Idan Drori
Ohad Sapir

Question #1

Solution:
($\Rightarrow$) We assume that $x^* \in \underset{x\in S}{\arg\min} f(x)$. Let $x \in S$, and let $w = x - x^*$. From differentiability of f we get:

$$\nabla f(x^*)^T \cdot w = \frac{\partial f}{\partial w}(x^*) = \lim_{h\to 0}\frac{f(x^* + h\cdot w) - f(x^*)}{h}$$

Because $f(x^*)$ is the minimum, $f(x) - f(x^*) \geq 0$ for every $x \in S$ and so it holds for $x = x^* + hw \in S$. Meaning:

$$\nabla f(x^*) \cdot w = \lim_{h\to 0}\frac{f(x^* + h\cdot w) - f(x^*)}{h} \geq 0$$

($\Leftarrow$) We assume that $\forall x \in S, \nabla f(x^*)^T(x - x^*) \geq 0$. Then from convexity of f we get:

$$\forall x \in S, f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*) \geq f(x^*)$$

Meaning $x^* \in \underset{x\in S}{\arg\min} f(x)$ $\square$

Question #2

(a) $S = \{x : \|x\|_2 \leq 1\}$

Solution: Let $y \in \mathbb{R}^d, y^* = \frac{1}{\max\{1,\|y\|_2\}}y$.
If $\|y\|_2 \leq 1$, then $y \in S$ and

$$y^* = \frac{1}{\max\{1, \|y\|_2\}}y = \frac{1}{1}y = y$$

So $y^* \in S$ and $y^* = \arg\min_{x\in S}\|y - x\|$.
Else, $\|y\|_2 > 1$ and $y^* = \frac{y}{\|y\|_2}$.
$\|y^*\| = \|\frac{y}{\|y\|}\| = \frac{\|y\|}{\|y\|} = 1$, then $y^* \in S$
Let $x \in S$.

$$\|y - x\|^2 - \|y - y^*\|^2 = \|y - x\|^2 - \|y - \frac{y}{\|y\|}\|^2$$
$$= \|y - x\|^2 - \left(1 - \frac{1}{\|y\|}\right)^2\|y\|^2 = \|y\|^2 - 2y\cdot x + \|x\|^2 - (\|y\|^2 - 2\|y\| + 1)$$
$$= \|x\|^2 - 2y\cdot x + 2\|y\| - 1 \geq \|x\|^2 - 2\|y\|\|x\| + 2\|y\| - 1$$
$$= \|x\|^2 + 2\|y\|(1 - \|x\|) - 1 \geq \|x\|^2 + 2(1 - \|x\|) - 1$$
$$= \|x\|^2 - 2\|x\| + 1 = (\|x\| - 1)^2 \geq 0$$

So $y^* = \arg\min_{x \in S} \|y - x\|$.
Meaning $\Pi_S[y] = y^* = \frac{1}{\max\{1, \|y\|_2\}} y$.

$\square$

(b) $S = \{x : \|x\|_\infty \leq 1\}$

Solution: Let $y \in \mathbb{R}^d, y_i^* = \frac{1}{\max\{1,|y_i|\}} y_i, x \in S$.
Let $i \in [d]$. If $|y_i| \leq 1$, then $|y_i^*| = |y_i| \leq 1$

$$|y_i - y_i^*| = |y_i - \frac{1}{1} y_i| = 0 \leq |y_i - x_i|$$

Else, $|y_i| > 1$ and $|y_i^*| = |\frac{y_i}{y_i^*}| = 1$.
Therefore, $\|y^*\| = \max_{i \in [d]} |y_i^*| \leq 1$. So $y^* \in S$.
In addition

$$\begin{aligned}
&|y_i - x_i| - |y_i - y_i^*| = |y_i - x_i| - |y_i - \frac{y_i}{|y_i|}| = |y_i - x_i| - |y_i(1 - \frac{1}{|y_i|})| \\
&= |y_i - x_i| - \left| \left(1 - \frac{1}{|y_i|}\right) \right| |y_i| = |y_i - x_i| - \left(1 - \frac{1}{|y_i|}\right) |y_i| \\
&\geq ||y_i| - |x_i|| - \left(|y_i| - \frac{|y_i|}{|y_i|}\right) = |y_i| - |x_i| + 1 - |y_i| \\
&= 1 - |x_i| \geq 0
\end{aligned}$$

