Mock Midterm 2 Solutions

EECS 245

Name: ______________________________

uniqname: ______________________________

UMID: ______________________________

Room: ◯ 1014 DOW ◯ At Home

Instructions

- This exam consists of 8 problems, worth a total of (N/A) points, spread across 14 pages (7 sheets of paper).
- You have 120 minutes to complete this exam, unless you have extended-time accommodations through SSD.
- Write your uniqname in the top right corner of each page.
- For free response problems, you must show all of your work (unless otherwise specified), and circle your final answer. We will not grade work that appears elsewhere, and you may lose points if your work is not shown.
- For multiple choice problems, completely fill in bubbles and square boxes; if we cannot tell which option(s) you selected, you may lose points.
  ◯ A bubble means that you should only select one choice.
  ☐ A square box means you should select all that apply.
- You may refer to a single two-sided handwritten notes sheet. Other than that, you may not refer to any other resources or technology during the exam (no phones, watches, or calculators).

You are to abide by the University of Michigan/Engineering Honor Code. To receive a grade, please sign below to signify that you have kept the Honor Code pledge.

I have neither given nor received aid on this exam, nor have I concealed any violations of the Honor Code.

This page has been intentionally left blank.

Problem 1: Inversion

Let $A = \begin{bmatrix} 2 & -3 \\ 8 & -10 \end{bmatrix}$.

a) Find $\det(A)$, the determinant of A.

$\det(A) =$ **4**

Solution: For a 2×2 matrix,

$$\det(A) = ad - bc$$

So

$$\det(A) = (2)(-10) - (-3)(8) = -20 + 24 = 4$$

b) Find the area of the four-sided polygon in $\mathbb{R}^2$ with vertices $(0,0)$, $(1,4)$, $(-2,-6)$, and $(-3,-10)$.

Area $=$ **2**

Solution: The vectors from the origin to two adjacent vertices are

$$\begin{bmatrix} 1 \\ 4 \end{bmatrix} \qquad \begin{bmatrix} -3 \\ -10 \end{bmatrix}$$

The parallelogram they form has area

$$|\det(A)| = 4$$

The four-sided polygon in the problem is half of that parallelogram, so its area is

$$2$$

c) Find A^{-1}, the inverse of A.

$A^{-1} = \begin{bmatrix} -5/2 & 3/4 \\ -2 & 1/2 \end{bmatrix}$

Solution: Use the 2×2 inverse formula:

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} -10 & 3 \\ -8 & 2 \end{bmatrix}$$

Since $\det(A) = 4$,

$$A^{-1} = \frac{1}{4} \begin{bmatrix} -10 & 3 \\ -8 & 2 \end{bmatrix} = \begin{bmatrix} -5/2 & 3/4 \\ -2 & 1/2 \end{bmatrix}$$

Problem 2: True or False?

Let A be an $n \times d$ matrix. For each statement in this problem, select True or False. For True statements, provide a short justification; for False statements, provide a counterexample (or a general argument for why the statement is false).

a) If A is square, and $\vec{x} \in \text{nullsp}(A)$ and $\vec{y} \in \text{colsp}(A)$, then $\vec{x} \cdot \vec{y} = 0$.
○ True ● False

Solution: This is false. In general, nullsp(A) is orthogonal to the **row space** of A, not necessarily the column space.
For a counterexample, take

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

Then

$$\vec{x} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \in \text{nullsp}(A) \qquad \vec{y} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \text{colsp}(A)$$

but

$$\vec{x} \cdot \vec{y} = -1$$

So the statement is false.

b) If A is square, and $\vec{x} \in \text{nullsp}(A)$ and $\vec{y} \in \text{colsp}(A^T)$, then $\vec{x} \cdot \vec{y} = 0$.
● True ○ False

Solution: If $\vec{y} \in \text{colsp}(A^T)$, then for some vector $\vec{b}$,

$$\vec{y} = A^T\vec{b}$$

Since $\vec{x} \in \text{nullsp}(A)$,

$$A\vec{x} = \vec{0}$$

So

$$\vec{x} \cdot \vec{y} = \vec{x}^T A^T \vec{b} = (A\vec{x})^T \vec{b} = \vec{0}^T \vec{b} = 0$$

