1

Introducción a la optimización convexa

1.1 Introducción

$$(P)\begin{cases} \min & f_0(x) \\ & \\ s.a. & f_1(x) \leq b_1 \\ & f_2(x) \leq b_2 \\ & \vdots \\ & f_m(x) \leq b_m. \end{cases}$$

Donde,

$$\begin{array}{rl} f_i: & \mathbb{R}^n \to \mathbb{R} \\ f_0: & \text{Función objetivo.} \\ f_j: & \text{Función Restricción donde } j = 1, \ldots, m. \end{array}$$

El objetivo de (P) es encontrar x^* el optimo (arg min) que cumpla:

$$f_0(x^*) \leq f_0(x),\ \forall x \in \mathbb{R}^n / f_j(x) \leq b_j,\ j = 1, \ldots, m.$$

Los Puntos factibles son los $x \in \mathbb{R}^n / f_j(x) \leq b_j,\ j = 1, \ldots, m.$

Cuando el problema sea de la forma convexa se llama optimización convexa. Al final, la habilidad es identificar las restricciones y convertirlas a convexas.

- Las funciones objetivos en economía se les puede llamar funciones de coste.
- Multiplicamos las desigualdades por (-1) para darle la forma que más nos convenga darle al problema de optimización.
- Maximizar será lo mismo que minimizar. En nuestro caso minimizaremos las funciones.
- Al valor $f_0(x^*)$ se le llamará valor optimo.
- En $f_i : \mathbb{R}^n \to \mathbb{R}$ existirán algunas funciones el cual su dominio sera "tramposo".

Nota 1.1. *Imaginemos que tenemos*

$$\{\min \quad f(x)$$

$$\{\min \quad f_0^2(x)$$

- *Si las función f_0 es positiva las dos formas son equivalentes.*
- *El valor optimo no será el mismo, pero el punto optimo lo será, ya que las funciones son monótomas crecientes.*
- *Si el valor al cuadrado simplificará la solución, entonces podemos utilizarla. Esto nos permite que si no tengamos una función convexa podamos convexificarla.*

Notación 1.1 Podemos escribir Ax como

$$\begin{aligned} Ax &= \underbrace{\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{k1} \end{pmatrix}}_{A^1} x_1 + \underbrace{\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{k2} \end{pmatrix}}_{A^2} x_2 + \cdots + \underbrace{\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{kn} \end{pmatrix}}_{A^n} x_n \\ &= x_1 A^1 + x_2 A^2 + \cdots + x_n A^n. \end{aligned}$$

A^1 = A super 1 como columna, y A_1 = A super 1 como fila.

Ahora, en términos de filas. Si escribimos los vectores A en columna

$$A = \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_k^T \end{pmatrix}$$

Donde,

$$Ax = \begin{pmatrix} A_1^T x \\ A_2^T x \\ \vdots \\ A_n^T x \end{pmatrix} = \begin{pmatrix} \langle A_1^T, x \rangle \\ \langle A_2^T, x \rangle \\ \vdots \\ \langle A_n^T, x \rangle \end{pmatrix}$$

Ejemplo 1.1 Sean $A \in \mathcal{M}_{k\times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathcal{R}^k$.

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \qquad \mathbf{x}^T = (x_1, x_2, \ldots, x_n).$$

El problema será una minimización global dada por: $\left\{ \begin{array}{ll} \min: & \|A\mathbf{x} - b\|_2^2 \\ s.a. & \emptyset. \end{array} \right.$

Solución.- Por diferenciabilidad se tiene,

$$f_0(x) = \|Ax - b\|_2^2 = \langle Ax - b, Ax - b \rangle.$$

Intentaremos demostrar el punto donde las parciales de $f_0 = 0$. Para

- Diremos que el un vector cualquiera sera vector columna.
- El subindice $_2$ significa la normal Euclidea. Que es la distancia normal que existe en $\mathbb{R}^2$.

ello, encontraremos

$$\begin{aligned} D_i f_0 &= D_i \left(\langle Ax - b, Ax - b \rangle \right) \\ &= \langle D_i \left(Ax - b \right), Ax - b \rangle + \langle Ax - b, D_i \left(Ax - b \right) \rangle \\ &= 2\langle Ax - b, D_i \left(Ax - b \right) \rangle. \end{aligned}$$

Veamos la parcial de $D_i\,(Ax - b)$.

$$D_i\,(Ax - b) = D_i \left(x_1 A^1 + x_2 A^2 + \cdots + x_n A^n - b \right) = A^i.$$

Dado que b que es constante vale cero, y donde todos los suman que no estén las x_i también valen cero. Por lo tanto,

$$D_i f_0 = 2\langle Ax - b, A^i \rangle.$$

Luego,

$$2\langle Ax - b, A^i \rangle = 0 \;\forall i = 1, \ldots, n \quad \Rightarrow \quad \langle Ax - b, A^i \rangle = 0, \quad \forall i = 1, \ldots, n.$$

Observemos que,

- En funciones convexas el extremo local será el mínimo global.

$$\overrightarrow{0} = \begin{pmatrix} \langle Ax - b, A^1 \rangle \\ \langle Ax - b, A^2 \rangle \\ \vdots \\ \langle Ax - b, A^n \rangle \end{pmatrix} = \begin{pmatrix} (A_1)^T \\ (A_2)^T \\ \vdots \\ (A_n)^T \end{pmatrix} (Ax - b) = A^T(Ax - b).$$

$$A^T(Ax - b) = \overrightarrow{0} \quad \Leftrightarrow \quad A^T Ax = A^T b.$$

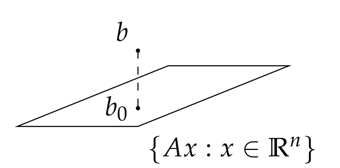
El cual es una ecuación normal.

Por último, veamos los argumentos geométricos. Notemos que,

$$\min \|Ax - b\|_2^2 = d(b_i, Ax)^2$$

Donde Ax tendrá la forma geométrica de un subespacio vectorial (en el caso de $\mathbb{R}^3$ será un plano). Entonces,

- Si $b \in \{Ax : x \in \mathbb{R}^n\} \;\Leftrightarrow\; x^* \in \mathbb{R}^n : Ax^* = b$. El valor optimo es $f_0(x^*) = 0$.
- Si $b \notin \{Ax : x \in \mathbb{R}^n\}$, $f_0(x^*) = d(b, b_0)^2$.



Ahora, cual es el optimo?; es decir cual es el x^*. Para ello,

$$x^* \in \mathbb{R}^n : Ax^* = b_0.$$

Aquí, b_0 está en el plano, si estamos en $\mathbb{R}^3$. ¿Cómo llegamos algebraicamente?:

$$\begin{aligned}
b - b_0 \perp \{Ax : x \in \mathbb{R}^n\} &\Leftrightarrow b - b_o \perp A^i,\ i = 1, \ldots, n \\
&\Leftrightarrow \langle b - b_0, A^i \rangle = 0,\ i = 1, \ldots, n \\
&\Leftrightarrow \langle b - Ax^*, A^i \rangle = 0,\ i = 1, \ldots, n \\
&\Leftrightarrow \langle Ax^* - b, A^i \rangle = 0,\ i = 1, \ldots, n \\
&\Leftrightarrow A^T A x^* = A^T b.
\end{aligned}$$

(Las ecuaciones normales vienen dadas por la perpendicularidad.)

