Fall 2022 *Instructor: Ashwin Pananjady*

7750: Mathematical Foundations of Machine Learning

Linear algebra and probability for data analysis

Homework 1

Released: Aug 24 *Due: Sep 6, 11:59pm ET*

Note. All external sources and collaborators must be acknowledged in your submission. As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.

Objective. To solidify your algebraic and geometric understanding of linear representations with basis functions. To build intuition for normed vector spaces by reviewing some finite-dimensional linear algebra.

Resources. Lectures, notes, and modules posted before Aug 31.

Notation: Capital boldface letters will be matrices, and small boldface letter will be vectors. Capital letters (not boldface) will be random variables (and sometimes random vectors), and small letters will typically be scalars. Dimensions of matrices and vectors will be specified when needed, but for the most part, you should be able to intuit these yourself (and doing this is a useful exercise).

Problem 1 (Visualizing quadratic functions and linear algebra review). 20 points: For $\mathbf{x} \in \mathbb{R}^d$, consider the function $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$ for a square and symmetric matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$, vector $\mathbf{b} \in \mathbb{R}^d$ and scalar c.

(a) Write the gradient of f. Suppose the matrix $\mathbf{A}$ is invertible. Is there a unique solution to $\nabla f(\mathbf{x}) = \mathbf{0}$, and if so, what is it?

Solution:
The gradient is expressed as:

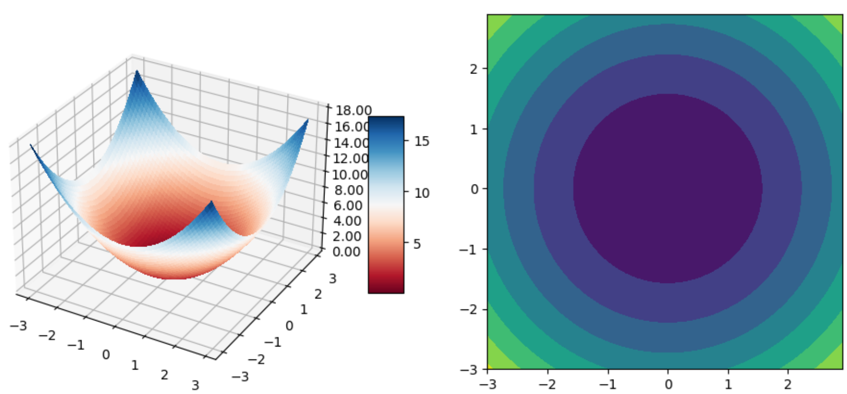
$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + \mathbf{b}$$

A unique solution exists due to A being invertible. It is found by setting $\nabla f(\mathbf{x}) = 0$,

$$\mathbf{x} = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$$

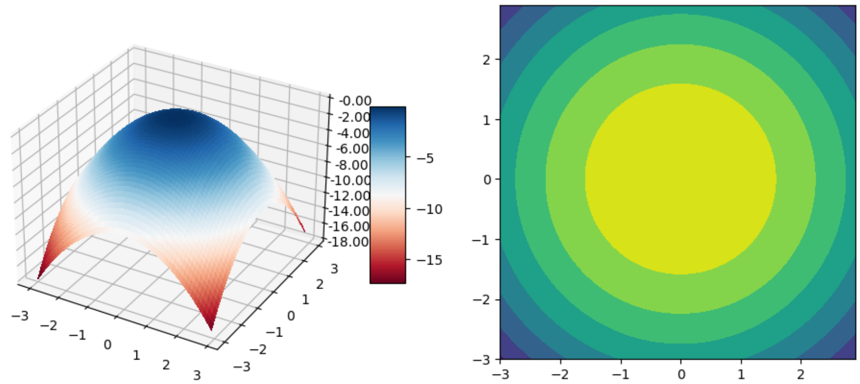
(b) You have been given an iPython notebook with starter code to generate 3D plots and contour plots of functions. Fill in the requisite lines of code and provide both 3D and contour plots of f for the following settings of $(\mathbf{A}, \mathbf{b}, c)$:

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



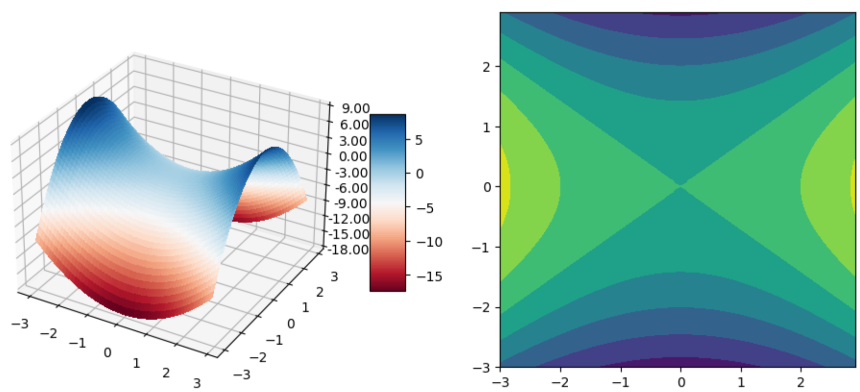
The contours are circular and this is because the eigenvalues of $\mathbf{A}$ are same. Second, the region near the origin is dark and the region away from the origin point is light. Correspondingly, in the 3D plot, the function value decreases when it goes close to the center and increases when we move away.

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



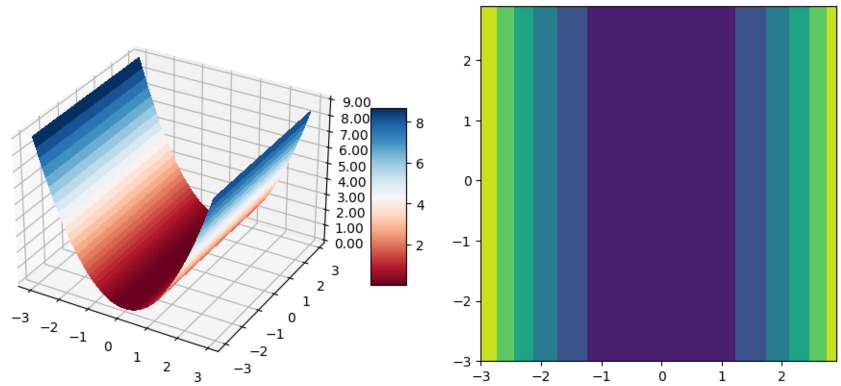
The contours are circular and this is because the eigenvalues of $\mathbf{A}$ are same (but negative this time). Second, the region near the origin is light and the region away from the origin point is dark. Correspondingly, in the 3D plot, the function value increases when it goes close to the center and decreases when we move away.

