



TRABAJO FIN DE GRADO
INGENIERÍA INFORMÁTICA Y MATEMÁTICAS

**Teoremas de la alternativa, optimización convexa,
valoración de activos financieros
y
procesamiento de nubes de puntos generadas por
escáner láser.**

Autor

Pedro Manuel Flores Crespo

Directores

Manuel Ruiz Galán

Juan Carlos Torres Cantero



Facultad de Ciencias



Granada, junio de 2020

Índice general

1. Resumen y palabras clave	5
2. Resumen extendido y palabras clave en inglés	7
3. Introducción	9
4. Objetivos	11
5. Desarrollo	13
6. Teoremas de la alternativa	15
6.1. Teorema de Mazur-Orlicz-König	15
6.2. Teorema de la alternativa de Gordan. Reformulaciones.	22
7. Alternativa y minimax	31
7.1. Una desigualdad minimax	31
7.2. Separación de convexos	34
8. Optimización	41
8.1. Lema de Farkas	41
8.2. Dualidad en programación lineal	43
8.3. Optimización con restricciones: Fritz John y Karush-Kuhn-Tucker	45
9. Valoración de activos financieros	51
9.1. Preliminares financieros	51
9.2. Primer teorema fundamental	54
9.3. Aplicación al modelo binomial	62
10. Conclusiones y vías futuras	69

Resumen y palabras clave

El trabajo fin de grado que presentamos supone la incursión a un campo dentro de la optimización, los teoremas de la alternativa, y sus aplicaciones, principalmente a la propia optimización y finanzas. El trabajo comienza con el teorema de Mazur-Orlicz-König, versión del conocido teorema de Hahn-Banach. Posteriormente, estudiamos el teorema de la alternativa de Gordan, esencial para el resto de resultados. Su primera aplicación se da en la teoría minimax, que nos conduce a resultados clásicos sobre separación de convexos. A continuación, deduciremos otro teorema de la alternativa, el de Farkas, y lo aplicaremos a la programación lineal. Para demostrar los resultados sobre optimización, volvemos al teorema de Gordan, que nos proporcionara los teoremas de Fritz John y Karush-Kuhn-Tucker. Finalmente, nos introducimos en el mundo de las matemáticas financieras obteniendo, gracias al teorema de separación, el primer teorema de asignación de precios para valorar opciones europeas. Concluimos realizando simulaciones de su valor en diferentes casos.

Palabras clave: teoremas de la alternativa, teorema de Hahn-Banach, minimax, optimización, matemáticas financieras.

Resumen extendido y palabras clave en inglés

The end of degree project that we present is an incursion into a field included in optimization, the theorems of the alternative, and their applications, mainly to the own optimization and finances.

Many of the theorems of the alternative are just reformulations of convex separation's theorems in certain contexts, so we start this memory with a study on Hahn-Banach's theorem. There are several equivalent versions of this result, collected for example in [7]. That is the reason why many authors talk about Hahn-Banach's theorems. We will focus on one of them, Mazur-Orlicz-König's theorem. Despite its geometrical charge, it is mainly an algebraical result. We will provide a proof of that theorem – the results in this work are self-contained – and we will use it to give a convex version of Gordan's theorem of alternative. Actually, it is an equivalent result due to S. Simons. This version is more general than the original Gordan's result and it will be extremely useful in the next important results.

Then, we will apply the convex Gordan's theorem of the alternative in the minimax theory. A minimax inequality guarantees, under certain hypothesis, that in a two variables function we can substitute $\inf \sup$ for $\sup \inf$. The power of minimax inequalities is clear when we use it – in one that we will deduce from the convex Gordan's theorem – to give classical convex separation's results.

Next, we will deduce from one separation's theorem the Farkas's lemma, which is one of the most known theorem of the alternative. This will allow us to prove, almost immediately, one of the key points of linear programming: the duality theorem. Considering more general optimization theorems, those in which the objective function and the inequalities constraints are differentiables, we will establish, using Gordan's theorem of the alternative, the theorems of Fritz John and Karush-Kuhn-Tucker.

We will conclude with a little foray into the field of financial mathematics. After introducing some of the main concepts, like a derivative security; we will prove, as a consequence of one of the convex separation's theorem, the result known as “First Fundamental Theorem of Asset Pricing”. It can be used in the pricing of European's options in the binomial model. Finally, we have programmed using *Sagemath 2.8*

some examples in order to know how their price varies according to its parametres.

Keywords: theorems of the alternative, Hahn-Banach's theorem, minimax, optimization, financial mathematics.

Introducción

El estudio de los teoremas de la alternativa hunde sus raíces en el teorema de separación de convexos de Hahn-Banach. Es más, algunas de sus versiones más conocidas, como el teorema de la alternativa de Gordan o el lema de Farkas, son precursoras de ese resultado fundamental del análisis funcional o la optimización. Y es precisamente ahí, en un contexto de optimización, donde surgen hace casi siglo y medio. Desde entonces han aparecido en gran cantidad, vinculados a problemas de optimización convexa en muchas ocasiones y mediante el uso de técnicas de separación. Además, su aplicabilidad no se ha circunscrito exclusivamente al campo de la optimización sino que ha trascendido dicha área: análisis convexo, análisis funcional, problemas de equilibrio ...

En esta memoria se aborda el estudio de los dos resultados de la alternativa mencionados, el teorema de Gordan y el lema de Farkas, dando incluso una versión más general del primero. Para establecerlos usamos una versión muy simple del teorema de Hahn-Banach y una mejora del mismo. Además, se aplican para establecer una desigualdad minimax que deriva, en particular, en una serie de teoremas de separación de convexos. También se ilustra su aplicabilidad en establecer resultados centrales en la optimización, como es el teorema de dualidad en programación lineal o los teoremas de Fritz John y Karush-Kuhn-Tucker en un contexto diferenciable. Finalmente, dedicamos todo un capítulo de la memoria a usar los teoremas de la alternativa, en una de sus formas equivalentes, para demostrar un resultado importante en el ámbito de las matemáticas financieras, el primer teorema fundamental de valoración de activos financieros. Ello requiere un bagaje previo – conceptos y resultados – que también se recoge en la memoria. Dicho teorema se aplica al caso de ciertos derivados muy populares, las opciones europeas, bajo un modelo binomial y se presentan algunas simulaciones numéricas realizadas con *SageMath* en su versión 2.8.

En definitiva, en esta memoria se plasma tanto el carácter convexo-funcional de los teoremas de la alternativa como su aplicabilidad a campos tan diversos como la optimización, el análisis convexo y las matemáticas financieras. Señalamos finalmente que las referencias usadas en la elaboración de esta memoria aparecen recogidas en el capítulo de Bibliografía. No obstante, los textos de [1], [2] y [8] han sido los esenciales.

Objetivos

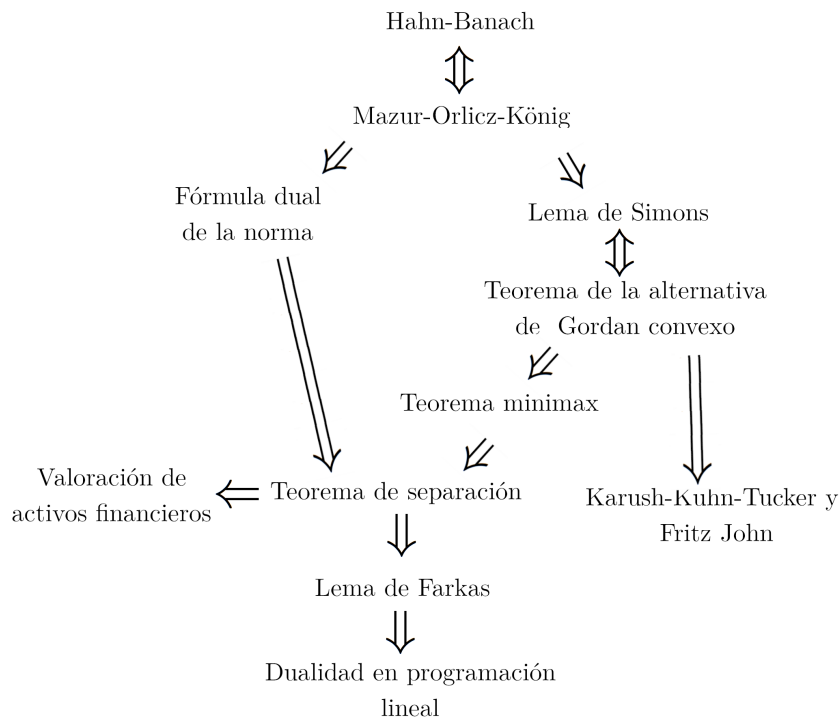
Los objetivos inicialmente previstos en la propuesta de TFG fueron:

- Realizar una recopilación de algunos teoremas de la alternativa.
- Teorema de dualidad en programación lineal.
- Teoremas de Karush-Kuhn-Tucker y Fritz John para programación convexa.
- Aplicación a finanzas: teorema fundamental de valoración de activos financieros en mercados finitos.

Sin embargo, nuestro tratamiento final ha sido algo más ambicioso, pues hemos incluido todo un capítulo de aplicaciones de los teoremas de la alternativa a la teoría minimax y a la separación de convexos. Además, en lugar de considerar los teoremas de Karush-Kuhn-Tucker y Fritz John en un ambiente convexo, los hemos establecido en un contexto no lineal y diferenciable. La idea que nos ha llevado a ello ha sido aumentar el número y tipología de aplicaciones de los teoremas de la alternativa, mostrando su versatilidad en diversos campos.

Desarrollo

El proceso seguido en el desarrollo del TFG ha sido, por un lado, recopilar material sobre el tema y analizarlo, y por otro, darle estructura totalmente auto-contenida, elaborando los diversos contenidos de forma jerarquizada en el sentido de que se deducen de los anteriores. A modo de esquema, los resultados se han estructurado atendiendo al siguiente esquema donde además se recoge la relación entre ellos:



Como puede observarse, las técnicas son de carácter convexo y analítico funcional.

Teoremas de la alternativa

Los teoremas de la alternativa constituyen una potente herramienta en optimización. A pesar de tener un claro carácter convexo, se aplican incluso a problemas no convexos, tal y como se comprobará a lo largo de esta memoria. Nuestro punto de partida es el resultado del análisis convexo, más importante, el teorema de Hahn-Banach. Es más, daremos una versión equivalente debida a H. König, conocida como el teorema de Mazur-Orlicz-König. Como consecuencia, obtendremos el teorema de la alternativa de Gordan, tanto en su versión clásica como una más general.

6.1. Teorema de Mazur-Orlicz-König

El objetivo principal de esta sección es demostrar una versión equivalente no muy conocida del clásico teorema de Hahn-Banach, el teorema de Mazur-Orlicz-König. Iremos de una versión básica y algebraica del teorema de Hahn-Banach al teorema de Mazur-Orlicz-König siguiendo como aparece en el texto de S. Simons [8].

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 conjuntos que usaremos asumiremos que son no vacíos.

Definición 6.1. Sea V un espacio vectorial. Decimos que el $P : V \rightarrow \mathbb{R}$ es sublineal si cumple las siguientes condiciones:

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

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

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

Por ejemplo, toda norma o incluso toda seminorma sobre V es un funcional sublineal. Así, dados $a, b \in \mathbb{R}$ con $a < b$ tenemos que $P : H^1(a, b) \rightarrow \mathbb{R}$ 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

$$P(x) = [x]^+ = \max\{0, x\}, \quad \forall x \in \mathbb{R}$$

obtenemos un funcional sublineal sobre \mathbb{R} .

El lema que exponemos a continuación, y que generalizaremos posteriormente en el lema 6.2, nos servirá para demostrar el teorema de Hahn-Banach.

Lema 6.1. *Sea V un espacio vectorial y $P : V \rightarrow \mathbb{R}$ un funcional sublineal. Fijamos un elemento $y \in V$ y para todo $x \in V$ tomamos*

$$P_y(x) := \inf_{\lambda > 0} [P(x + \lambda y) - \lambda P(y)].$$

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

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, $P_y(V) \subset \mathbb{R}$, esto es $P_y : V \rightarrow \mathbb{R}$.

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, la definición de P_y da

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

Tomando ínfimos 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:

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

Obtenemos que P_y es positivamente homogénea y como consecuencia sublineal.

Para demostrar que $P_y \leq P$, sea $x \in V$ y tomando $\lambda = 1$ en la definición de P_y ,

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

Como $x \in V$ es arbitrario, entonces $P_y \leq P$. Finalmente, razonando de manera similar,

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

■

El teorema de Hahn-Banach que hemos mencionado antes (básico y algebraico) se enuncia en estos términos:

Teorema 6.1 (Hahn-Banach). *Sea V un espacio vectorial y $P : V \rightarrow \mathbb{R}$ un funcional sublineal. Entonces existe un funcional lineal $L : V \rightarrow \mathbb{R}$ tal que $L \leq P$.*

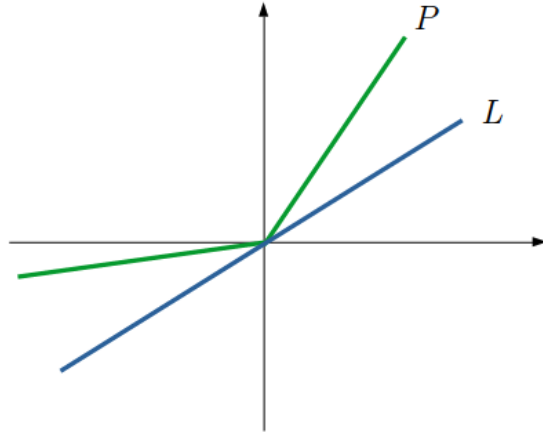


Figura 6.1: Teorema de Hahn-Banach.

Demostración. Sea $SUB = \{Q : V \rightarrow \mathbb{R} : Q \leq P\}$, es decir, el conjunto no vacío de funcionales sublineales sobre V que minoren a P . Nuestro propósito es emplear el lema de Zorn con objeto de probar que tiene un elemento minial y tal elemento será el funcional L que buscamos. Para ello, dados $T_1, T_2 \in SUB$ consideramos la relación de orden usual, es decir:

$$T_1 \leq T_2 \iff T_1(x) \leq T_2(x) \quad \forall x \in V.$$

Primero probaremos que todo subconjunto \mathcal{T} totalmente ordenado de SUB tiene una cota inferior en SUB . Definimos $Q(x) := \inf\{T(x) : T \in \mathcal{T}\}$ 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:

$$0 = T(0) = T(x - x) \leq T(x) + T(-x) \implies T(x) \geq -T(-x) \quad (6.1)$$

Por otro lado:

$$T \in \mathcal{Q} \implies T(x) \leq P(x) \implies -T(x) \geq -P(x) \quad (6.2)$$

Usando (6.1), (6.2) y tomando ínfimo sobre T llegamos a $Q(x) \geq -P(x) > -\infty$. Por lo tanto $Q : V \rightarrow \mathbb{R}$.

