Zadání zkoušky z konvexní optimalizace dne 10.2.2017

Termín odevzdání: do 9,00 hodin dne 11.2.2017, buď mailem nebo mně do rukou na katedře algebry v sobotu mezi 8,45-9,00

Celkově je ze zkoušky možné získat 52 bodů. Správné řešení příkladů 1,2,3 a správná formulace příkladu 4 jako úlohy lineárního programování jsou nutné podmínky pro složení zkoušky (nikoliv však postačující). Další nutnou podmínkou je získání aspoň 26 bodů. Všechny uvedené nutné podmínky jsou dohromady také postačující.

Dlouhé a složité výpočty nebudu luštit. Pište řešení jasně a přehledně.

Zkoušku řešte zcela samostatně, bez vzájemných konzultací. Používat můžete jakoukoliv literaturu.

Případné dotazy k formulacím úloh mi pošlete mailem na adresu `tumaatkarlin.mff.cuni.cz` do 11 hodin dne 10.2.2017. Odpovím pokud možno obratem.

Budete-li odevzdávat řešení mailem, pošlete každý příklad jako samostatný soubor s názvem `KO_prijmeni_cislo`, pozor na pořadí stránek! Všechny soubory pošlete zazipované v jednom balíku. Budete-li odevzdávat osobně na katedře algebry, naskenované odpovědi si také uložte do svého počítače.

V řešení by také měly být všechny programy, které u některých úloh napíšete a použijete.

Budu-li s někým z vás chtít hovořit v pondělí 13.2.2017 po obědě, napíšu Vám čas mailem nejpozději v neděli večer.

Příklad 1. *1 bod*

Pro $a = (a_1, a_2, \dots, a_k) \in \mathbb{R}^k$ označíme

$$f(t) = a_1 + a_2 t + \dots, a_k t^{k-1} \ .$$

Je množina $\{a \in \mathbb{R}^k : p(0) = -1, |p(t)| < 1 \text{ pro každé } t > 0\}$ konvexní? Odpověď dokažte.

Příklad 2. *1 bod*

Dokažte, že množina $D \subset \mathbb{R}^n$ je konvexní právě když platí

$$\alpha D + \beta D = (\alpha + \beta) D$$

pro každá dvě reálná čísla $\alpha, \beta \geq 0$.

Příklad 3. *3 body* Logaritmická bariéra pro kužel druhého řádu

Dokažte, že funkce $f(x,t) = -\log(t^2 - x^T x)$ definovaná na množině $\mathbf{dom}\ f = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : t > \|x\|_2\}$, tj. na kuželu druhého řádu, je konvexní. Můžete to udělat například tak, že spočítáte hessián funkce f a dokážete, že je v každém bodě definičního oboru f pozitivně semidefinitní. Nebo to můžete také udělat pomocí následující posloupnosti jednoduchých kroků. Každý z nich ale musíte pečlivě odůvodnit.

(1) Funkce $(u^T u)/t$ definovaná na $\mathbb{R}^n \times \mathbb{R}_{++}$ je konvexní,
(2) funkce $t - (u^T u)/t$ definovaná na $\mathbf{dom}\ f$ je konkávní,
(3) funkce $-\log\big(t - (u^T u)/t\big)$ definovaná na $\mathbf{dom}\ f$ je konvexní,
(4) funkce f je konvexní.

Příklad 4. *6 bodů* Optimalizace spotřeby energie

Uvažujeme lineární dynamický systém se stavovými vektory $x(t) \in \mathbb{R}^n$ a vnějšími vstupy $u(t) \in \mathbb{R}$ pro $t = 0, 1, \dots, N$. Jeho vývoj je daný rekurentní formulí

$$x(t+1) = A\,x(t) + b\,u(t), \quad t = 0, 1, \dots, N-1 \ ,$$

kde $A \in \mathbb{R}^{n \times n}$ je daná matice a $b \in \mathbb{R}$ je dané reálné číslo.

Mezi všemi možnými posloupnostmi vstupů $u = (u(0), u(1), \dots, u(N-1))^T \in \mathbb{R}^N$ najděte takovou, která minimalizuje spotřebu energie definovanou jako

$$E = \sum_{i=0}^{N-1} f(u(t))$$

za podmínek, že počáteční stav $x(0) = 0$ a cílový stav $x(N) = z \in \mathbb{R}^n$.