2

Conjuntos convexos

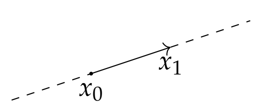
2.1 Conjuntos convexos de $\mathbb{R}^n$

El dominio serán conjuntos convexos o dominio efectivo.

- Cuando λ vale 1 será x_1 cuando valga cero x_0.
- Cuando es positivo irá a la derecha, cuando es negativo hacia la izquierda.
- Toda la recta nos da un concepto que denominamos Afín. Es cualquier punto que este entre x_0 y x_1 del gráfico de arriba.

Definición 2.1 **Lineal.**

$$L(x_0, x_1) := \{x_0 + \lambda(x_1 - x_0) : \lambda \in \mathbb{R}\}$$



Para la convexidad no es necesario tener la linea, solo necesitaremos un segmento definido por:

$$[x_0, x_1] := \{x_0 + \lambda(x_1 - x_0) : \lambda \in [0, 1]\} = \{(1 - \lambda)x_0 + \lambda x_1 : \lambda \in [0, 1]\}$$

- Manejar el concepto de afín con lineas es incomodo, por lo que se utiliza el concepto de combinación afín.
- La diferencia entre espacio vectorial y espacio afín es que el espacio afín esta desplazado; es decir, no necesariamente pasa por el cero como en un subespacio vectorial.

Definición 2.2 Sea $A \subseteq \mathbb{R}^n$. Se dice **Afín**, si $\forall x, y \in A$ se tiene que la $L(x, y) \subseteq A$. (Subespacios vectoriales desplazados).

- Un circulo no es afín ya que la linea es infinita.
- Un plano podría ser Afín.

- La recta es afín.
- Todo $\mathbb{R}^n$ es afín.
- Un único punto también es afín, dado que $x = y$.

- $(1-\lambda)x_0 + \lambda x_1$ es una combinación lineal de x_0 y x_1. Donde $(1-\lambda) + \lambda = 1$.
- Lo demás puntos fuera del segmento son las combinaciones lineales de x_0 y x_1.

Definición 2.3 Una **combinación afín** de los vectores $\{x_1, x_2, \ldots, x_k\}$ es un vector de la forma

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k.$$

tal que

$$\sum_{i=1}^{k} \lambda_i = 1.$$

- Este conjunto es estable para combinaciones lineales, muy similar al concepto de subespacio vectorial.

Teorema 2.1 A es afín sii A contiene toda combinación afín de sus puntos.

Demostración.- Primero, tomemos puntos arbitrarios $\{x_1, x_2, \ldots, x_k\}$ en A tal que

$$z = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k$$

donde $\sum\limits_{i=1}^{k} \lambda_i = 1$.

Ahora, consideremos dos puntos x_i, x_j de z. Dado que A es afín, entonces $L(x_i, x_j) \subseteq A$, para todo x_i, x_j. Esto implica que z está en A. Intuitivamente, si

$$\lambda_1 x_1 + \lambda_2 x_2, \quad \lambda_3 \lambda_3 + \lambda_4 \lambda_4, \quad \ldots, \quad \lambda_{k-1} x_{k-1} + \lambda_k x_k. \qquad \text{con } \sum_{i01}^{k} y_i = 1.$$

están en A. Entonces, z tendrá que estar en A.

Para demostrar la otra implicación, tomemos dos puntos cualesquiera x_1 e x_k en A. Queremos demostrar que

$$L(x_1, x_k) = \{x_1 + \lambda(x_k - x_1) : \lambda \in \mathbb{R}\}$$

está contenido en A. Por el hecho de que x_1 e x_k están en A, podemos considerar la línea $L(x_1, x_k)$. Cualquier punto en esta línea se puede expresar como

$$z = x_1 + \lambda(x_k - x_1),$$

donde $\lambda \in \mathbb{R}$. Ahora, notemos que $\lambda + (1-\lambda) = 1$, lo cual es la condición de combinación afín. Y dado que A contiene toda combinación afín de sus puntos, esto implica que z está en A. Por lo tanto, A es afín. ■

Nota 2.2. *La definición de subespacio se refiere a tomar dos escalares y dos vectores, realizar la* *combinación lineal, donde esta combinación lineal no se saldrá del conjunto dado.*

Notación 2.1 La suma de Minkowski es la operación de conjuntos; es decir, si $A, E \subseteq \mathbb{R}^n$. Entonces,

$$A = x_0 + E = \{x_0 + e : e \in E\} \quad \text{o} \quad E = A - x_0 = \{a - x_0 : a \in A\}.$$

Es sencillamente trasladar los puntos del plano y desplazarlos o moverlos.

Teorema 2.2 $A \subseteq \mathbb{R}^n$ es afín sii existe un subespacio vectorial $E \subseteq \mathbb{R}^n$ tal que $A = x_0 + E$ para todo $x_0 \in A$.

Demostración.- Supongamos que A es afín y fijamos $x_0 \in A$. Intentaremos probar que $E = A - x_0$ es un subespacio de $\mathbb{R}^n$, esto es equivalente a decir que:

$$\lambda, \mu \in \mathbb{R}, e_i, e_2 \subseteq E \quad \Rightarrow \quad \lambda e_i + \mu e_2 \in E.$$

Probemos que $\lambda e_1 + \mu e_2 \in E$; en otras palabras, probaremos que $\lambda e_1 + \mu e_2$ es $a - x_0$.

$$\begin{aligned} \lambda e_1 + \mu e_2 &= \lambda(a_1 - x_0) + \mu(a_2 - x_0) \\ &= \lambda a_1 + \lambda a_2 - \lambda x_0 - \mu x_0 \\ &= \lambda a_1 + \lambda a_2 - \lambda x_0 - \mu x_0 + x_0 - x_0 \\ &= \lambda a_1 + \lambda a_2 + (1 - \lambda - \mu)x_0 - x_0. \end{aligned}$$

Observemos que $\lambda a_1 + \lambda a_2 + (1 - \lambda - \mu)x_0$ está en A, dado a que $\lambda + \mu + (1 - \lambda - \mu) = 1$. Por lo tanto,

$$A - x_0 = E.$$

Es un subespacio vectorial.

Ahora, para demostrar que A es afín, probaré que cualquier combinación afín de elementos de A sigue estando en A (Teorema 1.1). Sean,

$$\{a_1, a_2, \ldots, a_k\}, \ \lambda_1, \lambda_2, \cdots, \lambda_k : \sum \lambda_i = 1.$$

De donde,

$$\begin{aligned} \lambda_1 a_1 + \lambda_2 a_2 + \cdots + \lambda_k a_k &= \lambda_1(e_1 - x_0) + \lambda_2(e_2 - x_0) + \cdots + \lambda_k(e_k - x_0) \\ &= \lambda_1 e_1 + \cdots + \lambda_k e_k + \left(\sum_{i=1}^{k} \lambda_i\right) x_0 \end{aligned}$$

Observemos que $\lambda_1 e_1 + \cdots + \lambda_k e_k$ es una combinación lineal afín el cual existe en E y por definición, $\left(\sum_{i=1}^{k} \lambda_i\right) = 1$. Por lo tanto,

$$E + x_0 = A.$$

■

Definición 2.4 **Envoltura Afín.**

La envolura afín de B, $\text{Aff}(B)$, es el menor conjunto afín que contiene a B. Esto implica que es el conjunto de las combinaciones afines de elementos de B o es la intersección de los conjuntos afines que contienen a B

Definición 2.5 Si A es Afín se llama "dimensión afín de A" a la dimensión de su espacio vectorial.

