Math Problem Set #5: Convex Analysis

OSM Lab - University of Chicago
Geoffrey Kocks

Problem 1: HJ 7.1.

Solution. Let S be a nonempty subset of V. Then conv(S) is defined by the set of all finite sums of the form: $\lambda_1\mathbf{x}_1 + ... + \lambda_k\mathbf{x}_k$ where $\mathbf{x}_i$ are the elements of S, $\lambda_i \geq 0$ and $\lambda_1 + ... + \lambda_k = 1$. Now let $\mathbf{x}$ and $\mathbf{y}$ each be elements of conv(S). Then we can express them as:

$$\mathbf{x} = \lambda_{11}\mathbf{x}_1 + ... + \lambda_{k1}\mathbf{x}_k$$

$$\mathbf{y} = \lambda_{12}\mathbf{x}_1 + ... + \lambda_{k2}\mathbf{x}_k$$

conv(S) is convex if for all λ such that $0 \leq \lambda \leq 1$ and all $\mathbf{x}, \mathbf{y}$ in conv(S): $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in$ conv(S). Then for any λ:

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y} = (\lambda\lambda_{11} + (1-\lambda)\lambda_{12})\mathbf{x}_1 + ... + (\lambda\lambda_{k1} + (1-\lambda)\lambda_{k2})\mathbf{x}_k.$$

We note that the form of this sum is the same as the form that defines the elements of conv(S). We also note that the coefficient in front of each $\mathbf{x}_i$ is at least zero because it it is made up of a sum of non-negative terms. Finally we show that this sum is part of conv(S) by showing that the coefficients sum to 1:

$$\begin{aligned}(\lambda\lambda_{11} + (1-\lambda)\lambda_{12}) + ... + (\lambda\lambda_{k1} + (1-\lambda)\lambda_{k2}) &= \lambda(\lambda_{11} + ...\lambda_{k1}) + (1-\lambda)(\lambda_{12} + ... + \lambda_{k2}) \\ &= \lambda + (1-\lambda) = 1.\end{aligned}$$

Therefore if S is a nonempty subset of V, then conv(S) is convex.

Problem 2: HJ 7.2.

Solution.

(i.) A hyperplane in V is defined as the set of the form $P = \{\mathbf{x} \in V | \langle \mathbf{a}, \mathbf{x} \rangle = b\}$. To show that a hyperplane is convex, we must show that for $\mathbf{x}, \mathbf{y}$ in C, we have that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in C$. By the definition of a hyperplane, $\langle \mathbf{a}, \mathbf{x} \rangle = b$ and $\langle \mathbf{a}, \mathbf{y} \rangle = b$. Then:

$$\langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle = \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle = \lambda(b) + (1-\lambda)(b) = b.$$

Therefore $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ is part of the set so a hyperplane is convex.

(ii.) A half space in V is defined as the set of the form $P = \{\mathbf{x} \in V | \langle \mathbf{a}, \mathbf{x} \rangle \leq b\}$. To show that a half space is convex, we must show that for $\mathbf{x}, \mathbf{y}$ in C, we have that $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in C$. By the definition of a half space, $\langle \mathbf{a}, \mathbf{x} \rangle \leq b$ and $\langle \mathbf{a}, \mathbf{y} \rangle \leq b$. Then:

$$\langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle = \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle \leq \lambda(b) + (1-\lambda)(b) = b.$$

Therefore $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ is part of the set so a half space is convex.

Problem 3: HJ 7.4.

Solution.

(i.)

$$\begin{aligned}\|\mathbf{x}-\mathbf{y}\|^2 &= \|\mathbf{x}-\mathbf{p}+\mathbf{p}-\mathbf{y}\|^2 = \langle \mathbf{x}-\mathbf{p}+\mathbf{p}-\mathbf{y}, \mathbf{x}-\mathbf{p}+\mathbf{p}-\mathbf{y}\rangle \\ &= \|\mathbf{x}-\mathbf{p}\|^2 + \langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \langle \mathbf{p}-\mathbf{y}, \mathbf{x}-\mathbf{p}\rangle + \|\mathbf{p}-\mathbf{y}\|^2 \\ &= \|\mathbf{x}-\mathbf{p}\|^2 + 2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \|\mathbf{p}-\mathbf{y}\|^2\end{aligned}$$