Funkce $f : \mathbb{R} \to \mathbb{R}$ udává energetickou náročnost vstupu $a \in \mathbb{R}$ předpisem

$$f(a) = \begin{cases} |a| & \text{je-li } |a| \le 1 \ , \\ 2|a| - 1 & \text{je-li } |a| > 1 \ . \end{cases}$$

Napřed úlohu zformulujte jako úlohu lineárního programování (LP) a potom ji pomocí CVX vyřešte pro konkrétní hodnoty vstupních dat

$$A = \begin{pmatrix} -1 & 0.5 & 0.8 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 0.2 \end{pmatrix}, \quad z = \begin{pmatrix} 7 \\ 2 \\ -6 \end{pmatrix}, \quad N = 30 \ .$$

Nakreslete hodnoty vstupů $u(t)$ do grafu, vodorovná osa je t, svislá $u(t)$. Třeba pomocí příkazu `plot` nebo něčeho podobného.

Příklad 5. *7 bodů* Volba rychlostí praotce Janečka

Pračlověk Janeček potřebuje postupně roznést vzkazy mezi spřátelené tlupy pralidí, se kterými se snaží vytvořit celoevropskou protineandrtálskou unii. Tlupy se nacházejí na $n+1$ různých místech označených $0, 1, 2, \ldots, n$. Vzdálenost mezi tlupami $i-1$ a i je d_i. Po trase délky d_i se Janeček pohybuje konstantní rychlostí v_i. Vzhledem k různým obtížnostem terénů mezi různými tlupami jsou možné konstantní rychlosti v_i omezené na intervaly $v_i^{\min} < v_i < v_i^{\max}$ pro $i = 1, 2, \ldots, n$.

Začíná u 0-té tlupy v čase $t = 0$. K tlupě 1 dorazí v čase $t_1 = d_1/v_1$. Zařve vzkaz a bez zastávky pokračuje v cestě k tlupě 2. K ní dorazí v čase $t_2 = t_1 + d_2/v_2$, atd. K poslední n-té tlupě tak dorazí v čase $t_n = t_{n-1} + d_n/v_n$. Vzhledem k tomu, že jde o kočovné tlupy, každá z nich se nachází na místě i pouze v časech t_i, které musí být v intervalech $t_i^{\min} < t_i < t_i^{\max}$, jinak se tlupa vydá na další pochod, aniž by ji pan Janeček zastihnul.

Během svých přesunů pan Janeček spotřebovává potraviny v závislosti na rychlosti svého pohybu, tato závislost je popsána nějakou rostoucí konvexní funkcí $f : \mathbb{R}_+ \to \mathbb{R}_{++}$.

Vstupní data pro úlohu jsou n, vektor vzdáleností $d = (d_1, d_2, \ldots, d_n)^T$, vektory $v^{\min}$, $v^{\max}$ omezující vektor rychlostí $v = (v_1, v_2, \ldots, v_n)^T$, vektory $t^{\min}$, $t^{\max}$ omezující možné časy $t = (t_1, t_2, \ldots, t_n)^T$, a funkce f. Vaším úkolem je navrhnout panu Janečkovi rychlosti v_i tak, aby se pohyboval povolenými rychlostmi v každém úseku, na každé místo dorazil v čase, kdy se na něm příslušná tlupa vyskytuje, a aby jeho celková spotřeba potravin během cesty byla co nejnižší.

(1) Formulujte tuto úlohu jako úlohu konvexní optimalizace. Pokud zavedete nové optimalizované proměnné, vysvětlete jak z nových proměnných dostanete proměnné $v_1, v_2, \ldots, v_n$. Pečlivě odůvodněte konvexitu účelové funkce a všech omezujících podmínek,

(2) vyřešte úlohu s konkrétními daty, která najdete v souboru `janecek_data.txt` a funkcí $f(v) = v^2 + 6v + 10$,

(3) uveďte jaká je celková spotřeba potravin pro optimální řešení nalezené v bodu (2),

(4) nalezený optimální vektor v znázorněte graficky např. pomocí funkce `step` z knihovny `mathplotlib` v Pythonu.

