



UNIVERSIDAD CENTRAL DE VENEZUELA  
FACULTAD DE CIENCIAS  
POSTGRADO EN MATEMÁTICA

# Problemas de máxima entropía y cálculo variacional

**Autor:** Lic. Andrés Pérez

**Tutor:** Dr. Ramón Bruzual

Trabajo de Grado de Maestría presentado ante la ilustre Universidad Central de Venezuela para optar al título de Magister Scientiarum, Mención Matemática .

Caracas, Venezuela

Julio - 2007

# Dedicatoria

Existen varias personas a las cuales les dedico mi esfuerzo y dedicación, a las cuales les debo en muy buena medida parte de mi existencia, de mi ser y de mis ganas de vivir. Estas personas, son hoy por hoy mi centro de atención, mi luz al final del túnel, el camino de la vida y el motor que impulsa cualquier intención de prosperar y desarrollarme.

Este trabajo, está dedicado a las personas que representan el sueño tranquilo, el despertar bonito, la paz interior y el amor puro, la fraternidad y el compañerismo.

Estas personas son justamente, mis hijos, Leonardo Andrés y Yeleand Alexandra, mi esposa y compañera de batallas Janeth del Carmen y una persona muy especial que no está presente en cuerpo, pero si en alma, mi abuela Olga Mercedes (q.e.p.d).

Gracias por simplemente creer en mí.

# Agradecimientos

Mis agradecimientos más sinceros a mi tutor, el Prof. Ramón Bruzual, por abrir una puerta cuando otros la cierran, por creer en mi y por su disposición al trabajo. Al Prof. José Rafael León, por estar siempre dispuesto a ayudar y a todos los amigos y compañeros que con su positivismo, influyeron de alguna u otra forma en la culminación exitosa de este trabajo.

# CONTENIDO

<b>1</b>	<b>Entropía de Shannon</b>	<b>4</b>
1.1	Entropía de Shannon . . . . .	4
1.2	Otras definiciones de entropía . . . . .	9
1.2.1	La entropía de Boltzmann . . . . .	10
1.2.2	La entropía de Gibbs . . . . .	11
<b>2</b>	<b>Ecuación de E - L en un intervalo acotado</b>	<b>12</b>
2.1	Caso finito dimensional . . . . .	12
2.1.1	Extremos incondicionados . . . . .	12
2.1.2	Extremos condicionados . . . . .	14
2.1.3	Método de los multiplicadores de Lagrange . . . . .	15
2.2	Caso general . . . . .	16
2.2.1	Extremos de funcionales . . . . .	16
2.2.2	Ecuación de Euler . . . . .	25
2.2.3	El problema isoperimétrico . . . . .	27
<b>3</b>	<b>Extensión de la ecuación de Euler-Lagrange</b>	<b>37</b>
3.1	Invarianza de la ecuación de Euler . . . . .	37
3.1.1	Cambio de variables generalizado . . . . .	37
3.1.2	Cambio de variables en la ecuación de Euler - Lagrange . . . . .	38
<b>4</b>	<b>Problemas de máxima entropía</b>	<b>44</b>
4.1	Teorema general . . . . .	45
4.1.1	Caracterización de las funciones de máxima entropía según sus restricciones . . . . .	47

*CONTENIDO*

v

**Bibliografía**

**52**

# Resumen

Se considera la definición de entropía dada por Shannon y luego, se hace una deducción de la fórmula de Shannon basándonos en el valor esperado de la función incertidumbre. Posteriormente, se extiende la ecuación de Euler - Lagrange a un intervalo no acotado y usando esta extensión, se resuelven problemas de máxima entropía en la recta considerando diferentes tipos de restricciones.

# Introducción

Originariamente, la entropía se ligó al segundo principio de la termodinámica o también principio de degradación de la energía, ya que, el calor se considera una forma más baja o “degradada” de energía, entendiendo por *degradación*, al paso de una forma de energía a otra que no puede ser acompañada de un proceso de inversión completo.

Entre mediados y finales del siglo XIX, surgieron los primeros trabajos alusivos a la entropía, con el físico francés Nicolas Léonard Sadi Carnot (1796 - 1832) y el físico alemán Rudolf Julius Emmanuel Clausius (1822 - 1888). Siendo el primero, el que dió una formulación de la teoría en 1824 y el segundo, una forma matemática rigurosa en 1850. Luego el físico austriaco Ludwig Boltzmann (1844 - 1906), logró proporcionar una profunda interpretación de esta ley, vinculando para ello, la teoría cinética con la termodinámica, llegando a la conclusión, de que la entropía de un sistema aislado está ligada a la probabilidad de su estado actual.

La magnitud asociada con el número de distintas formas en que podemos disponer las partículas de un sistema, es lo que denominamos *entropía* y matemáticamente la definimos como una cantidad proporcional al logaritmo del número de distribuciones posibles de los estados del sistema.

Formalmente, Claude Elwood Shannon (1916 - 2001), ingeniero eléctrico y matemático, recordado como “el padre de la teoría de la información”, en su artículo “A Mathematical Theory of Communication” [16], define la entropía, como una función que permite medir la cantidad de información asociada a un proceso aleatorio y allí establece que, cuanto mayor sea la probabilidad de que se produzca, menor será la información que aporta.

Matemáticamente, Shannon considera una variable aleatoria  $\mathcal{X}$  y a partir de ella define una nueva variable aleatoria  $I[\mathcal{X}] = -\log_2 p_i$ , llamada *cantidad de información*, donde  $p_i$  es la probabilidad del  $i$ -ésimo evento. Luego, define la entropía como la esperanza de la

cantidad de información, es decir

$$H[\mathcal{X}] = \mathbb{E}[I[\mathcal{X}]] = - \sum_{i=1}^n p_i \log_2 p_i.$$

Si en la entropía de Shannon, consideramos a  $\mathcal{X}$  como una variable aleatoria que toma valores en la recta real y  $f$  es la función de densidad asociada a  $\mathcal{X}$ , la entropía de Shannon se puede definir por el funcional

$$H[f] = - \int_{-\infty}^{\infty} f(x) \log f(x) dx \quad (1)$$

Es en este punto donde se centra el presente trabajo. Basicamente, lo que queremos es determinar las condiciones bajo las cuales podemos garantizar, la existencia y unicidad de las funciones de densidad que maximizan el funcional dado en (1), sometiéndolas a diversas restricciones, para así, finalmente poder caracterizar estas funciones. Para ello, hacemos uso de las técnicas que nos proporciona el cálculo variacional.

Este trabajo, consta de cuatro capítulos. En el primero, se realiza la construcción axiomática de la función incertidumbre (cantidad de información) y con ella se facilita la definición de la entropía de Shannon, pasando por otras definiciones de trascendencia a través de la historia, tales como, la entropía de Boltzmann y la de Gibbs.

El segundo capítulo, está dedicado al cálculo variacional, con la idea de plantear el problema fundamental del cálculo variacional y enfocarnos en la ecuación de Euler - Lagrange, pasando por el caso finito dimensional y luego haciendo un enfoque general hasta llegar al problema isoperimétrico sobre intervalos acotados.

En el tercer capítulo, hacemos la extensión de la ecuación de Euler - Lagrange a toda la recta, observando las condiciones mínimas necesarias para considerar transformaciones que nos permitan dejar invariante dicha ecuación, al pasar de un intervalo acotado a toda la recta.

En el último capítulo, enfocamos el problema de la entropía máxima, buscando caracterizaciones de las funciones de densidad, en virtud, de las restricciones a las cuales están sujetas, usando las técnicas expuestas en el capítulo 3.



# Capítulo 1

## Entropía

El concepto básico de entropía en Teoría de la Información tiene mucho que ver con la *incertidumbre* que existe en cualquier experimento o señal aleatoria, o se puede visualizar también, como la cantidad de “ruido” o “desorden” que contiene o libera un sistema. De esta forma, podremos hablar de la cantidad de información que lleva una señal. Frecuentemente, al remitirnos a la entropía en la formulación que hace de ella la Teoría de la información, encontramos que a esta entropía se le llama *Entropía de Shannon*, en honor de Claude E. Shannon.

Los *procesos irreversibles*, son aquellos en virtud de los cuales los sistemas se acercan al equilibrio. En un proceso irreversible, la entropía del sistema aumenta, mientras que un proceso reversible la entropía del sistema permanece constante, es decir, dicha entropía no disminuye nunca. Una vez que se ha alcanzado el estado de equilibrio, los sistemas, no pueden cambiar sin ser distorsionados por un agente externo.

Parece por tanto, que el predecir la producción y dirección de algunos cambios es cuestión sencilla, lo único que se debe tratar de hacer, es indagar si se cumplen o no los criterios que conducen a un estado de equilibrio, en vías de establecer caracterizaciones de los mismos.

### 1.1 Entropía de Shannon

Intuitivamente, la entropía se puede visualizar como una magnitud  $S$ , que es una función del estado del sistema, es decir, la entropía de una variable de un sistema es la medida esperada de la información que aportará la realización del evento asociado a la variable, no la propia variable.

Por tanto, la entropía se puede entender literalmente, como un número asociado a una variable aleatoria que es igual al valor esperado de la respuesta que recibiríamos al conseguir una comprensión de la variable aleatoria.

Usualmente, la entropía se asocia o está familiarizada a la química, a la termodinámica o a la teoría de la información y quizá de esta última, se pueda inferir directamente que la entropía resulta ser una medida del *desorden* de un sistema.

Shannon, ofrece una definición de entropía que satisface las siguientes afirmaciones:

- La medida de información debe ser proporcional (continua). Es decir, el cambio pequeño en una de las probabilidades de aparición de uno de los elementos de la señal debe cambiar poco la entropía.
- Si todos los elementos de la señal son equiprobables a la hora de aparecer, entonces, la entropía será máxima.

La elección de esta función inesperada hipotética del estado del sistema, a ser denotada por  $S$  que propone Shannon, la podemos fundamentar en que ella debe satisfacer algunos axiomas básicos. Estos son:

**Axioma 1.1.**  $S(1) = 0$ .

Este axioma deja ver en forma muy clara que, lo inesperado es cero cuando el evento es seguro.

El segundo axioma establece, que cuando un evento es mas improbable que otro, lo inesperado debe ser mas grande, es decir, que una entropía alta implica que la propia variable aporta una información pequeña. Veamos:

**Axioma 1.2.** *Si  $p > q$ , entonces  $S(p) > S(q)$ . Así  $S$ , es una función estrictamente decreciente.*

El tercer axioma, establece que pequeños cambios en la probabilidad de un evento determinan pequeñas variaciones en el valor inesperado, a saber:

**Axioma 1.3.**  $S(p)$  es una función continua en  $p$ .

El axioma final, está motivado por lo que ocurre cuando observamos que dos eventos independientes ocurren. Supongamos que el evento  $E$  tiene probabilidad  $p$  de ocurrir y que el evento  $F$  tiene probabilidad  $q$ . Entonces,  $P(EF) = pq$ , corresponde con la probabilidad

de eventos independientes. Por lo tanto, la sorpresa provocada por la ocurrencia tanto de E como de F es  $S(pq)$ . Ahora, la sorpresa provocada sólo por la ocurrencia de  $p$  es  $S(p)$  y por consiguiente, la sorpresa adicional provocada por  $q$  es  $S(pq) - S(p)$ . Sin embargo, si E y F son independientes, entonces, esta sorpresa sólo debe ser  $S(q)$ . Veamos

**Axioma 1.4.**  $S(pq) = S(p) + S(q)$ , para  $p, q \in (0, 1]$ .

Usando estos axiomas podemos derivar la forma funcional de  $S(p)$ . Notemos que el Axioma 1.4, usado en forma inductiva establece que

$$S(p^m) = mS(p)$$

luego para  $n \in \mathbb{Z}^+$ , también se cumple que

$$S(p) = \frac{1}{n}S(p^n)$$

por consiguiente,

$$S(p^{m/n}) = \frac{m}{n}S(p)$$

o equivalentemente

$$S(p^x) = xS(p)$$

cada vez que  $x \in \mathbb{Q}^+$ .

Ahora, necesitamos la continuidad de  $S(p)$ . Consideremos una sucesión  $\{q_n\}_{n \geq 1}$  que converge a  $q$ , entonces en forma análoga a la definición de funciones continuas, tenemos que

$$\lim_{n \rightarrow \infty} q_n = q \implies \lim_{n \rightarrow \infty} S(q_n) = S(q)$$

Escojamos entonces  $q = p^x$ , para cualquier  $x$  positivo y arbitrario y definamos la sucesión  $q_n = p^{\lfloor nx \rfloor / n}$ . Por tanto,

$$\lim_{n \rightarrow \infty} q_n = q = p^x$$

y notemos que  $S(q_n) = S(p^{\lfloor nx \rfloor / n}) = \frac{\lfloor nx \rfloor}{n}S(p)$ , luego

$$\lim_{n \rightarrow \infty} S(q_n) = \lim_{n \rightarrow \infty} S(p^{\lfloor nx \rfloor / n}) = \lim_{n \rightarrow \infty} \frac{\lfloor nx \rfloor}{n}S(p) = xS(p)$$

pero por la continuidad sabemos que

$$\lim_{n \rightarrow \infty} S(q_n) = S(q) = S(p^x)$$

Por consiguiente, para cualquier  $x$  positivo tenemos que

$$S(p^x) = xS(p)$$

Por supuesto, si  $x = 0$  por el Axioma 1.1, obtenemos que  $S(p^0) = S(1) = 0 = 0 \cdot S(p)$ .

Para cualquier  $p$ , tal que  $0 < p \leq 1$  y para cualquier  $a > 0$ , sea  $x = -\log_a p$ , tal que  $p = \left(\frac{1}{a}\right)^x$ , entonces

$$S(p) = S\left(\left(\frac{1}{a}\right)^x\right) = xS\left(\frac{1}{a}\right) = -C \log_a p$$

donde  $C = S\left(\frac{1}{a}\right) > S(1) = 0$  (por los Axiomas 1.1 y 1.2). Por lo tanto, estos axiomas determinan que la función de respuesta inesperada debe ser

$$S(p) = -S\left(\frac{1}{a}\right) \log_a p$$

y si denotamos por  $S_a(p)$  a la función que resulta de escoger  $S\left(\frac{1}{a}\right) = 1$ , tenemos que

$$S_a(p) = -\log_a p$$

Entonces, la entropía para una variable aleatoria discreta  $X$ , la cual tiene probabilidad  $p_i$ , cuando toma el valor  $x_i$ , es calculada usando la probabilidad de la función de masa de la variable aleatoria, esto es

$$H_a(X) = \mathbb{E}[S_a(p(X))] = -\sum_i p(x_i) \log_a p(x_i)$$

En la práctica, las comparaciones de entropía es lo que nos interesa mas que la cantidad total y el factor constante arbitrario  $a$ , no es importante, ya que del hecho que

$$\log_a x = \frac{1}{\log_b a} \log_b x$$

obtenemos que, para dos variables aleatorias  $X$  e  $Y$ , se cumple que

$$\frac{H_a(X)}{H_a(Y)} = \frac{H_b(X)}{H_b(Y)}$$

Otra manera de mirar este factor constante, es pensar en la relación con la unidad de medida, es decir, la base del logaritmo sólo está especificando las unidades en la que estamos midiendo la entropía, entonces, podemos establecer que la unidad de medida es *bit*, *nat*, *dec*, etc., dependiendo de si es 2,  $e$ , 10, etc.

Necesitamos entonces, hablar de la cantidad de información de un suceso. Cuanto mayor sea la probabilidad de que se produzca, menor será la información que aporta. Si el suceso es de probabilidad uno, la información que nos aporta su conocimiento es cero. En el caso límite contrario, un suceso de probabilidad cero nos aportaría una información infinita.

Precisamente la función logaritmo tiene unas propiedades muy buenas para cuantificar este extremo:  $\log x$  vale cero para  $x = 1$ , y va aumentando (en valor absoluto) hacia infinito conforme la  $x$  va desde la unidad hacia el cero. Definiremos por tanto, la cantidad de información asociada a un suceso aleatorio de la siguiente manera:

**Definición 1.5.** Sea  $x_i$ , el  $i$ -ésimo suceso de una variable aleatoria  $\mathcal{X}$ , entonces la cantidad de información que aporta este suceso está determinada por la expresión

$$I_i = -\log_2 P(x_i)$$

donde  $P(x_i)$  es la probabilidad de este evento.

El motivo del signo menos es que el logaritmo de todo número comprendido entre 0 y 1 es negativo. La elección de la base 2 para los logaritmos es de índole práctica e irrelevante para la explicación del concepto. Podríamos en principio poner cualquier base; simplemente es una cuestión de escala.

**Definición 1.6.** Sea  $\mathcal{X}$  una variable aleatoria, que toma valores  $\{x_1, x_2, \dots, x_n\}$  con probabilidades  $p_1, p_2, \dots, p_n$ , que aportan cantidades de información  $I_1, I_2, \dots, I_n$ , llamamos *variable aleatoria cantidad de información asociada a  $\mathcal{X}$*  a la variable aleatoria  $I[\mathcal{X}]$  y consideremos su valor esperado, o esperanza  $E[I[\mathcal{X}]]$ ; número que denotaremos por  $H[\mathcal{X}]$ . A este valor se le denomina *Entropía de Shannon*.

