki je lepo vidna linearna konvergenca.

- Taylorjev polinom drugega reda $\widehat{f}$ za funkcijo f v točki x je

$$\widehat{f}(x + \Delta x) = f(x) + \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x,$$

  ki je konveksna kvadratna funkcija spremenljivke Δx. Minimum je dosežen pri

$$\Delta x = -\nabla^2 f(x)^{-1} \nabla f(x).$$

  Glej sliko 4.4.

- Drugi način, kako izpeljemo Newtonovo metodo je preko linearizacije pogoja optimalnosti $\nabla f(x) = 0$. Denimo, da je x tekoči približek. Iščemo tak popravek Δx, da bo veljalo

$$\nabla f(x + \Delta x) = 0.$$

  Glej sliko 4.5. Ker pa je

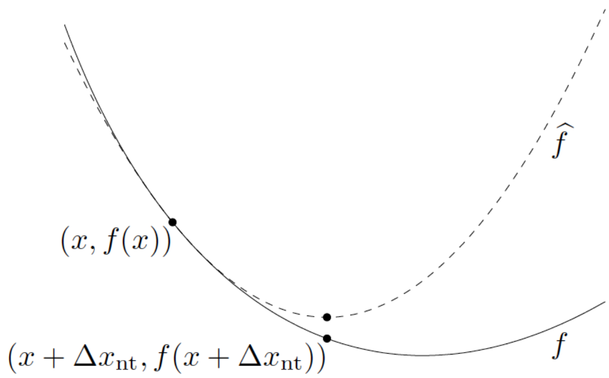
$$\nabla f(x + \Delta x) \approx \nabla \widehat{f}(x + \Delta x) = \nabla f(x) + \nabla^2 f(x) \Delta x = 0,$$

  dobimo rešitev $\Delta x = -\nabla^2 f(x)^{-1} \nabla f(x)$.

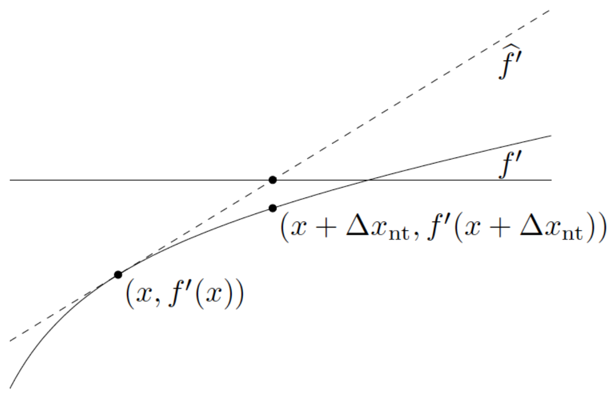
Definicija 27. *Vektorju*

$$\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

pravimo Newtonova smer.



Slika 4.4: Funkcijo f aproksimiramo v bližini točke x s Taylorjevim polinomom $\widehat{f}$ drugega reda. Iz točke x se moramo premakniti v Newtonovi smeri Δx_{nt}, da minimiziramo funkcijo $\widehat{f}$.



Slika 4.5: Na sliki sta narisana odvod f' funkcije f in njegova linearna aproksimacija $\widehat{f'}$ v točki x. Newtonova smer je razlika med ničlo funkcije $\widehat{f'}$ in točko x.

Zaradi pozitivne definitnosti Hessejeve matrike velja

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

razen, če je x že optimalen ($\nabla f(x) = 0$). Zato je Newtonova smer res smer spusta.

Interpretacija Newtonove metode. Če je f kvadratna funkcija, je $x + \Delta x_{nt}$ optimalna rešitev. Sicer v bližini trenutnega približka x funkcijo f aproksimiramo s kvadratno funkcijo $\widehat{f}$ in kot nov približek vzamemu minimum le-te. Ko smo blizu optimalne rešitve x^* ($\nabla f\,(x) \approx 0$), je kvadratični model $\widehat{f}$ zelo dobra aproksimacija funkcije f. Zato je $x + \Delta x_{nt}$ zelo blizu optimalne rešitve. Izkaže se, da ima Newtonova metoda *kvadratično konvergenco.*

Zaustavitveni pogoj Newtonove metode. Ker imamo na voljo informacijo o Hessejevi matriki, lahko za Newtonovo metodo izpeljemo boljši zaustavitveni pogoj, ki je neodvisen od konstante m (krepka konveksnost).

Radi bi naredili oceno za razliko $f(x) - p^*$. Ko je x blizu x^*, je minimum kvadratične aproksimacije $\widehat{f}$ zelo blizu p^* Zato razlika $f(x) - \inf_y \widehat{f}(y)$ podaja dobro oceno za $f(x) - p^*$. Dobimo

$$\begin{aligned}
f(x) - \inf_y \widehat{f}(y) &= f(x) - \widehat{f}(x + \Delta x_{nt}) \\
&= f(x) - f(x) + \nabla f(x)^T \Delta x_{nt} - \frac{1}{2}\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt} \\
&= f(x) - f(x) + \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) - \frac{1}{2} \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) \\
&= \frac{1}{2}\lambda(x)^2,
\end{aligned}$$

pri čemer smo označili

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

Po definiciji Newtonove smeri lahko tudi zapišemo

$$\lambda(x) = \sqrt{\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt}}. \tag{4.12}$$

Od tod dobimo *Newtonovo metodo.*

Algorithm 4.3.1: Newtonova metoda.

vhodni podatki: začetni približek $x^{(0)} \in \mathcal{D}_f$, toleranca $\varepsilon > 0$

for $k = 0, 1, \ldots,$ **do**
- Izračunaj Newtonovo smer $\Delta x_{nt}^{(k)} = -\nabla^2 f\left(x^{(k)}\right)^{-1} \nabla f\left(x^{(k)}\right)$ in $\lambda^2\left(x^{(k)}\right) = \nabla f\left(x^{(k)}\right)^T \nabla^2 f\left(x^{(k)}\right)^{-1} \nabla f\left(x^{(k)}\right)$
- **if** $\lambda^2\left(x^{(k)}\right)/2 \leq \varepsilon$ **then**
  - končaj
- Določi premik $t^{(k)} > 0$ (največji spust ali diskretni spust).
- Izračunaj novi približek $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x_{nt}^{(k)}$.

Izračun Newtonove smeri. Zanima nas, kako na učinkovit način izračunamo vektor Δx_{nt}.

Najprej izračunamo gradient $g = \nabla f(x)$ in Hessejevo matriko $H = \nabla^2 f(x)$ v dani točki x. Da določimo Newtonovo smer Δx_{nt} je potrebno rešiti sistem linearnih enačb

$$H \Delta x_{nt} = -g.$$

Ker je matrika H simetrična in pozitivno definitna, lahko uporabimo razcep Choleskega $H = LL^T$, kjer je L spodnje trikotna matrika. Najprej s premo substitucijo rešimo sistem $Ly = -g$, da dobimo $y = -L^{-1}g$. Nato z obratno substitucijo rešimo sistem $L^T \Delta x_{nt} = y$, da dobimo $\Delta x_{nt} = -H^{-1}g$. Razcep Choleskega uporabimo tudi pri izračunu λ:

$$\lambda^2 = g^T H^{-1} g = g^T L^{-T} L^{-1} g = \left\|L^{-1} g\right\|_2^2 = \|y\|_2^2.$$

Analiza konvergence Newtonove metode. Newtonova metoda temelji na kvadratični aproksimaciji kriterijske funkcije. Če je kriterijska funkcija kvadratična, potem Newtonova metoda skonvergira v enem koraku. Boljša kot je ta kvadratična aproksimacija, hitrejša bo Newtonova metoda. Matematično to najlažje izrazimo preko Lipschitzove lastnosti Hessejeve matrike.

Definicija 28. *Dana je funkcija $f\colon \mathcal{D}_f \to \mathbb{R}$. Pravimo, da je Hessejeva matrika $\nabla^2 f$* Lipschitzova *s konstanto L, če velja*

$$\left\|\nabla^2 f(y) - \nabla^2 f(x)\right\|_2 \leq L \left\|y - x\right\|_2$$

Koeficient L lahko interpretiramo kot zgornjo mejo za tretji odvod. Če je torej število L majhno, tj. Hessejeva matrika se ne spreminja veliko, potem se kvadratični model ne spreminja veliko in Newtonova metoda hitro skonvergira.

Lema 10. *Naj bo f dvakrat zvezno odvedljiva na $\mathbb{R}^n$. Potem velja*

$$\nabla f(y) - \nabla f(x) = \int_0^1 \left[\nabla^2 f\left(x + t(y-x)\right)\right](y-x)dt.$$

Dokaz. Definirajmo funkcijo $\phi(t) = \nabla f(x + t(y-x))$. Potem velja $\phi(0) = \nabla f(x)$, $\phi(1) = \nabla f(y)$ in $\phi'(t) = \left[\nabla^2 f\left(x + t(y-x)\right)\right](y-x)$. Po osnovnem izreku integralskega računa dobimo

$$\begin{aligned}\nabla f(y) - \nabla f(x) &= \phi(1) - \phi(0)\\ &= \int_0^1 \phi'(t)dt\\ &= \int_0^1 \left[\nabla^2 f\left(x + t(y-x)\right)\right](y-x)dt.\end{aligned}$$

□

Izrek 9 (O kvadratični konvergenci Newtonove metode)**.** *Naj bo f dvakrat zvezno odvedljiva na $\mathbb{R}^n$. Naj bo f krepko konveksna s konstanto m ($\nabla^2 f(x) \succeq mI$) in naj bo Hessejeva matrika funkcije f Lipschitzova s konstanto L. Potem za Newtonovo metodo velja*

$$\left\|x^{(k+1)} - x^*\right\|_2^2 \leq \frac{L}{2m}\left\|x^{(k)} - x^*\right\|_2^2.$$

Torej ima Newtonova metoda kvadratično konvergenco.

