AMATH 515: Homework 5
Sid Meka

1. Consider the Euclidean projection operator

$$P_\Omega(x) = \arg\min_{z\in\Omega} \|z - x\|^2,$$

for a closed convex set $\Omega \subset \mathbb{R}^n$. Show that P_Ω is non-expansive, i.e.,

$$\|P_\Omega(x) - P_\Omega(x')\| \leq \|x - x'\|, \qquad \forall x, x' \in \mathbb{R}^n.$$

We start by defining the projection operator. The Euclidean projection $P_\Omega(x)$ of a point $x \in \mathbb{R}^n$ onto the closed convex set Ω is defined as:

$$P_\Omega(x) = \arg\min_{z\in\Omega} \|z - x\|^2$$

Since Ω is closed and convex, the projection is well defined and unique for every x. Let $p = P_\Omega(x)$ and $p' = P_\Omega(x')$ meaning:

$$p = \arg\min_{z\in\Omega} \|z - x\|^2 \text{ and } p' = \arg\min_{z\in\Omega} \|z - x'\|^2$$

A key property of projections onto convex sets is that the projection satisfies the variational inequality:

$$\langle p - x, z - p\rangle \geq 0\,, \quad \forall z \in \Omega$$

We can apply the same property to p':

$$\langle p' - x', z - p'\rangle \geq 0\,, \quad \forall z \in \Omega$$

From $\langle p - x, z - p\rangle \geq 0$, we can apply $z = p'$ to get:

$$\langle p - x, p' - p\rangle \geq 0$$

Similarly, from $\langle p' - x', z - p'\rangle \geq 0$, we can apply $z = p$ to get:

$$\langle p' - x', p - p'\rangle \geq 0$$

Combining $\langle p - x, p' - p\rangle \geq 0$ and $\langle p' - x', p - p'\rangle \geq 0$, we get:

$$\langle p - x, p' - p\rangle + \langle p' - x', p - p'\rangle \geq 0$$

$$\langle p - x, p' - p\rangle - \langle p' - x', p' - p\rangle \geq 0$$

$$\langle (p - x) - (p' - x'), p' - p\rangle \geq 0$$

$$\langle p - p' - x + x', p' - p\rangle \geq 0$$

Using the linearity and symmetry of the inner product, we rewrite:

$$\langle p - p', p' - p\rangle - \langle x - x', p' - p\rangle \geq 0$$

$$\langle p - p', -(p - p')\rangle - \langle x - x', p' - p\rangle \geq 0$$

$$-\|p - p'\|^2 - \langle x - x', p' - p\rangle \geq 0$$

$$-\|p - p'\|^2 \geq \langle x - x', p' - p\rangle$$

$$\|p - p'\|^2 \leq -\langle x - x', p' - p\rangle$$
$$\|p - p'\|^2 \leq -\langle x - x', p - p'\rangle$$

Now, to complete the proof, we apply Cauchy-Schwarz inequality:

$$\langle x - x', p - p'\rangle \leq \|x - x'\|\|p - p'\|$$

Since we established $\|p - p'\|^2 \leq \langle x - x', p - p\rangle$, using Cauchy-Schwarz gives us:

$$\|p - p'\|^2 \leq \|x - x'\|\|p - p'\|$$

If $p = p'$, then the inequality trivially holds. Thus, we proceed with the non trivial case dividing both sides by $\|p - p'\|$:

$$\|p - p'\| \leq \|x - x'\|$$

Thus, we have that $\|p - p'\| \leq \|x - x'\|$ for both $p = p'$ and $p \neq p'$.

Thus, we conclude:

$$\|P_\Omega(x) - P_\Omega(x')\| \leq \|x - x'\|, \qquad \forall x, x' \in \mathbb{R}^n$$

This proves that the projection operator is indeed nonexpansive.

2. Let $\Omega = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ be the unit ball. Show that

$$P_\Omega(x) = \begin{cases} x & \text{if } \|x\| \leq 1 \\ \dfrac{x}{\|x\|} & \text{if } \|x\| > 1 \end{cases}$$

The projection $P_\Omega(x)$ of a point $x \in \mathbb{R}^n$ onto Ω is defined as the unique point in Ω that minimizes the Euclidean distance to x: $P_\Omega(x) = \arg\min_{y \in \Omega} \|y - x\|$. Since Ω is closed and convex, projections onto Ω are well defined and unique.

