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 2020-03-10 08:09.

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,

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

Solution: First, we note that when either a or b is equal to 0, the left hand side of the inequality is 0 while the right hand side is non-negative. Therefore, Young's inequality holds in this case. Now, assume the case that $a, b > 0$ and let $u = a^p$ and $w = b^q$. Now, from the concavity of $f(x) = \log x$ and using the fact that $1/p + 1/q = 1$, we have:

$$\log\left(\frac{1}{p} \cdot u + \frac{1}{q} \cdot w\right) \geq \frac{1}{p} \cdot \log u + \frac{1}{q} \cdot \log w.$$

By plugging in the definitions of u and w, we get using the monotonicity of f:

$$\log\left(\frac{a^p}{p}+\frac{b^q}{q}\right) \geq \log a + \log b \implies ab \leq \frac{a^p}{p}+\frac{b^q}{q}.$$

(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:

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

Solution: We first consider the case where $\|\vec{x}\|_p$ or $\|\vec{y}\|_q$ is equal to 0. In this case, note that either $\vec{x}$ or $\vec{y}$ must be 0. This means that the left hand side of Hölder's Inequality is 0 while the right hand side is non-negative which proves Hölder's Inequality in this case. Now consider the case where $\vec{x}$ and $\vec{y}$ are not equal to 0. With this setup, notice that Hölder's Inequality is equivalent to proving:

$$\left(\frac{\vec{x}}{\|\vec{x}\|_p}\right)^\top \left(\frac{\vec{y}}{\|\vec{y}\|_q}\right) \leq 1.$$

Let $\vec{u} = \frac{\vec{x}}{\|\vec{x}\|_p}$ and $\vec{w} = \frac{\vec{y}}{\|\vec{y}\|_q}$. We have:

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

Similarly, we have $\|\vec{w}\|_q = 1$. Now, we conclude the proof of Hölder's Inequality as follows:

$$\vec{u}^\top \vec{w} = \sum_{i=1}^n u_i w_i \leq \sum_{i=1}^n |u_i|\,|w_i| \leq \sum_{i=1}^n \left(\frac{|u_i|^p}{p}+\frac{|w_i|^q}{q}\right) = \frac{\|\vec{u}\|_p^p}{p}+\frac{\|\vec{w}\|_q^q}{q} = 1$$

where the second inequality follows from Young's Inequality proved in part (a).

(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 $\|\vec{y}\|_q = 1$.

Solution: Firstly, we get via Hölder's Inequality that:

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

We will now prove the other direction using a particular choice of y. First note that when $\vec{x} = 0$, the inequality is true as both the left and right sides of the equality are 0. Now, assume that $\vec{x} \neq 0$. In this case, $\|\vec{x}\|_p > 0$. Consider the vector, $\vec{y}$, defined as follows:

$$y_i = \operatorname{sgn}(x_i) \frac{|x_i|^{p-1}}{\|\vec{x}\|_p^{p-1}}.$$

With this definition of $\vec{y}$, we have:

$$\vec{x}^\top \vec{y} = \sum_{i=1}^n \frac{|x_i|^p}{\|\vec{x}\|_p^{p-1}} = \frac{\|\vec{x}\|_p^p}{\|\vec{x}\|_p^{p-1}} = \|\vec{x}\|_p .$$

We now need to verify that $\|\vec{y}\|_q = 1$. We do this via the following calculation using that fact that $\frac{1}{p} + \frac{1}{q} = 1$ implies that $pq = p + q$:

$$\|\vec{y}\|_q^q = \sum_{i=1}^n \frac{|x_i|^{pq-q}}{\|\vec{x}\|_p^{pq-q}} = \sum_{i=1}^n \frac{|x_i|^p}{\|\vec{x}\|_p^p} = 1.$$

By taking the q^{th} roots on both sides, this concludes the proof of the statement.

