Special Topics in Operations Research 26:711:685:01

Convex Analysis and Optimization

Fall 2023 Rutgers University Prof. Eckstein

Solutions to Homework 1

Only one student scored below 87 on this assignment. The most common difficulties, surprisingly, had to do with showing that a set is a linear subspace or that a function is linear. To show that a set $S \subseteq \mathbb{R}^n$ is a linear subspace, you have to show that it is closed under addition and scalar multiplication, that is,

$$\begin{array}{rrr} & (\forall\, x, y \in S) & x + y \in S \\ \text{and} & (\forall\, x \in S)\ (\forall\, c \in \mathbb{R}) & cx \in S. \end{array}$$

To show that a function $h : \mathbb{R}^n \to \mathbb{R}^m$ is linear, one similarly needs to show that

$$\begin{array}{rrr} & (\forall\, x, y \in \mathbb{R}^n) & h(x + y) = h(x) + h(y) \\ \text{and} & (\forall\, x \in \mathbb{R}^n)\ (\forall\, c \in \mathbb{R}) & h(cx) = ch(x). \end{array}$$

These conditions may be respectively condensed into

$$\begin{array}{rr} (\forall\, x, y \in S)\ (\forall\, c \in \mathbb{R}) & cx + y \in S \\ (\forall\, x, y \in \mathbb{R}^n)\ (\forall\, c \in \mathbb{R}) & h(cx + y) = ch(x) + h(y), \end{array}$$

although my experience is that doing so sacrifices clarity and rarely gains much in space.

Many students failed to demonstrate these conditions fully. For example, to show that a function is linear, it is *not* enough to show that

$$(\forall\, x, y \in \mathbb{R}^n)\ (\forall\, c \in \mathbb{R}) \qquad h\big(cx + (1 - c)y\big) = ch(x) + (1 - c)h(y);$$

this only shows that h is affine. If you can show this and also that $h(0) = 0$, then h must be linear, but that was essentially what I was asking you to show and so requires some more details to be filled in.

1. (a) Take any $y, y' \in AC + b$ and $\alpha \in [0, 1]$. If we can show that $\alpha y + (1 - \alpha)y' \in AC + b$ for any such y, y', α, then $AC + b$ must be convex. To that end, we note that we must have $x, x' \in C$ such that $y = Ax + b$ and $y' = Ax' + b$. Then the convexity of C gives that $x_\alpha = \alpha x + (1 - \alpha)x' \in C$, and then we must have $AC + b \ni Ax_\alpha + b = \alpha Ax + (1 - \alpha)Ax' + b = \alpha Ax + \alpha b + (1 - \alpha)Ax' + (1 - \alpha)b = \alpha(Ax + b) + (1 - \alpha)(Ax' + b) = \alpha y + (1 - \alpha)y'$.

   (b) Let $P = \{x \mid Ax + b \in D\}$ be the preimage in question. It is sufficient to show that for any $x, x' \in P$ and $\alpha \in [0, 1]$ that $x_\alpha = \alpha x + (1 - \alpha)x' \in P$. To show this, we note as in the previous question that $Ax_\alpha + b = \alpha Ax + (1 - \alpha)Ax' + b = \alpha Ax + \alpha b + (1 - \alpha)Ax' + (1 - \alpha)b = \alpha(Ax + b) + (1 - \alpha)(Ax' + b)$. Since $x, x' \in P$, we have $Ax + b \in D$ and $Ax' + b \in D$, and since D is convex, we have $Ax_\alpha + b = \alpha(Ax + b) + (1 - \alpha)(Ax' + b) \in D$.

2. Consider any function f obeying (1) for $\alpha \in [0,1]$. We first show that the same equation holds for any $\alpha \in \mathbb{R}$. Suppose we have $z = \alpha x + (1-\alpha)y$ for $\alpha > 1$. We can rearrange this equation into $\alpha x = z + (\alpha - 1)y$ and divide by α to obtain

$$x = \left(\tfrac{1}{\alpha}\right) z + \left(\tfrac{\alpha-1}{\alpha}\right) y.$$

From (1) with the substitution $\alpha \leftarrow 1/\alpha \in [0,1]$, we then obtain

$$f(x) = \left(\tfrac{1}{\alpha}\right) f(z) + \left(\tfrac{\alpha-1}{\alpha}\right) f(y),$$

which we can algebraically manipulate into $f(z) = \alpha f(x) + (1-\alpha)f(y)$, even though $\alpha > 1$.

A similar technique applies if $\alpha < 0$: we write $y = \left(\frac{1}{1-\alpha}\right) z + \left(\frac{-\alpha}{1-\alpha}\right) x$, apply (1), and then apply a reverse series of algebraic manipulations. Thus, we may consider (1) to hold for $\alpha \in \mathbb{R}$.