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 $\mu > 0$:

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

De este modo, Q es sublineal y como es claro que $Q \leq P \implies Q \in \text{SUB}$. 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, $L_y : V \rightarrow \mathbb{R}$ es sublineal, $L_y \leq L$ (como consecuencia $L_y \in \text{SUB}$) 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. ■

Nos disponemos a probar, a partir del teorema de Hahn-Banach el de Mazur-Orlicz-König, que supone un refinamiento. Destacamos que el teorema de Hahn-Banach tiene gran cantidad de versiones equivalentes. Véase, por ejemplo, [7]. Antes, necesitamos un resultado técnico, que constituye una especie de versión global sobre un convexo del lema 6.1.

Lema 6.2. 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, $Q(V) \subset \mathbb{R}$, $Q : V \rightarrow \mathbb{R}$ es sublineal, $Q \leq P$ y además $\forall d \in D$ se cumple $-Q(-d) \geq \beta$.

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

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

La primera desigualdad 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

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

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

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

por lo que $Q(V) \subset \mathbb{R}$. 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:

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

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. Para concluir, fijamos $d \in D$. Sea $x \in V$ arbitrario. Entonces, $\forall \lambda > 0$, $Q(x) \leq P(x) + \lambda(P(d) - \beta)$. Tomando $\lambda \rightarrow 0$, $Q(x) \leq P(x)$ y como consecuencia $Q \leq P$. Finalmente, sea $d \in D$ arbitrario y tomando $\lambda = 1$:

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

■

Ya estamos preparados para probar el teorema de Mazur-Orlicz-König, debido a H. König [4, 5]. No solo es una versión equivalente del teorema de Hahn-Banach, también de un teorema de Mazur-Orlicz [6].

Teorema 6.2 (Mazur-Orlicz-König). *Sea V un espacio vectorial y $P : V \rightarrow \mathbb{R}$ un funcional sublineal. Sea D un subconjunto no vacío y convexo de V . Entonces existe un funcional $L : V \rightarrow \mathbb{R}$ lineal tal que $L \leq P$ e $\inf_D L = \inf_D P$.*

Antes de pasar a la demostración, observemos que, aunque es un resultado puramente algebraico, tiene una fuerte interpretación geométrica: el teorema de Hahn-Banach garantiza, dado un funcional sublineal $P : V \rightarrow \mathbb{R}$, la existencia de un funcional lineal $L : V \rightarrow \mathbb{R}$ con $L \leq P$. En la figura 6.2 mostramos unos ejemplos

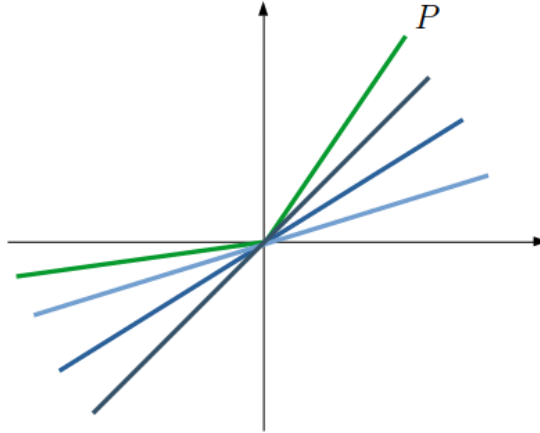


Figura 6.2: Ejemplos de funcionales aportados por el teorema de Hahn-Banach.

de funcionales que nos aporta el teorema de Hahn-Banach.

El teorema de Mazur-Orlicz-König, fijado un subconjunto convexo D de V , nos da solo los funcionales lineales que minoran a P y que cumplen

$$\inf_D L = \inf_D P.$$

En la imagen 6.3 se muestra un ejemplo del teorema.

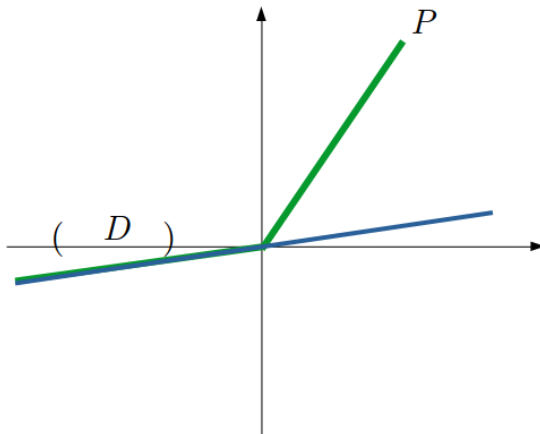


Figura 6.3: Teorema de Mazur-Orlicz-König.

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

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

Supongamos entonces que $\beta \in \mathbb{R}$. Definimos el funcional auxiliar Q como en el lema 6.2. 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 $L \leq P$:

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

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

Antes de probar la eficiencia del teorema de Mazur-Orlicz-König en el siguiente apartado, presentamos una consecuencia bien conocida. En particular, nos será de utilidad posteriormente para el teorema de separación. Recordemos dados E_1, E_2 dos espacios normados y $T : E_1 \longrightarrow E_2$ un operador lineal, entonces T es continuo si, y solo si, verifica la siguiente condición:

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

Consideramos el espacio vectorial dado por:

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

conocido como el espacio dual (topológico) de E . Para todo $T \in E^*$ definimos su norma como:

$$\|T\| = \min\{\alpha > 0 : |T(x)| \leq \alpha \|x\| \quad \forall x \in E\}.$$

De este modo, podemos escribir:

$$|T(x)| \leq \|T\| \|x\|$$

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

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

Para $x \in E_1 \setminus \{0\}$ tenemos que $|T(x)| / \|x\| = \|T(x/\|x\|)\|$ 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 $|T(x)| = \|x\| \|T(u)\| \leq \|T(u)\|$ ya que $\|x\| \leq 1$. De este modo, también tenemos que:

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

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

Corolario 6.1. *Dado un espacio normado E y $x \in E$, entonces se cumple que:*

$$\sup_{x^* \in B_{E^*}} x^*(x) = \|x\|. \quad (6.3)$$

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-König, teorema 6.2, y obtenemos que existe un funcional $L : E \longrightarrow \mathbb{R}$ lineal tal que $L \leq P$ e $\inf_D L = \inf_D P$. 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 decir, $x^* \in B_{E^*}$. Por su parte, como $\inf_D L = \inf_D x^* = \inf_D P$ 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. ■

6.2. Teorema de la alternativa de Gordan. Reformulaciones.

Una vez que disponemos de la herramienta fundamental, el teorema de Mazur-Orlicz-König, nos disponemos a dar una versión convexa del teorema de la alternativa de Gordan. Antes de ello, y haciendo uso de este teorema tipo Hahn-Banach, probaremos el lema de Simons, sobre cierta condición de optimalidad para funciones convexas.

Antes de comenzar, necesitamos hacer la siguiente definición. Para agilizar la lectura, cuando indicamos $N \in \mathbb{N}$ estamos suponiendo que $N \geq 1$. También, para $N \in \mathbb{N}$, notaremos con caracteres en negrita a los elementos de \mathbb{R}^N , por ejemplo, $\mathbf{t} = (t_1, \dots, t_N) \in \mathbb{R}^N$.

Definición 6.2. *Dado $N \in \mathbb{N}$ llamamos **símplex unitario** de \mathbb{R}^N al subconjunto de \mathbb{R}^N definido como:*

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

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

Antes de continuar veamos que Δ_N es convexo y compacto:

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

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

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

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

Por lo tanto, $\lambda \mathbf{t} + (1 - \lambda) \mathbf{s} \in \Delta_N$ 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. Pero claramente es acotado por lo que nos centraremos en probar que es cerrado. Sea $\{\mathbf{t}_n\}_{n \in \mathbb{N}}$ una sucesión de Δ_N y sea $\mathbf{t}_0 \in \mathbb{R}^N$ tal que $\{\mathbf{t}_n\}_{n \in \mathbb{N}} \rightarrow \mathbf{t}_0$. Tenemos que comprobar que $\mathbf{t}_0 \in \Delta_N$.

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

Así, hemos demostrado que $\mathbf{t}_0 \in \Delta_N$ 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 $L : \mathbb{R}^N \rightarrow \mathbb{R}$ es lineal, entonces es de la forma $L(\mathbf{x}) = a_1 x_1 + \dots + a_N x_N$, que se corresponde con $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{R}^N$. 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:

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{R}^N \times \mathbb{R}^N &\rightarrow \mathbb{R} \\ (\mathbf{x}, \mathbf{y}) &\mapsto \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^N x_i y_i. \end{aligned}$$

24 6.2. Teorema de la alternativa de Gordan. Reformulaciones.

Antes de enunciar el lema de Simons, hacemos otra identificación, esta vez, correspondiente a un símplex unitario:

Lema 6.3. Sea $N \in \mathbb{N}$ y $S : \mathbb{R}^N \rightarrow \mathbb{R}$ definida por:

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

Entonces, S es sublineal. Además, si $L : \mathbb{R}^N \rightarrow \mathbb{R}$ es un funcional lineal tal que $L \leq S$ entonces L es de la forma

$$L(\mathbf{x}) = t_1 x_1 + \dots + t_N x_N$$

con $(t_1, \dots, t_N) \in \Delta_N$. De hecho, el recíproco también es cierto, es decir, si $L = \mathbf{t} = (t_1, \dots, t_N) \in \Delta_N$ entonces $L \leq S$.

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

$$\begin{aligned} S(\mathbf{x} + \mathbf{y}) &= \max\{x_1 + y_1, \dots, x_N + y_N\} \\ &\leq \max\{x_1, \dots, x_N\} + \max\{y_1, \dots, y_N\} = S(\mathbf{x}) + S(\mathbf{y}). \end{aligned}$$

Por ello, S es sublineal. Para terminar veamos que

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

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

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

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

donde la primera desigualdad se debe a que $x_i \leq S(\mathbf{x})$ para todo x_i con $i = 1, \dots, N$ 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 $L = \mathbf{t} = (t_1, \dots, t_N) \in \mathbb{R}^N$ lineal tal que para todo $\mathbf{x} \in \mathbb{R}^N$ cumple que $\sum_{i=1}^N t_i x_i \leq S(\mathbf{x})$. 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 $i = 1, \dots, N$, entonces

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

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

$$\left. \begin{aligned} L(e) &= \sum_{i=1}^N t_i \leq \max\{1, \dots, 1\} = 1 \\ L(-e) &= -\sum_{i=1}^N t_i \leq \max\{-1, \dots, -1\} = -1 \end{aligned} \right\} \implies \sum_{i=1}^N t_i = 1.$$

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

■

Enunciamos ahora el lema de Simons [8]. Tal y como se ha mencionado antes, se trata de un resultado sobre funciones convexas y cierta condición de optimalidad: dos funciones distintas poseen el mismo ínfimo en un convexo. Esto sugiere el uso del teorema de Mazur-Orlicz-König.

Lema 6.4 (Simons). *Sea C un subconjunto convexo de un espacio vectorial y $N \in \mathbb{N}$. Dadas $f_1, \dots, f_N : C \rightarrow \mathbb{R}$, funciones convexas, entonces 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 $V = \mathbb{R}^N$. Definimos $S : V \rightarrow \mathbb{R}$ como

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

Por el lema 6.3, S es sublineal. Tomamos el subconjunto:

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

Comprobemos en primer lugar que D es un subconjunto convexo de V . Sean $\mathbf{x}, \mathbf{y} \in D$, por ello, existen $c_x, c_y \in C$ tales que $f_i(c_x) \leq x_i$ y $f_i(c_y) \leq y_i \quad \forall i = 1, \dots, N$. 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 y y esté en D . Así, para todo $i = 1, \dots, N$:

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

donde la primera desigualdad se debe a que las f_i son convexas y la segunda a que $\mathbf{x}, \mathbf{y} \in D$. 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-Orlicz-König, existe L funcional lineal en V tal que $L \leq S$ e $\inf_D L = \inf_D S$.

Nuevamente, por el lema 6.3 tenemos que $L = \mathbf{t} \in \Delta_N$. Finalmente:

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

e

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

por lo que

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

■

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

Teorema 6.3 (Teorema de la alternativa de Gordan). *Dados $N, M \in \mathbb{N}$, sean $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ con $\mathbf{x}_i \in \mathbb{R}^M$ para $i = 1, \dots, N$. Entonces una, y solo una, de la siguientes afirmaciones se cumple:*

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

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

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. Antes de continuar, destacar que la interpretación geométrica de este resultado se observa en la figura 6.4.

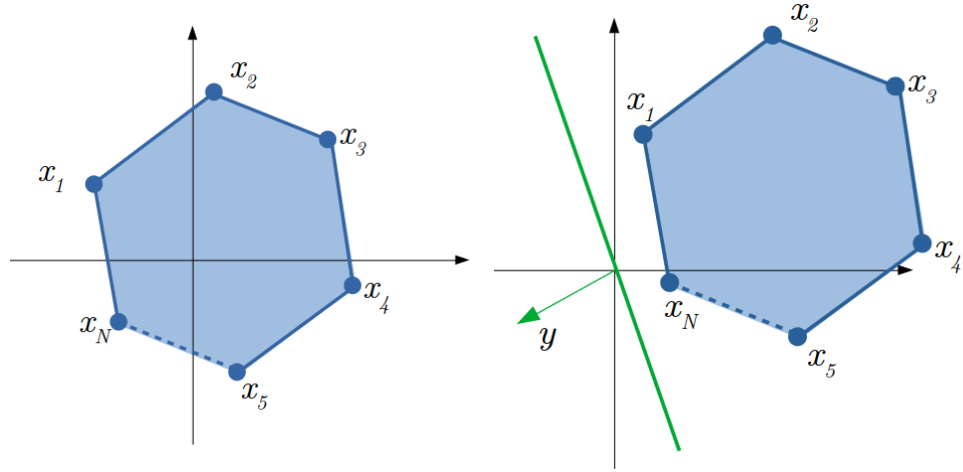


Figura 6.4: A la izquierda alternativa i^*) y a la derecha la alternativa ii^*).

Teorema 6.4 (Teorema de la Alternativa de Gordan-versión convexa). *Sean C un subconjunto convexo de un espacio vectorial, $N \in \mathbb{N}$ y $f_1, \dots, f_N : C \rightarrow \mathbb{R}$ funciones convexas. Entonces una, y solo una, de la siguientes afirmaciones se cumple:*

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

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

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

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

$$\inf_{c \in C} \sum_{i=1}^N t_i f_i(c) \leq \sum_{i=1}^N t_i f_i(c_0) < 0$$

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

\Rightarrow) Si aplicamos el lema de Simons, lema 6.4, a las funciones f_1, \dots, f_N obtenemos:

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

Entonces, de $\neg i)$ sabemos que, para este $\mathbf{t} \in \Delta_N$ (como para cualquier otro)

$$\inf_{c \in C} \sum_{i=1}^N t_i f_i(c) < 0$$

luego

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

como queríamos demostrar.

