CS/ECE/ME 532
Homework 2: Norms and Least Squares

1. Define the mapping $\Phi : \mathbb{R}^n \to \mathbb{R}$ as

$$\Phi(\boldsymbol{x}) \;=\; \max_{1 \le i \le n} i|x_i|.$$

Is $\Phi(\boldsymbol{x})$ a norm? **SOLUTION:** $\Phi(x)$ is a norm. To prove this, we must verify the three axioms.

a) $\Phi(x) = 0$ if and only if $x = 0$.
True by inspection.

b) $\Phi(\alpha x) = |\alpha|\Phi(x)$.

$$\Phi(\alpha \boldsymbol{x}) \;=\; \max_{1 \le i \le n} i|\alpha x_i| = |\alpha| \max_{1 \le i \le n} i|x_i| = |\alpha|\Phi(\boldsymbol{x}).$$

c) $\Phi(\boldsymbol{x}+\boldsymbol{y}) \le \Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y})$, the triangle inequality.

$$\Phi(x+y) = \max_i i|x_i + y_i| \le \max_i i(|x_i| + |y_i|) \le \max_i i|x_i| + \max_i i|y_i|$$

2. Equivalence of norms. For each case below find positive constants a and b (possibly different in each case) so that for every $\boldsymbol{x} \in \mathbb{R}^n$

(i) $a\|\boldsymbol{x}\|_1 \;\le\; \|\boldsymbol{x}\|_2 \;\le\; b\|\boldsymbol{x}\|_1$ (HINT: Use Cauchy-Schwarz inequality, which states that for any vectors u, v, we have: $u^T v \le \|u\|_2\|v\|_2$ and a vector of all ones.) **SOLUTION:**

The lower bound follows from the Cauchy-Schwarz inequality, which states that for any vectors u, v, we have: $u^T v \le \|u\|_2\|v\|_2$. Apply it here by letting

$$u = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T \qquad \text{and} \qquad v = \begin{bmatrix} |x_1| & |x_2| & \dots & |x_n| \end{bmatrix}^T$$

Then we have: $\|x\|_1 = u^T v$, and $\|u\|_2 = \sqrt{n}$, and $\|v\|_2 = \|x\|_2$. By Cauchy-Schwarz, we then have $\|x\|_1 \le \sqrt{n}\|x\|_2$. Equality occurs when all components of x are equal. The upper bound follows from the triangle inequality. Write:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{bmatrix}$$

Then the triangle inequality gives: $\|x\|_2 \le |x_1| + \cdots + |x_n| = \|x\|_1$. Equality occurs when x has a single nonzero component. Assembling the upper and lower bounds, we obtain:

$$\frac{1}{\sqrt{n}}\|x\|_1 \le \|x\|_2 \le \|x\|_1$$

(ii) $a\|\boldsymbol{x}\|_1 \;\leq\; \|\boldsymbol{x}\|_\infty \;\leq\; b\|\boldsymbol{x}\|_1$

SOLUTION: Recall that $\|x\|_\infty = \max_{1\leq i\leq n}|x_i|$. It is the largest component of x in absolute value. The upper bound is simply $\|x\|_\infty \leq \|x\|_1$. Equality occurs when x has a single nonzero component. The lower bound is found by reasoning that given a list of nonnegative numbers, the largest number is always greater than (or equal to) the average of the numbers. Equality occurs when all components are equal. Assembling both bounds, we obtain:

$$\frac{1}{n}\|x\|_1 \leq \|x\|_\infty \leq \|x\|_1$$

3. Let

$$\boldsymbol{A} \;=\; \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

a) What is the rank of $\boldsymbol{A}$?

SOLUTION: A has rank 3 (full rank) because of the triangular structure.

b) Suppose that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Derive an explicit formula for $\boldsymbol{x}$ in terms of $\boldsymbol{y}$.

