ℓ^0-norm, ℓ^1-norm and ℓ^4-norm in S^n

Kwan Ho Ryan Chan
Department of Electrical Engineer & Computer Science
University of California, Berkeley

December 12, 2019

Source Code

The code used to generate the following plots are uploaded to Github. Link: `https://github.com/ryanchankh/L0L1L4NormSn`

1 Introduction

Suppose we have an optimization problem as follows:

$$\underset{||\boldsymbol{x}||_2=1}{\operatorname{argmin}} \quad ||\boldsymbol{x}||_0 \tag{1}$$

In plain words, we are trying to find the sparsest solution (most 0's entries), such that the solution is a point on the sphere S^n. This problem is NP-hard. So a natural way to solve this is by relaxing the problem to another form. The goal of this article is to show how the solutions to (1) corresponds to solving the following:
Case 1:

$$\underset{||\boldsymbol{x}||_2=1}{\operatorname{argmin}} \quad ||\boldsymbol{x}||_1^2 \tag{2}$$

Case 2:

$$\underset{||\boldsymbol{x}||_2=1}{\operatorname{argmax}} \quad ||\boldsymbol{x}||_4^4 \tag{3}$$

Note that raising them to even powers does not change the solution. Notice for ℓ^1-norm, the problem is a minimization. But for ℓ^4-norm, it is a maximization. Miraculously, their solutions converge, and both corresponds to the best approximation to (1).

2 Prerequisites

For clarification, we say $\boldsymbol{x}$ is sparser than $\boldsymbol{y}$ if there are more 0 entries in $\boldsymbol{x}$ than in $\boldsymbol{y}$.

To make the proof make sense, let us first show a simple lemma:

Lemma 1. *For any $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$, the following expression is true.*

$$(\sum_{i=1}^{n} x_i)^2 = \sum_{i=1}^{n} x_i^2 + 2\sum_{i<j}^{n} x_i x_j \tag{4}$$

Proof. It is easy to see once we express the summation in an specific tabular way:

$$\begin{aligned}
(\sum_{i=1}^{n} x_i)^2 &= (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \\
&= x_1x_1 + x_1x_2 + \cdots + x_1x_n \\
&+ x_2x_1 + x_2x_2 + \cdots + x_2x_n \\
&+ \cdots \\
&+ x_nx_1 + x_2x_2 + \cdots + x_nx_n \\
&= \sum_{i=1}^{n} x_i^2 + 2\sum_{i<j}^{n} x_i x_j
\end{aligned}$$

by breaking the expression into the sum of diagonals and off-diagonals. □

Moreover, we can actually raise all x_i to a certain power, and generalize the expression to the following:

$$(\sum_{i=1}^{n} x_i^k)^2 = \sum_{i=1}^{n}((x_i)^k)^2 + 2\sum_{i<j}^{n}(x_i)^k(x_j)^k \tag{5}$$

Another lemma we need to know, is the following about square roots:

Lemma 2. *For any $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$, the following inequality is true:*

$$\sum_{i=1}^{n}\sqrt{(x_i)^2} \geq \sqrt{\sum_{i=1}^{n}(x_i)^2} \tag{6}$$

The proof is fairly simple, so we will leave this to the reader.

3 Case 1: ℓ^1-norm Minimization

3.1 Algebraic Explanation

To find the bound for $||\boldsymbol{x}||_1$, we expand the terms for $||\boldsymbol{x}||_1^2$, then we get:

$$
\begin{aligned}
||\boldsymbol{x}||_1^2 &= \sum_1^n |x_i| \\
&= \sum_1^n \sqrt{(x_i)^2} \\
&\geq \sqrt{\sum_1^n (x_i)^2} && \text{(by (6))} \\
&= ||x||_2 \\
&= 1 && \text{(by } S^n \text{ constraint)}
\end{aligned}
$$

Hence, we have shown that 1 is the lower bound for $||\boldsymbol{x}||_1$.

If we expand the expression using (1), we get:

$$
\begin{aligned}
||\boldsymbol{x}||_1^2 &= (\sum_{i=1}^n |x_i|)^2 \\
&= \sum_{i=1}^n |x_i|^2 + 2\sum_{i=1}^n |x_i||x_j|
\end{aligned}
$$

One can see that if we set $\boldsymbol{x}$ as any vector that is not a canonical vector, then $||\boldsymbol{x}||_1^2$ will be larger than 1.

3.2 Geometric Explanation

One can visualize the relationship of ℓ^1-norms by considering different vectors on S^1. For example, let's turn to figure 1. Assume we are living in $\mathbb{R}^2$. The geometry of ℓ^1-norm is a diamond. The red arrow is the vector $(1, 0)$, which is one of the solution to our optimization problem. And the green arrow is the vector $(\sqrt{2}, \sqrt{2})$. We can see that both red and green arrow lives in S^1, but when we graph out the relative ℓ^1-norm diamond, we can see that the green diamond is bigger than the red diamond. Moreover, this generalizes to any vector that lives in S^1.