Para finalizar, si $\inf_{c \in C} [\max_{i=1,\dots,N} f_i(c)] = -\infty$ es claro que estamos en el caso $ii)$ y se prueba como en el caso de que dicho ínfimo sea un valor real. ■

Destacamos las siguientes observaciones:

Observación 6.1. *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 $f_1, \dots, f_N : C \rightarrow \mathbb{R}$ definidas por

$$f_i(\mathbf{c}) := \langle \mathbf{c}, \mathbf{x}_i \rangle, \forall i = 1, \dots, N.$$

Notar que las funciones f_1, \dots, f_N son lineales por la izquierda y como consecuencia son convexas. En este caso, la alternativa $ii)$ implica $ii^*)$ ya que:

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

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

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

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

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

28 6.2. Teorema de la alternativa de Gordan. Reformulaciones.

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

Observación 6.2. *El lema de Simons (lema 6.4) y el teorema convexo de la Alternativa de Gordan (teorema 6.4) 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 $\alpha := \inf_{c \in C} [\max_{i=1, \dots, N} \{f_i(c)\}]$. Si $\alpha = -\infty$. Por el lema 6.3 sabemos que $\forall \mathbf{t} \in \Delta_N$ se cumple que $\sum_{i=1}^N t_i f_i(c) \leq \max_{i=1, \dots, N} \{f_i(c)\}$ para todo $c \in C$. Tomando ínfimos en C :

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

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

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

Supongamos ahora que $\alpha \in \mathbb{R}$. Sean las funciones $g_1, \dots, g_N : C \longrightarrow \mathbb{R}$ definidas como $g_i = f_i - \alpha$ con $i = 1, \dots, N$. Veamos que las funciones g_1, \dots, g_N son convexas como consecuencia de que f_1, \dots, f_N lo son. Sean $i = 1, \dots, N$, $c_1, c_2 \in C$ 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 $i = 1, \dots, N$. 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 $0 \leq \inf_{c \in C} \sum_{i=1}^N t_i g_i(c)$.
- ii) $\exists c \in C$ que cumple $\max_{i=1, \dots, N} \{g_i(c)\} < 0$.

Razonemos que no se puede dar ii). Si fuese así, tendríamos que $\exists c \in C$ tal que $\max_{i=1, \dots, N} \{g_i(c)\} = \max_{i=1, \dots, N} \{f_i(c) - \alpha\} < 0$. En particular, se cumpliría

$$f_j(c) - \alpha < 0 \implies f_j(c) < \alpha = \inf_{c \in C} \left[\max_{i=1, \dots, N} \{f_i(c)\} \right], \quad \forall j \in 1, \dots, N.$$

Esto es imposible por la propia definición de α . Por ello, afirmamos que $\exists \mathbf{t} \in \Delta_N$ tal que $0 \leq \inf_C \sum_{i=1}^N t_i g_i$. Desarrollando el sumatorio:

$$\begin{aligned}
 0 &\leq \inf_{c \in C} \sum_{i=1}^N t_i g_i(c) \\
 &= \inf_{c \in C} \sum_{i=1}^N t_i (f_i(c) - \alpha) \\
 &= \inf_{c \in C} \left[\sum_{i=1}^N t_i f_i(c) - \sum_{i=1}^N t_i \alpha \right] \\
 &= \inf_{c \in C} \left[\sum_{i=1}^N t_i f_i(c) - \alpha \sum_{i=1}^N t_i \right] \\
 &= \inf_{c \in C} \left[\sum_{i=1}^N t_i f_i(c) - \alpha \right] \\
 &= \inf_{c \in C} \left[\sum_{i=1}^N t_i f_i(c) \right] - \alpha.
 \end{aligned}$$

Por lo tanto:

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

El lema 6.3 nos aporta la otra desigualdad y llegamos nuevamente a que $\exists \mathbf{t} \in \Delta_N$ que cumple:

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

■

Mencionemos antes de terminar que el teorema de la alternativa de Gordan no solo es equivalente al lema de Simons. Otro tipo de teorema de la alternativa equivalente es el lema de Farkas que estudiaremos en el capítulo sobre optimización. Son muchos los teoremas de la alternativa que se conocen, y siempre vinculados a la optimización: véase, por ejemplo, la amplia muestra que aparece recogida en [3]. A modo de ejemplo, y usando la notación clásica matricial, enunciaremos algunos de los resultados más conocidos:

Teorema 6.5 (Motzkin). *Dados $N, M \in \mathbb{N}$, tomamos las matrices $A, B, D \in \mathbb{R}^{M \times N}$. Entonces una, y solo una, de las siguientes condiciones se cumple:*

m1) $\exists \mathbf{x} \in \mathbb{R}^N$ tal que $A\mathbf{x} = 0$, $B\mathbf{x} \geq 0$ y $D\mathbf{x} > 0$.

m2) $\exists \mathbf{y}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^M$ tales que $\mathbf{y}^T A + \mathbf{v}^T B + \mathbf{w}^T D = 0$, $\mathbf{v} \geq 0$ y $\mathbf{w} \geq 0$.

Teorema 6.6 (Tucker). *Dados $N, M \in \mathbb{N}$, tomamos $A, B, D \in \mathbb{R}^{M \times N}$. Entonces una, y solo una, de las siguientes condiciones se cumple:*

m1) $\exists \mathbf{x} \in \mathbb{R}^N$ tal que $A\mathbf{x} = 0$, $B\mathbf{x} \geq 0$ y $D\mathbf{x} \geq 0$.

m2) $\exists \mathbf{y}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^M$ tales que $\mathbf{y}^T A + \mathbf{v}^T B + \mathbf{w}^T D = 0$, $\mathbf{v} \geq 0$ y $\mathbf{w} > 0$.

Teorema 6.7 (Primer teorema de la alternativa de Fenchel). *Dados $N, M \in \mathbb{N}$, tomamos $A, B \in \mathbb{R}^{M \times N}$. Entonces una, y solo una, de las siguientes condiciones se cumple:*

m1) $\exists \mathbf{x}, \mathbf{z} \in \mathbb{R}^N$ tales que $A\mathbf{x} + B\mathbf{z} = 0$, $\mathbf{x} \geq 0$ y $\mathbf{z} \geq 0$.

m2) $\exists \mathbf{y} \in \mathbb{R}^M$ tal que $\mathbf{y}^T A \geq 0$ e $\mathbf{y}^T B > 0$.

Alternativa y minimax

Este capítulo se centra en el uso de los teoremas de la alternativa en la teoría minimax, surgida a finales de la década de 1920 de la mano de J. von Neumann en el seno de la teoría de juegos. Sin embargo, nosotros aplicaremos un teorema minimax deducido del teorema de la alternativa de Gordan para obtener resultados de separación de convexos

7.1. Una desigualdad 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 teorema 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 y $f : X \times Y \rightarrow \mathbb{R}$. Obviamente, esta igualdad no es cierta en general tal y como mostramos en el siguiente ejemplo. Definimos $f : \{0, 1\} \times \{0, 1\} \rightarrow \mathbb{R}$ 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 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 7.1. Sea X un espacio topológico. Decimos que $f : X \rightarrow \mathbb{R}$ es superiormente semicontinua en X 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por

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

Para funciones superiormente semicontinuas, podemos enunciar los siguientes resultados. En particular, el lema 7.2 es análogo al del caso continuo.

Lema 7.1. El ínfimo de una familia de funciones reales $(f_\lambda)_{\lambda \in \Lambda}$ definidas en el mismo espacio topológico X , si es finito, determina una función superiormente semicontinua.

Este hecho es inmediato de la definición pues dado $r \in \mathbb{R}$,

$$\begin{aligned} \{x \in X : \inf_{\lambda \in \Lambda} f_\lambda(x) \geq r\} &= \{x \in X : \lambda \in \Lambda \implies f_\lambda(x) \geq r\} \\ &= \bigcap_{\lambda \in \Lambda} \{x \in X : f_\lambda(x) \geq r\}, \end{aligned}$$

que es cerrado por ser intersección de cerrados.

Lema 7.2. Si X es un espacio topológico compacto y $f : X \rightarrow \mathbb{R}$ es superiormente semicontinua, entonces f alcanza su supremo en X .

La demostración de este resultado está recogida en [1].

En estos momentos nos encontramos en condiciones de enunciar y demostrar el siguiente teorema minimax, más general que el de von Neumann. Los ingredientes de la prueba son las condiciones topológicas que se suponen y el teorema de la alternativa de Gordan, versión convexa, en forma equivalente del lema de Simons. Seguimos la demostración dada por [8].

Teorema 7.1. 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 $f : X \times Y \rightarrow \mathbb{R}$ es:

- i) cóncava y superiormente semicontinua en X y
- ii) convexa en Y .

Entonces:

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

Demostración. En primer lugar, por lemas 7.1 y 7.2 podemos escribir máximo en ambos casos en vez de supremo ya que f es superiormente semicontinua en X , por ello $\inf_{y \in Y} f(x, y)$ también lo es y X es compacto. Como hemos explicado anteriormente, solo necesitamos la desigualdad

$$\inf_{y \in Y} \max_{x \in X} f(x, y) \leq \max_{x \in X} \inf_{y \in Y} f(x, y). \quad (7.1)$$

Definimos $\alpha := \inf_{y \in Y} \max_{x \in X} f(x, y)$. 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 (7.1) es equivalente a

$$\exists x_0 \in X : \alpha \leq \inf_{y \in Y} f(x_0, y),$$

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

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

o lo que es igual,

$$\bigcap_{y \in Y} \{x \in X : \alpha \leq f(x, y)\} \neq \emptyset. \quad (7.2)$$

Como f es superiormente semicontinua en X estamos ante una intersección de cerrados. Usando la propiedad de intersección finita (X es compacto) obtenemos que (7.2) equivale a

$$\begin{aligned} N \in \mathbb{N}, y_1, \dots, y_N \in Y &\implies \bigcap_{i=1}^N \{x \in X : \alpha \leq f(x, y_i)\} \neq \emptyset. \\ &\iff \\ N \in \mathbb{N}, y_1, \dots, y_N \in Y &\implies \exists x_0 \in X : \alpha \leq \min_{i=1, \dots, N} f(x_0, y_i). \\ &\iff \\ N \in \mathbb{N}, y_1, \dots, y_N \in Y &\implies \alpha \leq \max_{x \in X} \min_{i=1, \dots, N} f(x, y_i). \end{aligned} \quad (7.3)$$

Esta última condición es cierta. En efecto, sean $N \in \mathbb{N}$ e $\{y_1, \dots, y_N\} \in Y$. Aplicamos el lema de Simons, lema 6.4, tomando $C := X$ y $f_i : X \rightarrow \mathbb{R}$ 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 en X con $i = 1, \dots, N$. De este modo, existe $\mathbf{t} \in \Delta_N$ tal que

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

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

$$\begin{aligned} \inf_{x \in X} \max_{i=1, \dots, N} \{-f(x, y_i)\} &= \inf_{x \in X} \left[\sum_{i=1}^N t_i (-f(x, y_i)) \right] \\ &\Downarrow \\ \max_{x \in X} \min_{i=1, \dots, N} \{f(x, y_i)\} &= \max_{x \in X} \left[\sum_{i=1}^N t_i f(x, y_i) \right]. \end{aligned}$$

Al ser f convexa en Y :

$$\begin{aligned} \max_{x \in X} \min_{i=1, \dots, N} \{f(x, y_i)\} &= \max_{x \in X} \left[\sum_{i=1}^N t_i f(x, y_i) \right] \\ &\geq \max_{x \in X} \left[f(x, \sum_{i=1}^N t_i y_i) \right] \\ &\geq \inf_{y \in Y} \max_{x \in X} f(x, y) \\ &= \alpha. \end{aligned}$$

Hemos probado entonces la desigualdad (7.3) y con ello el teorema. ■

7.2. Separación de convexos

En esta sección introducimos algunos resultados sobre separación a partir del teorema minimax, teorema 7.1. En general, estos nos aportan herramientas para poder concluir cuándo dos subconjuntos convexos pueden ser separados mediante un hiperplano. En la siguiente figura 7.1 vemos un ejemplo sobre la situación en la que nos encontramos.

Exponemos un resultado sencillo en el que solo involucramos un conjunto.

Teorema 7.2. *Dado $N \in \mathbb{N}$, sea $C \subset \mathbb{R}^N$ convexo y definimos*

$$\delta := \inf \{\|\mathbf{c}\| : \mathbf{c} \in C\}.$$

Entonces, existe $\mathbf{x}_0 \in \mathbb{R}^N$ tal que

$$\mathbf{c} \in C \implies \delta \leq \langle \mathbf{x}_0, \mathbf{c} \rangle.$$

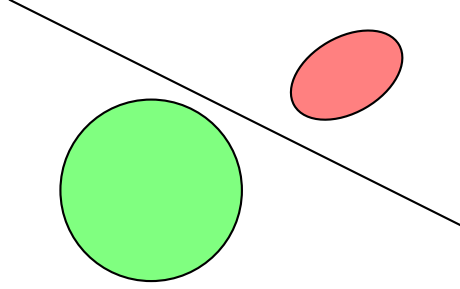


Figura 7.1: Situación de teoremas de separación

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

$$\exists \mathbf{x}_0 \in \mathbb{R}^N : \mathbf{c} \in C \implies \delta \leq \langle \mathbf{x}_0, \mathbf{c} \rangle.$$

$$\Updownarrow$$

$$\exists \alpha > 0, \exists \mathbf{x}_0 \in \alpha B_{\mathbb{R}^N} : \mathbf{c} \in C \implies \delta \leq \langle \mathbf{x}_0, \mathbf{c} \rangle.$$

$$\Updownarrow$$

$$\exists \alpha > 0, \exists \mathbf{x}_0 \in \alpha B_{\mathbb{R}^N} : \delta \leq \inf_{\mathbf{c} \in C} \langle \mathbf{x}_0, \mathbf{c} \rangle.$$

$$\Updownarrow$$

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

Llamamos $X := \alpha B_{\mathbb{R}^N}$ que es compacto y convexo, $Y := C$ convexo y f a la función real y continua con valores en $X \times Y$ definida como

$$f(\mathbf{x}, \mathbf{y}) := \langle \mathbf{x}, \mathbf{y} \rangle$$

(al ser f continua, en particular es superiormente semicontinua). Aplicamos el teorema minimax, teorema 7.1, y obtenemos que probar la última desigualdad es equivalente a probar que

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

Tenemos que $\max_{\mathbf{x} \in \alpha B_{\mathbb{R}^N}} \langle \mathbf{x}, \mathbf{c} \rangle = \alpha \|\mathbf{c}\|$ por el corolario 6.1. Así, debemos demostrar que

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

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

Ahora, vamos a hacer una generalización de este resultado.