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

Solution: We will do this by verifying that $\|\cdot\|_p$ satisfies the three properties of a norm:

i. For the first property, assume that $\|x\|_p = 0$. This means that $x_i = 0$ for all $1 \leq i \leq n$ as:

$$\|\vec{x}\|^p \geq |x_i|^p \geq 0.$$

When $\vec{x} = 0$, we see from the definition of $\|\cdot\|_p$ that $\|\vec{x}\|_p = 0$.

ii. For the second property, we have:

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

iii. From, part (c), we see that $\|\cdot\|_p$ is a convex function as it is a pointwise maximum of convex (linear) functions. Therefore, we get:

$$\left\| \frac{\vec{x} + \vec{y}}{2} \right\|_p \leq \frac{\|\vec{x}\|_p}{2} + \frac{\|\vec{y}\|_p}{2}.$$

Now, using the second property, we conclude:

$$\|\vec{x} + \vec{y}\|_p \leq \|\vec{x}\|_p + \|\vec{y}\|_p .$$

From the previous three parts, we get that $\|\cdot\|_p$ indeed defines a norm.

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 $\mathbb{R}$

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

This is strictly convex since $\dfrac{d^2 f(x)}{dx^2} = e^x > 0$ for all $x \in \mathbb{R}$.

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

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

This is neither convex nor concave. The Hessian of f is

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

which has eigenvalues ±1 which implies the Hessian is neither positive semidefinite nor negative semidefinite.

(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.

Solution: For $g(\mu)$ we have,

$$\begin{aligned}\nabla g(\mu) &= \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} \\ \nabla^2 g(\mu) &= -\frac{n}{\sigma^2} < 0.\end{aligned}$$

Since σ is finite, g is strictly concave (equivalently f is strictly concave in μ).

For $h(z)$ we have,

$$\begin{aligned}\nabla h(z) &= \frac{n}{2z} - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2} \\ \nabla^2 h(z) &= -\frac{n}{2z^2} < 0.\end{aligned}$$

Since z^2 is finite ($\sigma > 0$), h is strictly concave (equivalently f is strictly concave in σ^2). For $f(\mu, \frac{1}{\sigma^2})$, we find the second order partial derivatives and stack them in the Hessian. We have,

$$\nabla^2 f(\mu, \frac{1}{\sigma^2}) = \begin{bmatrix} -\frac{n}{\sigma^2} & \sum_{i=1}^{n}(x_i - \mu) \\ \sum_{i=1}^{n}(x_i - \mu) & -\frac{n\sigma^4}{2} \end{bmatrix}.$$

The determinant of the Hessian is given by,

$$\det(\nabla^2 f) = \frac{n^2\sigma^2}{2} - (\sum_{i=1}^{n}(x_i - \mu))^2.$$

and the trace of the Hessian is given by,

$$\operatorname{tr}(\nabla^2 f) = -\frac{n}{\sigma^2} - \frac{n\sigma^4}{2} < 0$$

Note that the trace is the sum of the eigenvalues, and the determinant is the product of the eigenvalues. Since the trace is always negative, if the determinant is negative it must imply that one eigenvalue is positive and another is negative; that is, we have f is neither convex nor concave. It is easy to see that $\det(\nabla^2 f)$ can sometimes be negative – for example, if we choose σ^2 to be close to zero and μ away from x_i, the second negative term dominates and make $\det(\nabla^2 f) \leq 0$.

Aside: Note however, in the maximum likelihood estimates, the Hessian is negative semi-definite implying that locally the function is concave. More concretely, at

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

we have $\nabla^2 f(\hat{\mu}, 1/\hat{\sigma}^2) \preceq 0$

(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?

Solution: We will do this by verifying the second order sufficient conditions for convexity. We have the derivatives of f can be computed using the chain rule as follows:

$$f'(x) = \frac{\partial f}{\partial x} = \frac{e^x}{1+e^x}$$
$$f''(x) = \frac{\partial^2 f}{\partial x^2} = \frac{e^x}{(1+e^x)^2} > 0.$$

Since we have $f''(x) > 0$ for all x, we conclude that the function f is convex.

