Solution 4: 19 March 2019
CS-526 Learning Theory

Exercise 1

1. $f(x) = \max_{1\le i\le m} f_i(\mathbf{x})$ where $f_i(\mathbf{x}) = \mathbf{a}_i^T\mathbf{x} + b_i$ is convex differentiable with gradient $\nabla f_i(\mathbf{x}) = \mathbf{a}_i$. By Claim 14.6, it follows that $\forall \mathbf{x} : \mathbf{a}_j \in \partial f(\mathbf{x})$ where $j \in \arg\max_i f_i(\mathbf{x})$.

2. $f(x) = \max_{1\le i\le m} f_i(\mathbf{x})$ where $f_i(\mathbf{x}) = |\mathbf{a}_i^T\mathbf{x} + b_i|$ is convex subdifferentiable. Fix $\mathbf{x}$, let $j \in \arg\max_i f_i(\mathbf{x})$ and choose $\mathbf{v} \in \partial f_j(\mathbf{x})$ as follows:

$$\mathbf{v} = \begin{cases} -\mathbf{a}_j & \text{if } \mathbf{a}_j^T\mathbf{x} + b_j < 0\,, \\ 0 & \text{if } \mathbf{a}_i^T\mathbf{x} + b_i = 0\,, \\ +\mathbf{a}_j & \text{if } \mathbf{a}_j^T\mathbf{x} + b_j > 0\,. \end{cases}$$

A straightforward generalization of Claim 14.6 shows that $\mathbf{v}$ is a subgradient of f at $\mathbf{x}$.

3. Note that the sup is really a maximum as $t \mapsto p(t,\mathbf{x})$ is a continuous function on a compact. Hence $f(\mathbf{x}) = \max_{t\in[0,1]} p(t,\mathbf{x})$ and $\forall t \in [0,1] : \nabla_{\mathbf{x}} p(t,\mathbf{x}) = [1, t, \dots, t^{n-1}]^T \in \mathbb{R}^n$. A straightforward generalization of Claim 14.6 shows that $[1, t(\mathbf{x}), \dots, t(\mathbf{x})^{n-1}]^T \in \partial f(\mathbf{x})$, where $t(\mathbf{x}) \in \arg\max_{t\in[0,1]} p(t,\mathbf{x})$.

Exercise 2

1. v is a subgradient of f at 0 if $\forall u > 0 : f(u) \ge f(0) + (u-0)v$, i.e.,

$$\forall u > 0 : 0 \ge 1 + uv\,. \tag{1}$$

Clearly v must be negative for the later to hold, and if v is negative then $0 \ge 1 + uv \Leftrightarrow u \ge 1/|v|$. Whatever v, (1) cannot hold on the whole interval $[0, +\infty)$. Hence f is not subdifferentiable at 0.
2. v is a subgradient of f at 0 if $\forall u > 0 : f(u) \ge f(0) + (u-0)v$, i.e.,

$$\forall u > 0 : -1 \ge \sqrt{u}v\,. \tag{2}$$

Clearly v must be negative for the later to hold, and if v is negative then $-1 \ge \sqrt{u}v \Leftrightarrow u \ge 1/v^2$. Whatever v, (2) cannot hold on the whole interval $[0, +\infty)$. Hence f is not subdifferentiable at 0.

Exercise 3

Fix $\mathbf{w}$, $\mathbf{u}$. The function f is λ-strongly convex, so for all $\alpha \in [0,1]$ we have:

$$
\begin{aligned}
& f((1-\alpha)\mathbf{w} + \alpha\mathbf{u}) \le (1-\alpha) f(\mathbf{w}) + \alpha f(\mathbf{u}) - \frac{\lambda}{2}\alpha(1-\alpha)\|\mathbf{w}-\mathbf{u}\|^2 \\
\Leftrightarrow \quad & f(\mathbf{w} + \alpha(\mathbf{u}-\mathbf{w})) - f(\mathbf{w}) \le \alpha\Big(f(\mathbf{u}) - f(\mathbf{w}) - \frac{\lambda}{2}(1-\alpha)\|\mathbf{w}-\mathbf{u}\|^2\Big) \qquad (3)
\end{aligned}
$$

