EECE 7200 Homework 2

Mathematical notations and abbreviations are defined in the document **Commonly Used Symbols** at the course web page.

1. The circle of a unit radius around the origin of a plane, briefly referred to as the *unit circle*, comprises of all the points whose distance from the origin is 1. This concept generalizes to the concept of the *unit sphere* in any normed vectors space: it is the set of of points whose distance from the origin, defined by the norm, is 1:

$$S = \{x \in \mathbf{V} \,:\, \|x\| = 1\}.$$

In order to illustrate this concept you are requested to use MATLAB to plot, in a single figure, the unit spheres (or circles)

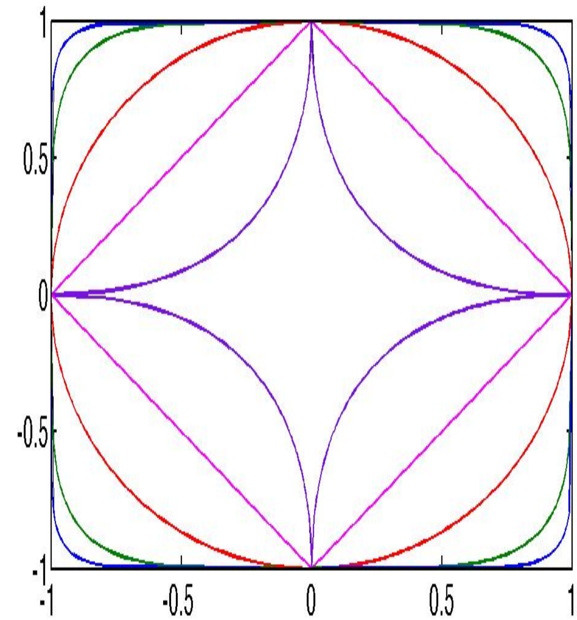

Figure 1: The unit circles in the plane, relative to the 0.5-“norm” (purple), 1-norm (violet), 2-norm (red), 5-norm (green), 10-norm (blue) and ∞-norm (black frame).

in the real plane, $\mathbb{R}^2$, as they are defined p norms for $p = 0.5$, 1, 2, 5, 10 and ∞.

Recall that the definitions of “p-norms” in the vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, \quad p < \infty, \;\text{ and }\; \|x\|_\infty = \max_{1\le i\le n} |x_i|$$

The plots are shown in Figure 1. Each curve is tagged with the respective value of p. The parameterization of the curves is obtained from the respective definitions of the norms. Thus, for $p < \infty$ we parameterize $x_2(x_1)$ as

$$x_2 = \pm\left(1 - |x_1|^p\right)^{\frac{1}{p}}, \quad x_1 \in [-1, 1].$$

For $p = \infty$, one of the two coordinates of $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, where the maximal amplitude is attained, must have an amplitude of 1, while the other's amplitude is at most 1. Thus S is the union of four intervals:

$$\{x_1 = \pm 1,\, x_2 \in [-1,\, 1]\} \cup \{x_1 \in [-1,\, 1],\, x_2 = \pm 1\}.$$

Challenge: Consider the three axioms that must be satisfied by a valid norm, and see if you can suggest from the plots for which values of p does the formal definition of $\|\cdot\|_p$ yield a valid norm and for which it is not? A good argument is good enough – you are not required to provide a formal proof.

The answer to this challenge is based on the following observation:
Observation. *Let $\|\cdot\|$ be a candidate for a norm on V, satisfying the first two axioms of a norm*[1]*. The triangle inequality is then equivalent to the requirement that the unit disk*[2] $D = \{x \in V \,:\, \|x\| \leq 1\}$*, is a* ***convex*** *set.*

The property of convexity means that if $x,\, y \in D$ then the entire interval that connects these two points is inside the disk. It i easy to see that the interval connecting two points, x and y is parameterized as

$$\mathcal{I}(x,y) = \{\lambda x + (1-\lambda)y \,:\, \lambda \in [0,\,1]\}.$$

Proof of the Observation. To prove an equivalence we need to establish two implications.

Assume the triangle inequality is satisfied. We have to show that D is convex. To do so, let $x,\, y \in D$ and $\lambda \in [0,\,1]$. Then

$$|\lambda x + (1-\lambda)y\| \underbrace{\leq}_{\substack{\text{triangle}\\ \text{inequality}}} \|\lambda x\| + \|(1-\lambda)y\| = \lambda \underbrace{\|x\|}_{\leq 1} + (1-\lambda)\underbrace{\|y\|}_{\leq 1} \leq \lambda + (1-\lambda) = 1 \quad \Rightarrow \quad \lambda x + (1-\lambda)y \in D.$$

This establishes that, indeed, D is convex.