So the statement is true.

c) If A is symmetric ($A = A^T$), and $\vec{x} \in \text{nullsp}(A)$ and $\vec{y} \in \text{colsp}(A)$, then $\vec{x} \cdot \vec{y} = 0$.
● True ○ False

Solution: If A is symmetric, then

$$\text{colsp}(A) = \text{colsp}(A^T)$$

So this reduces to part **b)**. Therefore $\vec{x} \cdot \vec{y} = 0$, and the statement is true.

d) If P projects vectors in $\mathbb{R}^n$ onto the column space of A, and $\vec{x}, \vec{y} \in \mathbb{R}^n$ are orthogonal, then $P\vec{x}$ and $P\vec{y}$ are also orthogonal.

○ True ● False

Solution: This is false. Take P to be projection onto the line spanned by

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and let

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

These vectors are orthogonal. But

$$P\vec{x} = P\vec{y} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

and these are certainly not orthogonal to each other. So the statement is false.

Problem 3: CR decomposition

Suppose $\vec{x}, \vec{y} \in \mathbb{R}^5$ and let

$$A = \begin{bmatrix} 3 & 0 & | & | & 3 \\ 1 & 1 & | & | & 2 \\ -1 & 1 & \vec{x} & \vec{y} & 0 \\ 0 & 1 & | & | & 1 \\ 2 & 0 & | & | & 2 \end{bmatrix}$$

Additionally, suppose A has the following CR decomposition:

$$A = \begin{bmatrix} 3 & 0 \\ 1 & 1 \\ -1 & 1 \\ 0 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & -1 & b_1 \\ 0 & 1 & 0 & 2 & b_2 \end{bmatrix}$$

a) Determine the following values.

$\text{rank}(A) =$ **2** , $b_1 =$ **1** , $b_2 =$ **1**

Solution: In a CR decomposition, the number of columns of C is the rank, so

$$\text{rank}(A) = 2$$

Also, the last column of R tells us how column 5 of A is written as a combination of the first two columns:

$$\text{col}_5(A) = b_1 \text{col}_1(A) + b_2 \text{col}_2(A)$$

Here

$$\begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

so

$$b_1 = 1 \qquad b_2 = 1$$

b) Find a basis for nullsp(A). Show your work, and circle your final answer, which should be a list of vectors. *Hint: One of those vectors can involve b_1 and b_2.*

Solution: From the CR decomposition, the non-pivot columns satisfy

$$\begin{aligned}
\text{col}_3(A) &= 3\text{col}_1(A) \\
\text{col}_4(A) &= -\text{col}_1(A) + 2\text{col}_2(A) \\
\text{col}_5(A) &= \text{col}_1(A) + \text{col}_2(A)
\end{aligned}$$

So the following vectors are in $\text{nullsp}(A)$:

$$\begin{bmatrix} 3 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} -1 \\ 2 \\ 0 \\ -1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$$

Since $\text{rank}(A) = 2$ and A has 5 columns,

$$\dim(\text{nullsp}(A)) = 5 - 2 = 3$$

So these three independent vectors form a basis. One basis is

$$\left\{ \begin{bmatrix} 3 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}$$

For the rest of the problem, let $\vec{z} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

c) Find **one** vector $\vec{w}'$ such that $A\vec{w}'$ is the orthogonal projection of $\vec{z}$ onto colsp(A). Show your work, and circle your final answer, which should be a vector with 5 entries.

Solution: Since colsp(A) = colsp(C), we can project $\vec{z}$ onto the column space of

$$C = \begin{bmatrix} 3 & 0 \\ 1 & 1 \\ -1 & 1 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}$$

Compute

$$C^T C = \begin{bmatrix} 15 & 0 \\ 0 & 3 \end{bmatrix} \qquad C^T \vec{z} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

So the coefficient vector for the projection is

$$\hat{\vec{w}} = (C^T C)^{-1} C^T \vec{z} = \begin{bmatrix} 1/15 & 0 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 4/15 \\ 1/3 \end{bmatrix}$$

One choice of $\vec{w}'$ that gives this same linear combination of columns of A is

$$\vec{w}' = \begin{bmatrix} 4/15 \\ 1/3 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

d) Describe the **complete** set of vectors $\vec{w}^*$ such that $A\vec{w}^*$ is the orthogonal projection of $\vec{z}$ onto colsp(A). Show your work, and circle your final answer, which should be a set of vectors described in {set notation}.

