The notion of studying norms is important because we want to define metrics by which we can assess the shape and size of vectors, as well as measure the difference between them, which is crucial for understanding optimization problems, stability, and sensitivity in numerical linear algebra.

1. Vector Norms

We start by defining the ℓ_p norms.

Given a vector $x \in \mathbb{R}^m$:

The ℓ_1 norm is defined by:

$$\|x\|_1 = \sum_{i=1}^{m} |x_i| \tag{1}$$

This is sometimes called the Manhattan distance, or taxicab distance, to express the total travel distance when moving along grid-like paths, where you can only move horizontally or vertically, much like a taxi navigating city streets.

The ℓ_2-norm is defined by:

$$\|x\|_2 = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2} \tag{2}$$

The ℓ_2 norm is essentially a generalization of the distance formula to higher-dimensional spaces, extending the concept of Euclidean distance beyond two or three dimensions.

Generally, the ℓ_p norm is defined by:

$$\|x\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p}$$

ℓ_p norms greater than 2 are not often used because they can lead to less intuitive geometric interpretations and are more computationally expensive, as they require higher powers and roots. Additionally, they don't have the same desirable properties as the ℓ_2 norm, such as differentiability, which is useful for optimization.

And finally the ℓ_∞ norm of a vector is defined by:

$$\|x\|_\infty = \max_{1 \le i \le m} |x_i|.$$

The unit ball represents all vectors that are "closest" to the origin according to the given norm, and its shape varies depending on whether we use the ℓ_1, ℓ_2, or ℓ_∞ norm. By studying the unit ball, we gain insights into the structure of the space and how distances between vectors are measured.

In mathematical terms this unit ball can be expressed as:

$$\left\{x \in \mathbb{R}^m \middle| \; \|x\| \leq 1\right\}$$

This translates to "the set of all points in a space $\mathbb{R}^m$ such that the norm is less than or equal to 1."

So to express the unit ball defined by the L_1 norm in $\mathbb{R}^2$, we can write:

$$\left\{x \in \mathbb{R}^2 \middle| \; \|x\|_1 \leq 1\right\}$$

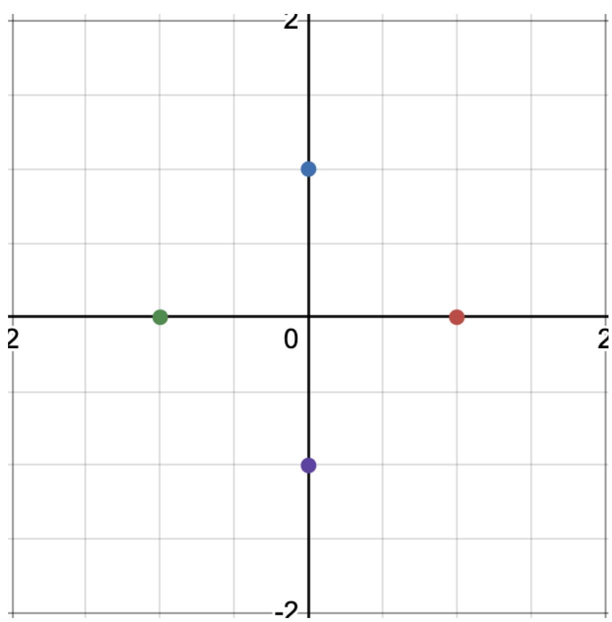
Lets try to figure out what this looks like.

Let's pick points that lie on the boundary. In other words, find vectors in $\mathbb{R}^2$ such that the norm is exactly equal to 1.