Por tanto, el número real  $H[\mathcal{X}]$ , es el valor esperado de la cantidad de información que obtendremos al obtener un resultado del experimento expresado por dicha variable aleatoria. Por consiguiente, la Entropía de Shannon de la variable dada tiene una expresión analítica dada por:

$$H[\mathcal{X}] = E[I[\mathcal{X}]] = \sum_{i=1}^n p(x_i) I(x_i)$$

luego

$$H[\mathcal{X}] = -\sum_{i=1}^n p(x_i) \log_2(x_i)$$

Si consideramos entonces,  $\mathcal{X}$  una variable aleatoria en  $\mathbb{R}$ , con densidad de probabilidad  $f : \mathbb{R} \rightarrow \mathbb{R}$ , tenemos que la entropía de Shannon de  $\mathcal{X}$ , es justamente

$$-\int_{-\infty}^{\infty} f(x) \log f(x) dx$$

con la convención de que  $0 \log 0 = 0$ .

## 1.2 Otras definiciones de entropía

Supongamos entonces, que tenemos un sistema aislado, por tanto, no puede intercambiar energía con el resto del universo. Como un sistema de este tipo no tiene alrededores, solamente podemos concluir que la entropía del sistema aislado o bien aumenta o permanece constante. Es decir, la entropía aumenta al sufrir un cambio irreversible y cuando el sistema eventualmente alcanza el equilibrio, su entropía permanece constante dejando de aumentar.

Podemos por tanto, decir, que en un sistema aislado en el cual la energía es constante, la condición de equilibrio es tal que, la entropía alcanza su valor máximo por encima del cual no puede aumentar (para mas detalles ver [11]).

Entonces, desde el punto de vista de la termodinámica, se puede establecer un concepto más elaborado de la entropía, mediante las interpretaciones de las variaciones de la misma en función del comportamiento molecular, estableciendo las conexiones entre las propiedades macroscópicas y microscópicas del sistema.

Las propiedades macroscópicas de un sistema en equilibrio son constantes. Esto indica, que hay un número muy elevado de estados microscópicos asociados al estado macroscópico al que dan lugar. La entropía entonces, es una medida de cómo muchos estados microscópicos diferentes son compatibles con un particular estado macroscópico, donde estos estados macroscópicos tienen entropías muy grandes.

Podemos concluir entonces, que el estado de equilibrio de un sistema físico aislado está relacionado con un número mayor de estados microscópicos que los estados de no equilibrio. Por consiguiente, a la larga, el sistema alcanzará el estado de equilibrio, simplemente porque es más probable que los estados de no equilibrio.

Ahora bien, a través de los años la entropía desde el punto de vista definición, ha tenido variaciones importantes dependiendo del sistema que se estudie, entonces, consideremos algunas de estas definiciones para ilustrar de alguna manera el concepto establecido.

Tal es el caso, de la definición más sencilla en dinámica de gases, a saber

$$S = \frac{R}{N} \ln \Omega$$

donde  $R$  representa la constante de los gases,  $\Omega$  representa el número de estados microscópicos asociados al sistema y  $N$  es el número de Avogadro, que no es más que una constante física, que representa el número de moléculas existentes en una molécula gramo, ésta es igual para todas las sustancias, a saber,  $6.0235 \times 10^{23}$  moléculas/mol en la escala química y  $6.025 \times 10^{23}$  moléculas/mol en la escala física.

Otras definiciones que han tenido trascendencia en el mundo científico, son la entropía de Boltzmann y la de Gibbs.

### 1.2.1 La entropía de Boltzmann

**Ludwig Eduard Boltzmann** (1844-1906), fué un físico austríaco que trabajó en la teoría cinética de los gases (teoría física que explica el comportamiento y propiedades macroscópicas de los gases, a partir de una descripción estadística de los procesos moleculares microscópicos). Boltzmann, empleó las leyes de la estadística para estudiar el comportamiento de las moléculas de un gas, descubriendo la relación existente entre entropía y probabilidad, para ello consideró un sistema, el cual en un tiempo  $t$ , tiene un microestado dado por el vector  $\mathbf{x}(t)$  en el espacio de fase  $\Gamma$  (el tiempo y el espacio pueden ser tanto continuos como discretos).

Los macroestados (estados observables), están definidos por un conjunto  $\Lambda$  de variables macroscópicas (las cuales pueden incluir variables termodinámicas tales como: volumen, número de partículas, etc.-pueden también incluir otras variables especificando, por ejemplo, el número de partículas en un conjunto de subvolúmenes. Ridderbos (2002), denota a esto con un nombre colectivo: *Variables Supra-termodinámicas*).

Sea  $\{\mu\}_\Lambda$  el conjunto de macroestados, entonces, para cualquier  $\Gamma$ , existe un macroestado denotado por  $\mu(\mathbf{x})$  y la aplicación  $\mathbf{x} \mapsto \mu(\mathbf{x})$  es biunívoca. Cualquier macroestado  $\mu$ , está asociado a su volumen  $\nu(\mu) = \nu_\mu$  en  $\Gamma$  (cuando  $\Gamma$  es continuo normalmente  $\nu_\mu$  coincide con la medida de Lebesgue y si es discreto con el número de puntos o partículas de  $\mu$ ).

Observemos entonces la aplicación  $\mathbf{x} \mapsto \mu(\mathbf{x}) \mapsto \nu_\mu(\mathbf{x})$  de  $\Gamma$  a  $\mathbb{R}^+$  ó  $\mathbb{N}$ . La *entropía de Boltzmann* está definida por:

$$S_B(\mathbf{x}) = K_B \ln \left[ \frac{\nu_\mu(\mathbf{x})}{\nu_{\min}} \right]$$

donde  $\nu_{\min}$ , es el volumen del macroestado de mínimo volumen y  $K_B$  es la constante de Boltzmann (constante de los gases para moléculas aisladas). En la teoría cinética de gases,

aparece en la relación que encontró Boltzmann entre entropía y probabilidad y su valor es  $1.3803 \times 10^{-23}$ J/molécula-gramo. Luego  $S_B$ , es una función de fase que depende de la escogencia de las variables macroscópicas  $\Lambda$ .

## 1.2.2 La entropía de Gibbs

Supongamos que tenemos un sistema con  $N$  microsistemas, donde el espacio de fase se denota por  $\Gamma_N$ . La *entropía de Gibbs* está dada por el funcional

$$S_{\text{FGG}}[\rho_N(\mathbf{t})] = -K_B \int_{\Gamma_N} \rho_N(\mathbf{x}, t) \ln[\rho_N(\mathbf{x}, t)] d\Gamma_N$$

donde  $\rho_N(\mathbf{x}, t)$  es la función de densidad de  $\Gamma_N$ .

Para una medida que mantiene el sistema, en el cual  $\rho_N(\mathbf{x}, t)$  satisface la ecuación de Liouville,  $S_{\text{FGG}}[\rho_N(\mathbf{t})]$  (Fine-Grained Gibbs Entropy-por su nombre en inglés-) se queda constante en el tiempo.

La resolución a este problema sugerido por Gibbs en 1902, es considerar al espacio de fase  $\Gamma_N$ , como un “graneado” grueso (coarse-grain), en la forma que los macroestados son usualmente obtenidos en la aproximación de Boltzmann. [ver [8]]



## Capítulo 2

# Ecuación de Euler - Lagrange en un intervalo acotado

En este capítulo, nos enfocaremos en dejar establecidos todos los requerimientos necesarios para poder plantear con bases teóricas sólidas el problema fundamental del cálculo variacional. A saber, nuestro norte directo consiste en abordar el problema isoperimétrico pasando inicialmente por el problema en el caso de dimensión finita (que lo enfocaremos a manera de recordatorio), para luego, a partir de esta información, considerar en forma análoga el caso general o infinito dimensional.

### 2.1 Caso finito dimensional

Necesitamos establecer, las formas que existen de localización de valores extremos de una función definida sobre espacios de dimensión finita, sujeta o no a ciertas restricciones, este estudio lo haremos en forma muy ligera, basándonos en los extremos incondicionados y seguidamente los extremos condicionados.

#### 2.1.1 Extremos incondicionados

Sea  $E^n$  un espacio euclideo  $n$ -dimensional y sea  $D \subset E^n$  un conjunto abierto del espacio. Consideremos  $\mathbf{x} \in D$ , tal que  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ . Sea  $f(x_1, x_2, \dots, x_n)$  o simplemente  $f(\mathbf{x})$  una función definida en  $D$ .

Diremos que  $f$  alcanza su valor máximo en el punto  $\mathbf{x}^0 \in D$  si  $f(\mathbf{x}) \leq f(\mathbf{x}^0)$ , para  $\mathbf{x} \in D$ . Análogamente, se define el mínimo de la función  $f$ . Si  $D$  es un conjunto cerrado y acotado y

$f$  es una función continua, el teorema de Bolzano-Weierstrass nos garantiza la existencia de los valores máximo y mínimo de  $f$  en  $D$ .

**Definición 2.1.** Sea  $f$  una función definida en un abierto  $D \subset E^n$ . Se dice  $f$  alcanza un máximo estricto en  $\mathbf{x}^0 = (x_1^0, x_2^0, \dots, x_n^0) \in D$ , si existe un entorno  $\Omega$  de  $\mathbf{x}^0$ , tal que la desigualdad  $f(\mathbf{x}) < f(\mathbf{x}^0)$  se cumple para todo  $\mathbf{x} \in \Omega \cap D$ , con  $\mathbf{x} \neq \mathbf{x}^0$ . A saber,

$$\Delta f = f(\mathbf{x}) - f(\mathbf{x}^0) < 0$$

para todo  $\mathbf{x} \in \Omega \cap D$ . En forma análoga se define el mínimo estricto.

A los puntos donde  $f$  alcanza sus máximos y sus mínimos, se les denomina *puntos extremos* de  $f$ .

Consideremos entonces, a  $f$  definida en  $\Omega$ , tal que  $\mathbf{x}^0 \in \Omega$  es un punto extremo. Supongamos que  $f$  es diferenciable en  $\Omega$ , por lo tanto,  $\frac{\partial f}{\partial x_i}$  existe para todo  $i = 1, 2, \dots, n$ . Si  $\mathbf{x}^0$  es un punto extremo de  $f$ , entonces, se debe cumplir que  $\frac{\partial f}{\partial x_i}(\mathbf{x}^0) = 0$ , para todo  $i = 1, 2, \dots, n$ . A este resultado se le conoce como *condición necesaria de extremo*.

A los puntos que cumplen la condición necesaria de extremo, se les conoce como *puntos críticos* y a los puntos  $\mathbf{x}^0$  tales que  $df(\mathbf{x}^0) = 0$ , se les conoce como *puntos estacionarios* de la función  $f$ .

La condición  $df(\mathbf{x}^0) = 0$ , es equivalente a la condición

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^0) = 0, \quad \text{para todo } i = 1, 2, \dots, n$$

Es importante hacer notar que la existencia de un punto crítico no garantiza la existencia de un extremo, para resolver este problema veamos las condiciones suficientes de extremo estricto.

**Definición 2.2.** Se dice que la forma cuadrática

$$A(\mathbf{x}) = A(x_1, x_2, \dots, x_n) = \sum_{i,j=1}^n a_{ij}x_ix_j, \quad a_{ij} = a_{ji}, \quad i, j = 1, 2, \dots, n$$

es definida positiva si  $A(\mathbf{x}) > 0$ , para todo  $\mathbf{x} \in E^n$ ,  $\mathbf{x} \neq 0$  y se anula sólo cuando  $\mathbf{x} = 0$ , es decir  $x_i = 0$  para todo  $i = 1, 2, \dots, n$ . (Análogo para definida negativa)

Sea entonces  $f$  una función de clase  $\mathcal{C}^2$  en  $\Omega$  y sea  $\mathbf{x}^0 \in \Omega$  un punto estacionario de  $f$ . Si la forma cuadrática

$$A(dx_1, dx_2, \dots, dx_n) = \sum_{i,j=1}^n \frac{\partial^2 f(\mathbf{x}^0)}{\partial x_i \partial x_j} dx_i dx_j$$

es decir, la segunda diferencial de la función  $f$  en el punto  $\mathbf{x}^0$ , es definida positiva, el punto  $\mathbf{x}^0$  es punto de mínimo estricto (Análogo para el máximo estricto). Si la forma cuadrática es indefinida entonces  $\mathbf{x}^0$  no es un punto extremo de  $f$ . A estas condiciones se les conoce como *condiciones suficientes de extremo estricto*.

### 2.1.2 Extremos condicionados

Sea  $z = f(x_1, x_2, \dots, x_n)$  una función de  $n$  variables definida en un abierto  $D \subset E^n$ . Supongamos, además, que las  $n$  variables están conectadas por  $m$  condiciones complementarias, donde  $m < n$  y  $c_i \in \mathbb{R}$  para  $i = 1, \dots, m$ , a saber

$$\begin{cases} \varphi_1(x_1, x_2, \dots, x_n) = c_1 \\ \varphi_2(x_1, x_2, \dots, x_n) = c_2 \\ \vdots \\ \varphi_m(x_1, x_2, \dots, x_n) = c_m \end{cases}$$

que se denominan *ecuaciones de enlace*. Sea  $\mathbf{x}^0 = (x_1^0, x_2^0, \dots, x_n^0)$  un punto interior de  $D$ . Se dice que  $f$  tiene máximo condicionado (mínimo condicionado, respectivamente) en el punto  $\mathbf{x}^0$  si la desigualdad

$$f(\mathbf{x}) \leq f(\mathbf{x}^0) \quad \text{ó} \quad f(\mathbf{x}) \geq f(\mathbf{x}^0)$$

respectivamente, se cumple en un entorno de  $\mathbf{x}^0$ , siempre que los puntos  $\mathbf{x}$  y  $\mathbf{x}^0$  verifiquen las ecuaciones de enlace.

Nuevamente, consideremos el caso  $n = 2$  para ilustrar el proceso. Supongamos que se busca el extremo de la función  $z = f(x, y)$  sujeta a la restricción  $\varphi(x, y) = c$ . Supongamos que para los valores determinados de  $x$  e  $y$ , en virtud del **Teorema de la Función Implícita**, la ecuación de enlace  $\varphi(x, y) = c$  determina a  $y$  como una función diferenciable univocamente definida  $y = \psi(x)$ , entonces,  $z = f(x, \psi(x)) = F(x)$ , donde el extremo incondicionado de la función  $F$ , será justamente, el extremo buscado de la función  $f$  con la respectiva condición de enlace. Entonces, para resolver el problema original de hallar los extremos de la función  $z = f(x_1, x_2, \dots, x_n)$ , sujeta a las ecuaciones de enlace, lo que se hace es combinar el hecho de que  $\nabla f(\mathbf{x}) = 0$ , con el teorema de la función implícita obteniendo como consecuencia el *método de los multiplicadores de Lagrange*.

### 2.1.3 Método de los multiplicadores de Lagrange

**Teorema 2.3** (Lagrange). Sea  $D \subset E^n$  un abierto y sean  $f, \varphi_k : D \rightarrow \mathbb{R}, k = 1, \dots, m$ , con  $m < n$ , funciones suaves dadas. Sea  $\mathbf{x}^0 \in D$ , tal que  $\varphi_k(\mathbf{x}^0) = c_k$ , con  $c_k = \text{cte}$ , para cada  $k = 1, \dots, m$  y sea  $S = \{\mathbf{x} \in D : \varphi_k(\mathbf{x}) = c_k\}$ . Supongamos que  $\nabla \varphi_k(\mathbf{x}^0)$  son linealmente independientes, para  $k = 1, \dots, m$ . Si  $f|_S$ , que denota a  $f$  restringida a  $S$ , tiene un máximo o un mínimo en  $S$ , alcanzado en  $\mathbf{x}^0$ , entonces existen números reales  $\lambda_k, k = 1, \dots, m$  tales que

$$\nabla f(\mathbf{x}^0) = \sum_{k=1}^m \lambda_k \nabla \varphi_k(\mathbf{x}^0)$$

A estos valores  $\lambda_k$  se les conoce como los **multiplicadores de Lagrange**.

*Observación 2.4.* La demostración del Teorema de Lagrange, la haremos más adelante utilizando como herramienta un resultado más general.

Podemos considerar un par de condiciones que se deben satisfacer para hallar los extremos de funciones dadas. Para ilustrar un poco el método de los multiplicadores de Lagrange, observemos dichas condiciones y a partir de estas podremos esquematizar el proceso, veamos

1. Las derivadas parciales de las funciones  $f(x_1, x_2, \dots, x_n)$  y  $\varphi_i(x_1, x_2, \dots, x_n)$ , con  $i = 1, \dots, m$ , son continuas en  $D$ .
2.  $m < n$ , siendo el rango de la matriz  $\left(\frac{\partial \varphi_i}{\partial x_j}\right)$ , con  $i = 1, \dots, m$  y  $j = 1, \dots, n$ , igual a  $m$  en todo punto de  $D$ .