3. Quadratic inequalities

Consider the set S defined by the following inequalities:

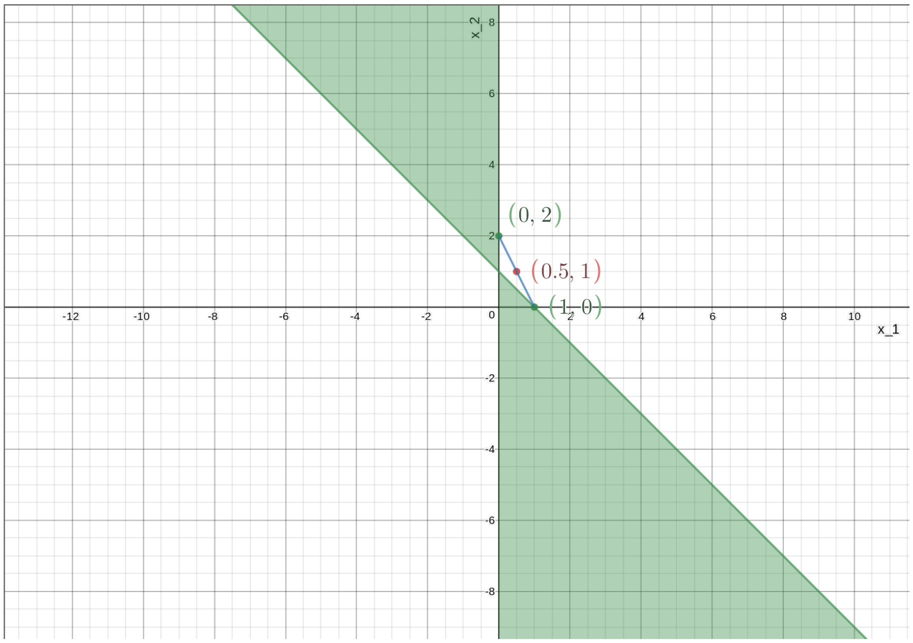
$$(x_1 \geq -x_2 + 1 \text{ and } x_1 \leq 0) \text{ or } (x_1 \leq -x_2 + 1 \text{ and } x_1 \geq 0)\,.$$

To be more precise,

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

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

Solution: The set S as shown in Fig. (a) is not convex. We can prove this by counterexample. $(0,2)$ and $(1,0)$ both belong to the set, but the midpoint $(1/2,1)$ does not.



(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?

Solution: Within set S, $x_1 + x_2 - 1 \geq 0$ when $x_1 \leq 0$ and $x_1 + x_2 - 1 \leq 0$ when $x_1 \geq 0$. It follows that $q(\vec{x}) = x_1(x_1 + x_2 - 1) \leq 0$ if and only if it is in the set. Expressing $q(\vec{x})$ in the desired form:

$$q(\vec{x}) = x_1^2 + x_1x_2 - x_1 = \vec{x}^\top A\vec{x} + 2\vec{b}^\top\vec{x} + c$$

where

$$A = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 0 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -1/2 \\ 0 \end{bmatrix}, \quad c = 0.$$

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

Solution: The convex hull of the set is the whole space, $\mathbb{R}^2$. To see this note than any point $z = (z_1, z_2) \in \mathbb{R}^2$ can be written as $z = \frac{x+y}{2}$ with $x, y \in S$ as follows:
$x = (2z_1, 1 - 2z_1), y = (0, 2(z_1 + z_2) - 1)$.

(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 $\vec{x} \in S_1$.

Solution: We first compute the unconstrained optimal value of f. Notice that f is a convex function. Therefore, we can compute its optimal value by computing its gradient and setting it to 0. Doing so, we obtain the optimal value of f to be 0 attained at the point $\vec{x}^* = (-1, 3)$. Now, since $\vec{x}^* \in S_1$, $\vec{x}^*$ is the solution to the constrained optimization problem as well.