- Dimensión 0 un punto.
- Dimensión 1 una recta.
- Dimensión 2 una plano.

Ejemplo 2.1 Dado $C \in \mathbb{R}^n$ afín. Siempre existirán una matriz $A \in \mathcal{M}_{p\times n}$ y $b \in \mathbb{R}^p$ tal que

$$C = \{x \in \mathbb{R}^n : Ax = b\}.$$

Solución.- El conjunto lineal asociado será el núcleo de la aplicación lineal. Es decir,

$$E = \{e \in \mathbb{R}^n : Ae = 0\},$$

cualquier solución de $x_0 \in C$ de modo que $Ax_0 = b$. Tomando un punto de C y otro de E, tenemos

$$A(x_0 + e) = Ax_0 + Ae = b + 0 = b.$$

Por lo tanto,

$$C = \{x \in \mathbb{R}^n : Ax = b\} = x_0 + E.$$

Así, el conjunto afín no es más que el traslado del espacio vectorial.

Definición 2.6 **Topología de $\mathbb{R}^n$.**

Sean $A \subseteq \mathbb{R}^n$ y $a \in \mathbb{R}^n$:

1) $a \in A$ está en el interior de A $\left(a \in \text{int}(A) \text{ o } a \in \mathring{A}\right)$, cuando existe $\delta > 0$ tal que $B(a,\delta) \subseteq A$.

$$B(a,\delta) = \{x \in \mathbb{R}^n : \|x-a\|_2 \leq \delta.\}.$$

2) A se dice abierto si $A = \text{int}(A) = \mathring{A}$.

3) Decimos que $c \in \mathbb{R}^n$ está en el cierre (o clausura) de A, cuando $\exists \{a_n\} \in A | a_n \to c$.

4) Decimos que A es cerrado cuando $A = \overline{A}$ donde

$$\{x \in \mathbb{R}^n : x \text{ está en el cierre de A}\}.$$

- El concepto de punto interior es importante, ya que podemos acercarnos al punto a de todas las direcciones.
- Si es un punto relativo interior nos acercaremos por todos los lados del conjunto.
- El punto de adherencia o clausura es un punto el cual me puedo acercar de alguna forma.

1) En $\mathbb{R}^2$ será un circulo y en $\mathbb{R}^3$ será una esfera.

2) Son las bolas que están completamente dentro del conjunto. Es decir, no tienen puntos frontera.

3) Es cualquier bola de c que corta al conjunto o los puntos que contienen a toda su frontera.

4) El cierre son los puntos interior y los puntos frontera.

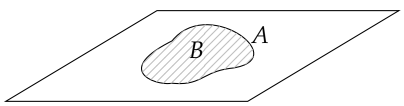
5) Cualquier bola está en el interior cómo en el exterior del conjunto.

5) Se llama frontera de A, ∂A a la intersección $\overline{A} \cap \left(\overline{\mathbb{R}\backslash A}\right) = \overline{A}\backslash \operatorname{int}(A)$ (Cualquier bola estará una parte en el interior y otra en el exterior del conjunto).

6) $a \in \operatorname{relint}(A)$ si existe $\delta > 0$ tal que $B(a,\delta) \cap Aff(A) \subseteq A$.

6) Imaginamos un corte transversal para proyectar una imagen.

Ejemplo 2.2 Dibujemos un plano ($\mathbb{R}^3$)



- A es la frontera.
- El objetivo será encontrar el punto optimo de un esfera que está proyectada en este plano.
- B es el interior con la frontera.
- El conjunto tendrá que ser convexo.

Veamos algunas propiedades de este conjunto.

1) A es cerrado | Cualquier punto que ponga en B me puedo acercar por puntos de B.
2) B cerrado.
3) $\mathring{A} = \emptyset$ | Si yo ponga una bola $\mathbb{R}^3$, se saldrá del conjunto A.
4) $\mathring{B} = \emptyset$ | Ya que no existirá en el plano ninguna esfera.
5) $\operatorname{relint}(A) = \emptyset$; $\operatorname{relint}(B) = B\backslash A$.

- La única diferencia entre combinación convexa y afín es que la combinación convexa es positiva.

Definición 2.7 **Combinación convexa.**

Sean $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ y $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ tales que

$$\lambda_i \geq 0 \text{ y } \sum_{i=1}^{k} \lambda_i = 1.$$

Al vector

$$\sum_{i=1}^{k} \lambda_i x_i = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k$$

se le llama combinación convexa de los puntos $\{x_1, \ldots, x_k\}$.

- En particular, las combinaciones convexas son los segmentos.

Observación 2.1 La definición para 2 puntos $\{x_1, x_2\}$ nos da las combinaciones convexas,

$$\lambda x_1 + (1-\lambda)x_2, \qquad \lambda \geq 0, (1-\lambda) \geq 0 \iff \lambda \in [0,1].$$

Esto es el segmento,

$$\{\lambda x_1 + (1-\lambda)x_2 : \lambda \in [0,1]\} = [x_1, x_2].$$

Nos quedamos con el segmento que los une, eso nos permitirá utilizar las propiedades de los números reales. Por lo que podremos realizar análisis.

- Decimos que C es cerrado para las combinaciones convexas. Es decir, no me salgo del conjunto.

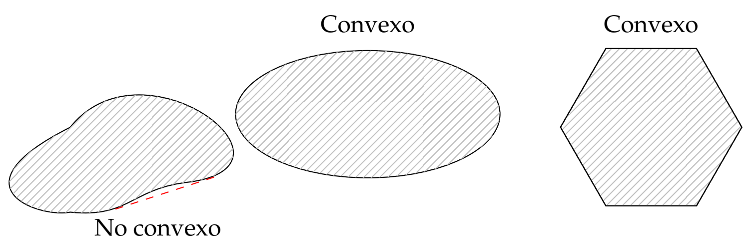
Definición 2.8 **Convexo.**

Un conjunto $C \in \mathbb{R}^n$ se dice convexo cuando C contiene las combinaciones convexas de sus puntos, si y sólo si

$$\forall x_1, x_2 \in C \Rightarrow [x_1, x_2] \subseteq C.$$

Un conjunto es convexo si dados dos puntos el segmento que los une se queda adentro.

Ejemplo 2.3



Del gráfico 1) ¿Cuál es el menor conjunto convexo que lo contiene?

Definición 2.9 se llama **envoltura convexa** de A al menor conjunto convexo que lo contiene o a la intersección de todos los convexos que contienen a A, denotado por $\text{co}(A)$. También es equivalente a decir que

$$\text{co}(A) = \{\text{Combinación convexa de puntos de A.}\}$$

Ejercicio 2.1 Demostrar que la intersección de conjuntos convexos es convexo.