Consideremos entonces una nueva función

$$\Phi(\mathbf{x}) = f(\mathbf{x}) + \sum_{k=1}^m \lambda_k \varphi_k(\mathbf{x})$$

conocida como *función de Lagrange*, donde los  $\lambda_k$ , son los factores constantes indeterminados (multiplicadores de Lagrange).

Seguidamente, analizamos los extremos incondicionados de la función  $\Phi$ , es decir, formamos el sistema de ecuaciones (condiciones necesarias de extremo)

$$\frac{\partial \Phi}{\partial x_1} = 0, \quad \frac{\partial \Phi}{\partial x_2} = 0, \dots, \quad \frac{\partial \Phi}{\partial x_n} = 0 \tag{2.1}$$

y a partir de este sistema y de las  $m$  ecuaciones de enlace, determinamos los valores de los parámetros  $\lambda_1, \lambda_2, \dots, \lambda_m$  y las coordenadas  $(x_1, x_2, \dots, x_n)$  de los posibles puntos de extremo.

Debido a que las ecuaciones (2.1) son condiciones necesarias de extremo, tanto para la función  $\Phi$ , como para la función inicial  $f$ , tenemos que si  $\mathbf{x}^0$  es un punto extremo de  $f$ , será a la vez, un punto estacionario de  $\Phi$ , es decir, en este punto  $\frac{\partial \Phi}{\partial x_j} = 0$ , con  $j = 1, 2, \dots, n$ . Para analizar el punto estacionario  $\mathbf{x}^0$  como un extremo condicionado de la función  $\Phi$ , se tendría que considerar la forma cuadrática

$$B(dx_1, dx_2, \dots, dx_{n-m}) = \sum_{i,j=1}^{n-m} b_{ij} dx_i dx_j$$

y de las posibles definiciones de esta forma cuadrática encontraremos los posibles extremos estrictos de la función  $\Phi$ .

## 2.2 Caso general

En la sección anterior, se realizó un recordatorio de los extremos de una función  $f$  a valores reales, definida en un espacio euclideo de dimensión finita, como lo es en efecto,  $E^n$ . En esta sección, hacemos una entrada formal al cálculo variacional, teniendo como primera variante, el hecho de trabajar en  $\mathcal{C}[\mathbf{a}, \mathbf{b}]$  que es un espacio de dimensión infinita, por tanto, trabajaremos con funciones de funciones o *funcionales*. Es relevante hacer notar, que nuestros funcionales están definidos sobre trayectorias continuas, en consecuencia, las derivadas parciales no se toman como en el caso anterior con respecto a una variable independiente de un espacio euclideo, sino con respecto a una trayectoria regular. Para tratar de hacer un poco más cómoda la presentación, esta sección la dividiremos en dos partes fundamentales tales como el caso abstracto y el cálculo variacional propiamente dicho.

### 2.2.1 Extremos de funcionales

Supongamos que  $E$  es una variedad lineal contenida en  $\mathcal{C}[\mathbf{a}, \mathbf{b}]$ . Al igual que en el caso de dimensión finita, se definen máximos, mínimos, extremos, etc.

**Definición 2.5.** Sea  $(E, \|\cdot\|_E)$  un espacio vectorial normado y consideremos  $\mathcal{M} \subset E$ . Un *funcional*, es una función  $F : \mathcal{M} \rightarrow \mathbb{R}$ . El conjunto  $\mathcal{M}$  formado por las funciones  $\psi$  en la que está definido el funcional, se denomina *campo de definición del funcional*.

Se denomina *variación o incremento*  $\Delta\psi$  del argumento  $\psi(t)$  del funcional  $F(\psi)$ , a la diferencia entre dos funciones  $\psi(t)$  y  $\varphi(t)$  pertenecientes al subconjunto  $\mathcal{M}$ , a saber;

$$\Delta\psi = \psi(t) - \varphi(t)$$

Supongamos entonces, que las curvas  $\psi(t)$  y  $\varphi(t)$  están definidas en el intervalo  $[a, b]$ . De la definición del incremento del funcional se intuye el hecho de estudiar la proximidad entre las curvas de una cierta familia, es decir, diremos que las curvas  $\psi(t)$  y  $\varphi(t)$  tienen *proximidad de orden cero*, si sobre el intervalo  $[a, b]$ , la magnitud  $\|\psi(t) - \varphi(t)\|_E$  es pequeña desde el punto de vista geométrico, esto significa, que son próximas las ordenadas de las curvas sobre el conjunto.

Supongamos ahora, que las curvas  $\psi(t)$  y  $\varphi(t)$  son al menos de clase  $\mathcal{C}^1$  sobre el intervalo  $[a, b]$ . Entonces, las curvas  $\psi(t)$  y  $\varphi(t)$ , son próximas de orden 1, si son pequeñas sobre  $[a, b]$ , las magnitudes

$$\|\psi(t) - \varphi(t)\|_E, \quad \|\psi'(t) - \varphi'(t)\|_E \quad (2.2)$$

Denominaremos *distancia de orden k*, con  $k = 0, 1$ , a la mayor de las expresiones dadas en (2.2), entonces representamos dicha distancia de la siguiente manera

$$\rho_k = \rho_k[\psi(t), \varphi(t)] = \max_{0 \leq i \leq k} \max_{a \leq t \leq b} \|\psi^{(i)}(t) - \varphi^{(i)}(t)\|_E, \quad k = 0, 1$$

Se denomina  $\varepsilon$ -*vecindad* de orden k de la curva  $\psi(t)$ ,  $a \leq t \leq b$ , al conjunto de todas las curvas  $\varphi(t)$  cuyas distancias de k-ésimo orden a la curva  $\psi(t)$  son menores que  $\varepsilon$ . Ahora bien, la  $\varepsilon$ -vecindad de orden cero, se denomina  *$\varepsilon$ -vecindad fuerte* de  $\psi(t)$  y la  $\varepsilon$ -vecindad de primer orden se denomina  *$\varepsilon$ -vecindad débil* de la curva  $\psi(t)$ .

A partir de este resultado, podemos definir la continuidad de un funcional, en forma análoga a la continuidad de funciones de variable real, veamos

**Definición 2.6.** Un funcional  $F(\psi)$  definido en el conjunto  $\mathcal{M}$  de funciones  $\psi(t)$ , se dice continuo en  $\varphi(t)$  en el sentido de proximidad de k-ésimo orden, si para todo  $\varepsilon > 0$ , existe  $\eta > 0$  tal que se verifica que  $|F(\psi) - F(\varphi)| < \varepsilon$ , para todas las funciones  $\psi \in \mathcal{M}$  que satisfacen  $\rho_k[\psi(t), \varphi(t)] < \eta$ .

Supongamos entonces, que tenemos un funcional definido sobre el conjunto  $\mathcal{M}$ , la magnitud

$$\Delta F = \Delta F(\psi) = F(\psi + \Delta\psi) - F(\psi)$$

se denomina *incremento o variación del funcional*  $F(\psi)$ , correspondiente, al incremento  $\Delta\psi$  del argumento.

Otra forma de medir el incremento del funcional, puede ser, considerar la derivada del funcional  $F$  en la dirección de  $\Delta\psi$  en el punto  $\psi_0$ , esto es

$$\Delta F = \left. \frac{\partial}{\partial \alpha} F(\psi_0 + \alpha \Delta\psi) \right|_{\alpha=0}$$

**Definición 2.7.** Diremos que el funcional  $F$  alcanza su *máximo local* en la curva  $\varphi$ , si los valores que toma  $F(\psi)$  en cualquier curva próxima a  $\varphi$  (en algún orden), no son mayores que  $F(\varphi)$ , a saber

$$\Delta F = F(\psi(t)) - F(\varphi(t)) \leq 0$$

Si  $\Delta F = 0$ , sólo cuando  $\psi(t) = \varphi(t)$ , diremos que el funcional alcanza un máximo estricto en la curva  $\varphi$ . Análogamente se define el *mínimo*.

Por otra parte, diremos que el funcional  $F$  alcanza un *máximo local fuerte* en  $\psi$ , si se cumple la desigualdad anterior, para todas las curvas pertenecientes a una  $\varepsilon$ -vecindad de orden cero de la curva  $\varphi$ . Alcanzará un *máximo local débil*, si se cumple la misma desigualdad, para todas las curvas pertenecientes a una  $\varepsilon$ -vecindad de primer orden de la curva  $\varphi$ . Análogamente, se definen los mínimos relativos fuertes y débiles.

Todos estos extremos, se denominan *extremos locales* del funcional, y resulta relevante acotar, que todo extremo fuerte es a su vez un extremo débil. El extremo del funcional referente a la totalidad de las funciones en la que está definida la misma, se denomina *extremo absoluto*, y por supuesto, todo extremo absoluto es a su vez un extremo local fuerte.

Se necesita ahora, evaluar las condiciones para que exista el extremo de un funcional y para formalizar este hecho, comenzaremos por hacer ver que en esencia la definición de diferenciabilidad de funcionales entre espacios normados es la misma que sobre espacios euclideos, con la excepción, del requerimiento de que la aproximación se realice por un funcional lineal continuo, como veremos en la siguiente definición.

**Definición 2.8.** Sean  $(E, \|\cdot\|_E)$  y  $(H, \|\cdot\|_H)$  espacios normados. El funcional  $F : E \rightarrow H$  es diferenciable en  $x \in E$  si y sólo si existe un funcional lineal continuo  $L : E \rightarrow H$ , tal que

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x) - L(h)}{\|h\|_E} = 0 \quad (2.3)$$

Si  $F : E \rightarrow H$  es diferenciable en  $x \in E$ , entonces al funcional lineal continuo que satisface (2.3), es llamado la diferencial de  $F$  en  $x$  y se denota por  $dF_x : E \rightarrow H$ .

En el caso de dimensión finita, es decir,  $E = \mathbb{R}^n$  y  $H = \mathbb{R}^m$  esta diferencial coincide con la matriz formada por las derivadas parciales, cuando estas están expresadas en las respectivas bases canónicas, pero, estamos principalmente interesados en las correspondencias entre espacios de dimensión infinita, donde la diferencial de los funcionales lineales no tienen una representación matricial.

Hemos hablado en líneas anteriores de la continuidad de un funcional lineal, en el siguiente teorema dejamos claro a que hacemos referencia, con la intención básica de poder dejar las bases sentadas, para así, lograr determinar el espacio tangente haciendo uso de la regla de la cadena sobre espacios normados.

**Teorema 2.9.** *Sea  $L : E \rightarrow H$  un funcional lineal, donde  $(E, \|\cdot\|_E)$  y  $(H, \|\cdot\|_H)$  son espacios vectoriales normados. Entonces las siguientes condiciones sobre  $L$  son equivalentes*

1. *Existe un número  $c > 0$ , tal que  $\|L(v)\|_H \leq c\|v\|_E$ , para todo  $v \in E$ .*
2.  *$L$  es continuo en todas partes.*
3.  *$L$  es continuo en  $0 \in E$ .*

*Demostración.* Supongamos primero que (1) se cumple. Entonces, dado  $x_0 \in E$  y  $\varepsilon > 0$ , tenemos que

$$\begin{aligned} \|x - x_0\|_E < \frac{\varepsilon}{c} &\implies \|L(x) - L(x_0)\|_H = \|L(x - x_0)\|_H \\ &\leq c\|x - x_0\|_E \\ &< \varepsilon \end{aligned}$$

luego obtenemos que  $L$  es continuo en  $x_0$ , para cada  $x_0 \in E$ , por tanto, tenemos que (1)  $\implies$  (2).

Evidentemente, (2)  $\implies$  (3). Veamos entonces que (3)  $\implies$  (1). Supongamos que  $L$  es continuo en 0 y escogemos  $\delta > 0$  tal que  $\|x\|_E \leq \delta \implies \|L(x)\|_H < 1$ . Entonces, dado  $v \neq 0 \in E$ , obtenemos que

$$\begin{aligned} \|L(v)\|_H &= L\left(\frac{\|v\|_E}{\delta} \cdot \frac{\delta}{\|v\|_E} v\right) \\ &= \frac{\|v\|_E}{\delta} L\left(\frac{\delta}{\|v\|_E} v\right) \quad \text{tomando } x = \frac{\delta}{\|v\|_E} v \\ &< \frac{1}{\delta} \|v\|_E \end{aligned}$$

y tomando  $c = \frac{1}{\delta}$ , obtenemos la implicación deseada. □

El siguiente resultado, permite calcular el diferencial de la composición de dos funcionales en términos de los diferenciales de cada uno de los funcionales.



**Teorema 2.10.** *Sea  $\mathcal{U}$  y  $\mathcal{V}$  subconjuntos abiertos de los espacios normados  $(E, \|\cdot\|_E)$  y  $(H, \|\cdot\|_H)$  respectivamente. Si los funcionales  $F : \mathcal{U} \rightarrow H$  y  $g : \mathcal{U} \rightarrow G$  (un tercer espacio vectorial normado), son diferenciables en  $x \in \mathcal{U}$  y  $F(x) \in \mathcal{V}$  respectivamente, entonces la composición  $h = g \circ F$  es diferenciable en  $x$  y se cumple que*

$$dh_x = dg_{F(x)} \circ dF_x$$

*Observación 2.11.* Este resultado se da sin demostración, ya que esta es análoga a la que se da en el caso finito dimensional.

Consideremos un subconjunto  $\mathcal{M}$  de un espacio vectorial normado  $E$ , el *espacio tangente*  $TM_x$  de  $\mathcal{M}$  en el punto  $x \in \mathcal{M}$ , es el conjunto de todos aquellos vectores  $v \in E$ , para los cuales, existe una trayectoria diferenciable  $\varphi : \mathbb{R} \rightarrow \mathcal{M}$  tal que  $\varphi(0) = x$  y  $\varphi'(0) = v$ . Por consiguiente,  $TM_x$  es simplemente el conjunto de todos los vectores velocidad en  $x$  de trayectorias diferenciables en  $\mathcal{M}$  que pasan por  $x$ . Es importante hacer notar, que si  $\mathcal{M}$  es abierto, entonces  $TM_x = E$  para todo  $x \in \mathcal{M}$ .

**Teorema 2.12.** *Sea  $(E, \|\cdot\|_E)$  un espacio normado y  $F : E \rightarrow \mathbb{R}$  un funcional diferenciable y sea  $\mathcal{M}$  un subconjunto de  $E$  y consideremos  $x \in \mathcal{M}$ . Si  $F$  es diferenciable en  $x$  y  $F|_{\mathcal{M}}$  alcanza un mínimo en  $x$ , entonces*

$$dF_x|_{TM_x} = 0$$

*Demostración.* Dado  $v \in TM_x$ , sea  $\varphi : \mathbb{R} \rightarrow \mathcal{M}$  una trayectoria diferenciable en  $E$ , tal que,  $\varphi(0) = x$  y  $\varphi'(0) = v$ . Entonces la función  $g : \mathbb{R} \rightarrow \mathbb{R}$ , definida por  $g = F \circ \varphi$  tiene un mínimo local en 0, tal que  $g'(0) = 0$ . Entonces por la regla de la cadena nos queda que

$$0 = g'(0) = dg_0(1) = dF_{\varphi(0)}(d\varphi_0(1)) = dF_x(\varphi'(0)) = dF_x(v)$$

□

*Observación 2.13.* Si  $\mathcal{M}$  es abierto entonces se obtiene directamente que  $dF_x = 0$ . Si  $\mathcal{M}$  es un conjunto cualquiera, entonces, para hallar los extremos es necesario hallar el espacio tangente a  $\mathcal{M}$ .

Como un resultado particular del teorema anterior, podemos ahora presentar la demostración del Teorema de Lagrange con múltiples restricciones.

**Teorema 2.14** (Lagrange). *Sea  $D \subset E^n$  un abierto y sean  $f, \varphi_k : D \rightarrow \mathbb{R}$ ,  $k = 1, \dots, m$ , con  $m < n$ , funciones suaves dadas. Sea  $\mathbf{x}^0 \in D$ , tal que  $\varphi_k(\mathbf{x}^0) = c_k$ , con  $c_k = \text{cte}$ , para cada  $k = 1, \dots, m$  y sea  $S = \{\mathbf{x} \in D : \varphi_k(\mathbf{x}) = c_k\}$ . Supongamos que  $\nabla\varphi_k(\mathbf{x}^0)$  son linealmente independientes, para  $k = 1, \dots, m$ . Si  $f|_S$ , que denota a  $f$  restringida a  $S$ , tiene un máximo o un mínimo en  $S$ , alcanzado en  $\mathbf{x}^0$ , entonces existen números reales  $\lambda_k$ ,  $k = 1, \dots, m$  tales que*

$$\nabla f(\mathbf{x}^0) = \sum_{k=1}^m \lambda_k \nabla \varphi_k(\mathbf{x}^0)$$

*Demostración.* Sea  $G : D \rightarrow \mathbb{R}^m$ , la función definida por  $G(\mathbf{x}) = (\varphi_1(\mathbf{x}), \dots, \varphi_m(\mathbf{x}))$ , entonces  $G(\mathbf{x}^0) = (c_1, \dots, c_m)$  y  $S = \{\mathbf{x} \in D : G(\mathbf{x}) = C\}$ , donde  $C = (c_1, \dots, c_m)$ .

