MATH3332 Data Analytic Tools

Ye Moe

HKUST Fall 2022

Introduction

The purpose of this course is to introduce some crucial mathematical analysis tools for data analysis/machine learning.

According to *Pedro Domingos*,

$$Learning = Representation + Evaluation + Optimization$$

1. Representation
    - How do we represent a learner? Which set should a learner be in? This set is called the hypothesis space of the learner. Some related tools are "space of functions".
    - How do we represent the input? Potential tools include vectors, graphs, manifolds, ...
2. Evaluation
    - How to pick the best learner from the hypothesis space? Needs calculus of "functions of functions" also known as functionals.
    - How to represent the input effectively? Needs Linear Algebra, Graph Theory, Manifolds Calculus, Harmonic Analysis, ...
3. Optimization
    - Numerical optimization solver - how to get the optimal solution numerically by a computer? Many of the resulting optimization is convex optimization and it is related to Convex Analysis.

So this course consists of some

- Basic functional analysis (calculus of functionals)
- Basic convex analysis
- Fourier analysis and Wavelet analysis (if time allowed)

Normed and Inner Product Space

2.1 Vector Spaces

Definition: A vector space over $\mathbb{R}$ is a set $\mathbb{V}$ together with two functions.

1. Vector addition: $+ : (\mathbb{V}, \mathbb{V}) \to \mathbb{V}$
   i.e. $\forall x, y \in \mathbb{V}, x + y \in \mathbb{V}$

2. Scalar multiplication: $. : (\mathbb{R}, \mathbb{V}) \to \mathbb{V}$
   i.e. $\forall \alpha \in \mathbb{R}, x \in \mathbb{V}, \alpha x \in \mathbb{V}$

These two functions should satisfy the following eight properties:

1. Associativity of addition: $x + (y + z) = (x + y) + z, \forall x, y, z \in \mathbb{V}$

2. Commutativity of addition: $x + y = y + x, \forall x, y \in \mathbb{V}$

3. Zero vector: $\exists$ an element, denoted by 0 in $\mathbb{V}$ s.t. $x + 0 = 0 + x = x, \forall x \in \mathbb{V}$

4. Negative vector: $\forall x \in \mathbb{V}, \exists$ an elements, denoted by $-x \in \mathbb{V}$ s.t. $x + (-x) = (-x) + x = 0$

5. $\forall x \in \mathbb{V}, 1 \cdot x = x$

6. $\forall x \in \mathbb{V}, \alpha, \beta \in \mathbb{R}, \alpha(\beta x) = (\alpha\beta)x$

7. $\forall x \in \mathbb{V}$ and $\alpha, \beta \in \mathbb{R}, (\alpha + \beta)x = \alpha x + \beta x$

8. $\forall x, y \in \mathbb{V}, \alpha(x + y) = \alpha x + \alpha y$

Remarks: We can define vector space over the complex domain $\mathbb{C}$, but since vector space over complex domain $\mathbb{C}$ is used very rarely, we will only consider vector space in the real domain $\mathbb{R}$.

Some examples of vector space include $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{m \times n}$, $\mathbb{R}^{m \times n \times l}$, $C[a, b]$ and L_∞.

Example: Prove that $\mathbb{R}^n$ is a vector space.
$\forall x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$,

$$x + y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \in \mathbb{R}^n$$

$$\alpha x = \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix} \in \mathbb{R}^n$$

Since it is closed under both vector addition and scalar multiplication, $\mathbb{R}^n$ is a vector space.

Example: Prove that $C[a, b]$ is a vector space.
$\forall f, g \in C[a, b]$ and $\alpha \in \mathbb{R}$,

$$f(t) + g(t) = (f + g)(t) \in C[a, b], \forall t \in [a, b]$$
$$\alpha f(t) = (\alpha f)(t) \in C[a, b], \forall t \in [a, b]$$

Since it is closed under both vector addition and scalar multiplication, $C[a, b]$ is a vector space.

Remarks: $C[a, b]$ is referred to as a function space, since any vector in this vector space is a function. It might be a hypothesis space of a learner with one input and one output, i.e. Find a $f \in C[a, b]$ s.t. $f(x_i) \approx f(y_i)$ for all i.

Example: Prove that L_∞ is a vector space.

$$L_\infty = \{ \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix} \mid \exists \text{ a finite number } c \text{ s.t.} \mid a_i \mid \leq c \text{ for any i} \}$$

$\forall a, b \in L_\infty$ and $\alpha \in \mathbb{R}$,

$$a + b = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \end{bmatrix} \in L_\infty$$

$$\alpha a = \alpha \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \end{bmatrix} \in L_\infty$$

Since it is closed under both vector addition and scalar multiplication, L_∞ is a vector space.

Remarks: This vector space can be used to model stock prices with a very fine time resolution.

Example: Consider the set of all strings.

$$\text{I' + 'am'} \neq \text{'am' + 'I'}$$

The set of all strings violates the commutative properties of a vector space, therefore it isn't a vector space. Hence, we cannot use vector space to model text data in this naïve way.

How do we "vectorize" the text data?
This is a fundamental question in text data analysis.

2.2 Normed and Banach Space

In order to do calculus on vector spaces, we need to define 'distance/closeness' between vectors.