Case 1: $\|x\| \leq 1$:
If x is already in Ω, then clearly x is its own projection since the function $\|y - x\|$ attains its minimum at $y = x$ as when $y = x$, $\|y - x\| = 0$. Hence, $P_\Omega(x) = x$ when $\|x\| \leq 1$.

Case 2: $\|x\| > 1$:
If $\|x\| > 1$, then $x \notin \Omega$. The closest point in Ω to x must be on the boundary $\partial\Omega = \{y \in \mathbb{R}^n \mid \|y\| = 1\}$. We now solve the constrained optimization problem: $\min_{y \in \Omega} \|y - x\|^2$ subject to $\|y\| \leq 1$. We define the Lagrangian: $L(y, \lambda) = \|y - x\|^2 + \lambda(\|y\|^2 - 1)$. We find $\nabla_y L = 2(y - x) + 2\lambda y$. We set $\nabla_y L$ to 0 in order to find the critical points of the Lagrangian, which correspond to the optimal solutions that

satisfy the constraint $\|y\| = 1$ for our projection. From setting $2(y - x) + 2\lambda y$ to 0, we get:

$$\begin{aligned}
2(y-x) + 2\lambda y &= 0 && \text{(We set this to 0.)} \\
y - x + \lambda y &= 0 \\
y + \lambda y &= x \\
y(1+\lambda) &= x \\
y &= \frac{x}{1+\lambda} && \text{(Not needed now, but will be useful later)} \\
\|y\|(1+\lambda) &= \|x\| && \text{(Taking norms on both sides as } \|y\| = 1\text{)} \\
1 + \lambda &= \|x\| && \text{(As } \|y\| = 1\text{)} \\
\lambda &= \|x\| - 1 && \text{(Solving for } \lambda\text{)} \\
y &= \frac{x}{1+\lambda} && \text{(Stated before)} \\
y &= \frac{x}{1 + \|x\| - 1} && \text{(Substituting } \lambda = \|x\| - 1\text{)} \\
y &= \frac{x}{\|x\|}
\end{aligned}$$

Because $y = \frac{x}{\|x\|}$, the Euclidean projection of x onto Ω for $\|x\| > 1$ is $P_\Omega(x) = \frac{x}{\|x\|}$.

From our cases of $\|x\| \leq 1$ and $\|x\| > 1$, and we found using them, we have shown that

$$P_\Omega(x) = \begin{cases} x & \text{if } \|x\| \leq 1 \\ \dfrac{x}{\|x\|} & \text{if } \|x\| > 1 \end{cases}$$

3. Given scalars $L \geq m > 0$ show that $\max\{|1 - \alpha m|, |1 - \alpha L|\}$ is minimized when $\alpha = \frac{2}{L+m}$.

   We will consider two cases: $L = m$ and $L > m$.

   Case 1: $L = m$
   When $L = m$, $\alpha = \frac{2}{2L}$, or more simply, $\alpha = \frac{1}{L}$. Furthermore, $\max\{|1 - \alpha m|, |1 - \alpha L|\} = \max\{|1 - \alpha L|, |1 - \alpha L|\}$. This means that $\max\{|1 - \alpha m|, |1 - \alpha L|\} = |1 - \alpha L|$ because the max of two identical values is simply that value itself. Furthermore, $|1 - \alpha L| = |1 - \frac{L}{L}| = |1 - 1| = 0$. Thus, in this trivial case, $\max\{|1-\alpha m|, |1-\alpha L|\}$ simply equals 0. We conclude that the expression $\max\{|1-\alpha m|, |1-\alpha L|\}$ is minimized at 0 in this trivial case.

   Case 2: $L > m$
   To achieve minimization, we should make both terms equal: $|1 - \alpha m| = |1 - \alpha L|$. Since in this case, $L > m > 0$, we have that $\alpha > 0$. We analyze when both absolute values are equal while still considering $L > m$ as this means $L \neq m$. This means we consider $1 - \alpha m = -(1 - \alpha L)$ instead of $1 - \alpha m = 1 - \alpha L$ as once again, $L \neq m$. So, we proceed with $1 - \alpha m = -(1 - \alpha L)$.

$$1 - \alpha m = -(1 - \alpha L)$$

$$1 - \alpha m = -1 + \alpha L$$

$$2 = \alpha m + \alpha L$$

$$\frac{2}{\alpha} = m + L$$

$$\alpha = \frac{2}{m+L}$$

We have that in both cases $\max\{|1-\alpha m|, |1-\alpha L|\}$ is minimized when $\alpha = \frac{2}{L+m}$. Thus, with our scalars $L \geq m > 0$, $\max\{|1-\alpha m|, |1-\alpha L|\}$ is minimized when $\alpha = \frac{2}{L+m}$.