Dokaz. Po Lemi 10 velja

$$\begin{aligned}x^{(k+1)} - x^* &= x^{(k)} - \nabla^2 f\left(x^{(k)}\right)^{-1}\nabla f\left(x^{(k)}\right) - x^*\\ &= x^{(k)} - x^* + \nabla^2 f\left(x^{(k)}\right)^{-1}\left(\nabla f\left(x^*\right) - \nabla f\left(x^{(k)}\right)\right)\\ &= x^{(k)} - x^* + \nabla^2 f\left(x^{(k)}\right)^{-1}\int_0^1 \left[\nabla^2 f\left(x^{(k)} + t(x^* - x^{(k)})\right)\right](x^* - x^{(k)})dt\\ &= \nabla^2 f\left(x^{(k)}\right)^{-1}\int_0^1 \left[\nabla^2 f\left(x^{(k)} + t(x^* - x^{(k)})\right) - \nabla^2 f\left(x^{(k)}\right)\right](x^* - x^{(k)})dt.\end{aligned}$$

Od tod dobimo oceno

$$\begin{aligned}
\left\|x^{(k+1)} - x^*\right\|_2 &\leq \left\|\nabla^2 f\left(x^{(k)}\right)^{-1}\right\|_2 \left\|\int_0^1 \left[\nabla^2 f\left(x^{(k)} + t(x^* - x^{(k)})\right) - \nabla^2 f\left(x^{(k)}\right)\right](x^* - x^{(k)})dt\right\|_2 \\
&\leq \left\|\nabla^2 f\left(x^{(k)}\right)^{-1}\right\|_2 \int_0^1 \left\|\left[\nabla^2 f\left(x^{(k)} + t(x^* - x^{(k)})\right) - \nabla^2 f\left(x^{(k)}\right)\right](x^* - x^{(k)})\right\|_2 dt \\
&\leq \left\|\nabla^2 f\left(x^{(k)}\right)^{-1}\right\|_2 \int_0^1 \left\|\nabla^2 f\left(x^{(k)} + t(x^* - x^{(k)})\right) - \nabla^2 f\left(x^{(k)}\right)\right\|_2 \left\|(x^* - x^{(k)})\right\|_2 dt \\
&\leq \frac{1}{m}\int_0^1 Lt\left\|(x^* - x^{(k)})\right\|_2^2 dt \\
&= \frac{L}{2m}\left\|x^{(k)} - x^*\right\|_2^2. \qquad \square
\end{aligned}$$

Z naslednjima primeroma bomo pokazali, zakaj je v splošnem potrebno uporabljati premike.

Naloga 32. (VAJE) *Newtonova metoda s fiksnim premikom $t = 1$ lahko divergira, če začetni približek ni dovolj blizu optimalne rešitve x^*. Dan je konveksni optimizacijski problem*

$$\text{minimiziraj} \quad f(x) = \log\left(e^x + e^{-x}\right).$$

Minimum je dosežen v točki $x^* = 0$. Oglej si delovanje Newtonove metode s fiksnim premikom $t = 1$, če za začetna približka vzameš $x^{(0)} = 1$ in $x^{(0)} = 1.1$.

Rešitev: *f je gladka strogo konveksna funkcija z $\mathcal{D}_f = \mathbb{R}$. Velja*

$$f'(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{in} \quad f''(x) = \frac{4}{(e^x + e^{-x})^2}.$$

Točka $x = 0$ je edina rešitev enačbe $f'(x) = 0$. Za prvi približek dobimo

$$x^{(1)} = x^{(0)} + s^{(0)} = -0.8134,$$

kjer je

$$s^{(0)} = -\frac{f'\left(x^{(0)}\right)}{f''\left(x^{(0)}\right)} = -\frac{1}{4}(e^2 - e^{-2}) = -1.8134.$$

V nadaljevanju dobimo naslednje približke.

k	$x^{(k)}$	$f(x^{(k)}) - p^\star$
1	$-8.134 \cdot 10^{-01}$	$4.338 \cdot 10^{-1}$
2	$4.094 \cdot 10^{-01}$	$2.997 \cdot 10^{-1}$
3	$-4.730 \cdot 10^{-02}$	$8.156 \cdot 10^{-2}$
4	$7.060 \cdot 10^{-05}$	$1.118 \cdot 10^{-3}$
5	$-2.346 \cdot 10^{-13}$	$2.492 \cdot 10^{-9}$

Če pa za začetni približek vzamemo $x^{(0)} = 1.1$, pa metoda divergira.

k	$x^{(k)}$	$f(x^{(k)}) - p^\star$
1	$-1.129 \cdot 10^0$	$5.120 \cdot 10^{-1}$
2	$1.234 \cdot 10^0$	$5.349 \cdot 10^{-1}$
3	$-1.695 \cdot 10^0$	$6.223 \cdot 10^{-1}$
4	$5.715 \cdot 10^0$	$1.035 \cdot 10^0$
5	$-2.302 \cdot 10^4$	$2.302 \cdot 10^4$

Naloga 33. *(VAJE) Oglej si delovanje Newtonove metode s fiksnim premikom $t = 1$ na konveksnem optimizacijskem problemu*

$$\text{minimiziraj} \quad f(x) = -\ln(x) + x,$$

če za začetni približek vzameš $x^{(0)} = 3$.

Rešitev: *f je gladka strogo konveksna funkcija z $\mathcal{D}_f = \mathbb{R}_{++}$. Velja*

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

Minimum je dosežen v točki $x^* = 1$. Če uporabimo fiksni premik $t = 1$, dobimo za prvi približek

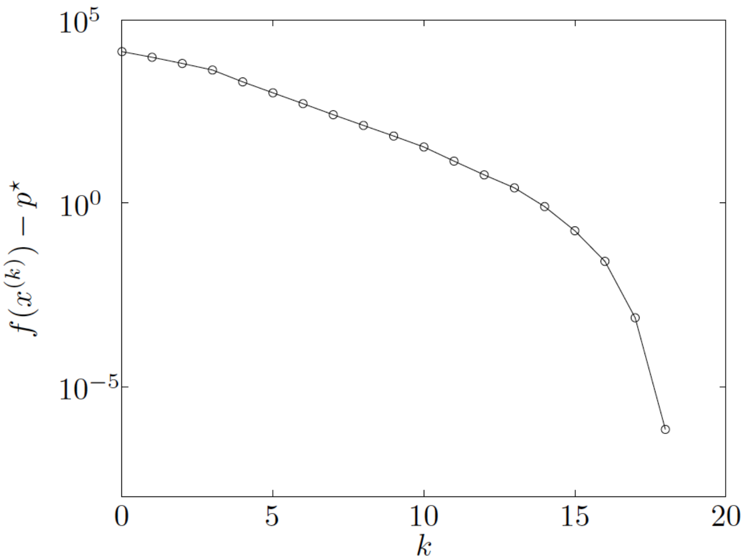
$$x^{(1)} = x^{(0)} - \frac{f'\left(x^{(0)}\right)}{f''\left(x^{(0)}\right)} = 3 - 6 = -3,$$

ki pa ne leži v domeni funkcije f.

Primer 47. *Za $n = 10000$ in $m = 100000$ je dan konveksni optimizacijski problem*

$$\text{minimiziraj} \quad -\sum_{i=1}^{n} \ln\left(1 - x_i^2\right) - \sum_{i=1}^{m} \ln\left(b_i - a_i^T x\right). \tag{4.13}$$

Za $i = 1, \ldots, m$ so vektorji $a_i \in \mathbb{R}^{10000}$ in števila b_i naključno generirani. Na Sliki 4.6 je prikazana konvergenca Newtonove metode, ki uporablja metodo diskretnea spusta s parametroma $\alpha = 0.01$ in $\beta = 0.5$. V prvi fazi opazimo linearno konvegenco. Ko pa smo dovolj blizu optimuma pa ima metoda kvadratično konvergenco.



Slika 4.6: Na sliki je prikazano obnašanje Newtonove metode, ki uporablja metodi največjega spusta in diskretnega spusta ($\alpha = 0.01, \beta = 0.5$) za optimizacijski problem (4.13). Skala na y osi je logaritemska. Čeprav gre za velik problem, Newtonova metoda potrebuje samo 18 iteracij, da skonvergira.

Poglavje 5

Konveksna optimizacija z enakostmi

V tem poglavju bomo predstavili metode za reševanje konveksnih optimizacijskih problemov z enakostmi. Podan je optimizacijski problem oblike

$$\begin{aligned} \text{minimiziraj} \quad & f(x) \\ \text{pri pogojih} \quad & Ax = b, \end{aligned} \tag{5.1}$$

kjer je $f\colon \mathcal{D}_f \subset \mathbb{R}^n \to \mathbb{R}$ dvakrat zvezno odvedljiva konveksna funkcija, $b \in \mathbb{R}^p$ in $A \in \mathbb{R}^{p\times n}$ s polnim rangom $p < n$. Predpostavimo, da optimalna rešitev x^* obstaja. Optimalno vrednost označimo s $p^* = f(x^*) = \inf_x \{f(x) \mid Ax = b\}$.

Po Posledici 3 ali če zapišemo KKT optimalnostne pogoje, vemo, da je $x^* \in \mathcal{D}_f$ optimalna rešitev problema 5.1 natanko tedaj, ko obstaja dualna spremenljivka $\nu^* \in \mathbb{R}^p$ tako da velja

$$Ax^* = b, \quad \nabla f(x^*) + A^T \nu^* = 0. \tag{5.2}$$

Reševanje problema 5.1 je tako ekvivalentno iskanju rešitve KKT sistema (5.2). To je v splošnem nelinearen sistem $n + p$ enačb za $n + p$ neznank x^* in ν^*. Samo v določenih primerih lahko sistem enačb (5.2) rešimo analitično. Najpomembnejši posebni primer je, kadar je kriterijska funkcija f kvadratična (glej Primer 19). Za $P \succeq 0$ in $q \in \mathbb{R}^n$ je dan optimizacijski problem

$$\begin{aligned} \text{minimiziraj} \quad & \frac{1}{2}x^T P x + q^T x \\ \text{pri pogojih} \quad & Ax = b. \end{aligned}$$

V tem primeru dobimo za KKT pogoje sistem linearnih enačb

$$Ax^* = b, \quad Px^* + A^T \nu^* = -q,$$

ki ga zapišemo v matrični boliki kot

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}. \tag{5.3}$$

Vemo, da je matrika tega sistema nesingularna natanko tedaj, ko velja $P + A^T A \succ 0$. Zadostni pogoj za obrnljivost je $P \succ 0$.

Če sistem (5.3) ni rešljiv, potem je optimizacijski problem navzdol neomejen ali nedopusten.

Reševanje konveksnih problemov z enakostmi.

- Vsak konveksni optimizacijski problem z enakostmi lahko pretvorimo na nevezan problem tako, da iz sistema linearnih enačb $Ax = b$ izrazimo katere od spremenljivk.

Primer 48. *Za dane konveksne funkcije f_i, $i = 1, \ldots, n$ in število b je podan optimizacijski problem*

$$\begin{array}{rl} \text{minimiziraj} & f_1(x_1) + \ldots + f_n(x_n) \\ \text{pri pogojih} & x_1 + \ldots + x_n = b. \end{array}$$

Če iz linearne enačbe elimiramo $x_n = b - x_1 - \ldots - x_{n-1}$, dobimo reduciran problem

$$\text{minimiziraj} \quad f_1(x_1) + \ldots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \ldots - x_{n-1})$$

v spremenljivkah $x_1, \ldots, x_{n-1}$.