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



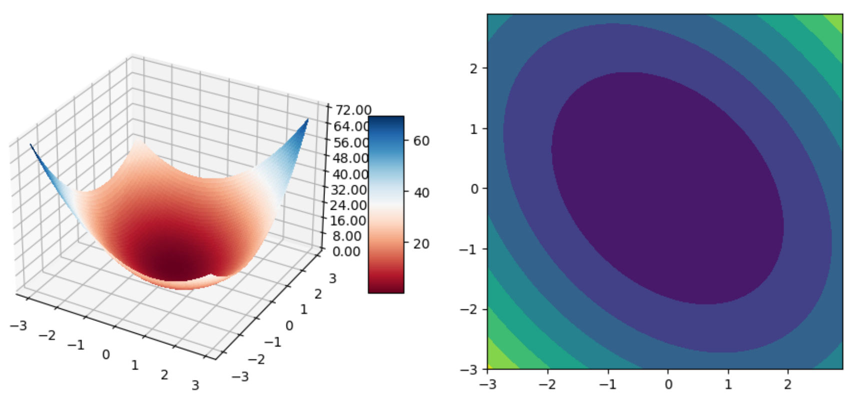
The contours are hyperbola and this is because one eigenvalue of $\mathbf{A}$ is positive and the other is negative. Along the x_1 axis, it becomes darker when the point is closer to the origin; while along the x_2 axis, it becomes lighter when the point is closer to the origin. Correspondingly, in the 3D plot, the function value decreases when it goes close to the origin along the x_1 axis and increases when it goes close to the original point along the x_2 axis.

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



> The contours are parallel lines and this is because one eigenvalue of **A** is positive and the other is zero. **A** is independent of x_2 axis. Moving along x_1 axis, the value of the function decreases, but away from the origin, it increases. Since **A** is independent of x_2, the value does not change along x_2 axis.

- $$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad c = 0.$$



> The contours are elliptical and not circular, and this is because the eigenvalues of **A** are not the same. We can also see the axes of the ellipse do not coincide with the x_1 and x_2 axes. Second, the region near the center is dark and the region away from it is light. Correspondingly, in the 3D plot, the function value decreases when it goes close to the origin and increases when it goes away from the origin.

(c) Show that we have the linear representation $f(\mathbf{x}) = \sum_{j=1}^{M} \alpha_j p_j(\mathbf{x})$ where $p_1, \ldots, p_M$ are all the monomials in $(x_1, \ldots, x_d)$ of degree at most 2. In particular, for each such monomial write down the coefficient that multiplies it.

Takeaway: This and the previous part should convince you that there are quite a few "shapes" taken by functions that can be represented as quadratic polynomials in two variables.

Solution:
Given $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$
Using index notation for matrix-vector multiplication, decompose the above operation into the following,

$$f(\mathbf{x}) = \sum_{i=1}^{d} \sum_{j=1}^{d} \underbrace{a_{i,j}}_{\text{quadratic coefficients}} x_i x_j + \sum_{i=1}^{d} \underbrace{b_i}_{\text{linear coefficients}} x_i + c.$$

In particular, for each i, j pair, the monomial $x_i x_j$ has coefficient $a_{i,j}$, the monomial x_i has coefficient b_i and the monomial 1 has coefficient c.

(d) For the fifth example (non-diagonal $\mathbf{A}$), use a package to compute its eigendecomposition. Compare the orthogonal matrix returned by this eigendecomposition to what you see on the corresponding contour plot. Is $\mathbf{A}$ positive semidefinite (PSD)?

Solution:
The eigenvalues of $\mathbf{A}$ are $\lambda_1 = 4$ and $\lambda_2 = 2$, and so $\mathbf{A}$ is PSD. The corresponding eigenvectors are $\mathbf{v}_1 = [0.7071, 0.7071]^\top$ and $\mathbf{v}_2 = [-0.7071, 0.7071]^\top$. These two eigenvectors form the principal axes of the ellipse in the corresponding contour plot.

To explain in more detail, the eigendecomposition of $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top$, where $\mathbf{V}$ is a *rotation matrix* satisfying $\mathbf{V}^\top\mathbf{V} = \mathbf{V}\mathbf{V}^\top = \mathbf{I}$. We may now perform the change of variables $\mathbf{y} = \mathbf{V}^\top\mathbf{x}$ to write $f(\mathbf{x}) = \mathbf{y}^\top\mathbf{\Lambda}\mathbf{y} + \mathbf{b}^\top\mathbf{V}^\top\mathbf{y} + c := g(\mathbf{y})$. In the coordinate system of $\mathbf{y}$, the function g is now an axis-aligned quadratic. To form the function f, we are performing a change a variables: the matrix $\mathbf{V}$ allows us to go back and forth between $\mathbf{x}$ and $\mathbf{y}$.

(e) Now return to the original problem in dimension d with general $(\mathbf{A}, \mathbf{b}, c)$. If $\mathbf{A}$ is PSD and invertible, i.e., positive definite[1], show that $\widehat{\mathbf{x}} = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$ is the unique global minimum of f.

(Hint: There are several ways to do this, but one way is to write the function as $f(\mathbf{x}) = (\mathbf{x} - \widehat{\mathbf{x}})^\top\mathbf{A}(\mathbf{x} - \widehat{\mathbf{x}}) +$ other terms and argue from there.)

[1] A positive definite matrix satisfies $\mathbf{v}^\top\mathbf{A}\mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{R}^d$, with $\mathbf{v}^\top\mathbf{A}\mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$

Solution: Let us showcase two solutions to this. For the first solution, compute the Hessian of $f(\mathbf{x})$ similar to how the gradient (or Jacobian) was computed in part (a). The Hessian of $f(\mathbf{x})$ is given by,

$$\mathbf{H}_f = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2} = \mathbf{A}.$$

The Hessian is positive definite which means f is strictly convex. Hence, the solution in part (a) is the unique global minimum of $f(\mathbf{x})$.
To see a second solution, write $\hat{\mathbf{x}} = -\mathbf{A}^{-1}\mathbf{b}/2$. Then adding and subtracting terms, we have

$$f(\mathbf{x}) = (\mathbf{x} - \hat{\mathbf{x}})^\top \mathbf{A}(\mathbf{x} - \hat{\mathbf{x}}) + \hat{\mathbf{x}}^\top \mathbf{A}\mathbf{x} + \mathbf{x}^\top \mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{x}}^\top \mathbf{A}\hat{\mathbf{x}} + \mathbf{b}^\top \mathbf{x} + c.$$

Note that $\mathbf{x}^\top \mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{x}}^\top \mathbf{A}\mathbf{x} = -\mathbf{b}^\top \mathbf{x}/2$, so rearranging terms yields

$$f(\mathbf{x}) = (\mathbf{x} - \hat{\mathbf{x}})^\top \mathbf{A}(\mathbf{x} - \hat{\mathbf{x}}) \underbrace{- \hat{\mathbf{x}}^\top \mathbf{A}\hat{\mathbf{x}} + c}_{\text{indpt. of } \mathbf{x}}.$$

Taking stock, we have proved the hint in the question. Now to argue that $\hat{\mathbf{x}}$ is the unique minimizer, note that $\mathbf{A}$ is positive definite, so the first term in the expansion of f is always non-negative. Furthermore, the unique value of $\mathbf{x}$ for which $(\mathbf{x} - \hat{\mathbf{x}})^\top \mathbf{A}(\mathbf{x} - \hat{\mathbf{x}}) = 0$ is $\mathbf{x} = \hat{\mathbf{x}}$.

