Teoremas de Hanh-Banach y de Mazur-Orlicz

El objetivo principal de esta sección es demostrar tanto una versión no tan conocida del clásico teorema de Hanh-Banach como el teorema de Mazur-Orlicz. Para ello, iremos utilizando una serie de lemas previos que nos facilitarán el proceso. Así, una vez vistos estos resultados, habremos construido la base para el lema de Simons.

En primer lugar vamos a recordar la definición de funcional sublineal sobre un espacio vectorial V. Notar que todos los espacios vectoriales que vamos a usar son reales. Del mismo modo, los espacios y conjuntos que usaremos asumiremos que son no vacíos.

Definición 1.1. *Sea V un espacio vectorial. Decimos que el $P : V \to \mathbb{R}$ es sublineal si cumple las siguientes condiciones:*

- *P es subaditiva: $x_1, x_2 \in V \Longrightarrow P(x_1 + x_2) \leq P(x_1) + P(x_2)$.*
- *P es positivamente homogénea: $x_1 \in V$ y $\lambda > 0 \Longrightarrow P(\lambda x) = \lambda P(x)$.*

Como consecuencia, podemos afirmar que $P(0) = 0$. En efecto:

$$P(0) = P(2 \times 0) = 2 \times P(0) \Longrightarrow P(0) = 0.$$

Por ejemplo, toda norma o seminorma sobre V es un funcional sublineal. Así, dados $a, b \in \mathbb{R}$ con $a < b$ tenemos que $P : H^1(a, b) \longrightarrow \mathbb{R}$ definida como $P(x) = \|x'\|_{L^2(a,b)}$ es una seminorma y por ello sublineal. También, si V= $\mathbb{R}$ y definimos la parte positiva $P(x) = (x)^+ = \text{máx}\{0, x\}, \forall x \in \mathbb{R}$ obtenemos un funcional sublineal sobre $\mathbb{R}$.

El lema que exponemos a continuación, y que generalizaremos posteriormente en el lema 1.2, nos servirá para demostrar el teorema de Hanh-Banach.

Lema 1.1. *Sea V un espacio vectorial y $P: V \to \mathbb{R}$ un funcional sublineal. Fijamos un elemento $y \in V$. Para todo $x \in V$ tomamos*

$$P_y(x) := \text{ínf}_{\lambda>0} [P(x + \lambda y) - \lambda P(y)]$$

Entonces, $P_y : V \to \mathbb{R}$, P_y es sublineal, $P_y \leq P$ y $P_y(-y) \leq -P(y)$.

Demostración. Fijamos $y \in V$. Sea $x \in V$ y $\lambda > 0$. Como P es sublineal tenemos:

$$\lambda P(y) = P(\lambda y) = P(\lambda y + x - x) \leq P(x + \lambda y) + P(-x).$$

Por lo tanto, se obtiene que $P(x + \lambda y) - \lambda P(y) \geq -P(-x)$. Tomando el ínfimo sobre $\lambda > 0$ llegamos a $P_y(x) \geq -P(-x) > -\infty$. Por consiguiente, $P_y : V \to \mathbb{R}$.

Probaremos ahora que P_y es sublineal. Empezamos viendo la subaditividad. Tomamos $x_1, x_2 \in V$ y sean $\lambda_1, \lambda_2 > 0$ arbitrarios. Entonces:

$$\begin{aligned}
[P(x_1 + \lambda_1 y) - \lambda_1 P(y)] + [P(x_2 + \lambda_2 y) - \lambda_2 P(y)]& \\
&\geq [P(x_1 + x_2 + (\lambda_1 + \lambda_2) y)] - (\lambda_1 + \lambda_2) P(y) \\
&\geq P_y(x_1 + x_2).
\end{aligned}$$

Tomando ínfimo sobre λ_1 y λ_2, $P_y(x_1) + P_y(x_2) \geq P_y(x_1 + x_2)$. Así, P_y es subaditiva. Para comprobar que es positivamente homogénea tomamos $x \in V$ y $\mu > 0$. Entonces:

$$\begin{aligned}
P_y(\mu x) &= \inf_{\lambda>0} [P(\mu x + \lambda y) - \lambda P(y)] = \mu \inf_{\lambda>0} [P(x + (\lambda/\mu) y) - (\lambda/\mu) P(y)] \\
&= \mu \inf_{\upsilon>0} [P(x + \upsilon y) - \upsilon P(y)] = \mu P_y(x).
\end{aligned}$$

Obtenemos que P_y es positivamente homogénea y como consecuencia sublineal.

Para demostrar que $P_y \leq P$, sea $x \in V$ y tomando $\lambda = 1$:

$$P_y(x) \leq P(x + y) - P(y) \leq P(x) + P(y) - P(y) = P(x) \Longrightarrow P_y \leq P.$$

Como $x \in$V es arbitrario, entonces $P_y \leq P$. Finalmente, actuando de manera similar al caso anterior:

$$P_y(-y) \leq P(-y + y) - P(y) = -P(y).$$

∎

Ahora procedemos a probar el teorema de Hanh-Banch para funcionales sublineales, el cual es uno de los resultados más importantes del análisis funcional.

Teorema 1.1 (Hanh-Banach)**.** *Sea V un espacio vectorial y $P : V \to \mathbb{R}$ un funcional sublineal. Entonces existe un funcional lineal L en V tal que $L \leq P$.*

Demostración. Sea $SUB = \{Q : \mathrm{V} \longrightarrow \mathbb{R} : Q \leq P\}$, es decir, el conjunto de funcionales sublineales sobre V que son menores o iguales que P. Nuestro propósito es emplear el lema de Zorn. Para ello, dados $T_1, T_2 \in \mathcal{Q}$ consideramos la relación de orden usual, es decir:

$$T_1 \leq T_2 \Longleftrightarrow T_1(x) \leq T_2(x) \quad \forall x \in \mathrm{V}.$$

Primero probaremos que todo subconjunto $\mathcal{T}$ totalmente ordenado de SUB tiene una cota inferior en SUB. Definimos $Q(x) := \text{ínf}\{T(x) : T \in \mathcal{T}\}$ y queremos ver que $Q(x) \in \mathbb{R}$. Si $x \in$V y $T \in \mathcal{T}$, como T es subaditiva obtenemos la siguiente relación:

$$0 = T(0) = T(x - x) \leq T(x) + T(-x) \Longrightarrow T(x) \geq -T(-x) \quad (1)$$

Por otro lado:

$$T \in \mathcal{Q} \Longrightarrow T(x) \leq P(x) \Longrightarrow -T(x) \geq -P(x) \quad (2)$$

Usando (1), (2) y tomando ínfimo sobre T llegamos a $Q(x) \geq -P(x) \geq -\infty$. Por lo tanto $Q : V \to \mathbb{R}$.

Ahora probaremos que Q es subaditiva. Para ello, tomamos $x_1, x_2 \in$V. Sean $T_1, T_2 \in \mathcal{T}$ arbitrarios. Si $T_1 \geq T_2$ (el caso de $T_2 \geq T_1$ es análogo.):

$$T_1(x_1) + T_2(x_2) \geq T_2(x_1) + T_2(x_2) \geq T_2(x_1 + x_2) \geq Q(x_1 + x_2).$$

Concluimos (en ambos casos) que $T_1(x_1) + T_2(x_2) \geq Q(x_1 + x_2)$. Tomando ínfimo en T_1 y T_2 obtenemos que $Q(x_1) + Q(x_2) \geq Q(x_1 + x_2)$. Así, Q es sublineal. Que sea positivamente homogénea es consecuencia de que T también lo es. Dado $\mu > 0$:

$$\begin{aligned} Q(\mu x) &= \text{ínf}\{T(\mu x) : T \in \mathcal{T}\} \\ &= \text{ínf}\{\mu T(x) : T \in \mathcal{T}\} \\ &= \mu\, \text{ínf}\{T(x) : T \in \mathcal{T}\} \\ &= \mu Q(x). \end{aligned}$$

De este modo, Q es sublineal y como es claro que $Q \leq P \Longrightarrow Q \in SUB$. Así, es directo que Q es el elemento minimal de $\mathcal{T}$ en SUB.

El lema de Zorn nos proporciona entonces un elemento minimal de SUB que llamaremos L. Vamos a comprobar que L es lineal y, por tanto, es el funcional buscado. Tomamos ahora $y\in$V. Con la notación del lema anterior, L_y :V$\longrightarrow \mathbb{R}$ es sublineal, $L_y \leq L$ (como consecuencia $L_y \in SUB$) y $L_y(-y) \leq L(-y)$. De hecho, como L es minimal en SUB, $L_y = L$ y por ello $L(-y) \leq L(-y)$. Por otro lado, como L es subaditiva, $L(-y) \geq -L(y)$. Combinando ambas desigualdades, $L(-y) = -L(y)$. Tomamos $x\in$V y $\lambda < 0$, usando la igualdad anterior llegamos a:

$$L(\lambda x) = L(-(-\lambda)x) = -L(-\lambda x) = -(-\lambda)L(x) = \lambda L(x)$$

obteniendo que L es homogénea. Si $x_1, x_2 \in$V, la subaditividad de L nos da $L(-x_1 - x_2) \leq L(-x_1) + L(-x_2)$. Usando la homogeneidad de L:

$$\begin{aligned} L(x_1 + x_2) &= L(-(-x_1 - x_2)) = -L(-x_1 - x_2) \\ &\geq -L(-x_1) - L(-x_2) = L(x_1) + L(x_2) \geq L(x_1 + x_2). \end{aligned}$$

Por ello, $L(x_1 + x_2) = L(x_1) + L(x_2)$ y por la arbitrariedad de $x_1, x_2 \in$V concluimos que L es lineal.

■

El siguiente resultado importante que demostraremos será el teorema de Mazur-Orlicz. Primero, veamos un lema previo.

Lema 1.2. *Sea V un espacio vectorial y $P : V \rightarrow \mathbb{R}$ un funcional sublineal. Sea D un subconjunto convexo de V y sea $\beta := \inf_D P \in \mathbb{R}$. Para todo $x \in V$ tomamos*

$$Q(x) := \inf_{d\in D, \lambda>0} [P(x + \lambda d) - \lambda\beta]$$

Entonces, $Q : V \rightarrow \mathbb{R}$, Qes sublineal, $Q \leq P$ y $\forall d \in D, -Q(-d) \geq \beta$.

Demostración. Si $x\in$V, $d \in D$ y $\lambda > 0$, entonces

$$P(x + \lambda d) - \lambda\beta \geq -P(-x) + \lambda P(d) - \lambda\beta \geq -P(-x) \geq -\infty.$$

La primera igualdad se deduce de la linealidad de P ya que:

$$\begin{aligned} \lambda P(d) &= P(\lambda d) \\ &= P(\lambda d + x - x) \\ &\leq P(x + \lambda d) + P(-x) \end{aligned}$$

por lo que $\lambda P(d) - P(-x) \leq P(x + \lambda d)$. La segunda se debe a que

$$\beta = \inf_D P \Longrightarrow \lambda P(d) \geq \lambda\beta \Longrightarrow \lambda P(d) - \lambda\beta \geq 0.$$

Tomando el ínfimo sobre $d \in D$ y $\lambda > 0$ llegamos a

$$Q(x) \geq -P(-x) > -\infty$$

por lo que $Q : V \to \mathbb{R}$. Es relativamente fácil probar que Q es positivamente homogénea por lo que para ver que es sublineal solo queda comprobar la subaditividad. Para ello, tomamos $x_1, x_2 \in V$. Sean $d_1, d_2 \in D$ y $\lambda_1, \lambda_2 > 0$ arbitrarios. Para simplificar la notación llamamos $x := x_1 + x_2$, $\lambda := \lambda_1 + \lambda_2$ y $d := (\lambda_1/\lambda)d_1 + (\lambda_2/\lambda)d_2$. Notar que $d \in D$ al ser este convexo. Entonces:

$$\begin{aligned}[P(x_1 + \lambda_1 d_1) - \lambda_1\beta] + [P(x_2 + \lambda_2 d_2) - \lambda_2\beta] &\geq P(x + \lambda_1 d_1 + \lambda_2 d_2) - \lambda\beta \\ &= P(x + \lambda d) - \lambda\beta \qquad . \\ &\geq Q(x) = Q(x_1 + x_2)\end{aligned}$$

Tomando ínfimo sobre d_1, d_2, λ_1 y λ_2, $Q(x_1) + Q(x_2) \geq Q(x_1 + x_2)$. Así, Q es subaditiva y como consecuencia sublineal. Fijamos $d \in D$. Sea $x \in$V arbitrario. Entonces, $\forall\lambda > 0$, $Q(x) \leq P(x) + \lambda\,[P(d) - \beta]$. Tomando $\lambda \longrightarrow 0$, $Q(x) \leq P(x)$ y como consecuencia $Q \leq P$. Finalmente, sea $d \in D$ arbitrario y tomando $\lambda = 1$:

$$Q(-d) \leq Q(-d + d) - \beta = -\beta \Longrightarrow -Q(-d) \geq \beta.$$

■

Visto este lema, estamos preparados para ver el resultado que nos interesa:

> **Teorema 1.2** (Mazur-Orlicz). *Sea V un espacio vectorial y $P : V \to \mathbb{R}$ un funcional sublineal. Sea D un subconjunto no vacío y convexo de V . Entonces existe un funcional $L : V \longrightarrow \mathbb{R}$ lineal tal que $L \leq P$ e $\inf_D L = \inf_D P$.*

Demostración. Sea $\beta := \inf_D P$. En el caso de que $\beta = -\infty$ por el teorema de Hanh-Banach tenemos que $\exists L$ sobre V tal que es lineal y $L \leq P$. Así:

$$L \leq P \Longrightarrow inf_D L \leq \inf_D P = -\infty \Longrightarrow inf_D L = \inf_D P.$$