- Drugi način, kako se lahko lotimo reševanja problema (5.1) je preko duala. Ker ni pogojev neenakosti, so pripadajoče dualne spremenljivke proste. Če je f_0 strogo koneksna, za vsak ν obstaja enolična vrednost dualne funkcije, zato je dualni program nevezan. Če je dualna funkcija dvakrat zvezno odvedljiva, lahko za maksimizacijo le-te uporabimo katero izmed metod iz Poglavja 4.

Primer 49. *Za dana pozitivna števila a_i, $i = 1, \ldots, n$, je podan konveksni optimizacijski problem*

$$\begin{array}{rl} \text{minimiziraj} & \displaystyle\sum_{i=1}^{n} a_i x_i^2 \\ \text{pri pogojih} & \displaystyle\sum_{i=1}^{n} x_i = 1. \end{array}$$

Iz Lagrangeove funkcije

$$L(x, \nu) = \sum_{i=1}^{n} a_i x_i^2 + \nu \left(1 - \sum_{i=1}^{n} x_i\right) = \nu + \sum_{i=1}^{n} \left(a_i x_i^2 - \nu x_i\right)$$

dobimo dualno funkcijo

$$g(\nu) = \inf_x \left(\nu + \sum_{i=1}^{n} \left(a_i x_i^2 - \nu x_i\right)\right) = \nu + \sum_{i=1}^{n} \inf_{x_i} \left(a_i x_i^2 - \nu x_i\right) = \nu - \frac{1}{4}\left(\sum_{i=1}^{n} \frac{1}{a_i}\right)\nu^2.$$

Pri tem smo uporabili dejstvo, da je minimum funkcije $f(x_i) = a_i x_i^2 - \nu x_i$ dosežen pri

$$x_i = \frac{\nu}{2a_i}. \tag{5.4}$$

Dualni problem je potem

$$\text{maksimiziraj} \quad \nu - \frac{1}{4}\left(\sum_{i=1}^{n} \frac{1}{a_i}\right)\nu^2,$$

kjer je $\nu \in \mathbb{R}$. Torej iščemo maksimum konkavne funkcije ene spremenljivke. Ker je dualna funkcija kvadratna, lahko poiščemo optimalno rešitev

$$\nu^* = \frac{2}{\sum_{i=1}^{n} \frac{1}{a_i}}$$

analitično. Optimalna vrednost duala je torej enaka

$$\nu^* - \frac{1}{4}\left(\sum_{i=1}^{n} \frac{1}{a_i}\right)(\nu^*)^2 = \frac{1}{\sum_{i=1}^{n} \frac{1}{a_i}}.$$

Optimalno rešitev primarnega problema potem izračunamo iz zveze (5.4)

$$x_i^* = \frac{1}{a_i}\frac{1}{\sum_{i=1}^{n} \frac{1}{a_i}}.$$

Naloga 34. *(VAJE) ZA dane podatke $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$ je podan konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & -\sum_{i=1}^{n} \ln x_i \\ \text{pri pogojih} \quad & Ax = b. \end{aligned}$$

Pri tem je $x > 0$ implicitna omejitev. Kako bi s pomočjo duala poiskal optimalno rešitev primarnega problema?

Rešitev: *Iz Lagrangeove funkcije*

$$L(x,\nu) = -\sum_{i=1}^{n} \ln x_i + \nu^T(Ax - b) = -b^T\nu + \left(A^T\nu\right)^T x - \sum_{i=1}^{n} \ln x_i$$

dobimo dualno funkcijo

$$g(\nu) = -b^T\nu + \inf_x \left(\left(A^T\nu\right)^T x - \sum_{i=1}^{n} \ln x_i\right) = -b^T\nu + n + \sum_{i=1}^{n} \ln\left(A^T\nu\right)_i.$$

Pri tem smo upoštevali, da je minimum funkcije $x \mapsto \left(A^T\nu\right)^T x - \sum_{i=1}^{n} \ln x_i$ dosežen pri $x = (x_1, \ldots, x_n)$ za katerega velja

$$\left(\frac{1}{x_1}, \ldots, \frac{1}{x_n}\right) = A^T\nu.$$

Dualni problem je potem

$$\text{maksimizíraj} \quad -b^T\nu + n + \sum_{i=1}^{n} \ln\left(A^T\nu\right)_i,$$

kjer je $A^T\nu > 0$ implicitna omejitev. Uporabimo katero od metod iz nevezane optimizacije, da poiščemo optimalno rešitev ν^* dualnega problema. Potem je

$$x_i = \frac{1}{\left(A^T\nu^*\right)_i}$$

optimalna rešitev primarnega problema.

- Najpogostejši pristop za reševanje problema (5.1) pa je posplošitev Newtonove metode iz nevezane optimizacije na probleme z linearnimi omejitvami. Ta metoda je še posebej primerna, ko lahko izkoristimo strukturo prvotnega problema, npr. razpršenost.

5.1 Newtonova metoda z linearnimi omejitvami

Radi bi posplošili Newtonovo metodo na konveksne optimizacijske probleme oblike

$$\begin{aligned} \text{minimiziraj} \quad & f(x) \\ \text{pri pogojih} \quad & Ax = b. \end{aligned} \tag{5.5}$$

Metodo lahko izpeljemo preko Taylorjeve vrste ali z linearizacijo pogojev optimalnosti. Denimo, da je x tekoči približek. Iščemo Newtonov premik Δ_{nt}.

- Taylorjev polinom drugega reda $\widehat{f}$ za funkcijo f v točki x je

  $$\widehat{f}(x + \Delta x) = f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x.$$

  Iščemo tako smer Δx, da bo nov približek $x + \Delta x$ dopusten. Torej mora veljati $A(x + \Delta x) = b$ oziroma $A\Delta x = 0$. Zato je smer Δx rešitev problema

  $$\begin{aligned} \text{minimiziraj} \quad & f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x \\ \text{pri pogojih} \quad & A\Delta x = 0. \end{aligned}$$

  Zapišimo optimalnostne pogoje. Premik Δx je optimalna rešitev zgornjega prirejenega problema natanko tedaj, ko obstaja w, da velja

  $$A\Delta x = 0, \quad \nabla f(x) + \nabla^2 f(x)\Delta x + A^T w = 0.$$

- Drugi način, kako izpeljemo metodo je preko linearizacije pogojev optimalnosti $A(x+\Delta x) = b$ in $\nabla f(x + \Delta x) + A^T w = 0$. Zaradi aproksimacije

$$\nabla f(x + \Delta x) + A^T w \approx \nabla f(x) + \nabla^2 f(x)\Delta x + A^T w = 0,$$

  iščemo torej tak popravek Δx, da bo veljalo

$$A\Delta x = 0, \quad \nabla f(x) + \nabla^2 f(x)\Delta x + A^T w = 0.$$

V obeh primerih dobimo, da Δx_{nt} in w rešita sistem linearnih enačb

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}. \tag{5.6}$$

Za zaustavitveni pogoj bomo spet uprabljali λ (glej (4.12)), ki je definirana kot

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

V splošnem ta vrednost ni več enaka $\sqrt{\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)}$.

Podobno kot v primeru nevezane optimizacije lahko izpeljemo, da velja

$$f(x) - \inf \left\{ \widehat{f}(x + \Delta x) \mid A\Delta x = 0 \right\} = \frac{1}{2}\lambda^2(x).$$

Zaradi pogojev optimalnosti (5.6) namreč velja $\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt} = -\nabla f(x)^T \Delta x_{nt}$.

Od tod dobimo *Newtonovo metodo z linearnimi omejitvami.*

Algorithm 5.1.1: Newtonova metoda z linearnimi omejitvami.

---

vhodni podatki: začetni približek $x^{(0)} \in \mathcal{D}_f$ za katerega velja $Ax^{(0)} = b$, toleranca $\varepsilon > 0$

for $k = 0, 1, \ldots,$ **do**
- Izračunaj Newtonovo smer $\Delta x_{nt}^{(k)}$ in $\lambda^2\left(x^{(k)}\right)$.
- **if** $\lambda^2\left(x^{(k)}\right)/2 \le \varepsilon$ **then**
  - končaj
- Določi premik $t^{(k)} > 0$ (največji spust ali diskretni spust).
- Izračunaj novi približek $x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x_{nt}^{(k)}$.

Ta metoda je spustna metoda, saj so rešitve v vsaki iteraciji dopustne in velja $f(x^{(k+1)} < f(x^{(k)})$, za vsak k, razen če je $x^{(k)}$ že optimum. V primerih ko je matrika sistema (5.6) posebne strukture, potem uporabimo bločno eliminacijo, da hitreje izračunamo naslednji približek.

Primer 50. *Za dane podatke in dvakrat zvezno odvedljive ter strogo konveksne funkcije ϕ_i, $i = 1, \ldots, n$ je podan konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & \sum_{i=1}^{n} \phi_i(x_i) \\ \text{pri pogojih} \quad & Ax = b. \end{aligned}$$

Pogoji optimalnosti dajo KKT sistem oblike

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ w \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix},$$

kjer je $g = \nabla f(x)$ in $H = diag\,(\phi_1''(x_1), \ldots, \phi_n''(x_n))$ diagonalna matrika s pozitivnimi elementi na diagonali. Do Newtonove smeri in dualne spremenljivke w pridemo preko reševanja manjšega sistema. Če iz prve enačbe eliminiramo Δx_{nt} in vstavimo v drugo enačbo, dobimo sistem za w

$$AH^{-1}A^T w = -AH^{-1}g.$$

Nato iz enačbe $H\Delta x_{nt} = -g - A^T w$ izračunamo Δx_{nt}.

