However, to show that the l^2-norm is indeed a norm, we must use the following famous inequality.

Theorem. *(Cauchy-Schwarz Inequality) For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have*

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i \leq \left\{\sum_{i=1}^{n} x_i\right\}^{1/2} \left\{\sum_{i=1}^{n} y_i\right\}^{1/2} = \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 .$$

Proof. We have nothing to prove if $\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{y} = \boldsymbol{0}$. So, we suppose both vectors are nonzero.

Consider the scaled version of the two vectors, $\boldsymbol{u} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}$ and $\boldsymbol{v} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2}$. The reason why we consider these is because the original statement is equivalent to show that

$$\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2} \leq 1.$$

So, it suffices to show that $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} \leq 1$.

Then, knowing that the dot product of a vector to itself is always nonnegative (yields the 2-norm squared, in fact), we have

$$\begin{aligned} 0 &\leq (\boldsymbol{u} - \boldsymbol{v})^{\mathrm{T}} (\boldsymbol{u} - \boldsymbol{v}) \\ &= \|\boldsymbol{u}\|_2^2 - 2\boldsymbol{u}^T\boldsymbol{v} + \|\boldsymbol{v}\|_2^2 \\ &= 2\left(1 - \boldsymbol{u}^T\boldsymbol{v}\right) \end{aligned}$$

This implies

$$1 - \boldsymbol{u}^T\boldsymbol{v} \geq 0 \implies \boldsymbol{u}^T\boldsymbol{v} \leq 1$$

Substituting $\boldsymbol{x}$ and $\boldsymbol{y}$ back, we have

$$\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right)^{\mathrm{T}} \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2} \leq 1$$

and thus

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 .$$

□

With Cauchy-Schwarz, we can now prove that $\|\cdot\|_2$ is indeed a norm. The only hard part is the triangle-inequality. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\begin{aligned} \|\boldsymbol{x} + \boldsymbol{y}\|_2^2 &= (\boldsymbol{x} + \boldsymbol{y})^{\mathrm{T}} (\boldsymbol{x} + \boldsymbol{y}) = \sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 \\ &\overset{\text{Cauchy-Schwarz}}{\leq} \|\boldsymbol{x}\|_2^2 + 2\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 + \|\boldsymbol{y}\|_2^2 = \left(\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2\right)^2 . \end{aligned}$$

Taking a square root of both sides, we obtain

$$\|\boldsymbol{x} + \boldsymbol{y}\|_2 \leq \|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2 .$$

Now, knowing that $\|\cdot\|_2$ is a norm, we can go on measuring the difference between two vectors in $\mathbb{R}^n$.

Definition. If $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ amd $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$ are vectors in $\mathbb{R}^n$, and the l^2 and l^∞ distances between $\boldsymbol{x}$ and $\boldsymbol{y}$ are defined by

$$\|\boldsymbol{x} - \boldsymbol{y}\|_2 = \left(\sum_{i=1}^{n} (x_i - y_i)^2\right)^{1/2}$$

and

$$\|\boldsymbol{x} - \boldsymbol{y}\|_\infty = \max_{1 \le i \le n} |x_i - y_i| .$$

Example. Suppose an approximate solution to a linear system, using 5-digit rounding, is

$$\tilde{\boldsymbol{x}} = (1.2001, 0.99991, 0.92538)^{\mathrm{T}}$$

and the true solution is

$$\boldsymbol{x} = (1, 1, 1)^{\mathrm{T}} .$$

Let's find the difference between them under the two settings of norm.

$$\begin{aligned}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_\infty &= \max\{|1 - 1.2001|, |1 - 0.99991|, |1 - 0.92538|\} \\ &= \max\{0.2001, 0.00009, 0.07462\} \\ &= 0.2001,\end{aligned}$$

and

$$\begin{aligned}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2 &= \left[(1 - 1.2001)^2 + (1 - 0.99991)^2 + (1 - 0.92538)^2\right]^{1/2} \\ &= 0.21356.\end{aligned}$$

From this, we see that, even though $\widetilde{x_2}$ and $\widetilde{x_3}$ are good approximations, $\widetilde{x_1}$ really drags down the error.

The concept of distance in $\mathbb{R}^n$ is also used to define a limit of sequence of vectors in this space.

Definition. A sequence $\left\{\boldsymbol{x}^{(k)}\right\}_{k=1}^{\infty}$ of vectors in $\mathbb{R}^n$ is said to **converge** to $\boldsymbol{x}$ with respect to the norm $\|\cdot\|$ if, given any $\epsilon > 0$, there exists an integer $N(\epsilon)$ such that

$$\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\right\| < \epsilon, \quad \text{for all } k \ge N(\epsilon) .$$

Remark. One should think this sequence getting very close to $\boldsymbol{x}$ past certain index,

$$\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(N)}, \boldsymbol{x}^{(N+1)}, \ldots$$

The statement is saying, given your tolerance level ϵ between the approximant $\boldsymbol{x}^{(k)}$ and the target $\boldsymbol{x}$, one can always find an exact index $N(\epsilon)$ such that this distance between the approximant and the target is within ϵ. A convergent sequence has this property.

On the other hand, a divergent sequence does not have this property. We state the negation of convergence. We say $\boldsymbol{x}^{(k)} \in \mathbb{R}^n$ diverges with respect to the norm $\|\cdot\|$ if for every $\boldsymbol{x} \in \mathbb{R}^n$, there exists $\epsilon > 0$ such that for every integer N, there exists an index $k \ge N$ such that

$$\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\right\| \ge \epsilon.$$

You may consider the simple sequence $\boldsymbol{x}^{(k)} = (k, 0) \in \mathbb{R}^2$. Eventually, k gets so big, say, to N, such that for every $\boldsymbol{x} \in \mathbb{R}^2$, one can look just a few more terms down the sequence to find $\left\|\boldsymbol{x}^{(N+j)} - \boldsymbol{x}\right\| \ge \epsilon$ where you are free to set ϵ. You can always find something larger to beat whatever tolerance you set, meaning that this sequence is indeed growing out of control.