Assume that D is convex. We have to show that the triangle inequality is satisfied To do so let $x,\, y \in V$. To avoid the trivial case, assume $x,\, y \neq 0$. Let us make the following definitions –

$$\tilde{x} = \frac{1}{\|x\|}x, \quad \tilde{y} = \frac{1}{\|y\|}y, \quad \lambda = \frac{\|x\|}{\|x\| + \|y\|}, \quad 1-\lambda = \frac{\|y\|}{\|x\| + \|y\|}.$$

Clearly, $\tilde{x},\, \tilde{y} \in S \subset D$ and $\lambda \in (0,1)$. The assumed convexity of D therefore implies that

$$\lambda\tilde{x} + (1-\lambda)\tilde{y} \in D.$$

Expanding this expression by the original notations -

$$1 \geq \|\lambda\tilde{x} + (1-\lambda)\tilde{y}\| = \left\| \frac{\|x\|}{(\|x\| + \|y\|)} \cdot \frac{1}{\|x\|}x + \frac{\|y\|}{(\|x\| + \|y\|)} \cdot \frac{1}{\|y\|}y \right\| = \frac{1}{(\|x\| + \|y\|)}\|x+y\|.$$

Therefore $\|x\| + \|y\| \geq \|x+y\|$ and the triangle inequality is established. $\square$

Now, from the plots in Figure 1 it is apparent that D is convex for $p \geq 1$, but that it is not convex for $p = 0.5$. (As a matter of fact, it is not convex for any $p < 1$.) The lack of convexity for $p = 0.5$ can also be demonstrated algebracaly, by computation for a single counter-example[3]: Let

$$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \;\; y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \;\; \text{and} \;\; z = 0.5\,x \,+\, 0.5\,y \,=\, \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}.$$

In this case, indeed, $x,\, y \in D$:

$$\|x\|_{0.5} \,=\, \|y\|_{0.5} \,=\, \left(1^{0.5} + 0^{0.5}\right)^{\frac{1}{0.5}} = 1.$$

However $z \notin D$:

$$\|z\|_{0.5} = \left(0.5^{0.5} + 0.5^{0.5}\right)^{\frac{1}{0.5}} = \left(\frac{2}{\sqrt{2}}\right)^2 = (\sqrt{2})^2 = 2 > 1.$$

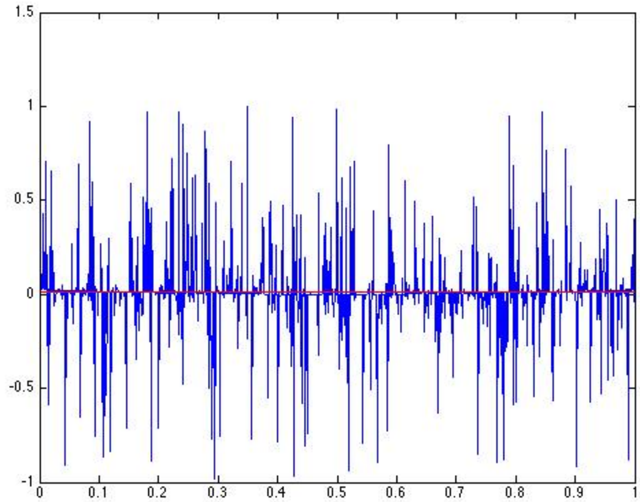

Figure 2: The data in Problem 2 (blue) and the optimal guess for y which are undistinguishable here between the 3 norms

The file `EECE7200HW2data.mat` contains data used questions 2.

2. The requested plot is provided in Figure 2.
The best approximation in the ∞-norm is the value that minimizes the distance from both the minimal and the maximal values of the data. That is, it is the median value of the data `yn`. With a 4 digits precision, it is 0.011.

The optimal values for the 1-norm and the 2-norm are estimated here by an iterative search algorithm. Each iteration comprises of the following steps:
Step 1. Create a grid of equally spaced guesses for the optimal value of `c`; initially the grid span `[min(yn), max(yn)]`

Step 2. Find the grid value that minimizes the error norm $\|yn - c\|_p$ (for the appropriate p.

Step 3. Narrow the search interval to the two neighbors of the grid point is the best guess.

Step 4. Returns to Step 1.

The iterations stop when the length of the search interval becomes shorter than the desired precision tolerance.

With a precision tolerance of 0.0001, the optimal value for p=1 is 0.0059 and for p=2 it is 0.011.
In this data set the differences between the 3 estimates are small relative to error sizes. In general, the ∞ norm stresses worst-case errors. Both the 1-norm and 2-norm consider the error in aggregate. The 2 norm sums the squared errors. Taking the square increases large numbers and decreases small ones. Therefore, when compared tot the 1-norm, optimization with respect to the 2-norm stresses the aggregate minimization of large errors and reduces the effect of small errors on the optimization.