Naloga 35. *(VAJE) Poišči aproksimacijo prvega reda za funkcijo $f\colon \mathcal{S}_{++}^n \to \mathcal{S}_{++}^n$ definirano s predpisom $f(X) = X^{-1}$.*

Rešitev: *Za $Z \in \mathcal{S}_{++}^n$, ki je blizu $X \in \mathcal{S}_{++}^n$ in $\Delta X = Z - X$ dobimo*

$$\begin{aligned} Z^{-1} &= (X + \Delta X)^{-1} \\ &= \left(X^{\frac{1}{2}}(I + X^{-\frac{1}{2}}\Delta X X^{-\frac{1}{2}})X^{\frac{1}{2}}\right)^{-1} \\ &= X^{-\frac{1}{2}}(I + X^{-\frac{1}{2}}\Delta X X^{-\frac{1}{2}})^{-1}X^{-\frac{1}{2}} \\ &\approx X^{-\frac{1}{2}}(I - X^{-\frac{1}{2}}\Delta X X^{-\frac{1}{2}})X^{-\frac{1}{2}} \\ &= X^{-1} - X^{-1}\Delta X X^{-1}. \end{aligned}$$

Pri tem smo upoštevali, da za male A velja $(I + A)^{-1} \approx I - A$. Torej za male ΔX velja

$$(X + \Delta X)^{-1} \approx X^{-1} - X^{-1}\Delta X X^{-1}. \tag{5.7}$$

Primer 51. *Za dane podatke $A_i \in \mathcal{S}^n$, $b_i \in \mathbb{R}$, $i = 1, \ldots, p$ je za optimizacijsko spremenljvko $X \in \mathcal{S}^n$ podan konveksni optimizacijski problem*

$$\begin{aligned} \text{minimiziraj} \quad & -\ln\det X \\ \text{pri pogojih} \quad & \text{sled}(A_i X) = b_i, \quad i = 1, \ldots, p. \end{aligned}$$

Predpostavimo, da so matrike A_i linearno neodvisne. Poleg implicitne omejitve $X^* \succ 0$ so optimalnostni pogoji enaki

$$-(X^*)^{-1} + \sum_{j=1}^{p} \nu_j^* A_j = 0, \quad \text{sled}(A_i X^*) = b_i, \quad i = 1, \ldots, p.$$

Pri tem smo upoštevali, da je $\nabla \ln\det X = X^{-1}$ in da je adjungiran operator $\mathcal{A}^T$ operatorja $\mathcal{A}(X) = (\text{sled}(A_1 X^*), \ldots, \text{sled}(A_p X^*))^T$ enak $\mathcal{A}^T(\nu) = \sum_{j=1}^{p} \nu_j A_j$. Naj bo X trenutni približek za optimalno rešitev. Po Newtonovi metodi Iščemo tak ΔX, ki bo izpolnil linearizirane optimalnostne pogoje

$$X^{-1}\Delta X X^{-1} + \sum_{j=1}^{p} w_j A_j = X^{-1}, \quad \text{sled}(A_i \Delta X) = 0, \quad i = 1, \ldots, p.$$

Pri tem smo uporabili (5.7). *To je sistem* $\frac{n(n+1)}{2} + p$ *enačb za prav toliko neznank* ΔX *in* w. *Iz prve enačbe lahko elimiramo spremenljivko* ΔX*:*

$$\Delta X = X - \sum_{j=1}^{p} w_j X A_j X. \tag{5.8}$$

Če to vstavimo v drugo enačbo, dobimo

$$\mathrm{sled}(A_i \Delta X) = \mathrm{sled}(A_i X) - \sum_{j=1}^{p} w_j \,\mathrm{sled}(A_i X A_j X) = 0, \quad i = 1, \ldots, p.$$

To je sistem p *linearnih enačb oblike* $Cw = d$ *za dualno spremenljivko* w. *Pri tem je* $c_{ij} = \mathrm{sled}(A_i X A_j X)$ *in* $d_i = \mathrm{sled}(A_i X)$. *Matrika* C *je simetrična in pozitivno definitna, saj velja*

$$c_{ij} = \mathrm{sled}(A_i X A_j X) = \mathrm{sled}(A_j X A_i X) = c_{ji}$$

in

$$\begin{aligned} x^T C x &= \sum_{i,j=1}^{n} c_{ij} x_i x_j = \sum_{i,j=1}^{n} x_i x_j \,\mathrm{sled}(A_i X A_j X) = \sum_{i,j=1}^{n} \mathrm{sled}(x_i A_i X x_j A_j X) \\ &= \left\langle \sum_{i=1}^{n} x_i A_i X, \sum_{j=1}^{n} x_j A_j X \right\rangle = \left\| \sum_{i=1}^{n} x_i A_i X \right\|_F^2 > 0 \end{aligned}$$

za poljuben neničeln vektor x. *Namreč matrika* $\sum_{i=1}^{n} x_i A_i X$ *je ničelna natanko tedaj, ko je matrika* $\sum_{i=1}^{n} x_i A_i$ *ničelna. To se pa zaradi linearne neodvisnosti matrik* A_i *zgodi le, če je vektor* x *enak 0. Ker je matrika* C *pozitivno definitna, lahko za reševanje sistem* $Cx = d$ *uporabimo razcep Choleskega. Za izračun premika* ΔX *uporabimo nato* (5.8). *Pri izbiri premika* t *je na koncu treba paziti, da je matrika* $X + t\Delta X$ *pozitivno definitna. To najlažej preverjamo z razcepom Choleskega.*

5.2 Newtonova metoda z eliminacijo enačb

Ker je matrika A polnega vrstičnega ranga, lahko množico rešitev $\{x \in \mathbb{R}^n \mid Ax = b\}$ lahko opišemo tudi kot $\{Fz + \hat{x} \mid z \in \mathbb{R}^{n-p}\}$, kjer je $\hat{x}$ poljubna dopustna rešitev za $Ax = b$ in $F \in \mathbb{R}^{n\times(n-p)}$ poljubna matrika, katere stolpci napenjajo jedro A.

Problem (5.1) lahko torej zapišemo kot

$$\mathrm{minimiziraj} \quad \tilde{f}(z) \;=\; f(Fz + \hat{x}) \tag{5.9}$$

Iz optimalne rešitve z^* lahko dobimo optimlano repitev osnovnega problema (5.1) kot $x^* = Fz^* + \hat{x}$. Iz x^* lahko izračunamo tudi optimalno dualno spremenljivko kot

$$\nu^* = -(AA^T)^{-1})A\nabla f(x^*).$$

Da je ν^* res optamlna dualna spremenljiva preverimo tako, da pokažemo drug pogoj iz (5.2), torej:

$$\nabla f(x^*) + A^T \nu^* = \nabla f(x^*) - A^T(AA^T)^{-1})A\nabla f(x^*) = 0.$$

Uporabimo, da je $AF = 0$ in $\nabla \tilde{f}(z^*) = F^T \nabla f(x^*) = 0$. Torej velja

$$\begin{bmatrix} F^T \\ A \end{bmatrix} \Big(\nabla f(x^*) - A^T (AA^T)^{-1}) A \nabla f(x^*) \Big) = 0.$$

Ker ima matrika

$$\begin{bmatrix} F^T \\ A \end{bmatrix}$$

poln vrstični rang, je torej obrnljiva, iz te enačbe sledi $\nabla f(x^*) - A^T (AA^T)^{-1}) A \nabla f(x^*) = 0$.

5.3 Konvergenca Newtonove metode

Konvergenco Newtonove metode (Algoritem 5.1.1) pokažemo tako, da pokažemo konvergenco za primer brez omejitev (5.9). Za dokaz konvergence ptrebujemo tri predpostavke

1. Množica $S = \{x \mid x \in \mathbf{dom} f,\ f(x) \leq f(x^{(0)},\ Ax = b\}$ je zaprta, kjer $x^{(0)} \in \mathbf{dom} f$ zadošča $Ax^{(0)} = b$.

2. Na množici S velja $\nabla^2 f(x) \preceq MI$ in

$$\left\| \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}^{-1} \right\|_2 \leq K$$

3. $\nabla^2 f$ je Lipschitzeva s konstanto L, torej $\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L \|x - y\|_2$, za vsaka $x, y \in S$.

Lemma 5.1. *Iz predpostavke 2 sledi, da je $\nabla^2 \tilde{f}(x) \succeq mI$ za*

$$m = \frac{\sigma_{\min}(F)^2}{K^2 M}$$

ki je pozitivno število.

Dokaz. Najprej utemeljimo, a je $\sigma_{\min}(F)^2$ poziitvno število. To sledi iz tega, da je F polnega stolpičnega ranga, torej je $\sigma_{\min}(F)^2 = \lambda_{\min}(F^T F$.

Glavnino leme pokažemo s protislovjem. Predposatavimo, da je $\nabla^2 \tilde{f}(x) = F^T H F \not\succeq mI$, kjer je $H = \nabla^2 f(x)$. Potem lahko najdemo enotski vekor u, da velja $u^T F^T H F u < m$, torej $\|H^{1/2} F u\|_2 < \sqrt{m}$. Ker je $AF = 0$, velja

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} Fu \\ 0 \end{bmatrix} = \begin{bmatrix} HFu \\ 0 \end{bmatrix}.$$

Ker je $\|Ux\|_2 \leq \|U\|_2 \|x\|_2$, za vsako matriko U in vsak vektor x in ker po predpostavki 2 obstaja inverz matrike

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}$$

sledi

$$\begin{bmatrix} Fu \\ 0 \end{bmatrix} = \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} HFu \\ 0 \end{bmatrix}.$$

Torej

$$\| \begin{bmatrix} Fu \\ 0 \end{bmatrix} \|_2 \leq \| \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}^{-1} \|_2 \| \begin{bmatrix} HFu \\ 0 \end{bmatrix} \|_2,$$

oziroma

$$\frac{\|Fu\|_2}{\|HFu\|_2} \leq \| \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}^{-1} \|_2.$$

Ker je $\|Fu\|_2 \geq \sigma_{\min}(F)$ in $\|HFu\|_2 \leq \|H^{1/2}\|_2 \|H^{1/2}Fu\|_2 < M^{1/2}m^{1/2}$, sledi

$$\| \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}^{-1} \|_2 > \frac{\|Fu\|_2}{\|HFu\|_2} M^{1/2}m^{1/2} = K,$$

kar je protislovje.

□

Posledica 11. *Naj bodo za optimizacijski problem* (5.1) *izpolnjene zgornje predpostavke 1-3. Tedaj velja, da Newtonova metoda za ta problem (Algoritem 5.1.1) konvergira kvadratično.*

Dokaz. Iz zgornjih predpostavk sledi, da optimizacijski problem, dobljen z eliminacijo (5.9), zadošča pogojem Izreka 9 ($\tilde{f}$ je dvakrat zvezno odvedljiva, strogo konveksna in Lipschitzova. . Torej Newtonova metoda, aplicirana na problem (5.9), ima kvadratično konveregenco. □

Poglavje 6

Metode notranjih točk

V tem poglavju, ki je zgoščena in skrajšana verzija poglavja 11 iz [BBV04], bomo predstavili družino metod, imenovano metode notranjih točk, za reševanje konveksnih optimizacijskih problemov z neenakostmi in z linearnimi enačbami:

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) \\ \text{pri pogojih} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{aligned} \tag{6.1}$$

kjer so $f_i : \mathcal{D}_i \subset \mathbb{R}^n \to \mathbb{R}$ konveksne dvakrat zvezno odvedljive funkcije, $A \in \mathbb{R}^{p \times n}$ s polnim vrstičnim rangom ($\operatorname{rang} A = p < n$). Označimo

$$\mathcal{D} := \bigcap_{i=0}^{m} \mathcal{D}_i.$$

Nadalje v tem poglavju privzamemo naslednje.