Supongamos entonces que $\beta \in \mathbb{R}$. Definimos el funcional auxiliar Q como en el lema 1.2. Del teorema de Hanh-Banach obtenemos que existe un funcional lineal L sobre V tal que $L \leq Q$ (como $Q \leq P$ tenemos que $L \leq P$). Sea $d \in D$, entonces:

$$L(d) = -L(-d) \geq -Q(-d) \geq \beta.$$

Tomando ínfimo sobre $d \in D$:

$$\inf_D L \geq \beta = \inf_D P.$$

Por otro lado, como $L \geq P$:

$$\inf_D L \leq \inf_D P.$$

Juntando ambas desigualdades obtenemos $\inf_D L = \inf_D P$. ■

Igualdad

Ahora, queremos obtener una igualdad que nos será de utilidad posteriormente para el Teorema de Separación. Recordemos dados E_1, E_2 dos espacios normados y $L : E_1 \longrightarrow E_2$ un operador lineal, entonces L es continuo si, y solo si, verifica la siguiente condición:

$$\exists \alpha > 0 : \|L(x)\| \leq \alpha \|x\| \quad \forall x \in E_1,$$

conocido como el espacio dual (topológico) de E. Consideramos el espacio vectorial dado por:

$$E^* = \{L : E \longrightarrow \mathbb{R} : L \text{ es lineal y continuo}\}.$$

Para todo $L \in E^*$ definimos su norma como:

$$\|L\| = \text{mín}\{\alpha > 0 : \|L(x)\| \leq \alpha \|x\| \quad \forall x \in E\}.$$

De este modo, podemos escribir:

$$\|L(x)\| \leq \|L\| \|x\|$$

siendo dicha desigualdad óptima. También podemos expresar su norma como el mínimo mayorante de un conjunto mayorado, es decir, el supremo:

$$\|L\| = \sup\{\|L(x)\| / \|x\| : \forall x \in E \setminus \{0\}\}.$$

Para $x \in E_1 \setminus \{0\}$ tenemos que $\|L(x)\| / \|x\| = \|L(x/\|x\|)\|$ y es claro que $\{x/\|x\| : x \in E_1 \setminus \{0\}\}$ es la esfera unidad de E que notamos como S_E. Si en vez de la esfera consideramos la bola unidad, B_E el supremo no varía. Efectivamente, si $x \in B_E$ se tiene que $x = \|x\| u$ con $u \in S_E$, y por ello $\|L(x)\| = \|x\| \|L(u)\| \leq \|L(u)\|$ ya que $\|x\| \leq 1$. De este modo, también tenemos que:

$$\|L\| = \sup_{x \in B_E} \|L(x)\|.$$

En este momento, estamos en disposición de enunciar y demostrar la igualdad que deseamos:

Proposición 2.1. *Dado un espacio normado E y $x \in E$, entonces se cumple que:*

$$\sup_{x^* \in B_{E^*}} x^*(x) = \|x\| \,. \tag{2.1}$$

Demostración. Sea $x_0 \in E$ y sea $D := \{x_0\}$ Consideramos el funcional

$$\begin{aligned} P : E &\longrightarrow \mathbb{R} \\ x &\longmapsto \|x\| \,. \end{aligned}$$

Es claro que P es sublineal y que D es convexo. Podemos aplicar el Teorema de Mazur-Orlicz, teorema 1.2, y obtenemos que existe un funcional $L : E \longrightarrow \mathbb{R}$ lineal tal que $L \leq P$ e $\operatorname{ínf}_D L = \operatorname{ínf}_D P$. Como $L \leq P$, entonces

$$\|L(x)\| \leq \|P(x)\| = \|x\| \quad \forall x \in E.$$

Concluimos que L es continua. Además, como tenemos que $L \in E^*$, llamamos $L = x^*$ y llegamos a que $\|L\| = \|x^*\| \leq 1$, es decri, $x^* \in B_{E^*}$. Por su parte, como $\operatorname{ínf}_D L = \operatorname{ínf}_D x^* = \operatorname{ínf}_D P$ y $D = \{x_0\}$ entonces, $x^*(x_0) = \|x_0\|$. De este modo, llegamos a que existe $x^* \in B_{E^*}$ tal que $x^*(x_0) = \|x_0\|$. Si tomamos cualquier elemento $y^* \in B_{E^*}$, entonces $\|y^*(x_0)\| \leq \|y^*\| \, \|x_0\| \leq \|x_0\|$ y podemos asegurar que:

$$\sup_{x^* \in B_{E^*}} x^*(x_0) = \|x_0\| \,.$$

Como $x_0 \in E$ es arbitrario, la desigualdad enunciada queda probada. ∎

Lema de Simons y Teorema de la Alternativa de Gordan

Este capítulo se centra en lema de Simons, el cual será de gran ayuda para demostrar el Teorema de la Alternativa de Gordan.

Antes de comenzar, necesitamos hacer la siguiente definición. Para agilizar la lectura, dado $N \in \mathbb{N}$ en las N-uplas y espacios vectoriales $\mathbb{R}^N$ supondremos que $N \geq 1$. También notaremos con caracteres en negrita a los elementos de espacios vectoriales reales, por ejemplo, $\boldsymbol{t} = (t_1, ..., t_N) \in \mathbb{R}^N$.

Definición 3.1. *Dado $N \in \mathbb{N}$ llamamos símplex unitario de $\mathbb{R}^N$ al conjunto de $\mathbb{R}^N$ dado por:*

$$\Delta_N := \left\{ \mathbf{t} \in \mathbb{R}^N : \sum_{i=1}^{N} t_i = 1, \;\; t_1, ..., t_N \geq 0 \right\}.$$

Recordamos ahora la noción de envolvente convexa de un conjunto cualquiera X, que notamos como $\mathrm{co}(X)$ y que se define como la intersección de todos los conjuntos convexos que contienen a X. Nótese que $\Delta_N = co\{e_1, ...e_N\}$.

Antes de continuar veamos que Δ_N es convexo y compacto:

- Convexo: tenemos que comprobar que dados $\boldsymbol{t}, \boldsymbol{s} \in \Delta_N$ y $\lambda \in [0, 1]$ entonces $\lambda \boldsymbol{t} + (1 - \lambda) \boldsymbol{s} \in \Delta_N$. En efecto, las coordenadas $\lambda t_i + (1 - \lambda) s_i \in \mathbb{R}^N$ verifican:

   i) $\lambda t_i + (1 - \lambda) s_i \geq 0$ para todo $i = 1, ..., N$, ya que $t_i, s_i \geq 0$ y $0 \leq \lambda \leq 1$.

ii)

$$\begin{aligned}\sum_{i=1}^{N}(\lambda t_i + (1-\lambda)s_i) &= \lambda \sum_{i=1}^{N} t_i + (1-\lambda)\sum_{i=1}^{N} s_i \\ &= \lambda + (1-\lambda) \\ &= 1,\end{aligned}$$

ya que $\sum_{i=1}^{N} t_i = \sum_{i=1}^{N} s_i = 1$.

Por lo tanto, $\lambda \boldsymbol{t} + (1-\lambda)\boldsymbol{s} \in \Delta_N$ para todo $\lambda \in [0,1]$.

- Compacto: al encontrarnos en $\mathbb{R}^N$ y aplicando el conocido Teorema de Heine-Borel basta y sobra ver que Δ_N es cerrado y acotado. Claramente es acotado por lo que nos centraremos en probar que es cerrado. Sea $\{\boldsymbol{t}_n\}_{n\in\mathbb{N}}$ una sucesión de Δ_N y sea $\boldsymbol{t}_0 \in \mathbb{R}^N$ tal que $\{\boldsymbol{t}_n\}_{n\in\mathbb{N}} \longrightarrow \boldsymbol{t}_0$. Tenemos que comprobar que $\boldsymbol{t}_0 \in \Delta_N$.

  i) Como todas las coordenadas de cada $\boldsymbol{t}_n$ para $n \in \mathbb{N}$ son no negativas las de $\boldsymbol{t}_0$ también lo son.
  ii) La función $f : \mathbb{R}^N \longrightarrow \mathbb{R}$ definida como $f(\boldsymbol{t}) = \sum_{i=1}^{N} t_i$ es continua. Claramente $f(\boldsymbol{t}) = 1$ para todo $\boldsymbol{t} \in \Delta_N$ y por ello $\{f(\boldsymbol{t}_n)\}_{n\in\mathbb{N}} \longrightarrow 1$. Por continuidad de f y unicidad de límite tenemos que $f(\boldsymbol{t}_0) = 1$ pero eso implica que la suma de sus componentes vale 1.

  Así, hemos demostrado que $\boldsymbol{t}_0 \in \Delta_N$ y por lo tanto Δ_N es compacto.

Antes de continuar, notamos que el espacio vectorial de todas las aplicaciones lineales de $\mathbb{R}^N$ a $\mathbb{R}$, con $N \in \mathbb{N}$, se puede identificar con $\mathbb{R}^N$. Esto se debe a que si $L : \mathbb{R}^N \longrightarrow \mathbb{R}$ es lineal, entonces es de la forma $L(\boldsymbol{x}) = a_1x_1 + \cdots + a_Nx_N$ que se corresponde con $\boldsymbol{a} = (a_1, \ldots, a_N) \in \mathbb{R}^N$. Del mismo modo, dado el vector tenemos la aplicación lineal asociada. En definitiva, el espacio dual de $\mathbb{R}^N$ se identifica con $\mathbb{R}^N$.

También presentamos el producto escalar usual en $\mathbb{R}^N$ que definimos como:

$$\begin{aligned}\langle \cdot, \cdot \rangle : \mathbb{R}^N \times \mathbb{R}^N &\longrightarrow \mathbb{R} \\ (\boldsymbol{x}, \boldsymbol{y}) &\longmapsto \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=0}^{N} x_i y_i\end{aligned}$$

Enunciamos este lema previo que usaremos posteriormente en la demostración del de Simons.

Lema 3.1. *Sea $N \in \mathbb{N}$ y $S : \mathbb{R}^N \longrightarrow \mathbb{R}$ definida por:*

$$S(\mathbf{x}) := \text{máx}\{x_1, ..., x_N\}.$$

Entonces, S es sublineal. Además, si $L : \mathbb{R}^N \longrightarrow \mathbb{R}$ es un funcional lineal tal que $L \leq S$ entonces L es de la forma $L(\mathbf{x}) = t_1x_1 + ... + t_Nx_N$ con $(t_1, ..., t_N) \in \Delta_N$. De hecho, el recíproco también es cierto, es decir, si $L = \mathbf{t} = (t_1, ..., t_N) \in \Delta_N$ entonces $L \leq S$.

Demostración. Claramente, S es positivamente homogénea. También es subaditiva ya que dados $x, y \in$V :

$$\begin{aligned} S(\boldsymbol{x} + \boldsymbol{y}) &= \text{máx}\{x_1 + y_1, ..., x_N + y_N\} \\ &\leq \text{máx}\{x_1, ..., x_N\} + \text{máx}\{y_1, ..., y_N\} = S(\boldsymbol{x}) + S(\boldsymbol{y}). \end{aligned}$$

Por ello, S es sublineal. Para terminar veamos que

$$\left\{L : \mathbb{R}^N \longrightarrow \mathbb{R} : L \text{ lineal y } L \leq S\right\} = \Delta_N$$

a través de la correspondencia mostrada anteriormente entre $\mathbb{R}^N$ y su dual.

$\supseteq$) Sea $\mathbf{t} \in \Delta_N$ definimos $L : \mathbb{R}^N \longrightarrow \mathbb{R}$ como $L(\boldsymbol{x}) := \langle \mathbf{t}, \boldsymbol{x} \rangle$. Es evidente que L es lineal en $\boldsymbol{x}$ al ser el producto escalar bilineal. Dado $\boldsymbol{x} \in \mathbb{R}^N$,

$$L(\boldsymbol{x}) = \sum_{i=1}^{N} t_i x_i \leq \sum_{i=1}^{N} t_i S(\boldsymbol{x}) = S(\boldsymbol{x}) \sum_{i=1}^{N} t_i = S(\boldsymbol{x})$$

donde la primera desigualdad se debe a que $x_i \leq S(\boldsymbol{x})$ para todo x_i con $i = 1, ..., N$ y a que $t_i \geq 0$ ya que $\mathbf{t} \in \Delta_N$. Esto también justifica la última igualdad ya que $\sum_{i=1}^N t_i = 1$.

$\subseteq$) Sea $L = \boldsymbol{t} = (t_1, ..., t_N) \in \mathbb{R}^N$ lineal tal que para todo $\boldsymbol{x} \in \mathbb{R}^N$ cumple que $\sum_{i=1}^N t_i x_i \leq S(\boldsymbol{x})$. Así, si tomamos $e_i \in \mathbb{R}^N$ donde e_i representa el i-ésimo elemento de la base usual de $\mathbb{R}^N$ con $i = 1, ..., N$, entonces:

$$L(-e_i) = -t_i \leq 0 \Longrightarrow t_i \geq 0 \quad \forall i = 1, ..., N.$$

Si ahora llamamos $e = \sum_{i=1}^N e_i$, obtenemos:

$$\left.\begin{array}{l} L(e) = \sum_{i=1}^N t_i \leq \text{máx}\{1, ..., 1\} = 1 \\ L(-e) = -\sum_{i=1}^N t_i \leq \text{máx}\{-1, ..., -1\} = -1 \end{array}\right\} \Longrightarrow \sum_{i=1}^{N} t_i = 1.$$

Concluimos entonces que $L = \boldsymbol{t} = (t_1, ..., t_N) \in \Delta_N$.