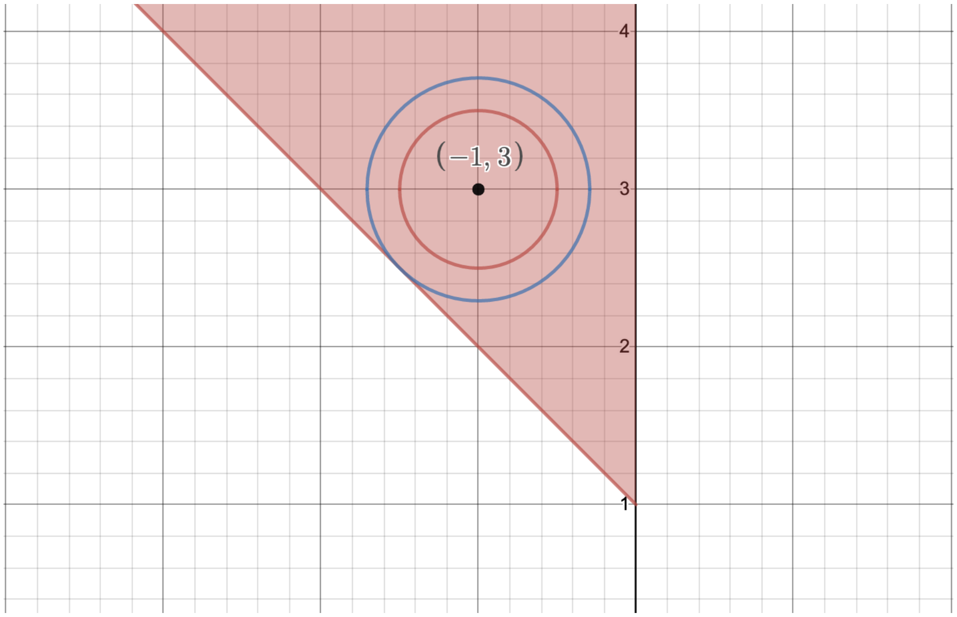


Figure 1: This figure illustrates the position of the optimum, $x^* = (-1, 3)$, and the level sets of the objective function, f, which are concentric circles around x^*.

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

Solution:

Proceeding as in the proof for the previous problem, we first find the solution to the unconstrained optimization problem. We get that the unconstrained problem is minimized at the point $\vec{x}_u^* = (-2, 2)$. However, this point is not in the feasible set, S_1. Therefore, the true optimum, $\vec{x}^*$, has one or more constraints active. Now, we will attempt to solve the problem with one active constraint. Suppose the one active constraint is $x_1 \geq -x_2 + 1$. Since this constraint is active, we must try and minimize $f(\vec{x})$ subject to $\vec{x}$ satisfying $x_1 = -x_2 + 1$. Note that any point on this line can be written in the form $(0, 1) + \alpha(-1, 1)$. Now consider the function, $g(\alpha)$:

$$g(\alpha) = f((0, 1) + \alpha(-1, 1)) = (\alpha - 2)^2 + (\alpha - 1)^2.$$

Note that the function, $f(\alpha)$, is convex in α. Therefore, we can minimize $g(\alpha)$ by taking its derivative and setting it to 0. By doing this, we get that $\alpha = 3/2$ is the unique minimizer of $g(\alpha)$. Therefore, the minimizer of f subject to $x_1 = -x_2 + 1$ is the point $(-3/2, 5/2)$. Similarly, the minimizer of f assuming the second constraint, $x_1 \leq 0$, is active is obtained at the point $(-2, 2)$. However, the point $(-2, 2)$ is not in S_1. The final possibility is that both constraints are active. However, the optimal value of f subject to both constraints being active will be greater than the value of f obtained at $(-3/2, 5/2)$ which is in S_1. Therefore, we get that $f(\vec{x})$ is minimized at the point $\vec{x}^* = (-3/2, 5/2)$ subject to $\vec{x} \in S_1$. There is one active constraint at $\vec{x}^*$.