Solution: If $\vec{w}'$ is one solution, then every solution has the form

$$\vec{w}^* = \vec{w}' + \vec{n}$$

where $\vec{n} \in \text{nullsp}(A)$.
Using part **b)** and the particular solution from part **c)**, the complete set is

$$\left\{ \begin{bmatrix} 4/15 \\ 1/3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 3 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ 0 \\ -1 \\ 0 \end{bmatrix} + u \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} : s, t, u \in \mathbb{R} \right\}$$

e) Recall, $\vec{z} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

Find a matrix P such that $P\vec{v}$ is the orthogonal projection of $\vec{v}$ onto $\vec{z}$. Show your work, and circle your final answer, which should be a matrix with no variables.

Solution: The projection matrix onto the span of $\vec{z}$ is

$$P = \frac{\vec{z}\vec{z}^T}{\vec{z}^T\vec{z}}$$

Here

$$\vec{z}\vec{z}^T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \vec{z}^T\vec{z} = 2$$

So

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

f) Find rank(PA), where P is the matrix you found in part **e)** and A is the matrix from the start of the problem.

○ 0 ● 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ Impossible to tell

Solution: The matrix P projects onto a 1-dimensional subspace, so

$$\text{rank}(P) = 1$$

That means

$$\text{rank}(PA) \leq 1$$

Also PA is not the zero matrix. For example, the first column of A is

$$\begin{bmatrix} 3 \\ 1 \\ -1 \\ 0 \\ 2 \end{bmatrix}$$

and

$$P \begin{bmatrix} 3 \\ 1 \\ -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \neq \vec{0}$$

So rank(PA) is not 0. Therefore

$$\text{rank}(PA) = 1$$

Problem 4: The Null Space

Consider the plane in $\mathbb{R}^3$ given by the equation $x - 3y + 4z = 0$.

In parts **a)**, **b)**, and **c)**, let A be a matrix whose **null space** is the plane defined above.

a) Fill in the blanks:

$\text{rank}(A) =$ **1** $\qquad \dim(\text{nullsp}(A)) =$ **2**

Solution: The null space is a plane through the origin in $\mathbb{R}^3$, so it is a 2-dimensional subspace:

$$\dim(\text{nullsp}(A)) = 2$$

Since A has 3 columns, rank-nullity gives

$$\text{rank}(A) + \dim(\text{nullsp}(A)) = 3$$

So

$$\text{rank}(A) = 1$$

b) Suppose $\dim(\text{nullsp}(A^T)) = 3$. Fill in the blanks:

A has **4** rows and **3** columns.

Solution: The null space of A is a plane in $\mathbb{R}^3$, so A must have 3 columns.
Also $\text{rank}(A) = 1$, and $\text{rank}(A^T) = 1$. Apply rank-nullity to A^T:

$$\text{rank}(A^T) + \dim(\text{nullsp}(A^T)) = \text{number of rows of } A$$

So

$$1 + 3 = 4$$

Therefore A has 4 rows and 3 columns.

c) Find one possible matrix A where $\dim(\text{nullsp}(A^T)) = 3$.

Solution: We want a 4×3 matrix of rank 1 whose null space is

$$\left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x - 3y + 4z = 0 \right\}$$

One easy choice is to make every row a multiple of $\begin{bmatrix} 1 & -3 & 4 \end{bmatrix}$. For example,

$$A = \begin{bmatrix} 1 & -3 & 4 \\ 1 & -3 & 4 \\ 1 & -3 & 4 \\ 1 & -3 & 4 \end{bmatrix}$$

Then $A\vec{v} = \vec{0}$ exactly when

$$\begin{bmatrix} 1 & -3 & 4 \end{bmatrix} \vec{v} = 0$$

which is the given plane.

d) Now, suppose B is a matrix whose **column space** is the plane defined above. Find one possible matrix B with 4 columns.

