CONVEX AND NONSMOOTH OPTIMIZATION: HW 1

TERRENCE ALSUP

(1) Let (x_1, t_1) and (x_2, t_2) be any two points in the quadratic cone and let $\theta \in (0,1)$ be arbitrary. By the triangle inequality we have

$$\|\theta x_1 + (1-\theta)x_2\|_2 \leq \theta\|x_1\|_2 + (1-\theta)\|x_2\|_2$$

where we have also used the fact that θ and $1-\theta$ are non-negative (to pull them out of the norm). However, since (x_1, t_1) and (x_2, t_2) are in the quadratic cone we have by definition that

$$\theta\|x_1\|_2 + (1-\theta)\|x_2\|_2 \leq \theta t_1 + (1-\theta)t_2$$

and since $\theta t_1 + (1-\theta)t_2 \geq 0$ we have that

$$\theta \begin{bmatrix} x_1 \\ t_1 \end{bmatrix} + (1-\theta) \begin{bmatrix} x_2 \\ t_2 \end{bmatrix}$$

is also in the quadratic cone. Hence, the quadratic cone is convex.

(2) Let A, B be any two convex sets and let $C = A \cap B$ be the intersection. Suppose $x_1, x_2 \in C$ and let $\theta \in (0,1)$ be arbitrary. We have that $x_1, x_2 \in A, B$ both. Since A, B are convex we have that $\theta x_1 + (1-\theta)x_2 \in A, B$ both as well. Hence $\theta x_1 + (1-\theta)x_2 \in C$ and C is still convex.

(3) Let S be a convex set and $f(y) = Ay + b$ an affine function. We are to show that $f(S)$ and $f^{-1}(S)$ are still convex sets. First let $x_1, x_2 \in f(S)$ and $\theta \in (0,1)$. There exist points $y_1, y_2 \in S$ (not necessarily unique) so that $f(y_1) = x_1$ and $f(y_2) = x_2$. Since $y_1, y_2 \in S$ we have that $\theta y_1 + (1-\theta)y_2 \in S$ by convexity of S. Then,

$$\begin{aligned} f\left(\theta y_1 + (1-\theta)y_2\right) &= A(\theta y_1 + (1-\theta)y_2) + b \\ &= \theta(Ay_1 + b) + (1-\theta)(Ay_2 + b) \\ &= \theta x_1 + (1-\theta)x_2 \in f(S) \end{aligned}$$

Date: February 4, 2020.

Hence $f(S)$ is convex.

Now we check that the inverse image is convex. Let $x_1, x_2 \in f^{-1}(S)$ so there exists $y_1, y_2 \in S$ such that $f(x_1) = y_1$ and $f(x_2) = y_2$. For any $\theta \in (0, 1)$ we have

$$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) &= \theta(Ax_1 + b) + (1-\theta)(Ax_2 + b) \\ &= \theta y_1 + (1-\theta)y_2 \end{aligned}$$

Since S is convex we have $\theta y_1 + (1-\theta)y_2 \in S$ and therefore, $\theta x_1 + (1-\theta)x_2 \in f^{-1}(S)$. Thus, $f^{-1}(S)$ is convex as well.

(4) Let C be a convex set. We know from the definition of convexity that when $k = 2$ we have $\theta_1 x_1 + \theta_2 x_2 \in C$ when $\theta_1 + \theta_2 = 1$ and $\theta_1, \theta_2 \geq 0$. In particular, $\theta_1 \in [0, 1]$ and $\theta_2 = 1 - \theta_1$. Now suppose it holds for some integer k that for $x_1, \ldots, x_k \in C$ and $\theta_1, \ldots, \theta_k \geq 0$ with $\theta_1 + \cdots + \theta_k = 1$ we have $\theta_1 x_1 + \cdots + \theta_k x_k \in C$. Now let $x_1, \ldots, x_{k+1} \in C$ and suppose that $\theta_1, \ldots, \theta_{k+1} \geq 0$ and $\theta_1 + \cdots + \theta_{k+1} = 1$. We can write

$$\begin{aligned} \sum_{i=1}^{k+1} \theta_i x_i &= \theta_{k+1} x_{k+1} + \sum_{i=1}^{k} \theta_i x_i \\ &= \theta_{k+1} x_{k+1} + \left(\frac{\sum_{i=1}^k \theta_i}{\sum_{i=1}^k \theta_i} \right) \sum_{i=1}^{k} \theta_i x_i \end{aligned}$$