Problem 2 (Monomial basis functions and invertibility). 20 points:

(a) In class, we claimed that given a dataset $(x_i, y_i)_{i=1}^n$ where $x_i, y_i \in \mathbb{R}$ were distinct across i, there is a *unique* polynomial of degree $n-1$ that interpolates these points. Without proving uniqueness, we showed that one such polynomial interpolator of degree $n-1$ was given by the Lagrange polynomial

$$p(x) = \sum_{k=1}^{n} y_k \cdot \prod_{\substack{1 \le j \le n \\ j \ne k}} \frac{x - x_j}{x_k - x_j}.$$

Argue formally that the Lagrange polynomial $p(x)$ is the *unique* polynomial of degree $n-1$ that interpolates the points $\{x_i, y_i\}_{i=1}^n$ *if and only if* $\mathbf{\Phi}$ is invertible.

Solution: Note that we need to establish both the *if and only if* directions. First, given that $\mathbf{\Phi}$ is invertible, show that the polynomial is unique. Second, given that the polynomial is unique, show $\mathbf{\Phi}$ is invertible.

Before we begin with the proof, we first construct the system of linear equations. Given $(x_i, y_i)_{i=1}^n$ distinct pairs, where $x_i \in \mathbb{R}$ are distinct, and a polynomial with coefficients α_i, $i = \{0, 1, \cdots, n-1\}$, we obtain

$$\begin{aligned}
\alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \cdots + \alpha_{n-1} x_1^{n-1} &= y_1 \\
\alpha_0 + \alpha_1 x_2 + \alpha_2 x_2^2 + \cdots + \alpha_{n-1} x_2^{n-1} &= y_2 \\
&\vdots \\
\alpha_0 + \alpha_1 x_n + \alpha_2 x_n^2 + \cdots + \alpha_{n-1} x_n^{n-1} &= y_n,
\end{aligned}$$

which can be written in the form

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & & & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$
$$\implies \mathbf{\Phi} \boldsymbol{\alpha} = \mathbf{y}$$

The matrix $\mathbf{\Phi}$ above is known as a Vandermonde matrix.

- Let us first show that the polynomial is unique given $\mathbf{\Phi}$ is invertible. Since $\mathbf{\Phi}$ is invertible, $\mathbf{\Phi}^{-1}$ is its unique inverse, implying $\boldsymbol{\alpha} = \mathbf{\Phi}^{-1}\mathbf{y}$ is unique. The resulting polynomial will be unique given the coefficient vector $\boldsymbol{\alpha}$ is unique. This polynomial must be the Lagrange polynomial given that this is one interpolator.

- Now, argue for the other direction. Suppose for the sake of contradiction $\mathbf{\Phi}$ does *not* have an inverse. This must imply that $\mathbf{\Phi}\boldsymbol{\alpha} = \mathbf{y}$ either has no solution or infinitely many solutions. However, we are given that the polynomial is unique and so $\boldsymbol{\alpha}$ is unique, which is a contradiction.

(b) Next, show that if $n = 2$ and $x_1 \neq x_2$, then there is a unique interpolating polynomial of degree 1 that interpolates these points.

Solution: Given $n = 2$, the Vandermonde matrix takes the following form,

$$\mathbf{\Phi} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}.$$

Compute the determinant to check invertibility of $\mathbf{\Phi}$ with $n = 2$: $\det(\mathbf{\Phi}) = x_2 - x_1$. Since $\det(\mathbf{\Phi}) \neq 0$, $\mathbf{\Phi}$ is invertible. Using part (a), given that $\mathbf{\Phi}$ is invertible, the resulting polynomial is unique with degree, $n - 1 = 1$, where $n = 2$. Another way to approach this is to observe that the columns of $\mathbf{\Phi}$ are linearly independent. Therefore, the matrix has full rank and is invertible.

(c) We will now try to set up a general way to attack the invertibility of $\mathbf{\Phi}$ by calculating its determinant. In particular, we will form another matrix $\mathbf{\Phi}'$. Let M_j denote the j-th column of a matrix $\mathbf{M}$ and execute the following loop:

for each $j = 0, \ldots n-1$: Set $\Phi'_{n-j} = \Phi_{n-j} - x_1 \cdot \Phi_{n-j-1}$.

Use the convention that $\Phi_0 = 0$. Argue that $\det(\mathbf{\Phi}) = \det(\mathbf{\Phi}')$.

Solution: Let us express the Vandermonde matrix in the following manner,

$$\mathbf{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_n \end{bmatrix},$$

where $\Phi_i \in \mathbb{R}^n$, $i = \{1, 2, \cdots, n\}$ is a column vector. You may have noticed that $\mathbf{\Phi}'$ was created from $\mathbf{\Phi}$ using elementary column operations. Writing the operation above out explicitly, we have

$$\mathbf{\Phi}' = \mathbf{\Phi} \cdot \begin{bmatrix} 1 & -x_1 & 0\ldots & & & & \\ 0 & 1 & -x_1 & 0 & \ldots & & \\ 0 & 0 & 1 & -x_1 & 0\ldots & & \\ & & & \vdots & & & \\ 0 & & & \ldots 0 & 1 & -x_1 & \\ 0 & & & \ldots & 0 & 1 & \end{bmatrix}.$$

Let Δ denote the second matrix above, and note that $\det(\Delta) = 1$ since it is a triangular matrix.
Consequently, we have

$$\det(\mathbf{\Phi}') = \det(\mathbf{\Phi}) \cdot \det(\Delta) = \det(\mathbf{\Phi}),$$

so that the determinant remains unchanged.

(d) Show that $\det(\mathbf{\Phi}) = (x_2 - x_1) \times (x_3 - x_1) \times \cdots \times (x_n - x_1) \times \det(\overline{\mathbf{\Phi}})$, where we have defined the $(n-1)\times(n-1)$ matrix

$$\overline{\mathbf{\Phi}} = \begin{bmatrix} 1 & x_2 & x_2^2 & \dots & x_2^{n-2} \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-2} \end{bmatrix}.$$

Solution: We begin with the Vandermonde matrix,

$$\mathbf{\Phi} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & & & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{bmatrix}.$$

Perform elementary operation on $\mathbf{\Phi}$: Subtract row $i>1$ with row $i=1$,

$$\mathbf{\Phi}_a = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & \cdots & x_2^{n-1} - x_1^{n-1} \\ \vdots & & & \ddots & \vdots \\ 0 & x_n - x_1 & x_n^2 - x_1^2 & \cdots & x_n^{n-1} - x_1^{n-1} \end{bmatrix}$$

Perform elementary operation on $\mathbf{\Phi}_a$: For column $j>1$, perform $\Phi_{a_j} - x_1\Phi_{a_{j-1}}$,

$$\mathbf{\Phi}_b = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & (x_2 - x_1)x_2 & \cdots & (x_2 - x_1)x_2^{n-2} \\ \vdots & & & \ddots & \vdots \\ 0 & x_n - x_1 & (x_n - x_1)x_n & \cdots & (x_n - x_1)x_n^{n-2} \end{bmatrix}$$

Taking the determinant of $\mathbf{\Phi}_b$, we obtain

$$\begin{aligned} \det(\mathbf{\Phi}_b) &= \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & x_2 - x_1 & (x_2 - x_1)x_2 & \cdots & (x_2 - x_1)x_2^{n-2} \\ \vdots & & & \ddots & \vdots \\ 0 & x_n - x_1 & (x_n - x_1)x_n & \cdots & (x_n - x_1)x_n^{n-2} \end{vmatrix} \\ &= (x_2 - x_1)\times(x_3 - x_1)\times\cdots\times(x_n - x_1)\times \begin{vmatrix} 1 & x_2 & x_2^2 & \dots & x_2^{n-2} \\ \vdots & & & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-2} \end{vmatrix} \\ &= \prod_{i=2}^{n}(x_i - x_1)\det(\overline{\mathbf{\Phi}}) \end{aligned}$$

From part (c), the determinants of the matrices $\mathbf{\Phi}$ and $\mathbf{\Phi}_b$ are the same. Hence, $\det(\mathbf{\Phi}) = \prod_{i=2}^{n}(x_i - x_1)\det(\overline{\mathbf{\Phi}})$.

