Math, Problem Set #5, Convex Optimization

OSM Lab, Dr. Barro

Due Friday, July 21 at 8:00am

7.1 Prove **Proposition 7.1.5**: If S is a nonempty subset of V, then $conv(S)$ is convex.

Proof Take any $y, z \in conv(S)$. We want to show that $\lambda y + (1-\lambda)z \in conv(S), \quad \forall \lambda \in (0,1)$. Now note next that since $y, z \in conv(S)$, both can be written as a finite sum of $x_i \in S$, say

$$y = \sum_{j=1}^{n} \lambda_{i_j} x_{i_j}, \quad z = \sum_{k=1}^{m} \lambda_{i_k} x_{i_k}, \quad \text{where} \quad \sum_{j=1}^{n} \lambda_{i_j} = \sum_{k=1}^{m} \lambda_{i_k} = 1$$

Now note that

$$\lambda y + (1-\lambda)z = \lambda \sum_{j=1}^{n} \lambda_{i_j} x_{i_j} + (1-\lambda) \sum_{k=1}^{m} \lambda_{i_k} x_{i_k} = \sum_{j=1}^{n} \lambda \lambda_{i_j} x_{i_j} + \sum_{k=1}^{m} (1-\lambda) \lambda_{i_k} x_{i_k}$$

Some x_i might appear in both sums, but the definition of a convex hull does not require x_i to be distinct (and it would be redundant to require it, since it does not change the condition $\sum_{i=1}^{n} \lambda_i = 1$). Hence without loss of generality, the above is

$$= \sum_{i=1}^{n+m} (\mathbb{1}\{i \leq n\} \lambda \lambda_{i_j} + \mathbb{1}\{i > n\}(1-\lambda)\lambda_{i_k}) x_i, \quad \text{where} \quad x_i \in S \quad \text{and}$$

$$\sum_{i=1}^{n+m} (\mathbb{1}\{i \leq n\} \lambda \lambda_{i_y} + \mathbb{1}\{i > n\}(1-\lambda)\lambda_{i_k}) = \sum_{j=1}^{n} \lambda \lambda_{i_j} + \sum_{i=1}^{m} (1-\lambda)\lambda_{i_k}$$

$$\lambda \sum_{j=1}^{n} \lambda_{i_j} + (1-\lambda) \sum_{k=1}^{m} \lambda_{i_k} = \lambda + (1-\lambda) = 1$$

Since $\lambda \in (0,1)$ was arbitrary, we thus have $\lambda y + (1-\lambda)z \in conv(S), \quad \forall \lambda \in (0,1)$, i.e. $conv(S)$ is convex. $\square$

7.2 Prove that

(i) A hyperplane is convex.

Proof By definition, a hyperplane is a set of the form

$$P = \{x \in V | \langle \mathbf{a, x} \rangle = \mathbf{b}\} \quad \text{where} \quad a \in V, a \neq 0, b \in \mathbb{R}.$$

Take any $y, z \in P$. We want to show that $\lambda y + (1-\lambda)z \in P, \quad \forall \lambda \in (0,1)$. Now note

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

Hence $\lambda y + (1-\lambda)z \in P$, i.e. a hyperplane is convex. □

(ii) A halfspace is convex.

Proof By definition, a halfspace is a set of the form

$$H = \{x \in V | \langle \mathbf{a, x} \rangle \leq \mathbf{b}\} \quad \text{where} \quad a \in V, a \neq 0, b \in \mathbb{R}.$$

Take any $y, z \in H$. We want to show that $\lambda y + (1-\lambda)z \in H, \quad \forall \lambda \in (0,1)$. Now note

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

Hence $\lambda y + (1-\lambda)z \in H$, i.e. a hyperplane is convex. □

7.4 Prove the following Theorem: Let $C \subset \mathbb{R}^n$ be nonempty, closed and convex. A point $p \in C$ is the projection of x onto C if and only if

$$\langle x - p, p - y \rangle \geq 0, \quad \forall y \in C. \qquad (7.14)$$

Prove the statements below and then write a complete proof of the theorem.