(ii.) Assume that $\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle \geq 0$. Then from the result of part (i):

$$\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}-\mathbf{p}\|^2 + 2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \|\mathbf{p}-\mathbf{y}\|^2 > \|\mathbf{x}-\mathbf{p}\|^2$$

By definition, each matrix norm is positive, so we can take the square root of each side to obtain: $\|\mathbf{x}-\mathbf{y}\| > \|\mathbf{x}-\mathbf{p}\|$

(iii.) Let $\mathbf{z} = \lambda\mathbf{y} + (1-\lambda)\mathbf{p}$. Then:

$$\begin{aligned}\|\mathbf{x}-\mathbf{z}\|^2 &= (\sqrt{\langle \mathbf{x}-\mathbf{z}, \mathbf{x}-\mathbf{z}\rangle})^2 = \langle \mathbf{x}-\lambda\mathbf{y}-\mathbf{p}+\lambda\mathbf{p}, \mathbf{x}-\lambda\mathbf{y}-\mathbf{p}+\lambda\mathbf{p}\rangle \\ &= \langle \mathbf{x}-\mathbf{p}\rangle + \langle \mathbf{x}-\mathbf{p}, -\lambda\mathbf{y}+\lambda\mathbf{p}\rangle + \langle -\lambda\mathbf{y}+\lambda\mathbf{p}, \mathbf{x}-\mathbf{p}\rangle + \langle -\lambda\mathbf{y}+\lambda\mathbf{p}, \lambda\mathbf{y}+\lambda\mathbf{p}\rangle \\ &= \|\mathbf{x}-\mathbf{p}\|^2 + \lambda\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \lambda\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + (-\lambda)^2\|\mathbf{y}-\mathbf{p}\|^2 \\ &= \|\mathbf{x}-\mathbf{p}\|^2 + 2\lambda\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \lambda^2\|\mathbf{y}-\mathbf{p}\|^2\end{aligned}$$

(iv.) Let $\mathbf{z} = \lambda\mathbf{y} + (1-\lambda)\mathbf{p}$. Then by the definition of a convex set, if $\mathbf{y}$ and $\mathbf{p}$ are both in the convex set C, then $\mathbf{z}$ is also part of the convex set C. By the definition of a projection, if $\mathbf{p}$ is the projection of $\mathbf{x}$ onto C, then $\|\mathbf{x}-\mathbf{p}\| \leq \|\mathbf{x}-\mathbf{y}\|$. Since $\mathbf{z}$ is also part of the set, $\|\mathbf{x}-\mathbf{p}\| \leq \|\mathbf{x}-\mathbf{z}\| \implies \|\mathbf{x}-\mathbf{z}\|^2 - \|\mathbf{x}-\mathbf{p}\|^2 \geq 0$. Then by the result of part (iii) of this problem:

$$2\lambda\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \lambda^2\|\mathbf{y}-\mathbf{p}\|^2 \geq 0$$

$$\implies 0 \leq 2\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle + \lambda\|\mathbf{y}-\mathbf{p}\|^2.$$

Therefore it follows that $\langle \mathbf{x}-\mathbf{p}, \mathbf{p}-\mathbf{y}\rangle \geq 0$ if $\mathbf{p}$ is a projection of $\mathbf{x}$ onto the convex set C.

Problem 4: HJ 7.6.

Solution. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function. Then for $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$: $f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$. Define the set C as $\{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) \leq c\} \in \mathbb{R}^n$. C is a convex set if $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ is also an element of C for any $\mathbf{x}, \mathbf{y}$ in C. $\mathbf{x}$ and $\mathbf{y}$ in C implies $f(\mathbf{x}) \leq c, f(\mathbf{y}) \leq c$. Then:

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) \leq \lambda(c) + (1-\lambda)(c) = c.$$

Therefore the combination of $\mathbf{x}$ and $\mathbf{y}$ is part of C, so f being a convex function implies that the set C is a convex set.

Problem 5: HJ 7.7.

