

CMPUT 466 Assignment 2

Arun Woosaree

September 26, 2021

Problem 1

Part 1

$$\{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$$

This set is not convex.

Proof. Assume the set is convex.

Let's draw a line from $(x_1, x_2) = (-1, 0)$ to $(1, 0)$

These two points are part of the set because $(-1)^2 + 0^2 = 1$,
and $1^2 + 0^2 = 1$

That is, $x_1 \in [-1, 1], x_2 = 0$

If this set is convex, then any point on this line must also be in the set.

Let's select the point $(0, 0)$. i.e., $\lambda = 0.5$

This point is clearly part of the line defined above, however, it is definitely not in the set because $0^2 + 0^2 = 0 \neq 1$.

Therefore, we have found a counterexample that proves that this set is not convex. \square

Part 2

$$\{(x_1, x_2) : |x_1| + |x_2| \leq 1\}$$

This set is convex.

We begin by looking at the definition:

let $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ be any two points in the set.

Using the definition, if the set is convex, then

$$\lambda x + (1 - \lambda)y, \lambda \in (0, 1)$$

is also in the set.

Applying this definition, we get the point:

$$\begin{pmatrix} \lambda x_1 + (1 - \lambda)y_1 \\ \lambda x_2 + (1 - \lambda)y_2 \end{pmatrix}$$

If this point is in the set, we need to prove that:

$$|\lambda x_1 + (1 - \lambda)y_1| + |\lambda x_2 + (1 - \lambda)y_2| \leq 1$$

Proof. Using the triangle inequality: $|x + y| \leq |x| + |y|$:

$$|\lambda x_1 + (1 - \lambda)y_1| + |\lambda x_2 + (1 - \lambda)y_2| \leq |\lambda x_1| + |(1 - \lambda)y_1| + |\lambda x_2| + |(1 - \lambda)y_2|$$

we can factor out λ and $(1 - \lambda)$ from the right side, because by definition, $\lambda \in (0, 1)$:

$$\lambda(|x_1| + |x_2|) + (1 - \lambda)(|y_1| + |y_2|)$$

Because \mathbf{x} and \mathbf{y} are in the set, we know that:

$$|x_1| + |x_2| \leq 1$$

$$|y_1| + |y_2| \leq 1$$

Because of the above and $\lambda \in (0, 1)$, we also know that:

$$\lambda(\dots) + (1 - \lambda)(\dots) \leq 1$$

(The values where the dots are were determined to be ≤ 1)

Thus,

$$\lambda(|x_1| + |x_2|) + (1 - \lambda)(|y_1| + |y_2|) \leq 1$$

And because:

$$|\lambda x_1 + (1 - \lambda)y_1| + |\lambda x_2 + (1 - \lambda)y_2| \leq \lambda(|x_1| + |x_2|) + (1 - \lambda)(|y_1| + |y_2|) \leq 1$$

We have finally proven that:

$$|\lambda x_1 + (1 - \lambda)y_1| + |\lambda x_2 + (1 - \lambda)y_2| \leq 1$$

, which means that the set is indeed convex. □

Problem 2

$$f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1x_2$$

a) x_1 variable, x_2 constant:

$$\begin{aligned} \nabla^2 f(x_1) &= \nabla^2 x_1 - 4x_2 \\ &= 2 \geq 0 \end{aligned}$$

Because the second order gradient of $f(x_1, x_2)$ is positive when x_2 is constant, we have determined that f is convex in x_1

b) x_2 variable, x_1 constant

$$\begin{aligned}\nabla^2 f(x_2) &= \nabla^2 x_2 - 4x_1 \\ &= 2 \geq 0\end{aligned}$$