(i) $\|x-y\|^2 = \|x-p\|^2 + \|p-y\|^2 + 2\langle x-p, p-y \rangle$

Proof Recall that in $\mathbb{R}^n$, the usual inner product is additive, linear in both arguments, and $\langle x, y \rangle = \langle y, x \rangle$.

$$\|x-y\|^2 = \langle x-y, x-y \rangle = \langle x-p+p-y, x-p+p-y \rangle$$

$$= \langle x-p, x-p \rangle + \langle p-y, x-p \rangle + \langle p-y, x-p \rangle + \langle p-y, p-y \rangle$$
$$= \|x-p\|^2 + \|p-y\|^2 + 2\langle x-p, p-y \rangle \quad \square$$

(ii) If (7.14) holds, then $\|x-y\| > \|x-p\|$ for all $y \in C$, $y \neq p$.

Proof (7.14) states that $\langle x-p, p-y\rangle \geq 0, \quad \forall y \in C$. Further note that for any inner product, $\langle x, x\rangle \geq 0$, with equality iff $x = 0$. Hence, since $y \neq p$, $\|y-p\|^2 > 0$.

Combining the two inequalities, it easily follows from (i) that

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

Since the inner product is non-negative, this shows that $\|x-y\|^2 > \|x-p\|^2$ $\square$

(iii) If $z = \lambda y + (1-\lambda)p$, where $\lambda \in [0,1]$, then

$$\|x-z\|^2 = \|x-p\|^2 + 2\lambda\langle x-p, p-y\rangle + \lambda^2\|y-p\|^2$$

Proof Observe that $p - z = p - \lambda y - (1-\lambda)p = \lambda(p-y)$, so with (i), we have

$$\|x-z\|^2 = \|x-p\|^2 + \|p-z\|^2 + 2\langle x-p, p-z\rangle$$

$$= \|x-p\|^2 + \|\lambda(p-y)\|^2 + 2\langle x-p, \lambda(p-y)\rangle$$
$$= \|x-p\|^2 + 2\lambda\langle x-p, p-y\rangle + \lambda^2\|p-y\|^2 \quad \square$$

.

(iv) If p is a projection of x onto the convex set C, then $\langle x-p, p-y\rangle \geq 0$ for all $y \in C$.

Proof Take $y \in C$. Define $z = \lambda y + (1-\lambda)p$, for some $\lambda \in [0,1]$. Observe by convexity, $z \in C$.

By definition, p is a projection of x onto the convex set C if and only if $\|x-p\| \leq \|x-z\|, \quad \forall z \in C$. By Theorem 7.1.15, the projection on our set unique, so the inequality is strict unless $z = p$.

Note that since from (iii), $\|x-z\|^2 = \|x-p\|^2 + 2\lambda\langle x-p, p-y\rangle + \lambda^2\|y-p\|^2$.

Combining the above, we get that

$$0 \leq \|x-z\|^2 - \|x-p\|^2 = 2\lambda\langle x-p, p-y\rangle + \lambda^2\|y-p\|^2$$

Hence $0 \leq \langle x-p, p-y\rangle + \frac{1}{2}\lambda\|y-p\|^2$. Now, by choosing $\lambda = 0$, we get $0 \leq \langle x-p, p-y\rangle$. $\square$

Theorem of Exercise 7.4 Let $C \subset \mathbb{R}^n$ be nonempty, closed and convex. A point $p \in C$ is the projection of x onto C if and only if

$$\langle x-p, p-y\rangle \geq 0, \quad \forall y \in C. \qquad (7.14)$$

Complete Proof ($\Rightarrow$) The forward direction is (iv); see the proof above.

($\Leftarrow$) By (ii), $\|x-y\| > \|x-p\|$ for all $y \in C$, $y \neq p$. This is the definition of a projection: p is a projection of x onto the convex set C if and only if $\|x-p\| \leq \|x-y\|, \quad \forall y \in C$, which completes the proof (in fact, the backward direction even shows that p is the unique projection). $\square$

7.6 Prove: if $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the set $\{x \in \mathbb{R}^n | f(x) \leq c\} \subset \mathbb{R}^n$ is a convex set.

Proof Let $A = \{x \in \mathbb{R}^n | f(x) \leq c\} \subset \mathbb{R}^n$. Take any $y, z \in A$. We want to show that $\lambda y + (1-\lambda)z \in A, \quad \forall \lambda \in (0,1)$.

