Math, Problem Set #5, Convex Optimization

Instructor: Jorge Barro

Due Friday, July 21 at 8:00am

Homework: 1, 2, 4, 5, 7, 13, 20, and 21 at the end of Chapter 7 of Humpherys et al. (2017)

7.1 Show that the convex hull of any non-empty set S is convex with $S \subset V$. The hull $H = conv(S)$ is given by the set of all finite sums of the form $\lambda_1 x_1 + ... + \lambda_n x_n$ for all $x \in S$, $n \in \mathbb{N}$ with $\sum \lambda_i = 1$ and $\lambda_i \geq 0$. H is convex if $\gamma y + (1-\gamma)x \in H \, \forall x, y \in H$ and $0 \leq \gamma \leq 1. \gamma y + (1-\gamma)x = \gamma(\lambda_1 y_1 + ... + \lambda_n y_n) + (1-\gamma)(\lambda_1' x_1 + ... + \lambda_k' x_k) \in H$ iff $\gamma \sum \lambda_i + (1-\gamma)\sum \lambda_j' = 1$. Since $x, y \in H$ by definition of H it follows that $\sum \lambda_i = \sum \lambda_j' = 1$. Finally, $\gamma 1 + (1-\gamma)1 = 1$ is true.

7.2i Claim: A hyperplane is convex.
Let x_a and x_b be any two arbitrary points in $P = \{x \in V | \langle a, x \rangle = b\}$. Then, $\lambda x_a + (1-\lambda)x_b = \lambda a_1 x_{a1} + ... + \lambda a_n x_{an} + ... + (1-\lambda)a_1 x_{b1} + ... + (1-\lambda)a_n x_{bn} = \lambda a_1 x_{a1} + a_1 x_{b1} - \lambda a_1 x_{b1} + ... + \lambda_a n x_{an} + a_n x_{bn} - \lambda a_n x_{bn} = b + \lambda b - \lambda b = b$. Since any convex combination of the two points is still in the hyperplane P, we know that the hyperplane is convex.

7.2ii Claim: Half-spaces are convex.
Let $H = \{x \in V | \langle a, x \rangle \leq b\}$ be a half-space, where $a \in V, a \neq 0$, and $b \in \mathbb{R}$. Then, for any two arbitrary points x_a and x_b in the half-space, we know that $\lambda(a_1 x_1 + ... + a_n x_n + (1-\lambda)(a_1 x_1' + ... + a_n x_n' = \lambda a_1 x_1 + a_1 x_1' - \lambda a_1 x_1' + ... + \lambda a_n x_n + a_n x_n' - \lambda a_n x_n' \leq \lambda b + b - \lambda b = b$. Since the convex combination of any two arbitrary points is in the half-space, we conclude that the half-space is convex.

7.4i Claim: $\|x-y\|^2 = \|x-p\|^2 + \|p-y\|^2 + 2\langle x-p, p-y \rangle$.
$\|x-y\|^2 = \langle x-y, x-y \rangle = \langle x-p+p-y, x-p+p-y \rangle = \langle x-p, x-p \rangle + \langle x-p, p-y \rangle + \langle p-y, x-p \rangle + \langle p-y, p-y \rangle = \|x-p\|^2 + \|p-y\|^2 + 2\langle x-p, p-y \rangle$.

7.4ii $\|x-p\| \leq \|x-p\| + \|p-y\|$ because $\|p-y\| \geq 0$. Therefore squaring both sides, we preserve the inequality and obtain $\|x-p\|^2 \leq \|x-p\|^2 + \|p-y\|^2 + 2\langle x-p, p-y \rangle = \|x-y\|^2$. Taking the squareroot of both sides now, we obtain $\|x-p\| \leq \|x-y\|$.

7.4iii Given that $z = \lambda y + (1-\lambda)p$, we can write $\|x-z\|^2 = \|x-p\|^2 + \|p-z\|^2 + 2\langle x-p, p-z \rangle = \|x-p\|^2 + \|p-\lambda y - p + \lambda p\|^2 + 2\langle x-p, p-\lambda y - p + \lambda p \rangle = \|x-p\|^2 + \|\lambda p - \lambda y\|^2 + 2\langle x-p, \lambda p - \lambda y \rangle = \|x-p\|^2 + \lambda^2 \|p-y\| + 2\lambda \langle x-p, p-y \rangle$.

