UNIVERSIDAD DE GRANADA

Facultad de Ciencias

Grado en Matemáticas

TRABAJO DE FIN DE GRADO

Título del trabajo

Presentado por:
Nombre apellidos

Curso académico 2023-2024

Título del trabajo

Nombre apellidos

Nombre apellidos *Título del trabajo*.
Trabajo de fin de Grado. Curso académico 2023-2024.

Responsable de tutorización

Nombre del tutor 1
Departamento del tutor 1

Nombre del tutor 2
Departamento del tutor 2

Grado en Matemáticas
Facultad de Ciencias
Universidad de Granada

Declaración de originalidad

D./Dña. Nombre apellidos

Declaro explícitamente que el trabajo presentado como Trabajo de Fin de Grado (TFG), correspondiente al curso académico 2023-2024, es original, entendido esto en el sentido de que no he utilizado para la elaboración del trabajo fuentes sin citarlas debidamente.

En Granada a 27 de mayo de 2024

Fdo: Nombre apellidos

Dedicatoria (opcional)
Ver archivo `preliminares/dedicatoria.tex`

Índice general

Agradecimientos

Agradecimientos (opcional, ver archivo `preliminares/agradecimiento.tex`).

Summary

An english summary of the project (around 800 and 1500 words are recommended).

File: `preliminares/summary.tex`

Introducción

De acuerdo con la comisión de grado, el TFG debe incluir una introducción en la que se describan claramente los objetivos previstos inicialmente en la propuesta de TFG, indicando si han sido o no alcanzados, los antecedentes importantes para el desarrollo, los resultados obtenidos, en su caso y las principales fuentes consultadas.

Ver archivo `preliminares/introduccion.tex`

Parte I.

Aproximación a Operadores No Lineales mediante Redes Neuronales

1. Herramientas matemáticas

1.1. Algunas cuestiones básicas de Teoría de la Medida

Definición 1.0.1. Dado un conjunto no vacío Ω, una σ-álgebra en Ω es una familia $\mathcal{A}$ de partes de Ω que contenga a Ω y sea estable por complementación y por unión numerable, esto es, $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ es una σ-álgebra en Ω si verifica las siguientes propiedades:

1. $\Omega \in \mathcal{A}$
2. $E \in \mathcal{A} \to \Omega \backslash E \in \mathcal{A}$
3. Si $E_n \in \mathcal{A}$ para todo $n \in \mathbb{N}$, entonces $\cup_{n\in\mathbb{N}} E_n \in \mathcal{A}$

Un espacio medible es un par $(\Omega, \mathcal{A})$ donde Ω es un conjunto no vacío y $\mathcal{A}$ es una σ-álgebra en Ω. Los elementos de $\mathcal{A}$ se suelen llamar conjuntos medibles.

Definición 1.0.2. Si $(X, \mathcal{T})$ es un espacio topológico, la σ-álgebra engendrada por la topología $\mathcal{T}$ recibe el nombre de σ-álgebra de Borel de X y sus elementos son los conjuntos Borel medibles (o borelianos) de X.

Definición 1.0.3. Sean $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$ espacios medibles. Decimos que una función $f : \Omega_1 \to \Omega_2$ es medible cuando la imagen inversa por f de cualquier conjunto medible en Ω_2 es medible en Ω_1.

Definición 1.0.4. Dado un espacio medible $(\Omega, \mathcal{A})$, una medida en él es una función $\mu : \mathcal{A} \to [0, \infty]$ verificando:

- $\mu(\varnothing) = 0$
- Si $E_n : n \in \mathbb{N}$ es una familia numerable de elementos disjuntos de $\mathcal{A}$ dos a dos, se tiene

$$\mu(\cup_{n\in\mathbb{N}} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$$

.

Definición 1.0.5. Un espacio de medida es una terna $(\Omega, \mathcal{A}, \mu)$ donde Ω es un conjunto, $\mathcal{A}$ una σ-álgebra en Ω y μ una medida definida en $\mathcal{A}$.

Teorema 1.1. *Existe una σ-álgebra $\mathcal{M}$ en $\mathbb{R}^d$ y m una medida definida en $\mathcal{M}$ con las siguientes propiedades:*

- *$\mathcal{M}$ contiene a la σ-álgebra de Borel en $\mathbb{R}^d$.*
- *m es invariante por traslaciones.*
- *Complitud: Si $E \in \mathcal{M}$, $m(E) = 0$ y $\mathcal{A} \subseteq E$, entonces $\mathcal{A} \in \mathcal{M}$.*
- *El par $(\mathcal{M}, m)$ verificando estas propiedades es único.*

Los elementos de $\mathcal{M}$ se llaman conjuntos de Lebesgue medibles en $\mathbb{R}^d$ y m es la medida de Lebesgue en $\mathbb{R}^d$.