Let V be a vector space. Let x, y $\in \mathbb{V}. Then,$

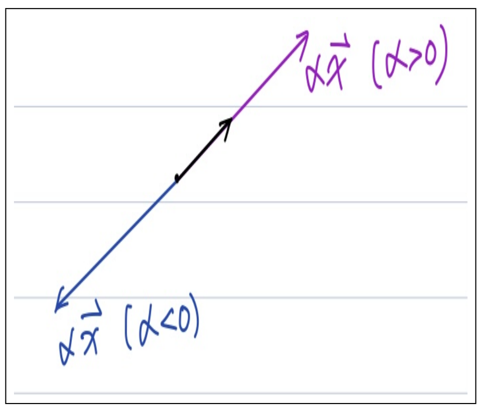
$$distance(x, y) = distance(x - y, y - y) = distance(x - y, 0) = \text{length of x - y}$$

Remarks: Distance should be shift invariant.
To define distance, we only need to define the length of vectors. Let x $\in \mathbb{V}$.
Denote $\|x\|$ be the length of x. Then $\|x\|$ should satisfy:

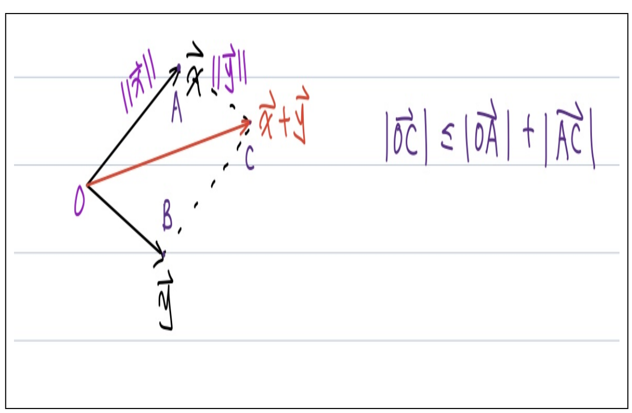
1. $\|x\| \geq 0$ (the length should be non-negative)
   Moreover, $\|x\| = 0 \iff x = 0$(only zero vector has a zero length)

2. $\|\alpha x\| = | \alpha | \|x\|$

(length of a scaling of a vector is a scaling of the length of the vector)



3. $\|x + y\| \leq \|x\| + \|y\|$ (also known as triangle inequality)
   (length of direct path should be smaller than the length of indirect path)



Definition: Let $\mathbb{V}$ be a vector space. A norm on V is a function $\|\cdot\|:\mathbb{V}\to\mathbb{R}$ such that:

1. $\|x\|\geq 0\ \forall x\in\mathbb{V}$ and $\|x\|=0 \iff x=0$
2. $\|\alpha x\| = \mid\alpha\mid \|x\|\ \forall\alpha\in\mathbb{R}, x\in\mathbb{V}$
3. $\|x+y\|\leq\|x\|+\|y\|\ \forall x,y\in\mathbb{V}$

Example: $\mathbb{R}$ is a vector space over $\mathbb{R}$.
Let $\|x\| = \mid x\mid\ \forall x\in\mathbb{R}$. Then it is a norm on $\mathbb{R}$.

Example: $\mathbb{R}^n$ is a vector space over $\mathbb{R}$.
There are many norms on $\mathbb{R}^n$.

- 2-norm: (Euclidean Norm)
  $\|x\|_2 = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$

  Question: Prove that $\|\cdot\|_2$ is indeed a norm for $\mathbb{R}^n$.
  $\forall x,y\in\mathbb{R}^n$ and $\alpha\in\mathbb{R}$,

$$\|x\|_2 = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}} \geq 0$$

$$\|x\|_2 = 0 \iff \sum_{i=1}^n x_i^2 = 0 \iff x_i^2 = 0,\ i=1,...,n$$

$$\iff x_i = 0,\ i=1,...,n \iff x=0$$

$$\|\alpha x\|_2 = (\sum_{i=1}^n(\alpha x_i)^2)^{\frac{1}{2}} = (\alpha^2\sum_{i=1}^n x_i^2)^{\frac{1}{2}} = \mid\alpha\mid(\sum_{i=1}^n x_i^2)^{\frac{1}{2}} = \mid\alpha\mid \|x\|_2$$

$$\|x+y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 + 2 < x,y >$$

$$\leq \|x\|_2^2 + \|y\|_2^2 + 2\|x\|_2\|y\|_2 \text{ (By Cauchy-Schwartz inequality)}$$

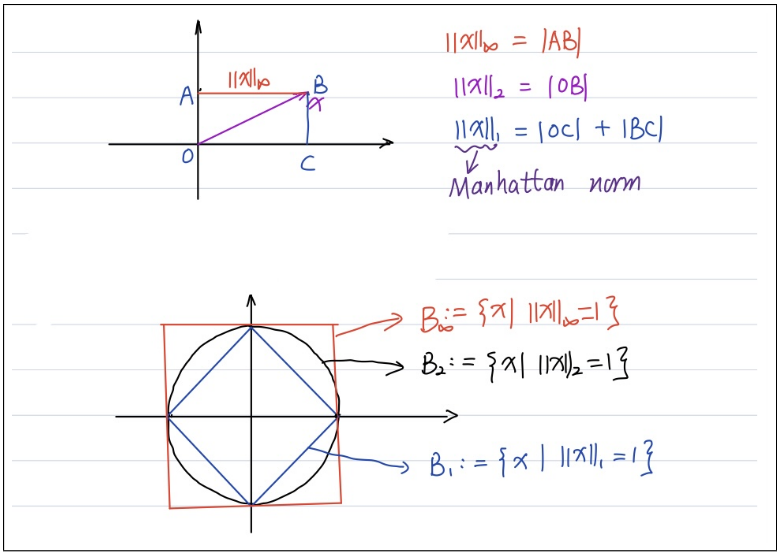
$$= (\|x\|_2 + \|y\|_2)^2$$

$$\|x+y\|_2 \leq \|x\|_2 + \|y\|_2$$

- 1-norm:
  $\|x\|_1 = \sum_{i=1}^n \mid x_i\mid$
- ∞-norm:
  $\|x\|_\infty = \max_{1\leq i\leq n}\mid x_i\mid$
- p-norm:
  $\|x\|_p = (\sum_{i=1}^n \mid x_i\mid^p)^{\frac{1}{p}}$

Fact: $\|x\|_p$ is a norm on $\mathbb{R}^n \iff$ p$\geq$ 1.