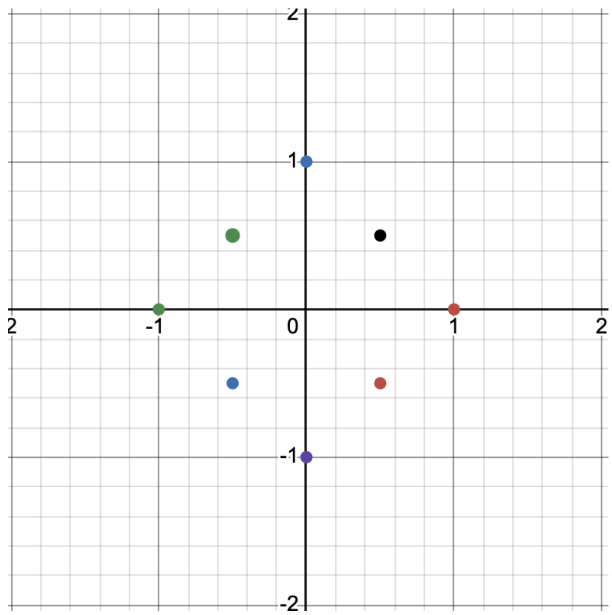
The simplest vectors with an L1 norm of 1 are: $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

You can verify that these satisfy the norm $\|x\|_1 = 1$ by plugging it into (1).

Plotting these points in $\mathbb{R}^2$ we have:



Now some intermediary points that also satisfy $\|x\|_1 = 1$: $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}, \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix}$.



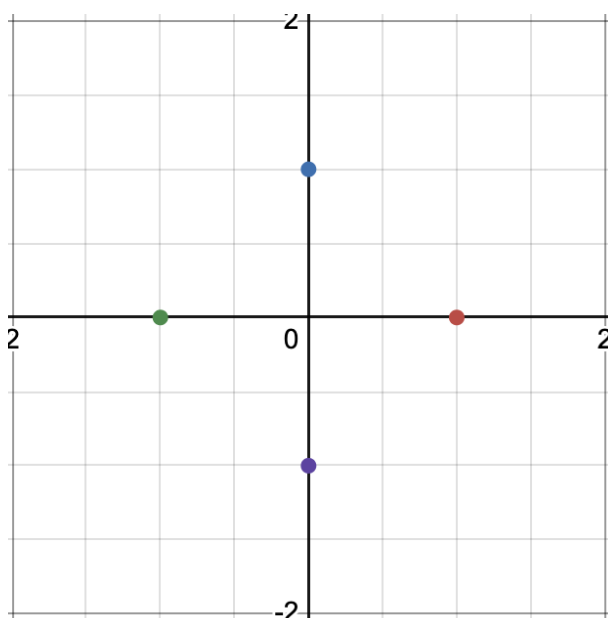
If you continue to search for vectors x in $\mathbb{R}^2$ such that $\|x\|_1 = 1$, you will notice that the boundary forms a square rotated 45 degrees from the origin. The unit ball, defined by the inequality, $\|x\|_1 \leq 1$, will be filled, as the condition allows for all points inside the square as well.

Let's do the same for the ℓ_2 norm in $\mathbb{R}^2$.

$$\left\{x \in \mathbb{R}^2 |\ \|x\|_2 \leq 1\right\}$$

Start by trying to find points such that the condition $\|x\|_2 = 1$ is met.

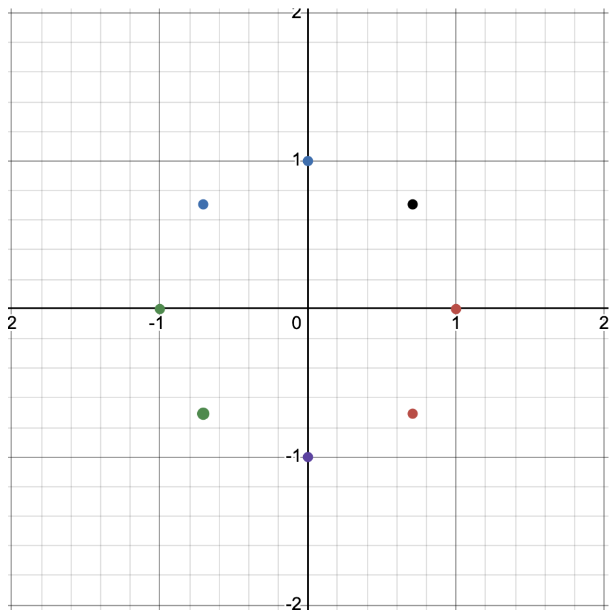
Note the same vectors from above $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ also satisfy the 2-norm boundary condition. Se we can start in the same position:



The following vectors satisfy the condition $\|x\|_2 = 1$

$$\begin{bmatrix} \sqrt{0.5} \\ \sqrt{0.5} \end{bmatrix}, \begin{bmatrix} -\sqrt{0.5} \\ \sqrt{0.5} \end{bmatrix}, \begin{bmatrix} \sqrt{0.5} \\ -\sqrt{0.5} \end{bmatrix}, \begin{bmatrix} -\sqrt{0.5} \\ -\sqrt{0.5} \end{bmatrix}$$

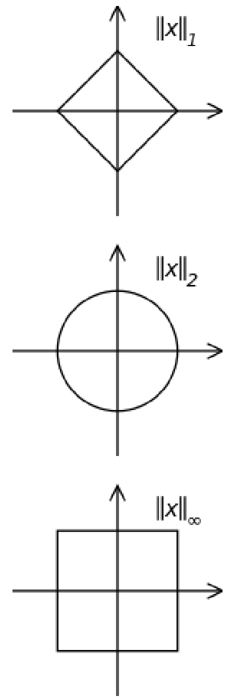
Again, you can check this by plugging it into (2).



If you do this for a sufficient amount of points, you will come to the conclusion that the boundary is defined by a circle.

Try this out yourself using the ℓ_∞!

You could try this with other norms besides 1 and 2, but it might be cumbersome to calculate them manually. Here's the visualization you should expect:



Induced Norms

Now we move on to the study of induced vector norms, sometimes called *operator norms*.
Often times we want to understand how a matrix A transforms a vector x. The goal of the induced vector norm is to quantify the maximum stretching or scaling that A applies to any vector x, based on the chosen vector norm. This helps us analyze the behavior of matrices in terms of how they affect the magnitude of vectors. For example, in solving linear systems, the induced norm can help estimate how much errors or small perturbations in input can affect the solution. We need a way to quantify how A, the matrix representing the system, amplifies or diminishes these errors.

Given a vector $x \in \mathbb{R}^n$ and a matrix $A \in \mathbb{R}^{m \times n}$, the induced matrix norm is defined as:

$$\|A\| = \sup_{\|x\|=1} \|Ax\| \tag{3}$$

The induced ℓ_1 norm is therefore:

$$\|A\|_1 = \sup_{\|x\|_1=1} \|Ax\|_1.$$

The induced ℓ_2 norm is:

$$\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2$$

and so on.

The induced norm can also be written as:

$$\sup_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|}. \tag{4}$$

(3) and (4) are equivalent because if we restrict the condition to $\|x\| = 1$, the denominator in the expression

$\frac{\|Ax\|}{\|x\|}$ becomes 1, simplifying the ratio to $\|Ax\|$. Since any nonzero vector x can be scaled to a unit vector $\widehat{x} = \frac{x}{\|x\|}$, the supremum over all nonzero vectors is the same as the supremum over unit vectors. Therefore, the two definitions capture the same maximum stretching effect, and the induced norm can be written as $\sup_{\|x\|=1} \|Ax\|$.

Two things to notice:
1) The use of $\sup$ instead of $\max$. What is the difference? We use the supremum (sup) instead of the maximum because the supremum is more general and can be defined even when the maximum doesn't exist. The supremum represents the least upper bound of a set, capturing the largest possible value that a function can approach, even if it never exactly reaches it. This is particularly important in continuous or infinite-dimensional spaces, where there may not be a specific vector that achieves the maximum, but the supremum still exists.

2) $\|x\| = 1$ on the bottom is specifying the set of all vectors x that have a norm of 1, essentially the unit ball in the given norm. When we calculate the induced norm of a matrix, we're looking at how much the matrix can stretch any vector within this unit ball.
Recall when we tried to plot the p-norms above we started by finding vectors which statisfy the condition $\|x\| = 1$. This is because these vectors represent the boundary points, the maximum distance from the origin according to the given norm.
The supremum is then the maximum amount that the matrix can stretch any of these unit vectors, capturing the "worst-case" scaling effect of the matrix on vectors in the unit ball.

