EECS 127/227AT Optimization Models in Engineering
Spring 2020 **Homework 6**

This homework is due Friday, March 6, 2020 at 23:00 (11pm).
Self grades are due Friday, March 13 2020 at 23:00 (11pm).

This version was compiled on 2020-02-29 07:12.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned) as well as your IPython notebook with solutions saved as a PDF.

1. **Proof of Hölder's Inequality** In this question, we will prove Hölder's Inequality using convexity and verify that the ℓ_p norms (defined subsequently) indeed satisfy the properties of a norm. Let $\vec{x} \in \mathbb{R}^n$, we now define the ℓ_p norm, denoted by $\|\cdot\|_p$ as follows:

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

for $p \geq 1$. Note, that when $p = 2$, the ℓ_p norm corresponds to the standard Euclidean norm of the vector $\vec{x}$. Hölder's states that for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $p, q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$, we have:

$$\vec{x}^\top \vec{y} \leq \|\vec{x}\|_p \|\vec{y}\|_q .$$

Notice that when $p = q = 2$, Hölder's Inequality recovers the standard Cauchy Schwarz inequality. We will now prove Hölder's Inequality via the following sequence of steps:

(a) Let $a, b \geq 0$. Using the concavity of the function, $f(x) = \log x$, prove the following statement:

$$a \cdot b \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

The above inequality is also known as *Young's Inequality.*
Hint 1: For the case where $a, b > 0$ it might be useful to denote $u = a^p, w = b^q$ and consider $\log\left(\frac{1}{p} \cdot u + \frac{1}{q} \cdot w\right)$.
Hint 2: We have,

$$\frac{1}{p} + \frac{1}{q} = 1.$$

(b) Use Young's inequality to conclude the proof of Hölder's Inequality.
Hint: When $\vec{x}, \vec{y} \neq 0$, define the vectors $\vec{u} = \frac{\vec{x}}{\|\vec{x}\|_p}$ and $\vec{w} = \frac{\vec{y}}{\|\vec{y}\|_q}$. Now, showing Hölder's Inequality is equivalent to proving:

$$\vec{u}^\top \vec{w} \leq 1.$$

(c) Now, we will show that Hölder's Inequality is tight i.e we can find $\vec{x}, \vec{y}$ such that $\vec{x}^\top \vec{y} = \|\vec{x}\|_p \|\vec{y}\|_p$. Let $p > 1$ and let q be such that $\frac{1}{p} + \frac{1}{q} = 1$. Prove that:

$$\|\vec{x}\|_p = \max_{\vec{y}:\|\vec{y}\|_q=1} \vec{x}^\top \vec{y}. \tag{1}$$

Note that this is equivalent to showing that Hölder's Inequality is tight because the optimal $\vec{y}^*$ from Equation (1) and $\vec{x}$ will satisfy Hölder's Inequality with equality.

Hint 1: That the right-hand side is a less than the left-hand side follows from Hölder's Inequality.

Hint 2: To show equality, choose vector $\vec{y}$ appropriately satisfying $\|\vec{y}\|_q = 1$, such that $\vec{x}^\top \vec{y} = \|\vec{x}\|_p$. Can you construct the entries of this $\vec{y}$? You might want to make sure that the sign of y_i matches that of x_i, and then appropriately pick the magnitude of x_i to have $\vec{x}^\top \vec{y} = \|\vec{x}\|_p$. Then check that $\|\vec{y}\|_q = 1$.

(d) Use part (c) to conclude that $\|\cdot\|_p$ indeed defines a norm. Recall that $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a valid norm if it satisfies the following three properties:

i. $\vec{x} = 0 \iff \|\vec{x}\| = 0$

ii. $\forall \alpha \in \mathbb{R},\ \vec{x} \in \mathbb{R}^n : \|\alpha \vec{x}\| = |\alpha| \|\vec{x}\|$

iii. $\forall \vec{x}, \vec{y} \in \mathbb{R}^n : \|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$.

2. Convex or Concave Determine whether the following functions are convex, strictly convex, concave, strictly concave, both or neither.

(a) $f(x) = e^x - 1$ on $\mathbb{R}$

(b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$

(c) The log-likelihood of a set of points $\{x_1, \ldots, x_n\}$ that are normally distributed with mean μ and finite variance $\sigma > 0$ is given by:

$$f(\mu, \sigma) = n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

i. Show that if we view the log likelihood for fixed σ as a function of the mean, i.e

$$g(\mu) = n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

then g is strictly concave (equivalently, we say f is strictly concave in μ).

ii. (Optional) Show that if we view the log likelihood for fixed μ as a function of the inverse of the variance, i.e

$$h(z) = n \log\left(\frac{\sqrt{z}}{\sqrt{2\pi}}\right) - \frac{z}{2} \sum_{i=1}^{n} (x_i - \mu)^2$$

then h is strictly concave (equivalently, we say f is strictly concave in $z = \frac{1}{\sigma^2}$). Note that we have used the dummy variable z to denote $\frac{1}{\sigma^2}$.

iii. (Optional) Show that f is not jointly concave in $\mu, \frac{1}{\sigma^2}$.
Hint: We say a function $w(x, y)$ with $x \in \mathcal{R}^m$ and $y \in \mathcal{R}^n$ is jointly convex if

$$w\left(\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)\right) \leq \lambda w((x_1, y_1)) + (1-\lambda) w((x_2, y_2)).$$

This is the same as letting $z = (x, y)$ and saying f is convex in z. We can define joint concavity in a similar fashion by reversing the inequalities.