El espacio tangente a  $S$  en  $\mathbf{x}^0$ , es la intersección de los hiperplanos tangentes a cada una de las variedades  $n - 1$  dimensionales definidas por

$$\{\mathbf{x} \in D : \varphi_k(\mathbf{x}) = c_k\}, \quad k = 1, \dots, m$$

Por lo tanto,  $S_{\mathbf{x}^0}$ , el espacio tangente a  $S$ , es el complemento ortogonal del conjunto  $\{\nabla\varphi_1(\mathbf{x}^0), \dots, \nabla\varphi_m(\mathbf{x}^0)\}$ . Si  $f|_S$ , alcanza un máximo en  $\mathbf{x}^0$ , entonces  $df_{\mathbf{x}^0}|_{S_{\mathbf{x}^0}} = 0$ .

Como,  $df_{\mathbf{x}^0}(\mathbf{x}) = \nabla f(\mathbf{x}^0) \cdot \mathbf{x}$ , se tiene que,  $\nabla f(\mathbf{x}^0) \cdot \mathbf{x} = 0$ , siempre que  $\mathbf{x} \in S_{\mathbf{x}^0}$ , es decir,  $\mathbf{x} \in \{\nabla\varphi_1(\mathbf{x}^0), \dots, \nabla\varphi_m(\mathbf{x}^0)\}^\perp$ , por consiguiente,  $\nabla f(\mathbf{x}^0) \in \{\{\nabla\varphi_1(\mathbf{x}^0), \dots, \nabla\varphi_m(\mathbf{x}^0)\}^\perp\}^\perp$  y como  $\{\{\nabla\varphi_1(\mathbf{x}^0), \dots, \nabla\varphi_m(\mathbf{x}^0)\}^\perp\}^\perp$  es la variedad lineal generada por  $\nabla\varphi_1(\mathbf{x}^0), \dots, \nabla\varphi_m(\mathbf{x}^0)$ , se tiene que existen  $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ , tales que

$$\nabla f(\mathbf{x}^0) = \sum_{k=1}^m \lambda_k \nabla \varphi_k(\mathbf{x}^0)$$

□

Consideraremos a partir de este momento  $E = \mathcal{C}^1[\mathbf{a}, \mathbf{b}]$ , es decir, consideraremos al espacio normado  $(\mathcal{C}^1[\mathbf{a}, \mathbf{b}], \|\cdot\|_{\mathcal{C}^1})$ , donde

$$\|\varphi\|_{\mathcal{C}^1} = \max_{a \leq t \leq b} |\varphi(t)| + \max_{a \leq t \leq b} |\varphi'(t)|$$

Sea  $\mathcal{M}$  el subconjunto formado por todas aquellas funciones  $\psi \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}]$ , que satisfacen las condiciones de borde  $\psi(\mathbf{a}) = \alpha$  y  $\psi(\mathbf{b}) = \beta$ . Si  $F$  es diferenciable en  $\varphi \in \mathcal{M}$  y  $F|_{\mathcal{M}}$  tiene un extremo local en  $\varphi$ , entonces por el teorema anterior tenemos que

$$dF_\varphi|_{T\mathcal{M}_\varphi} = 0 \tag{2.4}$$

por tanto, decimos que  $\varphi \in \mathcal{M}$  es una *extremal* de  $F$  sobre  $\mathcal{M}$ , si se satisface la condición necesaria (2.4).

Para poder averiguar si una función dada  $\varphi \in \mathcal{M}$  es o no un extremo de  $F$  en  $\mathcal{M}$ , a saber, si se cumple o no que  $dF_\varphi|_{T\mathcal{M}_\varphi} = 0$ , debemos calcular explícitamente  $dF_\varphi$  y determinar el espacio tangente  $T\mathcal{M}_\varphi$ .

El último problema, resulta sencillo de resolver. Si consideramos una trayectoria fija  $\varphi_0 \in \mathcal{M}$ , entonces dada  $\varphi \in \mathcal{M}$ , la diferencia  $\varphi - \varphi_0$  es un elemento del subespacio de  $\mathcal{C}^1[\mathbf{a}, \mathbf{b}]$  consistentes de aquellas funciones de clase  $\mathcal{C}^1$  sobre  $[\mathbf{a}, \mathbf{b}]$  que se anulan en los bordes, a saber;

$$\mathcal{C}_0^1[\mathbf{a}, \mathbf{b}] = \{\psi \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}] : \psi(\mathbf{a}) = \psi(\mathbf{b}) = 0\}$$

Contrariamente, si  $\psi \in \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ , entonces  $\varphi_0 + \psi \in \mathcal{M}$ . Por tanto,  $\mathcal{M}$  es un hiperplano en  $\mathcal{C}^1[\mathbf{a}, \mathbf{b}]$ , concretamente, la traslación por  $\varphi_0$  de  $\mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ , pero el plano tangente de un hiperplano en cualquier punto, es simplemente, el subespacio del cual es una traslación, por consiguiente

$$T\mathcal{M}_\varphi = \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$$

para cualquier  $\varphi \in \mathcal{M}$ . En el siguiente teorema, observaremos el cálculo de  $dF_\varphi$ .

**Teorema 2.15.** *Sea  $F : \mathcal{C}^1[\mathbf{a}, \mathbf{b}] \rightarrow \mathbb{R}$ , definida por*

$$F(\varphi) = \int_{\mathbf{a}}^{\mathbf{b}} f(\varphi(t), \varphi'(t), t) dt$$

con  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  de clase  $\mathcal{C}^2$ . Entonces,  $F$  es diferenciable y

$$dF_\varphi(\mathbf{h}) = \int_{\mathbf{a}}^{\mathbf{b}} \left[ \frac{\partial f}{\partial \mathbf{x}}(\varphi(t), \varphi'(t), t)\mathbf{h}(t) + \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t)\mathbf{h}'(t) \right] dt \quad (2.5)$$

para toda  $\varphi, \mathbf{h} \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}]$ .

*Demostración.* Si  $F$  es diferenciable en  $\varphi \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}]$ , entonces  $dF_\varphi(\mathbf{h})$  debe ser la parte lineal (en  $\mathbf{h}$ ) de la diferencia  $F(\varphi + \mathbf{h}) - F(\varphi)$ . Para investigar sobre esta diferencia, utilizamos la expansión de Taylor de segundo orden de  $f$  en el punto  $(\varphi(t), \varphi'(t), t) \in \mathbb{R}^3$ , veamos

$$\begin{aligned} f(\varphi(t) + \mathbf{h}(t), \varphi'(t) + \mathbf{h}'(t), t) - f(\varphi(t), \varphi'(t), t) &= \frac{\partial f}{\partial \mathbf{x}}(\varphi(t), \varphi'(t), t)\mathbf{h}(t) \\ &+ \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t)\mathbf{h}'(t) + r(\mathbf{h}(t)) \end{aligned} \quad (2.6)$$

donde

$$r(\mathbf{h}(t)) = \frac{1}{2!} \left[ \frac{\partial^2 f}{\partial \mathbf{x}^2}(\xi(t))(\mathbf{h}(t))^2 + 2 \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{y}}(\xi(t))\mathbf{h}(t)\mathbf{h}'(t) + \frac{\partial^2 f}{\partial \mathbf{y}^2}(\xi(t))(\mathbf{h}'(t))^2 \right]$$

para algún punto  $\xi(t)$ , del segmento de recta en el espacio  $\mathbb{R}^3$  que va desde  $(\varphi(t), \varphi'(t), t)$  a  $(\varphi(t) + \mathbf{h}(t), \varphi'(t) + \mathbf{h}'(t), t)$ . Si  $\mathcal{B}$  es una bola de radio suficientemente grande tal que contiene a la imagen de la trayectoria continua  $t \mapsto (\varphi(t), \varphi'(t), t)$ ,  $t \in [\mathbf{a}, \mathbf{b}]$  y  $P$  es el máximo de los valores absolutos de las derivadas parciales de segundo orden de  $f$  en puntos de  $\mathcal{B}$ , entonces obtenemos que

$$|\mathbf{r}(\mathbf{h}(t))| \leq \frac{P}{2} (|\mathbf{h}(t)| + |\mathbf{h}'(t)|)^2 \quad (2.7)$$

para toda  $t \in [\mathbf{a}, \mathbf{b}]$ , si  $\|\mathbf{h}\|_{\mathcal{C}^1}$  es suficientemente pequeño.

De (2.6) obtenemos que

$$F(\varphi + \mathbf{h}) - F(\varphi) = L(\mathbf{h}) + R(\mathbf{h})$$

donde

$$L(\mathbf{h}) = \int_{\mathbf{a}}^{\mathbf{b}} \left[ \frac{\partial f}{\partial \mathbf{x}}(\varphi(t), \varphi'(t), t) \mathbf{h}(t) + \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t) \mathbf{h}'(t) \right] dt$$

y

$$R(\mathbf{h}) = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{r}(\mathbf{h}(t)) dt$$

Para demostrar que  $F$  es diferenciable en  $\varphi$ , con  $dF_{\varphi}(\mathbf{h}) = L(\mathbf{h})$ , como se quiere, es suficiente notar que  $L : \mathcal{C}^1[\mathbf{a}, \mathbf{b}] \rightarrow \mathbb{R}$  es una función lineal continua y luego verificar que

$$\lim_{\|\mathbf{h}\|_{\mathcal{C}^1} \rightarrow 0} \frac{|R(\mathbf{h})|}{\|\mathbf{h}\|_{\mathcal{C}^1}} = 0 \quad (2.8)$$

donde,  $\|\mathbf{h}\|_{\mathcal{C}^1} = \max_{\mathbf{a} \leq t \leq \mathbf{b}} \{|\mathbf{h}(t)|, |\mathbf{h}'(t)|\}$ .

De (2.7), se sigue inmediatamente que

$$\begin{aligned} |R(\mathbf{h})| &\leq \int_{\mathbf{a}}^{\mathbf{b}} |\mathbf{r}(\mathbf{h}(t))| dt \\ &\leq \int_{\mathbf{a}}^{\mathbf{b}} \frac{P}{2} (2\|\mathbf{h}\|_{\mathcal{C}^1})^2 dt = 2P(\mathbf{b} - \mathbf{a})(\|\mathbf{h}\|_{\mathcal{C}^1})^2 \end{aligned}$$

lo que demuestra (2.8). □

Este teorema establece claramente que para que  $F$  sea diferenciable, necesariamente, se debe satisfacer que  $\mathbf{b} - \mathbf{a} < \infty$ . A continuación, enunciaremos un par de resultados que nos serán de mucha ayuda más adelante, en vías de demostrar el teorema que nos garantiza las condiciones necesarias para encontrar extremales de funcionales.

**Corolario 2.16.** Sea  $F : \mathcal{C}^1[\mathbf{a}, \mathbf{b}] \longrightarrow \mathbb{R}$ , definida por

$$F(\varphi) = \int_{\mathbf{a}}^{\mathbf{b}} f(\varphi(t), \varphi'(t), t) dt$$

con  $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$  de clase  $\mathcal{C}^2$ . Si  $\varphi$  es de clase  $\mathcal{C}^2$  en  $[\mathbf{a}, \mathbf{b}]$  y  $\mathbf{h} \in \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ , entonces

$$dF_{\varphi}(\mathbf{h}) = \int_{\mathbf{a}}^{\mathbf{b}} \left[ \frac{\partial f}{\partial \mathbf{x}}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t) \right] \mathbf{h}(t) dt \quad (2.9)$$

*Demostración.* Si  $\varphi$  es una función de clase  $\mathcal{C}^2$ , entonces  $\frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t)$  es una función de clase  $\mathcal{C}^1$  con  $t \in [\mathbf{a}, \mathbf{b}]$ . Integrando por partes, obtenemos

$$\begin{aligned} \int_{\mathbf{a}}^{\mathbf{b}} \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t) \mathbf{h}'(t) dt &= \left[ \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t) \mathbf{h}(t) \right]_{\mathbf{a}}^{\mathbf{b}} \\ &\quad - \int_{\mathbf{a}}^{\mathbf{b}} \frac{d}{dt} \left( \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t) \right) \mathbf{h}(t) dt \\ &= - \int_{\mathbf{a}}^{\mathbf{b}} \frac{d}{dt} \left( \frac{\partial f}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t) \right) \mathbf{h}(t) dt \end{aligned}$$

ya que  $\mathbf{h}(\mathbf{a}) = 0 = \mathbf{h}(\mathbf{b})$ , por consiguiente (2.9) se obtiene directamente de la fórmula (2.5).  $\square$

A partir de este momento,  $\mathcal{M}_0$  denotará una variedad lineal contenida en el conjunto

$$\{\psi \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}] : \psi(\mathbf{a}) = 0 = \psi(\mathbf{b})\}$$

**Lema 2.17.** Si  $\mathcal{M}_0$  es denso en  $L^1[\mathbf{a}, \mathbf{b}]$  y  $\varphi : [\mathbf{a}, \mathbf{b}] \longrightarrow \mathbb{R}$  es una función continua tal que

$$\int_{\mathbf{a}}^{\mathbf{b}} \varphi(t) \mathbf{h}(t) dt = 0$$

para cualquier  $\mathbf{h} \in \mathcal{M}_0$ , entonces,  $\varphi$  es idénticamente cero en  $[\mathbf{a}, \mathbf{b}]$ .

*Demostración.* Consideremos inicialmente el caso en que  $\mathcal{M}_0 = \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ . Supongamos que  $\varphi(t_0) \neq 0$ , para algún  $t_0 \in [\mathbf{a}, \mathbf{b}]$ . Entonces, como  $\varphi$  es continua, tenemos que  $\varphi$  no se anula en un intervalo que contiene a  $t_0$ . Supongamos entonces que  $\varphi(t) > 0$ , para  $t \in [t_1, t_2] \subset [\mathbf{a}, \mathbf{b}]$ . Si definimos  $\mathbf{h}$  en  $[\mathbf{a}, \mathbf{b}]$  como sigue

$$\mathbf{h}(t) = \begin{cases} (t - t_1)^2(t - t_2)^2 & \text{si } t \in [t_1, t_2], \\ 0 & \text{en otro caso.} \end{cases}$$

entonces  $h \in \mathcal{C}_0^1[a, b]$  y además

$$\int_a^b \varphi(t)h(t) dt = \int_{t_1}^{t_2} \varphi(t)(t - t_1)^2(t - t_2)^2 dt > 0$$

ya que el integrando es positivo salvo en los extremos del intervalo  $[t_1, t_2]$ , lo cual es una contradicción. Por consiguiente,  $\varphi \equiv 0$  en  $[a, b]$ .

Si  $\mathcal{M}_0$  es una variedad lineal contenida en

$$\{\psi \in \mathcal{C}^1[a, b] : \psi(a) = 0 = \psi(b)\}$$

como  $\mathcal{M}_0$  es denso en  $L^1[a, b]$ , obtendremos que existe una sucesión de funciones  $h_n \in \mathcal{M}_0$ , tales que  $h_n \rightarrow \varphi$  en  $L^1[a, b]$ , entonces

$$\int \varphi h_n \rightarrow \int \varphi^2$$

lo que indica que  $\varphi = 0$  en c.t.p, ya que  $\int \varphi h_n = 0$ , para todo  $n$ . □

### 2.2.2 Ecuación de Euler

Supongamos que se tiene una función  $f(x, y, t)$  con derivadas parciales continuas hasta de segundo orden, con respecto a todos sus argumentos. El problema elemental del cálculo variacional consiste en lo siguiente: entre todas las posibles funciones  $\varphi(t)$  que poseen derivada continua, se debe hallar la función que ofrece extremo débil al funcional

$$F(\psi) = \int_a^b f(\psi(t), \psi'(t), t) dt \tag{2.10}$$

sujeta a las condiciones de frontera

$$\psi(a) = \delta \quad y \quad \psi(b) = \gamma \tag{2.11}$$

Nuevamente, resulta necesario establecer las condiciones bajo las cuales es posible hallar las funciones extremales del funcional definido en (2.10). Supongamos que  $\mathcal{M}_1$  está contenido en el conjunto

$$\{\psi \in \mathcal{C}^1[a, b] : \psi(a) = \delta \quad y \quad \psi(b) = \gamma \}$$

y que existe una variedad lineal  $\mathcal{M}_0 \subset \mathcal{C}_0^1[a, b]$ , tal que  $\mathcal{M}_0 + \mathcal{M}_1 \subset \mathcal{M}_1$  y  $\mathcal{M}_0$  es denso en  $L^1[a, b]$ . En este caso el espacio tangente a  $\mathcal{M}_1$  en cualquier punto contiene a  $\mathcal{M}_0$ .

El siguiente teorema nos permite establecer la condición necesaria para encontrar extremales.

