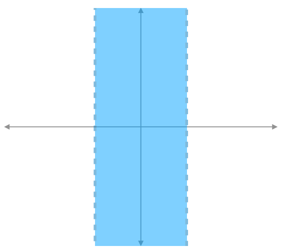
Math 4650/MSSC 5650 - Homework 3

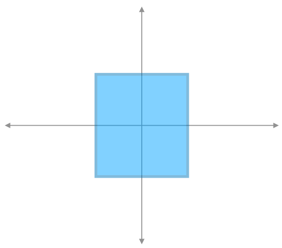
Henri Medeiros Dos Reis

Spring 2023

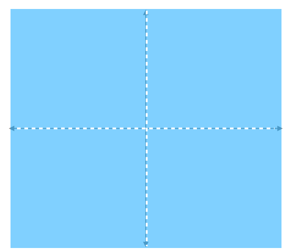
Problem 1 (5 pts)**.** For each of the four subsets of $\mathbb{R}^2$ shown below, select all terms that apply: *open*, *closed*, *bounded*, *unbounded*, *compact*. Briefly justify your selections.



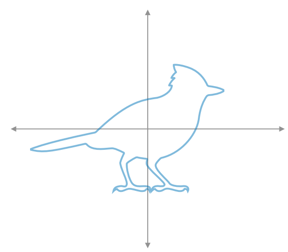
$A = \{(x, y) : |x| < 1\}$



$B = \{(x, y) : |x| \leq 1, |y| \leq 1\}$



$C = \{(x, y) : x \neq 0,\ y \neq 0\}$



$D = \{(x(t), y(t)) : t \in [0,1]\}$
where $x(t)$ and $y(t)$ are continuous

Solution 1. (a) Open, it is possible to take any point (x,y) in A and choose $\epsilon > 0$, such that the open ball $B((x, y), \epsilon)$ is contained in A.

Unbounded, since there are no limits on the y-axis.

(b) Closed, since all the limits of the set are included in the set.

Bounded, the set in contained in the square $[-1, 1] * [1, 1]$.

Compact, since it is closed and bounded.

(c) Open, it is possible to take any point (x,y) in A and choose $\epsilon > 0$, such that the open ball $B((x, y), \epsilon)$ is contained in C.

Unbounded, since there is no limit on this set.

(d) Closed, since we are including 0 and 1, in other words, we are including the limits of the set.

Bounded, we can fin an R large enough to make an open ball that would contain this entire set.

Compact, since it is closed and bounded.

Problem 2 (5 pts)**.** Compute the directional derivative of the function $f(x) = \sin\left(\frac{x_1+x_2}{2}\right)$ at the point $x = (0,0)^\top$ in the direction $d = (1,1)^\top$ directly from the limit definition:

$$f'(x;d) = \lim_{t\to 0^+} \frac{f(x+td) - f(x)}{t}$$

and verify this is the same as what you get from the formula $f'(x;d) = \nabla f(x)^\top d$.

Solution 2.

$$\begin{aligned} f'(x;d) &= \lim_{t\to 0^+} \frac{f(x+td) - f(x)}{t} \\ &= \lim_{t\to 0^+} \frac{sin(\frac{x_1+td_1+x_2+td_2}{2}) - sin(\frac{x_1+x_2}{2})}{t} \end{aligned}$$

And since $x = (0,0)^\top$ and $d = (1,1)^\top$, then:

$$\begin{aligned} f'(x;d) &= \lim_{t\to 0^+} \frac{sin(\frac{0+t+0+t}{2}) - sin(\frac{0}{2})}{t} \\ &= \lim_{t\to 0^+} \frac{sin(t) - sin(0)}{t} \\ &= \lim_{t\to 0^+} \frac{sin(t)}{t} \end{aligned}$$

And we can use l'Hospital's rule, since the numerator and denominator go to zero.

$$\begin{aligned} f'(x;d) &= \lim_{t\to 0^+} \frac{cos(t)}{1} \\ &= cos(0) = 1 \end{aligned}$$

Then it is possible to verify the formula $f'(x;d) = \nabla f(x)^\top d$:

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} \\ &= \begin{bmatrix} cos(\frac{x_1+x_2}{2})(1/2+0) \\ cos(\frac{x_1+x_2}{2})(0+1/2) \end{bmatrix} \\ &= \begin{bmatrix} cos(0)(1/2) \\ cos(0)(1/2) \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \end{aligned}$$

Then $\nabla f(x)^\top d = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 = f'(x;d)$

Problem 3 (5 pts)**.** Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x,y) = e^{-(2x^2+y^2)}$.