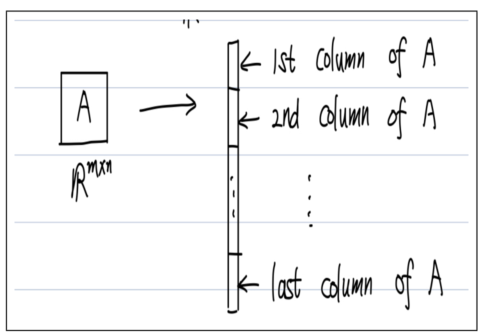
Geometric definition of different norms in $\mathbb{R}^n$



Note that $(\mathbb{R}^n, \|\cdot\|_1), (\mathbb{R}^n, \|\cdot\|_2), (\mathbb{R}^n, \|\cdot\|_\infty), \ldots$ are all different normed spaces. So for a given vector space, we can obtain various normed space by choosing different norms. Also, $\|x\|_p \leq \|x\|_q$ if $p \geq q$.

Example: $\mathbb{R}^{m\times n}$ is a vector space over $\mathbb{R}$.

1. $\mathbb{R}^{m\times n}$ can be viewed as $\mathbb{R}^{mn}$.



We can define vector p-norm for $\mathbb{R}^{m\times n}$.

- p = 1
  $\|A\|_{1,vec} = \sum_{i=1}^m \sum_{j=1}^n \mid a_{ij} \mid$

- p = 2
  $\|A\|_{2,vec} = (\sum_{i=1}^{m}\sum_{j=1}^{n} \mid a_{ij} \mid^2)^{\frac{1}{2}}$

  This norm is widely known as the Frobenius norm denoted as $\|A\|_F$.
- p = ∞
  $\|A\|_{\infty,vec} = \max_{i=1,\ldots,m} \max_{j=1,\ldots,n} \mid a_{ij} \mid$

2. $\mathbb{R}^{m\times n}$ can be viewed as linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$.
   We can define matrix p-norm for $\mathbb{R}^{m\times n}$.

$$\|A\|_p = \max_{x\neq 0, x\in\mathbb{R}^n} \frac{\|Ax\|_p}{\|x\|_p} = max_{\|x\|_p=1}\|Ax\|_p$$

   - p = 1
     $\|A\|_1 = max_{1\leq j\leq n}\sum_{i=1}^{n} \mid a_{ij} \mid$ = maximum absolute column sum
   - p = ∞
     $\|A\|_\infty = max_{1\leq i\leq n}\sum_{j=1}^{n} \mid a_{ij} \mid$ = maximum absolute row sum
   - p = 2
     $\|A\|_2$ = maximum singular value of A

3. We can also define other matrix norms.

   (a) We can use different norms in $\mathbb{R}^n$ and $\mathbb{R}^m$.

$$\|A\|_{p\to q} = \max_{\|x\|_p=1} \|Ax\|_q$$

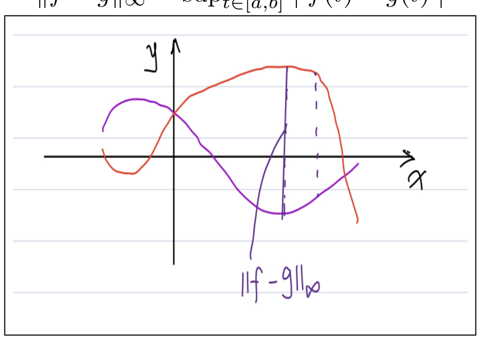
   (b) The nuclear norm $\|\cdot\|_*$

Example: $C[a,b]$ is a vector space over $\mathbb{R}$.
$\forall f \in C[a,b]$, define

$$\|f\|_\infty = \sup_{t\in[a,b]} \mid f(t) \mid$$

We can check that $\|\cdot\|_\infty$ is indeed a norm on $C[a,b]$.
The distance of two function $f, g \in C[a,b]$ is given by

$$\|f-g\|_\infty = \sup_{t\in[a,b]} \mid f(t) - g(t) \mid$$



Some other norms on $C[a,b]$.

1. $\|f\|_1 = \int_b^a \mid f(t) \mid dt$
2. $\|f\|_2 = (\int_b^a \mid f(t) \mid^2 dt)^{\frac{1}{2}}$
3. $\|f\|_p = (\int_b^a \mid f(t) \mid^p dt)^{\frac{1}{p}}$

Example: $L_\infty = \{a | a$ is a infinite sequence and $\exists\ c > 0$ s.t. $\mid a_i \mid \leq c, \forall i\}$

1. $\forall a \in L_\infty$, define

$$\|a\|_\infty = \sup_i \mid a_i \mid$$

   Remarks: You cannot replace sup here with max.

2. Define $\|a\|_p = (\sum_{i=1}^\infty \mid a_i \mid^p)^{\frac{1}{p}}\ \forall a \in L_\infty$ but this is not a norm on L_∞.

   e.g. $a = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \vdots \\ \frac{1}{i} \\ \vdots \end{bmatrix} \in L_\infty,$ but $\|a\|_1 = \sum_{i=1}^\infty \mid a_i \mid = \sum_{i=1}^\infty \frac{1}{i} = \infty$

   So, $\|\cdot\|_1$ is not a norm on L_∞.
   Instead, we consider

$$L_p = \{a \in L_\infty | \|a\|_p < \infty\} \subset L_\infty$$
$$\|\cdot\|_p \text{ is a norm on } L_p.$$

   e.g. $a = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \vdots \\ \frac{1}{i} \\ \vdots \end{bmatrix} \in L_\infty$

   $\|a\|_\infty = 1$, $\|a\|_2 = (\sum_{i=1}^\infty \frac{1}{i^2})^{\frac{1}{2}} = (\frac{\pi^2}{6})^{\frac{1}{2}} = \frac{\pi}{\sqrt{6}}, \|a\|_1 = \infty$
   So, $a \in L_\infty$, $a \in L_2$ but $a \notin L_1$. Indeed, $a \in L_p\ \forall p > 1$.