Teorema 7.3. Sean $N \in \mathbb{N}$ y A, B subconjuntos convexos de \mathbb{R}^N tal que A es cerrado, B es compacto y $A \cap B = \emptyset$. Entonces existe $\mathbf{x}_0 \in \mathbb{R}^N$ tal que

$$\sup_{\mathbf{a} \in A} \langle \mathbf{x}_0, \mathbf{a} \rangle < \inf_{\mathbf{b} \in B} \langle \mathbf{x}_0, \mathbf{b} \rangle.$$

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

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

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

$$\|\mathbf{a}_{\sigma(n)} - \mathbf{b}\| = \|\mathbf{a}_{\sigma(n)} - \mathbf{b}_{\sigma(n)} + \mathbf{b}_{\sigma(n)} - \mathbf{b}\| \leq \|\mathbf{a}_{\sigma(n)} - \mathbf{b}_{\sigma(n)}\| + \|\mathbf{b}_{\sigma(n)} - \mathbf{b}\|.$$

Entonces tenemos que $\|\mathbf{a}_{\sigma(n)} - \mathbf{b}\| \rightarrow 0$ ya que $\{\mathbf{b}_{\sigma(n)}\}_{n \in \mathbb{N}} \rightarrow \mathbf{b}$ y como se da $\|\mathbf{a}_n - \mathbf{b}_n\| \rightarrow 0$ también se cumple que $\|\mathbf{a}_{\sigma(n)} - \mathbf{b}_{\sigma(n)}\| \rightarrow 0$. Llegamos a que $\{\mathbf{a}_{\sigma(n)}\}_{n \in \mathbb{N}} \rightarrow \mathbf{b}$. Al ser A cerrado se debe cumplir que $\mathbf{b} \in A$ lo cual es imposible ya que $A \cap B = \emptyset$.

Ahora, aplicamos el teorema 7.2 a $C := B - A$. Notar que C es convexo por serlo A y B . Obtenemos entonces que existe $\mathbf{x}_0 \in \mathbb{R}^N$ tal que si $\mathbf{c} \in C$ se cumple que:

$$\delta = \inf_{\mathbf{c} \in C} \|\mathbf{c}\| \leq \langle \mathbf{x}_0, \mathbf{c} \rangle.$$

Por la definición de C , se tiene que:

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

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

$$\exists \mathbf{x}_0 \in \mathbb{R}^N : \mathbf{a} \in A, \mathbf{b} \in B \implies 0 < \delta \leq \langle \mathbf{x}_0, \mathbf{b} - \mathbf{a} \rangle = \langle \mathbf{x}_0, \mathbf{b} \rangle - \langle \mathbf{x}_0, \mathbf{a} \rangle.$$

$$\Updownarrow$$

$$\exists \mathbf{x}_0 \in \mathbb{R}^N : \mathbf{a} \in A, \mathbf{b} \in B \implies \langle \mathbf{x}_0, \mathbf{a} \rangle + \delta \leq \langle \mathbf{x}_0, \mathbf{b} \rangle, \text{ con } \delta > 0.$$

$$\Updownarrow$$

$$\exists \mathbf{x}_0 \in \mathbb{R}^N \implies \sup_{\mathbf{a} \in A} \langle \mathbf{x}_0, \mathbf{a} \rangle + \delta \leq \inf_{\mathbf{b} \in B} \langle \mathbf{x}_0, \mathbf{a} \rangle, \text{ con } \delta > 0.$$

$$\Updownarrow$$

$$\exists \mathbf{x}_0 \in \mathbb{R}^N : \sup_{\mathbf{a} \in A} \langle \mathbf{x}_0, \mathbf{a} \rangle < \inf_{\mathbf{b} \in B} \langle \mathbf{x}_0, \mathbf{a} \rangle.$$

Por ello, queda probado el teorema. ■

A partir de este resultado, llegamos al siguiente corolario que será necesario en la parte final del trabajo.

Corolario 7.1. Sean $N \in \mathbb{N}$, $L \leq \mathbb{R}^N$ subespacio vectorial y $K \subset \mathbb{R}^N$ un subconjunto compacto y convexo tales que $L \cap K = \emptyset$. Entonces, $\exists \mathbf{x}_0 \in \mathbb{R}^N$ tal que si $\mathbf{y} \in L$, entonces $\langle \mathbf{x}_0, \mathbf{y} \rangle = 0$ y

$$0 < \inf_{\mathbf{x} \in K} \langle \mathbf{x}_0, \mathbf{x} \rangle.$$

Demostración. Como $L \leq \mathbb{R}^N$ entonces L es convexo y además, como \mathbb{R}^N es finito dimensional, también es cerrado. Esto último se debe a que L es de la forma

$$L = \left\{ (x_1, \dots, x_N) \in \mathbb{R}^N : \begin{array}{l} a_{11}x_1 + \dots + a_{1N}x_N = 0 \\ \vdots \\ a_{N1}x_1 + \dots + a_{NN}x_N = 0 \end{array} \right\}$$

Para cada una de las ecuaciones anteriores definimos las funciones continuas para cada $i = 1, \dots, N$ como

$$\begin{aligned} f_i : \mathbb{R}^N &\longrightarrow \mathbb{R} \\ \mathbf{x} &\longmapsto f(\mathbf{x}) = a_{i1}x_1 + \dots + a_{iN}x_N. \end{aligned}$$

De este modo, el conjunto del puntos que cumple cada ecuación se puede escribir como $f_i^{-1}(\{0\})$, que es cerrado al ser la imagen inversa de un cerrado por una función continua. Por lo tanto, L es la intersección finita de cerrados y por ello cerrado. Estamos entonces bajo las condiciones del teorema 7.3, que nos aporta la existencia de un $\mathbf{x}_0 \in \mathbb{R}^N$ que cumple:

$$\sup_{\mathbf{y} \in L} \langle \mathbf{x}_0, \mathbf{y} \rangle < \inf_{\mathbf{x} \in K} \langle \mathbf{x}_0, \mathbf{x} \rangle. \quad (7.4)$$

Tomamos ahora $\mathbf{y} \in L$ con $\langle \mathbf{x}_0, \mathbf{y} \rangle \neq 0$ (si $\langle \mathbf{x}_0, \mathbf{y} \rangle = 0$ no hay nada que probar) y $\rho > 0$. Así,

$$\rho \langle \mathbf{x}_0, \mathbf{y} \rangle = \langle \mathbf{x}_0, \rho \mathbf{y} \rangle < \inf_{\mathbf{x} \in K} \langle \mathbf{x}_0, \mathbf{x} \rangle.$$

Como $\inf_{\mathbf{x} \in K} \langle \mathbf{x}_0, \mathbf{x} \rangle \in \mathbb{R}$, la arbitrariedad de $\rho > 0$ obliga a que $\langle \mathbf{x}_0, \mathbf{y} \rangle \leq 0$ ya que si no fuese así \mathbb{R} , estaría acotado. Como $L \leq \mathbb{R}^N$, cambiamos \mathbf{y} por $-\mathbf{y}$ y obtenemos que $-\langle \mathbf{x}_0, \mathbf{y} \rangle \leq 0$, por lo que $\langle \mathbf{x}_0, \mathbf{y} \rangle = 0$. Finalmente, el hecho de que

$$0 < \inf_{\mathbf{x} \in K} \langle \mathbf{x}_0, \mathbf{x} \rangle$$

se deduce de (7.4). ■

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. Esta vez, resulta del teorema de la alternativa de Gordan.

Teorema 7.4. Sea $N \in \mathbb{N}$ y A y B dos subconjuntos convexos y disjuntos de \mathbb{R}^N . Entonces existe $\mathbf{x}_0 \in \mathbb{R}^N \setminus \{0\}$ tal que

$$\sup_{\mathbf{a} \in A} \langle \mathbf{x}_0, \mathbf{a} \rangle \leq \inf_{\mathbf{b} \in B} \langle \mathbf{x}_0, \mathbf{b} \rangle.$$

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:

$$\exists \mathbf{x}_0 \in \mathbb{R}^N \setminus \{0\} : \sup_{\mathbf{c} \in C} \langle \mathbf{x}_0, \mathbf{c} \rangle \leq 0,$$

equivalentemente,

$$\exists \mathbf{x}_0 \in S_{\mathbb{R}^N} \setminus \{0\} : \sup_{\mathbf{c} \in C} \langle \mathbf{x}_0, \mathbf{c} \rangle \leq 0.$$

Pero esta afirmación no es más que

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

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

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

$$\Updownarrow$$

$$M \in \mathbb{N}, \{\mathbf{c}_1, \dots, \mathbf{c}_M\} \subset C \implies \exists \mathbf{x}_0 \in S_{\mathbb{R}^N} : \max_{i=1, \dots, M} \langle \mathbf{x}_0, \mathbf{c}_i \rangle \leq 0.$$

Esta condición se verifica, ya que si $M \geq 1$ y $\{\mathbf{c}_1, \dots, \mathbf{c}_M\} \subset C$, entonces, por ser C convexo se cumple que $\text{co}\{\mathbf{c}_1, \dots, \mathbf{c}_M\} \subset C$ y como $0 \notin C$, entonces tenemos que $0 \notin \text{co}\{\mathbf{c}_1, \dots, \mathbf{c}_M\}$. Por tanto, la versión clásica del Teorema de Gordan, teorema 6.3, garantiza

$$\exists \mathbf{z}_0 \in \mathbb{R}^N : \max_{i=1, \dots, M} \langle \mathbf{z}_0, \mathbf{c}_i \rangle < 0.$$

En particular, $\mathbf{z}_0 \neq 0$ y tomando $\mathbf{x}_0 = \frac{\mathbf{z}_0}{\|\mathbf{z}_0\|}$ queda probado el enunciado. ■

Destacamos ahora lo siguiente:

Observación 7.1. El Teorema de separación 7.3 implica el Teorema de Gordan, teorema 6.3.

En efecto, dados $\{\mathbf{x}_1, \dots, \mathbf{x}_M\} \subset \mathbb{R}^N$ con $M, N \in \mathbb{N}$ del enunciado de la alternativa de Gordan planteamos las siguientes alternativas excluyentes:

1. Si $0 \in \text{co}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ entonces es claro que se cumple la alternativa i^*).

2. Si $0 \notin \text{co}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ entonces $\{0\} \cap \text{co}\{\mathbf{x}_1, \dots, \mathbf{x}_M\} = \emptyset$. Al ser ambos conjuntos son compactos usamos el teorema de separación y obtenemos que:

$$\exists \mathbf{y} \in \mathbb{R}^M : \sup_{\mathbf{x} \in \text{co}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}} \langle \mathbf{y}, \mathbf{x} \rangle < 0,$$

luego, en particular $(\{\mathbf{x}_1, \dots, \mathbf{x}_M\} \subset \text{co}\{\mathbf{x}_1, \dots, \mathbf{x}_M\})$,

$$\exists \mathbf{y} \in \mathbb{R}^M : \max_{i=1, \dots, M} \langle \mathbf{y}, \mathbf{x}_i \rangle < 0.$$

■

Para finalizar este capítulo, recogemos las equivalencias que se han obtenido sobre diferentes resultados:

Teorema de Gordan clásico (teorema 6.3)

\Updownarrow

Teorema de Gordan convexo (teorema 6.4)

\Updownarrow

Teorema de separación de convexos en \mathbb{R}^N (teorema 7.3).

Optimización

Si en el capítulo anterior dábamos las primeras aplicaciones del teorema de Gordan en forma de teorema minimax y, consecuentemente, de teoremas de separación, ahora usamos los teoremas de la alternativa en su contexto original: a optimización. Tras presentar un nuevo resultado de la alternativa, el lema de Farkas, abordamos el estudio de dos problemas clásicos: uno de dualidad en programación lineal; otro de paso de un problema de optimización con restricciones (en este caso de tipo desigualdad) a otro sin restricciones, vía los teoremas de Fritz-John y Karush-Kuhn-Tucker. Para el primero de ellos usamos el de Farkas y para el segundo el teorema de Gordan en su versión clásica.

8.1. Lema de Farkas

En el capítulo anterior hemos visto 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 8.1. *Dados $M, N \in \mathbb{N}$ y $\{\mathbf{x}_1, \dots, \mathbf{x}_M\} \subset \mathbb{R}^N$, llamamos cono generado por $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ al conjunto de \mathbb{R}^N dado por:*

$$\text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\} := \left\{ \sum_{j=1}^N \mu_j \mathbf{x}_j : \mu_1, \dots, \mu_N \geq 0 \right\}.$$

Veamos que $\text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ es un subconjunto convexo y cerrado de \mathbb{R}^N :

- Convexo: tenemos que comprobar que dados $\mathbf{t}, \mathbf{s} \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ y $\lambda \in [0, 1]$ entonces $\lambda \mathbf{t} + (1 - \lambda) \mathbf{s} \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. En efecto:

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

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 $j = 1, \dots, M$. Así, podemos concluir que $\lambda\mathbf{t} + (1-\lambda)\mathbf{s} \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ y por tanto es un subconjunto convexo.

- Cerrado: sea $\{\mathbf{t}_n\}_{n \in \mathbb{N}}$ una sucesión de $\text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ y sea $\mathbf{t}_0 \in \mathbb{R}^N$ tal que $\{\mathbf{t}_n\}_{n \in \mathbb{N}} \rightarrow \mathbf{t}_0$. Si notamos $\mathbf{t}_n = \sum_{j=1}^N \mu_j^n \mathbf{x}_j$, entonces:

$$\{\mathbf{t}_n\}_{n \in \mathbb{N}} = \left\{ \sum_{j=1}^N \mu_j^n \mathbf{x}_j \right\}_{n \in \mathbb{N}} \rightarrow \sum_{j=1}^N \mu_j^0 \mathbf{x}_j = \mathbf{t}_0.$$

Como para cada \mathbf{t}_n cumple que $\mu_j^n \geq 0$ con $j = 1, \dots, N$ podemos asegurar que $\mu_j^0 \geq 0$ con $j = 1, \dots, N$. Hemos demostrado que \mathbf{t}_0 se expresa como combinación de $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ con coeficientes no negativos. Por lo tanto, $\mathbf{t}_0 \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ y por consiguiente es cerrado.

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

Lema 8.1 (Farkas). Sean $M, N \in \mathbb{N}$ y $\{\mathbf{x}_1, \dots, \mathbf{x}_M\} \subset \mathbb{R}^N$ y $\mathbf{b} \in \mathbb{R}^N$. Entonces una, y solo una, de la siguientes alternativas se cumple:

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

ii') $\exists \mathbf{z}_0 \in \mathbb{R}^N$ que cumple que:

1. $\max_{j=1, \dots, M} \langle \mathbf{z}_0, \mathbf{x}_j \rangle \leq 0$ y
2. $\langle \mathbf{z}_0, \mathbf{b} \rangle > 0$.