Solution: Pick two linearly independent vectors in the plane

$$x - 3y + 4z = 0$$

For example,

$$\begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 4 \\ 3 \end{bmatrix}$$

both lie in the plane, and they are linearly independent. So one possible choice is

$$B = \begin{bmatrix} 4 & 4 & 0 & 0 \\ 0 & 0 & 4 & 4 \\ -1 & -1 & 3 & 3 \end{bmatrix}$$

Its column space is exactly that plane.

Problem 5: Skincare Products

In this problem, we'll work with a dataset with information about skincare products. A sample of the dataset is shown below, but the full dataset has a total of n rows.

Price	Type	Brand	Sensitive	# Ingredients
55	Eye cream	PERRICONE MD	1	33
19	Cleanser	CLINIQUE	0	36
75	Eye cream	PETER THOMAS ROTH	1	42
38	Cleanser	PETER THOMAS ROTH	0	23
⋮	⋮	⋮	⋮	⋮

The `Sensitive` column contains either 1 or 0, corresponding to whether the product was designed for sensitive skin.

Our goal is to fit a multiple linear regression model (by minimizing mean squared error) that predicts **the number of ingredients** in a product given its `Price` and various other information.

a) Suppose we fit a model that uses an intercept term, `Price`, and `Sensitive` as features.

(i) Write the first two rows of the design matrix, X.

$$\begin{bmatrix} 1 & 55 & 1 \\ 1 & 19 & 0 \end{bmatrix}$$

(ii) Suppose that a solution to the normal equations is

$$\vec{w}^* = \begin{bmatrix} w_0^* \\ w_{\text{Price}}^* \\ w_{\text{Sensitive}}^* \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 8 \end{bmatrix}$$

This model is equivalent to **two** parallel lines in $\mathbb{R}^2$, each one of the form

$$h(\text{Price}_i) = a + b \cdot \text{Price}_i$$

where the line to use is determined by the value of `Sensitive`. Determine the values of a and b in both cases.

Solution: For part **(i)**, the features are intercept, price, and sensitive, so the first two rows are

$$\begin{bmatrix} 1 & 55 & 1 \\ 1 & 19 & 0 \end{bmatrix}$$

For part **(ii)**, the fitted model is

$$h(\text{Price}_i, \text{Sensitive}_i) = 5 - 2 \cdot \text{Price}_i + 8 \cdot \text{Sensitive}_i$$

If `Sensitive` $= 0$, then

$$h(\text{Price}_i) = 5 - 2 \cdot \text{Price}_i$$

so

$$a = 5 \qquad b = -2$$

If `Sensitive` $= 1$, then

$$h(\text{Price}_i) = 13 - 2 \cdot \text{Price}_i$$

so

$$a = 13 \qquad b = -2$$

The table below describes four possible models, each fit by minimizing mean squared error.

Model	Intercept?	Price?	Type?	Brand?	Sensitive?
Model 1	Yes	Yes	OHE, no categories dropped	No	No
Model 2	Yes	Yes	OHE, `"Eye cream"` dropped	No	No
Model 3	No	Yes	OHE, `"Eye cream"` dropped	OHE, no categories dropped	No
Model 4	No	Yes	OHE, `"Eye cream"` dropped	OHE, `"CLINIQUE"` dropped	No

In each part below, determine

- The number of columns in the design matrix X.
- The rank of the design matrix X.
- Whether or not the model's error vector $\vec{e}$ is guaranteed to sum to 0.

Assume that there are **8 unique values of `Type`** and **15 unique values of `Brand`**. **If it is impossible to determine the answer (e.g. if there are multiple possible answers), write "N/A"; don't just write one of them.**

b) Model 1

number of columns in X = **10** , rank(X) = **9**

Errors guaranteed to sum to 0? ● Yes ○ No

Solution: Model 1 uses

$$1 \text{ intercept} + 1 \text{ price} + 8 \text{ type columns} = 10 \text{ columns}$$

Since all 8 type columns are included, their sum equals the intercept column, so there is one guaranteed linear dependence. That gives rank 9.
The intercept column is present, so the residual vector is orthogonal to the all-ones vector. Therefore the errors are guaranteed to sum to 0.