Limit and Convergence on Normed Vector Space
To define calculus, we first need to define convergent sequence.
Let $\mathbb{V}$ be a normed vector space. Let $\{x^{(k)}\}_{k\in\mathbb{N}}$ be a sequence in $\mathbb{V}$, Let $x \in \mathbb{V}$. We say $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to x, denoted by $x^{(k)} \to x$, if

$$\lim_{k\to\infty} \|x^{(k)} - x\| = 0$$
$$\lim_{k\to\infty} \|x^{(k)} - x\| = 0 \iff x^{(k)} \to x$$

Example: Consider $\mathbb{R}^n$ with $\|\cdot\|_2$,

$$\text{Let } x^{(k)} = \begin{bmatrix} \frac{1}{k} \\ \frac{2}{k} \\ \vdots \\ \frac{n}{k} \end{bmatrix} \in \mathbb{R}^n \text{ and } x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n$$

$$\|x^{(k)} - x\|_2 = \|x^{(k)}\|_2 = (\textstyle\sum_{i=1}^n (\frac{i}{k})^2)^{\frac{1}{2}} = \frac{1}{k}(\sum_{i=1}^n i^2)^{\frac{1}{2}}$$
$$\lim_{k\to\infty} \|x^{(k)} - x\|_2 = \lim_{k\to\infty} \frac{1}{k}(\textstyle\sum_{i=1}^n i^2)^{\frac{1}{2}} = 0$$
$$x^{(k)} \to x$$

Unfortunately, the limit of a sequence may not always be in the same vector space as the original sequence. If this happen, we call this the normed vector space incomplete. Otherwise, it is a complete vector space also known as the Banach space.

Example of Banach space:

1. $\mathbb{R}^n$ with any norm
2. $\mathbb{R}^{m\times n}$ with any norm
3. Tensor space $\mathbb{R}^{m\times n\times l}$ with any norm
4. $C[a, b]$ with $\|\cdot\|_\infty$
5. L_p with p-norm, for $p \geq 1$ and $p = \infty$.

Cauchy Sequence
Definition: $\{x^{(k)}\}$ is a Cauchy sequence, if for any $\epsilon > 0$, there exists K such that for any $k, l > K$, $\|x^{(k)} - x^{(l)}\| < \epsilon$.
Facts:

1. If $x^{(k)} \to x$ in $(\mathbb{V}, \|\cdot\|)$, then $\{x^{(k)}\}$, then $\{x^{(k)}\}$ must also be a Cauchy sequence.
   Proof.
   $x^{(k)} \to x$ implies that $\forall\epsilon > 0$, $\exists k$, s.t. $k > K$ $\|x^{(k)} - x\| \leq \frac{\epsilon}{2}$. Therefore, $\|x^{(k)} - x^{(l)}\| \leq \|x^{(k)} - x\| + \|x^{(l)} - x\| \leq \epsilon$, $\forall k, l > K$

2. The reverse is **NOT** necessarily true.

Definition: A vector space $(\mathbb{V}, \|\cdot\|)$ is complete if the limit of all Cauchy sequences in $\mathbb{V}$ is in $\mathbb{V}$.

Remarks: We can always complete an incomplete normed vector space by including all limits of its Cauchy sequence.

Finite Dimensional Vector Space
In most cases, we are dealing with finite dimensional vector space such as $\mathbb{R}^n$, $\mathbb{R}^{m\times n}$ and $\mathbb{R}^{m\times n\times l}$.

Properties related to Finite Dimensional Vector Space:

- Any finite dimensional vector space with any norm is complete. That is, any finite dimensional vector space is Banach space.
- For a finite dimensional vector space $\mathbb{V}$, all norms are equivalent.
  Theorem: For any norms $\|\cdot\|_A$ and $\|\cdot\|_B$, $\exists c_1, c_2 > 0$ s.t.
  $c_1\|a\|_A \le \|a\|_B \le c_2\|a\|_A$, $\forall a \in \mathbb{V}$ (finite dimensional)

Example: Prove that $x^{(k)} \to x$ in $\|\cdot\|_A \iff x^{(k)} \to x$ in $\|\cdot\|_B$.
Since $x^{(k)} \to x$ in $\|\cdot\|_A$,

$$\lim_{k\to\infty} \|x^{(k)} - x\|_A = 0$$

Because of equivalence,

$$c_1\|x^{(k)} - x\|_A \le \|x^{(k)} - x\|_B \le c_2\|x^{(k)} - x\|_A$$

$$0 \le \lim_{k\to\infty} \|x^{(k)} - x\|_B \le c_2 \lim_{k\to\infty} \|x^{(k)} - x\|_A = 0$$

$$\lim_{k\to\infty} \|x^{(k)} - x\|_B = 0 \text{ (by squeeze theorem)}$$

$$x^{(k)} \to x \text{ under } \|\cdot\|_B$$

Similarly for the $\leftarrow$ direction.

Example: Consider $\mathbb{R}^n$ and $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$.

- $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent.

$$\|a\|_2 \le \|a\|_1 \le \sqrt{n}\|a\|_2, \forall a \in \mathbb{R}^n$$

- $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are equivalent.

$$\|a\|_\infty \le \|a\|_2 \le \sqrt{n}\|a\|_\infty, \forall a \in \mathbb{R}^n$$

- $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are equivalent.

$$\|a\|_\infty \le \|a\|_1 \le n\|a\|_\infty, \forall a \in \mathbb{R}^n$$

Remarks: Though they are equivalent, the speed at which they converge are different. In other words, the convergence speed depends on norms.

2.3 Inner Product and Hilbert Space

Question: How do we describe the correlation/alignment between two vectors? Norms are not able to describe it as they are 'scaling sensitive'.

A good answer would be to use angle. A good candidate would be to use inner product since it is 'scaling insensitive'.