Solution. If a function $f_i(x)$ is convex, this implies that for any $\mathbf{x}, \mathbf{y}$ in the domain: $f_i(\lambda\mathbf{x} - (1-\lambda)\mathbf{y}) \leq \lambda f_i(\mathbf{x}) + (1-\lambda)f_i(\mathbf{y})$. Let $f_1, ..., f_k$ be convex functions and define the function f as $f(x) = \sum_{i=1}^{k} \lambda_i f_i(x)$. Then for any $\lambda, \lambda_i \in \mathbb{R}_+$:

$$\lambda_i(f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq (\lambda_i)(\lambda f_i(\mathbf{x}) + (1-\lambda)f_i(\mathbf{y}))$$

$$\implies \sum_{i=1}^{k} \lambda_i f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda \sum_{i=1}^{k} \lambda_i f_i(\mathbf{x}) + (1-\lambda) \sum_{i=1}^{k} \lambda_i f_i(\mathbf{y}).$$

$$\implies f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}).$$

Therefore any nonnegative combination of convex functions is convex.

Problem 6: HJ 7.13.

Solution. Assume that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex and bounded above but not constant. By the definition of convexity: $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$. Now define $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ such that $f(\mathbf{z}) > f(\mathbf{y})$. Then the definition of convexity above implies:

$$f(\mathbf{z}) \leq \lambda f(\frac{\mathbf{z} - (1-\lambda)\mathbf{y}}{\lambda}) + (1-\lambda)f(\mathbf{y})$$

$$\implies \frac{f(\mathbf{z}) - (1-\lambda)f(\mathbf{y})}{\lambda} \leq f(\frac{\mathbf{z} - (1-\lambda)\mathbf{y}}{\lambda})$$

$$\implies \frac{f(\mathbf{z}) - f(\mathbf{y})}{\lambda} + f(\mathbf{y}) \leq f(\frac{\mathbf{z} - (1-\lambda)\mathbf{y}}{\lambda})$$

The left side of the inequality approaches ∞ as λ approaches 0, so on the right side of the inequality f will be unbounded. Therefore if f is convex and bounded above, f must be a constant.

Problem 7: HJ 7.20.

Solution. Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $-f$ both be convex. Therefore by the definition of convexity, we have for any $\mathbf{x}, \mathbf{y}$ in the domain:

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

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

$$\implies f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}).$$

Fix a value $\mathbf{y} = \mathbf{y}_0$ and we can then rearrange the equation as:

$$f(\mathbf{x}) = (1/\lambda)(f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}_0) - (1-\lambda)f(\mathbf{y}_0)).$$

We have therefore expressed $f(\mathbf{x})$ as a linear transformation of itself, so f is affine.

Problem 8: HJ 7.21.

Solution. Let $\mathbf{x}^*$ be the local minimizer for $f(\mathbf{x})$ subject to the constraints. This means that $f(\mathbf{x}^*) \leq f(\mathbf{y})$ for all $\mathbf{y}$ in the domain. Then because ϕ is a strictly increasing function, $\phi \circ f(\mathbf{x}^*) \leq \phi \circ f(\mathbf{y})$ This holds for any $\mathbf{y}$ in the domain, so that mean that $\mathbf{x}^*$ is also the local minimizer for $\phi \circ f(\mathbf{x})$.

Similarly, assume that $\mathbf{x}^*$ is the local minimizer for $\phi \circ f(\mathbf{x})$. If $\mathbf{x}^*$ were not also the local minimizer for $f(\mathbf{x})$, the reasoning from above would imply that $f(\mathbf{x}^{**}) \leq f(\mathbf{x}^*)$ which would mean that taking the composition of the function, $\mathbf{x}^*$ is not the local minimizer for $\phi \circ f(\mathbf{x})$. If $\mathbf{x}^*$ is the local minimizer for $\phi \circ f(\mathbf{x})$ it must also be the local minimizer for $f(\mathbf{x})$.

Therefore we have shown that $\mathbf{x}^*$ is the local minimizer for $\phi \circ f(\mathbf{x})$ if and only if it is the local minimizer for $f(\mathbf{x})$.