SOLUTION: Writing out the equation $y = Ax$, we have:

$$y = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} x \qquad \Longleftrightarrow \qquad \begin{aligned} y_1 &= x_1 + x_2 + x_3 \\ y_2 &= x_1 + x_2 \\ y_3 &= x_1 \end{aligned}$$

From the last equation, $x_1 = y_3$. Substituting this into the second equation, $x_2 = y_2 - y_3$. Substituting all of this into the first equation, $x_3 = y_1 - y_2$. So we conclude that:

$$\begin{aligned} x_1 &= y_3 \\ x_2 &= y_2 - y_3 \\ x_3 &= y_1 - y_2 \end{aligned} \qquad \Longleftrightarrow \qquad x = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix} y$$

4. Answer the following questions. Make sure to explain your reasoning.

a) Are the columns of the following matrix linearly independent?

$$\boldsymbol{A} \;=\; \begin{bmatrix} +0.92 & +0.92 \\ -0.92 & +0.92 \\ +0.92 & -0.92 \\ -0.92 & -0.92 \end{bmatrix}$$

SOLUTION: Yes. Since the first elements match, the only way one column can be a multiple of the other is if all elements match, which is not the case.

b) Are the columns of the following matrix linearly independent?

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

SOLUTION: Yes. This is clear because no column can be written as a weighted sum of the other columns. If you wish to see a more formulaic approach, this can be shown by bringing the matrix to reduced row echelon form:

$$\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \xrightarrow[r_2+r_3 \mapsto r_3]{} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 0 & 0 & -2 \end{bmatrix} \xrightarrow[r_1+r_2 \mapsto r_2]{} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \xrightarrow[r_1 - \frac{1}{2}r_2 + \frac{1}{2}r_3 \mapsto r_1]{} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

The final matrix is diagonal with nonzero entries so the matrix is full-rank (and therefore invertible in this case, since it's also square).

c) Are the columns of the following matrix linearly independent?

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 4 & 5 \\ 5 & 6 & 8 \end{bmatrix}$$

SOLUTION: No, the first column can be expressed as 1*(col 1) + (col 2)/2.

d) What is the rank of the following matrix?

$$A = \begin{bmatrix} +5 & +2 \\ -5 & +2 \\ +5 & -2 \end{bmatrix}$$

SOLUTION: The rank is at most 2 because there are 2 columns. The columns are not multiples of one another so they are independent and therefore rank$(A) = 2$.

e) Suppose the matrix in part c is used in the least squares optimization $\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$. Does a unique solution exist?

SOLUTION: The normal equations are $A^T Ax = A^T b$. And any solution to the normal equations is a solution (minimizer) of the least-squares problem. Because A does not have full column rank, $A^T A$ is not invertible, and so the normal equations do not have a unique solution. Consequently the least-squares problem does not have a unique solution.

5. Consider the following matrix and vector:

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

a) Find the solution $\widehat{\boldsymbol{x}}$ to $\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$.

SOLUTION: The normal equations are: $A^T Ax = A^T b$. Substituting A and b, they become:

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Solving yields $x_1 = 1/3$ and $x_2 = -1/3$.

b) Make a sketch of the geometry of this particular problem in $\mathbb{R}^3$, showing the columns of $\boldsymbol{A}$, the plane they span, the target vector $\boldsymbol{b}$, the residual vector and the solution $\widehat{\boldsymbol{b}} = \boldsymbol{A}\widehat{\boldsymbol{x}}$.

SOLUTION: analogous to class notes

6. Polynomial fitting. Suppose we observe pairs of points (a_i, b_i), $i = 1, \dots, m$. Imagine these points are measurements from a scientific experiment. The variables a_i are the experimental conditions and the b_i correspond to the measured response in each condition. Suppose we wish to fit a degree $d < m$ polynomial to these data. In other words, we want to find the coefficients of a degree d polynomial p so that $p(a_i) \approx b_i$ for $i = 1, 2, \dots, m$. We will set this up as a least-squares problem.

a) Suppose p is a degree d polynomial. Write the general expression for $p(a_i) = b$.

SOLUTION: A polynomial of degree d has the form: $p(z) = x_0 + x_1 z + x_2 z^2 + \cdots + x_d z^d$. Where $x_0, \dots, x_d$ are the coefficients. Note that there are $d+1$ coefficients. Therefore, the equation $p(a_i) = b_i$ is:

$$x_0 + x_1 a_i + x_2 a_i^2 + \cdots + x_d a_i^d = b_i$$

b) Express the $i = 1, \dots, m$ equations as a system in matrix form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Specifically, what is the form/structure of $\boldsymbol{A}$ in terms of the given a_i.