Demostración.- Demostremos por contradicción. Sean C_1 y C_2 dos conjuntos convexos. Y sea

$$C = C_1 \cap C_2.$$

no convexo. Esto significa que existen x e y tales que

$$\{\lambda x + (1-\lambda)y : \lambda \in \mathbb{R}\} \not\subseteq C.$$

Supongamos ahora que x e y están en C. Cómo ambos C_1 y C_2 son convexos, el segmento definido debe estar en ambos conjuntos. Es decir,

$$\{\lambda x + (1-\lambda)y : \lambda \in \mathbb{R}\} \subseteq C.$$

Lo que contradice nuestra suposición inicial. Por lo tanto, C es convexo. ■

- Contiene los rayos que pasan por el cero e intersecan a un punto dado.

Definición 2.10 **Cono.**

Un conjunto $C \subseteq \mathbb{R}^n$ se llama cono si y sólo si

$$\lambda x \in C \text{ si } x \in C, \ \lambda \geq 0.$$

Propiedad .1 **Propiedades de los conos.**

a) Un cono siempre contiene al origen.

b) La envoltura cónica de un conjunto es $\text{con}(A) = \{\lambda : \lambda \geq 0, a \in A\}$. La intersección de todos los conos que contiene a A.

c) Un cono C es convexo si y sólo si

$$\lambda_1, \lambda_2 \in C \Rightarrow \lambda_1 x_1 + \lambda_2 x_2 \in C, \ \forall \lambda_1, \lambda_2 \geq 0.$$

d) C es un cono convexo si

$$\sum_{i=1}^{m} \lambda_i x_i \in C$$

para $\lambda_i \geq 0$.

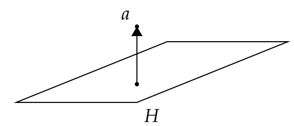
- El hiperplano es un caso particular del estudio convexo.

Definición 2.11 **Hiperplano.**

$H \subseteq \mathbb{R}^n$ es un **hiperplano** si existe $a \in \mathbb{R}^n \setminus \{0\}$ tal que

$$H = \left\{ x \in \mathbb{R}^n : \langle a, x \rangle = a^T x = 0 \right\} = a^{\perp}.$$

- Hiperplano:



- En $\mathbb{R}$ los hiperplanos son rectas.
- En $\mathbb{R}^n$, los hiperplanos serán uno menos de dimensión.

Proposición 2.1 $H \subseteq \mathbb{R}^n$ es un hiperplano si y sólo si, H es un subespacio de dimensión $n-1$.

Demostración.- Primero, supongamos que H es un hiperplano. Entonces, por definición, existe un vector $a \in \mathbb{R}^n \setminus \{0\}$ tal que

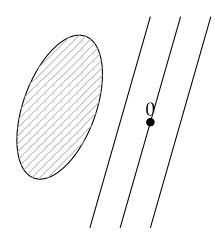
$$H = \left\{ x \in \mathbb{R}^n : a^T x = 0 \right\} = a^{\perp}$$

Esto significa que H es el conjunto de todos los vectores que son ortogonales a a. Recordemos que cualquier múltiplo escalar del primer vector también es ortogonal al segundo; además, si dos vectores son ortogonales a otro, entonces la suma de los dos primeros vectores también será ortogonal al tercer vector. Es decir, la ortogonalidad preserva las operaciones de suma y multiplicación por escalares. Por lo tanto, H es un subespacio de $\mathbb{R}^n$. Ahora bien, como a no es el vector cero, el conjunto $\{a\}$ es linealmente independiente (ya que no hay otros vectores), y por lo tanto forma una base para un subespacio de $\mathbb{R}^n$. Como este subespacio es ortogonal a H, la dimensión de H debe ser $n-1$.

Ahora supongamos que H es un subespacio de $\mathbb{R}^n$ de dimensión $n-1$. Entonces, existe un subespacio de $\mathbb{R}^n$ que es ortogonal a H y tiene dimensión 1. Este subespacio tiene una base formada por un único vector, digamos a. Entonces, para cualquier vector $x \in H$, tenemos que $a^T x = 0$, lo que significa que x es ortogonal a a. Por lo tanto, podemos escribir $H = \{x \in \mathbb{R}^n : a^T x = 0\}$, lo que significa que H es un hiperplano. ■

Observemos que la recta que pasa por el cero estará definida por:

$$\left\{ x \in \mathbb{R}^n : \langle a, x \rangle = a^T x = 0 \right\} = a^{\perp}.$$



Y todos los hiperplanos que serán paralelos a esa recta estarán dadas por:

$$\left\{ x \in \mathbb{R}^n : \langle a, x \rangle = a^T x = b \right\}.$$

Esto, nos dará dos semiespacios dados por:

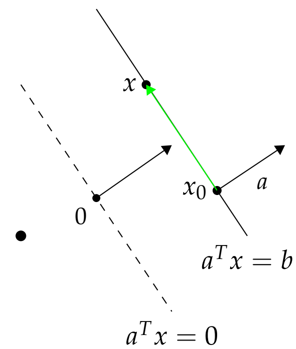
$$\left\{ x \in \mathbb{R}^n : a^T x \leq b \right\}$$

$$\left\{ x \in \mathbb{R}^n : a^T x > b \right\}$$

Esto nos divide el espacio en dos trozos. Que es la estrategia fundamental de análisis de datos. Por ejemplo cuando marcamos con lineas cuando existen datos por arriba y por abajo.

- Un hiperplano en $\mathbb{R}^2$ será sencillamente las rectas.
- Esa recta me va a definir dos semiespacios uno al lado del otro.
- Nos interesará desplazar esa recta que contiene al 0.
- La b nos dará una notación de distancia entre los hiperplanos.

Ejemplo 2.4



Para decidir en que dirección estará el punto, debemos tomar en cuenta a que lado apunto a, que nos marcará un punto perpendicular a ese conjunto. De donde,

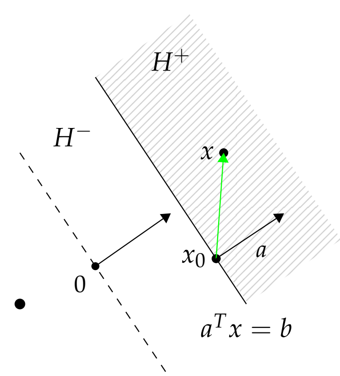
$$\langle (x - x_0), a \rangle = 0.$$

Si queremos en producto matricial se tiene,

$$\begin{aligned} a^T(x - x_0) &= 0 \\ a^T x - a^T x_0 &= 0 \\ a^T x &= a^T x_0 \\ a^T x &= b \end{aligned}$$

- Estas aplicaciones la llaman también aplicaciones del dual.