(e) Proceed iteratively and produce an expression for $\mathsf{det}(\mathbf{\Phi})$. Conclude that $\mathbf{\Phi}$ is invertible if and only if the points $\{x_i\}_{i=1}^n$ are distinct.

Note: The situation when $\mathbf{x}_i \in \mathbb{R}^d$ is significantly more delicate; indeed, uniqueness of the points does not suffice. However, if the points are sufficiently generic (say $\{\mathbf{x}_i\}_{i=1}^n$ are chosen i.i.d. from a standard Gaussian distribution $\mathsf{N}(0, \mathbf{I})$) then it can be shown that $\mathbf{\Phi}$ is indeed invertible. Bonus credit for proving this, but note that it is a challenging problem.

Solution: We showed in part (d) that $\det(\mathbf{\Phi}) = \prod_{i=2}^n (x_i - x_1) \det(\overline{\mathbf{\Phi}})$, where $\overline{\mathbf{\Phi}}$ had the form,

$$\overline{\mathbf{\Phi}} = \begin{bmatrix} 1 & x_2 & x_2^2 & \dots & x_2^{n-2} \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-2} \end{bmatrix}.$$

Let's rename $\overline{\mathbf{\Phi}} = \mathbf{\Phi}_1$. Repeat the same process for $\mathbf{\Phi}_1$.
We will get $\det(\mathbf{\Phi}_1) = \prod_{i=3}^n (x_i - x_2) \det(\mathbf{\Phi}_2)$, where $\mathbf{\Phi}_2$ has the form,

$$\mathbf{\Phi}_2 = \begin{bmatrix} 1 & x_3 & x_3^2 & \dots & x_3^{n-3} \\ \vdots & & & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-3} \end{bmatrix}.$$

Iterating in this manner will ultimately bring us to the form,

$$\det(\mathbf{\Phi}_{n-3}) = \prod_{i=n-1}^{n} (x_i - x_{n-2}) \det(\mathbf{\Phi}_{n-2}).$$

Hence,

$$\det(\mathbf{\Phi}) = \prod_{1 \le i < j \le n} (x_j - x_i).$$

Since all the $\{x_i\}_{i=1}^n$ points are distinct points, the determinant of $\mathbf{\Phi}$ will be non-zero. And given $\mathbf{\Phi}$ is a square matrix with non-zero determinant, it is invertible.

We will release a note with a solution for the bonus problem later on in the class. If you believe you have solved it, please reach out to the staff with a private Piazza message.

(f) The remaining parts are unrelated to the previous parts. We will now argue that the number of monomials of degree less than or equal to ℓ in d variables can be explicitly calculated. First argue that the number of such monomials can be counted by placing d "stars" and $\ell - 1$ "bars" in a sequence and counting the number of unique configurations that result when these are permuted.

> **Solution:** There were a few clarifications posted on Piazza for this question, and we are fine with any logical explanation that leads to the right count (you may need to adjust the number of stars/bars depending on your argument).
> We will set up the following counting mechanism to count the number of unique polynomials in d variables with degree exactly at most ℓ. Note that what we are really counting is the number of possible powers of these variables (call them $\lambda_1, \ldots, \lambda_d$ and recall that they are all positive integers satisfying $\sum_{i=1}^{d} \lambda_i \leq \ell$). There are two ways to count this: the first is to find the number of terms with degree exactly i and sum over i (which will lead to the correct answer as well), but we will showcase an approach that introduces a dummy variable.
> Notice that the above count is equivalent to asking for non-negative integers $\lambda_1, \ldots, \lambda_{d+1}$ such that $\sum_{i=1}^{d} \lambda_i = \ell$, where we have introduced a dummy variable λ_{d+1}. Counting this is now the standard stars and bars argument, where we place d of one and ℓ of another in sequence and count the number of unique permutations.

(g) How many unique configurations are there above? Your expression should be explicit, and a function of the pair (d, ℓ).

> **Solution:** Note that we have placed d stars and ℓ bars in sequence and are counting the number of unique permutations. This comes to
>
> $$\binom{d+l}{l}.$$
>
> Another way to arrive at this answer is to sum over i (where i is the exact degree), yielding $\sum_{i=1}^{\ell} \binom{d+i-1}{i-1}$.

Problem 3 (Splines and linear equations). 20 points: Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a second-order spline defined by the overlap of five B-splines:

$$f(x) = \sum_{k=0}^{4} \alpha_k b_2(x-k),$$

where $b_2(x)$ is defined as on page 11 of the Fall 20 notes:

$$b_2(x) = \begin{cases} (x+3/2)^2/2 & -3/2 \le x \le -1/2 \\ -x^2 + 3/4 & -1/2 \le x \le 1/2 \\ (x-3/2)^2/2 & 1/2 \le x \le 3/2 \\ 0 & |x| \ge 3/2 \end{cases}.$$

(a) Write a function which takes $\alpha = \{\alpha_0, \dots, \alpha_4\}$ and `x` as input and returns samples of $f(x)$ at the locations specified in the vector `x`. Turn in a plot of $f(x)$ for $\alpha = \{2, 1, -1, 3, -1\}$. Sample x densely enough (by specifying sufficiently many points in `x`) so that your plot looks like a smooth function.

Solution:

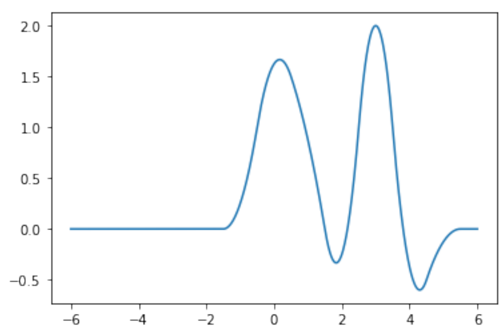


Figure 1: Plot for 3a

Code:

```
import numpy as np
import matplotlib.pyplot as plt

def piecepoly2(t, alpha):
    ft = np.zeros(len(t))
    t = np.array(t)
    for k, al in enumerate(alpha):
    ft = ft + al * bfun(t - k)
    return ft
```

```
def bfun(t):
    b = np.zeros(len(t))
    for idx, tt in enumerate(t):
    if tt >= -3./2. and tt <= -1./2.:
    bt = np.power((tt + 3./2.), 2) / 2.
    elif tt >= -1./2. and tt <= 1./2.:
    bt = -np.power(tt, 2) + 3/4.
    elif tt >= 1./2. and tt <= 3./2.:
    bt = np.power((tt - 3./2.), 2) / 2.
    else:
    bt = 0.
    b[idx] = bt
    return b

def getAlpha(t, k, f):
    f = np.array(f)
    t = np.array(t)
    A = np.zeros([len(f), k])
    for i in range(k):
    A[:, i] = bfun(t-i)
    alpha = np.linalg.solve(A, f)
    return alpha

if __name__ == "__main__":
    alpha = [3., 2., -1., 4., -1.]
    t = np.linspace(-6., 6., 1000)
    ft = piecepoly2(t, alpha)
    plt.plot(t, ft)
    plt.show()
    ft = [1., 2., -4., -5., -2.]
    t = [0., 1., 2., 3., 4.]
    alpha = getAlpha(t, 5, ft)
    print(alpha)
```

(b) Suppose

$$f(0) = 2, \quad f(1) = 2, \quad f(2) = -5, \quad f(3) = -5, \quad f(4) = -2.$$

What are the corresponding α_k? (Hint: you will have to construct a system of equations then solve it. You don't necessarily have to do this by hand.)