Note that the set $\mathbb{R}^n$ is convex. By definition of 'convex function', $\forall y, z \in \mathbb{R}^n$,

$$f(\lambda y + (1-\lambda)z) \leq \lambda f(y) + (1-\lambda)f(z) \leq \lambda c + (1-\lambda)c = c$$

where the first inequality comes from convexity of $\mathbb{R}^n$, and the last follows since we assume $y, z \in A$. A is convex. $\square$

7.7 Prove that any nonnegative combination of convex functions is convex. That is, for any convex set C, for any convex functions $f_1, \cdots, f_k$ taking C to $\mathbb{R}$, and for any $\lambda_1, \cdots, \lambda_k \in \mathbb{R}_+$, the function

$$f(x) = \sum_{i=1}^{k} \lambda_i f_i(x)$$

is convex.

Proof Note that since all f_i are convex, for any convex set C, and for any $x, y \in C$, $\lambda \in [0,1]$, we have

$$f_i(\lambda x + (1-\lambda)y) \leq \lambda f_i(x) + (1-\lambda)f_i(y)$$

Now take arbitrary convex C, $x, y \in C$, $\lambda \in [0,1]$, and observe

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

$$= \lambda \sum_{i=1}^{k} f_i(x) + (1-\lambda) \sum_{i=1}^{k} f_i(y) = \lambda f(x) + (1-\lambda)f(y) \quad \square$$

7.13 If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and bounded above, prove that f is constant.

Proof Suppose f is not constant, i.e. there exist some x_0, x_1 such that $f(x_1) > f(x_0)$, i.e. $f(x_1) - f(x_0) = e > 0$. Without loss of generality, $x_1 > x_0$, i.e. $x_1 - x_0 = d > 0$ For any $x > x_1$, we observe that

$$t = \frac{x - x_1}{x - x_0} \in (0, 1] \quad \text{and} \quad 1 - t = 1 - \frac{x - x_1}{x - x_0} = \frac{x_1 - x_0}{x - x_0} \in [0, 1)$$

.

where the closed sides of the bounds are approached as $x \to \infty$.

I chose this fraction to allows us to represent the middle point (x_1) as a convex combination of the outer ones, as we can see that

$$x_1 = \frac{x - x_0}{x - x_0} x_1 = \frac{xx_1 - xx_0 + xx_0 - x_1 x_0}{x - x_0} = \frac{x_1 - x_0}{x - x_0} x + \frac{x - x_1}{x - x_0} x_0 = (1 - t)x + tx_0$$

such that

$$f(x_1) = f(\frac{x_1 - x_0}{x - x_0} x + \frac{x - x_1}{x - x_0} x_0) \leq \frac{x_1 - x_0}{x - x_0} f(x) + \frac{x - x_1}{x - x_0} f(x_0)$$

Then note that

$$f(x) \geq \frac{x - x_0}{x_1 - x_0} f(x_1) - \frac{x - x_1}{x_1 - x_0} f(x_0) = \frac{x - x_0}{x_1 - x_0} f(x_1) - \frac{x - x_0 + x_0 - x_1}{x_1 - x_0} f(x_0)$$

$$= \frac{x - x_0}{x_1 - x_0}(f(x_1) - f(x_0)) + f(x_0) = \frac{x - x_0}{d} e + f(x_0) > f(x_0)$$

It is obvious that for any finite M as a bound of the function, we could choose a suitable x such that $f(x) > M$ (just to be rigorous, the $x > x_0 + \frac{d}{e}(M - f(x_0)$ will do). Thus, f is unbounded, a contradiction. This proves that, f must be constant. □

7.20 Prove Proposition 7.4.3 : If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $-f$ is also convex, then f is affine.

Proof Since we $f : \mathbb{R}^n \to \mathbb{R}$, we know that affine functions are of the form $f(x) = L(x) + b$, where L(x) is a linear transformation, and $b \in \mathbb{R}$.

By convexity of f, -f, we have that for any $y, z \in \mathbb{R}^n$, and $\forall \lambda \in [0, 1]$,

$$f(\lambda y + (1 - \lambda)z) \leq \lambda f(y) + (1 - \lambda)f(z), \quad -f(\lambda y + (1 - \lambda)z) \leq -\lambda f(y) - (1 - \lambda)f(z)$$

Combining these inequalities yields $f(\lambda y + (1 - \lambda)z) = \lambda f(y) + (1 - \lambda)f(z)$.