Predpostavka 1. *(a) Optimizacijski problem* (6.1) *ima optimalno rešitev: obstaja $x^* \in \mathcal{D}$, ki zadošča omejitvam in pri katerem doseže f_0 minimum, ki ga označimo s $p^* = f_0(x^*)$.*

(b) Obstaja strogo dopustna rešitev za (6.1)*, torej tak $x \in \operatorname{rel\,int} \mathcal{D}$, da velja: $f_i(x) < 0, \forall i = 1, \ldots, m$ in $Ax = b$*

Zaradi predpostavke 1 (b) velja Slaterjev pogoj, torej po izrekih 5 in 6 obstaja optimalna rešitev za dualni problem $\lambda^* \in \mathbb{R}^m, \nu^* \in \mathbb{R}^p$, da so izpolnjeni KKT pogoji:

$$\begin{aligned} Ax^* = b, \; f_i(x^*) \leq 0, \quad & i = 1, \ldots, m \\ \lambda_i^* \geq 0, \quad & i = 1, \ldots, m \\ \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + A^T \nu^* = 0, & \\ \lambda_i^* f_i(x^*) = 0, \quad & i = 1, \ldots, m \end{aligned} \tag{6.2}$$

Metode notanjih točk rešijo problem (6.1) (oz. KKT pogoje (6.2)) z uporabo Newtonove metode, s katero rešijo zaporedje KKT pogojev, dobljenih iz (6.2) z manjšo prilagoditvijo. Dve najbolj znani družini metod notranjih točk so t.i. *pregradne metode notranjih točk* in *primarno-dualne metode notranjih točk.*

Metode notranjih točk lahko razumemo kot naslednji nivo reševanje problemov konveksne optimizacije. Konveksni kvadratični problemi z linearnimi omejitvami so najpreprostejši problemi, ker imamo tam eskplicitno formulacijo rešitve. Newtonova metoda je naslednji nivo reševanja, saj omogoča reševanje konveksnih problemov z dvakrat zvezno odvedljivo kriterijsko funkcijo in z linearnimi omejitvami, kar naredi s kreiranjem zaporedja konveksnih kvadratičnih problemov z linearnimi omejitvami. Mnogo optimizacijskih problemov že ima obliko (6.1), kot npr. linearni, konveksni kvadratični problemi z linearnmi in/ali konveksnimi kvadratičnimi omejitvami,..., nekatere pa lahko pretvorimo v to obliko.

Primer 52 (Maksimizacija entropije)**.** *Problem maksimizacije entropije lahko zapišemo kot*

$$\begin{aligned} \text{minimiziraj} \quad & \sum_{i=1}^{n} x_i \ln(x_i) \\ \text{pri pogojih} \quad & Ax = b, \\ & Cx \leq d \end{aligned} \tag{6.3}$$

V tem primeru je domena $\mathcal{D} = \mathbb{R}^n_{++}$.

V nadaljevanju bomo uvedli logaritemsko pregradno funkcijo (angl. logarithmic barier function). Najprej bomo pretvorili problem (6.1) v problem, ki bo imel samo enačbe, saj na takem problemu lahko uporabimo Newtonovo metodo:

$$\begin{aligned} \text{minimiziraj} \quad & f_0(x) + \sum_{i=1}^{n} I_-(f_i(x)) \\ \text{pri pogojih} \quad & Ax = b, \end{aligned} \tag{6.4}$$

kjer je I_- indikatorska funkcija za nepozitivna realna števila, definirana kot:

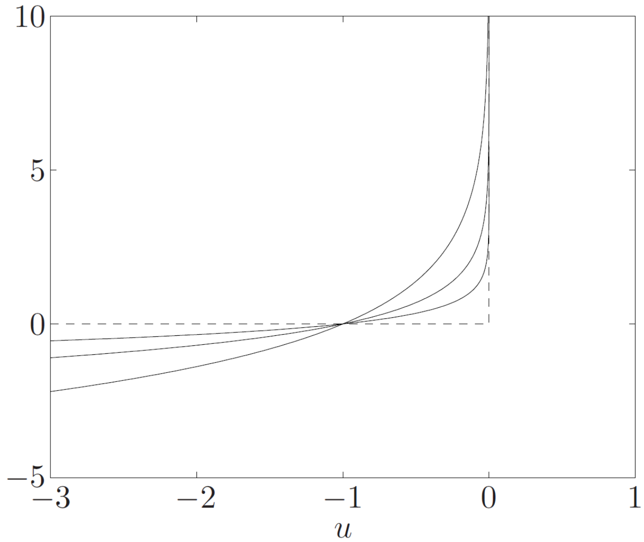
$$I_-(u) = \begin{cases} 0, & u \leq 0, \\ \infty & u > 0. \end{cases}$$

Funkcija $I_-(u)$ je konveksna in nepadajoča, ni pa diferenciabilna v točki $u = 0$. Problem (6.4) med omejitvami nima neenačb, vendar njegova kriterijska funkcija v splošnem ni diferenciabilna, kar pomeni, da ne moremo uporabiti Newtonove metode, zato najprej indikatorsko funkcijo aproksimiramo z

$$\widehat{I_-}(u) = -(1/t)\ln(-u), \quad t > 0,$$

katere domena je $-\mathbb{R}_{++}$. S parametrom t kontroliramo natančnost aproksimacije, saj večja vrednost t pomeni večjo natančnost aproksimacije. Funkcija $\widehat{I_-}(u)$ je tudi konveksna in nepadajoča, kot $I_-(u)$, za razliko od $I_-(u)$ pa je šepoljubnokrat odvedljiva in zaprta[1], saj narašča proti ∞, ko u narašča proti 0.

[1]Funkcija f je zaprta, če je za vsak $\alpha \in \mathbb{R}$ množica $\{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ zaprta, kar je ekvivalentno temu, da je epigraf funkcije f, definiran kot $\operatorname{epi} f = \{(x,t) \mid f(x) \leq t\}$, zaprta množica.



Slika 6.1: Prikaz natančnosti aprokcimacije funkcije $I_(u)$ s funkcijami $\widehat{I}_(u)$. Slika je vzeta iz [BBV04].

Če nadomestimo $I_i(u)$ z $\widehat{I}_-(u)$ v (6.4), dobimo

$$\begin{aligned} &\text{minimiziraj} \quad f_0(x) + \sum_{i=1}^{n} -(1/t)\ln(-f_i(x)) \\ &\text{pri pogojih} \quad Ax = b, \end{aligned} \tag{6.5}$$

V tem reformuliranem problemu je kriterijska funkcija konveksna in dvakrat zvezno odvedljiva. Če je izpolnjena zahteva po zaprtosti nove kriterijske funkcije, lahko uporabimo Newtonovo metodo. Funkcijo

$$\phi(x) = -\sum_{i=1}^{m} \ln(-f_i(x)) \tag{6.6}$$

imenujemo logaritemska pregradna funkcija za problem (6.1). Domena ϕ je $\mathbf{dom}\ \phi = \{x \in \mathbb{R}^n \mid f_i(x) < 0, i = 1, \ldots, m\} \subseteq \cap_{i=1}^{m} \mathcal{D}_i$, torej množica vseh točk, kjer so neenakosti iz (6.1) strogo izpolnjene. Zaradi predpostavke o strogi dopustnosti je $\mathbf{dom}\ \phi \neq \emptyset$. Nadalje vidimo, da ko za katerikoli i gre $f_i(x) \to 0$, narašča $\phi(x)$ proti ∞.

Za nadaljnje potrebe navedimo še gradient in Hessejevo metriko:

$$\begin{aligned} \nabla\phi(x) &= \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x), \\ \nabla^2\phi(x) &= \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x) \end{aligned}$$

Oglejmo si ponovno reformuliran problem (6.5) in ga zapišimo še malce drugače, tako da kriterijsko funkcijo pomnožimo s t, kar nam da naslednji ekvivalenten optimizacijski problem:

$$\begin{aligned} \text{minimiziraj} \quad & tf_0(x) + \phi(x) \\ \text{pri pogojih} \quad & Ax = b. \end{aligned} \tag{6.7}$$

Najprej pokažimo, da lahko ta modificiran problem rešimo z Newtonovo metodo.

Lema 11. *Naj bo x_0 strogo dopustna rešitev za* (6.1). *Če je za funkcijo $tf_0(x) + \phi(x)$ množica t.i. podnivojskih točk $\{x \mid tf_0(x) + \phi(x) < tf_0(x_0) + \phi(x_0)\}$ zaprta za vsak t, če je inverz prirejene KKT matrike obrnljiv in če Hessejeva matrika zadošča Lipschitzevemu pogoju, potem ima optimizacijski problem* (6.7) *enolično rešitev, ki jo lahko poljubno natančno izračunamo z Newtonovo metodo.*

Dokaz. To sledi iz Posledice 11. □

Za vsak $t > 0$ označimo z $x^*(t)$ optimalno rešitev (6.7). Središčna pot, prirejena problemu (6.1), je množica rešitev $\{x^*(t) \mid t > 0\}$. Točkam s središčne poti rečemo tudi središčne točke, ki so enolično določene s potrebnimi in zadostnimi pogoji za optimalnost:

1. $x^*(t)$ je strogo dopustna rešitev, torej $Ax^*(t) = b$ in $f_i(x^*(t)) < 0$, za vsak i.
2. obstaja $\hat{\nu} \in \mathbb{R}^p$, da velja:

$$\begin{aligned} 0 &= t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\nu} \\ &= t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t)}\nabla f_i(x^*(t)) + A^T\hat{\nu}. \end{aligned} \tag{6.8}$$