Příklad 6. *4 body*

Máme (primární) optimalizační úlohu

$$\begin{array}{lll} \text{minimalizujte} & f_0(x) & \text{kde } x \in \mathbb{R}^n \\ \text{za podmínek} & f_i(x) < 0, & i = 1, 2, \ldots, m \\ & l_i < x_i < u_i, & i = 1, 2, \ldots, n \ . \end{array}$$

Označíme $\mu \in \mathbb{R}^n_+$ vektor Lagrangeových multiplikátorů příslušných nerovnostem $x_i < u_i$ a $\nu \in \mathbb{R}^n_+$ vektor Lagrangeových multiplikátorů příslušných omezujícím nerovnostem $l_i < x_i$. Potom Lagrangián úlohy je

$$L(x, \lambda, \mu, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \mu^T(x - u) + \nu^T(l - x) \ ,$$

kde $u = (u_1, u_2, \ldots, u_n)^T$ a $v = (v_1, v_2, \ldots, v_n)^T$.

(1) Dokažte, že pro každé $x \in \mathbb{R}^n$ a každý vektor $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T \in \mathbb{R}^m$ existují vektory $\mu \in \mathbb{R}^n_{++}$ a $\nu \in \mathbb{R}^n_{++}$ takové, že x minimalizuje funkci $L(x, \lambda, \mu, \nu)$, a pomocí tohoto tvrzení popište množinu všech přípustných bodů pro duální úlohu (i.e. dual feasible points),

(2) sestrojte přípustný bod (λ, μ, ν) pro duální úlohu (i.e. dual feasible point) s využitím výsledku z bodu (1) pomocí volby $x = (l + u)/2$ a $\lambda = 0$. Odtud odvoďte dolní mez pro optimální hodnotu f^* primární úlohy. Ukažte, že tuto dolní mez můžete vyjádřit ve tvaru

$$f^* > f_0((l + u)/2) - ((u - l)/2)^T \, |\nabla f_0((l + u)/2)|$$

kde $|a|$ pro vektor $a = (a_1, a_2, \ldots, a_n)^T \in \mathbb{R}^n$ označuje vektor absolutních hodnot $(|a_1|, |a_2|, \ldots, |a_n|)^T$. Uměli byste tuto dolní mez také dokázat přímo?

Příklad 7. *2 body*

Uvažujeme optimalizační úlohu

$$\begin{array}{lll} \text{minimalizujte} & -\log\det X + \mathbf{trace}\,(SX) & \text{kde } X \in S^n_{++} \\ \text{za podmínky} & X \ \text{ je tridiagonální} & \end{array}$$

s danou maticí $S \in S^n$. Matice $A = (a_{ij}) \in S^n$ je tridiagonální, pokud obsahuje nenulové prvky pouze na hlavní diagonále, těsně nad hlavní diagonálou, a těsně pod hlavní diagonálou. Dokažte, že optimální hodnota matice X_{opt} splňuje

$$(X_{\text{opt}}^{-1})_{ij} = S_{ij}, \text{ pokud } |i - j| < 1 \ .$$

Příklad 8. *8 bodů* Po částech lineární regrese

Jsou-li dána reálná čísla $a_0 < a_1 < \cdots < a_K$, pak reálnou funkci f definovanou na uzavřeném intervalu $[a_0, a_K]$ nazýváme po částech lineární s uzlovými body $a_0, a_1, \ldots, a_K$, je-li lineární (nebo přesněji afinní) na každém uzavřeném intervalu $[a_i, a_{i+1}]$ pro $i = 0, 1, \ldots, K-1$. Nyní máme dána data $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, 2, \ldots, m$, splňující $a_0 \leq x_i \leq a_K$ pro každé i. Chceme najít konvexní po částech lineární funkci

f s uzlovými body $a_0, a_1, \ldots, a_K$, která tato data co nejlépe aproximuje ve smyslu metody nejmenších čtverců, tj. která minimalizuje reziduum

$$\sum_{i=1}^{m} (f(x_i) - y_i)^2 \ .$$

Návod. Je třeba najít vhodné optimalizované proměnné, které parametrizují funkci f, a omezující podmínky, které zajistí její konvexitu.