Definición 1.1.1. Llamamos función simple positiva a toda función $s : \Omega \to [0, \infty[$ medible cuya imagen sea un subconjunto finito de $\mathbb{R}^+$. Notamos por $\mathcal{S}$ al conjunto de las funciones simples positivas en Ω. Si $s(\Omega) = \alpha_1, ..., \alpha_n$ es una enumeración de los valores que toma s y, para cada $k \in 1, ..., n$, notamos $A_k = s^{-1}(\alpha_k)$, los conjuntos A_k son medibles, forman una partición de Ω y $s = \sum_{k=1}^{n} \alpha_k \chi_{\alpha_k}$. Esta expresión recibe el nombre de descomposición canónica de la función simple positiva s.

Definición 1.1.2. Sea $s \in \mathcal{S}$ y un conjunto medible E, llamamos integral de Lebesgue de s sobre E (con respecto a la medida μ), a

$$\int_E s\mu = \sum_{k=1}^{n} \alpha_k \mu(E \cap A_k) \in [0, \infty[,$$

donde $s = \sum_{k=1}^{n} \alpha_k \chi_{A_k}$ es la descomposición canónica de s. Dicha descomposición es única, salvo el orden de los sumandos, por lo que la definición anterior es correcta.

1.2. Resultados auxiliares de Análisis Matemático

Teorema 1.2 (Teorema de Hausdorff). *Toda biyección lienal entre dos espacios normados de dimensión finita es un isomorfismo.*

Teorema 1.3 (Teorema de derivación bajo el signo integral). *Sea U un subconjunto compacto de $\mathbb{R}^n \times \mathbb{R}^m$, y sea $f : U \to \mathbb{R}$ una función continua en todo U cuya derivada parcial $\frac{\partial f}{\partial y}$ existe y es contínua en U. Entonces, si A y B son sobconjuntos con volumen de $\mathbb{R}^n$ y $\mathbb{R}^m$ respectivamente, tales que B es abierto y $A \times B \subseteq U$, se tiene que la función $F : B \to \mathbb{R}$ definida por*

$$F(y) = \int_A f(x, y) dx$$

es diferenciable en B y

$$F'(y) = \int_A \frac{\partial f}{\partial y}(x, y) dx$$

para todo $y \in B$.

Definición 1.3.1. Un funcional sublineal en un espacio vectorial X es una función $p : X \to \mathbb{R}$ tal que

1. $p(x + y) \leqslant p(x) + p(y)$ para cualesquiera $x, y \in X$.
2. $p(\alpha x) = \alpha p(x)$, $\forall \alpha \geqslant 0$ y $\forall x \in X$.

Teorema 1.4 (Teorema de extensión de Hahn-Banach). *Sea X un espacio vectorial y p un funcional sublineal en X. Si M es un subespacio de X y g es un funcional en M verificando*

$$Reg(m) \leqslant p(m), \qquad (m \in M)$$

entonces existe un funcional lineal f en X cuya restricción a M coincide con g y que verifica

$$Ref(x) \leqslant p(x), \qquad (x \in X)$$

Teorema 1.5 (Teorema de representación de Riesz). *Sea X un espacio de Hausdorff localmente compacto. Sea Λ un funcional lineal positivo en el espacio de funciones continuas en X de soporte compacto, que denotaremos $C_c(X)$. Entonces existe una σ-álgebra M en X que contiene a todos los conjuntos de Borel en X, y existe una medida positiva μ en M que satisface:*

1. *Para todo $K \subseteq X$ compacto, $K \in M$ y $\mu(K) < \infty$.*
2. *Para todo $E \in M$, $\mu(E) = \inf\{\mu(V) : E \subset V, V \text{ abierto}\}$.*
3. *Para todo $E \in M$ tal que $\mu(E) < \infty$, $\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ compacto}\}$.*
4. *Para toda $f \in C_c(X)$, $\Lambda(f) = \int_X f\mu$.*