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

a) $\mathbf{b} \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. Estamos en el caso i') ya que:

$$\mathbf{b} \in \left\{ \sum_{j=1}^N \mu_j \mathbf{x}_j : \mu_1, \dots, \mu_N \geq 0 \right\}.$$

b) $\mathbf{b} \notin \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. Por su parte, esta alternativa equivale a ii'). En efecto:

ii') \implies b) Razonamos por reducción al absurdo. Suponemos por ello que el elemento $\mathbf{b} \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$, entonces, podemos expresar $\mathbf{b} = \sum_{j=1}^N \mu_j \mathbf{x}_j$ donde $\mu_1, \dots, \mu_N \geq 0$. Como se da ii'), en particular se cumple 1 y obtenemos, para el \mathbf{z}_0 cuya existencia se garantiza,

$$\begin{aligned} \langle \mathbf{z}_0, \mathbf{b} \rangle &= \langle \mathbf{z}_0, \sum_{j=1}^N \mu_j \mathbf{x}_j \rangle \\ &= \sum_{j=1}^N \mu_j \langle \mathbf{z}_0, \mathbf{x}_j \rangle \leq 0. \end{aligned}$$

Por otro lado, por 2 se tiene que $\langle \mathbf{z}_0, \mathbf{b} \rangle > 0$. Así, obtenemos que

$$\langle \mathbf{z}_0, \mathbf{b} \rangle \leq 0 < \langle \mathbf{z}_0, \mathbf{b} \rangle,$$

lo cual es imposible.

b) $\Rightarrow ii')$ Si $\mathbf{b} \notin \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ aplicamos el teorema de separación 7.3 a los conjuntos $\{\mathbf{b}\}$ que es compacto y convexo y a $\text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ que es cerrado y convexo. Obtenemos que:

$$\exists \mathbf{z}_0 \in \mathbb{R}^N : \sup_{a \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}} \langle \mathbf{z}_0, a \rangle < \langle \mathbf{z}_0, \mathbf{b} \rangle.$$

Por un lado, es obvio que $0 \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ y por ello:

$$0 = \langle \mathbf{z}_0, 0 \rangle \leq \sup_{a \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}} \langle \mathbf{z}_0, a \rangle < \langle \mathbf{z}_0, \mathbf{b} \rangle.$$

Hemos obtenido por tanto que es cierto 2. Para probar 1, fijamos $a_0 \in \text{co}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. Entonces, dado $\rho > 0$,

$$\rho \langle \mathbf{z}_0, a_0 \rangle = \langle \mathbf{z}_0, \rho a_0 \rangle \leq \sup_{a \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}} \langle \mathbf{z}_0, a_0 \rangle.$$

Llegamos a que el conjunto $\{\rho \langle \mathbf{z}_0, a_0 \rangle : \rho > 0\}$ está acotado y eso solo es posible si $\langle \mathbf{z}_0, a_0 \rangle \leq 0$. La arbitrariedad de a_0 nos aporta que

$$\sup_{a \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}} \langle \mathbf{z}_0, a \rangle = 0$$

de donde

$$\max_{j=1, \dots, M} \langle \mathbf{z}_0, \mathbf{x}_j \rangle \leq \sup_{a \in \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}} \langle \mathbf{z}_0, a \rangle \leq 0$$

ya que obviamente $\{\mathbf{x}_1, \dots, \mathbf{x}_M\} \subset \text{cone}\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. En particular, hemos probado 1. ■

8.2. Dualidad en programación lineal

Gracias al lema de Farkas podemos probar resultados como el teorema de dualidad en programación lineal que enunciamos a continuación. La prueba se recoge en [1].

Teorema 8.1. (*Teorema de dualidad en programación lineal*) Dados $N, M \in \mathbb{R}$ tomamos la matriz $A \in \mathbb{R}^{N \times M}$ y los vectores $\mathbf{b} \in \mathbb{R}^M$ y $\mathbf{c} \in \mathbb{R}^N$. Consideramos el problema de optimización primal:

$$\left. \begin{array}{l} p := \inf_{\mathbf{x}} \mathbf{c}^T \mathbf{x} \\ \text{s.a. } A^T \mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \in \mathbb{R}^N \end{array} \right\}, \quad (P)$$

donde $A^T \mathbf{x} \leq \mathbf{b}$ nos aporta un total de M restricciones. Sea el problema dual:

$$\left. \begin{array}{l} d := \sup_{\mathbf{y}} -\mathbf{b}^T \mathbf{y} \\ \text{s.a } A\mathbf{y} = -\mathbf{c} \\ \mathbf{y} \in \mathbb{R}_+^M \end{array} \right\}. \quad (D)$$

Entonces, la solución del problema primal y el dual coincide, o lo que es lo mismo, $p = d$.

Demostración. En primer lugar, veamos que $p \geq d$. Para ello, tomamos $\mathbf{x} \in \mathbb{R}^N$ e $\mathbf{y} \in \mathbb{R}^M$ factibles, es decir, cumplen las restricciones de sus respectivos problemas. Así:

$$\begin{aligned} & \left\{ \begin{array}{l} A^T \mathbf{x} - \mathbf{b} \leq 0 \\ \mathbf{y} \geq 0 \end{array} \right\} \\ & \Downarrow \\ & (A^T \mathbf{x} - \mathbf{b})^T \mathbf{y} = \mathbf{x}^T A\mathbf{y} - \mathbf{b}^T \mathbf{y} \leq 0 \\ & \Updownarrow \\ & -\mathbf{b}^T \mathbf{y} \leq \mathbf{c}^T \mathbf{x}, \end{aligned}$$

donde la última desigualdad se debe a que $A\mathbf{y} = -\mathbf{c}$. Tomando supremos en la izquierda e ínfimos en la derecha llegamos a $d \leq p$.

Supongamos ahora que $p \in \mathbb{R}$ ya que en caso contrario habríamos terminado. Consideramos el problema homogeneizado,

$$\left. \begin{array}{l} A^T \mathbf{x} - z\mathbf{b} \leq 0 \\ -\mathbf{c}^T \mathbf{x} + zp > 0 \\ \mathbf{x} \in \mathbb{R}^N, z \geq 0 \end{array} \right\}.$$

Este problema no tiene solución ya que en caso contrario, si $z > 0$, entonces $A^T \frac{\mathbf{x}}{z} < \mathbf{b}$ y, por la definición de p , $\mathbf{c}^T \frac{\mathbf{x}}{z} \geq p$ si, y solo si, $zp \leq \mathbf{c}^T \mathbf{x}$, lo que no es posible, pues debe ser $zp > \mathbf{c}^T \mathbf{x}$. En el caso $z = 0$, el problema queda

$$\left. \begin{array}{l} A^T \mathbf{x} \leq 0 \\ \mathbf{c}^T \mathbf{x} < 0 \end{array} \right\}. \quad (8.1)$$

Sea $\mathbf{u} \in \mathbb{R}^N$ un punto factible del problema primal, esto es,

$$A^T \mathbf{u} \leq \mathbf{b}.$$

Entonces, dado $\rho > 0$,

$$\begin{aligned} A^T(\rho\mathbf{x} + \mathbf{u}) &= \rho A^T \mathbf{x} + A^T \mathbf{u} \\ &\leq \mathbf{b} \end{aligned}$$

ya que $A^T \mathbf{x} \leq 0$, $\rho > 0$ y $A^T \mathbf{u} \leq \mathbf{b}$. Usando la definición de p ,

$$\begin{aligned} p &\leq \mathbf{c}^T(\rho\mathbf{x} + \mathbf{u}) \\ &= \rho \mathbf{c}^T \mathbf{x} + \rho \mathbf{c}^T \mathbf{u}. \end{aligned}$$

Como $p \in \mathbb{R}$, $\mathbf{c}^T \mathbf{x} < 0$ y $\rho > 0$ es arbitrario, llegamos a una contradicción (tómese límite cuando $\rho \rightarrow +\infty$). Aplicamos el lema de Farkas al problema homogeneizado, y al no tener el sistema solución observamos que no se puede dar la alternativa *ii'*). Entonces, tenemos que existen escalares $\mu_1, \dots, \mu_M, \beta \geq 0$ que cumplen

$$\sum_{j=1}^M \mu_j (A_j, -b_j) + \beta(0, -1) = (-\mathbf{c}, p),$$

donde para cada $j = 1, \dots, M$, A_j denota la fila j -ésima de A^T . Reescribiendo lo

anterior, llamamos $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_M \end{pmatrix}$ y obtenemos

$$\left. \begin{aligned} \sum_{j=1}^M \mu_j A_j &= A\boldsymbol{\mu} = -\mathbf{c} \\ \boldsymbol{\mu} &\in \mathbb{R}_+^M \\ -\mathbf{b}^T \boldsymbol{\mu} - \beta &= p \end{aligned} \right\}.$$

El vector $\boldsymbol{\mu}$ es una solución factible para el problema dual (D) con valor objetivo al menos p obteniendo entonces que $p \leq d$, quedando así probado el resultado. ■

8.3. Optimización con restricciones: Fritz John y Karush-Kuhn-Tucker

Concluimos este capítulo con un apartado dedicado a la optimización con restricciones. A pesar de tratarse de resultados no lineales, el teorema de Gordan clásico basta para establecer los resultados fundamentales. Al igual que en el apartado anterior, seguimos el texto [1]. En primer lugar, empezamos recordando la definición de derivada direccional.

Definición 8.2. Sea $D \subset \mathbb{R}^M$ con $M \in \mathbb{N}$ y sea la función $g : D \rightarrow \mathbb{R}$, definimos la derivada direccional de g en $\mathbf{x} \in D$ en la dirección del vector $\mathbf{d} \in \mathbb{R}^M$ como

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

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

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. De hecho, si una función g es diferenciable en $\mathbf{x}_0 \in D$ en el sentido de Fréchet, y notamos su derivada como $Dg(\mathbf{x}_0)$ entonces g también es diferenciable en el sentido de Gâteaux en \mathbf{x}_0 y además

$Dg(\mathbf{x}_0) = \nabla g(\mathbf{x}_0)$. El recíproco no es cierto tal y como mostramos en el siguiente ejemplo. Sea $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definida como:

$$f(x, y) = \begin{cases} \frac{xy^3}{x^2 + y^2}, & \text{si } (x, y) \neq (0, 0) \\ 0, & \text{si } (x, y) = (0, 0) \end{cases}.$$

Sabemos que si una función es Fréchet diferenciable en \mathbf{x}_0 , entonces es continua en \mathbf{x}_0 . Como f no es continua en $(0, 0)$, podemos afirmar que no es Fréchet diferenciable en dicho punto. Sin embargo, sí es Gâteaux diferenciable en $(0, 0)$. Para $\mathbf{d} = (d_1, d_2) \in \mathbb{R}^2$,

$$\begin{aligned} f'((0, 0); (d_1, d_2)) &= \lim_{t \rightarrow 0} \frac{f((0, 0) + t(d_1, d_2)) - f(0, 0)}{t} \\ &= \lim_{t \rightarrow 0} \frac{\frac{td_1(td_2)^3}{(td_1)^2 + (td_2)^2} - 0}{t} \\ &= \lim_{t \rightarrow 0} \frac{t^4 d_1 d_2^3}{t^3 d_1^2 + t^3 d_2^2} \\ &= \lim_{t \rightarrow 0} \frac{td_1 d_2^3}{d_1^2 + d_2^2} \\ &= 0. \end{aligned}$$

Como el límite existe y es lineal, la función es diferenciable en el sentido Gâteaux en $(0, 0)$.

A continuación, demosetremos 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 8.1. Sean $D \subset \mathbb{R}^M$ ($M \in \mathbb{N}$), $\bar{\mathbf{x}}$ un punto del interior de D y sean $g_1, \dots, g_N : D \rightarrow \mathbb{R}$ funciones continuas y diferenciables en $\bar{\mathbf{x}}$ donde $N \in \mathbb{N}$. Definimos $g : D \rightarrow \mathbb{R}$ como

$$g(\mathbf{x}) := \max_{i=1, \dots, N} \{g_i(\mathbf{x})\}$$

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

$$g'(\bar{\mathbf{x}}; \mathbf{d}) = \max_{i \in K} \langle \nabla g_i(\bar{\mathbf{x}}), \mathbf{d} \rangle, \quad \mathbf{d} \in \mathbb{R}^M. \quad (8.2)$$

Demostración. Podemos suponer sin pérdida de generalidad que $K = \{1, \dots, N\}$, por facilitar la notación. Para cada $i \in K$ tenemos la siguiente desigualdad:

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

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

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

Por otro lado, afirmamos que

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

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

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

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

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

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

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

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:

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

■

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

$$\begin{cases} \inf_{\mathbf{x} \in D} f(\mathbf{x}) \\ \text{s.a. } g_1(\mathbf{x}) \leq 0 \\ \quad \vdots \\ g_N(\mathbf{x}) \leq 0 \end{cases} \quad (8.3)$$

donde $D \subset \mathbb{R}^M$, f es la función objetivo y las restricciones g_i con $i = 1, \dots, N$ 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 $\bar{\mathbf{x}}$ definimos el *conjunto activo* como $I(\bar{\mathbf{x}}) = \{i : g_i(\bar{\mathbf{x}}) = 0\}$. El papel que juega este conjunto de índices no es trivial, apareciendo en los teoremas de Fritz John y Karush-Kuhn-Tucker como los únicos necesarios. Este trabajo ya se ha realizado en la proposición 8.1. Para este problema y asumiendo que $\bar{\mathbf{x}} \in D$, llamamos *vector de multiplicadores de Lagrange para $\bar{\mathbf{x}}$* a $\boldsymbol{\lambda} \in (\mathbb{R}^N)^+$ si $\bar{\mathbf{x}}$ es un punto crítico de:

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

es decir, se cumple que f, g_1, \dots, g_N son diferenciables y:

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

Además, $\lambda_i = 0$ si $i \notin I(\bar{\mathbf{x}})$, por tanto

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

También, se dice que un punto $\bar{\mathbf{x}} \in D$ es un *mínimo global* del problema 8.3 si es un punto factible y además $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para todo $\mathbf{x} \in D$ factible. Por otro lado, diremos que $\bar{\mathbf{x}} \in D$ es un *mínimo local* del problema 8.3 si es un punto factible y existe U entorno de $\bar{\mathbf{x}}$ tal que $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para todo $\mathbf{x} \in D \cap U$ factible. Los resultados que exponemos a continuación se centran en mínimos locales.

Teorema 8.2 (Teorema de Fritz John). *Supongamos que el problema (8.3) tiene un mínimo local en $\bar{\mathbf{x}} \in D$. Si las funciones f, g_i con $i \in I(\bar{\mathbf{x}})$ son diferenciables en $\bar{\mathbf{x}}$ entonces existen $\lambda_0, \lambda_i \in \mathbb{R}^+$ para $i \in I(\bar{\mathbf{x}})$, no todas cero, que satisfacen:*

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

Demostración. Consideramos la función

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

definida para todo punto $\mathbf{x} \in D$ y perteneciente a la región factible del problema (8.3). Como $\bar{\mathbf{x}}$ es un mínimo local de dicho problema también lo es de g . Supongamos que no fuera así, es decir, existe U entorno de $\bar{\mathbf{x}}$ existe un punto $\mathbf{x}_0 \in D \cap U$ factible tal que $g(\mathbf{x}_0) < g(\bar{\mathbf{x}})$. Como $\bar{\mathbf{x}}$ es un mínimo local del problema se tiene que $f(\mathbf{x}_0) - f(\bar{\mathbf{x}}) > 0$. En ese caso, como \mathbf{x} es factible (y por ello cumple las restricciones), se tiene que

$$\begin{aligned} g(\mathbf{x}_0) &= \max\{f(\mathbf{x}_0) - f(\bar{\mathbf{x}}), g_i(\mathbf{x}_0) : i \in I(\bar{\mathbf{x}})\} \\ &= f(\mathbf{x}_0) - f(\bar{\mathbf{x}}) \\ &> 0. \end{aligned}$$

Esto es imposible ya que se tendría que $0 < g(\mathbf{x}_0) < g(\mathbf{x}) = 0$. Por la proposición 8.1 tenemos que para toda dirección $\mathbf{d} \in \mathbb{R}^M$ se cumple:

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

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

Por lo tanto, el sistema

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

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 6.3, vemos que solo se puede dar la alternativa i*) y en ese caso obtenemos que:

$$\exists \mathbf{t} = (t_0, \dots, t_R) \in \Delta_{R+1} \text{ tal que } 0 = t_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I(\bar{\mathbf{x}})} t_i \nabla g_i(\bar{\mathbf{x}}),$$

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

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 y no influye en la información que se da. 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 $\mathbf{d}_0 \in \mathbb{R}^M$ que satisface que $\langle \nabla g_i(\bar{\mathbf{x}}), \mathbf{d}_0 \rangle < 0$ para todo índice $i \in I(\bar{\mathbf{x}})$. Enunciamos ahora otro teorema clásico que soluciona el problema que acabamos de comentar.