Note that we have just multiplied by 1 in the second line. If $\sum_{i=1}^k \theta_i = 0$ then $\theta_{k+1} = 1$ and x_{k+1} is obviously in C already. Therefore, we only consider $\sum_{i=1}^k \theta_i > 0$. We have

$$\sum_{i=1}^{k} \frac{\theta_i}{\sum_{i=1}^k \theta_i} x_i$$

is a convex combination however, since

$$\sum_{i=1}^{k} \frac{\theta_i}{\sum_{i=1}^k \theta_i} = 1, \qquad \text{and} \qquad \frac{\theta_i}{\sum_{i=1}^k \theta_i} \geq 0$$

By our inductive hypothesis, $y := \sum_{i=1}^k \frac{\theta_i}{\sum_{i=1}^k \theta_i} x_i \in C$. We now have

$$\sum_{i=1}^{k+1} \theta_i x_i = \theta_{k+1} x_{k+1} + \left(\sum_{i=1}^{k} \theta_i \right) y$$

Since $\theta_k + \left(\sum_{i=1}^k \theta_i \right) = 1$, $\theta_{k+1}, \left(\sum_{i=1}^k \theta_i \right) \geq 0$, and $x_{k+1}, y \in C$ we have that $\sum_{i=1}^{k+1} \theta_i x_i \in C$, which was to be shown.

(5) We show both results (for affine and convex sets) simultaneously since the proof is the same. First suppose that $C \subset \mathbb{R}^n$ is a convex (affine) set and let $L := \{tv + b : t \in \mathbb{R}\}$, where $v, b \in \mathbb{R}^n$, be an arbitrary line. If $x_1, x_2 \in L \cap C$ then we can write $x_1 = t_1 v + b$ and $x_2 = t_2 v + b$ for some $t_1, t_2 \in \mathbb{R}$. For any $\theta \in [0,1]$ ($\theta \in \mathbb{R}$) we have

$$\theta x_1 + (1-\theta)x_2 = \theta(t_1 v + b) + (1-\theta)(t_2 v + b) = (\theta t_1 + (1-\theta)t_2)v + b$$

Since $(\theta t_1 + (1-\theta)t_2) \in \mathbb{R}$ we have that $\theta x_1 + (1-\theta)x_2 \in L$. Moreover, since C is convex (affine) we have $\theta x_1 + (1-\theta)x_2 \in C$. Thus, $L \cap C$ is convex (affine).

We now show the reverse. Suppose that $L \cap C$ is convex (affine) for every line L. Let $x_1, x_2 \in C$ with $x_1 \neq x_2$. Let L be the line $L = \{(x_1 - x_2)t + x_2 : t \in \mathbb{R}\}$ that goes through x_1 and x_2. For any $\theta \in [0,1]$ ($\theta \in \mathbb{R}$ if $L \cap C$ is affine) we have

$$\theta x_1 + (1-\theta)x_2 = (x_1 - x_2)\theta + x_2 \in L$$

Since $x_1, x_2 \in L \cap C$ and $L \cap C$ is convex (affine) we must have $\theta x_1 + (1-\theta)x_2 \in L \cap C$ and therefore, $\theta x_1 + (1-\theta)x_2 \in C$. Thus, the set C is convex (affine). Therefore, a set is convex (affine) if and only if its intersection with every line is convex (affine).

(6) (a) Suppose that $A \succeq 0$. To make the problem simpler we will use the fact proven in the previous problem which is that C is convex if the intersection of C with every line is convex. Let $L = \{vt + v_0 : t \in \mathbb{R}\}$, with $v, v_0 \in \mathbb{R}^n$, be an arbitrary line. We have

$$L \cap C = \left\{vt + v_0 : (vt + v_0)^T A(vt + v_0) + b^T(vt + v_0) + c \leq 0\right\}$$

For brevity write

$$\phi(t) := (vt+v_0)^T A(vt+v_0)+b^T(vt+v_0)+c = t^2(v^T Av)+t(2v^T Av_0+b^T v)+(v_0^T Av_0+b^T v_0+c)$$

so that $L \cap C = \{vt + v_0 : \phi(t) \leq 0\}$. We are to show that $L \cap C$ is convex so suppose that $x_1, x_2 \in L \cap C$. There exist t_1, t_2 such that $x_1 = vt_1 + v_0$ and $x_2 = vt_2 + v_0$. For any $\theta \in [0,1]$ we have

$$\theta x_1 + (1-\theta)x_2 = \theta(vt_1 + v_0) + (1-\theta)(vt_2 + v_0) = v(\theta t_1 + (1-\theta)t_2) + v_0$$

Thus, we just need to show that $\phi(\theta t_1 + (1-\theta)t_2) \leq 0$. Note that $\phi(t)$ is just a quadratic function whose leading coefficient is nonnegative by the assumption that $A \succeq 0$. Since $\phi(t_1), \phi(t_2) \leq 0$ we must have that $\phi(s) \leq 0$ for any $s \in [t_1, t_2]$. In particular, this holds for $s = \theta t_1 + (1-\theta)t_2$ and so $\phi(\theta t_1 + (1-\theta)t_2) \leq 0$ meaning that $\theta x_1 + (1-\theta)x_2 \in L \cap C$ and C is convex.