Because the second order gradient of $f(x_1, x_2)$ is positive when x_1 is constant, we have determined that f is convex in x_2

c) $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

```

1 import numpy as np
2 from numpy import linalg as LA
3 # in a) and b) we found the values 2 and 2 in the Hessian matrix
4 # -4 is obtained because del/ (del x_1 del_x2) f(x_1, x_2) = -4
5 # same for del/ (del x_2 del_x1) because the matrix is symmetric
6 H = np.array([
7     [2, -4],
8     [-4, 2]
9 ])
10 eigenval, eigenvec = LA.eig(H)
11 print(f"eigenval: {eigenval} | eigenvector: {eigenvec} ")
12
13 if any(map(lambda x: x < 0, eigenval)):
14     print("Because one of the eigenvalues are negative, we conclude that the function is not")
15 else:
16     print("Because all of the eigenvalues are positive, we conclude that the function is convex")
eigenval: [ 6. -2.] | eigenvector: [[ 0.70710678  0.70710678]
[-0.70710678  0.70710678]]
Because one of the eigenvalues are negative, we conclude that the function is not convex

```

Problem 3

If f is a differentiable convex function, then f satisfies the first-order condition

Proof. Given f is a differentiable convex function, the definition holds:

$$\forall x, y \in \text{dom } f, \forall \lambda \in (0, 1), f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

Let's do some re-arranging:

$$f(\lambda x + y - \lambda y) \leq \lambda f(x) + f(y) - \lambda f(y)$$

$$f(\lambda(x - y) + y) \leq \lambda(f(x) - f(y)) + f(y)$$

$$\Rightarrow \frac{f(\lambda(x-y) + y) - f(y)}{\lambda} \leq f(x) - f(y)$$

The left side of this equation looks exactly like the definition of the derivative using limits.

As $\lambda \rightarrow 0$,

$$\begin{aligned}\Rightarrow \nabla[f(y)]^T(x-y) &\leq f(x) - f(y) \\ \nabla[f(y)]^T(x-y) + f(y) &\leq f(x)\end{aligned}$$

This is equivalent to the first-order condition:

$$f(y) \geq f(x) + \nabla[f(x)]^\top(y-x), \forall x, y \in \text{dom } f$$

□

Problem 4

Proof. Because

$$z = (1-\lambda)x + \lambda y$$

, we can find $\|y-x\|$ with some rearranging:

$$z = x - \lambda x + \lambda y = x + \lambda(y-x)$$

$$z - x = \lambda(y-x)$$

$$\|z-x\| = \|\lambda(y-x)\| = \lambda\|y-x\|$$

Using $\lambda = \frac{\epsilon}{2\|y-x\|}$:

$$\lambda\|y-x\| = \|z-x\| = \frac{\epsilon}{2}$$

So,

$$\lambda\|y-x\| < \epsilon$$

Because $\lambda \in (0, 1)$, λ cannot be greater than 1, so:

$$\|y-x\| < \epsilon$$

□

Problem 5

Intuitively, we know that if the gradient of f is $\mathbf{0}$ at a point, it is either a local minimum or local maximum of the function. We know that it must be a local minimum because no part of the function curves down. We know this, because if any neighboring point around \mathbf{x} does not have a gradient equal to $\mathbf{0}$, we can simply do gradient descent to get to the minimum, where the gradient is $\mathbf{0}$. In class, we went over a theorem with a proof that if \mathbf{x} is a local minimum, and f is convex, then \mathbf{x} is a global minimum. Because \mathbf{x} is a local minimum, by this theorem, \mathbf{x} is also a global minimum.