Inner Product
Definition: A function $< \cdot, \cdot >: (\mathbb{V}, \mathbb{V}) \to \mathbb{R}$ on a vector space $\mathbb{V}$ is called an inner product over R, if:

1. $\forall x \in \mathbb{V}, < x, x > \geq 0 \; and < x, x > = \; 0 \iff x = 0$
2. $< \alpha x_1 + \beta x_2, y >= \alpha < x_1, y > +\beta < x_2, y >$, $\forall \alpha, \beta \in \mathbb{R}, x_1, x_2, y \in \mathbb{V}$
3. $< x, y >=< y, x >, \forall x, y \in \mathbb{V}$

Remarks:

1. By 2 and 3, $< x, \alpha y_1 + \beta y_2 >= \alpha < x, y_1 > +\beta < x, y_2 >$, $\forall \alpha\beta \in \mathbb{R}$, $x_1, y_1, y_2 \in \mathbb{V}$. Therefore, $< \cdot, \cdot >$ is a bi-linear function, i.e., it is linear with respect to one of the variable with the other fixed.
2. For inner product of vector spaces on $\mathbb{C}$, we only need to change 3 to $< x, y >= \overline{< y, x >}$, where $\bar{\cdot}$ stands for complex conjugate.

Example: $\mathbb{R}^n$ is a vector space. We can define an inner product as

$$< x, y >= \sum_{i=1}^{n} x_i y_i = x^T y, \; \forall x, y \in \mathbb{R}^n.$$

Example: Another inner product in $\mathbb{R}^n$ is as follows. We can define a "weighted" inner product as $< x, y >_A = x^T A y$, where $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix.
Remarks: A is SPD $\iff A = A^T$ and $x^T A x > 0 \; \forall x \in \mathbb{R}^n$ and $x \neq 0$.

Example: $\mathbb{R}^{m \times n}$ is a vector space. We can define an inner product as

$$< A, B >= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}, \; \forall A, B \in \mathbb{R}^{m \times n}$$

Similarly, these are equal to $trace(A^T B)$, $trace(B^T A)$, $trace(AB^T)$ and $trace(BA^T)$, where $trace(A)$ is defined as the sum of the diagonal of matrix A.

Example: In L_2, we can define an inner product as

$$< a, b >= \sum_{i=1}^{\infty} a_i b_i, \; \forall a, b \in L_2$$

Example: In $C[a, b]$, we can define an inner product as

$$< f, g >= \int_a^b f(t) g(t) dt, \; \forall f, g \in C[a, b]$$

Cauchy-Schwartz Inequality
If $< \cdot, \cdot >$ is an inner product on $\mathbb{V}$, then, for any $x, y \in \mathbb{V}$,

$$| < x, y > |^2 \leq < x, x >< y, y >$$

The equality holds true if and only if $x = \alpha y$ or $y = \alpha x$ for some $\alpha \in \mathbb{R}$

Proof.
Let $\lambda \in \mathbb{R}$ be an arbitrary number,

$$0 \ \leq < x + \lambda y, x + \lambda y >$$

$$= < x, x > + \ \lambda < y, x > + \lambda < x, y > + \lambda^2 < y, y >$$

$$= < x, x > + \ 2 \ \lambda < x, y > + < x, x >$$

$$\text{Thus, } \lambda^2 < y, y > + 2\lambda < x, y > + < x, x > \geq \ 0, \forall \lambda \in \mathbb{R}$$

The left is a quadratic function of λ and is always non-negative. There is at most one root of the quadratic function, hence, the determinant $b^2 - 4ac \leq 0$.

$$\text{So, } (2 < x, y >)^2 - 4 < x, x >< y, y > \leq 0$$

$$\Longrightarrow < x, y >^2 \leq < x, x >< y, y >$$

Finally, when $< x, y >^2 = < x, x >< y, y >$, there is a root, i.e., $\exists$ a unique $\lambda \in \mathbb{R}$, $\lambda^2 < y, y > + 2\lambda < x, y > + < x, x > = 0$.

$$\Longleftrightarrow$$

$$\exists \text{ a unique } \lambda \in \mathbb{R}, < x + \lambda y, x + \lambda y > = 0.$$

$$\Longleftrightarrow$$

$$\text{a unique } \lambda \in \mathbb{R}, \ x + \lambda y = 0.$$

$$\Longleftrightarrow$$

$$\exists \text{ a unique } \lambda \in \mathbb{R}, \ x = -\lambda y.$$

With the Cauchy-Schwartz inequality, we can show that

$$\|x\| = (< x, x >)^{\frac{1}{2}} \text{ defines a norm.}$$

This is also called "norm induced by the inner product". This one above is for $\mathbb{R}^n$.

Proof.

$$\|x\| = (< x, x >)^{\frac{1}{2}} \geq 0 \text{ and } \|x\| = (< x, y >)^{\frac{1}{2}} = 0 \iff x = 0$$

$$\|\alpha x\| = (< \alpha x, \alpha x >)^{\frac{1}{2}} = (\alpha^2 < x, x >)^{\frac{1}{2}} = \mid \alpha \mid \|x\|$$

$$\begin{gathered}
\|x + y\|^2 = < x + y, x + y > \\
= < x, x > + < x, y > + < y, x > + < y, y > \\
= \|x\|^2 + \|y\|^2 + 2 < x, y > \\
\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \\
= (\|x\| + \|y\|)^2 \\
\|x + y\| \leq \|x\| + \|y\|
\end{gathered}$$

Remarks: In the proof above, we have used an alternative version of the Cauchy-Schwartz inequality.

$$\mid < x, y > \mid \leq \ \|x\|\|y\|$$

All kinds of induced norm

1. $\mathbb{R}^n$ with inner product $< \cdot, \cdot >$: $< x, y >= x^T y$
   The induced norm is
   $\|x\| = (< x, x >)^{\frac{1}{2}} = (x^T x)^{\frac{1}{2}} = (\sum_{i=1}^n x_i^2)^{\frac{1}{2}} = \|x\|_2$

2. $\mathbb{R}^n$ with weighted inner product $< \cdot, \cdot >_A$: $< x, y >_A = x^T Ay$
   The induced norm is
   $\|x\|_A = (x^T Ax)^{\frac{1}{2}} = (\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i xj)$

3. The p-norm of $\mathbb{R}^n$
   $\|x\|_p = (\sum_{i=1}^n \mid x_i \mid^p)^{\frac{1}{p}}$
   When $p = 2$, $\| \cdot \|_2$ is induced by $< \cdot, \cdot >$. It is not induced by inner product for all p except for 2.