Úlohu pak vyřešte pro data uložená v souboru `PL_data_fitting.txt`. Všechna čísla x_i jsou z uzavřeného intervalu $[0, 1]$. Úlohu vyřešte postupně pro $K = 1, 2, 3, 4$ s pravidelně rozmístěnými uzlovými body v intervalu $[u_0 = 0, u_K = 1]$. Například pro $K = 4$ jsou uzlové body $u_0 = 0, u_1 = 0.25, u_2 = 0.5, u_3 = 0.75$ a $u_4 = 1$. Uveďte vždy hodnotu rezidua nejlepší aproximace a nakreslete data v rovině spolu s optimální po částech lineární funkcí f, kterou jste našli.

Příklad 9. *6 bodů* Nejširší pás oddělující dvě množiny

Jsou dány dvě podmnožiny v $\mathbb{R}^n$: polyedr

$$C = \{x \in \mathbb{R}^n : A\,x \preceq b\} \ ,$$

kde matice $A \in \mathbb{R}^{m \times n}$ a vektor $b \in \mathbb{R}^m$, a dále elipsoid

$$D = \{P\,u + q : \|u\| \leq 1\}$$

definovaný maticí $P \in \mathbb{R}^{n \times n}$ a vektorem $q \in \mathbb{R}^n$. Předpokládáme, že množiny C, D jsou neprázdné a disjunktní. Zajímá nás optimalizační problém

$$\begin{array}{ll} \text{maximalizujte} & \inf\{a^T x : x \in C\} - \sup\{a^T x : x \in D\} \\ \text{za podmínky} & \|a\|_2 = 1 \ , \end{array}$$

vzhledem k proměnné $a \in \mathbb{R}^n$.

Vysvětlete, např. obrázkem, geometrickou interpretaci této úlohy. Zformulujte tento problém jako nějaký standardní optimalizační problém, např. LP (lineární programování), QP (kvadratické programování), SOCP (second-order cone programming), SDP (semidefinitní programování)...

Příklad 10. *6 bodů* Efektivní numerická metoda pro regularizované nejmenší čtverce

Řešíme úlohu na regularizované nejmenší čtverce s vyhlazováním (smoothing)

$$\text{minimalizujte} \quad \sum_{i=1}^{k} (a_i^T x - b_i)^2 + \delta \sum_{i=1}^{n-1} (x_i - x_{i+1})^2 + \eta \sum_{i=1}^{n} x_i^2 \ ,$$

kde $x \in \mathbb{R}^n$ je optimalizovaná proměnná, $a_i \in \mathbb{R}^n$ jsou dané vektory, a $\delta, \eta > 0$ jsou regularizační parametry.

(1) Vyjádřete podmínku optimality pro tuto úlohu ve tvaru soustavy lineárních rovnic v proměnné x (říká se jí také soustava normálních rovnic).

(2) Nyní budeme předpokládat, že $k \ll n$. Navrhněte efektivní metodu jak vyřešit soustavu normálních rovnic pro tuto úlohu, spočítejte přibližný odhad počtu aritmetických operací (flops) pro vaši metodu a porovnejte jej s odhadem počtu aritmetických operací v případě, že matice soustavy normálních rovnic je obecná regulární matice.

(3) **Dobrovolné**. Budete-li mít čas a chuť, vygenerujte si náhodně instanci této úlohy s parametry $k = 100$, $n = 4000$, $\delta = \eta = 1$. Vyřešte soustavu normálních rovnic pro vygenerovanou instanci problému pomocí vámi navrženého efektivního algoritmu a porovnejte časovou náročnost s řešením téže soustavy pomocí standardního algoritmu fungujícího pro jakoukoliv regulární matici řádu n (např. Gaussovo eliminací).