■

Enunciamos ahora lema de Simons [ST08] en que destacamos la ausencia de hipótesis topológicas, lo que será importante posteriormente.

Lema 3.2 (Lema de Simons)**.** *Sea C un subconjunto convexo de un espacio vectorial. Dadas $f_1, \dots, f_N : C \longrightarrow \mathbb{R}$, con $N \in \mathbb{N}$, funciones convexas, entones existe $\mathbf{t} \in \Delta_N$ que cumple*

$$\inf_{c\in C}\left[\max_{i=1\dots,N} f_i(c)\right] = \inf_{c\in C}\left[\sum_{i=1}^{N} t_i f_i(c)\right].$$

Demostración. Sea $V = \mathbb{R}^N$ con $N \in \mathbb{N}$. Definimos S :V$\longrightarrow \mathbb{R}$ como

$$S(x_1, ..., x_N) := \max\{x_1, ..., x_N\}.$$

Por el lema 3.1, S es sublineal. Tomamos el subconjunto:

$$D = \{(x_1, ..., x_N) \in V : \exists c \in C \quad \text{tal que} \quad \forall i = 1, ...N, \quad f_i(c) \leq x_i\}.$$

Comprobemos en primer lugar que D es un subconjunto convexo de V . Sean $x, y \in D$, por ello, existen $c_x, c_y \in C$ tales que $f_i(c_x) \leq x_i$ y $f_i(c_y) \leq y_i \quad \forall i = 1, ..., N$. Dado $\lambda \in [0, 1]$, llamamos $c := (1 - \lambda)c_x + \lambda c_y$ que pertenece a C por ser este convexo. Veamos que c es el elemento necesario de C para que cualquier combinación convexa de x e y esté en D. Así, para todo $i = 1, ..., N$:

$$f_i(c) = f_i((1 - \lambda)c_x + \lambda c_y) \leq (1 - \lambda)f(c_x) + \lambda f(c_y) \leq (1 - \lambda)x_i + \lambda y_i,$$

donde la primera desigualdad se debe a que las f_i son convexas y la segunda a que $x, y \in D$. Por ello, $(1 - \lambda)x_i + \lambda y_i \in D, \quad \forall \lambda \in [0, 1]$ por lo que D es convexo. Aplicando el Teorema de Mazur-Orlizc, existe L funcinal lineal en V tal que $L \leq S$ e $\inf_D L = \inf_D S$.

Nuevamente, por el lema 3.1 tenemos que $L = \boldsymbol{t} \in \Delta_N$. Finalmente:

$$\inf_D L = \inf_{c\in C}\left[t_1 f(c) + \cdots + t_N f(c)\right] = \inf_{c\in C}\left[\sum_{i=1}^{N} t_i f_i(c)\right]$$

e

$$\inf_D S = \inf_{c\in C}\left[\max_{i=1\dots,N} f_i(c)\right]$$

por lo que

$$\inf_{c\in C}\left[\max_{i=1\dots,N} f_i(c)\right] = \inf_{c\in C}\left[\sum_{i=1}^{N} t_i f_i(c)\right].$$

■

Enuciamos ahora el Teorema de la Alternativa de Gordan en su versión clásica.

Teorema 3.1 (Teorema de la alternativa de Gordan)**.** *Sean* $\{\mathbf{x}_1, ...\mathbf{x}_N\}$ *con* $\mathbf{x}_i \in \mathbb{R}^M$ *para* $i = 1, ..., N$ *y* $N, M \in \mathbb{N}$. *Entonces una, y solo una, de la siguientes afirmaciones se cumple:*

i)* $\exists \mathbf{t} \in \Delta_N$ *tal que* $0 = \displaystyle\sum_{i=1}^{N} t_i \mathbf{x}_i$.

ii)* $\exists \mathbf{y} \in \mathbb{R}^M$ *tal que cumple* $\displaystyle\max_{i=1,...,N} \langle \mathbf{y}, \mathbf{x}_i \rangle < 0$.

Omitimos esta demostración ya que a continuación mostramos la versión convexa del mismo y será la que probaremos. Después veremos que la versión convexa implica la clásica por lo que quedará probada.

Teorema 3.2 (Teorema de la Alternativa de Gordan-versión convexa)**.** *Sea* C *un subconjunto convexo de un espacio vectorial y sean* $f_1, ..., f_N : C \longrightarrow \mathbb{R}$ *funciones convexas con* $N \in \mathbb{N}$. *Entonces una, y solo una, de la siguientes afirmaciones se cumple:*

i) $\exists \mathbf{t} \in \Delta_N$ *tal que* $0 \leq \displaystyle\inf_{c \in C} \sum_{i=1}^{N} t_i f_i(c)$.

ii) $\exists c \in C$ *que cumple* $\displaystyle\max_{i=1...,N} f_i(c) < 0$.

Demostración. En primer lugar, veamos que las alternativas $i)$ y $ii)$ son excluyentes y exhaustivas. Para ello, veamos que $\neg i) \iff ii)$ suponiendo que $\inf_{c \in C} [\max_{i=1...,N} f_i(c)] \in \mathbb{R}$.

$\Leftarrow)$ Si existe $c_0 \in C$ tal que $\max_{i=1...,N} f_i(c) < 0$, entonces, dado $\boldsymbol{t} \in \Delta_N$,

$$\inf_{c \in C} \sum_{i=0}^{N} t_i f_i(c) \leq \sum_{i=0}^{N} t_i f_i(c_0) < 0$$

donde la última desigualdad se debe a que $\boldsymbol{t} \in \Delta_N$. Hemos probado entonces $\neg i)$.

$\Rightarrow)$ Si aplicamos el lema de Simons, lema 3.2, a las funciones $f_1, ..., f_N$ obtenemos:

$$\exists t \in \Delta_N : \inf_{c \in C} \left[\max_{i=1...,N} f_i(c) \right] = \inf_{c \in C} \sum_{i=1}^{N} t_i f_i(c).$$

Entonces, de $\neg i)$ sabemos que, para este $\boldsymbol{t} \in \Delta_N$

$$\inf_{c\in C} \sum_{i=0}^{N} t_i f_i(c) < 0$$

luego

$$\inf_{c\in C} \left[\max_{i=1\ldots,N} f_i(c) \right] < 0,$$

como queríamos demostrar.

Para finalizar, si $\inf_{c\in C} [\max_{i=1\ldots,N} f_i(c)] = -\infty$ es claro que estamos en el caso $ii)$. ∎

Destacamos las siguientes observaciones:

Observación 3.1. *Esta versión convexa del teorema implica la versión clásica del mismo.*

Para ello, basta aplicar la versión convexa del teorema a $C := \mathbb{R}^M$ y a las funciones $f_1, ..., f_N : C \longrightarrow \mathbb{R}$ definidas por $f_i(c) := \langle c, \boldsymbol{x}_i \rangle, \forall i = 1, ..., N$. Notar que las funciones $f_1, ..., f_N$ son lineales por la izquierda y como consecuencia son convexas. En este caso, la alternativa $ii)$ implica $ii*)$ ya que:

$$\exists c \in C = \mathbb{R}^M : \max_{i=1,\ldots,N} \langle c, x_i \rangle = \max_{i=1,\ldots,N} f_i(c) < 0.$$

Por su parte, la alternativa $i)$ nos da:

$$\exists \mathbf{t} \in \Delta_N : 0 \leq \inf_{c\in\mathbb{R}^M} \sum_{i=1}^{N} t_i f_i(c) = \inf_{c\in\mathbb{R}^M} \sum_{i=1}^{N} t_i \langle c, \boldsymbol{x}_i \rangle = \inf_{c\in\mathbb{R}^M} \langle c, \sum_{i=1}^{N} t_i \boldsymbol{x}_i \rangle.$$

Hemos obtenido por ello que $0 \leq \inf_{c\in\mathbb{R}^M} \langle c, \sum_{i=1}^{N} t_i \boldsymbol{x}_i \rangle$ lo que significa que $0 \leq \langle c, \sum_{i=1}^{N} t_i \boldsymbol{x}_i \rangle$ para todo $c \in \mathbb{R}^M$. Usando la bilinealidad del producto escalar:

$$0 \leq \langle -c, \sum_{i=1}^{N} t_i \boldsymbol{x}_i \rangle \iff 0 \leq -\langle c, \sum_{i=1}^{N} t_i \boldsymbol{x}_i \rangle \iff \langle c, \sum_{i=1}^{N} t_i \boldsymbol{x}_i \rangle \leq 0, \quad \forall c \in \mathbb{R}^M.$$

Juntando ambas desigualdades obtenemos que $0 = \langle c, \sum_{i=1}^{N} t_i \boldsymbol{x}_i \rangle, \quad \forall c \in \mathbb{R}^M$. Como la igualdad anterior se cumple para todo elemento de $\mathbb{R}^M$ entonces podemos deducir que $\sum_{i=1}^{N} t_i \boldsymbol{x}_i = 0$ ya que $\sum_{i=1}^{N} t_i \boldsymbol{x}_i \in (\mathbb{R}^M)^{\perp} = \{0\}$. Así pues, tenemos que $i)$ garantiza $i*)$.

Observación 3.2. *El lema de Simons (lema 3.2) y el Teorema de la Alternativa de Gordan (teorema 3.2) son equivalentes.*

Ya hemos visto que el Lema de Simons implica el Teorema de la Alternativa de Gordan. Veamos que el recíproco también es cierto.

Llamamos $\alpha := \text{ínf}_{c\in C}\left[\text{máx}_{i=1...,N}\{f_i(c)\}\right]$. Si $\alpha = -\infty$. Por el lema 3.1 sabemos que $\forall \boldsymbol{t} \in \Delta_N$ se cumple que $\sum_{i=1}^N t_i f_i(c) \leq \text{máx}_{i=1...,N}\{f_i(c)\}$ para todo $c \in C$. Tomando ínfimos en C:

$$\text{ínf}_{c\in C}\left[\sum_{i=1}^N t_i f_i(c)\right] \leq \text{ínf}_{c\in C}\left[\text{máx}_{i=1...,N}\{f_i(c)\}\right] = -\infty \Longrightarrow \text{ínf}_{c\in C}\left[\sum_{i=1}^N t_i f_i(c)\right] = -\infty$$

y por ello $\forall \boldsymbol{t} \in \Delta_N$ (en particular para uno cualquiera) se cumple que

$$\text{ínf}_{c\in C}\left[\sum_{i=1}^N t_i f_i(c)\right] = \text{ínf}_{c\in C}\left[\text{máx}_{i=1...,N}\{f_i(c)\}\right].$$

Supongamos ahora que $\alpha \in \mathbb{R}$. Sean las funciones $g_1, ..., g_N : C \longrightarrow \mathbb{R}$ definidas como $g_i = f_i - \alpha$ con $i = 1, ..., N$. Veamos que las funciones $g_1, ..., g_N$ son convexas como consecuencia de que $f_1, ..., f_N$ lo son. Sean $i = 1, \ldots, N$, $c_1, c_2 \in C$ y $\lambda \in [0, 1]$:

$$\begin{aligned}
g_i(\lambda c_1 + (1-\lambda)c_2) &= f_i(\lambda c_1 + (1-\lambda)c_2) - \alpha \\
&\leq \lambda f_i(c_1) + (1-\lambda) f_i(c_2) - \alpha \\
&= \lambda f_i(c_1) + (1-\lambda) f_i(c_2) - \lambda\alpha + (1-\lambda)\alpha \\
&= \lambda(f_i(c_1) - \alpha) + (1-\lambda)(f_i(c_2) - \alpha) \\
&= \lambda g_i(c_1) + (1-\lambda) g_i(c_2).
\end{aligned}$$

Obtenemos así que g_i es convexa para todo $i = 1, ..., N$. Si usamos el Teorema de la Alternativa de Gordan obtenemos que solo se puede dar una y solo de las siguientes posibilidades:

i) $\exists \mathbf{t} \in \Delta_N$ tal que $0 \leq \text{ínf}_{c\in C} \sum_{i=1}^N t_i g_i(c)$.

ii) $\exists c \in C$ que cumple $\text{máx}_{i=1...,N}\{g_i(c)\} < 0$.

Razonemos que no se puede dar *ii*). Si fuese así, tendríamos que $\exists c \in C$ tal que $\text{máx}_{i=1...,N}\{g_i(c)\} = \text{máx}_{i=1...,N}\{f_i(c) - \alpha\} < 0$. En particular, para todo índice $j \in 1, ..., N$ se cumpliría $f_j(c) - \alpha < 0 \Longrightarrow f_j(c) < \alpha = \text{ínf}_{c\in C}\left[\text{máx}_{i=1...,N}\{f_i(c)\}\right]$. Esto es imposible por la propia definición de α. Por ello, afirmamos que $\exists \mathbf{t} \in \Delta_N$ tal que $0 \leq \text{ínf}_C \sum_{i=1}^N t_i g_i$. Desarrollando el sumatorio:

$$0 \leq \text{ínf}_{c\in C} \sum_{i=1}^N t_i g_i(c) = \text{ínf}_{c\in C} \sum_{i=1}^N t_i (f_i(c) - \alpha) = \text{ínf}_{c|inC}\left[\sum_{i=1}^N t_i f_i(c) - \sum_{i=1}^N t_i \alpha\right]$$

$$= \text{ínf}_{c\in C}\left[\sum_{i=1}^N t_i f_i(c) - \alpha \sum_{i=1}^N t_i\right] = \text{ínf}_{c\in C}\left[\sum_{i=1}^N t_i f_i(c) - \alpha\right] = \text{ínf}_{c\in C}\left[\sum_{i=1}^N t_i f_i(c)\right] - \alpha.$$

Por lo tanto:

$$0 \leq \inf_{c\in C}\left[\sum_{i=1}^{N} t_i f_i(c)\right] - \alpha \Longleftrightarrow \inf_{c\in C}\left[\max_{i=1\dots,N}\{f_i(c)\}\right] = \alpha \leq \inf_{c\in C}\left[\sum_{i=1}^{N} t_i f_i(c)\right].$$

El lema 3.1 nos aporta la otra desigualdad y llegamos nuevamente a que $\exists \boldsymbol{t} \in \Delta_N$ que cumple:

$$\inf_{c\in C}\left[\max_{i=1\dots,N}\{f_i(c)\}\right] = \inf_{c\in C}\left[\sum_{i=1}^{N} t_i f_i(c)\right].$$

Teoremas de Fritz John y de Karush-Kuhn-Tucker

En primer lugar, empezamos recordando la definición de derivada direccional.