(a) Find the gradient $\nabla f(x, y)$ and the Hessian $\nabla^2 f(x, y)$.

(b) Find the 2nd-order Taylor series approximation of $f(x, y)$ centered at $(x, y) = (0, 0)$.

Solution 3. (a)

$$\begin{aligned}\nabla f(x,y) &= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \\ &= \begin{bmatrix} e^{-(2x^2+y^2)}(-4x) \\ e^{-(2x^2+y^2)}(-2y) \end{bmatrix}\end{aligned}$$

And the Hessian $\nabla^2 f(x, y)$

$$\begin{aligned}\nabla^2 f(x,y) &= \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial xy} \\ \frac{\partial^2 f}{\partial yx} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \\ &= \begin{bmatrix} -4e^{-(2x^2+y^2)} + 16x^2 e^{-(2x^2+y^2)} & 8xye^{-(2x^2+y^2)} \\ 8xye^{-(2x^2+y^2)} & -2e^{-(2x^2+y^2)} + 4y^2 e^{-(2x^2+y^2)} \end{bmatrix}\end{aligned}$$

(b) $f(0, 0) = e^0 = 1$, $\nabla f(0, 0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and $\nabla^2 f(0, 0) = \begin{bmatrix} -4 & 0 \\ 0 & -2 \end{bmatrix}$

Then the 2nd-order Taylor series approximation is:

$$\begin{aligned}f(x,y) &\approx f(0,0) + \nabla f(0,0)^T[(x,y) - (0,0)] + \frac{1}{2}[(x,y) - (0,0)]^T \nabla^2 f(0,0)[(x,y) - (0,0)] \\ &= 1 + \frac{1}{2}\begin{bmatrix} x \\ y \end{bmatrix}\begin{bmatrix} -4 & 0 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} x & y \end{bmatrix} \\ &= 1 + \frac{1}{2}(-4x^2 - 2y^2)\end{aligned}$$

Problem 4 (5 pts). Compute the gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$ for each definition of the function $f : \mathbb{R}^n \to \mathbb{R}$ given below. Express your answers as compactly as you can using matrix-vector notation. (*Hint: Expand these functions using the identity* $\|v\|^2 = v^\top v$)

(a) $f(x) = \frac{1}{2}\|x - a\|^2$, where $a \in \mathbb{R}^n$ is any fixed vector.

(b) $f(x) = \frac{1}{2}\|Ax\|^2$, where $A \in \mathbb{R}^{m \times n}$ is any fixed matrix.

(c) $f(x) = \frac{1}{2}\|Ax - b\|^2$, where $A \in \mathbb{R}^{m \times n}$ is any fixed matrix, and $b \in \mathbb{R}^m$ is any fixed vector.

Solution 4. (a)

$$\begin{aligned}f(x) &= \frac{1}{2}\|x - a\|^2 \\ &= \frac{1}{2}(x - a)^T(x - a) \\ &= \frac{1}{2}(x^T x - x^T a - a^T x - a^T a) \\ &= \frac{1}{2}(\|x\|^2 - 2x^T a - \|a\|^2)\end{aligned}$$

Now, let's take the gradient by using the linear property of the gradient.

$$\nabla \|x\|^2 = 2x, \ \nabla x^T a = a, \ \nabla \|a\|^2 = 0$$

Then $\nabla f(x) = \frac{1}{2}(2x - 2a) = x - a$. And now we can take the Hessian:

$$\nabla^2 f(x) = I, \text{ since } \frac{\partial^2 f}{\partial x_i x_j} = \frac{\partial f}{\partial x_i}(x_j - a_j) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