Příklad 11. *8 bodů* Steganografická síť

Náš nepřítel si vytvořil steganografickou síť sestávající z n uzlů (vrcholů) a m steganografických kanálů (hran), každý spojuje vždy dva různé uzly (vrcholy). Celou nepřítelovu síť tedy lze popsat jako orientovaný graf bez smyček s n vrcholy a m hranami. Pomocí této sítě chce předat zprávu ze zdrojového uzlu 1 do cílového uzlu n. Pro každý kanál (hranu) j zná pravděpodobnost p_j, že zpráva nebude při přenosu kanálem j odhalena. Odhalení zpráv v různých kanálech jsou nezávislé události, takže pravděpodobnost, že nepřítel nebude odhalen při přenášení zprávy z počátečního uzlu 1 do cílového uzlu n, se rovná $\prod_{j \in P} p_j$, kde P je množina použitých kanálů na cestě od 1 do n. Na základě známých pravděpodobností p_j si tak může spočítat maximální hodnotu pravděpodobnosti neodhalení zprávy $F^{\max}$ přes všechny možné cesty z 1 do n v jeho steganografické síti.

Pravděpodobnost neodhalení zprávy při přenosu kanálem j závisí na tom, kolik investujeme do stegoanalýzy tohoto kanálu. Použijeme jednoduchý model, kde investice $x_j \in \mathbb{R}_+$ do stegoanalýzy kanálu j znamená pravděpodobnost $p_j = e^{-a_j x_j}$, že zprávu procházející kanálem j neodhalíme, přičemž $a_j \in \mathbb{R}_{++}$ pro $j = 1, 2, \ldots, m$ jsou daná čísla. Dále máme omezení na maximální možnou investici $x_j^{\max}$ do stegoanalýzy kanálu j a celkový rozpočet B, do kterého se musíme vejít, tj. pro naše investice do stegoanalýzy kanálů v síti musí platit $\sum_{j=1}^{m} x_j \leq B$.

(1) Vysvětlete, jak vyřešit problém volby investic x do stegoanalýzy kanálů v síti tak, aby pravděpodobnost $F^{\max}$, že zprávu neodhalíme, byla co nejmenší. Oceníme, když použijete nějaký postup, který bude založený na vyčíslení všech možných cest z 1 do n v grafu sítě. Můžete to například zkusit tak, že najdete maximální hodnoty F_i součinů $\prod_{j \in P_i} p_j$, přes množinu P_i všech možných cest ze zdrojového uzlu 1 do uzlu i. Tj. $F^{\max} = F_n$.

(2) Cesty v orientovaném grafu bez smyček lze dobře popsat pomocí matice $C = (c_{ij})$, kde

$$c_{ij} = \begin{cases} -1 & \text{pokud } i \text{ je počáteční vrchol hrany } j \\ +1 & \text{pokud } i \text{ je koncový vrchol hrany } j \\ 0 & \text{v ostatních případech} \end{cases}$$

V každém sloupci matice C je tedy právě jeden prvek 1, právě jeden -1 a ostatní prvky jsou 0. Čemu se rovná j-tá složka součinu $C^T z$ pro vektor $z \in \mathbb{R}^m$?

(3) Použijte svoji metodu k nalezení vektoru $x \in \mathbb{R}^m$, který minimalizuje hodnotu $F^{\max}$ v případě konkrétní sítě s $n = 10$ uzly (vrcholy) $1, 2, \ldots, 10$ a $m = 20$ hranami $(1,2), (1,3), (1,4), (2,3), (2,4), (2,6), (3,5), (3,6), (4,6), (4,7), (5,7), (5,8), (6,7), (6,8), (7,8), (7,9), (7,10), (8,9), (8,10), (9,10)$. Hodnotu vektoru $a \in \mathbb{R}^m_{++}$ si zvolte náhodně tak, aby jeho jednotlivé složky byly vybrány uniformně z otevřeného intervalu $(0,1)$. Podobně vygenerujte vektor $x^{\max} \in \mathbb{R}^n_{++}$ ve tvaru $x^{\max} = \mathbf{1} + b \in \mathbb{R}^n_{++}$, kde $\mathbf{1} \in \mathbb{R}^m$ je vektor, který má všechny složky rovné 1 a b je opět náhodný vektor, jehož jednotlivé složky jsou vybrány uniformně z otevřeného intervalu $(0,1)$. Nakonec položte $B = m/2$. Porovnejte optimální hodnotu $F^{\max *}$ s hodnotou $F^{\max}$ v případě, kdy zvolíme vektor $x = (B/m)\mathbf{1}$, tj. když do stegoanalýzy každého kanálu investujeme stejnou částku.