Definición 4.1. *Sea $D \subset \mathbb{R}^M$ con $M \in \mathbb{N}$ y sea la función $g : D \longrightarrow \mathbb{R}$, definimos la derivada direccional de g en $x \in D$ en la dirección del vector $d \in \mathbb{R}^M$ como*

$$g'(x; d) = \lim_{t \to 0} \frac{g(x + td) - g(x)}{t}$$

siempre y cuando el límite exista. Diremos que g es diferenciable en el sentido de Gâteaux en x si $g'(x; \cdot) : \mathbb{R}^M \longrightarrow \mathbb{R}$ es lineal y en ese caso escribimos $\nabla g(x) = g'(x; \cdot)$, es decir, $g'(x; d) = \langle \nabla g(x), d \rangle$ con $d \in \mathbb{R}^M$.

Dentro del contexto de este trabajo, cuando decimos que una función es diferenciable nos referimos a que lo es en el sentido de Gâteaux. A estas funciones también las llamaremos Gâteaux diferenciables. Destacamos que este concepto de diferenciabilidad es más débil que el de Fréchet, que es el más frecuente dentro de nuestros estudios. De hecho, si una función g es diferenciable en $x_0 \in D$ en el sentido de Fréchet, y notamos su derivada como $Dg(x_0)$ entonces g también es diferenciable en el sentido de Gâteaux en x_0 y además $Dg(x_0) = \nabla g(x_0)$. El recíproco no es cierto tal y como mostramos en el siguiente ejemplo:

EJEMPLO

A continuación, demostremos cómo se calcula la derivada de la función máximo de un número finito de funciones diferenciables, lo que nos será útil en posteriores resultados.

Proposición 4.1. *Sean $D \subset \mathbb{R}^M$ ($M \in \mathbb{N}$) , $\bar{x}$ un punto del interior de D y sean $g_1, ..., g_N : D \longrightarrow \mathbb{R}$ funciones continuas y diferenciables en $\bar{x}$*

donde $N \in \mathbb{N}$. Definimos $g : D \longrightarrow \mathbb{R}$ como $g(x) := \max_{i=1,...,N}\{g_i(x)\}$ y el conjunto de índices $K = \{i : g_i(\bar{x}) = g(\bar{x})\}$. Entonces, para toda dirección $d \in \mathbb{R}^M$ la derivada direccional de g existe en todo $\mathbb{R}^M$ y viene dada por la siguiente expresión:

$$g'(\bar{x}; d) = \max_{i\in K}\{\langle \nabla g_i(\bar{x}), d\rangle\}, \quad d \in \mathbb{R}^M. \tag{4.1}$$

Demostración. Podemos suponer sin pérdida de generalidad que $K = \{1, ..., N\}$, ya que aquellas g_i que no alcancen el máximo no afectarán al cálculo de la derivada de g. Para cada $i \in K$ tenemos la siguiente desigualdad:

$$\liminf_{t\to 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} \geq \liminf_{t\to 0} \frac{g_i(\bar{x} + td) - g_i(\bar{x})}{t} = \langle \nabla g_i(\bar{x}), d\rangle.$$

La primera desigualdad se deduce de la definición de g ya que es el máximo de las g_i para $i = 1, ..., N$ y la segunda igualdad de que todas las g_i son diferenciables en $\bar{x}$ y por tanto existe el límite de la definición de derivada direccional y coincide con el límite inferior. Por lo tanto:

$$\liminf_{t\to 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} \geq \max_{i=1,...,N} \langle \nabla g_i(\bar{x}), d\rangle.$$

Por otro lado, afirmamos que

$$\limsup_{t\to 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} \leq \max_{i=1,...,N} \langle \nabla g_i(\bar{x}), d\rangle.$$

De lo contrario, existirían una sucesión $\{t_n\} \to 0$ y $\varepsilon > 0$ que cumplirían:

$$\frac{g(\bar{x} + t_n d) - g(\bar{x})}{t_n} \geq \max_{i=1,...,N} \langle \nabla g_i(\bar{x}), d\rangle + \varepsilon \quad \forall n \in \mathbb{N}.$$

Tomamos ahora una sucesión parcial $\{t_{\sigma(n)}\}_{n\in\mathbb{N}}$ con $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$ estrictamente creciente y $j \in K$ un índice fijo tal que para todo $k \in \{\sigma(n) : n \in \mathbb{N}\}$ se cumple que $g(\bar{x} + t_k d) = g_j(\bar{x} + t_k d)$. Tomando límite obtenemos que :

$$\begin{aligned}
\limsup_{t\to 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} &= \limsup_{t\to 0} \frac{g_j(\bar{x} + td) - g_j(\bar{x})}{t} \\
&= \langle \nabla g_j(\bar{x}), d\rangle \\
&\geq \max_{i=1,...,N} \langle \nabla g_i(\bar{x}), d\rangle + \varepsilon,
\end{aligned}$$

lo cual, es imposible. Finalmente, hemos obtenido que:

$$\limsup_{t\to 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} \leq \max_{i=1,...,N} \langle \nabla g_i(\bar{x}), d\rangle \leq \liminf_{t\to 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t}.$$

Como el límite inferior es siempre menor o igual que el superior concluimos que ambos coinciden y por lo tanto existe el límite y además:

$$\lim_{t\to 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} = g'(\bar{x}; d) = \max_{i=1,...,N} \langle \nabla g_i(\bar{x}), d\rangle.$$

∎

Nuestro objetivo ahora es encontrar soluciones a problemas del siguiente tipo:

$$\begin{cases} \text{ínf}_{x\in D} f(x) \\ \text{s.a } g_1(x) \leq 0 \\ \qquad\quad \vdots \\ \quad g_N(x) \leq 0 \end{cases} \tag{4.2}$$

donde $D \subset \mathbb{R}^M$, f es la función objetivo y las restricciones g_i con $i = 1, \ldots, N$ son funciones reales definidas en D y continuas. Si un punto satisface todas las restricciones diremos que es *factible* y como consecuencia llamamos *región factible* al conjunto de todos los puntos factibles. Para un punto factible $\bar{x}$ definimos el *conjunto activo* como $I(\bar{x}) = \{i : g_i(\bar{x}) = 0\}$. Para este problema y asumiendo que $\bar{x} \in C$, llamamos *vector de multiplicadores de Lagrange para* $\bar{x}$ a $\lambda \in (\mathbb{R}^N)^+$ si $\bar{x}$ es un punto crítico de:

$$L(x; \lambda) = f(x) + \sum_{i=1}^{N} \lambda_i g_i(x),$$

es decir, se cumple que (cuando $f, g_1, \ldots, g_N$ sean diferenciables):

$$\nabla f(\bar{x}) + \sum_{i=1}^{N} \lambda_i \nabla g_i(\bar{x}) = 0$$

y además $\lambda_i = 0$ si $i \notin I(\bar{x})$.

Teorema 4.1 (Teorema de Fritz John)**.** *Supongamos que el problema (4.2) tiene un mínimo local en $\bar{x} \in D$. Si las funciones f, g_i con $i \in I(\bar{x})$ son diferenciables en $\bar{x}$ entonces existen $\lambda_0, \lambda_i \in \mathbb{R}^+$ para $i \in I(\bar{x})$, no todas cero, que satisfacen:*

$$\lambda_0 \nabla f(\bar{x}) + \sum_{i\in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0.$$

Demostración. Consideramos la función

$$g(x) = \text{máx}\{f(x) - f(\bar{x}),\ g_i(x) : i \in I(\bar{x})\}.$$

Como $\bar{x}$ es un mínimo local del problema (4.2) también lo es de g. Esto se debe a que como $f(x) \geq f(\bar{x})$ entonces $f(x) - f(\bar{x}) \geq 0$ para todo x en un entorno de $\bar{x}$. Por otro lado, $g_i(x) \leq 0$ para todo $x \in D$ y como $i \in I(\bar{x})$ entonces $g_i(\bar{x}) = 0$. De este modo $g(x) \geq 0$ para todo x en un entorno de $\bar{x}$ y $g(\bar{x}) = 0$ por lo que efectivamente alcanza un mínimo local en $\bar{x}$. Por la proposición 4.1 tenemos que para toda dirección $d \in$V se cumple:

$$g'(\bar{x}; d) = \text{máx}\{\langle \nabla f(\bar{x}), d\rangle, \langle \nabla g_i(\bar{x}), d\rangle : i \in I(\bar{x})\} \geq 0,$$

ya que si $g'(\bar{x}; d) < 0$, para todo $t > 0$ suficientemente pequeño tendríamos que $g(\bar{x} + td) < g(\bar{x})$ lo que contradice que g alcanza de mínimo local en $\bar{x}$.

Por lo tanto, el sistema

$$\begin{cases} \langle \nabla f(\bar{x}), d \rangle < 0 \\ \langle \nabla g_i(\bar{x}), d \rangle < 0 \text{ con } i \in I(\bar{x}) \end{cases}$$

no tiene solución (para ninguna dirección) ya que al menos uno es no negativo. Si aplicamos el Teorema de la Alternativa de Gordan en su versión clásica, teorema 3.1, vemos que solo se puede dar la alternativa i*) y en ese caso obtenemos que:

$$\exists \boldsymbol{t} = (t_0, ..., t_M) \in \Delta_{M+1} \text{ tal que } 0 = t_0 \nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} t_i \nabla g_i(\bar{x}),$$

con M el cardinal del conjunto $I(\bar{x})$. La demostración concluye llamando $\lambda_0 = t_0$ y $\lambda_i = t_i$ con $i \in I(\bar{x})$. ■

El teorema de Fritz John nos aporta una gran desventaja y es que es posible que $\lambda_0 = 0$ por lo que la función objetivo es independiente de las restricciones. Por ello, necesitamos imponer algunas condiciones de regularidad extra. En esta situación diremos que se cumple la *condición de Mangasarian-Fromovitz* si existe una dirección $d_0 \in$V que satisface que $\langle \nabla g_i(\bar{x}), d_0 \rangle < 0$ para todo índice $i \in I(\bar{x})$. Enunciamos ahora otro teorema que soluciona el problema que acabamos de comentar.

Teorema 4.2 (Teorema de Karush-Kuhn-Tucker)**.** *Supongamos que el problema (4.2) tiene un mínimo local en $\bar{x} \in D$. Si las funciones f, g_i con $i \in I(\bar{x})$ son diferenciables en $\bar{x}$ y se cumple la condición de Mangasarian-Fromovitz entonces existe un vector de multiplicadores de Lagrange para $\bar{x}$.*

Demostración. Del teorema 4.1 de las condiciones de Fritz John obtenemos que:

$$\exists \boldsymbol{t} = (t_0, ..., t_M) \in \Delta_{M+1} \text{ tal que } 0 = t_0 \nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} t_i \nabla g_i(\bar{x})$$

con M el cardinal de $I(\bar{x})$. Si multiplicamos escalarmente la igualdad por d_0 (dirección del espacio vectorial que nos aporta la condición de Mangasarian-Fromovitz), obtenemos:

$$0 = t_0 \langle \nabla f(\bar{x}), d_0 \rangle + \sum_{i \in I(\bar{x})} t_i \langle \nabla g_i(\bar{x}), d_0 \rangle.$$

Entonces $t_0 \neq 0$. Razonemos por reducción al absurdo. Si no fuese así, tendríamos que

$$0 = \sum_{i \in I(\bar{x})} t_i \langle \nabla g_i(\bar{x}), d_0 \rangle.$$

Al tener $\boldsymbol{t} \in \Delta_{M+1}$ se cumple que todas sus componentes son no negativas y $\sum_{i \in I(\bar{x})} t_i = 1$ (estamos suponiendo que $t_0 = 0$ por lo que no influye en la suma de la definición de Δ_{M+1}) por lo que algún término es distinto de 0. Tenemos garantizado que $\langle g_i(\bar{x}), d_0 \rangle < 0 \quad \forall i \in I(\bar{x})$. Así tendríamos que:

$$0 = \sum_{i \in I(\bar{x})} t_i \langle \nabla g_i(\bar{x}), d_0 \rangle < 0,$$

lo cual es imposible. Por ello, concluimos que $t_0 \neq 0$. La demostración concluye tomando $\lambda_i = t_i/t_0$ para $i \in I(\bar{x})$. ■

La condición de Mangasarian-Fromovitz no es prescindible en el teorema de Karush-Kuhn-Tucker, tal y como pone de manifiesto este sencillo ejemplo.

$$\begin{cases} f(x,y) = 2x \\ \text{s.a } g_1(x) = 2y - 5x^3 \\ \qquad g_2(x) = -y \end{cases}$$

Es claro que el mínimo se alcanza en $\bar{x} = (0,0)$ y tenemos que $\nabla f(0,0) = (2,0)$, $\nabla g_1(0,0) = (0,2)$, $\nabla g_2(0,0) = (0,-1)$ e $I(\bar{x}) = \{1,2\}$. La condición de Mangasarian-Fromovitz no se cumple. Si existiese una dirección $d = (d_1, d_2) \in V$ que la cumpliese se deberían satisfacer las siguientes condiciones simultáneamente

$$\begin{cases} \langle \nabla g_1(0,0), (d_1,d_2) \rangle < 0 \iff \langle (0,2), (d_1,d_2) \rangle < 0 \iff 2d_2 < 0 \\ \langle \nabla g_2(0,0), (d_1,d_2) \rangle < 0 \iff \langle (0,-1), (d_1,d_2) \rangle < 0 \iff d_2 > 0 \end{cases}$$

lo que es imposible.