Ejemplo 2.5



El angulo de $(x - x_0, a)$ esta entre $-90°$ y $90°$. En términos de cosenos sería:

$$\cos[\text{ang}(x - x_0, a)] \in [0, 1]$$

Luego,

$$0 \leq \cos\left[\text{ang}(x - x_0, a)\right] = \frac{\langle (x - x_0), a\rangle}{\|x - x_0\|_2 \|a\|_2}$$

Por lo tanto,

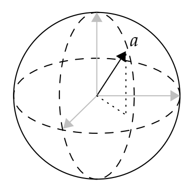
$$\begin{aligned} x \in H^+ \quad &\Leftrightarrow \quad \langle (x - x_0), a\rangle = \cos\left[\text{ang}(x - x_0), a\right] \|x - x_0\|_2 \|a\|_2 \geq 0 \\ &\Leftrightarrow \quad \langle (x - x_0), a\rangle \geq 0 \\ &\Leftrightarrow \quad \langle x, a\rangle \geq \langle x_0, a\rangle \\ &\Leftrightarrow \quad a^T x \geq a^T x_0 \\ &\Leftrightarrow \quad a^T x \geq b \end{aligned}$$

Ahora, si $x \in H^-$, entonces $a^T x \leq b$.

2.2 Bolas Euclideas

Tenemos que,

$$\begin{aligned} B(c, r) &= \{x \in \mathbb{R}^n : \|x - c\|_2 < r\} \\ &= \left\{x \in \mathbb{R}^n : (x - c)^T (x - c) < r^2\right\}. \end{aligned}$$



Ejercicio 2.2 Demostrar que $B(c, r)$ es convexo.

Demostración.- Sean x_0, x_1 en $B(c, r)$ y $\lambda \in [0, 1]$. Demostraremos que

$$(1 - t)x_0 + \lambda x_1$$

esta también en $B(c, r)$. Primero, notemos que

$$\|x_0 - c\|_2 < r \quad \text{y} \quad \|x_1 - c\|_2 < r.$$

Luego, por la definición de convexidad, y por la desigualdad triangular,

$$\begin{aligned} \|\left[(1 - \lambda)x_0 + \lambda x_1\right] - c\|_2 &= \|\lambda(x_0 - c) + (1 - \lambda)(x_1 - c)\|_2 \\ &\leq \lambda\|x_0 - c\|_2 + (1 - \lambda)\|x_1 - c\|_2 \\ &< \lambda r + (1 - \lambda)r. \end{aligned}$$

Por lo tanto,

$$\|(1 - \lambda)x_0 + \lambda x_1 - c\|_2 < r.$$

Concluimos que, $B(r, c)$ es convexo. (La demostración se basó en el libro de Boyd). ■

Propiedad .2 Mediante la suma de Minskowsky tenemos,

$$B(c,r) = c + rB(0,r).$$

- Si la bola está al rededor del cero u otro punto, la bola es la misma. Esto en distancias no tiene porque ser cierto.
- Todas las bolas que podamos dibujar podremos representarlos con centro cero, ya sean grandes o pequeñas.

2.3 Elipsoides

$$\begin{aligned}
\mathcal{E} &= \left\{x \in \mathbb{R}^n : (x-c)^T P^{-1}(x-c) \le 1\right\} && (1)\\
&= c + \left\{y \in \mathbb{R}^n : y^r P^{-1} y \le 1\right\} \qquad (y = x - c) && (2)\\
&= c + \left\{y \in \mathbb{R}^n : y^T LDL^T y \le 1\right\} && (3)\\
&= c + \left\{Lz : z^T Dz \le 1\right\} \qquad (z = L^T y) && (4)\\
&= c + L\left\{z : z^T Dz \le 1\right\}. && (5)
\end{aligned}$$

donde

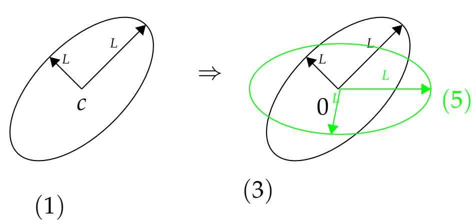
$$P = P^T > 0 \text{ (Simétrica y valores propios } > 0).$$

y

$$P^{-1} = LDL^T, \text{ Si } P \in \mathcal{M}_{n\times n}.$$

donde D es una matriz diagonal con valores propios de $1/P$.

- Cómo es simétrica y tiene valores propios positivos, se puede utilizar la diagonalización y escribir la matriz P cómo producto de una matriz diagonal por dos matrices de cambio que en realidad son ortonormales.
- $z^T Dz$, se puede hacer más grande o mas pequeña.
- Los vectores de L me dan los vectores que apuntan a la elipse.
- $z^T Dz$ es el circulo.
- D serán las curvaturas principales de la elipse.
- La L gira la elipsoide.
- Los elipsiodes se manejan para manejar imagenes donde incluye un objeto.



2.4 Bolas generales y conos asociados

Definición 2.12 **Norma.** $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}^+$.

Algunas condiciones:

i) $\|x\| = 0 \Leftrightarrow x = 0$.

ii) Si la norma de un vector se eleva al cuadrado, se esperara que la longitud sea el doble.

$$\|\lambda x\| = |\lambda|\|x\| \quad \forall \lambda \in \mathbb{R}.$$

iii) La desigualdad triangular.

$$\|x+y\| \leq \|x\| + \|y\|.$$

Recordemos que,

$$\begin{aligned} B(c,r) &= \{x \in \mathbb{R}^n : \|x-c\| \leq r\} \\ &= \left\{x \in \mathbb{R}^n : d_{\|\cdot\|}(x,c) \leq r\right\}. \end{aligned}$$

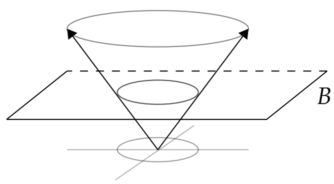
Donde $B(c,r)$ es convexa.

2.4.1 Cono asociado a una norma

- Una vez que conoces la $B(0,1)$ se conoce todas las demás.
- No se puede derivar en el origen, ya que estará en el pico del cono.
- Podemos analizar todo lo que está dentro del cono,
- A esto se le llama epigrafo de la función, es decir dibujar la función y pintar todo lo que hay arriba.
- La definición del cono es cerrada.

Definición 2.13

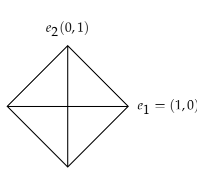
$$C_{\|\cdot\|} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \|x\| \leq t\}.$$



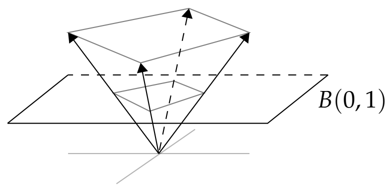
$$\left\{(x,t) : t = 1 \cap C_{\|\cdot\|}\right\} = \overline{B(x,t)}.$$

Tipos de norma

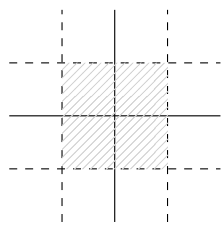
a) $\|\cdot\|_1 : \mathbb{R}^2 \to \mathbb{R}$, $\|(x,y)\|_1 = |x| + |y|$. Si lo definimos en $\mathbb{R}^8$ las sumas serán la suma de sus componentes. Si queremos dibujar $\overline{B(0,1)}$



Donde, e_1 y e_2 se llaman extremales.