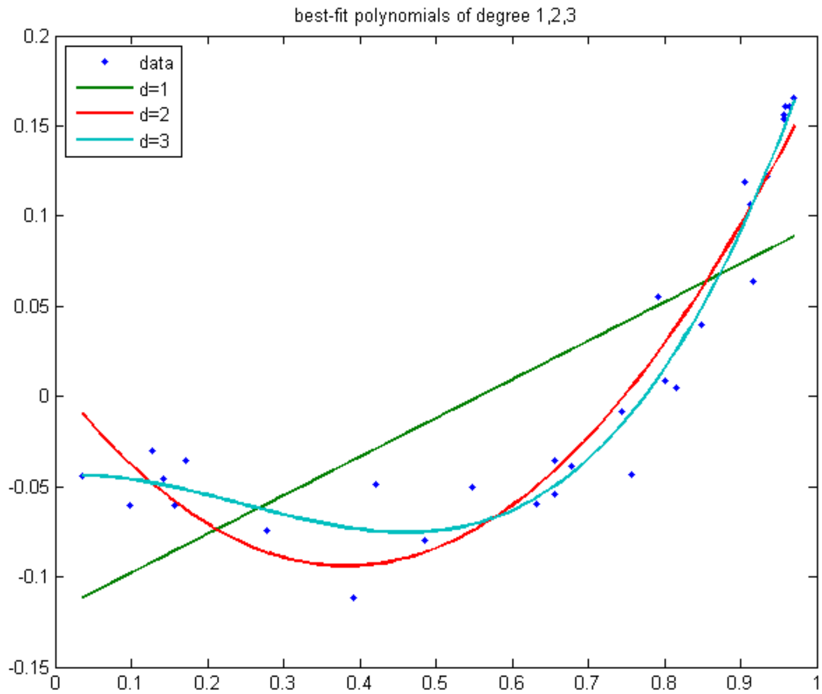
SOLUTION: Writing out the equations from the previous part for $i = 1, \dots, m$ and stacking them into a matrix, we obtain the equation:

$$\begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^p \\ 1 & a_2 & a_2^2 & \dots & a_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_m & a_m^2 & \dots & a_m^p \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Interesting bit of trivia: matrices of this form are known as *Vandermonde* matrices.

c) Write a Matlab or Python script to find the least-squares fit to the $m = 30$ data points in
`polydata.mat`. Plot the points and the polynomial fits for $d = 1, 2, 3$.

SOLUTION: Here is the output of the code:



And here is the Matlab code that produced the plot.

```
% load a and b vectors
load polydata.mat

m = numel(a);                      % number of data points
N = 100;                           % num points to use for interpolation
z = linspace(min(a),max(a),N);  % pts where interpolant is evaluated
y = zeros(3,N);                    % where we'll store polynomial values

for d = 1:3

   % generate A-matrix for this choice of d
   A = zeros(m,d+1);
   for i = 1:m
      for j = 1:d+1
         A(i,j) = a(i)^(j-1);
      end
   end

   % solve least-squares problem. x is the list of coefficients.
   % NOTE: a shortcut in matlab is to just type: x = A\b;
   x = inv(A'*A)*(A'*b);
```

```
    % evaluate best-fit polynomial at all points z. store result in y.
    % NOTE: you can do this in one line with the polyval command!
    for i = 1:N
        for j = 1:d+1
            y(d,i) = y(d,i) + x(j)*z(i)^(j-1);
        end
    end

end

% plot the data and the best-fit polynomials
figure(1)
plot(a,b,'.', z,y(1,:), z,y(2,:), z,y(3,:),'LineWidth',2)
legend('data','d=1','d=2','d=3','Location','NorthWest')
title('best-fit␣polynomials␣of␣degree␣1,2,3')
```

7. Recall the cereal calorie prediction problem discussed in class. The data matrix for this problem is

$$\boldsymbol{A} = \begin{bmatrix} 25 & 0 & 1 \\ 20 & 1 & 2 \\ 40 & 1 & 6 \end{bmatrix}.$$

Each row contains the grams/serving of carbohydrates, fat, and protein, and each row corresponds to a different cereal (*Frosted Flakes, Froot Loops, Grape-Nuts*). The total calories for each cereal are

$$\boldsymbol{b} = \begin{bmatrix} 110 \\ 110 \\ 210 \end{bmatrix}.$$

a) Write a small program (e.g., in Matlab or Python) that solves the system of equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Recall the solution $\boldsymbol{x}$ gives the calories/gram of carbohydrate, fat, or protein. What is the solution?