So $\forall i \in [d]. |y_i - y_i^*| \leq |y_i - x_i|$, meaning

$$\|y - x\| - \|y - y^*\| = \max_{i \in [d]} |y_i - x_i| - \max_{i \in [d]} |y_i - y_i^*| \geq 0$$

So $y^* = \arg\min_{x \in S} \|y - x\|$.
Meaning $\Pi_S[y] = y^* = \frac{1}{\max\{1,|y_i|\}} y_i$.

$\square$

Question #3

(a) *Solution:* Let $S \subseteq \mathbb{R}^d$ be a convex set such that $Q \subseteq S$. Let $x \in conv(Q)$. There exist $x_1, ...x_n \in Q, \lambda_1, ..., \lambda_n \in [0, 1]$ such than $\sum_{i=1}^n \lambda_i = 1$ and $\sum_{i=1}^n \lambda_i x_i = x$.
Since S is convex and $x_1, ...x_n \in Q \subseteq S$,

$$x = \sum_{i=1}^n \lambda_i x_i \in S$$

So $conv(Q) \subseteq S$

$\square$

(b) *Solution:* By the "obtuse angle" property

$$(x - \Pi_S[x]) \cdot (\Pi_S[y] - \Pi_S[x]) \leq 0$$

$$(y - \Pi_S[y]) \cdot (\Pi_S[x] - \Pi_S[y]) \leq 0$$

So from q1 in assignment 1:

$$\|\Pi_S[x] - \Pi_S[y]\| \leq \|x - y\|$$

□

(c) *Solution:* Let $U = -S := \{-x : x \in S\}, T := W + U$.
U is convex because S is convex. T is convex as the Minkowski sum of two convex sets.
Define $f(z) = \|z\|^2$ and $a = \arg\min_{z \in T} f(z)$. T and f are convex, so by constrained optimality conditions

$$\forall z \in T.\ 0 \leq \nabla f(a)(z - a) \cdot = 2a \cdot (z - a) \implies \|a\|^2 \leq a^\top z.$$

And since $S \cap W = \emptyset$, $a \neq 0$. Meaning $\|a\|^2 > 0$.
Let $x \in S, y \in W$. By T's definition, $(y - x) \in T$

$$0 < \|a\|^2 < a^\top (y - x) \implies a^\top x < a^\top y$$

□

Question #4

(a) *Solution:*
Projected Gradient Descent: Starting from $x_1 \in S$ with step size η_t, do:

$$x_{t+1} = \Pi_S \left(x_t - \frac{D}{\sqrt{2 \sum_{s=1}^{t} \|g_s\|^2}} g_t \right)$$

□

(b) *Solution:* In the recitation we showed:

$$\sum_{t=1}^{T} g_t \cdot (x_t - x^*) \leq \frac{1}{2} \sum_{t=1}^{T} (\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}) \|x_t - x^*\|^2 + \frac{1}{2} \sum_{t=1}^{T} \eta_t \|g_t\|^2$$