(d) $f(x) = \log(1+e^x)$. Note that this implies that $g(x) = -f(x) = \log \frac{1}{(1+e^x)}$ is concave. Compare this to $h(x) = \frac{1}{(1+e^x)}$, is $h(x)$ convex or concave?

3. Quadratic inequalities

Consider the set S defined by the following inequalities:

$$(x_1 \geq -x_2 + 1 \text{ and } x_1 \leq 0) \text{ or } (x_1 \leq -x_2 + 1 \text{ and } x_1 \geq 0).$$

To be more precise,

$$\begin{aligned} S_1 &= \{\vec{x} \in \mathbb{R}^2 \mid x_1 \geq -x_2 + 1, x_1 \leq 0\} \\ S_2 &= \{\vec{x} \in \mathbb{R}^2 \mid x_1 \leq -x_2 + 1, x_1 \geq 0\} \\ S &= S_1 \cup S_2. \end{aligned}$$

(a) Draw the set S. Is it convex?

(b) Show that the set S, can be described as a single quadratic inequality of the form $q(\vec{x}) = \vec{x}^\top A \vec{x} + 2\vec{b}^\top \vec{x} + c \leq 0$, for matrix $A = A^\top \in \mathbb{R}^{2\times 2}$, $\vec{b} \in \mathbb{R}^2$ and $c \in \mathbb{R}$ i.e S can be written as $S = \{\vec{x} \in \mathbb{R}^2 \mid q(\vec{x}) \leq 0\}$). Find $A, \vec{b}, c$.
Hint: Can you combine the constraints to make one quadratic constraint?

(c) What is the convex hull of this set?

(d) We will now consider some convex optimization problems over S_1 that illustrate the role of the constraints in the optimization problem. For each of the following optimization problems find the optimal point, $\vec{x}^*$. Describe the constraints that are active in attaining the optimal value. *Hint: Suppose that there exists a point $\vec{x}$ such that $\nabla f(\vec{x}) = 0$. From the first order characterization of a convex function $\vec{x}$ would be an optimum value for f subject to no constraints. If $\vec{x}$ is not in the constraint set S_1, then the optimum point must be on the boundary of the set, i.e. it satisfies at least one of the constraints defining S_1 with equality.*

i. Minimize $f(\vec{x}) = (x_1 + 1)^2 + (x_2 - 3)^2$ subject to $\vec{x} \in S_1$.
ii. Minimize $f(\vec{x}) = (x_1 + 2)^2 + (x_2 - 2)^2$ subject to $\vec{x} \in S_1$.
iii. Minimize $f(\vec{x}) = x_1^2 + x_2^2$ subject to $\vec{x} \in S_1$.

4. Gradient Descent Algorithm

Given a continuous and differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient of f at any point x, $\nabla f(x)$, is orthogonal to the level curve of f at point x, and it points in the increasing direction of f. In other words, moving from point x in the direction $\nabla f(x)$ leads to an increase in the value of f, while moving in the direction of $-\nabla f(x)$ decreases the value of f. This idea gives an iterative algorithm to minimize the function f: the gradient descent algorithm.

This problem is a light introduction to the gradient descent algorithm, which we will cover in more detail later in the class.

(a) Consider $f(x) = \frac{1}{2}(x-2)^2$, and assume that we use the gradient descent algorithm:

$$x[k+1] = x[k] - \eta \nabla f(x[k]) \quad \forall k \geq 0,$$

with some random initialization $x[0]$, where $\eta > 0$ is the step size (or the learning rate) of the algorithm. Write $(x[k]-2)$ in terms of $(x[0]-2)$, and show that $x[k]$ converges to 2, which is the unique minimizer of f, when $\eta = 0.2$.

(b) What is the largest value of η that we can use so that the gradient descent algorithm converges to 2 from all possible initializations in $\mathbb{R}$? What happens if we choose a larger step size?

(c) Now assume that we use the gradient descent algorithm to minimize $f(\vec{x}) = \frac{1}{2}\|A\vec{x} - \vec{b}\|_2^2$ for some $A \in \mathbb{R}^{m\times n}$ and $\vec{b} \in \mathbb{R}^m$, where A has full column rank. First show that $\nabla f(\vec{x}) = A^\top A\vec{x} - A^\top \vec{b}$. Then, write $(\vec{x}[k] - (A^\top A)^{-1}A^\top \vec{b})$ in terms of $(\vec{x}[0] - (A^\top A)^{-1}A^\top b)$ and find the largest step size that we can use (in terms of A and $\vec{b}$) so that the gradient descent algorithm converges for all possible initializations. Your largest step size should be a function of $\lambda_{\max}(A^\top A)$, the largest eigenvalue of $A^\top A$.

5. Homework process

Whom did you work with on this homework? List the names and SIDs of your group members.