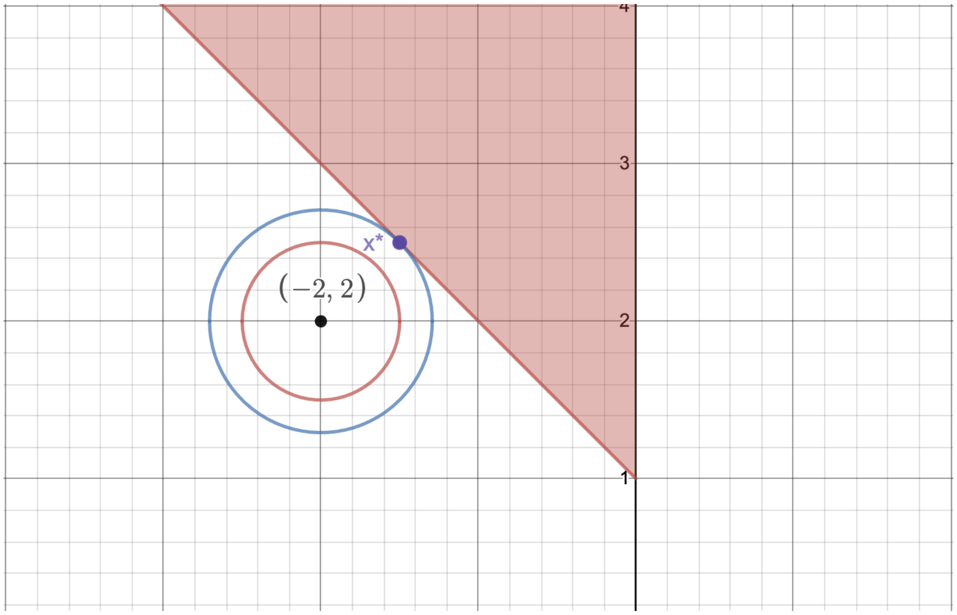


Figure 2: This figure illustrates the position of the optimum, $x^* = (-3/2, 5/2)$, and the level sets of the objective function, f, which are concentric circles around $(-2, 2)$. Note that in this case, the unconstrained optimum does not lie in the set, S_1 and the optimal point lies on the boundary of one of the constraints.

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

Solution: Proceeding as before, we first check the case where 0 constraints are active. However, the unconstrained minimizer of f is $(0, 0)$ which is not in S_1. Now, we check the cases where one of the constraints is active. Assume that the constraint $x_1 \leq 0$ is active. In this case the optimizer is again obtained at the point $(0, 0)$ which is not in S_1. We then consider the case where the constraint $x_1 \geq -x_2 + 1$ is active. As before, we define the function, $g(\alpha)$ as:

$$g(\alpha) = f((0, 1) + \alpha(-1, 1)) = \alpha^2 + (\alpha + 1)^2.$$

By optimizing over α by setting its gradient with respect to α and setting it to 0, we get the optimal setting of α is $-1/2$. However, note that the point $(1/2, 1/2)$ does not belong to S_1 either. Therefore, the only remaining possibility is the possibility that both constraints are active. This can happen solely at the point $(0, 1)$. At this point, the value of the function f is 1, the optimizer $\vec{x}^* = (0, 1)$ and both constraints are active at $\vec{x}^*$.