4. $\mathbb{R}^{m \times n}$ with inner product $< \cdot, \cdot >$: $< A, B >= \sum_{ij} a_{ij} bij$
   The induced norm is
   $\|A\| = (< A, A >)^{\frac{1}{2}} = (\sum_{ij} a_{ij}^2)^{\frac{1}{2}} = \|A\|_F = \|A\|_{vec,2}$

5. Infinite sequence with inner product $< \cdot, \cdot >$: $< a, b >= \sum_{i=1}^{\infty} a_i b_i$
   $\|a\| = (\sum_{i=1}^{\infty} a_i^2)^{\frac{1}{2}} = \|a\|_2$

6. $C[a, b]$ with inner product $< \cdot, \cdot >$: $< f, g >= \int_a^b f(t)g(t)dt$
   $\|f\| = (\int_a^b (f(t))^2 dt)^{\frac{1}{2}} = \|f\|_2$

Angle in inner product spaces
By Cauchy-Schwartz inequality,

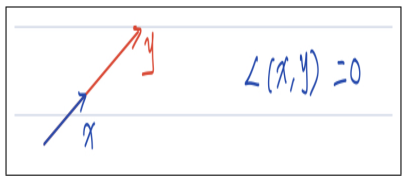
$$| < x, y > | \leq \|x\| \|y\| \ \forall x, y \in \mathbb{V}$$

Then,

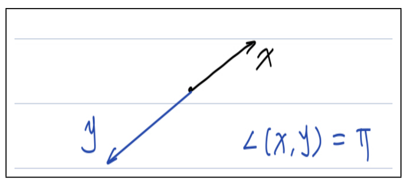
$$-\|x\|\|y\| \ \leq < x, y > \leq \ \|x\|\|y\|$$

$$-1 \ \leq \ \frac{<x,y>}{\|x\|\|y\|} \ \leq \ 1 \text{ if } x, y \neq 0$$

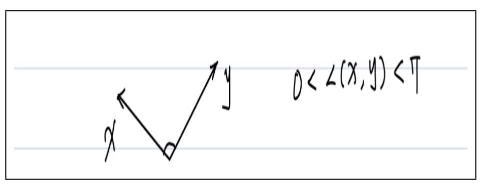
If $\frac{<x,y>}{\|x\|\|y\|} = 1$, then $x = \alpha y$ with $\alpha > 0$. Otherwise, if $\alpha \leq 0$, then $< x, y >= \alpha < y, y >= \alpha \|y\|^2 \leq 0$. $(Contradiction)$.



If $\frac{<x,y>}{\|x\|\|y\|} = -1$, then $x = \alpha y$ with $\alpha < 0$.



If $-1 \ < \ \frac{<x,y>}{\|x\|\|y\|} \ < \ 1$, then



Then we define

$$\boxed{L(x, y) = \arccos \frac{<x,y>}{\|x\|\|y\|}}$$

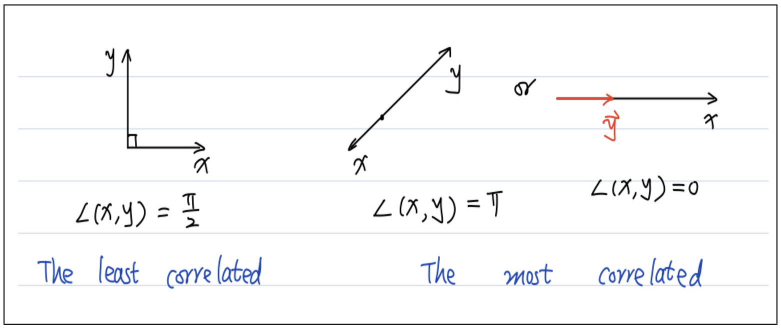
This definition is consistent with the observation above and the angles of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Orthogonality

Let $\mathbb{V}$ be a vector space and $<\cdot,\cdot>$ be the inner product.

- If $\frac{<x,y>}{\|x\|\|y\|} = 1$ or -1, then x and y are the most correlated.
- If $\frac{<x,y>}{\|x\|\|y\|} = 0$, then x and y are the least correlated.

If $<x,y>= 0$, then we say x and y are orthogonal.



Pythagorean theorem

Definition: Let x, y be two vectors in an inner product space $\mathbb{V}$.
Then $x \perp y \iff \|x+y\|^2 = \|x\|^2 + \|y\|^2$.

Proof.

$$\|x+y\|_2 =< x+y, x+y >$$

$$= \|x\|^2 + \|y\|^2 + 2 < x, y > \textbf{(1)}$$

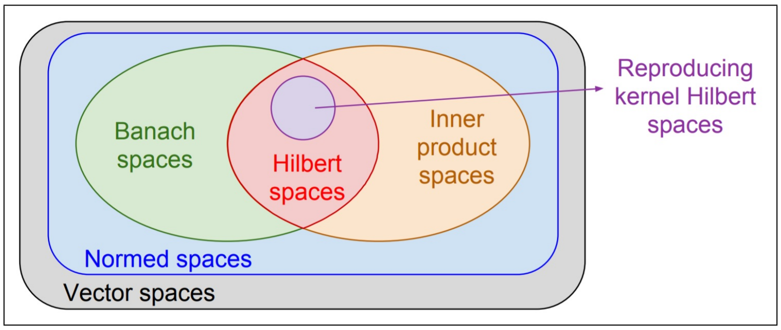
If $x \perp y$, then $< x, y >= 0$.

$$\implies \|x+y\|^2 = \|x\|^2 + \|y\|^2$$

If $\|x+y\|^2 = \|x\|^2 + \|y\|^2$, together with **(1)**, we have $< x, y >= 0$.

Hilbert Space

Definition: A Hilbert space is a Banach space in which the norm is induced by an inner product.