4. For a constant vector $c \in \mathbb{R}^n$ and variable $x \in \mathbb{R}^n$ find the minimizer of $c^T x$ over Ω, where Ω is each of the following sets:

   (a) The unit ball $\{x \mid \|x\| \leq 1\}$
   Since $c^T x$ represents the dot product of c and x, the function $c^T x$ is minimized when x is in the opposite direction of c. This is because the dot product is maximized when x is aligned with c and minimized when x is in the opposite direction. The constraint requires that x lies within or on the boundary of the unit ball meaning $\|x\| \leq 1$. To achieve the smallest possible value of $c^T x$, we should take x to be in the direction of $-c$ and have the largest possible norm, where $\|x^\star\| = 1$ because $\left\|-\frac{c}{\|c\|}\right\| = \frac{\|c\|}{\|c\|}$, and $\frac{\|c\|}{\|c\|}$ is simply 1. Thus, for the unit ball, the optimal choice is $x^\star = -\frac{c}{\|c\|}$, which satisfies $\|x^\star\| = 1$. With our minimizer $x^\star = -\frac{c}{\|c\|}$, we have that our minimum value is calculated from $c^T x^\star$, which is $c^T \left(-\frac{c}{\|c\|}\right)$, which simplifies to $-\|c\|$. Thus, the minimum value is $-\|c\|$ with minimizer $x^\star = -\frac{c}{\|c\|}$.

   (b) A box $\{x \mid 0 \leq x_i \leq 1, \quad i = 1, \ldots, n\}$
   The constraint means that each component x_i is bounded between 0 and 1. The function $f(x) = c^T x$ is linear, so it is minimized at a vertex of the box since a linear function attains its extreme values at the boundary of a convex set because a linear function's minimum will occur at an extreme point of the feasible region as a linear function attains its extreme values at the boundary of a convex set since it has no curvature and always increases and decreases in a fixed direction and in a monotone manner. The function $c^T x$ is the sum: $f(x) = \displaystyle\sum_{i=1}^{n} c_i x_i$. Each x_i should be set to 0 if $c_i > 0$ to minimize the contribution to $c^T x$, and each x_i should be set to 1 if $c_i < 0$ to minimize the contribution to $c^T x$. Thus, the optimal solution is: $x_i^\star = \begin{cases} 0 & \text{if } c_i > 0 \\ 1 & \text{if } c_i < 0 \end{cases}$. Thus, our minimizer is $x_i^\star = \begin{cases} 0 & \text{if } c_i > 0 \\ 1 & \text{if } c_i < 0 \end{cases}$ with the minimum value $-\displaystyle\sum_{i=1}^{n} c_i^{\smiley}$, where $c_i^{\smiley}$ is defined as $c_i^{\smiley} = \begin{cases} -c_i \text{ when } c_i < 0 \\ 0 \text{ when } c_i > 0 \end{cases}$. Note that $c_i = 0$ the term $c_i x_i = 0$ regardless of the choice of x_i, so any value of $x_i \in [0, 1]$ is optimal.

Hint: try to visualize the problems in 2D before attempting a proof.

5. Consider the projected steepest descent update formula

$$x^{(k+1)} = P_\Omega(x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)}))$$

with step size $\alpha^{(k)} > 0$, a closed convex set $\Omega \subset \mathbb{R}^n$, and a continuously differentiable f. Show that

$$x^{(k+1)} = \arg\min_{x \in \Omega} f(x^{(k)}) + \nabla f(x^{(k)})^T (x - x^{(k)}) + \frac{1}{2\alpha^{(k)}} \|x - x^{(k)}\|^2$$

The projection operator P_Ω projects a given point onto the closed convex set Ω by solving the minimization problem: $P_\Omega(y) = \arg\min_{x \in \Omega} \frac{1}{2}\|x - y\|^2$. For our given update rule, we set: $y = x^{(k)} -$