Primer 53 (Linearni program z neenakostmi)**.** *Za linearni problem z neenakostmi*

$$\begin{aligned} \text{minimiziraj} \quad & c^T x \\ \text{pri pogojih} \quad & Ax \le b. \end{aligned} \tag{6.9}$$

je logaritemska pregradna funkcija enaka

$$\phi(x) = -\sum_{i=1}^{m} \ln\bigl(b_i - a_i^T x\bigr)$$

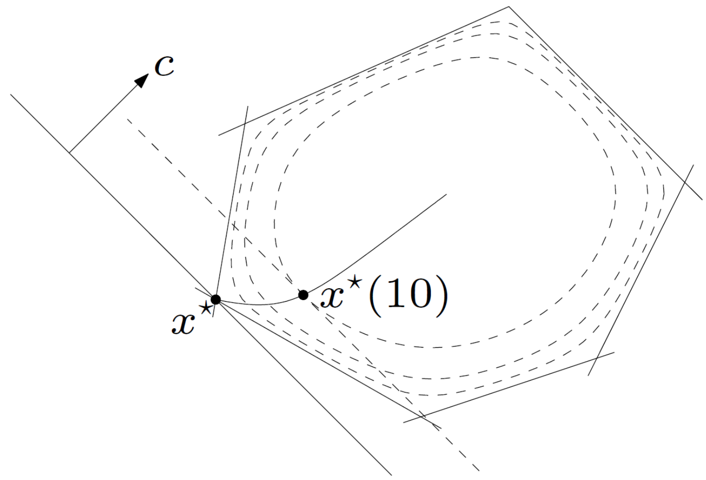
z domeno $\mathbf{dom}\phi = \{x \mid Ax < b\}$, torej množico vseh točk, pri katerih so vse neenakosti $Ax \le b$ strogo izpolnjene. V definiciji $\phi(x)$ so a_i^T vrstice matrike A. Gradient in Hessejeva matrika za ϕ sta

$$\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i, \quad \nabla^2\phi(x) = A^T\mathbf{Diag}(d)^2 A,$$

kjer so komponente vektorja $d \in \mathbb{R}^m$ dane z $d_i = 1/(b_i - a_i^T x)$. Ker je $x \in \mathbf{dom}\phi$ strogo dopusten, velja $d_i > 0$, torej je Hessejeva matrika nesingularna natanko tedaj, ko je A matrika ranga n. Središčni pogoj (6.8) *se v tem primeru posenostavi v*

$$tc + \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = tc + A^T d = tc + \nabla\phi(x) = 0. \tag{6.10}$$

Središčni pooj ima lepo geometrijsko interpretacijo: v vsaki točki $x^*(t)$ s središčne poti mora biti gradient $\nabla\phi(x^*(t))$, ki je pravokoten na nivojnico ϕ skozi $x^*(t)$, vzporeden z $-c$. Z drugimi besedami, hiperravnina $c^T x = c^T x^*(t)$ mora biti tangentna na nivojnico ϕ skozi $x^*(t)$, glej sliko 6.2



Slika 6.2: Središčna pot za primer linearnega programa z neenakostmi. Črtkane krivulje prikazujejo tri nivojnice za logaritemsko pregradno funkcijo ϕ. Središčna pot konvergira k optimalni točki, ko gre t proti neskončnosti. Središčni pogoj 6.10 lahko preverimo geometrijsko: premice $c^T x = c^T x^*(t)$ so res tangentne na nivojnice ϕ. Slika je vzeta iz [BBV04].

6.1 Izračun dobrih dualnih rešitev, povezanih s središčno potjo

V nadaljevanju pokažimo, kako lahko za vsako točko na središčni poti, ki je seveda (strogo) dopustna za primarni problem (6.1), najdemo še dopustno dualno točko. Najprej za poljuben $t > 0$ definirajmo

$$\lambda_i^*(t) = -\frac{1}{tf_i(x^*(t))},\ i = 1, \ldots, m,\ \ \nu^*(t) = \hat{\nu}/t. \tag{6.11}$$

Pokazali bomo, da je par $\lambda^*(t), \nu^*(t)$ dopusten za dualni problem. Očitno je $\lambda_i^* > 0,\ \forall i$, saj je $f_i(x^*(t)) < 0$, za vsak i. Optimalni pogoj (6.8) lahko preoblikujemo v

$$\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*(t)) + A^T \nu^*(t) = 0, \tag{6.12}$$

Upoštevajoč, da je Lagrangeova funkcija od (6.1) enaka

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^T (Ax - b),$$

sledi, da $x^*(t)$ minimizira Lagrangeovo funkcijo pri $\lambda_i = \lambda^*(t)$ in $\nu = \nu^*(t)$, kar pomeni, da sta $\lambda^*(t), \nu^*(t)$ dualni dopustni par za (6.1), torej je vrednost dualne funkcije $g(\lambda^*(t), \nu^*(t))$ končna in

$$g(\lambda^*(t), \nu^*(t)) = f_0(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(x^*(t)) + \nu^*(t)^T (Ax^*(t) - b) = f_0(x^*(t) - m/t.$$

Torej je dualni razmik primarno dualnega para $x^*(t)$ in $\lambda^*(t), \nu^*(t)$ enak m/t, oz., z drugimi besedami:

$$f_0(x^*(t)) - p^* \leq m/t,$$

torej je $x^*(t)$ največ za m/t slabši od optimuma. Torej, ko $t \to \infty$, $x^*(t)$ konvergira k optimlani točki originalnega problema (6.1).

Primer 54. *Obravnavajmo dualni linearni program za linearni program iz Primera 53, ki je oblike*

$$\begin{aligned} \text{maksimiziraj} \quad & -b^T\lambda \\ \text{pri pogojih} \quad & A^T\lambda + c = 0,\ \lambda \geq 0. \end{aligned}$$

Iz pogojev optimalnosti (6.10) *sledi, da je*

$$\lambda_i^*(t) = \frac{1}{t(b_i - a_i^T x^*(t))},\ i = 1, \ldots, m$$

dualno dopustna rešitev, pri kateri je vrednost dualne kriterijske funkcije enaka

$$-b^T\lambda^*(t) = c^T x^*(t) + (Ax^*(t) - b)^T\lambda^*(t) = c^T x^*(t) - m/t.$$

6.2 Interpretacija središčne poti preko KKT pogojev

Pogoja 1–2, ki določata središčno pot, lhko interpretiramo kot zvezno deformacijo KKT pogojev (6.2), saj lahko pogoja 1–2 zapišemo na naslednji način: točka x je enaka optimalni točki $x^*(t)$ (torej točki na na središčni poti) natanko tedaj, ko obstajata λ, ν, da velja

$$\begin{aligned} Ax = b,\ f_i(x) \leq 0, \quad & i = 1, \ldots, m \\ \lambda_i \geq 0, \quad & i = 1, \ldots, m \\ \nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T\nu = 0, & \\ -\lambda_i f_i(x) = 1/t, \quad & i = 1, \ldots, m. \end{aligned} \tag{6.13}$$

Zgornji pogoji se razlikujejo od pogojev (6.2) le v zadnji enačbi. Pogoju komplementarnosti $-\lambda_i f_i(x) = 0$ spremenimo desno stran in dobimo zadnji pogoj v (6.13)

$$-\lambda_i f_i(x) = 1/t.$$

Sprememba je za velike t zelo majhna, kar pomeni, da za velike t optimalne rešitve $x^*(t), \lambda^*(t)$ in $\nu^*(t)$ skoraj zadoščajo KKT pogojem (6.13).

6.3 Pregradna metoda notranjih točk

Zgoraj smo pokazali, da je $x^*(t)$, ki leži na središčni poti, m/t podoptimalna, kar lahko potrdimo s parom dualnih dopustnih rešitev $\lambda^*(t), \nu^*(t)$. Iz tega naravno sledi ideja, da bi rešili optimizacijski problem (6.1) do natančnosti ε na naslednji način: vzamemo $t = m/\varepsilon$ in rešimo problem

$$\begin{aligned} \text{minimiziraj} \quad & (m/\varepsilon) f_0(x) + \phi(x) \\ \text{pri pogojih} \quad & Ax = b. \end{aligned}$$

z uporabo Newtonove metode do želene natančnosti, pri čemer lahko enačbe odpravimo in uporabimo Newtonovo metodo za problem brez omejitev, ali pa jih ohranimo in uporabimo Newtonovo metodo za optimizacijske probleme z linearnimi enačbami. Oboje smo opisali poglavjih 5 in **??**. Ta metoda deluje vredu, če imamo dobre začetne točke, če zahtevana natančnost ε ni zelo majhna in če je dimenzija problema (m, n) majhna. POsledično se ta metoda v praksi ne uporablja, ampak se uporablja razširitev te metode, ki jo opišemo v nadaljevanju.

Razširitev zgoraj opisane metode je v resnici zelo preprosta, a vseeno bistveno doprinese k učinkovitosti metode. Njeno bstvo je v reševanju zaporedja optimizacijskih problemov (6.7) pri vedno večjih vrednostih t, pri tem pa v vsakem koraku vzamemo kot začetno točko rešitev iz prejšnjega koraka. Končamo, ko je $t > m/\varepsilon$. Tej metodi dejansko rečemo *pregradna metoda* ali tudi *metoda, ki sledi središčni poti* in je predstavljena kot algoritem 6.3.1.

Algorithm 6.3.1: Pregradna metoda za (6.1)

---

vhod : strogo dopusten x, $t := t^{(0)}$, $\mu > 1$, toleranca $\varepsilon > 0$
izhod: x

1 **dokler** $m/t \geq \varepsilon$ **ponavljaj**
2 izračunaj $x^*(t)$, ki je minimum (6.7), z uporabo začetne točke x;
3 posodobi $x := x^*(t)$;
4 $t := \mu t$;
5 **vrni** x

Algoritem lahko vrne tudi dualen dopusten par λ^*, ν^*, ki skupaj s primarno rešitvijo x daje dualni razmik največ ε.

Korak 2 imenujemo zunanja iteracija ali tudi korak proti sredini, saj v njem izračunamo točko na središčni poti, ki pripada trenutnemu t. V tem koraku lahko uporabimo katerokoli metodo, ki zna rešiti problem (6.7), vendar bomo v nadaljevanju opisali, kako ta korak narediti z Newtonovo metodo, iteracije te metode znotraj koraka 2 pa imenujmo notranje iteracije. Pri tem je pomembno, da računanje optimuma v zunanji iteraciji do visoke natančnosti ni potrebno, saj je edina naloga optimumov $x^*(t)$, da služijo kot začetna rešitev v naslednji zunanji iteraciji in da postopoma vodijo k optimalni rešitvi osnovnega problema, ko t narašča proti neskončnosti. Torej si želimo poenostaviti računanje in minimum od (6.7) v zunanji iteraciji izračunati le približno, z nekaj notranjimi iteracijami. To pomeni, da je par $\lambda^*(t), \nu^*(t)$, ki ga iz $x^*(t)$ izračunamo z uporabo (6.11), le približno dopusten za dualen problem. Če je x, ki ga izračunamo v zunanji iteraciji, vseeno dovolj blizu središčne poti, potem je mogoče zagotoviti dualnost para λ, ν tako, da za izračun uporabimo namesto (6.11) rahlo spremenjeno formulo, glej [BBV04, Primer 11.9].