**Teorema 2.18** (Condición necesaria). *Sea*

$$\mathcal{M}_1 \subset \{\psi \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}] : \psi(\mathbf{a}) = \delta \quad y \quad \psi(\mathbf{b}) = \gamma \}$$

y  $\mathcal{M}_0 \subset \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ , tal que,  $\mathcal{M}_0 + \mathcal{M}_1 \subset \mathcal{M}_1$  y  $\mathcal{M}_0$  es denso en  $L^1[\mathbf{a}, \mathbf{b}]$ . Para que el funcional

$$F(\psi) = \int_{\mathbf{a}}^{\mathbf{b}} f(\psi(t), \psi'(t), t) dt \quad (2.12)$$

definido en el conjunto de todas las funciones  $\psi \in \mathcal{M}_1$ , alcance su valor máximo, es necesario que la función  $\psi(t)$  verifique la ecuación de Euler

$$f_x - \frac{d}{dt} f_y = 0 \quad (2.13)$$

*Demostración.* Sea  $F$  un funcional definido como en (2.12) y consideremos  $h \in \mathcal{M}_0$ , donde  $\mathcal{M}_0$  es una variedad lineal contenida en  $\mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ . En virtud del Corolario 2.16, sabemos que

$$dF_{\varphi}(h) = \int_{\mathbf{a}}^{\mathbf{b}} \left[ \frac{\partial f}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial f}{\partial y}(\varphi(t), \varphi'(t), t) \right] h(t) dt$$

y como  $\mathcal{TM}_1 \supset \mathcal{M}_0$  por el Teorema 2.12, obtenemos que

$$\int_{\mathbf{a}}^{\mathbf{b}} \left[ \frac{\partial f}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial f}{\partial y}(\varphi(t), \varphi'(t), t) \right] h(t) dt = 0 \quad (2.14)$$

para toda función  $h \in \mathcal{M}_0$ . Pero en virtud del Lema 2.17, la ecuación (2.14) implica que

$$f_x - \frac{d}{dt} f_y = 0$$

□

*Observación 2.19.* Las curvas integrales que verifican la ecuación (2.13), se llaman *curvas extremales o curvas de Lagrange*, de allí que esta ecuación usualmente sea llamada ecuación de Euler-Lagrange.

Si desarrollamos la ecuación (2.13), da como resultado

$$\psi''(t) f_{yy}(\psi(t), \psi'(t), t) + \psi'(t) f_{xy}(\psi(t), \psi'(t), t) + f_{ty}(\psi(t), \psi'(t), t) - f_x(\psi(t), \psi'(t), t) = 0 \quad (2.15)$$

la cual, representa una ecuación diferencial de segundo orden, de modo que su solución general comprenderá dos constantes arbitrarias, cuyos valores se determinan a partir de las condiciones de frontera  $\psi(\mathbf{a}) = \delta$  y  $\psi(\mathbf{b}) = \gamma$ . El desarrollo dado en (2.15), es posible en virtud del siguiente teorema.

**Teorema 2.20.** *Sea  $\psi(t)$  una solución de la Ecuación de Euler-Lagrange. Si la función  $f(\psi(t), \psi'(t), t)$  tiene derivadas parciales continuas hasta de segundo orden inclusive, entonces la función  $\psi(t)$  tiene segunda derivada continua en todos los puntos para los cuales*

$$f_{yy}[\psi(t), \psi'(t), t] \neq 0$$

*Demostración.* Consideremos la diferencia

$$\begin{aligned} \Delta f_y &= f_y(x + \Delta x, y + \Delta y, t + \Delta t) - f_y(x, y, t) \\ &= \Delta t \bar{f}_{yt} + \Delta x \bar{f}_{yx} + \Delta y \bar{f}_{yy} \end{aligned}$$

donde las “barras” indican que las correspondientes derivadas son evaluadas en ciertas curvas intermedias. Si dividimos esta diferencia por  $\Delta t$  y consideremos el límite de la expresión

$$\bar{f}_{yt} + \frac{\Delta x}{\Delta t} \bar{f}_{yx} + \frac{\Delta y}{\Delta t} \bar{f}_{yy}$$

cuando  $\Delta t \rightarrow 0$ . Este límite existe, ya que  $f_y$  tiene derivada con respecto a  $t$ , la cual de acuerdo con la ecuación de Euler es igual a  $f_x$ . Además, por hipótesis las segundas derivadas de  $f(x, y, t)$  son continuas, entonces como  $\Delta t \rightarrow 0$ ,  $\bar{f}_{yt}$  converge a  $f_{yt}$ . Se sigue de la existencia de  $y$  y la continuidad de la segunda derivada  $f_{yy}$  que el segundo término también tiene límite cuando  $\Delta t \rightarrow 0$ . Pero entonces el tercer término también tiene límite, ya que el límite de la suma de los tres existe, a saber, el límite

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t} \bar{f}_{yy}$$

existe. Cuando  $\Delta t \rightarrow 0$ ,  $\bar{f}_{yy}$  converge a  $f_{yy} \neq 0$  y en consecuencia

$$\lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t} = y'(t)$$

existe. Finalmente a partir de la ecuación de Euler podemos encontrar una expresión para  $y'$ , en la cual, es claro que  $y'$  es continua cuando  $f_{yy} \neq 0$ .  $\square$

### 2.2.3 El problema isoperimétrico

Trataremos de ilustrar el problema isoperimétrico con un ejemplo.

En el siglo IX antes de Cristo, llega a las tierras del Norte de África (Túnez) la princesa fenicia Dido, huyendo de su hermano Pigmalión, el cual había asesinado a su marido. Al querer la princesa Dido comprar tierra para establecerse con su pueblo, el rey de aquellas



tierras solamente le consiente comprar la parcela de tierra que pueda ser aislada usando la piel de un toro. Dido cortó la piel en finas tiras formando una larga cuerda (de unos 1000 ó 2000 metros) y la dispuso de manera que cubriese la mayor parte de terreno posible...

Dido resolvió el siguiente problema: “Encontrar, entre todas las curvas cerradas de longitud fija, aquella que delimita la superficie más grande”.

Si consideremos dos puntos  $A = (a, 0)$  y  $B = (b, 0)$  en el eje  $x$ , donde la distancia entre ellos está dada. Es decir  $d(A, B) = l$ . El problema de hallar una curva que maximice el área entre ella y el eje  $x$  sería:

Hallar una función  $f(x)$  de modo que,

$$I[f] = \int_a^b f(x) dx = \max$$

con la restricción

$$G[f] = \int_a^b \sqrt{1 + [f'(x)]^2} dx = l \quad (\text{longitud de arco})$$

con  $f(a) = f(b) = 0$ .

El problema isoperimétrico, desde un enfoque general, consiste en conseguir las funciones extremales de la funcional dada en (2.10), sujeta a las condiciones de frontera de (2.11) y con la restricción

$$G(\psi) = \int_a^b g(\psi(t), \psi'(t), t) dt = c, \quad c \in \mathbb{R} \quad (2.16)$$

donde  $f, g : \mathbb{R}^3 \rightarrow \mathbb{R}$ , son funciones de clase  $\mathcal{C}^2[a, b]$ . Sea  $\mathcal{M}$  el hiperplano en  $\mathcal{C}^1[a, b]$  que contiene a todas aquellas funciones  $\psi : [a, b] \rightarrow \mathbb{R}$  de clase  $\mathcal{C}^1$ , tales que  $\psi(a) = \delta$  y  $\psi(b) = \gamma$ , entonces nuestro problema consiste en localizar los extremos locales de la funcional  $F$  sobre el conjunto  $\mathcal{M} \cap G^{-1}(c)$ . La idea es generalizar el método de los multiplicadores de Lagrange a espacios de dimensión infinita.

Consideremos el siguiente caso más simple: Sean  $F, G : \mathbb{R}^n \rightarrow \mathbb{R}$  de clase  $\mathcal{C}^1$ , tales que  $G(\mathbf{0}) = 0$  y  $\nabla G(\mathbf{0}) \neq \mathbf{0}$ . Si  $F$  tiene un mínimo o un máximo local en  $\mathbf{0}$ , sujeto a la restricción  $G(\mathbf{x}) = 0$ , entonces debe existir un número  $\lambda$  tal que

$$\nabla F(\mathbf{0}) = \lambda \nabla G(\mathbf{0}) \quad (2.17)$$

Por tanto, los diferenciales  $dF_{\mathbf{0}}, dG_{\mathbf{0}} : \mathbb{R}^n \rightarrow \mathbb{R}$ , están dados por

$$dF_{\mathbf{0}}(\mathbf{v}) = \nabla F(\mathbf{0}) \cdot \mathbf{v} \quad \text{y} \quad dG_{\mathbf{0}}(\mathbf{v}) = \nabla G(\mathbf{0}) \cdot \mathbf{v}$$

luego la ecuación (2.17), se puede reescribir así

$$dF_0 = \Lambda \circ dG_0 \quad (2.18)$$

donde  $\Lambda : \mathbb{R} \rightarrow \mathbb{R}$  es la función lineal definida por  $\Lambda(t) = \lambda t$ .

El siguiente resultado general permite justificar la ecuación (2.18) y será útil para generalizar el Teorema de Lagrange.,

**Lema 2.21.** Sean  $\alpha$  y  $\beta$  dos funciones lineales a valores reales en el espacio vectorial  $E$ , tales que

$$\mathbf{Ker} \alpha \supset \mathbf{Ker} \beta \quad \text{y} \quad \mathbf{Imagen} \beta = \mathbb{R}$$

Entonces existe  $\lambda \in \mathbb{R}$ , tal que  $\alpha = \lambda\beta$ .

Es decir, existe una función lineal  $\Lambda : \mathbb{R} \rightarrow \mathbb{R}$ , tal que  $\alpha = \Lambda \circ \beta$ , que es lo mismo que decir que el diagrama

$$\begin{array}{ccc} & \mathbb{R} & \\ \beta \nearrow & & \searrow \Lambda \\ E & \xrightarrow{\alpha} & \mathbb{R} \end{array}$$

conmuta.

*Demostración.* Dado  $t \in \mathbb{R}$ , tomamos  $x \in E$  tal que  $\beta(x) = t$  y definimos

$$\Lambda(t) = \alpha(x)$$

Para demostrar que  $\Lambda$  está bien definido, debemos ver que si  $y \in E$ , tal que  $x \neq y$ , con la condición que  $\beta(y) = t$ , entonces  $\alpha(x) = \alpha(y)$ . Pero si  $\beta(x) = \beta(y) = t$ , entonces  $x - y \in \mathbf{Ker} \beta \subset \mathbf{Ker} \alpha$ , es decir,  $\alpha(x - y) = 0$ , lo cual de forma inmediata implica que  $\alpha(x) = \alpha(y)$ .

Si  $\beta(x) = s$  y  $\beta(y) = t$ , entonces

$$\begin{aligned} \Lambda(as + bt) &= \alpha(ax + by) \\ &= a\alpha(x) + b\alpha(y) \\ &= a\Lambda(s) + b\Lambda(t) \end{aligned}$$

por consiguiente,  $\Lambda$  es lineal. □

Sean  $E, F$  y  $G$ , tres espacios vectoriales normados completos. Consideremos la aplicación  $f : E \times F \rightarrow G$  diferenciable, entonces, para cada  $\mathbf{a} \in E$  y cada  $\mathbf{b} \in F$ , las aplicaciones

$$\varphi : E \rightarrow G \quad \text{y} \quad \psi : F \rightarrow G$$

definidas por  $\varphi(\mathbf{x}) = f(\mathbf{x}, \mathbf{b})$  y  $\psi(\mathbf{y}) = f(\mathbf{a}, \mathbf{y})$ , son diferenciables en  $\mathbf{a} \in E$  y  $\mathbf{b} \in F$ , respectivamente. Entonces, las diferenciales parciales  $\mathbf{d}_x f_{(\mathbf{a}, \mathbf{b})}$  y  $\mathbf{d}_y f_{(\mathbf{a}, \mathbf{b})}$  de  $f$  en  $(\mathbf{a}, \mathbf{b})$ , con respecto a  $\mathbf{x} \in E$  y  $\mathbf{y} \in F$  respectivamente, se definen por

$$\mathbf{d}_x f_{(\mathbf{a}, \mathbf{b})} = \mathbf{d}\varphi_{\mathbf{a}} \quad \text{y} \quad \mathbf{d}_y f_{(\mathbf{a}, \mathbf{b})} = \mathbf{d}\psi_{\mathbf{b}}$$

Por consiguiente,  $\mathbf{d}_x f_{(\mathbf{a}, \mathbf{b})}$  es el diferencial de la aplicación de  $E$  en  $G$ , obtenida de la aplicación  $f : E \times F \rightarrow G$ , fijando  $\mathbf{y} = \mathbf{b}$  y análogamente,  $\mathbf{d}_y f_{(\mathbf{a}, \mathbf{b})}$  es el diferencial de la aplicación de  $F$  en  $G$ , obtenida de la aplicación  $f : E \times F \rightarrow G$ , fijando  $\mathbf{x} = \mathbf{a}$ .

**Definición 2.22.** Sean  $E$  y  $F$ , espacios vectoriales normados. La aplicación  $g : E \rightarrow F$ , se dice continuamente diferenciable o de clase  $\mathcal{C}^1$ , si es diferenciable y  $\mathbf{d}g_{\mathbf{x}}(\mathbf{v})$  es una función continua de  $(\mathbf{x}, \mathbf{v})$ , es decir, la aplicación  $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{d}g_{\mathbf{x}}(\mathbf{v})$  de  $E \times E$  en  $F$  es continua.

*Observación 2.23.* El siguiente resultado, corresponde con la versión del teorema de la función implícita sobre espacios normados, el cual presentaremos sin demostración. Para más detalles, se puede consultar [3].

**Teorema 2.24** (Teorema de la función implícita). Sean  $E, F$  y  $G$  tres espacios vectoriales normados completos y sea  $f : E \times F \rightarrow G$  una aplicación de clase  $\mathcal{C}^1$ . Supongamos que  $f(\mathbf{a}, \mathbf{b}) = \mathbf{0}$  y que

$$\mathbf{d}_y f_{(\mathbf{a}, \mathbf{b})} : F \rightarrow G$$

es un isomorfismo. Entonces, existe un entorno  $\mathcal{U}$  de  $\mathbf{a}$  en  $E$  y un entorno  $\mathcal{W}$  de  $(\mathbf{a}, \mathbf{b})$  en  $E \times F$  y una aplicación  $\varphi : \mathcal{U} \rightarrow F$  de clase  $\mathcal{C}^1$ , tal que: Si  $(\mathbf{x}, \mathbf{y}) \in \mathcal{W}$  y  $\mathbf{x} \in \mathcal{U}$ , entonces  $f(\mathbf{x}, \mathbf{y}) = \mathbf{0}$  si y sólo si  $\mathbf{y} = \varphi(\mathbf{x})$ .

**Teorema 2.25.** Sean  $F$  y  $G$  funciones a valores reales de clase  $\mathcal{C}^1$ , definidas en un espacio vectorial normado completo  $E$ , con  $G(\mathbf{0}) = 0$  y  $\mathbf{d}G_{\mathbf{0}} \neq 0$ . Si  $F : E \rightarrow \mathbb{R}$  tiene un extremo local en  $\mathbf{0}$ , sujeto a la restricción  $G(\mathbf{x}) = 0$ , entonces existe una función lineal  $\Lambda : \mathbb{R} \rightarrow \mathbb{R}$  tal que se satisface la ecuación  $\mathbf{d}F_{\mathbf{0}} = \Lambda \circ \mathbf{d}G_{\mathbf{0}}$ .

*Demostración.* Podremos utilizar el resultado del Lema 2.21, con  $\alpha = dF_0$  y  $\beta = dG_0$ , si logramos demostrar que  $\mathbf{Ker} dF_0$  contiene a  $\mathbf{Ker} dG_0$ . Supongamos que, dado  $v \in \mathbf{Ker} dG_0$ , existe una trayectoria diferenciable  $\gamma : (-\varepsilon, \varepsilon) \rightarrow E$ , cuya imagen está contenida en  $G^{-1}(0)$ , tal que,  $\gamma(0) = 0$  y  $\gamma'(0) = v$  (esta existencia la justificaremos luego).

Entonces la composición  $h = F \circ \gamma : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$  tiene un extremo local en 0, donde  $h'(0) = 0$ , entonces la regla de la cadena nos proporciona el siguiente hecho

$$0 = h'(0) = dh_0(1) = dF_{\gamma(0)}(d\gamma_0(1)) = dF_0(\gamma'(0)) = dF_0(v)$$

como se quería demostrar. □

*Observación 2.26.* Para justificar la existencia de la trayectoria diferenciable  $\gamma$ , usaremos el Teorema de la función implícita (Teorema 2.24).