Now, for a more formal proof:

Proof. From the first-order condition for convex functions we have:

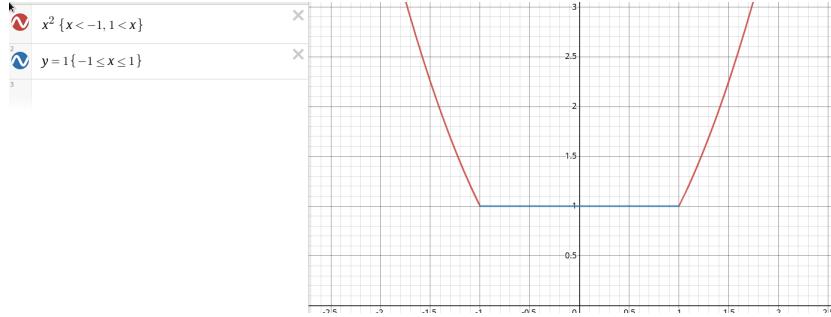
$$\forall x, y \in \text{dom} f, f(y) \geq f(x) + [\nabla_x f(x)]^\top (y - x)$$

Because $\nabla f(x) = \mathbf{0}$,

$$\forall x, y \in \text{dom} f, f(y) \geq f(x)$$

This is precisely the definition of a global optimum. \square

The global optimum is not necessarily unique. Picture a convex function where the minimum is a straight line. Any point along this line would be the minimum of the function, while still satisfying the condition $\forall x, y \in \text{dom} f, f(y) \geq f(x)$. For example, the plot below:



Problem 6

$$J = \sum_{m=1}^M \left(\underbrace{\sum_{i=0}^n w_i x_i^{(m)}}_{{X}_w} - \underbrace{t^{(m)}}_{\bar{t}} \right)^2 + \underbrace{\sum_{i=0}^n w_i^2}_{w^T w}$$

$$(X_w - t)^T (X_w - t) + w^T w$$

$$(w^T X^T - t^T)(Xw - t) + w^T w$$

$$\underbrace{w^T X^T X w}_\text{EIR} - \underbrace{t^T X w}_\text{EIR} - \underbrace{w^T X t + t^T t}_\text{EIR} + \underbrace{w^T w}_\text{EIR}$$

$$(w^T x^T t)^T = t^T X w$$

Scalar transpose is itself.

$$J = w^T X^T X w - 2t^T X w + t^T t + w^T w$$

$$\nabla J(w) = 2X^T X w - 2(t^T X)^T + 2w$$

$$= 2X^T X w - 2X^T t + 2w$$

$$\nabla \nabla J(w) = 2(X^T X)^T + 2$$

$$= 2\underbrace{X^T X}_{\geq 0} + 2$$

$\underbrace{\quad}_{\text{definitely}} > 0$

because this value is greater than 0, J is convex in w

Problem 7

Prove $J(w - \alpha u) \leq J(w)$ for small enough α
Taylor approximation:

$$J(w - \alpha u) \approx J(w) + [\nabla J(w)]^T [-\alpha u]$$

$$+ o(\alpha^2)$$

$\underbrace{\quad}_{\text{some other higher order}}$

terms

as $\alpha \rightarrow 0$

$$= J(w) - \alpha [\nabla J(w)]^T u + \underbrace{o(\alpha^2)}$$

small, can ignore as
 $\alpha \rightarrow 0$

$$J(w - \alpha u) + \alpha [\nabla J(w)]^T u = Jw + \underbrace{o(\alpha^2)}_{\geq 0} \text{ and } \alpha \geq 0$$

so,

$$J(w - \alpha u) \leq Jw$$

Problem 8

In #6 we computed the gradient:

$$\nabla J(w) = 2X^T X w - 2X^T t + 2w$$

Gradient descent:

initialize $w^{(0)}$ randomly

for iteration $t=1, 2, 3, \dots$ until satisfied

Compute gradient $\nabla J(w)(w=w^{(t-1)})$

$$g(t) = 2X^T X w^{(t-1)} - 2X^T t + 2w^{(t-1)}$$

update parameters $w^{(t)} = w^{(t-1)} - \alpha \nabla J(w) \Big|_{w=w^{(t-1)}}$

$$w^{(t)} = w^{(t-1)} - \alpha \left[2X^T X w^{(t-1)} - 2X^T t + 2w^{(t-1)} \right]$$

↓
step size

return last $w^{(t)}$