Name: Uzair Akbar ID: 03697290

Convex Optimization for Computer Vision

Exercise 1

a) $\bigcap_{C \in \mathcal{C}} C$

Let $x_1, x_2 \in \bigcap_{C \in \mathcal{C}} C$

then, the convex combination, x, is given by:

$$x = \theta x_1 + (1-\theta) x_2, \quad \theta \in [0,1]$$

however, this convex combination, x, should also be contained in all $C \in \mathcal{C}$ (definition of convex set), since both x_1 and x_2 are also in all $C \in \mathcal{C}$ (definition of intersection).

$$\because x \in C, \forall C \in \mathcal{C}$$

$$\therefore x \in \bigcap_{C \in \mathcal{C}} C \quad \text{(by definition of intersection).}$$

Hence proved that $\bigcap_{C \in \mathcal{C}} C$ is convex for convex family $\mathcal{C}$.

b) $P := \{x \in \mathbb{R}^n : Ax \leq b\}$

Here, we assume the definition of $\leq$ to be component/element wise inequality (sometimes represented as $\preceq$).

Let $x_1, x_2 \in P$.

$\Rightarrow Ax_1 \leq b \quad , \quad Ax_2 \leq b$ — (i)

Now, take x to be the convex combination of x_1, x_2:

$\Rightarrow x = \theta x_1 + (1-\theta)x_2 \quad , \quad \theta \in [0,1]$

$\Rightarrow Ax = A(\theta x_1 + (1-\theta)x_2)$

$\Rightarrow \quad = \theta Ax_1 + (1-\theta)Ax_2$

$\Rightarrow \quad \leq \theta b + (1-\theta)b$ [from (i)]

$\therefore Ax \leq b$

Scanned with CamScanner

c) $C_1 + C_2 := \{x+y : x \in C_1, y \in C_2\}$ (Minkowski sum)

Let $x, y \in C_1 + C_2$ such that $x = x_1 + x_2$, $y = y_1 + y_2$, where $x_i, y_i \in C_i$:.

Then, the convex combination of x and y is:

$\therefore z = \theta x + (1-\theta) y$, $\theta \in [0,1]$

$\Rightarrow \quad = \theta(x_1 + x_2) + (1-\theta)(y_1 + y_2)$

$\Rightarrow \quad = \{\theta x_1 + (1-\theta) y_1\} + \{\theta x_2 + (1-\theta) y_2\}$

$\Rightarrow z = z_1 + z_2$ —— (ii)

where z_i is the convex combination of points in C_i, therefore $z_i \in C_i$ (definition of convex set).

$\because z_1 \in C_1$, $z_2 \in C_2$

$\therefore z \in C_1 + C_2$ [from (ii) and definition of Minkowski sum]

Name: Uzair Akbar ID: 03697290

d) $\lambda C_1 := \{\lambda x : x \in C_1\}$ (λ-dilation)

let $x_1, x_2 \in C_1$, then convex combination is given as

$\Rightarrow x = \theta x_1 + (1-\theta) x_2 \in C_1$ [$\because C_1$ convex]

$\Rightarrow \lambda x = \lambda(\theta x_1 + (1-\theta) x_2)$ ~~(iii)~~

$\Rightarrow \quad = \theta(\lambda x_1) + (1-\theta)(\lambda x_2)$

$\Rightarrow x' = \theta x_1' + (1-\theta) x_2'$, where $x' = \lambda x$, $x_i' = \lambda x_i$ (by definition). Hence,

where $x', x_i' \in \lambda C_1$

λC_1 is convex.

Exercise 2

$\because \sum_{i=1}^{N} \lambda_i = 1,\ \lambda_i \geq 0$ definition of convex set

For $N=2$,

$$\sum_{i=1}^{2} \lambda_i x_i = \lambda_1 x_1 + \lambda_2 x_2 = \lambda_1 x_1 + (1-\lambda_1) x_2 \in C$$

Hence, the statement holds for $N=2$.

Now assume that the statement holds for some N. Then, for the $N+1$ case

$$\Rightarrow \sum_{i=1}^{N+1} \lambda_i x_i = \sum_{i=1}^{N} \lambda_i x_i + \lambda_{N+1} x_{N+1}$$

$$= (1-\lambda_{N+1}) \sum_{i=1}^{N} \frac{\lambda_i x_i}{(1-\lambda_{N+1})} + \lambda_{N+1} x_{N+1}$$

$$= (1-\lambda_{N+1}) \sum_{i=1}^{N} \lambda_i' x_i + \lambda_{N+1} x_{N+1}$$

$\left[\lambda_i' = \lambda_i / (1-\lambda_{N+1}) \Rightarrow \lambda_i' \in [0,1] \right]$ and $\sum_N \lambda_i = 1$

$$= (1-\lambda_{N+1}) x' + \lambda_{N+1} x_{N+1}$$

$\left[x' = \sum_{i=1}^{N} \lambda_i' x_i \Rightarrow x' \in C \right.$ as per our assumption. for N.$\left.\right]$

$$\therefore \sum_{i=1}^{N+1} \lambda_i x_i = (1-\lambda_{N+1}) x' + \lambda_{N+1} x_{N+1} \in C$$

Because $x', x_{N+1} \in C$ and $\lambda_{N+1} \in [0,1]$ and the statement already holds ~~for~~ for $N=2$.