Solution: Using the code above, we can solve for the system of equations and obtain the corresponding α_k: $\boldsymbol{\alpha} = [2.1044733, 3.37316017, -6.34343434, -5.31255411, -1.78124098]$

(c) To generalize this, suppose that f is now a superposition of N B-splines, with

$$f(x) = \sum_{n=0}^{N-1} \alpha_n b_2(x-n).$$

Describe how to construct the $N \times N$ matrix that maps the coefficients α to the N samples $f(0), \ldots, f(N-1)$. That is, find a matrix $\mathbf{A}$ such that

$$\begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ f(N-1) \end{bmatrix} = \mathbf{A} \cdot \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \end{bmatrix}$$

Solution: The inner product of a matrix and a vector can be viewed as the weighted summation of the columns. Therefore the j^{th} column of A is $b_2(n-j)$. $A_{ij} = b_2(i-j)$.

(d) Argue that the matrix $\mathbf{A}$ from part (c) is invertible for all values of N.
(Hint: Observe that the matrix $\mathbf{A}$ is banded and has large values on the diagonal. How might you use this to argue invertibility?)

Solution: Matrix $\boldsymbol{A}$ will be a tridiagonal matrix with 3/4 for every main diagonal entry and 1/8 for every off-diagonal entry, e.g.

$$\boldsymbol{A} = \begin{bmatrix} 3/4 & 1/8 & \ldots & 0 & 0 \\ 1/8 & 3/4 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 3/4 & 1/8 \\ 0 & 0 & \ldots & 1/8 & 3/4 \end{bmatrix}$$

One decomposition for $\boldsymbol{A}$ is then $\boldsymbol{A} = (3/4)\mathbf{I} + (1/8)\boldsymbol{G}$ where $\boldsymbol{G}$ includes only off-diagonal ones, e.g.

$$\boldsymbol{G} = \begin{bmatrix} 0 & 1 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

Let $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{v}_1 = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_N \\ 0 \end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix}$.

$$
\begin{aligned}
\boldsymbol{A x} &= ((3/4)\mathbf{I} + (1/8)\boldsymbol{G})\boldsymbol{x} \\
&= (3/4)\mathbf{I}\boldsymbol{x} + (1/8)\boldsymbol{G x} \\
&= (3/4)\boldsymbol{x} + (1/8)\boldsymbol{G x} \\
&= (3/4)\boldsymbol{x} + (1/8)(\boldsymbol{v}_1 + \boldsymbol{v}_2)
\end{aligned}
$$

Computing the norm $\|\boldsymbol{A x}\|_2$ yields

$$
\begin{aligned}
\|\boldsymbol{A x}\|_2 &= \|(3/4)\boldsymbol{x} + (1/8)(\boldsymbol{v}_1 + \boldsymbol{v}_2)\|_2 && (1) \\
&\geq \|(3/4)\boldsymbol{x}\|_2 - \|(1/8)(\boldsymbol{v}_1 + \boldsymbol{v}_2)\|_2 && (2) \\
&\geq (3/4)\|\boldsymbol{x}\|_2 - (1/8)\|\boldsymbol{v}_1 + \boldsymbol{v}_2\|_2 && (3) \\
&\geq (3/4)\|\boldsymbol{x}\|_2 - (1/8)(\|\boldsymbol{v}_1\|_2 + \|\boldsymbol{v}_2\|_2) && (4) \\
&\geq (3/4)\|\boldsymbol{x}\|_2 - (1/8)(\|\boldsymbol{x}_1\|_2 + \|\boldsymbol{x}_2\|_2) && (5) \\
\|\boldsymbol{A x}\|_2 &\geq (1/2)\|\boldsymbol{x}\|_2 && (6)
\end{aligned}
$$

where equation 2 results from reverse triangle inequality, equation 3 results from homogeneity of norms, equation 4 results from triangle inequality, and 5 results from the inequalities $\|\boldsymbol{v}_1\|_2 \leq \|\boldsymbol{x}\|_2$, $\|\boldsymbol{v}_2\|_2 \leq \|\boldsymbol{x}\|_2$. From equation 6, we now know that $\|\boldsymbol{A x}\|_2 = 0 \implies \boldsymbol{x} = \mathbf{0}$, which proves $\boldsymbol{A}$ is invertible for all values of N.

(e) To take this even further, suppose that

$$
f(x) = \sum_{n=-\infty}^{\infty} \alpha_n b_2(x-n),
$$

so f is described by the (possibly infinite) sequence of numbers $\{\alpha_n\}_{n\in\mathbb{Z}}$. Show that there is a convolution operator that maps the sequence $\{\alpha_n\}$ to the sequence $\{f(n)\}$. That is, find a sequence of numbers $\{h_n\}_{n\in\mathbb{Z}}$ such that $f(n) = \sum_{\ell=-\infty}^{\infty} h_\ell\, \alpha_{n-\ell}$.

$$
\begin{aligned}
f(n) = \sum_{l=-\infty}^{\infty} \alpha_l b_2(n-l) = \sum_{(n-l)=-\infty}^{\infty} \alpha_{n-l} b_2(n-(n-l)) &= \sum_{(n-l)=-\infty}^{\infty} \alpha_{n-l} b_2(l) \\
&= \sum_{l=-\infty}^{\infty} \alpha_{n-l} b_2(l)
\end{aligned}
$$

Therefore, $h_l = b_2(l)$. Since $b_2(l)$ is only nonzero for 3 values, we obtain:

$$
h_l = \begin{cases} 3/4, & l = 0 \\ 1/8, & |l| = 1 \\ 0, & \text{else} \end{cases}
$$

Problem 4. (Fun with norms and inner products). 20 points:

(a) For parts (a-e), $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{R}^N$. Prove that $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_\infty \cdot \|\mathbf{y}\|_1$.