Teorema 8.3 (Karush-Kuhn-Tucker). *Supongamos que el problema (8.3) tiene un mínimo local en $\bar{\mathbf{x}} \in D$. Si las funciones f, g_i con $i \in I(\bar{\mathbf{x}})$ son diferenciables en $\bar{\mathbf{x}}$ y se cumple la condición de Mangasarian-Fromovitz entonces existe un vector de multiplicadores de Lagrange para $\bar{\mathbf{x}}$.*

Demostración. Del teorema 8.2 de Fritz John obtenemos que:

$$\exists \mathbf{t} = (t_0, \dots, t_R) \in \Delta_{R+1} \text{ tal que } 0 = t_0 \nabla f(\bar{\mathbf{x}}) + \sum_{i \in I(\bar{\mathbf{x}})} t_i \nabla g_i(\bar{\mathbf{x}})$$

con R el cardinal de $I(\bar{\mathbf{x}})$. Si multiplicamos escalarmente la igualdad por \mathbf{d}_0 (dirección del espacio vectorial que nos aporta la condición de Mangasarian-Fromovitz), obtenemos:

$$0 = t_0 \langle \nabla f(\bar{\mathbf{x}}), \mathbf{d}_0 \rangle + \sum_{i \in I(\bar{\mathbf{x}})} t_i \langle \nabla g_i(\bar{\mathbf{x}}), \mathbf{d}_0 \rangle.$$

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

$$0 = \sum_{i \in I(\bar{\mathbf{x}})} t_i \langle \nabla g_i(\bar{\mathbf{x}}), \mathbf{d}_0 \rangle.$$

Al tener $\mathbf{t} \in \Delta_{R+1}$ se cumple que todas sus componentes son no negativas y $\sum_{i \in I(\bar{\mathbf{x}})} t_i = 1$ (estamos suponiendo que $t_0 = 0$ por lo que no influye en la suma de la definición de Δ_{R+1}) por lo que algún término es distinto de 0. Tenemos garantizado que $\langle g_i(\bar{\mathbf{x}}), \mathbf{d}_0 \rangle < 0 \quad \forall i \in I(\bar{\mathbf{x}})$. Así tendríamos que:

$$0 = \sum_{i \in I(\bar{\mathbf{x}})} t_i \langle \nabla g_i(\bar{\mathbf{x}}), \mathbf{d}_0 \rangle < 0,$$

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

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

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

Es claro que el mínimo se alcanza en $\bar{\mathbf{x}} = (0, 0)$ y tenemos que $\nabla f(0, 0) = (2, 0)$, $\nabla g_1(0, 0) = (0, 2)$, $\nabla g_2(0, 0) = (0, -1)$ e $I(\bar{\mathbf{x}}) = \{1, 2\}$. La condición de Mangasarian-Fromovitz no se cumple. Si existiese una dirección $\mathbf{d} = (d_1, d_2) \in \mathbb{R}^2$ que la cumpliera 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. Además, si $\lambda, \mu \geq 0$ fuesen multiplicadores de Lagrange para $(0, 0)$, entonces

$$\nabla f(0, 0) + \lambda \nabla g_1(0, 0) + \mu \nabla g_2(0, 0) = (0, 0)$$

$$\Updownarrow$$

$$(2, 2\lambda - \mu) = (0, 0),$$

que no se puede dar.

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. Destacar también que en el caso convexo e imponiendo ciertas condiciones de regularidad podemos probar resultados análogos sin hipótesis de diferenciabilidad [1].

Valoración de activos financieros

En este último capítulo del trabajo no centramos en las *matemáticas financieras*, más concretamente, nuestro objetivo es obtener precio de un activo financiero determinado. Para ello, será esencial el teorema de separación de convexos. De este modo, primero introduciremos los conceptos financieros que son necesarios para la comprensión de nuestro problema. Destacan el denominado *principio de arbitraje* y las *opciones compra*, que son los activos que queremos valorar. En la segunda sección, estableceremos el modelo matemático de nuestro mercado siguiendo el propuesto por [2]. Al acabar la misma, y gracias al teorema de separación, demostraremos el “primer teorema fundamental de asignación de precios”. Finalmente, se ha incluido una sección donde aplicaremos este último resultado para valorar las *opciones de compra* e ilustraremos cómo varía ese valor en función de sus parámetros.

9.1. Preliminares financieros

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 o financieros. Una vez se haya completado este apartado estaremos dispuestos a enunciar y probar la aplicación de los teoremas de la alternativa 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 diferenciar 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 en una cuenta bancaria 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 debido a muchas causas como, por ejemplo, la falta de riesgo absoluto antes mencionadas. 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 el retorno en vez del precio aunque se puede pasar de una a otra fácilmente. Podemos suponer una situación con gran número de posibles retornos 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 el retorno tiene el mismo comportamiento. El retorno para un instante se suele definir como:

$$R_t = \frac{S_t}{S_{t-1}},$$

donde S_t es la variable aleatoria por la que se rige el precio del activo. También se puede definir el retorno como $K_t = R_t - 1 = (S_t - S_{t-1})/S_{t-1}$. 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, $\Omega = \{\omega_1, \omega_2\}$. Tenemos por tanto la variable aleatoria $R_t : \Omega \rightarrow (-1, \infty)$ definida como:

$$R_t = \begin{cases} 1 + u & \text{con probabilidad } p \\ 1 + d & \text{con probabilidad } 1 - p \end{cases}$$

cumpliendo $-1 < d < u$ y $0 < p < 1$. La primera condición es importante ya que garantiza que todos los precios van a ser positivos tal y como especificaremos más adelante. Si por ejemplo $d < 0$, significa que $R_t < 1$ y por tanto $S_t < S_{t-1}$, es decir, hemos perdido dinero. Si por el contrario es positivo, significa que hemos obtenido beneficio. No se tiene que dar que $d < 0 < u$ ya que podríamos estar en una situación en la que siempre perdamos dinero o lo ganemos. El valor d significa la menor ganancia y u la mayor. Deberíamos denotar como $R_t(\omega)$ al retorno obtenida en el paso t si el mercado sigue el escenario $\omega \in \Omega$. El árbol de probabilidad para dos pasos se observa en la figura 9.1.

Por lo tanto, el precio el instante 1 viene dado por

$$S_1 = \begin{cases} S_0(1 + u) & \text{con probabilidad } p \\ S_0(1 + d) & \text{con probabilidad } 1 - p \end{cases},$$

donde S_0 es el precio actual del activo y que es conocido. Siguiendo un razonamiento

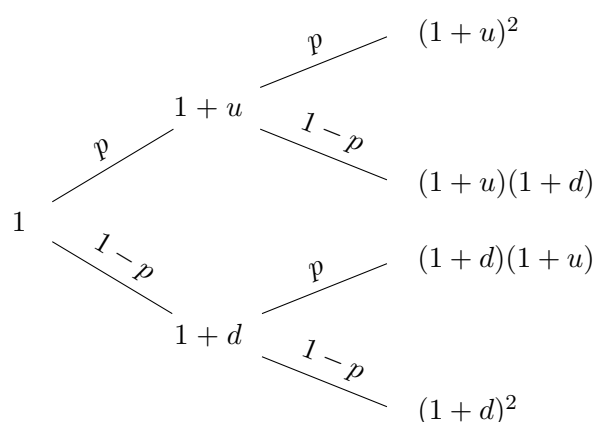


Figura 9.1: Ganancias en un árbol binomial de dos pasos.

“binomial”, el precio S_t de un activo se calcula de la siguiente forma:

$$S_t = S_0(1+u)^i(1+d)^{t-i} \text{ con probabilidad } \binom{t}{i} p^i (1-p)^{t-i}, \quad (9.1)$$

para $i = 1, \dots, t$, donde i es el número de escenarios en los que la ganancia es u (por ello hay $t-i$ escenarios en los que la ganancia es d).

Los activos que hasta ahora hemos presentado son 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, como el de Chicago.
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 esta memoria nos centraremos en las opciones europeas.

Si la situación es la que se ha explicado anteriormente, el propietario 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* o ganancia obtenida en una opción *call* es $C_T = \max\{S_T - K, 0\} = [S_T - K]^+$ y de una opción *put* $P_T = \max\{K - S_T, 0\} = [K - S_T]^+$. Por eso, el comprador debe pagar una prima 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 de las opciones *call* y *put* se representan en las gráficas 9.2 y 9.3 respectivamente.

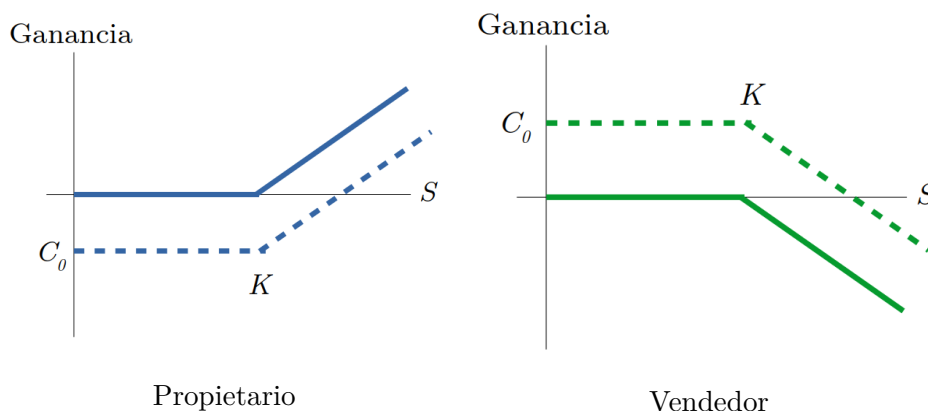
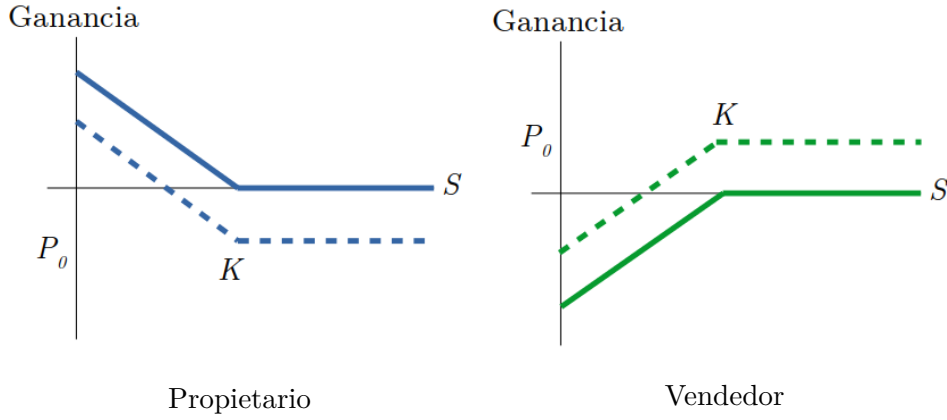


Figura 9.2: Ganancia opción *call*.

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. Para más información acerca de conceptos financieros, se pueden consultar [2] y el manual para universitarios publicado por la Comisión Nacional del Mercado de Valores que se puede consultar en <https://www.cnmv.es/DocPortal/Publicaciones/Guias/ManualUniversitarios.pdf>.

9.2. Primer teorema fundamental

Una vez explicada de manera simplificada el contexto financiero debemos formalizar matemáticamente el modelo del mercado. Para ello, seguimos [2]. Fijamos

Figura 9.3: Ganancia opción *put*.

el conjunto de tiempos $\mathbb{T} = \{0, 1, \dots, T\}$ con T el instante en el que finaliza el modelado de nuestra actividad económica. También fijamos el espacio de probabilidad (Ω, \mathcal{F}, P) . 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 sucesión finita y creciente de sub- σ -álgebras de \mathcal{F} , $\mathcal{F}_0 \subset \dots \subset \mathcal{F}_T = \mathcal{F}$ con \mathcal{F}_0 trivial, es decir, contiene solamente conjuntos con medida 0 o 1. A esta secuencia se le denomina filtración y se denota como $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$. El sentido financiero de esta filtración es indicar la información conocida hasta el momento acerca de la evolución de los precios. Por ejemplo, si comenzamos a modelar nuestro mercado hoy, \mathcal{F}_0 nos aporta únicamente el valor actual. Sin embargo, mañana tendremos la información de los precios en ese momento junto con la obtenida hoy, por lo que es claro que $\mathcal{F}_0 \subset \mathcal{F}_1$. También es claro que al llegar al instante final tenemos toda la información acerca del modelo. 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 $S = \{S_t^i : t \in \mathbb{T}, i = 0, \dots, d\}$ donde cada S_t^i es una variable aleatoria. Todos son activos con riesgos menos el marcado por 0. Asumimos que el proceso S_t^i es \mathcal{F}_t -medible para $i = 0, \dots, d$, es decir, se adapta a la filtración \mathbb{F} . Eso significa que para cada $t \in \mathbb{T}$ se conocen los precios de cada activo hasta ese instante como se ha explicado anteriormente. 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 $\beta_t < 1$ y notamos al valor actual como $\tilde{S}_t = \beta_t S_t$. 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 . Podríamos considerar que tomar $S^0(0) = 1$ es una simple normalización de los precios. El factor de descuento se puede ver como $\beta_t = (1 + r)^{-t}$. Para entender esta fórmula podemos ver la situación contraria: calcular una cantidad con ciertos intereses. De este modo, r representa el interés considerado o cómo aumentará el valor del dinero

en el futuro. Así, el valor de una cantidad S en el futuro es $(1+r)S$ considerando solo un paso o $(1+r)^t S$ si consideramos varios. Por eso, al actualizar pasamos dicho término dividiendo y llegamos a la fórmula de β_t expuesta. Observemos que en ocasiones, como en el caso continuo, el papel de $(1+r)^t$ es una función exponencial.