Hence, proved. ~~that if~~

Exercise 3

Let X be a closed set.

Hence $X^c := \mathbb{R}^n \setminus X$ is open.

$$\Rightarrow \forall x \in X^c, \exists \varepsilon > 0 \text{ s.t. } B_\varepsilon(x) \subset X^c.$$

$$\Rightarrow B_\varepsilon(x) \cap X = \phi \quad \text{—— (iii)}$$

Now, for a convergent sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ with limit $x_n \to x, n \to \infty$, by definition,

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, x_{n_0} \in B_\varepsilon(x) \quad \text{—— (iv)}$$

(where x is limit of x_n)

But, $x_n \in X$ (Given, statement 2 in question). —— (v)

Hence, $x \notin X^c$ as (iii), (iv) and (v) present a contradiction. $\therefore x \in X$

Name: Uzair Akbar ID: 03697290

Exercise 4

Name: Uzair Akbar ID: 03697290

a)

$$\nabla d(u) = \nabla \sqrt{u^T A u}$$

$$= \frac{1}{2\sqrt{u^T A u}} \nabla (u^T A u) \quad \text{[chain rule]}$$

$$= \frac{1}{2\sqrt{u^T A u}} \nabla \{(u^T)(Au)\}$$

$$= \frac{1}{2\sqrt{u^T A u}} \left[\nabla u^T (\underline{Au}) + \nabla \underline{u}^T A u \right]$$

[$\underline{A}$ denotes that A is taken as constant in calculation of the gradient.]

$$= \frac{1}{2\sqrt{u^T A u}} \left[(Au)^T + u^T A \right]$$

[$\because \nabla x^T A = A^T$
$\because \nabla a^T x = a^T, \nabla Ax = A$]

$$= \frac{1}{2\sqrt{u^T A u}} u^T (A^T + A)$$

$$\therefore \nabla d(u) = \frac{u^T(A^T + A)}{2\sqrt{u^T A u}}$$

b)

If A is not full rank, then it has a nontrivial kernel, i.e. $\exists u$, $Au = 0$, or $u^T A u = 0$ and therefore the gradient will not be defined for $u \in \ker(A)$.

c)

we shall make use of the following identity:

$$\nabla(\|Az\|^2) = 2A^T A z$$

Therefore:

$$\Rightarrow \nabla(\|-z - \mathbb{1}\langle x, Az\rangle\|^2) = \nabla(\|-(z + \mathbb{1}x^T Az)\|^2)$$

$$\Rightarrow \qquad = \nabla(\|z + \mathbb{1}x^T Az\|^2)$$

$$\Rightarrow \qquad = \nabla(\|(I + \mathbb{1}x^T A)z\|^2)$$

$$\therefore \nabla(\|-z - \mathbb{1}\langle x, Az\rangle\|^2) = 2(I + \mathbb{1}x^T A)^T (I + \mathbb{1}x^T A)z$$

Now, we calculate the Jacobian matrix of $R(z)$ as $R(z)$ is a many-to-many mapping.

$$J_R = \frac{J_z}{f^2}\sqrt{f^2\|Az\|^2 + \|-z-\mathbb{1}\langle x, Az\rangle\|^2 + \varepsilon} + \frac{z}{f^2}\left(\nabla\sqrt{f^2\|Az\|^2 + \|-z-\mathbb{1}\langle x, Az\rangle\|^2 + \varepsilon}\right)^T$$

$$\Rightarrow \quad = \frac{I}{f^2}\sqrt{f^2\|Az\|^2 + \|-z-\mathbb{1}\langle x, Az\rangle\|^2 + \varepsilon} + \frac{z}{f^2}\left(\frac{f^2\nabla(\|Az\|^2) + \nabla(\|-z-\mathbb{1}\langle x, Az\rangle\|^2)}{2\sqrt{f^2\|Az\|^2 + \|-z-\mathbb{1}\langle x, Az\rangle\|^2 + \varepsilon}}\right)^T$$

$$\therefore J_R = \frac{I}{f^2}\sqrt{f^2\|Az\|^2 + \|-z-\mathbb{1}\langle x, Az\rangle\|^2 + \varepsilon} + \frac{z}{f^2}\left(\frac{f^2 A^T A z + (I + \mathbb{1}x^T A)^T (I + \mathbb{1}x^T A) z}{\sqrt{f^2\|Az\|^2 + \|-z-\mathbb{1}\langle x, Az\rangle\|^2 + \varepsilon}}\right)^T$$