b) $\|\cdot\|_\infty : \mathbb{R}^n \to \mathbb{R}$, $\|x,y\|_\infty = \max\{|x|,|y|\}$. Compara y coge la más grande. Veamos una vista transversal con respecto del cono:

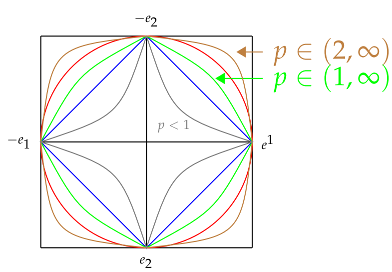


c) $\|\cdot\|_P : \mathbb{R}^n \to \mathbb{R}$, $p \in (1,\infty)$. (Existe una especie de promedio)

$$\|(x,y)\| = \sqrt[p]{|x|^p + |y|^p}$$

En particular:

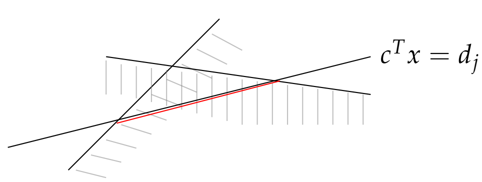
$$\|(x,y)\|_2 = \sqrt{|x|^2 + |y|^2}.$$



- La norma infinito es cómo la madre de todas las demás normas. Representamos con un cuadrado sobre la base canónica.
- La norma 1 sera el rombo que ya dibujamos.
- La Euclídea será un círculo.
- Cuando $p < 1$ no será una función convexa. (No se cumplirá la desigualdad triangular). Ya no será una bola.

2.5 Poliedros

$$P = \left\{ x \in \mathbb{R}^n : a_j^T x \le b_j, j = 1,\ldots,n, c_j^T x \le d_j, k = 1,\ldots,p \right\}.$$



- La idea de los semiespacios afines, de tomar un hiperplano de un lado y el otro, en realidad se puede hacer con hiperplanos o cualquier subespacio afín.
- Un polihedro será prácticamente una cantidad finita de caras. Es decir, son hiperespacio de dimensión $n-1$ que determinan la frontera, tomando desigualdades $a_j^T x \le b_j$. El cual es un semiespacio determinado por la dirección del hiperplano a_j. Y el b_j es una traslación.

El poliedro generaliza:

- Linea,
- segmento,
- semiespacio,
- subespacio afín.

Sin embargo no todo conjunto convexo se puede definir cómo un poliedro.

- Cuando se pide varias condiciones es la intersección de semiespacios.
- $c_j^T x = d_j$, me fija hiperplanos, que justo corte por un lugar específico.
- Los poliedros son aplicados a optimización lineal.

2.6 Operaciones que conservan la convexidad

Algunas propiedades que conservan la convexidad:

1. Intersección de convexos es convexo.
2. Si tenemos una función $f : \mathbb{R}^n \to \mathbb{R}^m$, f afín. Es decir,
$$(f(x) = Ax + b, A \in \mathcal{M}_{m\times n} \text{ y } b \in \mathbb{R}^m)\,.$$
3. Si tengo un conjunto $C \subseteq \mathbb{R}^n$ convexo. Entonces, $f(c)$ es convexo. Si $B \subseteq \mathbb{R}^n$ convexo, entonces $(f^A(B))$ la anti-imagen de B es también convexo.

- Las aplicaciones afines cuando trabajamos en $\mathbb{R}^n$ y $\mathbb{R}^m$ son tan sencillas como multiplicar una matriz y sumar un número. Es una aplicación lineal y sumar una constante, es decir trasladar.
- El hecho que funcione para adelante y para atrás, nos permite que para que una función sea convexa yo puedo demostrar que su imagen de f es convexa. Donde se nos simplifica las cosas.
- Si un convexo esta lejos del cero, dado que la traslación es afín, podemos trasladar a cero demostrarlo y llevarlo a su estado original. Así sin perdida de generalidad podemos asumir que el cero está en el conjunto.

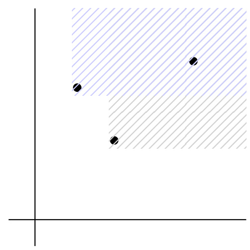
Algunos ejemplos particulares de afín:

- Homotecias: Multiplicar por un escalar.
- Translaciones.
- Proyecciones: Aplicaciones lineales importantes.
- Suma de Minskowsky. Empezamos con dos conjuntos y definimos la suma de Minkowski como el conjunto de los A, B tal que la A está en A y la A en B, lo podemos ver como la imagen aplicación afín de un conjunto convexo.
- Producto de dos conjuntos.

- Si 2., entonces se cumple 3. Lo que se puede ver es que los primeros tres ejemplos son estables, la imagen por una homotecia o la anti-imagen, etc.
- La suma de dos cosas convexas es convexa.
- El producto cartesiano de dos cosas convexas es convexa.
- Estas dos últimas son un poco más difíciles de demostrar porque hay que pensar que funciones afines nos da la suma de Minskowsky y que función afín nos da el producto de dos conjuntos. En realidad, se utiliza la pre-imagen.

2.7 Desigualdad generalizadas

¿Cual de los dos puntos es más importante?. Es los puntos donde están pintados, dependiendo si es máximo o mínimo.



Ahora, cual es mejor de estos dos conjuntos.

Esto se define como:

$$\begin{aligned} x, y \in \mathbb{R}^n, x \leq y \quad &\Leftrightarrow \quad x_i \leq y_i, \quad i = 1, \ldots, n \\ &\Leftrightarrow \quad 0 \leq y_i - x_i, \quad i = 1, \ldots, n \\ &\Leftrightarrow \quad y - x \in \mathbb{R}^n_+ \\ &\Leftrightarrow \quad y \in x + \mathbb{R}^n_+. \quad \text{(Minkowski)} \end{aligned}$$

Ahora, en vez de utilizar $\mathbb{R}^n_+$ puede utilizarse otro conjunto K. Ahora, ¿Qué propiedades debería tener K, para tener un conjunto con orden?. Necesito asignar una serie de propiedades que tiene $\mathbb{R}^n$.

Por lo que definimos de cono propios

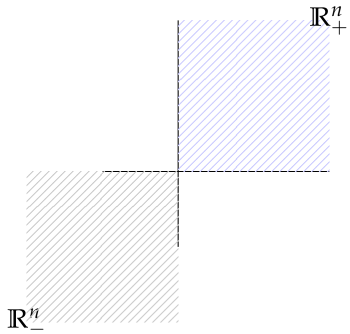
Definición 2.14 $K \subseteq \mathbb{R}^n$ es cono propio si es un cono:

i) Convexo: Cualquier parte de puntos el segmento estará en el mismo conjunto. Buen comportamientos de las SUMAS.

ii) Cerrado: Vamos a definir con el menor o igual.

iii) Sólido ($\text{int}(K) \neq \emptyset$): Que pueda decidir en todo $\mathbb{R}^n$. Para ello, debo tener al menos un punto interior.

iv) Apuntado ($x \in K, -x \in K \Rightarrow x = 0$): No me deja que contenga una recta entera.