We now show that f is affine. Define $g : \mathbb{R}^n \to \mathbb{R}$ as $g(x) = f(x) - f(0)$. Observe:

$$g(0) = f(x) - f(0) = 0, \text{ and } g(\lambda y + (1 - \lambda)z) = \lambda g(y) + (1 - \lambda)g(z) \quad (*)$$

Let y be arbitrary, $z = 0$. Then we have that $g(\lambda y) = \lambda g(y)\ \forall \lambda \in [0, 1]$. This is enough to show that g is a linear function, since we can easily observe that since $x \mapsto \frac{x}{\lambda}$ is a bijection, and $\frac{1}{\lambda} \in [0, 1]$ for $\lambda > 1$, such that for any $y = \frac{x}{\lambda} \in \mathbb{R}^n$, we have that $f(\frac{x}{\lambda}) = \frac{1}{\lambda} f(x)$, i.e. $f(\lambda y) = \lambda f(y)$. Linearity in negative values follows similarly with the bijection $x \mapsto -x$. Thus, combining linearity with $(*)$ yields that for any $y, z \in \mathbb{R}^n$, $a, b \in \mathbb{R}$, $g(ay + bz) = ag(y) + bg(z)$, which is the definition of a linear transformation.

We have thus shown that f(x) = g(x) + f(0), so f is affine. □

7.21 Prove Proposition 7.4.11 : If $D \subset \mathbb{R}$, and $f : \mathbb{R}^n \to D$ is a strictly increasing function, then x^* is a local minimizer for the problem

$$\text{minimize} \quad \phi \circ f(x)$$
$$\text{subject to} \quad G(x) \preceq \mathbf{0}$$
$$H(x) = 0$$

if and only if x^* is a local minimizer for the problem

$$\text{minimize} \quad f(x)$$
$$\text{subject to} \quad G(x) \preceq \mathbf{0}.$$
$$H(x) = 0$$

Proof Note that the constraints stance out a convex set K. We're proving a local property in $\mathbb{R}$, and the constraints are invariant across the two problems.

$\Rightarrow$ Suppose x^* is a minimizer of the second problem. We have that $f(x^*) \leq f(x)$ for all x in some open neighborhood $\mathscr{O} = B_\epsilon(x^*) \cap f(K)$, $\epsilon > 0$ (without loss of generality, since we are in Euclidean space). Since ϕ is strictly increasing, we have $\phi(y) > \phi(z)$ whenever $y > z$. Thus for all $x \in \mathscr{O}$, since $f(x) \geq f(x^*)$, we have $\phi \circ f(x) \geq \phi \circ f(x^*)$. This shows that x^* remains a minimizer in $\mathscr{O}$ under optimization of $\phi \circ f(x)$. $\square$

$\Leftarrow$ Suppose x^* is a minimizer of the first problem. We have that $\phi \circ f(x^*) \leq \phi \circ f(x)$ for all x in some open neighborhood $\mathscr{O}$. For the second problem, there are four mutually exclusive cases we need to distinguish for $\mathscr{O}$:

(i) There are no minimizers in $\mathscr{O}$ and subsets thereof.

(ii) x^* is the local minimizer in some $\mathscr{U} \subset \mathscr{O}$.

(iii) There are multiple, finite minimizers in some $\mathscr{U} \subset \mathscr{O}$; none of them are x^*.

(iv) There are infinitely many minimizers, none of which are x^*.

(i) is impossible since it would imply that for each x in any $\mathscr{U}$, there exist some y, z such that $f(y) < f(x) < f(z)$. If x^* in some $\mathscr{U}$, this would imply by ($\Rightarrow$) that $\phi \circ f(y) < \phi \circ f(x^*) < \phi \circ f(z)$, contradicting our assumptions.

Observe that by shrinking the open set $\mathscr{O}_\epsilon$ to a sufficiently small ϵ, case (iii) can be reduced to case (i).

(iv) is either reduced to (i), or we must observe that for any $\epsilon > 0$, there is a sequence of minimizers converging to x^*. In the latter case, we observe that since for any choice of $\mathscr{U} \ni x^*$, there exists a minimizer x' such that $f(x') \leq f(x)\ \forall x \in \mathscr{U}$. Hence it follows from ($\Rightarrow$) that $\phi \circ f(x') \leq \phi \circ f(x^*)$, and in fact, by our assumption on x^* as a minimizer of the first problem that $\phi \circ f(x') = \phi \circ f(x^*)$. But then since ϕ is strictly increasing, $\phi(y) = \phi(z) \Rightarrow y = z$. Hence $f(x) \geq f(x') = f(x^*)$, for all $x \in \overline{\mathscr{U}}$, and x^* is a minimizer of the second problem. $\square$