Examples of Hilbert Space

1. $\mathbb{R}^n$ with $< \cdot, \cdot >$ is a Hilbert space.
2. $\mathbb{R}^n$ with $< \cdot, \cdot >_A$ is a Hilbert space.
3. $\mathbb{R}^{m \times n}$ with $< \cdot, \cdot >$ is a Hilbert space.
4. $L_2 = \{a \mid \|a\|_2 < \infty$ and a is a infinite sequence$\}$ with $< \cdot, \cdot >$ is a Hilbert space.
5. $C[a, b]$ with $< \cdot, \cdot >$ is **NOT** a Hilbert space, because it is not a Banach space. In other words, the limit of a convergent sequence in $C[a, b]$ may not be in $C[a, b]$. To complete $C[a, b]$ under the norm $\| \cdot \| = (< \cdot, \cdot >)^{\frac{1}{2}}$, we need to extend the Riemann integral to the so-called Lebesgue integral, and the resulting Hilbert space is $L^2(a, b)$.

In the following chapters, we will consider calculus on Hilbert/Banach spaces.

Linear and Differentiable Functions

3.1 Linear Function

Definition: Let $\mathbb{V}$ be a vector space and $f : \mathbb{V} \to \mathbb{R}$ be a function. f is a linear function if

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y), \forall \alpha, \beta \in \mathbb{R} \text{ and } x, y \in \mathbb{V}$$

Example: The mean of a vector in $\mathbb{R}^n$.

$\forall x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, $f(x) = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$ is a linear function because

$$f(\alpha x + \alpha y) = \frac{\sum_{i=1}^n (\alpha x_i + \beta y_i)}{n} = \alpha \frac{\sum_{i=1}^n x_i}{n} + \beta \frac{\sum_{i=1}^n y_i}{n} = \alpha f(x) + \beta f(y)$$

Example: The maximum entry of a vector in $\mathbb{R}^n$.

$\forall x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, $f(x) = \max_{i=1,\cdots,n} x_i$ is not a linear function.

One counter example:

$$x = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, y = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \alpha = 1, \beta = 1$$

$$f(\alpha x + \beta y) = f(\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}) = 1 \text{ but } \alpha f(x) = 1 \text{ and } \beta f(y) = 1$$

Hence, $f(\alpha x + \beta y) \neq \alpha f(x) + \beta f(y)$.

Example: $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = < a, x >$, where $a \in \mathbb{R}^n$ is a fixed vector in $\mathbb{R}^n$ is linear.

Example: $F : C[-1,1] \to \mathbb{R}$ defined by $F(f) = f(0)$ is linear because

$$F(\alpha f + \beta g) = (\alpha f + \beta g)(0) = \alpha f(0) + \beta g(0) = \alpha F(f) + \beta F(g)$$

Example: $F : C[a,b] \to \mathbb{R}$ defined by $F(f) = \int_a^b f(t)dt$ is linear because

$$F(\alpha f + \beta g) = \int_a^b (\alpha f + \beta g)(t)dt$$

$$= \int_a^b (\alpha f(t) + \beta g(t))dt$$

$$= \alpha \int_a^b f(t)dt + \beta \int_a^b g(t)dt$$

$$= \alpha F(f) + \beta F(g)$$

Example: Let $\mathbb{V}$ be an inner product space with inner product $< \cdot, \cdot >$. Let $a \in \mathbb{V}$ and $f : \mathbb{V} \to \mathbb{R}$ defined by $f(x) =< a, x >$ is linear.

Example: A norm function on $\mathbb{V}$ is **NOT** linear.

Proof: Let $\| \cdot \| : \mathbb{V} \to \mathbb{R}$. Then $\| -x\| = \|x\|$ by norm property.
If $\| \cdot \|$ is linear, then

$$\| -x\| = \| -x + 0 \cdot x = -1\|x\| + 0\|x\| = -\|x\| \text{ (A contradiction)}$$

Properties of Linear Function:

1. *Homogeneity*:

$$f(\alpha x) = \alpha f(x),\ \forall \alpha \in \mathbb{R},\ x \in \mathbb{V}$$

$$\text{because } f(\alpha x) = f(\alpha x + 0 \cdot y) = \alpha f(x) + 0 \cdot f(y) = \alpha f(x)$$

$$\text{Choosing } \alpha = 0, \text{ then we obtain } f(0) = 0.$$

2. *Additivity*:

$$f(x + y) = f(x) + f(y)$$

$$f(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k)$$

$$= \alpha_1 f(x_1) + f(\alpha_2 x_2 + \cdots + \alpha_k x_k)$$

$$= \alpha_1 f(x_1) + \alpha_2 f(x_2) + f(\alpha_3 x_3 + \cdots + \alpha_k x_k)$$

$$= \cdots$$

$$= \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k$$

Linear Function on Hilbert Space
For simplicity, let's consider a linear function on $\mathbb{R}^n$ equipped with the standard inner product $< x, y = x^T y >$ and the induced norm $\|x\|_2 = (< x, x >)^{\frac{1}{2}}$.

- From one of the examples above,

  For any give $a \in \mathbb{R}^n$, the function $f(x) =< a, x >$ is linear.

- The reverse is true, i.e.,

  Any linear function $f : \mathbb{R}^n \to \mathbb{R}$ must be in the form of $f(x) =< a, x >$ for some $a \in \mathbb{R}^n$.

We are assuming that $\mathbb{H} = \mathbb{R}^n$ for simplicity, but this theorem actually holds for any forms of Hilbert Space $\mathbb{H}$.

Theorem: For any linear function $f : \mathbb{R}^n \to \mathbb{R}$, there exists a unique $a \in \mathbb{R}^n$ s.t. $f(x) =< a, x >$, $\forall x \in \mathbb{R}^n$.

Proof: Let e_1, e_2, .. e_n be the natural basis of $\mathbb{R}^n$ where e_i is a vector where the i-th entry is 1 and 0 elsewhere.