With this in mind, set $g(x) = f(x) - f(0)$. We show that $g : \mathbb{R}^n \to \mathbb{R}$ must be a linear form. For any $\lambda \in \mathbb{R}$, we have

$$\begin{aligned}
g(\lambda x) &= f(\lambda x) - f(0) \\
&= f(\lambda x + (1-\lambda)0) - f(0) \\
&= \lambda f(x) + (1-\lambda) f(0) - f(0) && \text{[by (1)]} \\
&= \lambda f(x) - \lambda f(0) \\
&= \lambda g(x).
\end{aligned}$$

Now take any $x, y \in \mathbb{R}^n$. We then observe that

$$\begin{aligned}
g(x+y) &= g\big(2\left(\tfrac{1}{2}x + \tfrac{1}{2}y\right)\big) \\
&= 2g\big(\tfrac{1}{2}x + \tfrac{1}{2}y\big) && \text{[since } g(\lambda x) = \lambda g(x)\text{]} \\
&= 2\left(\tfrac{1}{2}g(x) + \tfrac{1}{2}g(y)\right) && \text{[by (1)]} \\
&= g(x) + g(y).
\end{aligned}$$

So g is a linear functional. In $\mathbb{R}^n$, it is then very easy to prove that we must have $g(x) = \langle a, x\rangle$ for some $a \in \mathbb{R}^n$.[1] Setting $b = -f(0)$, we obtain from $g(x) = f(x) - f(0)$ that $f(x) = g(x) + f(0) = \langle a, x\rangle + b$.

The converse result is straightforward. Take $f(x) = \langle a, x\rangle + b$ and any $\alpha \in \mathbb{R}$. Then

$$\begin{aligned}
\alpha f(x) + (1-\alpha) f(y) &= \alpha(\langle a, x\rangle + b) + (1-\alpha)(\langle a, y\rangle + b) \\
&= \alpha\langle a, x\rangle + (1-\alpha)\langle a, y\rangle + \alpha b + (1-\alpha) b \\
&= \langle a, \alpha x + (1-\alpha) y\rangle + b \\
&= f(\alpha x + (1-\alpha) y).
\end{aligned}$$

[1] For those of you familiar with infinite-dimensonal spaces, this result is also true in any Hilbert space by the classic Riesz representation theorem. It may fail in more exotic infinite-dimensonal spaces.

3. Let Y be the set of all convex combinations of points from X. As in class, the convex hull $\operatorname{conv}(X)$ is the intersection of all convex sets containing X. First, we show that Y must be convex. Take any $y, y' \in Y$ and $\alpha \in (0,1)$. By construction, we have

$$y = \sum_{i=1}^{m} \beta_i x_i \qquad\qquad y' = \sum_{i=1}^{m'} \beta_i' x_i'$$

where $\beta_1, \ldots, \beta_m, \beta_1', \ldots, \beta_{m'}' \geq 0$, $x_1, \ldots, x_m, x_1', \ldots, x_{m'}' \in X$, $\sum_{i=1}^{m} \beta_i = 1$, and $\sum_{i=1}^{m'} \beta_i' = 1$. We then write

$$\alpha y + (1-\alpha)y' = \alpha\beta_1 x^1 + \cdots + \alpha\beta_m x^m + (1-\alpha)\beta_1' x_1' + \cdots (1-\alpha)\beta_{m'}' x_{m'}'.$$

Since $x_1, \ldots, x_m, x_1', \ldots, x_{m'}' \in X$ and

$$\begin{aligned}
\alpha\beta_1 + \cdots + \alpha\beta_m + (1-\alpha)\beta_1' + \cdots (1-\alpha)\beta_{m'}' &= \alpha \sum_{i=1}^{m} \beta_i + (1-\alpha) \sum_{i=1}^{m'} \beta_i' \\
&= \alpha \cdot 1 + (1-\alpha) \cdot 1 \\
&= 1,
\end{aligned}$$

it follows that $\alpha y + (1-\alpha)y$ is a convex combination of points from X and is thus in Y. Since y^1, y^2, and α were arbitrary, Y is convex. Since individual points from X may be considered convex combinations with $m = 1$, Y contains X. Thus Y is a convex set containing X, and so $Y \supseteq \operatorname{conv}(X)$.

Conversely, consider any $y \in Y$ and any convex set C containing X. We have $y = \sum_{i=1}^{m} \alpha_i x^i$, with $x_1, \ldots, x_m \in X$, $\alpha_1, \ldots, \alpha_m \geq 0$, and $\sum_{i=1}^{m} \alpha_i = 1$. Since each $x^i \in X$ and $C \supseteq X$, $x^i \in C$ for all i. Thus, y is a convex combination of points from C. As proved in class and on page 35 of the text, we have from the convexity of C that $y \in C$. Since $C \supseteq X$ was arbitrary, y is a member of *all* convex sets containing X, and thus must be a member of $\operatorname{conv}(X)$. Since $y \in Y$ was arbitrary, $Y \subseteq \operatorname{conv}(X)$. In view of the reverse inclusion above, $Y = \operatorname{conv}(X)$.