Finalmente, notamos que los teoremas de Fritz John y de Karush-Kuhn-Tucker que acabamos de demostrar también se pueden considerar consecuencia del lema de Farkas, que veremos posteriormente en este trabajo. Destacar también que en el caso convexo e imponiendo ciertas condiciones de regularidad podemos probar que el recíproco de dichos resultados también es cierto (REFERENCIAR).

Minimax

En esta sección llegaremos a otro de los resultados clave del trabajo. Será uno de los denominados teoremas minimax. A rasgos generales y a modo introductorio, podemos decir que un teroema minimax es un resultado que afirma, bajo ciertas hipótesis, que:

$$\inf_{y\in Y}\sup_{x\in X} f(x,y) = \sup_{x\in X}\inf_{y\in Y} f(x,y),$$

donde X e Y son conjuntos no vacíos y $f : X \times Y \longrightarrow \mathbb{R}$. Obviamente, esta igualdad no es cierta en general tal y como mostramos en el siguiente ejemplo. Definimos $f : \{0,1\} \times \{0,1\} \longrightarrow \mathbb{R}$ como:

$$f(x,y) = \begin{cases} 0, & \text{si } x = y \\ 1, & \text{si } x \neq y \end{cases}.$$

Por un lado tenemos

$$\inf_{y\in Y}\sup_{x\in X} f(x,y) = \min_{y\in Y}\max_{x\in X} f(x,y) = \min_{y\in Y}\{1\} = 1,$$

y por otro

$$\sup_{x\in X}\inf_{y\in Y} f(x,y) = \max_{x\in X}\min_{y\in Y} f(x,y) = \min_{y\in Y}\{0\} = 0.$$

Notemos que la desigualdad

$$\inf_{y\in Y}\sup_{x\in X} f(x,y) \geq \sup_{x\in X}\inf_{y\in Y} f(x,y)$$

siempre se da ya que si $(x_0, y_0) \in X \times Y$, entonces:

$$\sup_{x\in X} f(x,y_0) \geq f(x_0,y_0) \geq \inf_{y\in Y} f(x_0,y).$$

Por lo ello, los teoremas minimax solo nos aportan la otra desigualdad necesaria, de ahí que este tipo de resultados se conozcan también como desigualdad minimax.

Antes de continuar, exponemos la siguiente definición que aparecerá posteriormente en el teorema. Se trata de una propiedad más débil que la continuidad para funciones reales.

Definición 5.1. *Sea X un espacio topológico. Decimos que $f : X \longrightarrow \mathbb{R}$ es superiormente semicontinua si para todo $r \in \mathbb{R}$ se cumple que el conjunto $\{x \in X : f(x) \geq r\}$ es cerrado.*

Es claro que toda función continua en X es superiormente semicontinua, aunque el recíproco no es cierto, tal y como prueba la función $f : \mathbb{R} \longrightarrow \mathbb{R}$ dada por

$$f(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1, & \text{si } x \geq 0 \ . \end{cases}$$

En estos momentos nos encontramos en condiciones de enunciar y demostrar el siguiente teorema minimax.

Teorema 5.1. *Sean X, Y subconjuntos convexos de sendos espacios vectoriales (no tienen que ser el mismo) tal que X está dotado de una topología que lo hace compacto. Supongamos además que $f : X \times Y \longrightarrow \mathbb{R}$ es:*

i) cóncava y superiormente semicontinua en X y

ii) convexa en Y.

Entonces:

$$\operatorname*{ínf}_{y\in Y} \operatorname*{máx}_{x\in X} f(x,y) = \operatorname*{máx}_{x\in X} \operatorname*{ínf}_{y\in Y} f(x,y).$$

Demostración. En primer lugar, podemos escribir máximo en ambos casos en vez de supremo ya que f es superiormente semicontinua en X, por ello $\operatorname{ínf}_{y\in Y} f(x,y)$ también lo es (referenciar), y X es compacto (referenciar). Como hemos explicado anteriormente, solo necesitamos la desigualdad

$$\operatorname*{ínf}_{y\in Y} \operatorname*{máx}_{x\in X} f(x,y) \leq \operatorname*{máx}_{x\in X} \operatorname*{ínf}_{y\in Y} f(x,y). \tag{5.1}$$

Definimos $\alpha := \operatorname{ínf}_{y\in Y} \operatorname{máx}_{x\in X} f(x,y)$. Si $\alpha = -\infty$ no hay nada que probar, por lo que supondremos que $\alpha \in \mathbb{R}$. Primero vamos a reescribir el resultado a probar. La desigualdad (5.1) es equivalente a

$$\exists x_0 \in X : \ \alpha \leq \operatorname*{ínf}_{y\in Y} f(x_0,y),$$

ya que si existe un elemento en X que lo cumpla el máximo también lo cumplirá y recíprocamente. O lo que es lo mismo:

$$\exists x_0 \in X : \ y \in Y \Longrightarrow \alpha \leq f(x_0,y),$$

o lo que es igual,

$$\bigcap_{y\in Y}\{x\in X : \alpha \le f(x,y)\} \neq \emptyset. \tag{5.2}$$

Como f es superiormente semicontinua en X estamos ante una intersección de cerrados. Usando la propiedad de intersección finita (X es compacto) obtenemos que 5.2 equivale a

$$N\in\mathbb{N},\ y_1,\dots,y_N\in Y \Longrightarrow \bigcap_{i=1}^{N}\{x\in X : \alpha \le f(x,y_i)\} \neq \emptyset.$$

$$\Updownarrow$$

$$N\in\mathbb{N},\ y_1,\dots,y_N\in Y \Longrightarrow \exists x_0\in X : \alpha \le \min_{i=1\dots,N} f(x_0,y_i).$$

$$\Updownarrow$$

$$N\in\mathbb{N},\ y_1,\dots,y_N\in Y \Longrightarrow \alpha \le \max_{x\in X}\min_{i=1\dots,N} f(x,y_i). \tag{5.3}$$

Esta última condición es cierta. En efecto, sean $N\in\mathbb{N}$ e $\{y_1,...,y_N\}\in Y$. Aplicamos el lema de Simons, lema (3.2), tomando $C := X$ y $f_i : X \longrightarrow \mathbb{R}$ definidas como

$$f_i(x) := -f(x,y_i), \quad i=1,\dots,N.$$

Como f es cóncava respecto a X tenemos que las f_i son convexas para X con $i=1,\dots,N$. De este modo, existe $\boldsymbol{t}\in\Delta_N$ tal que

$$\inf_{x\in X}\left[\max_{i=1\dots,N}\{f_i(x)\}\right] = \inf_{x\in X}\left[\sum_{i=1}^{N} t_i f_i(x)\right].$$

Si ponemos la igualdad en función de f y recordamos que alcanza el supremo en X,

$$\inf_{x\in X}\left[\max_{i=1\dots,N}\{-f(x,y_i)\}\right] = \inf_{x\in X}\left[\sum_{i=1}^{N} t_i(-f(x,y_i))\right]$$

$$\Downarrow$$

$$\max_{x\in X}\left[\min_{i=1\dots,N}\{f(x,y_i)\}\right] = \max_{x\in X}\left[\sum_{i=1}^{N} t_i f(x,y_i)\right].$$

Al ser f convexa en Y:

$$\max_{x\in X}\left[\min_{i=1\dots,N}\{f(x,y_i)\}\right] \ge (\text{¿=? pero es convexa}) \max_{x\in X}\left[f(x,\sum_{i=1}^{N} t_i y_i)\right] \ge \inf_{y\in Y}\max_{x\in X} f(x,y) = \alpha.$$

Hemos probado entonces la desigualdad (5.3) y con ello el teorema. ■

Teorema de Separación

En esta sección introducimos algunos resultados sobre separación. En general, estos nos aportan herramientas para poder concluir cuándo dos subconjuntos convexos pueden ser separados mediante un hiperplano. En la siguiente imagen vemos un ejemplo sobre la situación en la que nos encontramos:

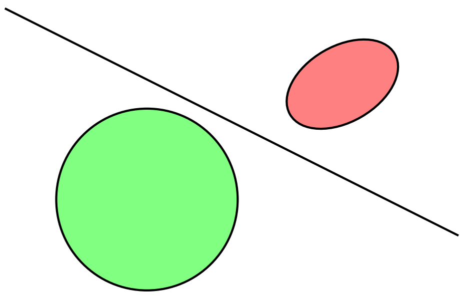

Figura 6.1: Situación de teoremas de separación

Exponemos un resultado sencillo en el que solo involucramos un conjunto

Teorema 6.1. *Dado $N \in \mathbb{N}$, sea $C \subset \mathbb{R}^N$ convexo y definimos*

$$\delta := \text{ínf}\{\|c\| : c \in C\}.$$

Entonces, existe $x_0 \in \mathbb{R}^N$ tal que si $c \in C$ se cumple que $\delta \leq \langle x_0, c\rangle$.

Demostración. En primer lugar, vamos a reescribir la tesis del teorema. Queremos ver que:

$$\exists x_0 \in \mathbb{R}^N : c \in C \Longrightarrow \delta \leq \langle x_0, c\rangle.$$

$$\Updownarrow$$

$$\exists \alpha > 0, \exists x_0 \in \alpha B_{\mathbb{R}^N} : c \in C \Longrightarrow \delta \leq \langle x_0, c\rangle.$$

$$\Updownarrow$$

$$\exists \alpha > 0, \exists x_0 \in \alpha B_{\mathbb{R}^N} : \delta \leq \inf_{c \in C} \langle x_0, c \rangle.$$

$$\Updownarrow$$

$$\exists \alpha > 0 : \delta \leq \max_{x \in \alpha B_{\mathbb{R}^N}} \inf_{c \in C} \langle x, c \rangle.$$

Llamamos $X := \alpha B_{\mathbb{R}^N}$ que es compacto y conexo, $Y := C$ convexo y f a la función real y continua con valores en $X \times Y$ definida como $f(x, y) := \langle x, y \rangle$ (al ser f continua, en particular es superiormente semicontinua). Aplicamos el teorema Minimax, teorema (5.1), y obtenemos que probar la última desigualdad es equivalente a probar que

$$\exists \alpha > 0 : \delta \leq \inf_{c \in C} \max_{x \in \alpha B_{\mathbb{R}^N}} \langle x_0, c \rangle.$$

Tenemos que $\max_{x \in \alpha B_{\mathbb{R}^N}} \langle x, c \rangle \leq$ (¿=? es desigualdad C-S)$\alpha \|c\|$ por la desigualdad de Cauchy–Schwarz. Así, debemos demostrar que

$$\exists \alpha > 0 : \delta \leq \inf_{c \in C} \alpha \|c\| \leq \alpha \delta.$$

Pero esta desigualdad es cierta tomando, por ejemplo, $\alpha = 1$. ■

Ahora, vamos a hacer una generalización de este resultado.

Teorema 6.2. *Sean $N \in \mathbb{N}$ y A, B subconjuntos convexos de $\mathbb{R}^N$ para tal que A es cerrado, B es compacto y $A \cap B = \emptyset$. Entonces existe $x_0 \in \mathbb{R}^N$ tal que*

$$\sup_{a \in A} \langle x_0, a \rangle < \inf_{b \in B} \langle x_0, b \rangle.$$

Demostración. En primer lugar, veamos que $\operatorname{dist}(A, B) > 0$ donde la distancia viene dada por

$$\operatorname{dist}(A, B) = \inf\{\|a - b\| : a \in A, b \in B\}.$$

Para ello, razonemos por reducción al absurdo. Suponemos $\operatorname{dist}(A, B) = 0$, entonces existe una sucesión $\{u_n\}_{n \in \mathbb{N}} \subset A - B$ tal que $\{u_n\}_{n \in \mathbb{N}} \longrightarrow 0$. Para $n \in \mathbb{N}$ tenemos que $u_n = a_n - b_n$ con $a_n \in A$ y $b_n \in B$. De este modo obtenemos las sucesiones $\{a_n\}_{n \in \mathbb{N}} \subset A$ y $\{b_n\}_{n \in \mathbb{N}} \subset B$. Como B es compacto, existe una sucesión parcial convergente, es decir, existe $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$ estrictamente creciente tal que $\{b_{\sigma(n)}\}_{n \in \mathbb{N}} \longrightarrow b$ con $b \in B$. Así,

$$\left\|a_{\sigma(n)} - b\right\| = \left\|a_{\sigma(n)} - b_{\sigma(n)} + b_{\sigma(n)} - b\right\| \leq \left\|a_{\sigma(n)} - b_{\sigma(n)}\right\| + \left\|b_{\sigma(n)} - b\right\|.$$

Entonces tenemos que $\left\|a_{\sigma(n)} - b\right\| \longrightarrow 0$ ya que $\{b_{\sigma(n)}\}_{n \in \mathbb{N}} \longrightarrow b$ y como $\|a_n - b_n\| \longrightarrow 0$ también se cumple que $\left\|a_{\sigma(n)} - b_{\sigma(n)}\right\| \longrightarrow 0$. Llegamos a que $\{a_{\sigma(n)}\}_{n\in\mathbb{N}} \longrightarrow b$. Al ser A cerrado se debe cumplir que $b \in A$ lo cual es imposible ya que $A \cap B = \emptyset$.