$\forall x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, $x = x_1e_1 + x_2e_2 + \cdots + x_ne_n$

So $f(x) = f(x_1e_1 + x_2e_2 + \cdots + x_ne_n)$

$= x_1f(e_1) + x_2f(e_2) + \cdots + x_nf(e_n)$

$=< \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} f(e_1) \\ f(e_2) \\ \vdots \\ f(e_n) \end{bmatrix} >$

$=< a, x >$ where $a = \begin{bmatrix} f(e_1) \\ f(e_2) \\ \vdots \\ f(e_n) \end{bmatrix}$

Now we prove the uniqueness of this theorem.
Suppose a is **NOT** unique, $\exists a, b \in \mathbb{R}^n$ s.t.

$$f(x) =< a, x >=< b, x >, \forall x \in \mathbb{R}^n$$

Then choose $x = e_i$, $i = 1, \cdots, n$

$$f(e_i) =< a, e_i >=< b, e_i > \implies a_i = b_i,\ i = 1, \cdots, n$$

$$\implies a = b$$

(A contradiction)

Therefore, a must be unique.

Riesz Representation Theorem

Extending previous theorem to the entirety of Hilbert space $\mathbb{H}$.

Theorem:
Let $\mathbb{H}$ be a Hilbert space. Let $f : \mathbb{H} \to \mathbb{R}$. Then f is linear and bounded if and only if $f(x) =< a, x >$ for some unique $a \in \mathbb{H}$.

Example: We know that mean of a vector on $\mathbb{R}^n$ is linear.

$$f(x) = \frac{x_1+x_2+\cdots+x_n}{n} =< \frac{1}{n}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}, x >$$

Example: Let $\mathbb{H}$ be a Hilbert space and $\|\cdot\|$ is **NOT** linear. So there is no such $a \in \mathbb{H}$ s.t. $\|x\| =< a, x >, \forall x \in \mathbb{H}$.

Example: $\mathbb{R}^{n\times n}$ with inner product

$$< A, B >= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij} \ \forall A, B \in \mathbb{R}^{n\times n}$$

Define $trace(A) = \sum_{i=1}^{n} a_{ii} \forall A \in \mathbb{R}^{n\times n}$ is linear. We have

$$trace(A) =< A, I >$$

Remarks:

1. In finite dimensional Hilbert space, linear $\iff$ linear and bounded.
2. In infinite dimensional Hilbert space, there exists linear but unbounded function.

Example: $L^2(-1,1)$ - the completion of $C[-1,1]$ under the inner product $< f, g >= \int_{-1}^{1} f(t)g(t)dt$ and $\|f\|_2 = (< f, f >)^{\frac{1}{2}} = (\int_{-1}^{1} \mid f(t) \mid^2 dt)^{\frac{1}{2}}$.
Consider $F(f) = f(0), \forall f \in L^2(-1,1)$
But $F(f)$ is unbounded since

$$\exists f \in L^2(-1,1) \text{ s.t. F(f)} = \infty$$

$$\text{e.g. } f(t) = \begin{cases} 1 & t \neq 0 \text{ and } t \in (-1,1) \\ \infty & t = 0 \end{cases}$$

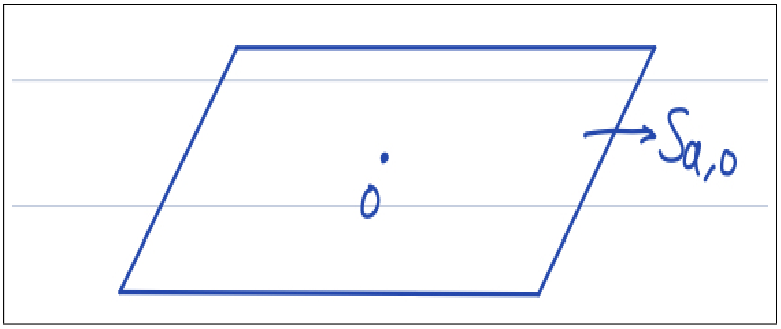
There exists no inner product representation for $F(f) = f(0)$.

Example: $L^2(-1,1)$
Consider $G : L^2(-1,1) \to \mathbb{R}$

$$G(f) = \int_{-1}^{1} f(t)dt$$

G is linear.
G is bounded because for any $f \in L^2(-1,1)$,
$G(f) = \int_{-1}^{1} f(t)dt = \int_{-1}^{1} f(t) \cdot 1dt =< f, 1 > \leq \|f\|_2(\int_{-1}^{1} 1^2 dt)^{\frac{1}{2}} \leq 2\|f\|_2$
Riesz $\implies g \in L^2(-1,1)$ s.t. $G(f) =< f, g >$. Indeed, $g(t) = 1, \forall t \in (-1,1)$.

Hyperplane
Let $\mathbb{H}$ be a Hilbert space and $a \in \mathbb{H}$.
Consider $S_{a,0} = \{x \in H | < a, x >= 0\} \subset \mathbb{H}$,
Then $\forall \alpha, \beta \in \mathbb{R}$ and $\forall x, y \in S_{a,0}$
$< a, \alpha x + \beta y >= \alpha < a, x > +\beta < a, y >= 0$. That is,
$\alpha x + \beta y \in S_{a,0} \implies S_{a,0}$ is a linear space (subspace of H).



$S_{a,b} = \{x \in \mathbb{H} | < a, x >= b\} \subset \mathbb{H}$ Let $x_0 \in S_{a,b}$, then $< a, x_0 >= b$.

1. $\forall a \in S_{a,b}$
   $< a, x - x_0 >=< a, x > - < a, x_0 >= b - b = 0$
   $\implies x - x_0 \in S_{a,0} \implies x \in x_0 + S_{a,0} \implies S_{a,b} \subset x_0 + S_{a,0}$

2. $\forall x \in S_{a,0}$
   $< a, x + x_0 >=< a, x > + < a, x_0 >= 0 + b = b$
   $\implies x + x_0 \in S_{a,b} \implies S_{a,0} + x_0 \subset S_{a,b}$

(1) and (2) $\implies S_{a,b} = S_{a,0} + x_0$
$S_{a,b}$ is a shift of a subspace.