Solution:

$$
\begin{aligned}
|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| &= |\sum_{k=0}^{N} x_k y_k| \leq \sum_{k=0}^{N} |x_k||y_k| \leq \sum_{k=0}^{N} |y_k| \max_{n \in \{1,\cdots,N\}} |x_n| \\
|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| &\leq \sum_{k=0}^{N} |y_k| \max_{n \in \{1,\cdots,N\}} |x_n| = \max_{n \in \{1,\cdots,N\}} |x_n| \sum_{k=0}^{N} |y_k| = \|\boldsymbol{x}\|_\infty \|\boldsymbol{y}\|_1
\end{aligned}
$$

(b) Prove that $\|\mathbf{x}\|_1 \leq \sqrt{N} \cdot \|\mathbf{x}\|_2$. (Hint: Cauchy–Schwarz)

Solution:
Let $\boldsymbol{m} = [|x_1|, \cdots, |x_n|^T]$ and $\boldsymbol{n} = [1, \cdots, 1]^T$.

$$\langle \boldsymbol{m}, \boldsymbol{n} \rangle = \sum_{i=1}^{N} |x_i| = \|\boldsymbol{x}\|_1$$

$$\|\boldsymbol{m}\|_2 \cdot \|\boldsymbol{n}\|_2 = (\sum_{i=1}^{N} |x_i|^2)^{\frac{1}{2}} \cdot (\sum_{i=1}^{N} 1^2)^{\frac{1}{2}} = \sqrt{N} \cdot \|\boldsymbol{x}\|_2$$

By Cauchy-Schwarz ($|\langle \boldsymbol{m}, \boldsymbol{n} \rangle| \leq \|\boldsymbol{m}\|_2 \cdot \|\boldsymbol{n}\|_2$),

$$\|\boldsymbol{x}\|_1 \leq \sqrt{N} \cdot \|\boldsymbol{x}\|_2$$

(c) Let B_2 be the unit ball for the ℓ_2 norm in $\mathbb{R}^N$. Show that

$$\max_{\mathbf{x} \in B_2} \langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{y}\|_2.$$

Describe the vector $\mathbf{x}$ (as a function of $\mathbf{y}$) which achieves the maximum.

Solution:

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|_2 \cdot \|\boldsymbol{y}\|_2$$

In the Ball B_2,

$$\max_{\boldsymbol{x} \in B_2} \|\boldsymbol{x}\|_2 = 1.$$

So,

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{y}\|$$

If $\boldsymbol{y} = \mathbf{0}$, then $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| = \|\boldsymbol{x}\|_2 \cdot \|\boldsymbol{y}\|_2 = 0$, regardless of the choice of $\boldsymbol{x}$. If $\boldsymbol{y} \neq \mathbf{0}$, let $\boldsymbol{x} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2}$.

Then,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{\langle \boldsymbol{y}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|_2} = \frac{\|\boldsymbol{y}\|_2^2}{\|\boldsymbol{y}\|_2} = \|\boldsymbol{y}\|_2$$

and $\|\boldsymbol{x}\|_2 = 1$.
Therefore, If $\boldsymbol{y} \neq \boldsymbol{0}$, then $\boldsymbol{x} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2}$ achieves the maximum. If $\boldsymbol{y} = \boldsymbol{0}$, any vector $\boldsymbol{x} \in B_2$ achieves the maximum.
In either case,

$$\max_{\boldsymbol{x} \in B_2} \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{y}\|_2$$

(d) Let B_∞ be the unit ball for the ℓ_∞ norm in $\mathbb{R}^N$. Show that

$$\max_{\mathbf{x} \in B_\infty} \langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{y}\|_1.$$

Describe the vector $\mathbf{x}$ (as a function of $\mathbf{y}$) which achieves the maximum.

Solution:
From part a, we know that

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|_\infty \cdot \|\boldsymbol{y}\|_1.$$

Since $\boldsymbol{x} \in B_\infty$,

$$\|\boldsymbol{x}\|_\infty \leq 1$$

So,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq \|\boldsymbol{y}\|_1.$$

Let

$$\operatorname{sign}(\alpha) \;=\; \begin{cases} 1 & \alpha > 0 \\ -1 & \alpha < 0 \\ 0 & \alpha = 0 \end{cases}$$

Suppose $x_n = \operatorname{sign}(y_n)$ for all $n \in \{1, \cdots, N\}$. Then,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{N} \operatorname{sign}(y_i) y_i = \sum_{i=1}^{N} |y_i| = \|\boldsymbol{y}\|_1$$

So, $x_n = \operatorname{sign}(y_n)$ for all $n \in \{1, \cdots, N\}$ achieves the maximum, and

$$\max_{\boldsymbol{x} \in B_\infty} \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{y}\|_1$$

(e) Let B_1 be the unit ball for the ℓ_1 norm in $\mathbb{R}^N$. Show that

$$\max_{\mathbf{x}\in B_1}\langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{y}\|_\infty.$$

Describe the vector $\mathbf{x}$ (as a function of $\mathbf{y}$) which achieves the maximum.

Solution:
From part a, we know that

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| = |\langle \boldsymbol{y}, \boldsymbol{x}\rangle| \leq \|\boldsymbol{y}\|_\infty \cdot \|\boldsymbol{x}\|_1$$

$$\|\boldsymbol{y}\|_\infty = \max_{n\in\{1,\cdots,N\}} |y_n|$$

$$\|\boldsymbol{x}\|_1 \leq 1$$

So,

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{y}\|_\infty = \max_{n\in\{1,\cdots,N\}} |y_n|$$

Select a $k \in \{1, \cdots, N\}$ such that $|y_k| = \max_{n\in\{1,\cdots,N\}} |y_n|$. Let

$$\text{sign}\,(\alpha) \;=\; \begin{cases} 1 & \alpha > 0 \\ -1 & \alpha < 0 \\ 0 & \alpha = 0 \end{cases}$$

Now, let

$$x_n \;=\; \begin{cases} \text{sign}\,(y_n) & n = k \\ 0 & n \neq k \end{cases}$$

Then,

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{i=1}^{N} x_i y_i = \text{sign}\,(y_k) y_k = |y_k| = \|\boldsymbol{y}\|_\infty$$

Therefore,

$$x_n \;=\; \begin{cases} \text{sign}\,(y_n) & n = k \\ 0 & n \neq k \end{cases}$$

achieves the maximum, and

$$\max_{\boldsymbol{x}\in B_1}\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \|\boldsymbol{y}\|_\infty$$

(f) For parts (f-i), suppose you are given an $N \times N$ matrix $\mathbf{Q}$, and set

$$\langle \mathbf{x}, \mathbf{y}\rangle_Q = \mathbf{y}^\top \mathbf{Q}\mathbf{x},$$

for vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$. Prove that if $\mathbf{Q}$ has an entry along its diagonal that is nonpositive, then $\langle \cdot, \cdot\rangle_Q$ cannot be a valid inner product on $\mathbb{R}^N$.

Solution:
If the i_{th} diagonal entry of $Q \in \mathbb{R}^{N\times N}$ is non-positive, we can define a vector x $x \in \mathbb{R}^N$, $x = [0 \quad 0 \cdots 1 \cdots 0 \quad 0]$ is 1 on the i_{th} entry and 0 otherwise. Then we have $\langle x, x\rangle_Q = Q_{ii} \leq 0$. This violates the property that $\langle x, x\rangle \geq 0$, and $\langle x, x\rangle = 0$ only when $x = 0$.

(g) Prove that if $\mathbf{Q}$ is not symmetric, then $\langle \cdot, \cdot\rangle_Q$ cannot be valid inner product on $\mathbb{R}^N$.