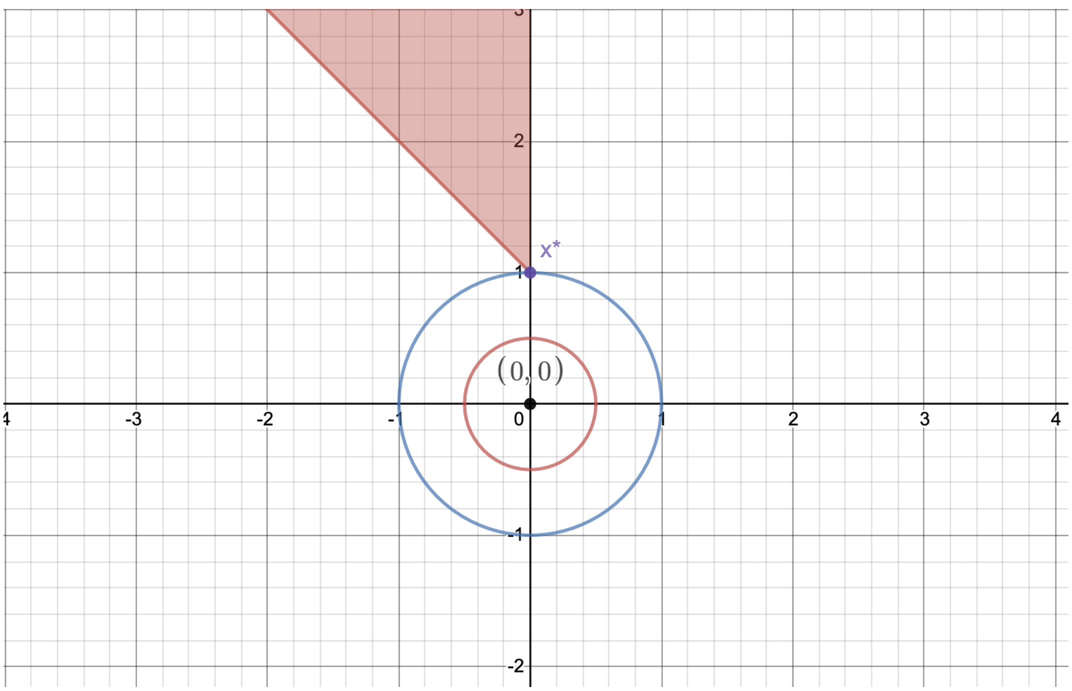


Figure 3: This figure illustrates the position of the optimum, $x^* = (0,1)$, and the level sets of the objective function, f, which are concentric circles around $(0,0)$. Note that in this case, the unconstrained optimum does not lie in the set, S_1 and the optimal point lies on the boundary of *both* of the constraints.

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 $\eta = 0.2$.

Solution: For the given function, we have $\nabla f(x) = (x-2)$; therefore, the gradient descent algorithm gives

$$x[k+1] = x[k] - \eta(x[k]-2).$$

By subtracting 2 from both sides, we obtain

$$(x[k+1]-2) = (1-\eta)(x[k]-2) \implies (x[k]-2) = (1-\eta)^k(x[0]-2).$$

Given $\eta = 0.2$, we have

$$|x[k] - 2| = 0.8^k |x[0] - 2| \to 0 \quad \text{as } k \to \infty,$$

which shows that $x[k]$ converges to 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?

Solution: From the solution for part (a), we have

$$|x[k] - 2| = |1 - \eta||x[0] - 2| \quad \forall k \in \mathbb{N}.$$

For convergence of the algorithm for every initialization, it is necessary and sufficient to have $|1 - \eta| < 1$, which is equivalent to $\eta \in (0, 2)$. If η is chosen larger than 2, $x[k]$ oscillates around 2 while $|x[k]|$ grows unboundedly.

(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 $A^\top A$.

Solution: We can write $f(\vec{x}) = \frac{1}{2}(\vec{x}^\top A^\top A\vec{x} - \vec{x}^\top A^\top \vec{b} - \vec{b}^\top A\vec{x} + \vec{b}^\top \vec{b})$, so

$$\nabla f(\vec{x}) = A^\top A\vec{x} - A^\top \vec{b}.$$

Then the gradient descent algorithm gives

$$\vec{x}[k+1] = \vec{x}[k] - \eta\left(A^\top A\vec{x}[k] - A^\top b\right) = \vec{x}[k] - \eta A^\top A\left(\vec{x}[k] - (A^\top A)^{-1}A^\top \vec{b}\right).$$

By subtracting $(A^\top A)^{-1}A^\top \vec{b}$ from both sides, we obtain

$$\left(\vec{x}[k+1] - (A^\top A)^{-1}A^\top \vec{b}\right) = \left(I - \eta A^\top A\right)\left(\vec{x}[k] - (A^\top A)^{-1}A^\top \vec{b}\right)$$

and consequently,

$$\left(\vec{x}[k] - (A^\top A)^{-1}A^\top \vec{b}\right) = \left(I - \eta A^\top A\right)^k \left(\vec{x}[0] - (A^\top A)^{-1}A^\top \vec{b}\right).$$

For the algorithm to converge, we need the largest eigenvalue of $(I - \eta A^\top A)$ to be less than 1 in absolute value. Since any eigenvector v of $A^\top A$ is also eigenvector of $I - \eta A^\top A$ with eigenvalue $1 - \eta\lambda$, we have

$$\forall i, \ |1 - \eta\lambda_i(A^\top A)| < 1.$$

Unwrapping the absolute values, this implies

$$\forall i, \ -1 < 1 - \eta\lambda_i(A^\top A) < 1.$$

Therefore,

$$\forall i, \ 0 < \eta\lambda_i(A^\top A) < 2$$

$$\forall i, \eta\lambda_i(A^\top A) < 2 \iff \eta\lambda_{\max}(A^\top A) < 2 \iff \eta < \frac{2}{\lambda_{\max}(A^\top A)}.$$

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