3. Find an example of a sequence $\{x_k\}_{k=1}^{\infty} \subset \ell$ such that -

1. $\lim_{k\to\infty} \|x_k\|_1 = \infty$

[1]That is, $\forall x \in V$, $\|x\| \geq 0$, $\|x\| > 0 \Leftrightarrow x \neq 0$, and $\forall \alpha \in \mathbb{F}$, $\|\alpha x\| = |\alpha|\|x\|$.
[2]The open unit disk comprises the points inside the unit circle, and the closed disk, D, is the union of the unit circle and its interior.
[3]While to prove a general statement requires to address its complete generality, a single counter example suffices to falsify it!

2. $\forall k, \quad \|x_k\|_2 = 1$

3. $\lim_{k\to\infty} \|x_k\|_\infty = 0$

Recall: "$x_k \in \ell$" means that $x_k = (\alpha_{k,1}, \alpha_{k,2}, \dots, \alpha_{k,i}, \dots)$, for each k. To do so you have to write down the definitions of the three norms and consider what properties of the entries $\alpha_{k,j}$ are needed for the three properties to simultaneously hold. Answer: We list below the intuitive considerations in selecting an example. We shall later use these intuitive observations to construct and rigorously justify the sought example.

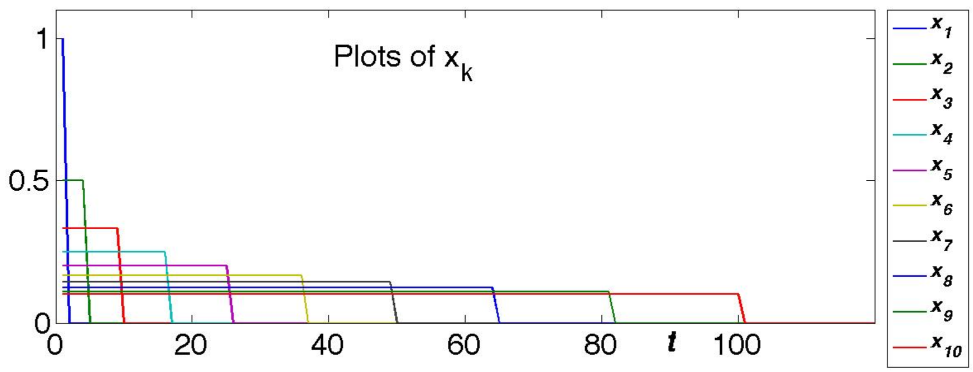

Figure 3: Plots of x_k, $k = 1, \dots, 10$, in the answer to Problem 4. (Continuous curves are used for visual clarity.)

— Requirement (iii) means that the functions $|x_k|$ become increasingly shallow as k grows.

— The requirements that $x_k \in \ell_p$, $p = 1, 2$, means that $\|x_k\|_p < \infty$, for $p = 1, 2$, and all $k = 1, 2, \dots$. These requirements imply that, for each k, the function x_k declines to zero at infinity: $x_k(t) \underset{t\to\infty}{\longrightarrow} 0$.

— Requirement (i) is that the area under the graph of $|x_k|$ grows indefinitely, as k grows. This means that while the graph of $|x_k|$ becomes more shallow, it also becomes wider as k grows. That is, the decay to zero of $|x_k(t)|$, as t grows, occurs later and later, as k grows.

— Requirement (ii) provides a handle for quantifiable relationships between "shallow" and "wide".

Arguably the simplest example one can construct is of a sequence of boxcar functions:

$$x_k(t) = \begin{cases} h_k & 1 \le t \le t_k \\ 0 & t > t_k \end{cases}, \qquad k, t \in \mathbb{N}.$$

A quantitative examination of what Requirements (i-iii) imply regarding the selections of h_k and t_k is shown in Table 1. Plots of x_k, $k = 1, \dots, 10$, are shown in Figure 3 for the selections $h_k = 1/k$.

Requirement	Property	Implication
(iii) $\|x_k\|_\infty \underset{k\to\infty}{\longrightarrow} 0$	$\|x_k\|_\infty = h_k$	$h_k \underset{k\to\infty}{\longrightarrow} 0$
(ii) $\|x_k\|_2 = 1$	$\|x_k\|_2^2 = h_k^2 \cdot t_k$	$t_k = 1/h_k^2$
(i) $\|x_k\|_1 \underset{k\to\infty}{\longrightarrow} \infty$	$\|x_k\|_1 = h_k \cdot t_k = h_k \cdot (1/h_k^2) = 1/h_k$	guaranteed by property (iii)

Table 1: The derivation of guidelines for selecting h_k and t_k in the answer to Problem 3.