Solution:
If Q is not symmetric, then there exists at least one pair of entries $Q_{ij} \neq Q_{ji}$. Define two vectors $x, y \in \mathbb{R}^N$, $x = [0 \quad 0 \cdots 1 \cdots 0 \quad 0]$ has 1 at the ith location and $y = [0 \quad 0 \cdots 1 \cdots 0 \quad 0]$ has 1 at the jth location. Then we have $\langle x, y\rangle_Q = Q_{ij} \neq Q_{ji} = \langle y, x\rangle_Q$, which violates the symmetry property.

(h) Recall that $\mathbf{Q}$ is *symmetric positive definite* if it is symmetric and obeys

$$\mathbf{x}^\top \mathbf{Q}\mathbf{x} > 0, \quad \text{for all } \ \mathbf{x} \in \mathbb{R}^N, \ \mathbf{x} \neq 0.$$

Prove that $\langle \cdot, \cdot\rangle_Q$ is a valid inner product on $\mathbb{R}^N$ if and only if $\mathbf{Q}$ is symmetric positive definite.

Solution:
Proof of this statement requires showing: i) if $\boldsymbol{Q}$ is sym+def, then $\langle \cdot, \cdot\rangle_Q$ is a valid inner product, and ii) if $\langle \cdot, \cdot\rangle_Q$ is a valid inner product, then $\boldsymbol{Q}$ is sym+def.
i) First property: $\boldsymbol{Q}$ is symmetric, so

$$\langle x, y\rangle_Q = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{y} = \langle y, x\rangle_Q$$

.
The second property of inner product will be trivially satisfied by any $\boldsymbol{Q}$, as matrix multiplication is a linear operator.

Third property: As $\boldsymbol{Q}$ is positive definite,

$$\langle x, x\rangle_Q = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} > 0, \quad \text{for all } \ \boldsymbol{x} \in \mathbb{R}^N, \ \boldsymbol{x} \neq \mathbf{0}$$

When $\boldsymbol{x} = 0, \langle x, x\rangle_Q = \mathbf{0}^{\mathrm{T}}\boldsymbol{Q}\mathbf{0} = 0$.

All properties of an inner product are satisfied, so $\langle \cdot, \cdot \rangle_Q$ is a valid inner product.

ii) We showed in b) that if $\boldsymbol{Q}$ is not symmetric, $\langle \cdot, \cdot \rangle_Q$ is not a valid inner product on $\mathbb{R}^N$. Then, $\langle \cdot, \cdot \rangle_Q$ is a valid inner product $\Rightarrow \boldsymbol{Q}$ is symmetric.

As $\boldsymbol{Q}$ is positive definite, $\langle x, x \rangle_Q = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{x} \geq 0, \ \forall \ \boldsymbol{x} \in \mathbb{R}^N$, and equality only holds when $x = \mathbf{0}$. This is exactly the definition of Q being positive definite.

We have separately showed Q has to be both symmetric and positive definite, so Q is sym+def.

(i) Define the norm on $\mathbb{R}^2$

$$\|\mathbf{x}\| = \|\mathbf{A}\mathbf{x}\|_2, \quad \mathbf{A} = \begin{bmatrix} 3 & 3 \\ -1/2 & 1/2 \end{bmatrix}.$$

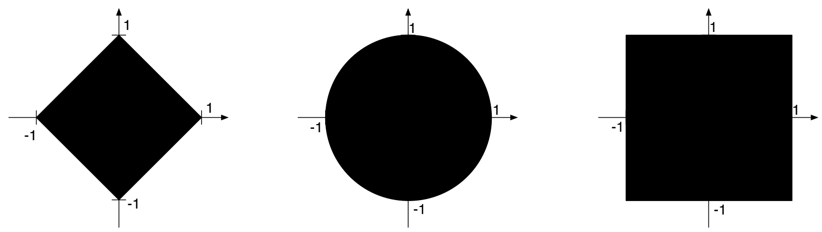
Find $\mathbf{Q}$ so that $\langle \cdot, \cdot \rangle_Q$ induces this norm.

Solution:

$$\boldsymbol{||x|| = ||Ax||_2 = \sqrt{(Ax)^T Ax} = \sqrt{x^T A^T Ax}}$$

$$\boldsymbol{Q = A^T A} = \frac{1}{4} \begin{bmatrix} 37 & 35 \\ 35 & 37 \end{bmatrix} \tag{7}$$

Problem 5. (Visualizing norm balls). 20 points One way to visualize a norm in $\mathbb{R}^2$ is by its *unit ball*, the set of all vectors such that $\|\mathbf{x}\| \leq 1$. For example, we have seen that the unit balls for the ℓ_1, ℓ_2, and ℓ_∞ norms look like:



$B_1 = \{\mathbf{x} : \|\mathbf{x}\|_1 \leq 1\}$ $B_2 = \{\mathbf{x} : \|\mathbf{x}\|_2 \leq 1\}$ $B_\infty = \{\mathbf{x} : \|\mathbf{x}\|_\infty \leq 1\}$

Given an appropriate subset of the plane, $B \subset \mathbb{R}^2$, it might be possible to define a corresponding norm using

$$\|\mathbf{x}\|_B = \text{minimum value } r \geq 0 \text{ such that } \mathbf{x} \in rB, \tag{8}$$

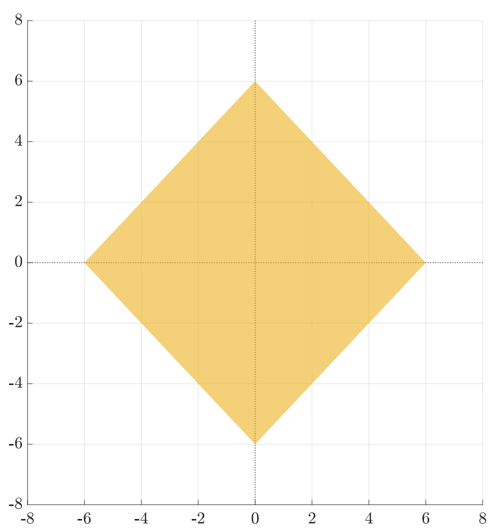
where rB is just a scaling of the set B:

$$\mathbf{x} \in B \quad \Rightarrow \quad r \cdot \mathbf{x} \in rB.$$

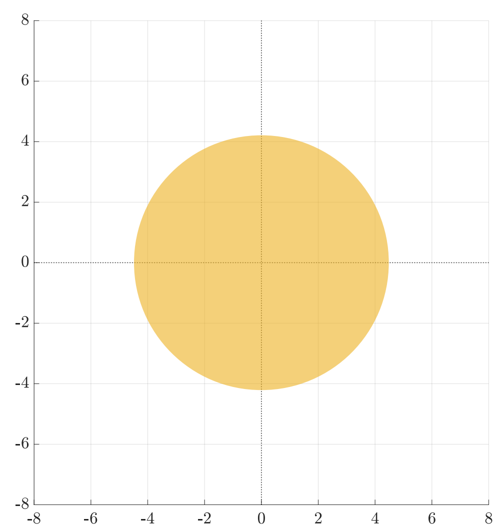
(a) Let $\mathbf{x} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$. For $p = 1, 2, \infty$, find $r = \|\mathbf{x}\|_p$, and sketch $\mathbf{x}$ and rB_p (use different axes for each of the three values of p).