$\alpha^{(k)}\nabla f(x^{(k)})$. Thus, we have $x^{(k+1)} = \underset{x\in\Omega}{\arg\min}\,\dfrac{1}{2}\|x - (x^{(k)} - \alpha^{(k)}\nabla f(x^{(k)}))\|^2$ from $x^{(k+1)} = P_\Omega(x^{(k)} - \alpha^{(k)}\nabla f(x^{(k)}))$.

$$
\begin{aligned}
x^{(k+1)} &= P_\Omega(x^{(k)} - \alpha^{(k)}\nabla f(x^{(k)})) \\
x^{(k+1)} &= \underset{x\in\Omega}{\arg\min}\,\frac{1}{2}\|x - (x^{(k)} - \alpha^{(k)}\nabla f(x^{(k)}))\|^2 \\
x^{(k+1)} &= \underset{x\in\Omega}{\arg\min}\,\frac{1}{2}\|x - x^{(k)} + \alpha^{(k)}\nabla f(x^{(k)})\|^2 \\
x^{(k+1)} &= \underset{x\in\Omega}{\arg\min}\,\frac{1}{2}\left(\|x - x^{(k)}\|^2 + 2\alpha^{(k)}\nabla f(x^{(k)})^T(x - x^{(k)}) + (\alpha^{(k)})^2\|\nabla f(x^{(k)})\|^2\right) \\
x^{(k+1)} &= \underset{x\in\Omega}{\arg\min}\,\frac{1}{2}\left(\|x - x^{(k)}\|^2\right) + \left(\alpha^{(k)}\nabla f(x^{(k)})^T(x - x^{(k)})\right) + \frac{1}{2}\left((\alpha^{(k)})^2\|\nabla f(x^{(k)})\|^2\right)
\end{aligned}
$$

Notice how the term $\frac{1}{2}\left(\alpha^{(k)})^2\|\nabla f(x^{(k)})\|^2\right)$ does not influence which x minimizes the expression as when minimizing over x, $f(x^{(k)})$ does not influence the choice of x and $f(x^{(k)})$ only shifts the objective function by a constant amount because $f(x^{(k)})$ is independent of x and does not contribute to the gradient or curvature of the function being minimized. Since optimization depends only on the terms that vary with x, constants can be ignored without affecting the minimization. Thus, $f(x^{(k)})$ and resultantly, $\frac{1}{2}\left((\alpha^{(k)})^2\|\nabla f(x^{(k)})\|^2\right)$ do not alter the minimization of the argument of our given objective function that we are set to minimize. Thus, we proceed by minimizing $\frac{1}{2}\left(\|x - x^{(k)}\|^2\right) + \left(\alpha^{(k)}\nabla f(x^{(k)})^T(x - x^{(k)})\right)$ and proceed with $x^{(k+1)} = \underset{x\in\Omega}{\arg\min}\,\dfrac{1}{2}\left(\|x - x^{(k)}\|^2\right) + \left(\alpha^{(k)}\nabla f(x^{(k)})^T(x - x^{(k)})\right)$.

Now, we proceed with the minimization problem:

$$x^{(k+1)} = \underset{x\in\Omega}{\arg\min}\,\frac{1}{2}\left(\|x - x^{(k)}\|^2\right) + \left(\alpha^{(k)}\nabla f(x^{(k)})^T(x - x^{(k)})\right)$$

Rearranging, we get:

$$x^{(k+1)} = \underset{x\in\Omega}{\arg\min}\,\alpha^{(k)}\nabla f(x^{(k)})^T(x - x^{(k)}) + \frac{1}{2}\|x - x^{(k)}\|^2$$

Now, we divide by our step size, $\alpha^{(k)}$ to get:

$$x^{(k+1)} = \underset{x\in\Omega}{\arg\min}\,f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) + \frac{1}{2\alpha^{(k)}}\|x - x^{(k)}\|^2$$

Thus, we have shown from our projected steepest descent update formula: $x^{(k+1)} = P_\Omega(x^{(k)} - \alpha^{(k)}\nabla f(x^{(k)}))$, we can get $x^{(k+1)} = \underset{x\in\Omega}{\arg\min}\,f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) + \dfrac{1}{2\alpha^{(k)}}\|x - x^{(k)}\|^2$.