4. (a) First, let X be an affine set as defined in the text. Take any $x^1, \ldots, x^m \in X$ and $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ with $\sum_{i=1}^{m} \alpha_i = 1$. Then, for each i, we can write $x^i = x + s^i$, where $s^i \in S$, and then

$$\sum_{i=1}^{m} \alpha_i x^i = \sum_{i=1}^{m} \alpha_i (x + s^i) = \left(\sum_{i=1}^{m} \alpha_i \right) x + \sum_{i=1}^{m} \alpha_i s^i = x + \sum_{i=1}^{m} \alpha_i s^i.$$

Since S is a linear subspace, the last summation above is a member of S. Thus $\sum_{i=1}^{m} \alpha_i x^i \in x + S = X$. Thus, X has the property stated in the problem.

Conversely, suppose X has the property described in the problem. To complete the proof, we must show it is of the form $X = x + S$, where S is a linear subspace. If X is empty, the result is vacuously true, so long as one considers $\emptyset$ to be a linear subspace. If X is nonempty, take some $x \in X$ and set $S = X - x = \{x' - x \mid x' \in X\}$. It is immediate that $X = x + S$, so it remains only to show that S is a linear subspace. Consider any point $s \in S$, which must be of the form $s = y - x$, where $y \in X$. Then, for any $\lambda \in \mathbb{R}$,

$$\lambda s = \lambda(y - x) = \lambda y - \lambda x = \lambda y + (1 - \lambda)x - x.$$

Since x and y are both in X, the assumption on X implies $y' = \lambda y + (1 - \lambda)x \in X$. Thus, λs is of the form $y' - x$ for $y' \in X$, and so $\lambda s \in S$. Next, consider $s, t \in S$; we would like to prove $s + t \in S$. By construction, there exist $y, z \in X$ such that $s = y - x$ and $t = z - x$. Then $s + t = y - x + z - x$. Since $x, y, z \in X$, we have, noting that $1 + (-1) + 1 = 1$, that $w \doteq y - x + z = 1y + (-1)x + 1z$ is an affine combination of points from X, and thus in X. So, $s + t = w - x$, where $w \in X$, and so $s + t \in S$. Together with the previous result, we conclude that S is a linear subspace.

(b) From this point, we can proceed much as in question 3. Let Z denote the set of all affine combinations of elements of Y (meaning that Z is the set on the right in displayed equation for this part of the assignment). Consider any affine set X containing Y. From part (a), X contains all affine combinations of its elements, and in particular all affine combinations from Y. Therefore, $X \supseteq Z$. Furthermore, since the affine set $X \supseteq Y$ was arbitrary, Z is contained in *all* affine sets containing Y, and we have $Z \subseteq \text{aff}(Y)$.

To complete the proof, we will show that Z is an affine set. This, along with the obvious fact that $Z \supseteq Y$, establishes that $Z \supseteq \text{aff}(Y)$, since $\text{aff}(Y)$ is the intersection of all affine sets containing Y. In view of the opposite inclusion above, we then conclude $Z = \text{aff}(Y)$.

To show that Z is affine, we consider any affine combination $w = \alpha_1 z^1 + \cdots + \alpha_m z^m$ of points $z^1, \ldots, z^m \in Z$, where $\alpha_1 + \cdots + \alpha_m = 1$. If we can show that any such w is in Z, then part (a) will assert that Z is affine. Now, for all $i = 1, \ldots, m$, we have from the construction of Z that

$$z^i = \sum_{j=1}^{n_i} \beta_{ij} y^{ij}, \qquad y_{i1}, \ldots, y_{in_i} \in Y \qquad \sum_{j=1}^{n_i} \beta_{ij} = 1.$$

Thus, we can write

$$w = \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{n_i} \beta_{ij} y^{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\alpha_i \beta_{ij}) y^{ij}.$$

Noting that

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} \alpha_i \beta_{ij} = \sum_{i=1}^{m} \alpha_i \left(\sum_{j=1}^{n_i} \beta_{ij} \right) = \sum_{i=1}^{m} \alpha_i \cdot 1 = \sum_{i=1}^{m} \alpha_i = 1,$$

it is clear that w is an affine combination of the points $y^{ij} \in Y$, and is hence a member of Z.

5. As suggested in the hint, consider the function $f : (0, \infty) \to \mathbb{R}$ given by $f(x) = -\log x$ (I will use "log" to stand for the natural logarithm). From elementary calculus, we find that

$f''(x) = 1/x^2$, which is positive for all $x > 0$. Using Proposition 1.2.6 with $n = 1$ and $C = (0, \infty)$, we conclude that f is strictly convex over $(0, \infty)$. Jensen's inequality, formula (1.7) from the text, with $n = 1$ and $X = (0, \infty)$, tells us that

$$f\left(\sum_{i=1}^{m} \alpha_i x_i\right) \leq \sum_{i=1}^{m} \alpha_i f(x_i).$$

Substituting the definition of f and multiplying by -1, we obtain

$$\log\left(\sum_{i=1}^{m} \alpha_i x_i\right) \geq \sum_{i=1}^{m} \alpha_i \log x_i.$$

Applying the monotonic function e^x to both sides of this inequality produces

$$\sum_{i=1}^{m} \alpha_i x_i \geq \prod_{i=1}^{m} x_i^{\alpha_i},$$

which is equivalent to the desired result. In the construction of Jensen's inequality, it can also be seen that if f is strictly convex, then the inequality will be strict unless $x_1 = \cdots = x_m$.