La cartera de inversión o portafolio para el instante $t \in \mathbb{T}$ viene dada por la variable aleatoria $d+1$ dimensional $\theta_t = (\theta_t^0, \dots, \theta_t^d)$. Cada θ_t^i indica el número de activos del tipo i que tiene un inversor en el instante t . El valor de la cartera 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 selecciona la cartera de inversión del instante t una vez se conocen los precios del momento anterior $t-1$. La estrategia de inversión $\theta = \{\theta_t : t = 1, \dots, T\}$ es el conjunto de todas las carteras de inversión *predecibles*. Decimos que una cartera θ_t es predecible si depende solamente de los precios de los activos hasta el instante $t-1$. Cuando cada cartera de inversión θ_{t+1} se puede financiar completamente con las ganancias o pérdidas actuales decimos que la estrategia es *autofinanciada*, es decir,

$$\langle \theta_{t+1}, S_t \rangle = \langle \theta_t, S_t \rangle,$$

Notamos $\Delta X_t = X_t - X_{t-1}$ para cualquier función sobre \mathbb{T} . Si una cartera es autofinanciada, 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

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

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

$$V_t(\theta) = V_0(\theta) + G_t(\theta), \quad \forall t \in \mathbb{T}. \quad (9.2)$$

Destacamos que todas la definiciones anteriores se pueden definir también en base a los valores actualizados \tilde{S}_t .

También es necesario imponer o aclarar una serie de suposiciones sobre nuestro modelo financiero:

1. Ninguna de las transacciones conlleva un coste extra.
2. Positividad: todos los precios son positivos, es decir,

$$S_t^i > 0 \text{ para } t \in \mathbb{T}, i = 0, \dots, d.$$

3. Divisibilidad, liquidez y *short selling* (venta en corto): el número de activos que posee un inversor puede ser cualquier valor real. Así,

$$\theta_t^i \in \mathbb{R} \text{ para } t \in \mathbb{T}, i = 0, \dots, d.$$

La divisibilidad hace referencia a que θ_t^i puede ser una fracción. Claramente, no podemos tener, por ejemplo, media acción. Sin embargo, cuando estamos trabajando con un gran número de acciones podemos considerar que tienen cifras decimales para trabajar con números menores. El hecho de que tome valores en \mathbb{R} también significa que podemos comprar o vender el número de acciones deseado, es decir, no imponemos ninguna restricción. Este hecho se conoce como liquidez. Claramente, en mercados reales sí existe dicha restricción en el volumen de las transacciones pero nosotros estamos trabajando con un modelo idealizado. Finalmente, cuando dichos valores son positivos decimos que el inversor tiene una posición larga o *long position*. Si por el contrario estos son negativos, tiene una posición corta o *short position*, por ejemplo, cuando vende algún tipo de bono. Estas acciones se suelen llevar a cabo, por ejemplo, en las denominadas ventas en corto o *short selling* que se usan cuando se prevé que un determinado activo va a bajar de valor. Así, suponemos que tenemos una acción cuyo precio actual es 100 cuyo valor suponemos que va a disminuir. Antes de que baje más, la vendemos, es decir, tomamos una posición *short*. Pasado un periodo de tiempo, la acción bajan a 70 por lo que tomamos una posición *long* y las compramos de nuevo. De este modo, tenemos las misma cantidad de activos que al principio pero hemos obtenido un beneficio de 70. En estas operaciones también corremos el riesgo de perder dinero ya que es posible que el valor de los activos crezca en vez de disminuir.

4. Solvencia: el valor de todas las carteras de inversión debe ser siempre positivo,

$$V_t(\theta) \geq 0 \text{ para todo } t \in \mathbb{T}.$$

En este caso, decimos que la cartera es *admisable*. A la clase de estrategias autofinanciadas y admisibles la denotaremos como Θ_a .

5. El espacio Ω es finito, es decir, las variables aleatorias S_t^i no pueden tomar infinitos valores. Así, tenemos que $\Omega = \{\omega_1, \dots, \omega_n\}$.

Ya tenemos el modelo financiero sobre el que vamos a trabajar. Exponemos ahora unos resultados que nos servirán posteriormente para probar el teorema principal del capítulo.

Lema 9.1. Dada V_0 una función \mathcal{F}_0 -medible y para $d \in \mathbb{N}$ sean los procesos reales y predecibles $\theta^1, \dots, \theta^d$, el único proceso predecible θ^0 que convierte a $\theta = (\theta^0, \theta^1, \dots, \theta^d)$ en una estrategia autofinanciada con valor $V_0(\theta) = V_0$ viene dado por

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

Demostración. Claramente θ^0 es predecible. Para ver que la estrategia es autofinanciada, por (9.2) 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*, esencial para lo que sigue. 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 $\theta \in \Theta_a$. La viabilidad del mercado impone la siguiente restricción.

Lema 9.2. *Si el modelo de mercado es viable, las ganancias actualizadas a cualquier proceso predecible $\hat{\theta} \in \mathbb{R}^d$ no pueden pertenecer a*

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

Demostración. En primer 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 $\theta \in \Theta_a$ tal que $V_0(\theta) = 0$ entonces

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

Por el lema 9.1, dados los procesos predecibles $\hat{\theta} = (\theta^1, \dots, \theta^d)$, existe un único proceso real θ^0 tal que $\theta = (\theta^0, \theta^1, \dots, \theta^d)$ 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 entonces que $V_T(\theta)$ 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 9.1. 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

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

También es importante destacar que, por la linealidad de la esperanza, cualquier combinación lineal de martingalas es una martingala.

Que los precios en el mercado sigan una martingala no debe de ser extraño. Recordemos que para $t \in \mathbb{T}$, \mathcal{F}_t indica la cantidad de información acerca de los precios de los activos hasta dicho momento. Por lo tanto, la esperanza condicionada solo nos indica que estamos calculando el valor esperado a partir de lo que conocemos hasta ahora. Además, para que el mercado sea “justo”, dicho valor en el futuro debería ser, en media, el que tenemos ahora.

El siguiente resultado es meramente técnico y nos servirá posteriormente.

Teorema 9.1. Un proceso real M es una martingala si, y solo si,

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

para todo proceso ϕ predecible y acotado.

Demostración. Si M es una martingala, también lo es la transformada $X = \phi \cdot M$ 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. Para $s > 0$, sea $A \in \mathcal{F}_s$ y definimos el proceso predecible ϕ como $\phi_{s+1} = 1_A$ y $\phi_t = 0$ para el resto de $t \in \mathbb{T}$. 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:

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

donde E_Q significa la esperanza respecto de la probabilidad (medida) Q . Sea $\theta \in \Theta_a$ una estrategia admisible tal que los procesos de los precios actualizados son integrables respecto a Q . Por (9.2) tenemos que:

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

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 9.2. Una probabilidad Q que sea equivalente a P como media ($Q \sim P$) 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 cada $i = 0, \dots, d$ tenemos que \bar{S}^i es una (\mathbb{F}, Q) -martingala.

Todo lo presentado anteriormente, nos aporta el siguiente resultado que acabamos de demostrar:

Proposición 9.1. Si existe una medida martingala equivalente para S , entonces el modelo de mercado discreto es viable, es decir, no contiene ninguna oportunidad de arbitraje.

El recíproco de esta afirmación es cierto, tal y como exponemos en el siguiente teorema que es el objetivo final de este capítulo y en el que usaremos el teorema de la alternativa de Gordan en forma equivalente del teorema de separación establecido en el corolario 7.1. Recibe el nombre de “primer teorema fundamental de asignación de precios”.

Teorema 9.2. *(Primer teorema fundamental de asignación de precios) 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 9.1 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} . Sea C es el cono convexo de todas las variables aleatorias reales ϕ en (Ω, \mathcal{F}) tales que $\phi(\omega) \geq 0$ casi seguramente y $\phi(\omega_i) > 0$ para al menos un $\omega_i = \Omega = \{\omega_1, \dots, \omega_n\}$. Asumimos $p_i = P(\{\omega_i\}) > 0$. Por el lema 9.2, hemos visto que para 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

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

es un subespacio vectorial del espacio 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 $K = \{X \in C : E_P(X) = 1\}$ gracias al corolario 7.1. Sea $\mathbf{x}^0 \in \mathbb{R}^d$ el elemento que proporciona dicho resultado. Tomamos $\xi_i = (0, \dots, \frac{1}{p_i}, \dots, 0)$ para $i \leq n$. Vemos que $E_P(\xi_i) = \frac{p_i}{p_i} = 1$ por lo que $\xi_i \in K$ y $\langle \mathbf{x}^0, \xi_i \rangle = \frac{x_i^0}{p_i} > 0$. Por ello, $x_i^0 > 0$ para todo $i = 1, \dots, n$.

Definimos ahora el funcional lineal $g(\mathbf{x}) = \frac{\langle \mathbf{x}^0, \mathbf{x} \rangle}{\alpha}$ donde $\alpha = \sum_{i=1}^n x_i^0$. Sea $p^* \in \mathbb{R}^d$ un vector 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 $\Omega = \{\omega_1, \dots, \omega_n\}$ haciendo $P^*(\{\omega_i\}) = p_i^* > 0$. Veamos que P^* es la martingala equivalente deseada. En efecto, sea $E^*(\cdot)$ que denota la esperanza relativa a P^* . Nuevamente, por el corolario 7.1, tenemos que $g(\mathbf{x}) = \frac{1}{\alpha} \langle \mathbf{x}^0, \mathbf{x} \rangle = 0$ para todo $\mathbf{x} \in L$. En particular, esta situación se da para $\bar{G}_T(\hat{\theta})$ con $\hat{\theta} = (\theta^1, \dots, \theta^d)$ un vector de procesos predecibles por lo que.

$$E^*(\bar{G}_T(\hat{\theta})) = 0.$$

Hemos conseguido una estrategia auto-financiada θ con $V_0(\theta) = 0$. Como $\bar{V}_t(\theta) = V_0(\theta) + \bar{G}_T(\theta)$, implica que $E^*(\bar{V}_t(\theta)) = 0$ para cada θ . Por el lema 9.1 podemos generar tal θ para cada proceso predecible de n dimensiones. En particular, lo podemos hacer para $(0, \dots, \theta^i, \dots, 0)$ donde $i \leq n$. Por lo tanto

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

se da para cada proceso predecible y acotado $(\theta^i)_{i=1, \dots, T}$. El teorema 9.1 implica que cada S^i es una martingala bajo P^* . Por ello, $P^* \sim P$. ■

9.3. Aplicación al modelo binomial

Usemos toda la información obtenida para llegar a una fórmula que nos permitirá establecer el precio de una opción europea. Para ello, utilizaremos un modelo binomial con más de un paso en el que suponemos que $d = 1$. De este modo, tenemos un activo sin riesgo con precios S^0 y uno con riesgo con precios S^1 los cuales denotaremos por S . También, como mencionamos anteriormente, sin pérdida de generalidad asumimos que $S_0^0 = 1$ por lo que $S_t^0 = (1 + r)^t$ para $t \in \mathbb{T}$. Los precios de nuestro activo con riesgo vienen determinados por $S_{t-1}(1 + u)$ o $S_{t-1}(1 + d)$ cumpliendo $-1 < d < u$, es decir, siguen una distribución de Bernoulli. El valor inicial S_0 es constante y conocido. Tomamos entonces como espacio de probabilidad $\Omega = \{1 + u, 1 + d\}^T$ con la filtración \mathbb{F} dada por $\mathcal{F}_0 = \{\emptyset, \Omega\}$ y $\mathcal{F}_t = \sigma(S_u^1 : u \leq t)$ para $t > 0$. Para $\omega = \{\omega_1, \dots, \omega_n\}$ en Ω definimos

$$P(\{\omega\}) = P\left(\frac{S_t}{S_{t-1}} = \omega_t, t = 1, \dots, T\right).$$

Notamos a $\frac{S_t}{S_{t-1}}$ como R_t tal y como ya hemos hecho anteriormente. Veamos la relación necesaria entre d, u y r para que tengamos una medida martingala equivalente.

Lema 9.3. *Para tener una EMM en el modelo binomial se debe cumplir que $d < r < u$.*

Demostración. Sea Q una probabilidad en (Ω, \mathcal{F}) . Para que sea una martingala equivalente se debe dar:

$$E_Q(\bar{S}_t | \mathcal{F}_{t-1}) = \bar{S}_{t-1}.$$

$$\Updownarrow$$

$$E_Q(R_t | \mathcal{F}_{t-1}) = E_Q\left(\frac{\bar{S}_t}{\bar{S}_{t-1}} | \mathcal{F}_{t-1}\right) = E_Q\left(\frac{\beta_t}{\beta_{t-1}} | \mathcal{F}_{t-1}\right) = \frac{\beta_t}{\beta_{t-1}} = 1 + r.$$

Como R_t solo puede tomar valores $1 + u$ y $1 + d$, su valor medio solo puede $1 + r$ si, y solo si, $d < r < u$. ■

A continuación, vamos a construir dicha medida Q mediante el siguiente lema:

Lema 9.4. *El proceso \bar{S} es una (\mathbb{F}, Q) -martingala si, y solo si, las variables aleatorias R_t son independientes e idénticamente distribuidas con $Q(R_1 = 1 + u) = \frac{r-d}{u-d}$ y $Q(R_1 = 1 + d) = 1 - \frac{r-d}{u-d}$. Además, si el modelo es viable, Q es única.*

Demostración. Para simplificar la notación, llamamos $q = \frac{r-d}{u-d}$.

\Leftarrow) Como R_t son independientes:

$$E_Q(R_t|\mathcal{F}_{t-1}) = E_Q(R_t) = q(1+u) + (1-q)(1+d) = q(u-d) + 1 + d = 1 + r.$$

Siguiendo un proceso similar a la demostración del lema 9.3, vemos que efectivamente es una (\mathbb{F}, Q) -martingala.

\Rightarrow) En este caso, $E_Q(R_t|\mathcal{F}_{t-1}) = 1 + r$. Como solo puede tomar valores $1 + u$ y $1 + d$, se debe cumplir que

$$(1 + d)Q(R_1 = 1 + d|\mathcal{F}_{t-1}) + (1 + u)Q(R_1 = 1 + u|\mathcal{F}_{t-1}) = 1 + r$$

y

$$Q(R_1 = 1 + d|\mathcal{F}_{t-1}) + Q(R_1 = 1 + u|\mathcal{F}_{t-1}) = 1.$$

Llamando $q = Q(R_1 = 1 + u|\mathcal{F}_{t-1})$ obtenemos

$$(1 + d)(1 - q) + (1 + u)q = 1 + r \implies q = \frac{r - d}{u - d}.$$

Con un argumento inductivo, para $\omega = (\omega_1, \dots, \omega_T) \in \Omega$ vemos que

$$Q(R_1 = \omega_1, \dots, R_t = \omega_t) = \prod_{i=1}^t q_i$$

donde

$$q_i = \begin{cases} q & \text{si } \omega_i = 1 + u \\ 1 - q & \text{si } \omega_i = 1 + d \end{cases}.$$

Por lo tanto, las variables R_t son independientes e idénticamente distribuidas.

