MATH 173A: Homework #4

Due on Nov 10, 2024 at 23:59pm

Professor Cloninger

Ray Tsai

A16848188

Problem 1

(a) Find an expression for the orthogonal projection of a point $x \in \mathbb{R}^n$ onto the convex set

$$B = \{z \in \mathbb{R}^n : 0 \le z_i \le 1 \text{ for each } i = 1, ..., n\}.$$

You need to show your work, and justify your answer. The expression can be written piecewise, and per dimension if it's easier / more compact. **Hint:** It might be helpful to sketch B, when $n = 2$ (i.e., in 2 dimensions), and use the sketch to help you figure out what the projection should be.

Proof. For $x \in \mathbb{R}^n$, we need to find $\Pi_B(x) = \arg\min_{z \in B} \|z - x\| = \arg\min_{z \in B} \sum_i (z_i - x_i)^2$. Notice that we may decouple this minimization problem across n dimension by minimizing each z_i independently. That is, for all i

$$z_i = \arg\min_{a \in [0,1]} (a - x_i)^2 = \min(\max(0, x_i), 1).$$

□

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by

$$f(x) = \|Ax\|_2^2 + a^T x$$

where $A \in \mathbb{R}^{n \times n}$ is a positive definite matrix, and $a \in \mathbb{R}^n$. Write a projected gradient descent algorithm to solve

$$\min_{x \in \Omega} f(x)$$

for $\Omega = B$, with B from part (a). You do not need to specify the step size for this problem.

Proof. Note that

$$\nabla f(x) = 2A^T Ax + a,$$

and thus the projected gradient descent algorithm is

$$x^{(k+1)} = \Pi_\Omega \left(x^{(k)} - \mu \nabla f(x^{(k)}) \right) = \Pi_B \left(x^{(k)} - \mu(2A^T Ax^{(k)} + a) \right).$$

More explicitly, for all i,

$$x_i^{(k+1)} = \min\left(\max\left(0, x_i^{(k)} - 2\mu(A^T Ax^{(k)} + a)_i \right), 1 \right).$$

□

(c) Repeat part (b) but for $\Omega = B_2^n = \{z \in \mathbb{R}^n : \|z\|_2 \le 1\}$.

Proof. Notice

$$\Pi_\Omega(x) = \begin{cases} \frac{x}{\|x\|_2} & \text{if } \|x\| > 1, \\ x & \text{if } \|x\| \le 1. \end{cases}$$

Hence, the projected gradient descent algorithm is

$$x^{(k+1)} = \Pi_\Omega \left(x^{(k)} - \mu \nabla f(x^{(k)}) \right) = \Pi_B \left((I - 2\mu A^T A)x^{(k)} - \mu a \right),$$

which is $\frac{(I-2\mu A^T A)x^{(k)} - \mu a}{\|(I-2\mu A^T A)x^{(k)} - \mu a\|_2}$ if $\|x\| > 1$ and $(I - 2\mu A^T A)x^{(k)} - \mu a$ otherwise. □

Problem 2

Consider the *hollow* sphere S in $\mathbb{R}^n$, i.e., the set $S := \{x \in \mathbb{R}^n : \|x\|_2^2 = 1\}$. Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = x^T Q x$$

where Q is an $n \times n$ symmetric matrix. For this problem you may use the fact that $\nabla f(x) = 2Qx$.

(a) For an arbitrary point $y \in \mathbb{R}^n$, $\Pi(y)$ be the projection of y onto S. Find an expression for $\Pi(y)$ and give a short argument (i.e., proof) for why this is the correct expression. Make sure to handle the case $y = 0$ (i.e., the zero vector).

Proof. I claim that $\Pi(y) = \frac{y}{\|y\|_2}$ if $y \neq 0$ and $\Pi(0)$ can be any point in S. Note that the reverse triangle-inequality yields a lower bound

$$\|x - y\|_2 \geq |\|x\|_2 - \|y\|_2| = |1 - \|y\|_2| \,,$$

for $x \in \Omega$. Obvisouly, any $x \in \Omega$ achieves the lower bound when $y = 0$. Suppose $y \neq 0$. Obviously $\frac{y}{\|y\|_2} \in \Omega$. Since

$$\left\| \frac{y}{\|y\|_2} - y \right\| = \left\| \left(\frac{1}{\|y\|_2} - 1 \right) y \right\| = \|y\|_2 \left| \frac{1}{\|y\|_2} - 1 \right| = |1 - \|y\|_2|$$

achieves the lower bound, $\Pi(y) = \frac{y}{\|y\|_2}$. □

(b) Is S a convex set?

Proof. S is not a convex set. Consider $x = (1, 0)$ and $y = (-1, 0)$. Then $0 = \frac{1}{2}(1, 0) + \frac{1}{2}(-1, 0) \notin S$. □

(c) Write a projected gradient descent algorithm, with constant step size μ, for

$$\min_{x \in \mathbb{R}^n} x^T Q x \qquad \text{subject to} \qquad \|x\|_2^2 = 1.$$

Proof. Note that $\nabla f(x) = 2Qx$, and thus the projected gradient descent algorithm is

$$x^{(k+1)} = \Pi_S \left((I - 2\mu Q) x^{(k)} \right),$$

which is equal to $\frac{(I - 2\mu Q)x^{(k)}}{\|(I - 2\mu Q)x^{(k)}\|}$ if $x^{(k)} \neq 0$ and any point in S if $x^{(k)} = 0$. □

(d) Is the projected gradient descent algorithm guaranteed to converge to the solution for small enough μ? If not, can you give an example of Q and an initialization $x^{(0)}$ where the algorithm won't converge?

Proof. Consider $Q = \operatorname{diag}(1, 0)$ and $x^{(0)} = (1, 0)$. Then $x^{(k+1)} = \Pi_S \left(\begin{bmatrix} 1 - 2\mu & 0 \\ 0 & 1 \end{bmatrix} x^{(k)} \right)$. Since $x^{(0)}$ only have the first entry non-zero,

$$x^{(k+1)} = \frac{1 - 2\mu}{|1 - 2\mu|} x^{(k)} = \left(\frac{1 - 2\mu}{|1 - 2\mu|} \right)^{k+1} x^{(0)},$$

which is equal to $x^{(0)}$ for small enough μ. But then $f(x^{(0)}) = 1$ and $f\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = 0$, so the algorithm fails to converge. □