Let $\mathbf{v} \in \partial f(\mathbf{w})$. Then, $\forall \alpha \in [0,1] : f(\mathbf{w} + \alpha(\mathbf{u}-\mathbf{w})) \ge f(\mathbf{w}) + \langle \alpha(\mathbf{u}-\mathbf{w}), \mathbf{v}\rangle$. Combining this inequality and (3) gives:

$$
\begin{aligned}
& \langle \alpha(\mathbf{u}-\mathbf{w}), \mathbf{v}\rangle \le \alpha\Big(f(\mathbf{u}) - f(\mathbf{w}) - \frac{\lambda}{2}(1-\alpha)\|\mathbf{w}-\mathbf{u}\|^2\Big) \\
\Leftrightarrow \quad & \langle \mathbf{u}-\mathbf{w}, \mathbf{v}\rangle \le f(\mathbf{u}) - f(\mathbf{w}) - \frac{\lambda}{2}(1-\alpha)\|\mathbf{w}-\mathbf{u}\|^2 \\
\Leftrightarrow \quad & \langle \mathbf{w}-\mathbf{u}, \mathbf{v}\rangle \ge f(\mathbf{w}) - f(\mathbf{u}) + \frac{\lambda}{2}(1-\alpha)\|\mathbf{w}-\mathbf{u}\|^2
\end{aligned}
$$

Taking the limit $\alpha \to 0+$ ends the proof: $\langle \mathbf{w}-\mathbf{u}, \mathbf{v}\rangle \ge f(\mathbf{w}) - f(\mathbf{u}) + \frac{\lambda}{2}\|\mathbf{w}-\mathbf{u}\|^2$.

Exercise 4

To prove that $\pi_C(\cdot)$ is Lipschiztian, we first show an important property of projection onto a closed convex set:

Lemma 1. *If C is a non-empty closed convex subset of a Hilbert space H then $\forall(\mathbf{x},\mathbf{y}) \in H \times C : \langle \mathbf{x} - \pi_C(\mathbf{x}), \mathbf{y} - \pi_C(\mathbf{x})\rangle \le 0$.*

Proof. Let $\alpha \in (0,1)$. By definition of $\pi_C(\cdot)$, we have:

$$
\begin{aligned}
0 &\le \|\mathbf{x} - (1-\alpha)\pi_C(\mathbf{x}) - \alpha\mathbf{y}\|^2 - \|\mathbf{x} - \pi_C(\mathbf{x})\|^2 \\
&= \|\mathbf{x} - \pi_C(\mathbf{x}) - \alpha(\mathbf{y} - \pi_C(\mathbf{x}))\|^2 - \|\mathbf{x} - \pi_C(\mathbf{x})\|^2 \\
&= \alpha^2\|\mathbf{y} - \pi_C(\mathbf{x})\|^2 - 2\alpha\langle \mathbf{x} - \pi_C(\mathbf{x}), \mathbf{y} - \pi_C(\mathbf{x})\rangle .
\end{aligned}
$$

Dividing the final inequality by α and taking the limit $\alpha \to 0$ ends the proof. □

We can now prove that $\pi_C(\cdot)$ is 1-Lipschitz. $\forall \mathbf{x}_0, \mathbf{x}_1$:

$$
\begin{aligned}
\|\pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\|^2 &= \langle \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1), \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\rangle \\
&= \underbrace{\langle \pi_C(\mathbf{x}_0) - \mathbf{x}_0, \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\rangle}_{\le 0} + \langle \mathbf{x}_0 - \pi_C(\mathbf{x}_1), \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\rangle \\
&\le \langle \mathbf{x}_0 - \pi_C(\mathbf{x}_1), \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\rangle \\
&\le \underbrace{\langle \mathbf{x}_1 - \pi_C(\mathbf{x}_1), \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\rangle}_{\le 0} + \langle \mathbf{x}_0 - \mathbf{x}_1, \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\rangle \\
&\le \langle \mathbf{x}_0 - \mathbf{x}_1, \pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\rangle \\
&\le \|\mathbf{x}_0 - \mathbf{x}_1\|\|\pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\| \qquad \text{(Cauchy-Schwarz inequality)}
\end{aligned}
$$

It directly implies $\|\pi_C(\mathbf{x}_0) - \pi_C(\mathbf{x}_1)\| \le \|\mathbf{x}_0 - \mathbf{x}_1\|$. Note that for $\mathbf{x}_0, \mathbf{x}_1 \in C$ this inequality is an equality, hence the it cannot be improved.