Ahora, aplicamos el teorema (6.1) a $C := B - A$. Notar que C es convexo por serlo A y B. Obtenemos entonces que existe $x_0 \in \mathbb{R}^N$ tal que si $c \in C$ se cumple que:

$$\delta = \inf_{c\in C} \|c\| \leq \langle x_0, c\rangle.$$

Por la definición de C, se tiene que:

$$\delta = \inf\{\|b-a\| : a \in A, b \in B\} = \text{dist}(A, B) > 0.$$

Del mismo modo, $c = b - a$ para todo $c \in C$ con $a \in A$ y $b \in B$. Así.

$$\exists x_0 \in \mathbb{R}^N : a \in A, b \in B \Longrightarrow 0 < \delta \leq \langle x_0, b-a\rangle = \langle x_0, b\rangle - \langle x_0, a\rangle.$$

$$\Updownarrow$$

$$\exists x_0 \in \mathbb{R}^N : a \in A, b \in B \Longrightarrow \langle x_0, a\rangle + \delta \leq \langle x_0, b\rangle, \text{ con } \delta > 0.$$

$$\Updownarrow$$

$$\exists x_0 \in \mathbb{R}^N \Longrightarrow \sup_{a\in A}\langle x_0, a\rangle + \delta \leq \inf_{b\in B}\langle x_0, a\rangle, \text{ con } \delta > 0.$$

$$\Updownarrow$$

$$\exists x_0 \in \mathbb{R}^N : \sup_{a\in A}\langle x_0, a\rangle < \inf_{b\in B}\langle x_0, a\rangle.$$

Por ello, queda probado el teorema. ■

El teorema de separación anterior podemos escribirlo no solo en $\mathbb{R}^N$ sino que se puede generalizar a cualquier espacio normado. A continuación, exponemos otro teorema de separación válido solo en el contexto finito dimensional, con tesis e hipótesis más débiles.

Teorema 6.3. *Sea $N \in \mathbb{N}$ y A y B dos subconjuntos convexos y disjuntos de $\mathbb{R}^N$. Entonces existe $x_0 \in \mathbb{R}^N \setminus \{0\}$ tal que*

$$\sup_{a\in A}\langle x_0, a\rangle \leq \inf_{b\in B}\langle x_0, b\rangle.$$

Demostración. Al igual que en el teorema anterior, podemos reducir la prueba al caso en que C es un subconjunto de $\mathbb{R}^N$ convexo de forma que $0 \notin C$, demostrando que:

$$\exists x_0 \in \mathbb{R}^N \setminus \{0\} : \quad \sup_{c\in C}\langle x_0, c\rangle \leq 0,$$

equivalentemente,

$$\exists x_0 \in S_{\mathbb{R}^N} \setminus \{0\} : \quad \sup_{c \in C} \langle x_0, c \rangle \leq 0.$$

Pero esta afirmación no es más que

$$\bigcap_{c \in C} \{x \in S_{\mathbb{R}^N} : \langle x, c \rangle \leq 0\} \neq \emptyset.$$

Usamos que $S_{\mathbb{R}^N}$ es compacta y por ello tenemos, gracias a la propiedad de intersección finita, que esto equivale a

$$\emptyset \neq C_0 \subset C \text{ finito} \Longrightarrow \bigcap_{c \in C_0} \{x \in S_{\mathbb{R}^N} : \langle x, c \rangle \leq 0\} \neq \emptyset.$$

$$\Updownarrow$$

$$M \in \mathbb{N},\ \{c_1, \ldots, c_M\} \subset C \Longrightarrow \exists x_0 \in S_{\mathbb{R}^N} : \quad \max_{i=1,\ldots,M} \langle x_0, c_i \rangle \leq 0.$$

Esta condición se verifica, ya que si $M \geq 1$ y $\{c_1, \ldots, c_M\} \subset C$, entonces, por ser C convexo se cumple que $\mathrm{co}\{c_1, \ldots, c_M\} \subset C$ y como $0 \notin C$, entonces $0 \notin \mathrm{co}\{c_1, \ldots, c_M\}$. Por tanto, la versión clásica del Teorema de Gordan, teorema (3.1), garantiza

$$\exists z_0 \in \mathbb{R}^N : \quad \max_{i=1,\ldots,M} \langle z_0, c_i \rangle < 0.$$

En particular, $z_0 \neq 0$ y tomando $x_0 = \frac{z_0}{\|z_0\|}$ queda probado el enunciado. ■

Destacamos ahora lo siguiente:

Observación 6.1. *El Teorema de separación (6.2) implica el Teorema de Gordan, teorema (3.1).*

En efecto, dados $\{x_1, \ldots, x_M\} \subset \mathbb{R}^N$ con $M, N \in \mathbb{N}$ del enunciado de la alternativa de Gordan planteamos las siguientes alternativas excluyentes:

1. Si $0 \in \mathrm{co}\{x_1, \ldots, x_M\}$ entonces es claro que se cumple la alternativa $i*$).

2. Si $0 \notin \mathrm{co}\{x_1, \ldots, x_M\}$ entonces $\{0\} \cap \mathrm{co}\{x_1, \ldots, x_M\} = \emptyset$. Al ser ambos conjuntos son compactos usamos el teorema de separación y obtenemos que:

$$\exists y \in \mathbb{R}^M : \quad \sup_{x \in \mathrm{co}\{x_1, \ldots, x_M\}} \langle y, x \rangle < 0,$$

   luego, en particular ($\{x_1, \ldots, x_M\} \subset \mathrm{co}\{x_1, \ldots, x_M\}$),

$$\exists y \in \mathbb{R}^M : \quad \max_{i=1,\ldots,M} \langle y, x_i \rangle < 0.$$

Lema de Farkas

En la sección anterior hemos vimos como uno de los teoremas de separación implica el único teorema de la alternativa que hemos visto hasta el momento. Ahora, vamos a deducir de manera parecida otro teorema de la alternativa. Antes exponemos la siguiente definición.

Definición 7.1. *Dados $M, N \in \mathbb{N}$ y $\{x_1, \ldots, x_M\} \subset \mathbb{R}^N$, llamamos cono generado por $\{x_1, \ldots, x_M\}$ al conjunto de $\mathbb{R}^N$ dado por:*

$$\text{cone}\{x_1, \ldots, x_M\} := \left\{ \sum_{j=1}^{N} \mu_j x_j : \ \mu_1, ..., \mu_N \geq 0 \right\}.$$

Veamos que cone$\{x_1, \ldots, x_M\}$ es convexo y compacto:

- Convexo: tenemos que comprobar que dados $\boldsymbol{t}, \boldsymbol{s} \in \text{cone}\{x_1, \ldots, x_M\}$ y $\lambda \in [0, 1]$ entonces $\lambda \boldsymbol{t} + (1 - \lambda)\boldsymbol{s} \in \text{cone}\{x_1, \ldots, x_M\}$. En efecto:

$$\begin{aligned}
\lambda \boldsymbol{t} + (1 - \lambda)\boldsymbol{s} &= \lambda \sum_{j=1}^{N} \mu_j^t x_j + (1 - \lambda) \sum_{j=1}^{N} \mu_j^s x_j \\
&= \sum_{j=1}^{N} [\lambda \mu_j^t x_j + (1 - \lambda)\mu_j^s x_j] \\
&= \sum_{j=1}^{N} [\lambda \mu_j^t + (1 - \lambda)\mu_j^s] x_j.
\end{aligned}$$

  Como μ_j^t y μ_j^s son no negativos, entonces $\lambda \mu_j^t + (1 - \lambda)\mu_j^s$ también es una cantidad no negativa para $j = 1, \ldots, M$. Así, podemos concluir que $\lambda \boldsymbol{t} + (1 - \lambda)\boldsymbol{s} \in \text{cone}\{x_1, \ldots, x_M\}$ y por tanto es un subconjunto convexo.

- Cerrado: sea $\{\boldsymbol{t}_n\}_{n\in\mathbb{N}}$ una sucesión de cone$\{x_1,\ldots,x_M\}$ y sea $\boldsymbol{t}_0 \in \mathbb{R}^N$ tal que $\{\boldsymbol{t}_n\}_{n\in\mathbb{N}} \longrightarrow \boldsymbol{t}_0$. Si notamos $\boldsymbol{t}_n = \sum_{j=1}^N \mu_j^n x_j$, entonces:

$$\{\boldsymbol{t}_n\}_{n\in\mathbb{N}} = \{\sum_{j=1}^N \mu_j^n x_j\}_{n\in\mathbb{N}} \longrightarrow \sum_{j=1}^N \mu_j^0 x_j = \boldsymbol{t}_0.$$

  Como para cada $\boldsymbol{t}_n$ cumple que $\mu_j^n \geq 0$ con $j=1,\ldots,N$ podemos asegurar que $\mu_j^0 \geq$ con $j=1,\ldots,N$. Hemos demostrado que $\boldsymbol{t}_0$ se expresa como combinación de $\{x_1,\ldots,x_M\}$ con coeficientes no negativos. Por lo tanto, $\boldsymbol{t}_0 \in$ cone$\{x_1,\ldots,x_M\}$ y por consiguiente es cerrado.

Enunciamos ahora otro de los teoremas de la alternativa más conocidos.

Lema 7.1 (Lema de Farkas)**.** *Sean $M, N \in \mathbb{N}$ y $\{x_1,\ldots,x_M\} \subset \mathbb{R}^N$ y $b \in \mathbb{R}^N$. Entonces una, y solo una, de la siguientes afirmaciones se cumple:*

i') $\exists \mu_1,\ldots,\mu_M \geq 0$ tal que $b = \sum_{j=1}^M \mu_j x_j$.

ii') $\exists z_0 \in \mathbb{R}^N$ que cumple que:

1. $\max_{j=1,\ldots,M}\langle z_0, x_j\rangle \leq 0$ *y*
2. $\langle z_0, b\rangle > 0$.

Demostración. Planteamos las siguientes alternativas, que obviamente son excluyentes, y que implican las de la tesis del lema:

a) $b \in$ cone$\{x_1,\ldots,x_M\}$. Estamos en el caso $i')$ ya que:

$$b \in \left\{\sum_{j=1}^N \mu_j x_j : \ \mu_1,\ldots,\mu_N \geq 0\right\}.$$

b) $b \notin$ cone$\{x_1,\ldots,x_M\}$. Por su parte, esta alternativa equivale a $ii')$. En efecto:

$ii') \Longrightarrow b)$ Razonamos por reducción al absurdo. Suponemos que $b \in$ cone$\{x_1,\ldots,x_M\}$, entonces, podemos expresar $b = \sum_{j=1}^N \mu_j x_j$ con $\mu_1,\ldots,\mu_N \geq 0$. Como se da $ii')$, en particular se cumple 1 y obtenemos

$$\begin{aligned}\langle z_0, b\rangle &= \langle z_0, \sum_{j=1}^N \mu_j x_j\rangle \\ &= \sum_{j=1}^N \mu_j \langle z_0, x_j\rangle \leq 0.\end{aligned}$$

Por otro lado, por 2 se tiene que $\langle z_0, b\rangle > 0$. Así, obtenemos que

$$\langle z_0, b\rangle \leq 0 < \langle z_0, b\rangle,$$

lo cual es imposible.

$b) \Longrightarrow ii')$ Si $b \notin \mathrm{cone}\{x_1, \ldots, x_M\}$ aplicamos el teorema de separación (6.2) a los conjuntos $\{b\}$ que es compacto y convexo y a $\mathrm{cone}\{x_1, \ldots, x_M\}$ que es cerrado y convexo. Obtenemos que:

$$\exists z_0 \in \mathbb{R}^N : \sup_{a \in \mathrm{cone}\{x_1, \ldots, x_M\}} \langle z_0, a\rangle < \langle z_0, b\rangle.$$

Por un lado, es obvio que $0 \in \mathrm{cone}\{x_1, \ldots, x_M\}$ y por ello:

$$0 = \langle z_0, 0\rangle \leq \sup_{a \in \mathrm{cone}\{x_1, \ldots, x_M\}} \langle z_0, a\rangle < \langle z_0, b\rangle.$$

Hemos obtenido por tanto que es cierto 2. Para probar 1, fijamos $a_0 \in \mathrm{co}\{x_1, \ldots, x_M\}$ (es obvio si $\mathrm{co}\{x_1, \ldots, x_M\} = \{0\}$). Entonces, dado $\rho > 0$,

$$\rho\langle z_0, a_0\rangle = \langle z_0, \rho a_0\rangle \leq \sup_{a \in \mathrm{cone}\{x_1, \ldots, x_M\}} \langle z_0, a_0\rangle.$$

Llegamos a que el conjunto $\{\rho\langle z_0, a_0\rangle : \ \rho > 0\}$ está acotado y eso solo es posible si $\langle z_0, a_0\rangle \leq 0$. La arbitrariedad de a_0 nos aporta que

$$\sup_{a \in \mathrm{cone}\{x_1, \ldots, x_M\}} \langle z_0, a\rangle = 0$$

de donde

$$\max_{j=1,\ldots,M} \langle z_0, x_j\rangle \leq \sup_{a \in \mathrm{cone}\{x_1, \ldots, x_M\}} \langle z_0, a\rangle = 0$$

ya que obviamente $\{x_1, \ldots, x_M\} \subset \mathrm{cone}\{x_1, \ldots, x_M\}$. En particular, hemos probado 1.