(b)

$$\begin{aligned} f(x) &= \frac{1}{2}\|ax\|^2 \\ &= \frac{1}{2}(Ax)^T Ax \\ &= \frac{1}{2}x^T A^T Ax \end{aligned}$$

And since $A^T A$ is always a symmetric matrix, then the gradient is:

$$\begin{aligned} \nabla f(x) &= \frac{1}{2}(2A^T Ax) \\ &= A^T Ax \end{aligned}$$

And the Hessian $\nabla^2 f(x) = A^T A$

(c)

$$\begin{aligned} f(x) &= \frac{1}{2}\|Ax - b\|^2 \\ &= \frac{1}{2}(Ax - b)^T (Ax - b) \\ &= \frac{1}{2}(x^T A^T - b^T)(Ax - b) \\ &= \frac{1}{2}(x^T A^T Ax - 2b^T Ax + b^T b \end{aligned}$$

Then, let's take the gradient by using the linear property of the gradient.

$$\nabla(x^T A^T Ax) = 2A^T Ax$$

$$\nabla(b^T Ax) = (b^T A)^T = A^T b \text{ since } \nabla(c^T x) = c\text{for a vector c, and } b^T A \text{ is a vector}$$

$$\nabla(b^T b) = 0$$

Therefore, $\nabla f(x) = 1/2(2A^T Ax - 2A^T b) = A^T Ax - A^T b$

And the Hessian $\nabla^2 f(x) = A^T A$

Problem 5 (5 pts)**.** Compute the gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$ of the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = e^{-\|x\|^2}$. Express your answers as compactly as you can using matrix-vector notation.

Solution 5.

$$f(x) = e^{-\|x\|^2} = e^{-x^T x}$$

Then we can compute the gradient

$$\nabla f(x) = e^{-x^T x}(-2x) = -2xe^{-\|x\|^2}$$

And the Hessian

$$\begin{aligned}\nabla^2 f(x) &= e^{-x^T x}(-2I) + e^{-x^T x}(-2x)(-2x^T) \\ &= -2Ie^{-\|x\|^2} + 4xx^T e^{-\|x\|^2}\end{aligned}$$

Where I is the $n \times n$ identity matrix.

Problem 6 (MSSC, 5pts)**.** A function f defined over an open set $U \subset \mathbb{R}^n$ is defined to be *continuous at the point* $x \in U$ if for all sequences of points $\{x_k\}$ contained in U with $\lim_{k\to\infty} x_k = x$ we have $\lim_{k\to\infty} f(x_k) = f(x)$. Using this definition, show that the function

$$f(x,y) = \begin{cases} \frac{x^2-y^2}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ c & \text{if } (x,y) = (0,0) \end{cases}$$

is *not* continuous at the point $(x,y) = (0,0)$ *for any value of* $c \in \mathbb{R}$. (In other words, there is no value c we can assign f at the origin to make the function continuous there).

Solution 6. To show that $f(x,y)$ is not continuous at $(0,0)$ for any value of c, we need to find 2 sequences of points x_k and y_k that converge to the point $(0,0)$ but the $\lim_{k\to\infty} f(x_k) \neq \lim_{k\to\infty} f(y_k)$, which implies that the function is not continuous at $(0,0)$.

Let $x_k = (\frac{1}{k}, \frac{1}{k})$ then $\lim_{k\to\infty} x_k = (0,0)$ and

$$\begin{aligned}\lim_{k\to\infty} f(x_k) &= \lim_{k\to\infty} \frac{\frac{1}{k}^2 - \frac{1}{k}^2}{\frac{1}{k}^2 - \frac{1}{k}^2} \\ &= \lim_{k\to\infty} \frac{0}{2(\frac{1}{k}^2)} \\ &= 0\end{aligned}$$

Therefore $\lim_{k\to\infty} f(x_k) = 0$.