7.4iv Claim: If p is a projection of x onto the convex set C, then $\langle x-p, p-y\rangle \geq 0 \forall y \in C$.
Suppose p is the projection of x onto the convex set C. Then we know that $\|x-z\|^2 = \|x-p\|^2 + 2\lambda\langle x-p, p-y\rangle + \lambda^2\|y-p\|^2$. We know that the right hand side of the equation is greater than $\|x-p\|^2$ since p is a projection onto C and z is a point in C (z is in C because C is convex and z is a convex linear combination of points in C). Moreover, the right hand side can be rewritten as $\|x-p\|^2 + \lambda(2\langle x-p, p-y\rangle + \lambda\|y-p\|^2)$ where $\lambda\|y-p\|^2 \geq 0$. Since the expression has to be greater than or equal to $\|x-p\|^2$, it follows that $2\langle x-p, p-y\rangle + \lambda\|y-p\|^2 \geq 0$ for all $y \in C$ and $\lambda \in [0,1]$. Thus, we can let $\lambda = 0$ and see that $2\langle x-p, p-y\rangle \geq 0$.

7.6 Claim: If f is a convex function, then the set $\{x \in \mathbb{R}^n | f(x) \leq c\}$ is a convex set.
Suppose f is a convex function. Let x_a and x_b be arbitrary elements of $S = \{x \in \mathbb{R}^n | f(x) \leq c\}$. It remains to show that $f(\lambda x_a + (1-\lambda)x_b) \leq c$. $f(\lambda x_a + (1-\lambda)x_b) \leq \lambda f(x_a) + (1-\lambda)f(x_b) \leq \lambda c + (1-\lambda)c = \lambda c - \lambda c + c = c$, as desired.

7.7 To show that f(x) is conex, we need to show that for all $x_1, x_2 \in C$, $f(\mu x_1 + (1-\mu)x_2) \leq \mu f(x_1) + (1-\mu)f(x_2)$. $f(\mu x_1 + (1-\mu)x_2) = \sum_{i=1}^k \lambda_i f_i(\mu x_1 + (1-\mu)x_2) \leq \sum_{i=1}^k \lambda_i[\mu f_i(x_1) + (1-\mu)f_i(x_2)] = \mu \sum_{i=1}^k \lambda_i f_i(x_i) + (1-\mu)\sum_{i=1}^k \lambda_i f_i(x_2) = \mu f(x_1) + (1-\mu)f(x_2)$.

7.13 Claim: If f is convex and bounded above, then f is constant.
Suppose f is convex and bounded above, and suppose to the contrary that there exists x, y where $f(x) \geq f(y)$. Then for any $x_1, x_2 \in C$ and $0 \leq \lambda \leq 1$, we have $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$. Thus, we know $f(x) \leq \lambda f(\frac{x-(1-\lambda)y}{\lambda}) + (1-\lambda)f(y)$, where $x = \lambda x_1 + (1-\lambda)x_2$ and $y = x_2$. Rearranging the inequality, we obtain $\frac{f(x)-(1-\lambda)f(y)}{\lambda} \leq f(\frac{x-(1-\lambda)y}{\lambda} \leq b$, where b is a finite upper bound for f. As $\lambda \to 0^+$, $\frac{f(x)-(1-\lambda)f(y)}{\lambda} \to \infty$, which contradicts the assumption that f is bounded above. Therefore, it follows that f is constant, that is $f(x) = f(y)$ for all $x, y \in C$.

7.20 Claim: If f is convex and -f is also convex, then f is affine.
Suppose f is convex. Then, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. Suppose -f is convex. Then $-f(\lambda x + (1-\lambda)y) \leq -\lambda f(x) + -(1-\lambda)f(y)$. Multiplying the last inequality by -1 throughout and see that $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$. The only way that both inequalities can be true is if $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$. Therefore f is linear and thus affine.

7.21 Suppose $f(x*) \leq f(y) \forall y \in \mathbb{R}^n$. Since ϕ is strictly increasing, we know that

$f(x^*)$ minimizes ϕ over the range of f, which means that x^* minimizes $\phi \circ f$. Suppose $\phi \circ f(x^*) \leq \phi \circ f(y) \forall y \in \mathbb{R}^n$. Then because ϕ is strictly increasing, $f(x^*) \leq f(y) \forall y \in C$, which means that x^* minimizes f.