■

Finanzas y primer teorema de asignación de precios

Nos introducimos ahora en el mundo de las denominadas *matemáticas financieras*. En las secciones anteriores hemos expuesto todas las herramientas matemáticas necesarias para el trabajo y en esta vamos a explicar los conceptos económicos. Una vez se haya completado este apartado estaremos dispuestos a enunciar y probar el enunciado culmen de este trabajo que nos servirá para la valoración de activos financieros en mercados finitos.

Nos preguntamos entonces: "¿qué es un *activo*?". Un activo o título de valor se define como un recurso con valor que alguien posee con el fin de obtener un beneficio en el futuro. Podemos diferenciar entre activos seguros, como depósitos en el banco o bonos del estado, y activos con riesgo, como las acciones. Uno de los conceptos más importantes que tenemos es nuestro modelo del mercado financiero es el *principio de no arbitraje*. Este principio intuitivamente nos dice que no podemos obtener beneficio si no corremos algún riesgo. Puede resultar un poco confuso ya que acabamos de diferencias entre activos seguros y con riesgo. Esto se debe a que un activo, aunque se llame seguro, no significa que tenga el beneficio asegurado. Por ejemplo, si tenemos nuestro dinero depositado es posible que el banco quiebre y perdamos todos nuestros ahorros. Así, en la realidad estas oportunidades, llamadas de arbitraje, son muy raras y cuando se dan suponen una ganancia muy pequeña en comparación con la cantidad de dinero que se está manejando globalmente.

Cuando nos movemos en el ámbito financiero también debemos de tener en cuenta el *valor del dinero*. Nuestro dinero se va devaluando con el paso del tiempo. Es preferible obtener una cantidad de dinero en este momento que en el futuro ya que no tendremos el mismo poder adquisitivo. Por eso, cuando alguien tiene una deuda debe devolver el dinero con cierto interés porque, de otro modo, sería injusto para la persona que presta el dinero. Además, dicho interés es en cierta medida una estimación ya que no se puede saber con seguridad el precio en un futuro. Lo mismo pasa con los activos con riesgo, solo sabemos el precio que tienen en este momento. Por tanto, es posible que el precio en el futuro sea mayor que el actual o menor. Matemáticamente, podemos representar su valor mediante una variable aleatoria que generalmente mide la ganancia en vez del precio aunque se puede pasar de una otra fácilmente. Podemos suponer una situación con gran número de posibles ganancias intentando abarcar la mayor cantidad de situaciones que nos podríamos encontrar. Sin embargo, el caso binomial en el que solo existen dos posibilidades es el más habitual ya que es lo suficientemente simple de manejar y además refleja bastantes situaciones del mercado financiero real. Este modelo también supone que en cada paso la ganancia tiene el mismo comportamiento. Tenemos por tanto la variable aleatoria $K(n) : \Omega \longrightarrow (-1.\infty)$ definida como:

$$K(n) = \begin{cases} u & \text{con probabilidad } p \\ d & \text{con probabilidad } 1-p \end{cases}$$

cumpliendo $-1 < d < u$ y $0 < p < 1$. La primera condición es importante ya que garantiza que todos los precios van a ser positivos. El espacio de probabilidad Ω denota todos los posibles escenarios $\omega \in \Omega$ en los que varía el precio. Como nos hemos restringido al caso binomial tenemos que $\Omega = \{\omega_1, \omega_2\}$. Deberíamos denotar como $K(n, \omega)$ a la ganancia obtenida en el paso n si el mercado sigue el escenario $\omega \in \Omega$.

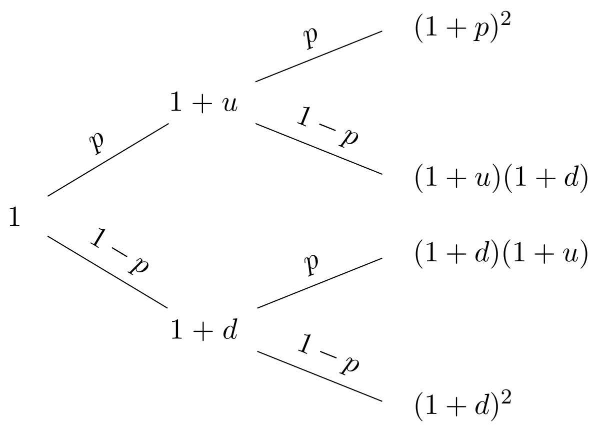

Figura 8.1: Ganancias en un árbol binomial de dos pasos.

Por lo tanto, si denotamos al precio de una activo en el paso $n \in \mathbb{N}$ como $S(n)$ tenemos que:

$$S(n) = S(0)(1+u)^i(1+d)^{n-i} \text{ con probabilidad } \binom{n}{i} p^i(1-p)^{n-i},$$

donde $S(0)$ es el precio actual del activo.

Los activos que hasta ahora hemos presentado con denominados *primarios* porque son independientes de otros títulos de valor. Por otro lado tenemos los activos *derivados* que son aquellos cuyo valor cambia en función de otros activos denominados subyacentes que pueden ser primarios u otros derivados. Ejemplos de activos derivados son:

1. Contrato forward (a plazo): es un acuerdo entre dos partes para comprar o vender cierto activo con riesgo a un precio fijo en un momento determinado en el futuro.

2. Contrato de futuros: es un tipo de contrato forward pero que está estandarizado y negociado en un mercado organizado.

3. Opciones: es un contrato mediante el cual el comprador de la opción adquiere el derecho pero no la obligación de comprar o vender un activo subyacente al vendedor de la misma. El precio al que se puede ejercer el derecho de compra o de venta del activo se denomina precio de ejercicio o también strike price. Existen dos tipos de opciones:

    a) Europeas: solo pueden ser ejercidas en la fecha de vencimiento.

    b) Americanas: se puede ejercer en cualquier momento hasta la fecha de vencimiento.

    A su vez, distinguimos entre:

    - Opciones de compra (*call*): otorga al poseedor de la misma la posibilidad de comprar el activo.
    - Opciones de venta (*put*): da al poseedor de la misma la posibilidad de vender el activo.

    En este trabajo nos centraremos en las opciones europeas.

Si la situación es la que se ha explicado anteriormente, el comprador de una opción puede obtener un beneficio sin riesgo. Por ejemplo, supongamos que tenemos un opción que nos otorga el derecho de comprar un bien a un determinado precio. Si en el momento de ejercer dicha opción el precio de mercado es más bajo, no ejerzo el derecho y la compro a precio de mercado. Por otro lado, si es más alto puedo ejercer la opción y pagar menos dinero. De este modo, si $S(t)$ representa la variable aleatoria que modela el precio del activo y K es el precio acordado en la opción, el *payoff* obtenido en una opción call es $C(T) = \text{máx}\{S(T) - K, 0\}$ y de una opción put $C(T) = \text{máx}\{K - S(T), 0\}$. Por eso, el comprados debe pagar una especie de tasa para obtener una opción. Este precio no puede ser ni demasiado bajo ni demasiado alto ya que, de ese modo, nadie compraría la opción. Si representamos por C_0 el precio de una opción y no tenemos en cuenta que el precio del dinero cambia con el tiempo, las ganancias o pérdidas se representan en las gráficas 8 y 8.

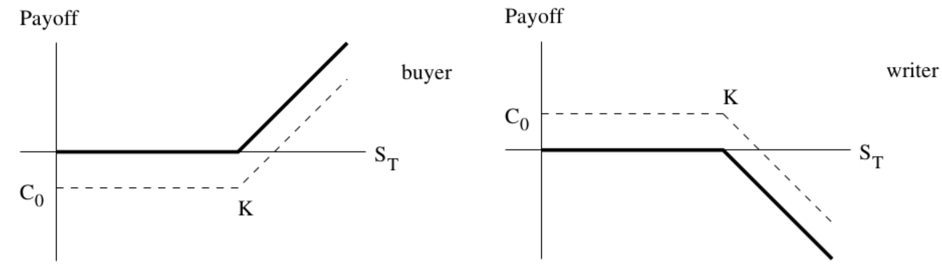

Figura 8.2: *Payoff* opción call.

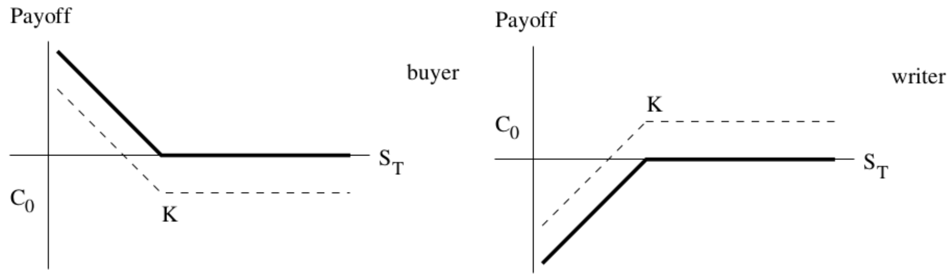

Figura 8.3: *Payoff* opción put.

Uno de los problemas más importantes es determinar de manera única el precio "justo" de una opción en un momento determinado para que ambas partes estén de acuerdo.

Una vez explicada de manera simplificada el contexto financiero debemos formalizar matemáticamente modelo del mercado. Fijamos el conjunto de las marcas de tiempo $\mathbb{T} = \{0, 1, \ldots, T\}$ con T el instante en el que finaliza el modelado de nuestra actividad económica. También fijamos el espacio de probabilidad $(\Omega, \mathcal{F}, P)$ con Ω finito. Este espacio contiene todos los posibles estados del mercado. La información de la que disponen los inversores acerca de la estructura en cada momento $t \in \mathbb{T}$ viene dada por una secuencia finita y creciente de sub-σ-álgebras de $\mathcal{F}$, $\mathcal{F}_0 \subset \cdots \subset \mathcal{F}_T = \mathcal{F}$ con $\mathcal{F}_0$ trivial, es decir, contiene solamente conjuntos con medida 0 o 1. A este secuencia se le denomina filtración y se denota como $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$. Llamamos d a la dimensión de nuestro mercado, es decir, el número de activos que manejamos. La evolución de los precios de los activos viene dada por el proceso estocástico $S = \{S_t^i : \ t \in \mathbb{T}, \ i = 0, \ldots, d\}$. Todos son activos con riesgos menos el marcado por 0. Asumimos que este proceso se adapta a la filtración $\mathbb{F}$, es decir, para cada $t \in \mathbb{T}$ se conocen los precios de cada activo hasta ese instante. En ese caso, decimos que S_t^i es $\mathcal{F}_t$-medible para $i = 0, \ldots, d$. En general, primero se escoge S y se toma $\mathbb{F}$ como la filtración que genera. La tupla $(\Omega, \mathcal{F}, P, \mathbb{F}, S)$ es la que modela nuestro mercado de valores.

El poder adquisitivo de una misma cantidad decrece con el tiempo, consideramos el factor de descuento (o de actualización) en el instante t dado por $(0 <?)\beta_t < 1$ y notamos al valor actual como $\bar{S}_t = \beta_t S_t$. Sin pérdida de generalidad se puede asumir que $S^0(0) = 1$ por lo que podemos expresar todas las unidades en función a dicho valor. En ese caso, el factor $\beta_t = 1/S_t^0$ es la cantidad de dinero que necesitamos para invertir en bonos en el instante 0 para tener una unidad en el instante t.

La cartera de inversión o portafolios para el instante $t \in \mathbb{T}$ viene dada por la variable aleatoria $d+1$ dimensional $\theta_t = (\theta_t^0, \ldots, \theta_t^d)$. Su valor viene determinado por $V_t(\theta)$ donde

$$V_0(\theta) = \langle \theta_1, S_0 \rangle, \quad V_t(\theta) = \langle \theta_t, S_t \rangle = \sum_{i=0}^{d} \theta_t^i S_t^i \text{ para } t \in \mathbb{T}, t \geq 1.$$

Cada inversor seleccionan la cartera de inversión del instante t una vez se conocen los precios del momento anterior $t-1$. La estrategia de inversión $\theta = \{\theta_t : t = 1, \ldots, T\}$ es el conjunto de todos los portafolios *predecibles*. Decimos que un portafolio θ_t es predecible si depende solamente de los precios de los activos hasta el instante $t-1$.(Resto de suposiciones??). Cuando cada cartera de inversión θ_t se puede financiar completamente por la ganancia actual decimos que la estrategia es *autofinanciada*, es decir,

$$\langle \theta_{t+1}, S_t \rangle \leq \langle \theta_t, S_t \rangle, \tag{8.1}$$

con igualdad si se ha usado todo el valor. Notamos $\Delta X_t = X_t - X_{t-1}$ para cualquier función sobre $\mathbb{T}$. Si suponemos que en 8.1 se da la igualdad, escribimos la ganancia de una cartera de inversión entre los instantes $(t-1, t]$ como

$$\Delta V_t(\theta) = \langle \theta_t, S_t \rangle - \langle \theta_{t-1}, S_{t-1} \rangle = \langle \theta_t, S_t \rangle - \langle \theta_t, S_{t-1} \rangle = \langle \theta_t, \Delta S_t \rangle.$$

De este modo, la ganancia asociada a la estrategia θ hasta el instante t viene dada por

$$G_0(\theta) = 0, \quad G_t(\theta) = \langle \theta_1, \Delta S_1 \rangle + \cdots + \langle \theta_t, \Delta S_t \rangle \text{ para } t \in \mathbb{T}, t \geq 1.$$

Vemos entonces que una estrategia es autofinanciada si, y solo si

$$V_t(\theta) = V_0(\theta) + G_t(\theta), \ \forall t \in \mathbb{T}. \tag{8.2}$$

Dentro de las estrategias autofinanciadas nos interesan las denominadas *admisibles* que son aquellas tales que $V_t(\theta) \geq 0$ para todo $t \in \mathbb{T}$. A la clase de estrategias autofinanciadas y admisibles la denotaremos como Θ_a. Destacaos que todas la definiciones anteriores se pueden definir también en base a los valores actualizados $\bar{S}_t$.