c) Model 2

number of columns in X = **9** , rank(X) = **9**

Errors guaranteed to sum to 0? ● Yes ○ No

Solution: Model 2 uses

$$1 \text{ intercept} + 1 \text{ price} + 7 \text{ type columns} = 9 \text{ columns}$$

Dropping one type removes the built-in redundancy from Model 1, so the rank is 9. Again, the intercept column is present, so the errors are guaranteed to sum to 0.

d) Model 3

number of columns in $X =$ **23**, $\text{rank}(X) =$ **N/A**

Errors guaranteed to sum to 0? ● Yes ○ No

Solution: Model 3 uses

$$1 \text{ price} + 7 \text{ type columns} + 15 \text{ brand columns} = 23 \text{ columns}$$

The rank is `N/A`, because it depends on relationships in the dataset between the type columns and the brand columns.
Even though there is no explicit intercept, all 15 brand columns are included, and their sum is the all-ones vector. So the column space contains an intercept column implicitly, which means the errors are guaranteed to sum to 0.

e) Model 4

number of columns in $X =$ **22**, $\text{rank}(X) =$ **N/A**

Errors guaranteed to sum to 0? ○ Yes ● No

Solution: Model 4 uses

$$1 \text{ price} + 7 \text{ type columns} + 14 \text{ brand columns} = 22 \text{ columns}$$

The rank is again `N/A`, because it depends on dataset-specific relationships between the type and brand columns.
Here there is no explicit intercept, and unlike Model 3 there is no guaranteed implicit intercept either, because one brand column is dropped. So the errors are not guaranteed to sum to 0.

Problem 6: Gradients

Let

$$h(\vec{x}) = (C\vec{x}) \cdot (C\vec{x} + \vec{b})$$

where $\vec{x} \in \mathbb{R}^d$, C is an $n \times d$ matrix, and $\vec{b} \in \mathbb{R}^n$.

a) Find $\nabla h(\vec{x})$, the gradient of $h(\vec{x})$. circle your final answer, which should be an expression in terms of $\vec{x}$, C, $\vec{b}$, and/or constants.

Solution: Rewrite h as

$$\begin{aligned} h(\vec{x}) &= (C\vec{x})^T (C\vec{x} + \vec{b}) \\ &= \vec{x}^T C^T C\vec{x} + \vec{x}^T C^T \vec{b} \end{aligned}$$

Differentiate term-by-term:

$$\nabla h(\vec{x}) = 2C^T C\vec{x} + C^T \vec{b}$$

Equivalently,

$$\nabla h(\vec{x}) = C^T (2C\vec{x} + \vec{b})$$

b) Let

$$g(\vec{x}) = \log\left((h(\vec{x}))^2\right)$$

where log is the natural logarithm. Find $\nabla g(\vec{x})$, the gradient of $g(\vec{x})$. circle your final answer, which should be an expression in terms of $\vec{x}$, C, $\vec{b}$, and/or constants.

Solution: As long as $h(\vec{x}) \neq 0$, the chain rule gives

$$\nabla g(\vec{x}) = \frac{2}{h(\vec{x})} \nabla h(\vec{x})$$

Using part **a)**,

$$\nabla g(\vec{x}) = \frac{2(2C^T C\vec{x} + C^T \vec{b})}{h(\vec{x})}$$

Since

$$h(\vec{x}) = \vec{x}^T C^T C\vec{x} + \vec{x}^T C^T \vec{b}$$

we can also write

$$\nabla g(\vec{x}) = \frac{2(2C^T C\vec{x} + C^T \vec{b})}{\vec{x}^T C^T C\vec{x} + \vec{x}^T C^T \vec{b}}$$

Problem 7: Gradient Descent

Suppose $\vec{u} \in \mathbb{R}^3$, and let

$$q(\vec{u}) = (u_1 + u_2 + u_3)^2 + (u_1 - u_2)^2 + (u_2 - u_3)^2$$

We write code that implements gradient descent in order to minimize q, using some initial guess, $\vec{u}^{(0)}$, and learning rate/step size, π. In our code, we add print statements that show us the values of $\vec{u}^{(t)}$ and $\nabla q(\vec{u}^{(t)})$ (the gradient of q) after each iteration.