Si  $\mathcal{X} = \mathbf{Ker} dG_0$ , entonces,  $dG_0 : E \rightarrow \mathbb{R}$  es una función continua y  $\mathcal{X}$  es un subespacio cerrado de  $E$  y por consiguiente completo. Escogiendo  $\omega \in E$ , de tal manera que,  $dG_0(\omega) = 1$  y denotando por  $\mathcal{Y}$  al subespacio cerrado de  $E$ , que contiene a todos los múltiplos escalares de  $\omega$ , podemos afirmar que  $\mathcal{Y}$  es una copia de  $\mathbb{R}$ .

Por otra parte,  $\mathcal{X} \cap \mathcal{Y} = 0$  y si  $e \in E$  y  $\alpha \in dG_0(e) \in \mathbb{R}$ , entonces

$$dG_0(e - \alpha\omega) = dG_0(e) - \alpha dG_0(\omega) = 0$$

por tanto,  $e - \alpha\omega \in \mathcal{X}$  y en consecuencia,  $e = x + y$ , donde  $x \in \mathcal{X}$  y  $y = \alpha\omega \in \mathcal{Y}$ . Por consiguiente,  $E$  es la suma directa algebraica de los subespacios  $\mathcal{X}$  y  $\mathcal{Y}$ , más aún, la norma sobre  $E$  es equivalente a la norma producto sobre  $\mathcal{X} \times \mathcal{Y}$ , así que podemos escribir  $E = \mathcal{X} \times \mathcal{Y}$ .

Para poder aplicar el Teorema de función implícita, necesitamos ver que  $d_y G_0 : \mathcal{Y} \rightarrow \mathbb{R}$ , es un isomorfismo. De el hecho que  $\mathcal{Y} \approx \mathbb{R}$ , debemos demostrar que  $d_y G_0 \neq 0$ . Pero, dado  $(r, s) \in \mathcal{X} \times \mathcal{Y} = \mathcal{E}$ , obtenemos que

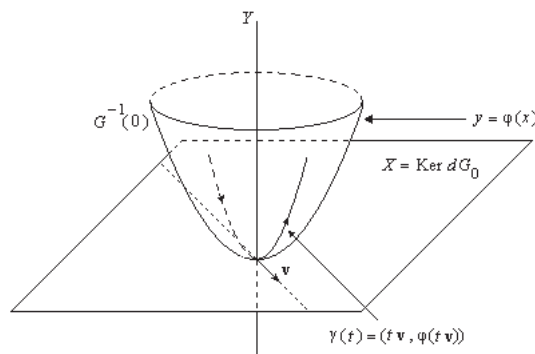
$$dG_0(r, s) = dG_0(r, 0) + dG_0(0, s) = dG_0(0, s) = d_y G_0(s)$$

Así que si se supone que  $d_y G_0 = 0$ , obtendremos que  $dG_0 = 0$ , lo cual contradice las hipótesis del teorema.

Consecuentemente, el Teorema de la función implícita proporciona una función diferenciable  $\varphi : \mathcal{X} \rightarrow \mathcal{Y}$ , cuyo gráfico  $y = \varphi(x)$  en  $\mathcal{X} \times \mathcal{Y} = \mathcal{E}$ , coincide con  $G^{-1}(0)$ , en algún entorno de 0. Si  $H(x) = G(x, \varphi(x))$ , entonces  $H(x) = 0$ , para  $x$  cercano a 0, por eso

$$0 = dH_0(u) = d_x G_0(u) + d_y G_0(d\varphi_0(u)) = d_y G_0(d\varphi_0(u))$$

para todo  $u \in \mathcal{X}$ , por tanto se obtiene que  $d\varphi_0 = 0$ , ya que  $d_y G_0$  es un isomorfismo.



Finalmente, dado  $v = (u, 0) \in \mathbf{Ker} \, dG_0$ , definimos  $\gamma : \mathbb{R} \longrightarrow E$  por  $\gamma(t) = (tu, \varphi(tu))$ . Entonces  $\gamma(0) = 0$  y  $\gamma(t) \in G^{-1}(0)$ , para  $t$  suficientemente pequeño y además

$$\gamma'(0) = (u, d\varphi_0(u)) = (u, 0) = v$$

Lo anteriormente expuesto, conjuntamente con el teorema de la función implícita, nos garantiza el siguiente resultado para el problema isoperimétrico.

**Teorema 2.27.** Sean  $F$ ,  $G_1$  y  $G_2$ , los funcionales sobre  $\mathcal{C}^1[a, b]$  definidos por

$$F(\psi) = \int_a^b f(\psi(t), \psi'(t), t) \, dt$$

y

$$G_i(\psi) = \int_a^b g_i(\psi(t), \psi'(t), t) \, dt - c_i$$

donde  $i = 1, 2$  y además  $f$  y  $g_i$  son funciones de clase  $\mathcal{C}^2$  en  $\mathbb{R}^3$ , para  $i = 1, 2$ . Sea  $\varphi \in \mathcal{M}$  una función de clase  $\mathcal{C}^2$  que no es una extremal de  $G_i$ . Si  $F$  tiene un extremo local en  $\varphi$  sujeto a las condiciones

$$\psi(a) = \alpha, \quad \psi(b) = \beta, \quad \text{y} \quad G_i(\psi) = 0$$

entonces existen números reales  $\lambda_1$  y  $\lambda_2$  tales que  $\varphi$  satisface la ecuación de Euler - Lagrange para la función  $h = f - \lambda_1 g_1 - \lambda_2 g_2$ , a saber,

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0$$

para todo  $t \in [a, b]$ .

*Demostración.* Supongamos que  $\varphi \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}]$  y es de clase  $\mathcal{C}^2$ , tal que F tiene un extremo local sobre  $\mathcal{M} \cap [\mathbf{G}_1^{-1}(0) \cap \mathbf{G}_2^{-1}(0)]$ , pero ahora, consideremos las funciones a valores reales  $F \circ \mathbf{T}$ ,  $\mathbf{G}_1 \circ \mathbf{T}$  y  $\mathbf{G}_2 \circ \mathbf{T}$  sobre  $\mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$  y del hecho de que F tiene un extremo local sobre  $\mathcal{M} \cap [\mathbf{G}_1^{-1}(0) \cap \mathbf{G}_2^{-1}(0)]$  en  $\varphi$ , obtendremos que  $F \circ \mathbf{T}$  tiene un extremo local en 0, sujeto a las condiciones  $\mathbf{G}_1 \circ \mathbf{T}(\psi) = 0$  y  $\mathbf{G}_2 \circ \mathbf{T}(\psi) = 0$ .

Por hipótesis,  $\varphi$  no es un extremal para  $\mathbf{G}_i$ , donde  $i = 1, 2$  sobre  $\mathcal{M}$ , es decir,

$$d\mathbf{G}_{i\varphi}|_{T\mathcal{M}_\varphi} \neq 0$$

así  $d(\mathbf{G}_i \circ \mathbf{T})_0 \neq 0$ . Entonces, si consideramos una función lineal  $\Lambda : \mathbb{R} \rightarrow \mathbb{R}$ , por el Teorema 2.25, tenemos que

$$d(F \circ \mathbf{T})_0 = \Lambda \circ d([\mathbf{G}_1 \circ \mathbf{T}] + [\mathbf{G}_2 \circ \mathbf{T}]_0)$$

para cada  $i = 1, 2$  y donde  $d\mathbf{T}_0$  es la aplicación identidad sobre  $\mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ . La regla de la cadena nos da en consecuencia que

$$dF_\varphi = \Lambda \circ (d\mathbf{G}_{1\varphi} + d\mathbf{G}_{2\varphi})$$

en  $\mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ . Escribiendo ahora  $\Lambda(t) = (\lambda_1 + \lambda_2)t$  y aplicando el Corolario 2.16 para los diferenciales  $dF_\varphi$  y  $d\mathbf{G}_{i\varphi}$ , concluimos que

$$\begin{aligned} & \int_a^b \left[ \frac{\partial f}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial f}{\partial y}(\varphi(t), \varphi'(t), t) \right] u(t) dt \\ &= \lambda_1 \int_a^b \left[ \frac{\partial g_1}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial g_1}{\partial y}(\varphi(t), \varphi'(t), t) \right] u(t) dt \\ &+ \lambda_2 \int_a^b \left[ \frac{\partial g_2}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial g_2}{\partial y}(\varphi(t), \varphi'(t), t) \right] u(t) dt \end{aligned}$$

para toda  $u \in \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ .

Si  $h : \mathbb{R}^3 \rightarrow \mathbb{R}$  está definida por

$$h(x, y, t) = f(x, y, t) - \lambda_1 g_1(x, y, t) - \lambda_2 g_2(x, y, t)$$

obtenemos que

$$\int_a^b \left[ \frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) \right] u(t) dt = 0$$

para toda  $u \in \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$ . En virtud del Lema 2.17, nos queda que

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0$$

□

**Teorema 2.28.** *Sea*

$$\mathcal{M}_1 \subset \{\psi \in \mathcal{C}^1[\mathbf{a}, \mathbf{b}] : \psi(\mathbf{a}) = \alpha \quad y \quad \psi(\mathbf{b}) = \beta \}$$

*y sean F y G, los funcionales sobre  $\mathcal{C}^1[\mathbf{a}, \mathbf{b}]$  definidos por*

$$F(\psi) = \int_{\mathbf{a}}^{\mathbf{b}} f(\psi(t), \psi'(t), t) dt \quad y \quad G(\psi) = \int_{\mathbf{a}}^{\mathbf{b}} g(\psi(t), \psi'(t), t) dt - c$$

*donde f y g son funciones de clase  $\mathcal{C}^2$  en  $\mathbb{R}^3$ . Sea  $\varphi \in \mathcal{M}_1$  una función de clase  $\mathcal{C}^2$  que no es una extremal de G. Si F tiene un extremo local en  $\varphi$  sujeto a las condiciones*

$$\psi(\mathbf{a}) = \alpha, \quad \psi(\mathbf{b}) = \beta, \quad y \quad G(\psi) = 0$$

*entonces existe un número real  $\lambda$  tal que  $\varphi$  satisface la ecuación de Euler - Lagrange para la función  $h = f - \lambda g$ , a saber,*

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0$$

*para todo  $t \in [\mathbf{a}, \mathbf{b}]$ .*

*Demostración.* Supongamos que  $\varphi \in \mathcal{M}_1$  y es de clase  $\mathcal{C}^2$ , tal que F tiene un extremo local sobre  $\mathcal{M}_1 \cap G^{-1}(0)$ . Sabemos que  $\mathcal{M}_0 \subset \mathcal{C}_0^1[\mathbf{a}, \mathbf{b}]$  y que  $\mathcal{M}_1$  es la traslación por cualquier elemento fijo de este subespacio. Sea entonces  $T : \mathcal{M}_0 \longrightarrow \mathcal{M}_1$ , la traslación definida por

$$T(\psi) = \psi + \varphi$$

notando nuevamente que  $T(0) = \varphi$ , de donde

$$dT_0 : \mathcal{M}_0 \longrightarrow \mathcal{M}_0 = T\mathcal{M}_{1_\varphi}$$

es la aplicación identidad.

Ahora, consideremos las funciones a valores reales  $F \circ T$  y  $G \circ T$  sobre  $\mathcal{M}_0$  y del hecho de que F tiene un extremo local sobre  $\mathcal{M}_1 \cap G^{-1}(0)$  en  $\varphi$ , obtenemos que  $F \circ T$  tiene un extremo local en 0, sujeto a la condición  $G \circ T(\psi) = 0$ .

Por hipótesis,  $\varphi$  no es un extremal para G sobre  $\mathcal{M}_1$ , es decir,

$$dG_\varphi|_{T\mathcal{M}_{1_\varphi}} \neq 0$$

así  $d(G \circ T)_0 \neq 0$ . Entonces, si consideramos una función lineal  $\Lambda : \mathbb{R} \longrightarrow \mathbb{R}$ , por el Teorema 2.25, tenemos que

$$d(F \circ T)_0 = \Lambda \circ d(G \circ T)_0$$

donde  $d\mathcal{T}_0$  es la aplicación identidad sobre  $\mathcal{M}_0$ , la regla de la cadena nos da en consecuencia que

$$dF_\varphi = \Lambda \circ dG_\varphi$$

en  $\mathcal{M}_0$ . Escribiendo  $\Lambda(t) = \lambda t$  y aplicando el Corolario 2.16 para los diferenciales  $dF_\varphi$  y  $dG_\varphi$ , concluimos que

$$\begin{aligned} & \int_a^b \left[ \frac{\partial f}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial f}{\partial y}(\varphi(t), \varphi'(t), t) \right] u(t) dt \\ &= \lambda \int_a^b \left[ \frac{\partial g}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial g}{\partial y}(\varphi(t), \varphi'(t), t) \right] u(t) dt \end{aligned}$$

para toda  $u \in \mathcal{M}_0$ .

Si  $h : \mathbb{R}^3 \rightarrow \mathbb{R}$  está definida por

$$h(x, y, t) = f(x, y, t) - \lambda g(x, y, t)$$

obtenemos que

$$\int_a^b \left[ \frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) \right] u(t) dt = 0$$

para toda  $u \in \mathcal{M}_0$ . En virtud del Lema 2.17, nos queda que

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0$$

□

**Teorema 2.29.** Sean  $F$ ,  $G_1$  y  $G_2$ , los funcionales sobre  $\mathcal{C}^1[a, b]$  definidos por

$$F(\psi) = \int_a^b f(\psi(t), \psi'(t), t) dt$$

y

$$G_i(\psi) = \int_a^b g_i(\psi(t), \psi'(t), t) dt - c_i$$

donde  $i = 1, 2$  y además  $f$  y  $g_i$  son funciones de clase  $\mathcal{C}^2$  en  $\mathbb{R}^3$ , para  $i = 1, 2$ . Sea  $\varphi \in \mathcal{M}_1$  una función de clase  $\mathcal{C}^2$  que no es una extremal de  $G_i$ . Si  $F$  tiene un extremo local en  $\varphi$  sujeto a las condiciones

$$\psi(a) = \alpha, \quad \psi(b) = \beta, \quad y \quad G_i(\psi) = 0$$



entonces existen números reales  $\lambda_1$  y  $\lambda_2$  tales que  $\varphi$  satisface la ecuación de Euler - Lagrange para la función  $h = f - \lambda_1 g_1 - \lambda_2 g_2$ , a saber,

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0$$

para todo  $t \in [a, b]$ .

*Demostración.* La demostración es inmediata usando conjuntamente los Teoremas 2.27 y 2.28. □

## Capítulo 3

# Extensión de la ecuación de Euler-Lagrange a un intervalo no acotado

Supongamos que tenemos un funcional definido sobre un subconjunto que consta de funciones de clase  $\mathcal{C}^1$ , en un intervalo semi-infinito de la forma  $(-\infty, a]$  ó  $[b, \infty)$  o inclusive sobre toda la recta y necesitamos hallar las funciones extremales de este funcional. Para ello, podemos hacer uso de dos herramientas que nos ayudarían en la consecución de nuestro objetivo. La primera, consiste en usar el hecho de que la ecuación de Euler es invariante bajo transformaciones regulares y la segunda, consiste en realizar manipulaciones de la ecuación de Euler haciendo uso de la regla de la cadena, veamos:

### 3.1 Invarianza de la ecuación de Euler

#### 3.1.1 Cambio de variables generalizado

Si el funcional dado en (2.10) se transforma efectuando una sustitución de variable independiente o una sustitución simultánea de la función incógnita y de la variable independiente, las extremales continúan determinándose de la ecuación de Euler que se obtiene a partir del integrando transformado.

Para ello consideremos,  $t = t(u, v)$  y  $x = x(u, v)$ , con la particularidad de que

$$\begin{vmatrix} t_u & t_v \\ x_u & x_v \end{vmatrix} \neq 0$$

y consideremos adicionalmente a  $u$  como variable independiente, entonces, por regla de la cadena se obtiene que

$$y = \frac{dx}{dt} = \frac{x_u + x_v v'}{t_u + t_v v'}$$

luego

$$\int f(x, y, t) dt = \int f \left[ x(u, v), \frac{x_u + x_v v'}{t_u + t_v v'}, t(u, v) \right] (t_u + t_v v') du = \int h(v, v', u) du$$

y las extremales de la ecuación inicial se determinan de la ecuación de Euler para la funcional  $\int h(v, v', u) du$ , es decir, la función debe satisfacer la ecuación (2.10), a saber;

$$h_v - \frac{d}{du} h_{v'} = 0$$

(Para más detalles, ver [6] pág. 30)

### 3.1.2 Cambio de variables en la ecuación de Euler - Lagrange

Observemos el resultado fundamental del Teorema 2.27: Si el funcional

$$F(\psi) = \int_a^b f(\psi(t), \psi'(t), t) dt \tag{3.1}$$

sujeto a la restricción

$$G(\psi) = \int_a^b g(\psi(t), \psi'(t), t) dt = k, \quad k = \text{ctte} \tag{3.2}$$

alcanza un máximo o un mínimo en  $\varphi(t)$ , entonces existe  $\lambda \in \mathbb{R}$  tal que  $\varphi(t)$  satisface la ecuación

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0$$

para  $h = f - \lambda g$  (función de Lagrange), donde  $f$  y  $g$  son funciones de clase  $\mathcal{C}^2$  en  $\mathbb{R}^3$  y  $\varphi \in \mathcal{M}$  es una función de clase  $\mathcal{C}^2$  que no es una extremal de  $G$ , y además se considera siempre que  $-\infty < a < b < \infty$ .