(b) We will use the same approach as in the previous part which is to show that the intersection with any line is convex. Let H denote the hyperplane $H = \{x \in \mathbb{R}^n : g^T x + h = 0\}$. If $L = \{vt + v_0 : t \in \mathbb{R}\}$ is again an arbitrary line we have that

$$L \cap C \cap H = \{vt + v_0 : t^2(v^T Av) + t(2v^T Av_0 + b^T v) + (v_0^T Av_0 + b^T v_0 + c) \leq 0, \\ g^T(vt + v_0) + h = 0\}$$

Note that

$$\lambda(g^T(vt + v_0) + h)^2 = \lambda t^2(g^T v)^2 + 2t\lambda(g^T v)(g^T v_0 + h) + \lambda(g^T v_0 + h)^2 = 0$$

Since $g^T v$ is a scalar we have $g^T v = v^T g$ so $(g^T v)^2 = v^T gg^T v$. Since the whole expression is 0 we can add it to the other constraint.

$$L \cap H \cap C = \{vt + v_0 : t^2(v^T Av) + t(2v^T Av_0 + b^T v) + (v_0^T Av_0 + b^T v_0 + c) \\ + \lambda t^2(g^T v)^2 + 2t\lambda(g^T v)(g^T v_0 + h) + \lambda(g^T v_0 + h)^2 \leq 0\}$$

Define

$$\phi(t) = t^2(v^T Av) + t(2v^T Av_0 + b^T v) + (v_0^T Av_0 + b^T v_0 + c) \\ + \lambda t^2(g^T v)^2 + 2t\lambda(g^T v)(g^T v_0 + h) + \lambda(g^T v_0 + h)^2$$

so that $\phi(t)$ is a quadratic function with leading coefficient $v^T Av + \lambda v^T gg^T v = v^T(A + \lambda gg^T)v$. Since $A + \lambda gg^T \succeq 0$ we know that the leading coefficient is positive just as in the previous problem. Now let $x_1, x_2 \in L \cap H \cap C$ so that we can write $x_1 = vt_1 + v_0$ and $x_2 = vt_2 + v_0$ for some $t_1, t_2 \in \mathbb{R}$. Therefore, $\phi(t_1), \phi(t_2) \leq 0$ and by the same argument as before we have $\phi(\theta t_1 + (1-\theta)t_2) \leq 0$ meaning that $\theta x_1 + (1-\theta)x_2 \in L \cap H \cap C$. Hence $H \cap C$ is convex.

(7) Let $z, \tilde{z} \in S$ so $z = (x, y_1 + y_2)$ for $(x, y_1) \in S_1$ and $(x, y_2) \in S_2$ and similarly, $\tilde{z} = (\tilde{x}, \tilde{y}_1 + \tilde{y}_2)$ for $(\tilde{x}, \tilde{y}_1) \in S_1$ and $(\tilde{x}, \tilde{y}_2) \in S_2$. Let $\theta \in [0, 1]$. We have that

$$\theta z + (1-\theta)\tilde{z} = \begin{bmatrix} \theta x + (1-\theta)\tilde{x} \\ \theta y_1 + (1-\theta)\tilde{y}_1 + \theta y_2 + (1-\theta)\tilde{y}_2 \end{bmatrix}$$

To simplify notation, let $x' = \theta x + (1-\theta)\tilde{x}$, $y_1' = \theta y_1 + (1-\theta)\tilde{y}_1$, and $y_2' = \theta y_2 + (1-\theta)\tilde{y}_2$. Then,

$$\theta z + (1-\theta)\tilde{z} = \begin{bmatrix} x' \\ y_1' + y_2' \end{bmatrix}$$

and therefore, we just need to show that $(x', y_1') \in S_1$ and $(x', y_2') \in S_2$. However, S_1, S_2 are convex and so

$$\begin{bmatrix} x' \\ y_1' \end{bmatrix} = \theta \begin{bmatrix} x \\ y_1 \end{bmatrix} + (1-\theta) \begin{bmatrix} \tilde{x} \\ \tilde{y}_1 \end{bmatrix} \in S_1$$

and similarly,

$$\begin{bmatrix} x' \\ y_2' \end{bmatrix} = \theta \begin{bmatrix} x \\ y_2 \end{bmatrix} + (1-\theta) \begin{bmatrix} \tilde{x} \\ \tilde{y}_2 \end{bmatrix} \in S_2$$

Thus, S is convex.