Here's what we see:

After 1 iteration, $\vec{u}^{(1)} = \begin{bmatrix} 0 \\ -1 \\ -3 \end{bmatrix}$, $\nabla q(\vec{u}^{(1)}) = \begin{bmatrix} -6 \\ -6 \\ -12 \end{bmatrix}$, $q(\vec{u}^{(1)}) = 21$

After 2 iterations, $\vec{u}^{(2)} = \begin{bmatrix} 1.2 \\ 0.2 \\ -0.6 \end{bmatrix}$, $\nabla q(\vec{u}^{(2)}) = \begin{bmatrix} 3.6 \\ 1.2 \\ 0 \end{bmatrix}$, $q(\vec{u}^{(2)}) = 2.28$

a) What is value of π? Give your answer as a fraction or decimal.

$\pi =$ **1/5**

Solution: Gradient descent updates via

$$\vec{u}^{(2)} = \vec{u}^{(1)} - \pi \nabla q(\vec{u}^{(1)})$$

Use the second component:

$$\begin{aligned} 0.2 &= -1 - \pi(-6) \\ 0.2 &= -1 + 6\pi \\ 1.2 &= 6\pi \\ \pi &= \frac{1}{5} \end{aligned}$$

b) What is the value of $\vec{u}_2^{(0)}$, i.e. what is the second component of the **initial guess vector**, $\vec{u}^{(0)}$? Circle your final answer, which should be an expression in terms of π and/or constants. (If your answer does not involve π, we cannot give you partial credit in case your answer to (a) was incorrect.)

Solution: We only need the second component of the gradient. Differentiate:

$$\begin{aligned}\frac{\partial q}{\partial u_2} &= 2(u_1 + u_2 + u_3) - 2(u_1 - u_2) + 2(u_2 - u_3) \\ &= 6u_2\end{aligned}$$

So the second component of the gradient descent update is

$$u_2^{(1)} = u_2^{(0)} - \pi(6u_2^{(0)}) = u_2^{(0)}(1 - 6\pi)$$

We are told that $u_2^{(1)} = -1$, so

$$-1 = u_2^{(0)}(1 - 6\pi)$$

Hence

$$u_2^{(0)} = \frac{-1}{1 - 6\pi}$$

If you substitute $\pi = 1/5$, this becomes 5.

uniqname: ____________________

Problem 8: Convexity

Let $f : \mathbb{R} \to \mathbb{R}$ be convex, and suppose $f(0) = 0$.

Prove that for all **non-negative** values of x and y,

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

Hints:

- First, handle the easy case, $x + y = 0$.
- If $x + y > 0$, define $t = \frac{x}{x+y}$. Why must $t \in [0, 1]$?
- Rewrite x as $x = t(x + y)$ using the definition of t above. Then use the fact from Lab 10 that for a convex function with $f(0) = 0$,

$$f(tu) \leq t\, f(u) \qquad \text{for } t \in [0, 1]$$

- Do the same for y by writing $y = (1 - t)(x + y)$.
- Add the two inequalities you get.

Solution: If $x + y = 0$, then $x = y = 0$ because both are non-negative. So

$$f(x + y) = f(0) = 0 = f(x) + f(y)$$

Now assume $x + y > 0$, and define

$$t = \frac{x}{x + y}$$

Because $x, y \geq 0$ and $x + y > 0$, we have $t \in [0, 1]$.
Also,

$$x = t(x + y) \qquad y = (1 - t)(x + y)$$

Since f is convex and $f(0) = 0$, we know that for any u and any $s \in [0, 1]$,

$$f(su) \leq sf(u)$$

Apply this with $u = x + y$:

$$f(x) = f(t(x + y)) \leq tf(x + y)$$

and similarly,

$$f(y) = f((1 - t)(x + y)) \leq (1 - t)f(x + y)$$

Add the two inequalities:

$$\begin{aligned} f(x) + f(y) &\leq tf(x + y) + (1 - t)f(x + y) \\ &= f(x + y) \end{aligned}$$

So for all non-negative x and y,

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

uniqname: ______________________

Congrats on finishing Mock Midterm 2!

Feel free to draw us a picture about EECS 245 in the box below.