Finalmente veamos la unicidad. Es claro que $q \in (0, 1)$ si, y solo si, $d < r < u$. Por ello, usando el teorema 9.2 y el lema 9.3, obtenemos que el modelo admite una única EMM que además es Q . ■

Al valor $q = (r - d)/(u - d)$ obtenido en el resultado anterior se le suele denominar *probabilidad de riesgo neutral*. Es un objeto matemático abstracto que no tiene que coincidir con la probabilidad real del mercado o estar relacionada con la misma y que de hecho, se conoce *a priori*. Sin embargo, para la valoración de opciones se utiliza q ya que es la que hace que los precios actualizados se comporten según una martingala tal y como acabamos de probar. De este modo, la ganancia $C_T = [S_T - K]^+$ de una opción *call* en el instante $t \in \mathbb{T}$ viene dada por

$$V_t(C_T) = \frac{1}{\beta_t} E_Q(\beta_T C_T | \mathcal{F}_t).$$

Usando la definición de R_t , vemos que se cumple la siguiente relación:

$$S_T = S_t \prod_{u=t+1}^T R_u.$$

Podemos calcular su media ya que S_t es \mathcal{F}_t -medible y R_u son independientes de \mathcal{F}_t si $u > t$. De ese modo,

$$\begin{aligned} V_t(C_T) &= \frac{\beta_T}{\beta_t} E_Q \left(\left[S_t \prod_{u=t+1}^T R_u - K \right]^+ \mid \mathcal{F}_t \right) \\ &= (1+r)^{t-T} E_Q \left(\left[S_t \prod_{u=t+1}^T R_u - K \right]^+ \mid \mathcal{F}_t \right) \\ &= (1+r)^{t-T} \sum_{v=0}^{T-t} \binom{T-t}{v} q^v (1-q)^{T-t-v} [S_t(1+u)^v(1+d)^{T-t-v} - K]^+, \end{aligned}$$

donde en el último paso se ha usado la fórmula (9.1). Finalmente, en el instante 0, el valor de la opción es:

$$V_0(C_T) = (1+r)^{-T} \sum_{v=A}^T \binom{T}{v} q^v (1-q)^{T-v} [S_0(1+u)^v(1+d)^{T-v} - K],$$

donde A denota el primer entero k que cumple $S_0(1+u)^k(1+d)^{T-k} > K$. Esto se debe a que

$$[S_0(1+u)^j(1+d)^{T-j} - K]^+ = 0, \text{ para } j < k.$$

Esta fórmula nos aporta el valor de una opción *call* o de compra. En la explicación de estos activos derivados dijimos que también existían las denominadas opciones *put* o de venta. Gracias a la ausencia de arbitraje podemos calcular el precio de una en función del precio de la otra. A este hecho se le denomina *paridad put-call*. Si denominamos por C_t al precio de la opción *call* y como P_t al de la opción *put* y tenemos en cuenta las ganancias como muestran las gráficas 9.2 y 9.3, se da la siguiente relación.

$$C_T - P_T = (S_T - K)^+ - (K - S_T)^+ = S_T - K.$$

Esta relación se debe dar para todo $t \in \mathbb{T}$ y si consideramos la actualización del valor de K en cada instante mediante el interés r llegamos a:

$$C_t - P_t = S_t - (1+r)^{-(T-t)}K,$$

donde el $(1+r)^{-(T-t)}$ es el factor de actualización entre los instantes t y T . Este hecho es bastante conocido y aparece recogido, por ejemplo, en [2].

Una vez obtenida la fórmula y hecha la aclaración sobre la relación de precios entre los dos tipos de opciones, podemos realizar una serie de simulaciones para ver cómo varía el precio de una opción de compra según los distintos parámetros (K, T, S, d, r, u) . Estas simulaciones se han realizado en *SageMath* 2.8. Para ello se han implementado las siguientes funciones como se observa en 9.1: una para obtener el valor de A llamada *getA* y otro el de la propia opción, *getV*.

Código 9.1: Definición en Sage de las funciones

```
1 sage: def getA(K, T, S, d, u):
2     k = 0
```



```

3         while S*(1+u)^(k)*(1+d)^(T-k) < K:
4             k+=1
5         return k
6 sage: def getV(K, T, S, d, r, u):
7         A = getA(K, T, S, d, u)
8         q = (r-d)/(u-d)
9         if r <= d or r >= u:
10            return 0
11        else:
12            return (1+r)^(-T)*sum(binomial(T,v)*q^v*(1-q)^(T-v)*(S*(1+u)
            ^v)*(1+d)^(T-v)- K) for v in range(A,T))
    
```

Para ello, partiremos del siguiente problema general del que iremos variando los diferentes parámetros en el que tomamos una unidad monetaria (u.m.) cualquiera.

“Las acciones de una empresa tienen valor actual 25 u.m. cada una. Queremos comprar acciones de dicha empresa pero no estamos seguros de obtener una gran ganancia con ellas en el futuro. Por eso, decidimos realizar una opción de compra de aquí a un año en el que se nos ofrece un precio de 27 u.m. Los tiempos se mueven en el intervalo de meses. Según las estimaciones suponemos que en cada mes el retorno podría ser de 1.2 ($u = 0.2$) pero también podemos perder dinero llegando este a tomar el valor 0.72 ($d = -0.28$). Consideramos que el precio del dinero se incrementa de manera constante cada mes un 0.1 (de este modo el mercado es viable).”

El precio en esta situación es aproximadamente de 12,9325 u.m. Veamos cómo varía en función de cada parámetro manteniendo el resto como en el enunciado. En primer lugar, consideremos el valor como función de T y tomamos valores desde un mes hasta dos años, es decir, 24 meses. Los resultados se obtienen en la figura 9.4. También se incluye un ejemplo del código en 9.2 para dibujar la gráfica en este caso junto con la definición de los parámetros generales⁵. Para el resto de parámetros se hace de manera análoga.

Código 9.2: Código en Sage para dibujar una gráfica

```

1 sage: K = 27
2 T = 12
3 S = 25
4 d = -0.28
5 r = 0.1
6 u = 0.2
7 sage: point([(T, getV(K, T, S, d, r, u)) for T in xrange(1, 25,1) ],
    rgbcolor = (1,0,0), size = 35, axes_labels=['T$', '$V_0(C_T)$'])
    
```

Vemos que tendríamos que poner una fecha de compra superior a los 4 meses ya que durante los tres primeros $A = T$, lo que significa que lo más seguro es que no ejerciéramos nuestro derecho a la opción. Observamos que los valores crecen en función de este parámetro. Veamos qué pasa si variamos el precio de la acción fijado en la opción de compra, es decir, K . Se ha obtenido de este modo la gráfica 9.5 en la que K varía de 23 a 33. Vemos que en este caso es decreciente. Esta situación podemos considerar que es la esperada. Cuanto menor sea el precio más posibilidades tiene el propietario de ejercerla ya que es raro que en el futuro la acción tenga ese valor tan bajo. Por ello, el vendedor tiene que exigir un mayor. Por otro lado, cuanto mayor sea el precio acordado es el comprador el que corre un mayor riesgo

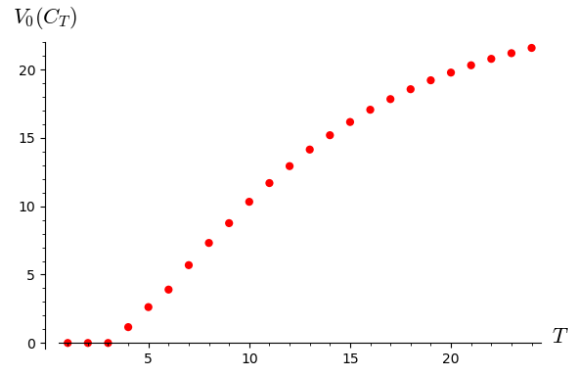


Figura 9.4: Valor de la opción en función del número de pasos.

de no ejercerla en el futuro por lo que tiene que pagar menos por ella ahora.

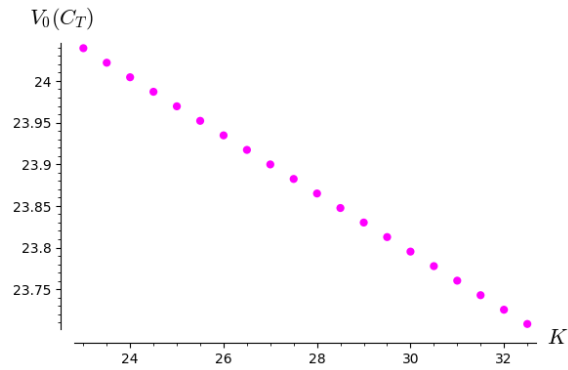


Figura 9.5: Valor de la opción en función del precio acordado.

Obtengamos ahora el precio en función del valor actual S_0 . Hemos variado dicho valor entre 23 y 33. Los resultados se muestran en la gráfica 9.6.

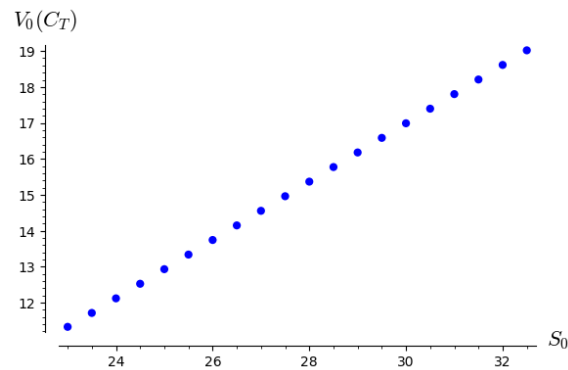


Figura 9.6: Valor de la opción en función del valor actual.

En este caso vuelven a ser crecientes, lo que era de espera, r ya que en la fórmula vemos que el precio de la opción es proporcional al valor actual de la misma. Finalmente, variemos los valores de u , d y r teniendo en cuenta las restricciones del lema 9.3. En la gráfica 9.7 se muestran los resultados. La variación de este valor

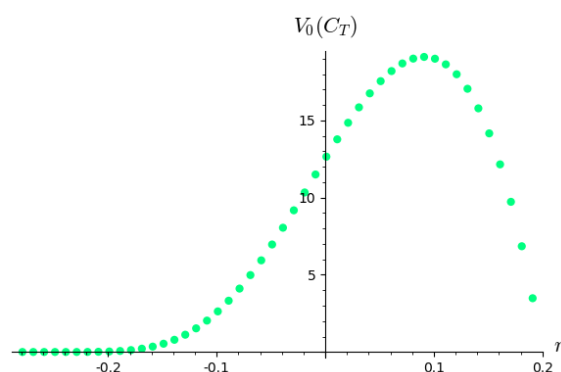


Figura 9.7: Valor de la opción en función del interés.

es quizás la más interesante desde el punto de vista del vendedor de la opción. El parámetro r es en cierto modo subjetivo ya que es una estimación del valor que tendrá el dinero en el futuro. Por ello, según los valores fijados de d y u puede ver con qué interés obtiene el mayor beneficio. Es claro que si $r < 0$ significa que el dinero en el futuro tendrá menos valor y es una situación que en principio no la contemplamos ya que suponemos que el valor del dinero aumentará en un futuro. De todos modos, se han usado valores negativos para ver su evolución. Para $r \geq 0$ vemos que en este caso la función alcanza un máximo que como hemos dicho es el valor más favorable al vendedor. Si consideramos un interés mayor al mismo le estamos exigiendo demasiada rentabilidad y eso hace bajar el precio de la opción. Veamos cómo varía ahora para valores de u entre $-0,4$ y $0,1$ mostrados en la gráfica 9.8. En este caso también es creciente lo que es normal ya que cuanto mayor sea

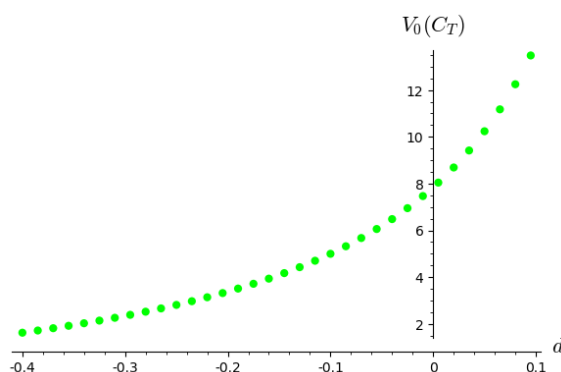


Figura 9.8: Valor de la opción en función de d .

d mayor será la posible ganancia del comprador. Por último veamos cómo varía en función de u . Para ello se han tomado valores entre $0,1$ y $0,55$. Se muestran en la

gráfica 9.9.

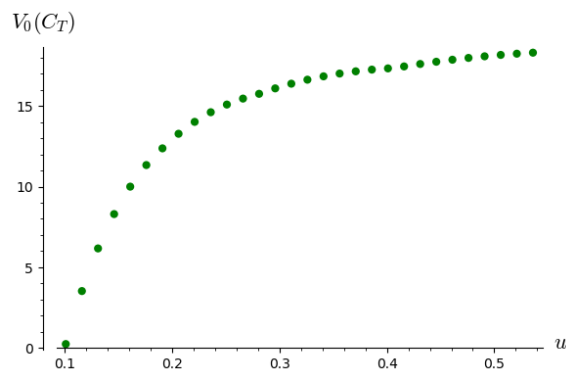


Figura 9.9: Valor de la opción en función de u .

Observamos que para valores cercanos al interés el valor de la opción es relativamente bajo ya que la ganancia que obtendremos sería menor. El precio va aumentando aunque llega un punto que dicho crecimiento se reduce.

Conclusiones y vías futuras

Los objetivos que nos marcamos en la propuesta inicial se han alcanzado satisfactoriamente. Es más, se ha realizado una incursión no prevista en la teoría minimax de manos de los teoremas de la alternativa. Ello ha permitido obtener una visión completa de las técnicas y aplicaciones de los teoremas de la alternativa.

Bibliografía

- [1] J.M. Borwein, A.S. Lewis, *Convex analysis and nonlinear optimization: theory and examples*, Second Edition, CMS Books in Mathematics/Ouvrages Mathématiques de la SMC 3, Springer, New York, 2006.
- [2] R.J. Elliot, R.E. Kopp, *Mathematics of financial markets*, Second Edition, Springer Finance, New York, 2005.
- [3] G. Giorgi, A. Guerraggio, J. Thierfelder, *Mathematics of optimization: smooth and nonsmooth case*, Elsevier Science B.V., Amsterdam, 2004.
- [4] H. König, *Über das von Neumannsche minimax-theorem*, Archiv der Mathematik 19(1968), 482-487.
- [5] H. König, *Sublinear functionals and conical measures*, Archiv der Mathematik 77(2001), 56-64.
- [6] S. Mazur, W. Orlicz, *Sur les espaces métriques linéaires II*, Studia Mathematica 13(1953), 137-179.
- [7] E. Schechter, *Handbook of analysis and its foundations*, Academic Press, Inc., San Diego, CA, 1999.
- [8] S. Simons, *From Hahn-Banach to monotonicity*, 2nd edition, Lectures notes in Mathematics 1693, Springer, New York, 2008.