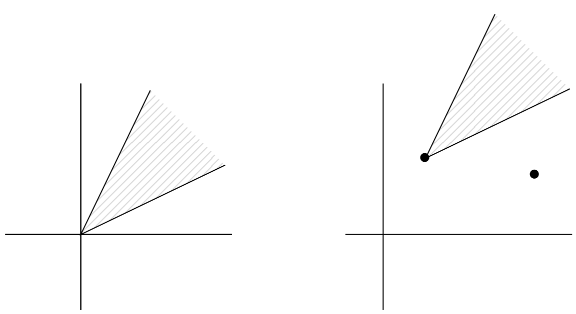
- Está claro que es convexo.
- Ya que no se tiene dos $++$, en $\mathbb{R}^n_+$. Entonces es cerrado.
- Sólido, ya que existe cualquier punto interior. Los bordes no son puntos interiores.
- Apuntado, ya que, cuando tengo un punto tiene opuesto. Básicamente tiene que pasar siempre en cero.

Dado un cono propio $K \subseteq \mathbb{R}^n$ se define $\leq_K$ un orden en $\mathbb{R}^n$ como:

$$\begin{aligned} x \leq_K y \quad &\Leftrightarrow \quad y - x \in K \\ &\Leftrightarrow \quad y \in x + K. \end{aligned}$$

- Cada K que fijemos será una lección de multicriterio.
- Si estoy en $\mathbb{R}^n$ por ejemplo, entonces tengo cinco conos para elegir, y según un criterio de esas 5 variables definiremos cual es mejor o cual es peor.
- Estos criterios lo puedo definir mediante conos.
- Me asegura un sistema sistemático de orden.

Ejemplo: Si tomamos dos puntos. Entonces,



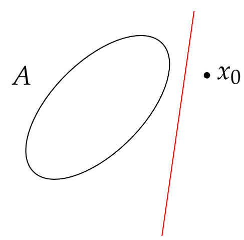
Aquí no puedo asegurar que $x \nleq_K y$.

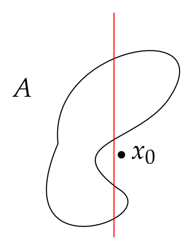
Ejercicio 2.3 Demuestra que $\leq_K$ es un orden.

i) Reflexiva, $x \leq_K x \; \forall x$.
ii) Transitiva, $x \leq_K y, y \leq_K z \Rightarrow x \leq_K z$.
iii) Antisimétrica, $x \leq_K y, t \leq_k x \Rightarrow x = y$..
iv) Estable para sumas: $x \leq_K y, z \leq_K w \Rightarrow y \pm w$.
v) Estable para productos positivos: $x \leq_K y \Rightarrow \lambda x \leq_K \lambda y, \lambda > 0$.
vi) Estable para límites: $x_n \leq_K y, x_n \to x \Rightarrow x \leq_K y$.

Demostración.- Para demostrar debemos utilizar las propiedades de cono propio. Lo que le pido que al orden se comporte bien con las sumas, con los productos por escalares positivos y que se comporte bien con los límites por lo que pido que el cono sea cerrado. ■

2.8 Teorema de separación y extensión





- La idea fundamental es que un conjunto convexo siempre se puede aproximar por tangentes.
- Los hiperplanos nos ayudaran a cortar el espacio, donde a un lado nos dejaran al conjunto y al otro nada.
- Moviendo esos hiperplanos, alrededor del conjunto podemos recuperar toda la forma del conjunto.
- Esto nos dan pie a unos teoremas que nos dice que si tenemos un conjunto y un punto que está fuera, yo siempre puedo encontrar una linea recta que los separa.
- Si el conjunto no es convexo y el punto a separar está en la región "no convexa". Entonces, no podemos separarlas con hiperplanos o semiespacios.

Lema 2.1 **Lema de Zorn.** Si tengo un conjunto $(\mathscr{A}, \leq)$ parcialmente ordenado, tal que si $\mathscr{C} \subseteq \mathscr{A}$ totalmente ordenado $(x, y \in \mathscr{C} \Rightarrow x \leq y$ ó $y \leq x)$, tiene una cota superior en $\mathscr{A}$ $(\exists a \in \mathscr{A} : x \leq a \; \forall x \in \mathscr{C})$. Entonces, $\mathscr{A}$ tiene un maximal $(\exists a \in \mathscr{A} : b \in \mathscr{A},\; b \leq a \Rightarrow b = a)$. $[\mathscr{A} \notin \emptyset]$. ■

- El conjunto $\mathscr{A}$ es no vacío.
- Se toma un conjunto $\mathscr{A}$ parcialmente ordenado, dentro un conjunto $\mathscr{C}$ todos en fila y este conjunto $\mathscr{A}$ tiene la propiedad de que cualquier conjunto de esa forma $\mathscr{C}$ que tome, siempre va a estar acotado superiormente.
- Maximal significa que no tiene un elemento por encima. Si hay otro por encima entonces es el mismo.

Este lema nos permite encontrar elementos maximales.

Lema 2.2 Si un espacio vectorial $H + [x_c]$ y unas funciones $g : H \to \mathbb{R}$ lineal y $p : H + [x_c] \to \mathbb{R}$ lineal. Cumple que $g(h) \leq p(h)$ $\forall h \in H$. Entonces, $\exists\; \overline{g} : H + [x_0] \to \mathbb{R}$ lineal, tal que $\overline{g}(h) = g(h)$ y $\overline{g}(x) = g(x)$ $\forall x \in H + [x_0]$.

Demostración.- Dado $g : H \to \mathbb{R}$ lineal, tal que $g(h) \leq p(h)$ $\forall h \in H$. Quiero definir $\overline{g} : H + [x_0] \to \mathbb{R}$ lineal y que $g(x) \leq p(x)$ $\forall x \in H + [x_0]$.

- Este lema es el mismo enunciado del teorema de abajo, pero no para cualquier subespacio sino para un espacio de co-dimensión. Yo se demostrar que si tengo una función en un hiperplano yo puedo extenderla a una dimensión.
- Con este lema 2.2 podemos realizar operaciones finitas sin necesitad de lema de Zorn.
- Si tenemos un espacio y tenemos un hiperplano y en el hiperplano tenemos las condiciones del teorema. Entonces somos capaces de encontrar una extensión a todo el espa-

Dado que $[x_0]$ es linealmente independiente a $g(h)$, por lo tanto

$$\overline{g}\,(h+\lambda x_0) = \overline{g}(h) + \lambda\overline{g}(x_0) = \alpha = g(h) + \lambda\alpha.$$

cio que sigue siendo lineal que extiende a la g y además está controlada por p.