Ahora bien, se hace necesario precisar la invarianza de las ecuaciones de Euler-Lagrange, para ello, consideremos el siguiente problema: Hallar los máximos o los mínimos del funcional

$$F(\psi) = \int_c^d f(\psi(t), \psi'(t), t) dt$$

sujeto a la restricción

$$G(\psi) = \int_c^d g(\psi(t), \psi'(t), t) dt = k, \quad k = \text{cte}$$

con la particularidad de que  $c$  y  $d$  pueden tomar los valores  $-\infty$  y/o  $\infty$  respectivamente.

Estos resultados nos permiten entonces enunciar un resultado análogo al Teorema 2.27, pero en la versión infinita. Antes de enunciar el teorema definiremos al conjunto donde trabajaremos.

Consideraremos al conjunto

$$\mathcal{D} = \left\{ \psi \in \mathcal{C}^1(\mathbb{R}) : \lim_{|t| \rightarrow \infty} |t|^n \psi(t) = 0, \text{ para todo } n \in \mathbb{N} \right\}$$

y lo utilizaremos para demostrar el siguiente teorema.

**Teorema 3.1.** Sean  $F$  y  $G$ , los funcionales definidos por

$$F(\psi) = \int_{-\infty}^{\infty} f(\psi(t), \psi'(t), t) dt \quad (3.3)$$

y

$$G(\psi) = \int_{-\infty}^{\infty} g(\psi(t), \psi'(t), t) dt - c \quad (3.4)$$

con  $\psi \in \mathcal{D}$  y donde,  $f$  y  $g$  son funciones de clase  $\mathcal{C}^2$ . Existen constantes  $\alpha_{1f}, \alpha_{1g} \in (0, 1)$  y  $\alpha_{2f}, \alpha_{2g} \in [1, \infty)$  y  $\beta_{1f}, \beta_{1g}, \beta_{2f}, \beta_{2g}, k_1, k_2, k_3, k_4 > 0$ , tales que

$$|f(x, y, t)| \leq k_1 |x|^{\alpha_{1f}} |t|^{\beta_{1f}} \quad y \quad |g(x, y, t)| \leq k_2 |x|^{\alpha_{1g}} |t|^{\beta_{1g}}$$

si  $|x| < 1$  y  $t \in \mathbb{R}$  y además

$$|f(x, y, t)| \leq k_3 |x|^{\alpha_{2f}} |t|^{\beta_{2f}} \quad y \quad |g(x, y, t)| \leq k_4 |x|^{\alpha_{2g}} |t|^{\beta_{2g}}$$

si  $|x| \geq 1$  y  $t \in \mathbb{R}$ . Sea  $\varphi \in \mathcal{D}$  una función que no es una extremal de  $G$ . Si  $F$  tiene un extremo local en  $\varphi$ , entonces, existe un número real  $\lambda$ , tal que  $\varphi$  satisface la ecuación de Euler - Lagrange para la función  $h = f - \lambda g$ , es decir,

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0 \quad (3.5)$$

para todo  $t \in \mathbb{R}$ .

*Demostración.* Sea  $(\mathbf{a}, \mathbf{b})$  un intervalo finito y sea  $\mathbf{u} : (\mathbf{a}, \mathbf{b}) \rightarrow (-\infty, \infty)$  una transformación creciente de clase  $\mathcal{C}^\infty$ , tal que:

$$\lim_{s \rightarrow \mathbf{a}^+} \mathbf{u}(s) = -\infty \quad , \quad \lim_{s \rightarrow \mathbf{b}^-} \mathbf{u}(s) = \infty$$

Sea  $s$  un punto interior del intervalo  $(\mathbf{a}, \mathbf{b})$ . Considerando el cambio de variables  $\mathbf{t} = \mathbf{u}(s)$ , las expresiones (3.1) y (3.2), nos proporcionan un nuevo problema variacional, dado por

$$F(\psi) = \int_{\mathbf{a}}^{\mathbf{b}} f(\psi(\mathbf{u}(s)), \psi'(\mathbf{u}(s)), \mathbf{u}(s)) \mathbf{u}'(s) \, ds \quad (3.6)$$

$$G(\psi) = \int_{\mathbf{a}}^{\mathbf{b}} g(\psi(\mathbf{u}(s)), \psi'(\mathbf{u}(s)), \mathbf{u}(s)) \mathbf{u}'(s) \, ds = K, \quad K = \text{ctte} \quad (3.7)$$

Resulta importante hacer notar que

$$\begin{aligned} \int_{-\infty}^{\infty} |f(\psi(\mathbf{t}), \psi'(\mathbf{t}), \mathbf{t})| \, d\mathbf{t} &= \int_{|\psi(\mathbf{t})| < 1} |f(\psi(\mathbf{t}), \psi'(\mathbf{t}), \mathbf{t})| \, d\mathbf{t} + \int_{|\psi(\mathbf{t})| \geq 1} |f(\psi(\mathbf{t}), \psi'(\mathbf{t}), \mathbf{t})| \, d\mathbf{t} \\ &\leq \int_{-\infty}^{\infty} k_1 |\psi(\mathbf{t})|^{\alpha_1 f} |\mathbf{t}|^{\beta_1 f} \, d\mathbf{t} + \int_{-\infty}^{\infty} k_3 |\psi(\mathbf{t})|^{\alpha_2 f} |\mathbf{t}|^{\beta_2 f} \, d\mathbf{t} \end{aligned}$$

ya que por hipótesis, se tiene que

$$|f(\mathbf{x}, \mathbf{y}, \mathbf{t})| \leq k_1 |\mathbf{x}|^{\alpha_1 f} |\mathbf{t}|^{\beta_1 f}, \quad \text{si } |\mathbf{x}| < 1$$

y además

$$|f(\mathbf{x}, \mathbf{y}, \mathbf{t})| \leq k_3 |\mathbf{x}|^{\alpha_2 f} |\mathbf{t}|^{\beta_2 f}, \quad \text{si } |\mathbf{x}| \geq 1$$

y como

$$\lim_{|\mathbf{t}| \rightarrow \infty} |\mathbf{t}|^n f(\mathbf{t}) = 0$$

para todo  $n \in \mathbb{N}$ , las dos integrales anteriores convergen.

Por tanto, la integral que define  $F(\psi)$ , en la ecuación (3.6) siempre converge. Análogamente, la integral que define  $G(\psi)$ , en la ecuación (3.7) también converge.

Consideremos ahora,  $\gamma$  una función definida en  $(\mathbf{a}, \mathbf{b})$ , tal que  $\gamma(s) = \psi(\mathbf{u}(s))$ , tendremos entonces que

$$\gamma'(s) = \psi'(\mathbf{u}(s)) \mathbf{u}'(s)$$

es decir,

$$\psi'(\mathbf{u}(s)) = \frac{\gamma'(s)}{\mathbf{u}'(s)}$$

por consiguiente, el problema se ha reducido a hallar los extremos del funcional

$$F(\gamma) = \int_a^b f\left(\gamma(s), \frac{\gamma'(s)}{u'(s)}, u(s)\right) u'(s) ds \quad (3.8)$$

sujeto a la restricción

$$G(\gamma) = \int_a^b g\left(\gamma(s), \frac{\gamma'(s)}{u'(s)}, u(s)\right) u'(s) ds = K, \quad K = \text{cte} \quad (3.9)$$

Sea  $\omega(s)$  un extremo, entonces,  $\omega$  debe satisfacer la ecuación

$$\frac{\partial \tilde{h}}{\partial x}(\omega(s), \omega'(s), s) - \frac{d}{ds} \frac{\partial \tilde{h}}{\partial y}(\omega(s), \omega'(s), s) = 0$$

donde  $\tilde{h} = \tilde{f} - \lambda \tilde{g}$  y adicionalmente

$$\tilde{f}(x, y, s) = f\left(x, \frac{y}{u'(s)}, u(s)\right) u'(s) \quad y \quad \tilde{g}(x, y, s) = g\left(x, \frac{y}{u'(s)}, u(s)\right) u'(s)$$

Si definimos  $\varphi : (c, d) \rightarrow \mathbb{R}$ , por

$$\varphi(t) = \omega(u^{-1}(t))$$

entonces  $\varphi$  es un extremo del problema inicial y además  $\varphi(u(s)) = \omega(s)$ , donde

$$\omega'(s) = \varphi'(u(s)) \cdot u'(s)$$

Sea  $h = f - \lambda g$ , donde

$$\tilde{h}(x, y, s) = h\left(x, \frac{y}{u'(s)}, u(s)\right) u'(s)$$

entonces

$$\frac{\partial \tilde{h}}{\partial x}(\omega(s), \omega'(s), s) = \frac{\partial h}{\partial x}(\varphi(u(s)), \varphi'(u(s)), u(s)) u'(s) \quad (3.10)$$

y por otra parte

$$\frac{\partial \tilde{h}}{\partial y}(\omega(s), \omega'(s), s) = \frac{\partial h}{\partial y}\left(x, \frac{y}{u'(s)}, u(s)\right) \quad (3.11)$$

luego tomando la derivada con respecto a  $s$  de la expresión (3.11), obtenemos:

$$\begin{aligned} \frac{d}{ds} \frac{\partial \tilde{h}}{\partial y}(\omega(s), \omega'(s), s) &= \frac{d}{ds} \frac{\partial h}{\partial y}\left(\omega(s), \frac{\omega'(s)}{u'(s)}, u(s)\right) \\ &= \frac{d}{ds} \frac{\partial h}{\partial y}(\varphi(u(s)), \varphi'(u(s)), u(s)) \\ &= \nabla \frac{\partial h}{\partial y}(\varphi(u(s)), \varphi'(u(s)), u(s)) \cdot (\varphi'(u(s))u'(s), \varphi''(u(s))u'(s), u'(s)) \\ &= u'(s) \nabla \frac{\partial h}{\partial y}(\varphi(u(s)), \varphi'(u(s)), u(s)) \cdot (\varphi'(u(s)), \varphi''(u(s)), 1) \end{aligned} \quad (3.12)$$

igualando entonces las ecuaciones (3.10) y (3.12), nos queda

$$\frac{\partial h}{\partial x}(\varphi(\mathbf{u}(s)), \varphi'(\mathbf{u}(s)), \mathbf{u}(s))\mathbf{u}'(s) = \mathbf{u}'(s)\nabla \frac{\partial h}{\partial \mathbf{y}}(\varphi(\mathbf{u}(s)), \varphi'(\mathbf{u}(s)), \mathbf{u}(s)).(\varphi'(\mathbf{u}(s)), \varphi''(\mathbf{u}(s)), \mathbf{1})$$

dividiendo entre  $\mathbf{u}'(s)$  y tomando el cambio  $\mathbf{t} = \mathbf{u}(s)$ , obtenemos finalmente

$$\begin{aligned} \frac{\partial h}{\partial x}(\varphi(\mathbf{t}), \varphi'(\mathbf{t}), \mathbf{t}) &= \nabla \frac{\partial h}{\partial \mathbf{y}}(\varphi(\mathbf{t}), \varphi'(\mathbf{t}), \mathbf{t}).(\varphi'(\mathbf{t}), \varphi''(\mathbf{t}), \mathbf{1}) \\ &= \frac{d}{dt} \frac{\partial h}{\partial \mathbf{y}}(\varphi(\mathbf{t}), \varphi'(\mathbf{t}), \mathbf{t}) \end{aligned}$$

donde obtenemos la misma ecuación que en un intervalo finito.  $\square$

A continuación, presentaremos el resultado con múltiples restricciones, cuya demostración es omitida, por ser totalmente análoga a la demostración del teorema anterior.

**Teorema 3.2.** Sean  $F$  y  $G_j$ , con  $j = 1, \dots, n$ , los funcionales definidos por

$$F(\psi) = \int_{-\infty}^{\infty} f(\psi(t), \psi'(t), t) dt$$

y

$$G_j(\psi) = \int_{-\infty}^{\infty} g_j(\psi(t), \psi'(t), t) dt - c_j$$

con  $\psi \in \mathcal{D}$  y donde,  $f$  y  $g_j$ , para cada  $j = 1, \dots, n$ , son funciones de clase  $\mathcal{C}^2$ . Existen constantes  $\alpha_{1f}, \alpha_{1g_j} \in (0, 1)$ ,  $j = 1, \dots, n$  y  $\alpha_{2f}, \alpha_{2g_j} \in [1, \infty)$ ,  $j = 1, \dots, n$  y  $\beta_{1f}, \beta_{2f}, k_1, k_3 > 0$  y  $\beta_{1g_j}, \beta_{2g_j}, k_{2j}, k_{4j} > 0$ ,  $j = 1, \dots, n$ , tales que

$$|f(x, y, t)| \leq k_1|x|^{\alpha_{1f}}|t|^{\beta_{1f}} \quad y \quad |g_j(x, y, t)| \leq k_{2j}|x|^{\alpha_{1g_j}}|t|^{\beta_{1g_j}}$$

si  $|x| < 1$  y  $t \in \mathbb{R}$ , con  $j = 1, \dots, n$  y además

$$|f(x, y, t)| \leq k_3|x|^{\alpha_{2f}}|t|^{\beta_{2f}} \quad y \quad |g_j(x, y, t)| \leq k_{4j}|x|^{\alpha_{2g_j}}|t|^{\beta_{2g_j}}$$

si  $|x| \geq 1$  y  $t \in \mathbb{R}$ , con  $j = 1, \dots, n$ . Sea  $\varphi \in \mathcal{D}$  una función que no es una extremal de  $G_j$ , para  $j = 1, \dots, n$ . Si  $F$  tiene un extremo local en  $\varphi$ , entonces, existen números reales  $\lambda_j$ , con  $j = 1, \dots, n$  tales que,  $\varphi$  satisface la ecuación de Euler - Lagrange para la función  $h = f - \sum_{j=1}^n \lambda_j g_j$ , es decir,

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial \mathbf{y}}(\varphi(t), \varphi'(t), t) = 0$$

para todo  $t \in \mathbb{R}$ .

*Observaciones 3.3.* Podemos encontrar funciones de distribución, que maximizan al funcional que define la entropía, que no están definidos sobre toda la recta, entonces, en estos casos podemos usar los resultados expuestos en los Teoremas 3.1 y 2.27, de la siguiente manera:

1. Si el intervalo donde está definido el funcional y las restricciones es de la forma  $(\mathbf{a}, \infty)$ , con  $\mathbf{a} \in \mathbb{R}$ , tenemos que el conjunto  $\mathcal{D}$ , se define como

$$\mathcal{D} = \left\{ \psi \in \mathcal{C}^1(\mathbb{R}) : \lim_{t \rightarrow \infty} t^n \psi(t) = 0, \text{ para todo } n \in \mathbb{N} \right\}$$

2. Si el intervalo donde está definido el funcional y las restricciones es de la forma  $(-\infty, \mathbf{a})$ , con  $\mathbf{a} \in \mathbb{R}$ , tenemos que el conjunto  $\mathcal{D}$ , está definido por la expresión

$$\mathcal{D} = \left\{ \psi \in \mathcal{C}^1(\mathbb{R}) : \lim_{t \rightarrow -\infty} t^n \psi(t) = 0, \text{ para todo } n \in \mathbb{N} \right\}$$



## Capítulo 4

# Cálculo variacional y problemas de máxima entropía

Sea  $X$  una variable aleatoria que toma valores en la recta real. La probabilidad que  $X$  tome valores menores o iguales que un número real dado  $x$ , se obtiene integrando a la función de densidad  $\rho$ , a saber

$$P(X \leq x) = \int_{-\infty}^x \rho(t) dt$$

Como  $X$  puede tomar cualquier valor, obtenemos que

$$\int_{\mathbb{R}} \rho(t) dt = 1$$

En muchos problemas, el interés principal radica en lograr determinar la función de densidad  $\rho$ , teniendo como base el conocimiento previo de ciertos valores esperados.