Pomembno vpašanje pri pregradni metodi je tudi izbira parametra μ, s katerim povečujemo t, saj z njim uravnavamo število zunanjih in otranjih iteracij. Večji kot je μ, manjše je število zunanjih iteracij, saj hitreje zmanjšujemo dualni razmik, ki je enak m/t, je pa zato večje število notranjih, saj je prejšnja rešitev (6.7) slaba začetna točka za nov problem (6.7). Obratno velja, če je μ zelo blizu 1. V praksi se za vrednosti μ med 3 in 100 ta dva pojava nekako poravnata in je skupno število notranjih iteracij približno konstantno. Če pa želimo imeti najboljšo najslabšo iteracijsko zahtevnost, potem so priporočene vrednosti μ blizu 1.

Pomembna je tudi izbira $t^{(0)}$. Če izberemo preveliko vrednost, bomo za prvo zunanjo iteraciji porabili zelo veliko notranjih (Newtonovih iteracij), bomo pa v nadaljevanju z malo zunanjimi iteracijami prišli do želene rešitve; če pa je $t^{(0)}$ zelo majhen, bomo potrebovali veliko zunanih iteracj, začetne zunanje iteracije pa bodo mogoče terjale malce manj notranjih iteracij.

Ker je pri izbranem $t^{(0)}$ število $m/t^{(0)}$ zgornja meja za začetni dualni razmik, je smiselno izbrati $t^{(0)}$ tako, da je $m/t^{(0)}$ enakega velikostnega razreda kot ocena začetnega dualnega razmika, ki jo dobimo npr. z izračunom $f_0(x^{(0)}) - g(\lambda, \nu)$, če poznamo začetno dualno dopustno rešitev.

Druga možnost je, da izberemo $t^{(0)}$ tako, da minimiziramo izraz

$$\|t\nabla f_0(x^{(0)}) + \nabla\phi(x^{(0)}) + A^T\nu\|_2, \tag{6.14}$$

ki ga lahko interpretiramo kot mero za odklon $x^{(0)}$ od središčne točke $x^*(t)$. Minimum lahko izračunamo po t in ν, kar je problem minimalne vsote kvadratov in je enostavno izračunljiv.

Če za začetno rešitev $x^{(0)} \in \mathcal{D}$ velja, da je strogo dopustna za neenačbe, torej $f_i(x^{(0)}) < 0, \forall i$, ni pa dopustna za enačbe $Ax = b$, potem v nekem koraku prve zunanje iteracije pregradne metode naredimo polni Newtonov korak, od tega koraka dalje pa so vse trenutne rešitve v notranjih in zunanjih iteracijah dopustne in nadaljujemo s standardno pregradno metodo.

6.3.1 Število zunanjih iteracij pregradne metode

V lemi 11 pokažemo, da ima optimizacijski problem (6.7), ki ga rešujemo v zunanjih iteracijah pregradne metode, enolično rešitev, ki jo lahko poljubno natančno izračunamo z Newtonovo metodo. Če torej v začetni in nadaljnjih korakih pregradne metode doovlj natančno izračunamo optimum (6.7) za $t = t^{(0)}, \mu t^{(0)}, \mu^2 t^{(0)}, \ldots$, potem dualni razmik pada in je po $(k+1)$ korakih, ko imamo rešitev $x^*(t)$ za $t = \mu^k t^{(0)}$, enak $m/(\mu^k t^{(0)})$. Da dosežemo natančnost ε, potrebujemo torej

$$1 + \left\lceil \frac{\ln\big(m/(\varepsilon t^{(0)})\big)}{\ln\mu} \right\rceil$$

korakov pregradne metode.

6.3.2 Newtonov korak za prilagojene KKT pogoje

Newtonov korak, ki je ključen element Newtonove metode pri reševanju (6.7) v koraku 2 pregradne metode, je definiran z naslednjo matrično enačbo v spremenljivkah $\Delta x_{\mathrm{nt}}, \nu_{\mathrm{nt}}$:

$$\begin{bmatrix} t\nabla^2 f_0(x) + \nabla^2\phi(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ \nu_{\mathrm{nt}} \end{bmatrix} = - \begin{bmatrix} t\nabla f_0(x) + \nabla\phi(x) \\ 0 \end{bmatrix} \tag{6.15}$$

Ta matrična enačba je ravno enačba za Newtonov korak za reševanje prilagojenih KKT pogojev (6.13), katerih ključni del tukaj ponovimo. Če je x dopusten, potem je x optimum (6.7) (torej točka na središčni poti) natanko tedaj, ko obstaja λ, ν, da veljao

$$\begin{aligned} \nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T\nu = 0, \\ -\lambda_i f_i(x) = 1/t, \quad i = 1, \ldots, m \\ Ax = b \end{aligned} \tag{6.16}$$

Ta sistem enačb tvori $n + p + m$ nelinearnih enačb v $n + p + m$ spremenljivkah x, ν, λ. Če najprej eliminiramo λ_i z uporabo $\lambda_i = -1/(tf_i(x))$, dobimo naslednji poenostavljen sistem $n + p$ enačb v $n + p$ spremenljivkah.

$$\nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-tf_i(x)} \nabla f_i(x) + A^T \nu = 0, \;\; Ax = b. \tag{6.17}$$

Newtonov korak za reševanje sistema enačb (6.17) je sestavljen iz aproksimacije nelinearnih členov v prvi enačbi s Taylorjevo vrsto prvega reda:

$$\begin{aligned}
&\nabla f_0(x+v) + \sum_{i=1}^{m} \frac{1}{-tf_i(x+v)} \nabla f_i(x+v) \\
&\approx \nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-tf_i(x)} \nabla f_i(x) + \nabla^2 f_0(x) v \\
&+ \sum_{i=1}^{m} \frac{1}{-tf_i(x)} \nabla^2 f_i(x) v + \sum_{i=1}^{m} \frac{1}{-tf_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T v.
\end{aligned}$$

Newtonov korak za reševanje (6.17) pomeni, da nelinearni del nadomestimo s to Taylorjevo aproksimacijo in dobimo sistem linearnih enačb:

$$Hv + A^T \nu = -g, \;\; Av = 0, \tag{6.18}$$

kjer so

$$\begin{aligned}
H &= \nabla^2 f_0(x) + \sum_{i=1}^{m} \frac{1}{-tf_i(x)} \nabla^2 f_i(x) + \sum_{i=1}^{m} \frac{1}{tf_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T \\
g &= \nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-tf_i(x)} \nabla f_i(x)
\end{aligned}$$

Opazimo lahko, da je

$$H = \nabla^2 f_0(x) + (1/t) \nabla^2 \phi(x), \;\; g = \nabla f_0(x) + (1/t) \nabla \phi(x),$$

Koraka $\Delta x_{\mathrm{nt}}, \nu_{\mathrm{nt}}$ v Newtonovem koraku (6.15) pregradne metode torej izračunamo z rešitvijo sistema linearnih enačb

$$tH \Delta x_{\mathrm{nt}} + A^T \nu_{\mathrm{nt}} = -tg, \;\; A \Delta x_{\mathrm{nt}} = 0.$$

Če primerjamo z (6.18), vidimo, da je

$$v = \Delta x_{\mathrm{nt}}, \;\; \nu = (1/t) \nu_{\mathrm{nt}},$$

torej lahko Newtnov korak za reševanje problema v koraku 2 interpretiramo (po množenju dualne spremenljivke ν_{nt} z $1/t$) s kot Newtonov korak pri reševanju prilagojenega in reduciranega KKT sistema (6.17).

Primer 55. *SDP in pregradna metoda*

6.4 Primarno dualne metode notranjih točk

V tem razdelku predstavimo osnovno verzijo primarno dualne metode notranjih točk. Ta metoda je zelo podobna pregradni metodi, a z nekaj pomembnimi razlikami:

- Metoda ima samo eno zanko, torej nima zunanjih in notranjih iteracij, kot je to pri pregradni metodi. V vsaki iteraciji metoda posodobi primarno in dualno dopustno rešitev.
- Iskalne smeri za novo primarno in dualno dopustno rešitev so dobljene z Newtonovo metodo, uporabljo na prilagojenih KKT pogojih. Te iskalne smeri so podobne kot pri barienri metodi, niso pa enake.
- Primarne in dualne rešitve, izračunane tekom iteracij, niso nujno dopustne za originalen problem.

Primarno dualne metode so pogosto bolj učinkovite kot pregradna metoda, še posebej, če nas zanima visoka natančnost rešitve in čas, potreben za izračun rešitve, saj lahko dosežejo konvergenco, hitrejšo od linearne. To še posebej velja za standardne družine problemov, kot so linearno, kvadratično, semidefinitno programiranje,...

6.4.1 Primarno dualne iskalne smeri

Tudi primarno dualna metoda začne s KKT pogoji (6.16), ki jih zapišemo kot $r_t(x,\lambda,\nu)=0$, kjer je $t>0$ in

$$r_t(x,\lambda,\nu)=\begin{bmatrix}\nabla f_0(x)+Df(x)\lambda+A^T\nu\\ -\mathbf{Diag}(\lambda)f(x)-1/t\mathbf{1}\\ Ax-b.\end{bmatrix} \tag{6.19}$$

V tem izrazu sta $f:\mathbb{R}^n\to\mathbb{R}^m$ in pripadajoča matrika odvodov $Df\in\mathbb{R}^{m\times n}$ definirana kot:

$$f(x)=\begin{bmatrix}f_1(x)\\ f_2(x)\\ \vdots\\ f_m(x)\end{bmatrix},\quad Df(x)=\begin{bmatrix}\nabla f_1(x)^T\\ \nabla f_2(x)^T\\ \vdots\\ \nabla f_m(x)^T\end{bmatrix}.$$

Če je x strogo dopusten in x,λ,ν zadoščajo pogoju $r(x,\lambda,\nu)=0$, potem je ta trojica na središčni poti, torej $x=x^*(t),\lambda=\lambda^*(t)$ in $\nu=\nu^*(t)$. Torej je par λ,ν dualno dopusten in dualni razik je enak m/t.

Prvo bločno komponento r_t iz (6.19)

$$r_{\text{dual}}=\nabla f_0(x)+Df(x)^T\lambda+A^T\nu$$

imenujemo tudi dualni ostanek, tretjo bločno komponento $r_{\text{pri}}=Ax-b$ imenujemo primarni ostanek, središčno komponento $r_{\text{cent}}=-\mathbf{Diag}(\lambda)f(x)-1/t\mathbf{1}$ pa središčni ostanek, saj predstavlja vektor vrednosti, za katere rešitev (x,λ,ν) odstopa od zrahljanega pogoja komplementarnosti $-\lambda_i f_i(x)=1/t$.