Solution:

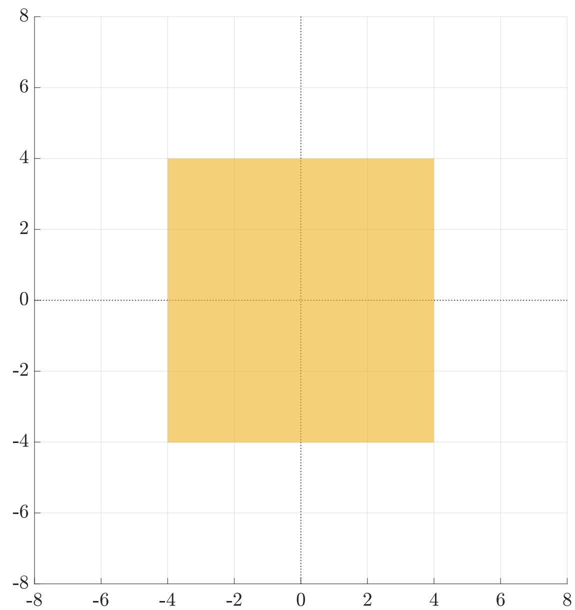
$$\|\boldsymbol{x}\|_1 = 6 \quad \|\boldsymbol{x}\|_2 = 4.47 \quad \|\boldsymbol{x}\|_\infty = 4$$



$6B_1$

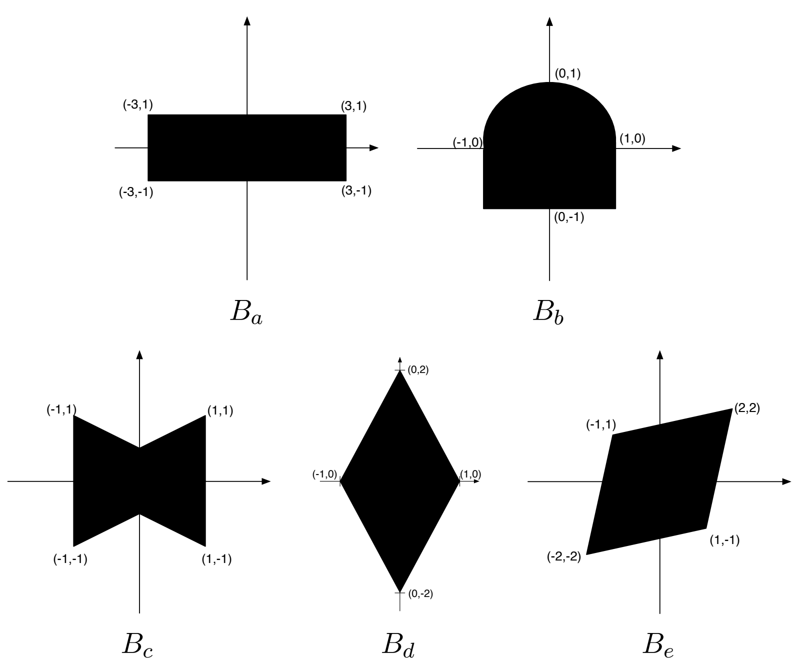


$4.47B_2$



$4B_\infty$

(b) Consider the 5 shapes below.



Explain why $\|\cdot\|_{B_b}$ and $\|\cdot\|_{B_c}$ are **not** valid norms. The most convincing way to do this is to find vectors for which one of the three properties of a valid norm are violated.

Solution:
For B_b, we can see that $\boldsymbol{x}_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ has norm $\|\boldsymbol{x}_1\|_{B_b} = 1$, while $\|-\boldsymbol{x}_1\|_{B_b} > 1$. So homogeneity is violated.
For B_c, consider $\boldsymbol{x}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. It is clear that $\|\boldsymbol{x}_1\|_{B_c} = \|\boldsymbol{x}_2\|_{B_c} = 1$. But note that $\boldsymbol{x}_1 + \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, and $\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_{B_c} > 2$. This means that $\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_{B_c} > \|\boldsymbol{x}_1\|_{B_c} + \|\boldsymbol{x}_2\|_{B_c}$. Triangle inequality is violated.

(c) Give a concrete method for computing $\|\mathbf{x}\|_{B_a}$, $\|\mathbf{x}\|_{B_d}$, and $\|\mathbf{x}\|_{B_e}$ for any given vector $\mathbf{x}$. (For example: for B_1, which corresponds to the ℓ_1 norm, we would write $\|\mathbf{x}\|_1 = |x_1| + |x_2|$.) Using you expressions, show that these are indeed valid norms.

B_a:

$$\|\boldsymbol{x}\|_{B_a} \;=\; \max\{\frac{|x_1|}{3}, |x_2|\} \;=\; \left\| \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{x} \right\|_\infty$$

B_d:

$$\|\boldsymbol{x}\|_{B_d} \;=\; |x_1| + \frac{1}{2}|x_2| \;=\; \left\| \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \boldsymbol{x} \right\|_1$$

B_e:

$$\|\boldsymbol{x}\|_{B_e} \;=\; \frac{1}{4}|x_1 + x_2| + \frac{1}{2}|-x_1 + x_2| \;=\; \left\| \begin{bmatrix} 1/4 & 1/4 \\ -1/2 & 1/2 \end{bmatrix} \boldsymbol{x} \right\|_1$$

So, all three of these norm expressions can be written in the form:

$$\|\boldsymbol{x}\|_{B_k} = \|\boldsymbol{A}\boldsymbol{x}\|_p$$

for some matrix $\boldsymbol{A}$ and norm p. In this form, triangle inequality and homogeneity follow immediately from the linearity of $\boldsymbol{A}$ and that the l_p norm satisfies triangle inequality and homogeneity.

$$\begin{aligned}
\|\alpha\boldsymbol{x}\|_{B_k} &= \|\boldsymbol{A}(\alpha\boldsymbol{x})\|_p = \|\alpha\boldsymbol{A}\boldsymbol{x}\|_p = |\alpha|\,\|\boldsymbol{A}\boldsymbol{x}\|_p = |\alpha|\,\|\boldsymbol{x}\|_{B_k}, \\
\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_{B_k} &= \|\boldsymbol{A}(\boldsymbol{x}_1 + \boldsymbol{x}_2)\|_p \\
&= \|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_2\|_p \le \|\boldsymbol{A}\boldsymbol{x}_1\|_p + \|\boldsymbol{A}\boldsymbol{x}_2\|_p \\
&= \|\boldsymbol{x}_1\|_{B_k} + \|\boldsymbol{x}_2\|_{B_k}.
\end{aligned}$$

To show that $\|\boldsymbol{x}\|_{B_k} = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$, we need to ensure that there is no $\boldsymbol{x} \neq \mathbf{0}$ such that $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$. This is true if $\boldsymbol{A}$ is invertible (i.e. $\boldsymbol{A}$ is square and has linearly independent columns). In all three cases above, our matrix is invertible, so clearly our expressions for $\|\boldsymbol{x}\|_{B_b}$, $\|\boldsymbol{x}\|_{B_d}$, and $\|\boldsymbol{x}\|_{B_e}$ are valid norms.