SOLUTION: The solution is: $x = \begin{bmatrix} 4.2500 \\ 17.5000 \\ 3.7500 \end{bmatrix}$. Matlab code that produces it is:

```
A = [ 25 0 1
      20 1 2
      40 1 6 ];
b = [ 110 110 210 ]';
x = inv(A)*b
```

b) The solution may not agree with the known calories/gram, which are 4 for carbs, 9 for fat and 4 for protein. We suspect this may be due to rounding the grams to integers, especially the fat grams. Assuming the true value for calories/gram is

$$\boldsymbol{x}^\star = \begin{bmatrix} 4 \\ 9 \\ 4 \end{bmatrix},$$

and that the total calories, grams of carbs, and grams of protein are correctly reported above, determine the "correct" grams of fat in each cereal.

SOLUTION: Let the unknown grams of fat be f_1, f_2, f_3. The new equations are:

$$\begin{bmatrix} 25 & f_1 & 1 \\ 20 & f_2 & 2 \\ 40 & f_3 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 9 \\ 4 \end{bmatrix} = \begin{bmatrix} 110 \\ 110 \\ 210 \end{bmatrix}$$

Rearranging and solving for the f's, we obtain:

$$f_1 = \frac{6}{9} \approx 0.667 \qquad f_2 = \frac{22}{9} \approx 2.444 \qquad f_3 = \frac{26}{9} \approx 2.889$$

c) Now suppose that we predict total calories using a more refined breakdown of carbohydrates, into total carbohydrates, complex carbohydrates and sugars (simple carbs). So now we will have 5 features to predict calories (the three carb features + fat and protein). So let's suppose we measure the grams of these features in 5 different cereals to obtain this data matrix

$$\boldsymbol{A} = \begin{bmatrix} 25 & 15 & 10 & 0 & 1 \\ 20 & 12 & 8 & 1 & 2 \\ 40 & 30 & 10 & 1 & 6 \\ 30 & 15 & 15 & 0 & 3 \\ 35 & 20 & 15 & 2 & 4 \end{bmatrix},$$

and the total calories in each cereal

$$\boldsymbol{b} = \begin{bmatrix} 104 \\ 97 \\ 193 \\ 132 \\ 174 \end{bmatrix}.$$

Can you solve $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$? Carefully examine the situation in this case. Is there a solution that agrees with the true calories/gram?

SOLUTION: Since the total carbs are equal to the complex carbs plus the sugars, the first column of the A matrix is equal to the sum of the second and third columns. So the

A matrix is *not* full rank. Therefore, we can't simply invert it and solve for x. To see whether the system of equations still has a solution, we can use Matlab to check that:

$$\text{rank}(A) = 4 \qquad \text{and} \qquad \text{rank}(\begin{bmatrix} A & b \end{bmatrix}) = 4$$

since adding the column b to the matrix A does not increase its rank, b must lie in the span of the columns of A (i.e. there is a solution to $Ax = b$). But how do we find it? One way to do this is to remove the first column of A, which is redundant. We then have:

$$\begin{bmatrix} 15 & 10 & 0 & 1 \\ 12 & 8 & 1 & 2 \\ 30 & 10 & 1 & 6 \\ 15 & 15 & 0 & 3 \\ 20 & 15 & 2 & 4 \end{bmatrix} z = \begin{bmatrix} 104 \\ 97 \\ 193 \\ 132 \\ 174 \end{bmatrix}$$

Solving the associated least-squares problem (which is now full-rank), we obtain the solution:

$$z = \begin{bmatrix} 4 \\ 4 \\ 9 \\ 4 \end{bmatrix}$$

But this isn't just the least-squares solution... it's an exact solution to the problem! (the residual is zero). Moreover, it agrees with the true calories per gram! Due to the rank deficiency of A, the solution to the original problem is not unique. The most general solution to the original problem is:

$$x = \begin{bmatrix} \alpha \\ 4 - \alpha \\ 4 - \alpha \\ 9 \\ 4 \end{bmatrix} \qquad \text{for any } \alpha \in \mathbb{R}$$

and the solution that agrees with the true carbs/gram is when $\alpha = 0$. Note that this problem is very special. Usually in such situations (e.g. if b is chosen randomly but we keep the same A matrix), there won't exist a solution to $Ax = b$ and we would have to resort to solving the least squares problem (i.e. finding an x that is "close" to being a solution).