Now, let's consider $y_k = (\frac{1}{k}, -\frac{1}{k})$ then $\lim_{k\to\infty} y_k = (0,0)$. However,

$$\begin{aligned}
\lim_{k\to\infty} f(y_k) &= \lim_{k\to\infty} \frac{\frac{1}{k}^2 - (-\frac{1}{k}^2)}{\frac{1}{k}^2 - (-\frac{1}{k}^2)} \\
&= \lim_{k\to\infty} \frac{\frac{1}{k}^2 + \frac{1}{k}^2}{\frac{1}{k}^2 + \frac{1}{k}^2} \\
&= \lim_{k\to\infty} \frac{2}{k^2}\frac{k^2}{2} \\
&= 1
\end{aligned}$$

Therefore $\lim_{k\to\infty} f(y_k) = 1$.

And since $\lim_{k\to\infty} f(x_k) \neq \lim_{k\to\infty} f(y_k)$, $f(x,y)$ is not continuous at $(0,0)$ for any value of c.

Problem 7 (MSSC, 5pts)**.** Let $Q(x; x_0)$ be the quadratic approximation of a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $x_0 \in \mathbb{R}^n$ defined by

$$Q(x; x_0) = f(x_0) + \nabla f(x_0)^\top (x - x_0) + \frac{1}{2}(x - x_0)^\top \nabla^2 f(x_0)(x - x_0) \quad \text{for all} \quad x \in \mathbb{R}^n, \qquad (\star)$$

Recall from lecture that the error between $f(x)$ and $Q(x; x_0)$ obeys the bound

$$|f(x) - Q(x; x_0)| = o(\|x - x_0\|^2)$$

for all points x close to x_0 where $o(\cdot)$ is a function such that $o(t)/t \to 0$ as $t \to 0^+$.

For this problem you will demonstrate the above error bound numerically for the function

$$f(x) = e^{-\|x\|^2} \text{ for } x \in \mathbb{R}^{10}$$

at the point $x_0 = (1, 1, ..., 1)^\top \in \mathbb{R}^{10}$. In particular, you will show that for $x = x_0 + \epsilon d$ where $d \in \mathbb{R}^{10}$ is a random unit vector and ϵ is a positive scalar, we have $\frac{|f(x) - Q(x;x_0)|}{\epsilon^2} \to 0$ as $\epsilon \to 0$.

- First, define a function `f(x)` in MATLAB that computes $f(x)$ above, such that the input `x` to the function is a 10×1 column vector.

- Now define a MATLAB function `Q(x,x0)` that takes in a pair of 10×1 column vectors `x` and `x0` and gives the output indicated in equation $(\star)$. (Reminder: Problem 5 has you compute the gradient and Hessian of f.)

- Define the vector x_0 in MATLAB using the command `x0 = ones(10,1);`, and define a random unit vector `d` using the commands

```
rng(1); %fix random seed
d = randn(10,1);
d = d/norm(d);
```

Also, initialize an error array with `err = [];`.

- Next, define a for-loop over `k=1:200`. Inside the for-loop, define `ep = 1/k;` and `x = x0 + ep*d;`, then compute and store the scaled absolute error with `err(end+1) = abs(f(x)-Q(x,x0))/ep^2;`
- Finally, show that the scaled absolute error decays to zero as ϵ goes to zero by plotting the scaled error array `err` on a log-scale using the command `semilogy(err)`; include this plot in your write-up.

**Include in your write-up for this problem a print-out/screenshot of your MATLAB code.

Solution 7. .

```
%% Define a function f(X), that takes as input a column vector
f = @(x) exp(1)^(-norm(x)^2);
%% Helper funcions
grad = @(x) -2*x*exp(-norm(x)^2);
Hessian = @(x) -2*eye(10)* ...
    exp(1)^(-norm(x)^2)+4*x*x'*exp(1)^(-norm(x)^2);
%% Define a function Q(x,x0), that the output is the equation *
Q = @(x,x0) f(x0)+grad(x0)'* ...
    (x-x0)+1/2*(x-x0)'*Hessian(x0)*(x-x0);
%% Define the vector x0 and define a random unit vector
x0 = ones(10,1);
rng(1); %fix random seed
d = randn(10,1);
d = d/norm(d);
err = [];
%% For loop to compute and store scaled absolute error
for k=1:200
    ep = 1/k;
    x = x0+ep*d;
    err(end+1)= abs(f(x)-Q(x,x0))/ep^2;
end
%% show that the scaled absolute error decays to zero as ep
   goes to zero
semilogy(err);
```