En general, lo que se quiere es hallar los extremos del funcional

$$F(\psi) = \int_{-\infty}^{\infty} f(\psi(t), \psi'(t), t) dt \quad (4.1)$$

sujeto a las restricciones

$$\begin{aligned} G_1(\psi) &= \int_{-\infty}^{\infty} g_1(\psi(t), \psi'(t), t) dt = c_1 \\ &\vdots \\ G_k(\psi) &= \int_{-\infty}^{\infty} g_k(\psi(t), \psi'(t), t) dt = c_k \end{aligned} \quad (4.2)$$

Ahora bien, si  $\varphi$  es un extremo del funcional dado en (4.1), sujeto a las restricciones observadas en (4.2) y en virtud de los Teoremas 3.1 y 2.29 deben existir  $\lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{R}$ ,

tales que  $\varphi$  satisface la ecuación de Euler - Lagrange para la función  $h$  definida por

$$h(x, y, t) = f(x, y, t) - \sum_{k=1}^n \lambda_k g_k(x, y, t)$$

es decir,

$$\frac{\partial h}{\partial x}(\varphi(t), \varphi'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\varphi(t), \varphi'(t), t) = 0$$

## 4.1 Teorema general

**Teorema 4.1.** Sean  $h_1, \dots, h_n : \mathbb{R} \rightarrow \mathbb{R}$  funciones medibles, tales que,  $|h_k(t)| \leq m_k |t|^{n_k}$ ,  $m_k, n_k > 0$ , para cada  $k = 1, \dots, n$  y  $c_1, \dots, c_n \in \mathbb{R}$ . Supóngase que existen  $a_1, \dots, a_n \in \mathbb{R}$  tales que la función  $q : \mathbb{R} \rightarrow \mathbb{R}$  definida por

$$q(t) = \exp(a_0 + a_1 h_1(t) + \dots + a_n h_n(t))$$

es una densidad de probabilidad y cada una de las funciones  $h_k q$ ,  $k = 1, \dots, n$  es integrable en  $\mathbb{R}$  y

$$\int_{-\infty}^{+\infty} h_k(t) q(t) dt = c_k, \quad k = 1, \dots, n.$$

Entonces, el funcional  $F$  definido por

$$F(p) = - \int_{-\infty}^{+\infty} p(t) \log p(t) dt,$$

con dominio el conjunto de todas las funciones medibles  $p : \mathbb{R} \rightarrow \mathbb{R}$  tales que

$$p(t) \geq 0, \quad \int_{-\infty}^{+\infty} p(t) dt = 1 \quad y \quad \int_{-\infty}^{+\infty} h_k(t) p(t) dt = c_k, \quad k = 1, \dots, n,$$

alcanza un extremo en  $q$ .

*Demostración.* Consideremos el funcional, dado por

$$F(p) = \int_{-\infty}^{\infty} p(t) \log p(t) dt \tag{4.3}$$

sujeto a las restricciones

$$\int_{-\infty}^{\infty} h_k(t) p(t) dt = c_k, \quad k = 0, \dots, n \tag{4.4}$$

Consideremos además, las funciones  $f$  y  $g_k$ , con  $k = 0, \dots, n$ , definidas de la siguiente manera

$$f(x, y, t) = -x \log x, \quad g_k(x, y, t) = x h_k(t), \quad \text{con } k = 1, \dots, n$$

donde

$$|f(x, y, t)| = |-x \log x| = |x| |\log x| \leq |x|^{3/2}, \text{ si } |x| < 1 \quad \text{y} \quad |f(x, y, t)| \leq |x|^2, \text{ si } |x| > 1$$

y

$$|g_k(x, y, t)| = |x h_k(t)| \leq |x| |t|^{n_k}$$

además,  $h_0(t) \equiv 1$ , para todo  $t \in \mathbb{R}$  y  $c_0 = 1$ .

Ahora bien, si  $q$  es un extremo del funcional dado en (4.3), sujeto a las restricciones observadas en (4.4), deben existir  $\lambda_0, \lambda_1, \dots, \lambda_n \in \mathbb{R}$ , tales que  $q$  satisface la ecuación de Euler - Lagrange para la función  $z$  definida por

$$z(x, y, t) = f(x, y, t) - \sum_{k=0}^n \lambda_k g_k(x, y, t)$$

es decir,

$$\frac{\partial z}{\partial x}(q(t), q'(t), t) - \frac{d}{dt} \frac{\partial z}{\partial y}(q(t), q'(t), t) = 0$$

Por consiguiente, tenemos que

$$\frac{\partial z}{\partial x} = -\log x - 1 - \sum_{k=0}^n \lambda_k h_k(t) \quad \text{y} \quad \frac{\partial z}{\partial y} = 0$$

luego la ecuación de Euler - Lagrange para  $q$ , viene dada por

$$-\log q(t) - 1 - \sum_{k=0}^n \lambda_k h_k(t) = 0$$

y por tanto

$$\log [q(t)] = -1 - \sum_{k=0}^n \lambda_k h_k(t)$$

de donde se obtiene que

$$q(t) = \exp \left\{ -1 - \sum_{k=0}^n \lambda_k h_k(t) \right\} = \exp \{ a_0 + a_1 h_1(t) + \dots + a_n h_n(t) \}$$

donde  $a_k = -\lambda_k$ , para cada  $k = 1, \dots, n$  y  $a_0 = -1 - \lambda_0$ . □

*Observación 4.2.* Si nos encontramos en una situación que amerite la descripción de un fenómeno o modelado de un proceso y en consecuencia, por propiedades del fenómeno mismo, encontráramos que existe un único extremo, este debe corresponder con la distribución de máxima entropía.

A continuación, presentaremos como casos particulares, algunos tipos de funciones de densidad que generan máxima entropía en ciertos sistemas dependiendo de sus restricciones y nos dedicaremos a encontrar estas funciones usando las técnicas del cálculo variacional.

### 4.1.1 Caracterización de las funciones de máxima entropía según sus restricciones

Consideremos a  $\mathcal{X}$  una variable aleatoria que toma valores en algún subconjunto  $I$  de la recta real. Sea  $\rho : I \rightarrow \mathbb{R}$ , la función de densidad de  $\mathcal{X}$ , es decir,  $\rho$  cumple las condiciones siguientes:

- i)  $\rho(t) > 0$ , para todo  $t \in I$  y  $\rho(t) = 0$  en otro caso.
- ii)  $\int_I \rho(t) dt = 1$ .

A partir de las técnicas del cálculo variacional, trataremos de generar algunas de las funciones “típicas” de densidad que maximizan al funcional

$$S[\rho] = - \int_I \rho(t) \log \rho(t) dt \quad (4.5)$$

En general, trabajaremos con restricciones adicionales a las presentadas en (i) y (ii), ya que, supondremos además que la función  $\rho$ , satisface condiciones de la forma

$$\int_I p_k(t) \rho(t) dt = c_k, \quad c_k \in \mathbb{R}, \quad k = 1, 2, \dots \quad (4.6)$$

donde  $p_1(t) = 1$ , para todo  $t \in I$  y  $c_1 = 1$ .

#### Caso 1: Exponencial

Supongamos entonces que  $I = (0, \infty)$  y  $\rho$  es una función que satisface las condiciones (i) y (ii), además de que  $E(\mathcal{X}) = \alpha$ . Por consiguiente, el problema se centra en maximizar el

funcional (4.5), sujeto a las restricciones

$$\int_0^{\infty} \rho(t) dt = 1 \quad \text{y} \quad \int_0^{\infty} t\rho(t) dt = \alpha, \quad \alpha > 0 \quad (4.7)$$

donde  $p_2(t) = t$  y  $c_2 = \alpha$ .

Consideremos entonces a las funciones

$$f(x, y, t) = -x \log x, \quad g_1(x, y, t) = x \quad \text{y} \quad g_2(x, y, t) = tx$$

Luego las funciones  $g_1$  y  $g_2$ , satisfacen que

$$|g_1(x, y, t)| \leq |x| \quad \text{y} \quad |g_2(x, y, t)| \leq |t||x|$$

Entonces, por los Teoremas 3.1 y 2.29, existen  $\lambda_1, \lambda_2 \in \mathbb{R}$ , tales que  $\rho$  satisface la ecuación de Euler - Lagrange para la función  $h$  definida por

$$h(x, y, t) = f(x, y, t) - \lambda_1 g_1(x, y, t) - \lambda_2 g_2(x, y, t)$$

es decir,

$$\frac{\partial h}{\partial x}(\rho(t), \rho'(t), t) - \frac{d}{dt} \frac{\partial h}{\partial y}(\rho(t), \rho'(t), t) = 0 \quad (4.8)$$

para todo  $t \in (0, \infty)$ .

En virtud de que

$$\frac{\partial h}{\partial x} = -\log x - 1 - \lambda_1 - \lambda_2 t \quad \text{y} \quad \frac{\partial h}{\partial y} = 0$$

tenemos que la ecuación (4.8), se puede representar por

$$-\log x - 1 - \lambda_1 - \lambda_2 t = 0$$

Por consiguiente, si  $\rho$  es la función extremal, obtenemos en consecuencia

$$\rho(t) = C e^{-\lambda_2 t}, \quad C = e^{1+\lambda_1}$$

Luego usando las condiciones dadas en (4.7), obtenemos que

$$1 = \int_0^{\infty} C e^{-\lambda_2 t} dt = C \left[ -\frac{1}{\lambda_2} e^{-\lambda_2 t} \right]_0^{\infty} = \frac{C}{\lambda_2}$$

por tanto,  $C = \lambda_2$ .

Por otro lado,

$$\alpha = \int_0^{\infty} C t e^{-\lambda_2 t} dt = C \int_0^{\infty} t e^{-\lambda_2 t} dt = C \left[ -\frac{t}{\lambda_2} e^{-\lambda_2 t} \right]_0^{\infty} + \frac{1}{\lambda_2} \int_0^{\infty} e^{-\lambda_2 t} dt = C \left[ \frac{1}{\lambda_2 C} \right] = \frac{1}{\lambda_2}$$

Luego,  $\rho(t) = \frac{1}{\alpha} e^{-\frac{1}{\alpha} t}$ . Si tomamos  $\frac{1}{\alpha} = a$ , obtenemos el resultado esperado.

**Caso 2: Gamma**

Supongamos nuevamente que  $I = (0, \infty)$ , y que la función  $\rho$  maximiza al funcional dado en (4.5), sujeto a las restricciones

$$\int_0^\infty \rho(t) dt = 1, \quad \int_0^\infty t\rho(t) dt = \alpha, \quad \alpha > 0 \quad \text{y} \quad \int_0^\infty \log(t)\rho(t) dt = \beta \quad (4.9)$$

Entonces, considerando ahora la función

$$h(x, y, t) = f(x, y, t) - \lambda_1 g_1(x, y, t) - \lambda_2 g_2(x, y, t) - \lambda_3 g_3(x, y, t)$$

donde

$$f(x, y, t) = -x \log x, \quad g_1(x, y, t) = x, \quad g_2(x, y, t) = tx \quad \text{y} \quad g_3(x, y, t) = x \log t$$

Por el caso 1, sabemos que  $g_1$  y  $g_2$ , así definidas cumplen con las condiciones de los Teoremas 3.1 y 2.29, luego, considerando  $g_3(x, y, t) = x \log t$ , es fácil ver que

$$|g_3(x, y, t)| = |x \log t| \leq |t||x|, \quad \text{para cada } t \in (0, \infty)$$

lo cual verifica las condiciones de los Teoremas 3.1 y 2.29 y realizando un cálculo análogo, nos queda que

$$-\log x - 1 - \lambda_1 - \lambda_2 t - \lambda_3 \log t = 0$$

de donde obtenemos que

$$\rho(t) = C e^{-\lambda_2 t} t^{-\lambda_3} \quad (4.10)$$

Usando las restricciones dadas en (4.9), obtenemos en principio que

$$1 = \int_0^\infty C e^{-\lambda_2 t} t^{-\lambda_3} dt = C \int_0^\infty e^{-u} \frac{u^{-\lambda_3}}{\lambda_2^{-\lambda_3}} \frac{du}{\lambda_2}, \quad \text{si } u = \lambda_2 t \quad (4.11)$$

Ahora bien, si  $p - 1 = -\lambda_3$ , entonces (4.11), se puede escribir así

$$1 = \frac{C}{\lambda_2^p} \int_0^\infty e^{-u} u^{p-1} du = \frac{C}{\lambda_2^p} \Gamma(p) \quad \implies \quad C = \frac{\lambda_2^p}{\Gamma(p)}$$

y de la segunda condición de (4.9), obtenemos que

$$\alpha = \int_0^\infty t C e^{-\lambda_2 t} t^{-\lambda_3} dt = C \int_0^\infty \frac{u}{\lambda_2} e^{-u} \frac{u^{-\lambda_3}}{\lambda_2^{-\lambda_3}} \frac{du}{\lambda_2} = \frac{C}{\lambda_2^{p+1}} \int_0^\infty e^{-u} u^p du = \frac{C}{\lambda_2^{p+1}} p \Gamma(p) = \frac{1}{\lambda_2} p$$

por tanto,  $\lambda_2 = \frac{p}{\alpha}$ . Si  $\frac{p}{\alpha} = a$ , tendremos entonces que (4.10), se puede escribir como

$$\rho(t) = \frac{a^p}{\Gamma(p)} e^{-at} t^{p-1}$$

*Observación 4.3.* Si usamos la tercera condición de (4.9), se obtiene que  $p = 1$ , por lo que la función que acabamos de obtener se transforma en la función del primer caso ya que  $\Gamma(1) = 1$ .

### Caso 3: Laplace

Si consideramos ahora,  $I = \mathbb{R}$ , con las condiciones

$$\int_{-\infty}^{\infty} \rho(t) dt = 1 \quad \text{y} \quad \int_{-\infty}^{\infty} |t|\rho(t) dt = \alpha, \quad \alpha > 0 \quad (4.12)$$

obtendremos que la función de densidad, está dada por

$$\rho(t) = Ce^{-\lambda_2|t|}$$

Luego como la función  $\rho$  en este caso es par, por el caso 1, podemos llegar facilmente a que

$$\rho(t) = \frac{1}{2}ae^{-a|t|}$$

### Caso 4: Normal

Consideremos ahora,  $I = \mathbb{R}$ , con las condiciones

$$\int_{-\infty}^{\infty} \rho(t) dt = 1 \quad \text{y} \quad \int_{-\infty}^{\infty} t^2\rho(t) dt = \sigma^2 \quad (4.13)$$

Consideremos además, a las funciones  $f$ ,  $g_1$  y  $g_2$  como sigue

$$f(x, y, t) = -x \log x, \quad g_1(x, y, t) = x \quad \text{y} \quad g_2(x, y, t) = t^2x$$

y tomando además la función  $h$  que satisface la ecuación de Euler-Lagrange como

$$h(x, y, t) = f(x, y, t) - \lambda_1 g_1(x, y, t) - \lambda_2 g_2(x, y, t)$$

tenemos entonces que

$$\frac{\partial h}{\partial x} = -\log x - 1 - \lambda_1 - \lambda_2 t^2 \quad \text{y} \quad \frac{\partial h}{\partial y} = 0$$

luego la ecuación de Euler - Lagrange para la función extremal  $\rho$  viene dada por

$$-\log \rho(t) - 1 - \lambda_1 - \lambda_2 t^2 = 0$$

y usando propiedades del logaritmo, nos queda

$$\log [C\rho(t)] = -\lambda_2 t^2 \implies \rho(t) = \frac{1}{C} e^{-\lambda_2 t^2}$$

donde  $C = e^{1+\lambda_1}$ .

Luego, usando las restricciones presentes en (4.13) y el hecho de que

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$$

llegamos finalmente a que

$$\rho(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right)$$



# Bibliografía

- [1] AKHIEZER, N., The Calculus of Variations, *Blaisdell Publishing Company* (1962).
- [2] EDWARDS, C., Advanced Calculus, *Prentice Hall* (1963), Vol. II.
- [3] CARTAN, H., Calculo Diferencial, *Ediciones Omega*, Barcelona, (1972). 30
- [4] CHOVER, J., On Normalized Entropy and the Extensions of a Positive - Definite Functions, *Journal of Mathematics and Mechanics*.(1963), Vol. 10, **6**, (1961), pp. 927 - 945.
- [5] FIGUEROA, J., Brief Notes on the Calculus of Variations, <http://www.maths.ed.ac.uk/~jmf/Teaching/Lectures/CoV.pdf>
- [6] GELFAND, I Y FOMIN, S., Calculus of Variations, *Prentice Hall, INC.* (1963). 38
- [7] KAGAN, A. M., LINNIK, Y., RAO, C., Characterization problems in mathematical statistics, *Wiley*. New York, (1973).
- [8] LAVIS, D. A. , Some Examples of Simple Systems, <http://www.mth.kcl.ac.uk/~dlavis/papers/examples-ver2.pdf>. 11
- [9] MARSDEN, J., Elementary Classical Analysis, *W. H. Freeman and Company* (1974).
- [10] MARSDEN, J Y TROMBA, A., Cálculo Vectorial, *Addison - Wesley Iberoamericana* (1991).
- [11] MAHAN, B., Termodinámica Química Elemental, *Ed, Reverté, S.A.* (1976). 9
- [12] MAKARENKO, G, KRASNOV, M Y KISELIOV, A., Cálculo Variacional, *Ed. Mir Moscú* (1976).
- [13] RUDIN, W., Functional Analysis, *Mac - Graw Hill, Inc.* (1974).

- [14] RAO, C., Linear statistical inference and its applications, *Wiley*. 2<sup>da</sup> ed. New York, (1973).
- [15] RESNICK, R Y HALLIDAY, D., Física, *Compañía Editorial Continental, S.A.* Parte I, (1980).
- [16] SHANNON, C., A Mathematical Theory of Communication, *The Bell System Technical Journal*, Vol. 27, pp. 623 - 656, Octubre, 1948. 2