Oglejmo si, kako izgleda Newtonov korak za reševanje KKT pogojev (6.19), v neki točki (x, λ, ν), za katero velja $f_i(x) < 0, \lambda_i > 0$, za vsak i.

Označimo trenutko točko z $y = (x, \lambda, \nu)$ in Newtonov korak z $\Delta y = (\Delta x, \Delta\lambda, \Delta\nu)$. Newtonov korak je definiran z linearnimi enačbami

$$r_t(y + \Delta y \approx r_t(y) + Dr_t(y)\Delta y,$$

torej $\Delta y = -Dr_t(y)^{-1} r_t(y)$, oziroma v polni obliki

$$\begin{bmatrix} \nabla^2 f_0(x) + \sum_i \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\ -\mathbf{Diag}(\lambda) Df(x) & -\mathbf{Diag}(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{bmatrix} = - \begin{bmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{pri}} \end{bmatrix}. \tag{6.20}$$

Primarno dualni iskalni korak je definiran kot $\Delta y_{\text{pd}} = (\Delta x_{\text{pd}}, \Delta\lambda_{\text{pd}}, \Delta\nu_{\text{pd}})$ je torej definiran kot rešitev sistema linearnih enačb (6.20). Opazimo lahko, da je primarno dualna iskalna smer prepletena, kar pomeni, da je izračun primarnega iskalnega koraka Δx_{pd} odvisen od trenutne dualne rešitve λ, ν in seveda tudi od x. Ter obratno za izračun dualnega iskalenga koraka. Opomniti velja še to, da če je primarni ostanek $r_{\text{pri}} = 0$, potem tretja bločna enačba v (6.20) pomeni $A\Delta x_{\text{pd}} = 0$, torej je $A(x + s\Delta x_{\text{pd}}) = b$ za vsak s.

V primarno dualnih metodah notranjih točk rešitve x^k, λ^k, ν^k, ki jih dobimo tekom iteracij, niso nujno primarno dualno dopustne, kar pomeni, da ne moremo enostavno ovrednotiti dualnega razmika, kot lahko to naredimo pri pregradni metodi. Zato za vsak x, λ z $f_i(x) < 0, \forall i$, $\lambda \geq 0$ definiramo nadomestek dualnega razmika z

$$\hat{\eta}(x, \lambda) := -f(x)^T \lambda. \tag{6.21}$$

Ta nadomestek je kar enak dualnemu razmiku, če je x primarno dopusten in λ, ν dualno dopuste, torej, če je $r_{\text{pri}} = 0, r_{\text{dual}} = 0$. Vrednost parametra t, ki pripada nadomestku $\hat{\eta}$, je enaka $t = m/\hat{\eta}$.

Primarno dualno metodo notranjih točk lahko predstavimo z algoritmom 6.4.1

Algorithm 6.4.1: Primarno dualna metoda notranjih točk za (6.1)

vhod : x in λ z $f_i(x) < 0, \lambda_i > 0, \forall i$, $\mu > 1$, toleranci $\varepsilon_{\text{feas}} > 0, \varepsilon > 0$
izhod: x, λ, ν

ponavljaj
    izračunaj $t := \mu m/\hat{\eta}$;
    izračunaj primarno dualno iskalno smer Δy_{pd};
    izračunaj dolžino koraka $s > 0$ in posodobi $y = y + s\Delta y_{\text{pd}}$;
dokler $\|r_{\text{pri}}\|_2 \leq \varepsilon_{\text{feas}}, \|r_{\text{dual}}\|_2 \leq \varepsilon_{\text{feas}}, \hat{\eta} \leq \varepsilon$;
vrni x, λ, ν

6.4.2 Primerjava med pregradno metodo in primarno dualno metodo notranjih točk

Iskalna smer v primarno dualnih metodah notranjih točk je tesno povezana z iskalnimi smermi v pregradni metodi, nista pa ti smeri enaki. V nadaljevaju pokažemo, kje so razlike med obema smerema.

Iz bločne enačbe (6.20) lahko najprej odstranimo Δ_{pd} z uporabo druge bločne enačbe:

$$\Delta\lambda_{\mathrm{pd}} = -\mathbf{Diag}(f(x))^{-1}\mathbf{Diag}(\lambda)Df(x)\Delta x_{\mathrm{pd}} + \mathbf{Diag}(f(x))^{-1}r_{\mathrm{cent}}.$$

Ko s tem izrazom zamenjamo $\Delta\lambda_{\mathrm{pd}}$ v prvi bločni enačbi, dobimo, da se (6.20) poenostavi v

$$\begin{aligned} \begin{bmatrix} H_{\mathrm{pd}} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{pd}} \\ \Delta\nu_{\mathrm{pd}} \end{bmatrix} &= -\begin{bmatrix} r_{\mathrm{dual}} + Df(x)^T\mathbf{Diag}(f(x))^{-1}r_{\mathrm{cent}} \\ r_{\mathrm{pri}} \end{bmatrix} \\ &= -\begin{bmatrix} \nabla f_0(x) + \frac{1}{t}\sum_{i=1}^m \frac{1}{-f_i(x)}\nabla f_i(x) + A^T\nu \\ r_{\mathrm{pri}} \end{bmatrix} \end{aligned} \tag{6.22}$$

kjer je

$$H_{\mathrm{pd}} = \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) + \sum_{i=1}^m \frac{\lambda_i}{-f_i(x)}\nabla f_i(x)\nabla f_i(x)^T \tag{6.23}$$

Sedaj poglejmo enačb (6.15), ki definirajo iskalno smer v pregradni metodi. Te enačbe lahko zapišemo kot

$$\begin{aligned} \begin{bmatrix} H_{\mathrm{pre}} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{pre}} \\ \nu_{\mathrm{pre}} \end{bmatrix} &= -\begin{bmatrix} t\nabla f_0(x) + \nabla\phi(x) \\ r_{\mathrm{pri}} \end{bmatrix} \\ &= \;\; -\begin{bmatrix} t\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)}\nabla f_i(x) \\ r_{\mathrm{pri}} \end{bmatrix} \end{aligned} \tag{6.24}$$

Če je trenutna rešitev x primarno dopustna, potem je $r_{\mathrm{pri}} = 0$ in dobimo natanko enačbe (6.15). Zgoraj je

$$H_{\mathrm{pre}} = t\nabla^2 f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)}\nabla^2 f_i(x) + \sum_{i=1}^m \frac{1}{f_i(x)^2}\nabla f_i(x)\nabla f_i(x)^T. \tag{6.25}$$

Opazimo lahko, da sta sistema enačb (6.22) in (6.24) zelo podobna. Koeficienti v obeh sistemih imajo enako strukturo in matriki $H_{\mathrm{pd}}, H_{\mathrm{pre}}$ sta obe linearni kobinaciji matrik $\nabla^2 f_i(x)$, $\nabla f_i(x)\nabla f_i(x)^T$, za $i = 0, 1, \ldots, m$. Torej so ključni koraki izračuna obeh iskalnih smeri enaki. V nadaljevanju bomo natančneje osvetlili povezave med obema iskalnima smerema.

Če delimo prvi blok enačb (6.24) s t in definiramo $\Delta\nu_{\mathrm{pre}} = (1/t)\nu_{\mathrm{pre}} - \nu$ (ν je poljuben), dobimo:

$$\begin{bmatrix} (1/t)H_{\mathrm{pre}} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{pre}} \\ \Delta\nu_{\mathrm{pre}} \end{bmatrix} = -\begin{bmatrix} \nabla f_0(x) + (1/t)\sum_{i=1}^m \frac{1}{-f_i(x)}\nabla f_i(x) + A^T\nu \\ r_{\mathrm{pri}} \end{bmatrix}$$

V tej obliki je desna stran identična desni strani (6.22). Matriki koeficientov (levi strani enačb) pa se razlikujeta le v levem zgornjem bloku:

$$\begin{aligned} H_{\mathrm{pd}} &= \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) + \sum_{i=1}^m \frac{\lambda_i}{-f_i(x)}\nabla f_i(x)\nabla f_i(x)^T \\ (1/t)H_{\mathrm{pre}} &= \nabla^2 f_0(x) + \sum_{i=1}^m \frac{1}{-tf_i(x)}\nabla^2 f_i(x) + \sum_{i=1}^m \frac{1}{tf_i(x)^2}\nabla f_i(x)\nabla f_i(x)^T. \end{aligned}$$

Torej, če x in λ zadoščata $-f_i(x)\lambda_i = 1/t$ (če je torej $r_{\text{cent}} = 0$), potem sta ti dve matriki enaki, s tem pa tudi iskalni smeri.

Pri primarno dualnih metoda notranjih točk je pomembno vprašanje tudi določitev dolžine koraka, ki ga naredimo v isklani smeri - torej korak 4 v Algoritmu 6.4.1. Če je x, λ, ν trenutna točka, nova točka pa je nova točka ob koraku s enaka

$$x^+ = x + s\Delta x_{\text{pd}}, \quad \lambda^+ = \lambda + s\Delta\lambda_{\text{pd}}, \quad \nu^+ = \nu + s\Delta\nu_{\text{pd}}.$$

Korak s določimo v dveh fazah. V prvi fazi poščiščemo največji $s \leq 1$, ki zagotavlja $\lambda^+ \geq 0$:

$$\begin{aligned} s^{\max} :=& \{s \in [0,1] \mid \lambda + s\Delta\lambda \geq 0\} \\ =& \min\{1, \min\{-\lambda_i/\Delta\lambda_i \mid \Delta\lambda_i < 0\}\}. \end{aligned}$$

V drugi fazi korigiramo $s^{\max}$ z vračanjem nazaj, dokler ne dosežemo $f_i(x^+) < 0, \forall i$, kar naredimo tako, da najprej vzamemo $s_0 = 0.99 s^{\max}$, nato pa v k-tem koraku izračunamo $s_k = \beta s_{k-1}$, dokler ne velja $f_i(x^+) < 0, \forall i$. Ko to dosežemo, nadaljujemo z množenjem z β, dokler ne dosežemo še tretjega pogoja:

$$\|r_t(x^+, \lambda^+, \nu^+)\|_2 \leq (1 - \alpha s)\|r_t(x, \lambda, \nu)\|_2.$$

Izbira za α, β je poodbna kot pri klasični Newtonovi metodi, stroka svetuje $\alpha \in [0.01, 01]$, $\beta \in [0.3, 0.8]$.

Primer 56 (Linearno programiranje)**.**