Ahora, debemos definir $[x_0]$ para que funcione la cosa. Para ello buscamos $\alpha \in \mathbb{R}$ tal que $g(h) + \lambda x_0 \leq p(h) + \lambda x_0\ \forall h \in H,\ \forall \lambda \in \mathbb{R}$. Esto será equivalente A:

$$\left\{ \begin{array}{lcl} g(x)+\alpha & \leq & p(h+x_0),\ \forall h \in H \\ g(h)-\alpha & \leq & p(h-x_0),\ \forall h \in H \end{array} \right. \qquad (1)$$

$$\Updownarrow$$

$$g(x) - p(x-x_0) \leq \alpha \leq p(y-x_0) - g(y)\ \forall x, y \in H.$$

Se puede poner x, y en vez h ya que las desigualdades de (1) son independientes. ■

- Extender funciones lineales es sumamente sencillo: Por ejemplo si estamos en $\mathbb{R}^3$ y tenemos una función lineal definida en el plano, lo único que tenemos que hacer es tomar una base del plano, completarla a la base del espacio y en ente caso como es $\mathbb{R}^3$ y $\mathbb{R}^2$, sería solamente un vector. Entonces, definimos una aplicación lineal que coincida con g en el plano y en el vector que sobra le damos el valor cero.

Teorema 2.3 **Teorema de Hahn-Banach (Extensión o separación.)** Sea E un espacio vectorial y $p : E \to \mathbb{R}^+$ sublineal. Sea $E_0 \subseteq E$ y $g : E_0 \to \mathbb{R}$ lineal y $g(x) \leq p(x)\,\forall x \in E_0$. Entonces existe $\overline{g} : E \to \mathbb{R}$ lineal, tal que $\overline{g}(x) = g(x)\,\forall x \in E_0$ ($\overline{g}$ extiende a g) y $\overline{g}(x) \leq p(x),\ \forall x \in E$.

- La idea es extender combinaciones lineales.
- Es **sublineal** si $p(x+y) \leq p(x) + p(y)$ y $p(\lambda x) = \lambda p(x)$ si $\lambda > 0$.
- La función g lineal es controlada por esa función norma p. Quiere decir que es continua en la métrica g de p.
- La idea es la siguiente: tengo un subespacio, una función lineal definida en el subespacio y que está contralada por p.
- Yo tengo una distancia en todo mi espacio, una aplicación lineal definida abajo y puedo extenderla.

Demostración.- Si yo quiero comenzar por una función g, en un espacio pequeño y quiero extenderlo a un espacio grande, debo comenzar por tomar todas las distribuciones.

Empezamos definiendo de objetos, en este caso extensiones como,

$$\mathscr{F} = \left\{ (E_1, g_1) : \begin{array}{l} g_1 : E_1 \to \mathbb{R} \text{ lineal} \\ g_1(x) = g(x)\ \forall x \in E_0 \\ E_0 \subseteq E_1 \\ g_1(x) \leq p(x)\ \forall x \in E_1. \end{array} \right\}$$

Lo primero que justifico es el vacío. Es decir,

$$\mathscr{F} \notin \emptyset, \text{ ya que } (E_0, g) \in \mathscr{F}.$$

Ahora, introducimos un orden en esta familia, de forma que una de estas extensiones sea mejor que otra. Dos los dos elementos definidos,

$$(E_1, g_1) \leq (E_2, g_2).$$

- $\mathscr{F}$ es no vacío porque la función inicial g y el subespacio E_0 están ahí.
- La idea de este teorema se basa en que tengo una g que la extiendo por muchas ramas, donde cada extensión las puedo extender de nuevo por más ramas. Pero cuando tomo la cadena totalmente ordenado solo me estaré yendo por una rama. Donde

Al definir esto, no decimos que todos los elementos se puedan comparar. Lo que decimos es que si yo tomo estos elementos se cumplirán siempre que se cumpla las siguientes condiciones:

las g van coincidiendo a través de las ramas. Básicamente la $\mathscr{F}$ es un árbol y la g es una rama bien ordenada.

- $E_1 \subseteq E_2$.
- $g_1(x) = g_2(x)\ \forall x \in E_1$.

El conjunto

$$(\mathscr{F}, \leq)$$

se llama parcialmente ordenado.

Ahora demostraremos que que este conjunto parcialmente ordenado, toda cadena $\mathscr{C} \subseteq \mathscr{A}$ tiene una cota superior, y por el lema de Zorm tiene un elemento maximal.

Si $\mathscr{C} \subseteq \mathscr{F}$ es una cadena totalmente ordenado. Lo que queremos encontrar un elemento no necesariamente en $\mathscr{C}$ pero si en $\mathscr{F}$ que sea una cota superior. Para ello,

$$\mathscr{C} = \{(E_i, g_i)\}_{i \in I}$$

Ahora, definimos

$$E_{\mathscr{C}} = \bigcup_{i \in I} E_i.$$

Dados dos que yo fije, uno mayor que otro, siempre tendré inclusión. Por eso nos sirve la unión. Porque están contenido uno debajo del otro.

Definimos también

$$g_{\mathscr{C}} : E_{\mathscr{C}} \to \mathbb{R}.$$

Lo que tendrá que coincidir con las g_i. Por lo que,

$$g_{\mathscr{C}}(x) = g_i(x) \text{ si } x \in E_i.$$

Dado que g_i cumplen $g_1(x) \leq p(x)\ \forall x \in E_1$. Entonces, en lo particular

$$g_{\mathscr{C}}(x) \leq p(x)\ \forall x \in E_{\mathscr{C}}.$$

Así,

$$\left(E_{\mathscr{C}, g_{\mathscr{C}}}\right) \in \mathscr{F}.$$

Y además,

$$(E_{\mathscr{C}}, g_{\mathscr{C}}) \geq (E_i, g_i)\ \forall i \in I.$$

Donde $\left(E_{\mathscr{C}, g_{\mathscr{C}}}\right)$ es una cota superior de $\mathscr{C}$.

Por lo tanto, por el lema de Zorm existe $(\overline{E}, \overline{g}) \in \mathscr{F}$ que es un elemento maximal de $\mathscr{F}$.

Ahora, ¿qué haremos con ese elemento maximal?. Lo ideal sería que $\overline{E}$ sea todo el E, pero podría ser que no. ¿Qué pasa si $\overline{E}$ se encuentra por debajo de E?. En otras palabras, no llegue a extender a todo el espacio vectorial.

Ahora, demostraremos que $\overline{E}$ es efectivamente la E. Para ello, utilizaremos una reducción al absurdo. Supongamos que $\overline{E} \subsetneq E$. Entonces existe un $x_0 \in \overline{E} \backslash E$. Por lo que puedo ampliar $\overline{E}$ de la siguente forma:

Cuando tomamos $[x_0]$ quiere decir que tomo la linea generada por x_0 (span).

$$\overline{E} \subsetneq \overline{E} + [x_0] \subseteq E.$$

Lo que intentaremos probar por contradicción es que si $\overline{E}$ no fuese E. $\overline{E}$ no puede ser un elemento maximal. Por el lema 2.2, aplicado a $H = \overline{E}$ y a $\overline{g} : E \to \mathbb{R}$. Existe una extensión $\overline{g} : E + [x_0] \to \mathbb{R}$ tal que

$$\overline{g}(x) \leq p(x) \ \forall x \in \mathbb{E} + [x_0].$$

Ahora,

$$\left(\overline{E} + [x_0], \overline{g}\right) \in \mathscr{F} \lneq \overline{E}, \overline{g}.$$

Por lo que es imposible pues $\left(\overline{E}, \overline{g}\right)$ es un elemento maximal. Luego, $\overline{E} = E$ y $\left(\overline{E}, \overline{g}\right)$ es la extensión buscada. ■