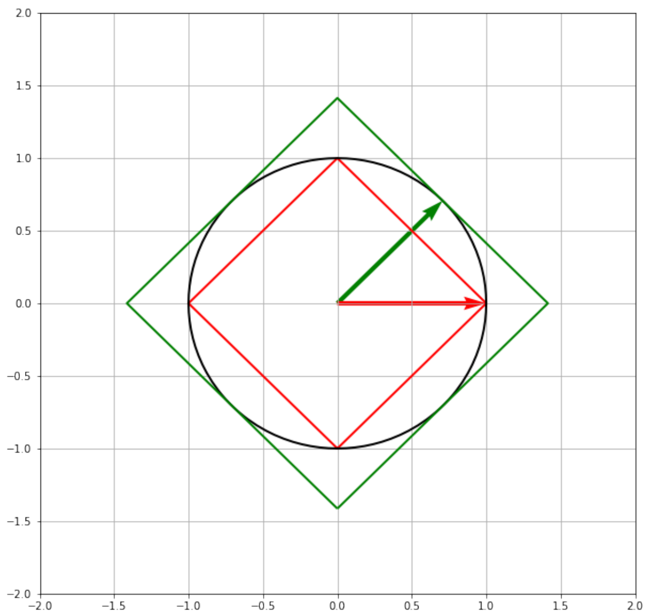

Figure 1: ℓ^1-norm minimization visualization

4 Case 2: ℓ^4-norm Maximization

As we will see later, the proofs are nearly identical as the ℓ^1 case.

4.1 Algebraic Explanation

To find the bound for $||\boldsymbol{x}||_1$, we expand the terms for $||\boldsymbol{x}||_1^2$, then we get:

$$\begin{aligned} ||\boldsymbol{x}||_4^4 &= \sum_{i=1}^{n} |x_i|^4 && (7)\\ &\leq \sum_{i=1}^{n} ((x_i)^2)^2 + 2\sum_{i<j}^{n} x_i^2 x_j^2 && (8)\\ &= (\sum_{i=1}^{n} (x_i)^2)^2 && (9)\\ &= 1 && (10) \end{aligned}$$

Although this seem less intuitive, we can observe the properties of the expressions we derived. Here, the $||\boldsymbol{x}_4||^4$ is now upper bounded by 1. Moreover, we show equation (8) = 1, and $2\sum_{i<j}^{n} x_i^2 x_j^2$ is minimum if and only if

$\sum_{i=1}^{n}((x_i)^2)^2$ is maximum. Since $2\sum_{i<j}^{n} x_i^2 x_j^2$ is at minimum if it has 0's, this implies that $||\boldsymbol{x}||_4^4$ is maximum when $\boldsymbol{x}$ is a canonical vector.

4.2 Geometric Explanation

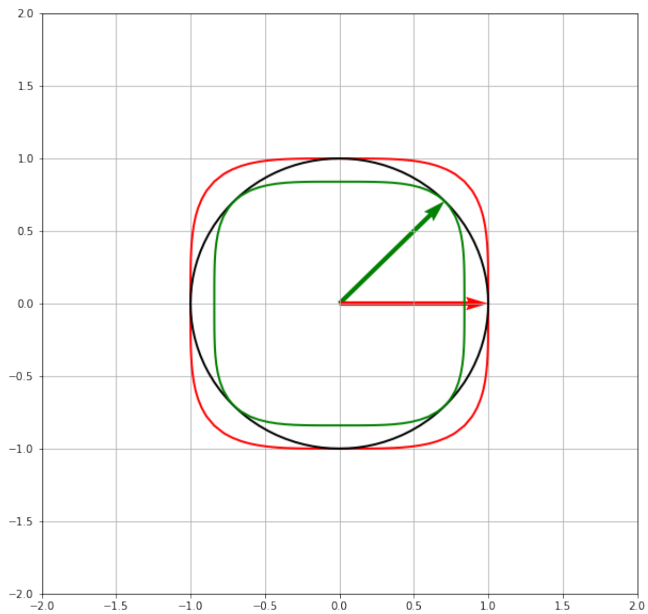

Figure 2: ℓ^4-norm maximization visualization

Similar to our previous visualization for the ℓ^1-norm, the red vector here is the canonical vector, and the green vector is the diagonal. The shape of ℓ^4-ball is a rounded square. One can see that if we choose a vector that is not a canonical vector, like the green vector, the green rounded square is always smaller than the red rounded square.

5 Conclusion

What we presented here might seem like very simple concepts. ℓ^p-norm is an operation used all over optimization, linear algebra, metric differential geometry, etc. It is important to be to see the same topic from all kinds of perspective. I hope this short write-up reveals something interesting to you, as this is also a reminder to me just in case if I forget about this in the future lol.