Matrix Norms

The induced matrix ℓ_1 Norm:
$A \in \mathbb{R}^{m\times n}$

$$\|A\|_1 = \max_{1\le j\le n} \|a_j\|_1$$

Proof: *Consider a matrix $A \in \mathbb{R}^{m\times n}$ and define its columns as $a_j \in \mathbb{R}^m$. Now consider the L_1 norm unit ball $\left\{x \in \mathbb{R}^n \mid \Sigma_{j=1}^n \|x_j\|_1 \le 1\right\}$. We want to define $\|Ax\|_1$ in terms of an induced norm (the measures of the largest possible scaling effect on some vector x). $\|Ax\|_1 = \|\Sigma_{j=1}^n a_j x_j\|_1$ (note x_j are components of x and a_j are column vectors of A). By the triangle inequality in L_1 norm, we have the upper bound $\|\Sigma_{j=1}^n a_j x_j\|_1 \le \Sigma_{j=1}^n |x_j| \|a_j\|_1$. Since $\Sigma_{j=1}^n |x_j| \le 1$ from the unit ball definition, then $\|\Sigma_{j=1}^n a_j x_j\|_1 \le \Sigma_{j=1}^n |x_j| \|a_j\|_1 \le \max_{1\le j\le n} \|a_j\|_1$. Equality is attained by choosing $x = e_j$ (basis vectors) and thus* $\|Ae_j\|_1 = \|A\|_1 = \max_{1\le j\le n} \|a_j\|_1$.

This matrix L_1 norm can be understood as the maximum column sum.

The induced matrix ℓ_∞ Norm:

$$\|A\|_1 = \max_{1 \leq j \leq m} \|a_j^T\|_1$$

Proof: Left as an exercise to the reader (; (follows same structure as proof above but for rows instead of columns).

The matrix ℓ_∞ norm can be understood as the maximum row sum.

Finally, I will define two more norms, one of which is *not* defined via an induced norm.
The Frobenius norm:

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{n}\|a_j\|_2^2\right)^{\frac{1}{2}} = \sqrt{\operatorname{tr}\left(A^T A\right)} = \sqrt{\sum_i \sigma_i^2}$$

*where σ_i are the singular values of the matrix A.

tr() here denotes *trace*. The trace of a matrix is simply the sum of the diagonal elements of a matrix. Convince yourself that these equalities hold by performing the operations defined after each consecutive equality on a simple 2 by 2 matrix. The frobenius norm is analogous to the Euclidean norm for vectors, but applied to matrices. This is not to be confused with the *induced* ℓ_2-norm.

The ℓ_2-norm for vectors is the Euclidian analog for vectors but the frobenius norm is the Euclidean analog for matrix. We call the induced induced ℓ_2-norm the spectral norm, and it is defined as follows:

The spectral Norm (induced ℓ_2-norm):

$$\|A\|_2 = \max_i \ \sigma_i$$

The proof for the induced ℓ_2-norm is somewhat convoluted and involves advanced concepts, so I will leave it out for now.

Summary
Vector Norms:
1. ℓ_1-norm (Manhattan norm): $\|\mathbf{x}\|_1 = \sum_{i=1}^{n}|x_i|$ (Measures the sum of the absolute values of the components of the vector.)
2. ℓ_2-norm (Euclidean norm): $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$ (Measures the length (distance from the origin) of the vector in Euclidean space.)
3. ℓ_∞-norm (Maximum norm): $\|\mathbf{x}\|_\infty = \max_i |x_i|$ (Measures the largest absolute value among the components of the vector.)

Matrix Norms:

1. Frobenius norm: $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$ (Measures the square root of the sum of the squares of all the entries in the matrix.)

2. Induced ℓ_1-norm: $\|A\|_1 = \max_j \sum_i |a_{ij}|$ (Measures the maximum absolute column sum of the matrix.)

3. Induced ℓ_2-norm (Spectral norm): $\|A\|_2 = \sigma_{\max}(A)$ (Measures the largest singular value of the matrix, representing its maximum stretching factor.)

4. Induced ℓ_∞-norm: $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ (Measures the maximum absolute row sum of the matrix.)