(8) Consider the following two sets.

$$S_1 := \left\{(x, y) \in \mathbb{R}^2 : y \leq 0\right\}$$
$$S_2 := \left\{(x, y) \in \mathbb{R}^2 : y \geq e^{-x}\right\}$$

Both sets are closed, and they are disjoint because $e^{-x} > 0$ for all $x \in \mathbb{R}$. The first set is convex since it is just the lower half-plane and the second set is convex because it is the epigraph of a convex function (see BV pg. 75 at the bottom). The point is that these two sets become arbitrarily close so they cannot be strictly separated. Suppose by way of a contradiction that we could strictly separate them so that there are $a_1, a_2, b \in \mathbb{R}$ with $(a_1, a_2) \neq (0, 0)$ such that $a_1 x + a_2 y + b > 0$ for all $(x, y) \in S_1$ and $a_1 x + a_2 y + b < 0$ for all $(x, y) \in S_2$. We must necessarily have $a_1 = 0$ otherwise the line will intersect the lower half-plane S_1 and these sets will not separate S_1 and S_2. Therefore, the separating hyperplane would have to be of the form $y = c$. However, for any $c \leq 0$ we intersect S_1 and for any $c > 0$ we intersect S_2. Thus, these sets cannot be strictly separated.

(9) The set C is essentially just an n-dimensional cube. For each point $\hat{x}$ on the boundary of C, define a vector $v \in \{-1, 0, 1\}^n$ by

$$v_i = \begin{cases} 1 & \hat{x}_i = 1, \\ 0 & |\hat{x}_i| < 1, \\ -1 & \hat{x}_i = -1. \end{cases}$$

This vector just indicates which face, edge, or corner $\hat{x}$ is on. Since $\hat{x}$ is on the boundary we have that $v \neq \vec{0}$ and that

$$v^T \hat{x} = \sum_{i=1}^{n} v_i \hat{x}_i = \sum_{i=1}^{n} |v_i| \geq \sum_{i=1}^{n} v_i x_i$$

for all $x \in C$ (since $\|x\|_\infty \leq 1$ if $x \in C$). Therefore, the sets $\{x \in \mathbb{R}^n : v^T\hat{x} = v^T x\}$ are supporting hyperplanes for C at $\hat{x}$. We now show that these are the only ones. Let $a \in \mathbb{R}^n$ with $a^T\hat{x} \geq a^T x$ for all $x \in C$. Without loss of generality we may assume that $\|a\|_\infty = 1$ since otherwise we could just divide both sides by $\|a\|_\infty$. We have by definition that

$$\sum_{i=1}^n a_i\hat{x}_i = a^T\hat{x} \geq \sup_{x\in C} a^T x = \sup_{x\in C}\sum_{i=1}^n a_i x_i = \sum_{i=1}^n |a_i|$$

since we can take $x_i = \operatorname{sgn}(a_i)$. In particular this tells us that

$$\sum_{i=1}^n a_i\hat{x}_i = \sum_{i=1}^n |a_i|$$

since $\hat{x} \in C$ as well. Therefore, if $|\hat{x}_i| < 1$ we must take $a_i = 0$ and if $|x_i| = 1$ we have to take $a_i = \operatorname{sgn}(x_i)$. Thus, we have recovered the vector v and these are all of the supporting hyperplanes for C at $\hat{x}$.

(10) Suppose that x is in the hyperbolic cone, then $f(x) = (P^{1/2}x, c^T x) =: (z, t)$. We have $t = c^T x \geq 0$ and

$$z^T z = x^T(P^{1/2})^T P^{1/2}x = x^T Px \leq (c^T x)^2 = t^2$$

where we have used the fact that $P \in \mathbb{S}^n_+$ and therefore has a symmetric square root. Thus, $f(x)$ is in the quadratic cone and the hyperbolic cone is a subset of the inverse image of the second-order cone under the given affine transformation. Now we show the reverse. Let x be such that $f(x) = (P^{1/2}x, c^T x) =: (z, t)$ is in the second-order cone. In other words, x is in the inverse image. We have

$$x^T Px = (P^{1/2}x)^T(P^{1/2}x) = z^T z \leq t^2 = (c^T x)^2$$

Since $t = c^T x \geq 0$ as well, we know that x is in the hyperbolic cone. Therefore, the hyperbolic cone is exactly the inverse image of the second-order cone under the given affine transformation.