Substituting $\eta_t = \frac{D}{\sqrt{2\sum_{s=1}^{t}\|g_s\|^2}}$, we get:

$$\sum_{t=1}^{T} g_t \cdot (x_t - x^*) \leq \frac{1}{2}\sum_{t=1}^{T}(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}})\|x_t - x^*\|^2 + \frac{1}{2}\sum_{t=1}^{T}\frac{D}{\sqrt{2\sum_{s=1}^{t}\|g_s\|^2}}\|g_t\|^2$$

$$\leq \frac{D^2}{2\eta_T} + \frac{1}{2}\sum_{t=1}^{T}\eta^t\|g_t\|^2 \leq \frac{D^2}{\frac{2D}{\sqrt{2\sum_{s=1}^{T}\|g_s\|^2}}} + \frac{1}{2}\sum_{t=1}^{T}\frac{D\|g_t\|^2}{\sqrt{2\sum_{s=1}^{t}\|g_s\|^2}}$$

$$\leq \frac{D}{2}\sqrt{2\sum_{s=1}^{T}\|g_s\|^2} + \frac{1}{2}\sum_{t=1}^{T}\frac{D\|g_t\|^2}{\sqrt{2\sum_{s=1}^{t}\|g_s\|^2}}$$

□

(c) *Solution:* Let $\pi = (S_n)_{n=0}^{T}$ be a partition of $[0, S_T]$, and so for $f(x) = \frac{1}{\sqrt{x}}$ we get:

$$\sum_{t=1}^{T}\frac{\alpha_t}{\sqrt{S_t}} = \sum_{t=1}^{T}\frac{S_t - S_{t-1}}{\sqrt{S_t}} = \underline{S}(f,\pi) \leq \int_0^{S_T} f(x)\,dx = 2\sqrt{x}\Big|_0^{S_T} = 2\sqrt{S_T}$$

As required. □

(d) *Solution:* From Jensen's inequality we get:

$$f(\overline{x}) - f(x^*) \leq \frac{1}{T}\sum_{t=1}^{T} f(x_t) - f(x^*) = \frac{1}{T}\sum_{t=1}^{T}(f(x_t) - f(x^*))$$

From convexity:

$$\frac{1}{T}\sum_{t=1}^{T}(f(x_t) - f(x^*)) \leq \frac{1}{T}\sum_{t=1}^{T}\nabla f(x_t)\cdot(x_t - x^*)$$

From (b):

$$\frac{1}{T}\sum_{t=1}^{T}\nabla f(x_t)\cdot(x_t - x^*) \leq \frac{D}{2T}\left(\sqrt{2\sum_{t=1}^{T}\|\nabla f(x_t)\|^2} + \frac{1}{\sqrt{2}}\sum_{t=1}^{T}\frac{\|\nabla f(x_t)\|^2}{\sqrt{\sum_{s=1}^{t}\|\nabla f(x_s)\|^2}}\right)$$

From (c):

$$\frac{D}{2T}\left(\sqrt{2\sum_{t=1}^{T}\|\nabla f(x_t)\|^2} + \frac{1}{\sqrt{2}}\sum_{t=1}^{T}\frac{\|\nabla f(x_t)\|^2}{\sqrt{\sum_{s=1}^{t}\|\nabla f(x_s)\|^2}}\right)$$

$$\leq \frac{D}{2T}\left(\sqrt{2}\sqrt{\sum_{t=1}^{T}\|\nabla f(x_t)\|^2} + \frac{1}{\sqrt{2}}2\sqrt{\sum_{t=1}^{T}\|\nabla f(x_t)\|^2}\right) \leq \frac{D}{T}\sqrt{2\sum_{t=1}^{T}\|\nabla f(x_t)\|^2}$$

In summary we get:

$$f(\overline{x}) - f(x^*) \leq \frac{D}{T}\sqrt{2\sum_{t=1}^{T}\|\nabla f(x_t)\|^2}$$

As required. $\square$

(e) *Solution:* We've seen that given a differentiable function $f : S \to \mathbb{R}$ that's G-Lipschitz it follows that for all $x \in S : \|\nabla f(x)\| \leq G$, so we get:

$$f(\overline{x}) - f(x^*) \leq \frac{D}{T} \cdot \sqrt{2\sum_{t=1}^{T}\|\nabla f(x)\|^2} \leq \frac{D}{T}\sqrt{2TG^2} = \frac{\sqrt{2}DG}{\sqrt{T}}$$

$\square$