Exponemos ahora unos resultados que nos servirán posteriormente para probar el teorema principal del trabajo.

Lema 8.1. *Dada V_0 una función $\mathcal{F}_0$-medible y para $d \in \mathbb{N}$ sean los procesos reales y predecibles $\theta^1, \ldots, \theta^d$, el único proceso predecible θ^0 que convierte a $\theta = (\theta^0, \theta^1, \ldots, \theta^d)$ en una estrategia autofinanciada con valor $V_0(\theta) = V_0$ viene dado por*

$$\theta_t^0 = V_0 + \sum_{i=1}^{t-1} (\theta_i^1 \Delta \bar{S}_i^1 + \cdots + \theta_i^d \Delta \bar{S}_i^d) - (\theta_t^1 \bar{S}_{t-1}^1 + \cdots + \theta_t^d \bar{S}_{t-1}^d).$$

Demostración. Claramente θ^0 es predecible. Para ver que la estrategia es autofinanciada, por 8.2 solo necesitamos ver que θ_t^0 es la única solución predecible de la ecuación

$$\begin{aligned} \bar{V}_t(\theta) &= \theta_t^0 + \theta_t^1 \bar{S}_t^1 + \cdots + \theta_t^d \bar{S}_t^d \\ &= V_0 + \sum_{i=1}^{t} (\theta_i^1 \Delta \bar{S}_i^1 + \cdots + \theta_i^d \Delta \bar{S}_i^d) \end{aligned}$$

■

Introducimos ahora el concepto de *viabilidad*. Decimos que un mercado es viable si para toda estrategia admisible y autofinanciada no contiene ninguna oportunidad de arbitraje. La ausencia de arbitraje significa que si el valor inicial de una cartera de inversión es $V_0(\theta) = 0$ entonces $V_T(\theta) = 0$ con probabilidad 1 para toda $\theta \in \Theta_a$.

Lema 8.2. *Si el modelo de mercado es viable, las ganancias actualizadas a cualquier proceso $\hat{\theta} \in \mathbb{R}^d$ no pueden pertenecer a*

$$C = \{Y \in \mathbb{R}^n : Y_i \geq 0 \text{ para } i = 1, \ldots, n \text{ y } \exists i \text{ tal que } Y_i > 0\}.$$

Demostración. En primero lugar vemos que C es el octante positivo de $\mathbb{R}^n$ sin el origen que claramente es un cono y es convexo. La ausencia de arbitraje significa que para toda estrategia admisible $\theta \in \Theta_a$ tenemos que si $V_0(\theta) = 0$ entonces

$$\bar{V}_t(\theta) = \bar{G}_t(\theta) \notin C.$$

Por el lema (8.1), dados los procesos predecibles $\hat{\theta} = (\theta^1, \ldots, \theta^d)$, existe un único proceso real θ^0 tal que $\theta = (\theta^0, \theta^1, \ldots, \theta^d)$ es autofinanciada y $V_0(\theta) = 0$. Las ganancias con los valores actualizados viene dada por

$$\bar{G}_t(\hat{\theta}) = \sum_{j=1}^{t} \langle \theta_j, \Delta \bar{S}_j \rangle = \sum_{j=1}^{t} \left(\sum_{i=1}^{d} \theta_j^i \Delta \bar{S}_j^i \right).$$

Supongamos que $\bar{G}_t(\hat{\theta}) \in C$, si β_T denota el factor de descuento en el instante T,

$$V_T(\theta) = \beta_T^{-1} \bar{V}_T(\theta) = \beta_T^{-1}(V_0(\theta) + \bar{G}_T(\theta)) = \beta_T^{-1} \bar{G}_T(\theta).$$

Vemos que entonces $V_T(\theta)$ es no negativa y estrictamente positiva con probabilidad no nula, lo que contradice la viabilidad al existir arbitraje. ■

Nuestro objetivo es caracterizar la viabilidad de un mercado en términos de los incrementos de $\bar{S}$. Para ello son necesarias las martingalas.

Definición 8.1. *Un proceso $\mathbb{F}$-adaptado $M = (M_t)_{t \in \mathbb{T}}$ es una $(\mathbb{F}, P)$-martingala si $E(|M_t|) < \infty$ para todo $t \in \mathbb{T}$ y*

$$E(M_{t+1} | \mathcal{F}_t) = M_t \text{ para todo } t \in \mathbb{T} \setminus T.$$

Si $M = (M_t)$ es una martingala y $\phi = (\phi_t)_{t \in \mathbb{T}}$ es un proceso predecible en $(\Omega, \mathcal{F}, P, \mathbb{F}, S)$ entonces al proceso $X = \phi \cdot M$ dado por

$$X_0 = 0, \quad X_t = \phi_1 \Delta M_1 + \cdots + \phi_t \Delta M_t \quad t \geq 1$$

se le denomina martingala transformada de M por ϕ.

Notamos que M es una martingala si, y solo, si

$$E(\Delta M_{t+1} | \mathcal{F}_t) = 0 \text{ para todo } t \in \mathbb{T} \setminus \{T\}.$$

Teorema 8.1. *Un proceso real M es una martingala si, y solo si,*

$$E((\phi \cdot M)_t) = E(\sum_{i=1}^{t} \phi_i \Delta M_i) = 0, \ \forall t \in \mathbb{T} \setminus \{0\}$$

para todo proceso predecible acotado ϕ.

Demostración. Si M es un martingala, también lo es la tranfomada $X = \phi \cdot M$ y $X_0 = 0$ ya que

$$E(\Delta X_{t+1} | \mathcal{F}_t) = E(\phi_{t+1} \Delta M_{t+1} | \mathcal{F}_t) = \phi_{t+1} E(\Delta M_{t+1} | \mathcal{F}_t) = 0.$$

Por ello, $E((\phi \cdot M)_t) = 0$ para todo $t \geq 1$ en $\mathbb{T}$. Demostremos ahora la otra implicación. Sea $A \in \mathcal{F}_s$ dada y definimos el proceso predecible ϕ como $\phi_{s+1} = 1_A$ y $\phi_t = 0$ para todo $t \in \mathbb{T}$. Entonces, para $t > s$ se tiene que

$$0 = E((\phi \cdot M)_t) = E(1_A(M_{s+1} - M_s)).$$

Como es cierto para todo $A \in \mathcal{F}_s$, se cumple que $E(\Delta M_{s+1}|)\mathcal{F}_s = 0$ por lo que M es una martingala. ■

Nos encontramos ahora un contexto general donde no asumimos que el modelo sea finito o que $\mathbb{F}$ sea generada por S. Supongamos que el proceso de los precios actualizados $\bar{S}$ es una martingala bajo una probabilidad Q, esto es:

$$E_Q(\Delta \bar{S}_t^i | \mathcal{F}_{t-1}) = 0, \text{ para } t \in \mathbb{T} \setminus \{0\} \text{ e } i = 0, \dots, d.$$

Sea $\theta \in \Theta_a$ una estrategia admisible tal que los procesos de los precios actualizados son integrables respecto a Q, por (8.2) tenemos que:

$$\begin{aligned} \bar{V}(\theta) &= V_0(\theta) + \bar{G}_t(\theta) \\ &= \langle \theta_1, S_0 \rangle + \sum_{u=1}^{t} \langle \theta_u, \Delta \bar{S}_u \rangle \\ &= \sum_{i=1}^{d} (\theta_1^i S_0^i + \sum_{u=1}^{t} \theta_u^i \Delta \bar{S}_u^i). \end{aligned}$$

Vemos entonces que $\bar{V}(\theta)$ es una constante más una suma finita de martingalas transformadas, por lo que también es una martingala con valor inicial $V_0(\theta)$. Entonces, tenemos que $E(\bar{V}_t(\theta)) = E(V_0(\theta)) = V_0(\theta)$. Esta situación imposibilita la existencia de arbitraje. Si sabemos de antemano que los procesos de los precios actualizados son integrables respecto a Q, supongamos que $V_0(\theta) = 0$ y $V_T(\theta) \geq 0$ casi seguramente (respecto a Q). Como $E_Q = (\bar{V}_t(\theta)) = 0$ se sigue que $V_T(\theta) = 0$ casi seguramente (respecto a Q). Esto sigue siendo verdadero casi seguramente para P, demostrando que P y Q tienen los mismos conjuntos vacíos. Llegamos entonces a la siguiente definición:

Definición 8.2. *Una probabilidad Q que sea equivalente a P $(Q \sim P)$ es una medida martingala equivalente (EMM) para S si el proceso de los precios actualizados $\bar{S}$ es una martingala bajo Q para la filtración $\mathbb{F}$. Es decir, para casa $i \leq d$ tenemos que $\bar{S}^i$ es una $(\mathbb{F}, Q)$-martingala.*

Todo lo presentado anteriormente, nos aporta el siguiente resultado que ya se ha demostrado:

Proposición 8.1. *Si existe una medida martingala equivalente para S, entonces el modelo de mercado discreto es viable.*

Exponemos ahora el siguiente teorema que es el objetivo final de este trabajo y en el que usaremos las herramientas que hemos ido desarrollando anteriormente.

Teorema 8.2. *Un modelo de mercado discreto es viable si, y solo si, existe una medida de martingala equivalente para S.*

Demostración. Ya sabemos por la proposición (8.1) que la existencia de una medida de martingala equivalente garantiza la viabilidad del modelo por lo que solo tenemos que probar la otra implicación.

Suponemos entonces que el modelo es viable. Necesitamos construir una medida $Q \sim P$ en la que los precios son martingalas relativas a la filtración $\mathbb{F}$. Recordamos que C es el cono convexo de todas las variables aleatorias reales ϕ en $(\Omega, \mathcal{F})$ tales que $\phi(\omega) \geq 0$ casi seguramente y $\phi(\omega_i) > 0$ para al menos un $\omega_i = \Omega = \{\omega_i, \ldots, \omega_n\}$. Asumimos $p_i = P(\{\omega_i\}) > 0$. Por el lema (8.2), hemos visto que en un mercado viable debemos tener $\bar{G}_t(\hat{\theta}) \notin C$ para todos los procesos predecibles $\hat{\theta} \in \mathbb{R}^d$. Por otro lado, el conjunto definido por tales ganancias

$$L = \{\bar{G}_t(\hat{\theta}) : \hat{\theta} = (\theta^1, \ldots, \theta^d), \text{ con } \theta^i \text{ predecible para } i = 1, \ldots, d\},$$

es un subespacio lineal del espacio vectorial de todas las funciones reales y $\mathcal{F}$-medibles en Ω. Como L y C son disjuntos, podemos separar L y el subconjunto compacto de C definido como $K = \{X \in C : E_P(X) = 1\}$ mediante el teorema 6.2???. Sea $\boldsymbol{x}^0 \in \mathbb{R}^d$ el elemento que proporciona dicho resultado. Tomamos $\xi_i = (0, \ldots, \frac{1}{p_i}, \ldots, 0)$ para $i \leq n$. Vemos que $E_P(\xi_i) = \frac{p_i}{p_i} = 1$ por lo que $\xi_i \in K$ y por ello $\langle \boldsymbol{x}^0, \xi_i \rangle = \frac{x_i^0}{p_i} > 0$. Por ello, $x_i^0 > 0$ para todo $i = 1, \ldots, n$.

Definimos ahora el funcional lineal $g(\boldsymbol{x}) = \frac{\langle \boldsymbol{x}^0, \boldsymbol{x} \rangle}{\alpha}$ donde $\alpha = \sum_{i=1}^n x_i^0$. Esto es realizado por el vector p^* con $p_i^* = \frac{x_i^0}{\alpha}$ por lo que $\sum_{i=1}^n p_i^* = 1$. Usamos el vector p^* para inducir una probabilidad P^* en $\Omega = \{\omega_1, \ldots, \omega_n\}$ haciendo $P^*(\{\omega_i\}) = p_i^* > 0$ por lo que $P^* \sim P$. Sea $E^*(\cdot)$ que denota la esperanza relativa a P^*. Como $g(\boldsymbol{x}) = \frac{1}{\alpha}\langle \boldsymbol{x}^0, \boldsymbol{x} \rangle = 0$ para todo $\boldsymbol{x} \in L$, tenemos $E^*(\bar{G}_T(\hat{\theta})) = 0$ para cada vector $\hat{\theta}$ creando una estrategia auto-financiada con $V_0(\theta) = 0$. Como $\bar{V}_t(\theta) = V_0(\theta) + \bar{G}_T(\hat{\theta})$, implica que $E^*(\bar{V}_t(\theta)) = 0$ para cada θ. Por el lema (8.1) podemos generar tal θ para cada proceso predecible de n dimensiones. En particular, lo podemos hacer para $(0, \ldots, \theta^i, \ldots, 0)$

donde $i \leq n$. Por lo tanto

$$E^*(\sum_{t=1}^{T} \theta_t^i \Delta \bar{S}_t^i) = 0$$

se da para cada proceso predecible $(\theta^i)_{i=1,\dots,T}$. El teorema (8.1) implica que cada S^i es una martingala bajo P^*. Por ello, P^* es la medida martingala equivalente deseada. ■

Bibliografía

[BL10] Jonathan Borwein y Adrian S Lewis. *Convex analysis and nonlinear optimization: theory and examples.* Springer Science & Business Media, 2010.

[EK05] RJ Elliot y PE Kopp. Mathematics of financial markets springer. *New York*, 2005.

[ST08] Stephen Simons y F Takens. *From Hahn-Banach to Monotonicity*, volumen 1693. Springer, 2008.