Teorema 1.6 (Teorema de extensión de Tietze). *Sea X un espacio normado. Sea $f : A \to \mathbb{R}$ una función continua de un subconjunto cerrado A de X en $\mathbb{R}$ con la topología estándar. Entonces, existe una extensión contínua de f a X, esto es, existe una función $F : A \to \mathbb{R}$ continua en todo X con $F(a) = f(a)$ para todo $a \in A$. Además, se puede escoger F tal que $\sup\{|f(a)| : a \in A\} = \sup\{|F(x)| : x \in X\}$. Esto es, si f está acotada entonces F puede escogerse acotada.*

Lema 1.6.1. *Sea K un compacto en R^n, $f \in C(K)$. Entonces, existe una función continua $E(f) \in C(R^n)$) tal que:*

1. $f(x) = E(f)(x), \forall x \in K$
2. $sup_{x\in R^n}\{|E(f)(x)|\} \leqslant sup_{x\in K}\{|f(x)|\}$
3. *Existe una constante c tal que*

$$\sup_{|x'-x''|<\delta}\{|E(f)(x') - E(f)(x'')|\} \leqslant c \sup_{|x'-x''|<\delta}\{|f(x') - f(x'')|\} \quad (x', x'' \in K)$$

Lema 1.6.2 (Teorema de Ascoli-Azela). *V es un conjunto compacto en C(K) si, y solo si:*

1. *V es cerrado en C(K).*
2. *Existe una constante M tal que $\|f(x)\|_{C(K)} \leqslant M, \forall f \in V$.*
3. *V es equicontinuo.*

Lema 1.6.3 (Lema de Riemann-Lebesgue). *Si $f \in L^1(\mathbb{R}^n)$, entonces*

$$\mathcal{F}(f)(\xi) = \int_{\mathbb{R}^n} f(x)e^{i\xi\cdot x}dx \to 0, \text{ cuando } |\xi \to \infty$$

es decir, $\mathcal{F}(f) \in C_0(\mathbb{R}^n)$.

2. Espacios Vectoriales Topológicos

El camino hacia la Teoría de Distribuciones comienza con la presentación del tipo de espacios en el que viven estos objetos: los Espacios Vectoriales Topológicos. Como su nombre indica, estos espacios permiten la definición y estudio de propiedades tanto topológicas como algebraicas. Esta estructura resulta especialmente útil en Análisis Funcional, ya que todos los espacios de funciones tienen estructura de espacio vectorial y en ellos se introducen nociones de convergencia para los elementos del espacio. No es de extrañar, por tanto, que el lugar donde viven las "funciones generalizadas"posea la misma estructura.

2.0.1. Presentación y construcción de EVT

Definición 2.0.1. Sea τ una topología sobre un espacio vectorial X cumpliendo que

- Cada punto de X es un cerrado en (X, τ).
- Las aplicaciones suma y producto por escalares son continuas con respecto a τ.

Entonces, decimos que τ es una topología vectorial en X y X con la topología τ es un Espacio Vectorial Topológico (EVT).

Definición 2.0.2. Diremos que un subconjunto U de un espacio vectorial X es absorbente si $X = \mathbb{R}^+ U$. Diremos que U es equilibrado si $\mathbb{D}U \subseteq U$, donde $\mathbb{D}$ es la bola unitaria en X. Dado un conjunto cualquiera V de X, llamaremos envolvente equilibrada de V al mínimo subconjunto equilibrado de X que contiene a V, esto es, $\mathbb{D}V$.

Observación 2.0.1. Que un conjunto sea equilibrado no implica que sea absorbente, ya que le pueden faltar vectores linelrmente independientes al resto para llegar a cubrir toda la esfera unidad de X. Recíprocamente, si un conjunto es absorbente no tiene por qué ser equilibrado (es decir, no tiene por qué ser invariante por rotaciones).

Definición 2.0.3. Sean X un espacio vectorial y $E \subseteq X$ absorvente. Definimos el Funcional de Minkowski de E, $v_E : X \to [0, \infty[$ como :

Definición 2.0.4. Una familia $\mathcal{B}$ de subconjuntos no vacíos de un conjunto X es una base de filtro si verifica que para cada $U, V \in \mathcal{B}$ existe algún $W \in \mathcal{B}$ tal que $W \subset U \cap V$.

Teorema 2.1 (Caracterización de las bases de entornos de cero en un EVT)**.** *Todo entorno de cero U en un EVT es absorbente y contiene un entorno de cero equilibrado V tal que $V + V \subset U$. Recíprocamente, sea X un espacio vectorial y $\mathcal{B}$ una base de filtro en X formada por conjuntos absorbentes y equilibrados, verificando que para cada $U \in \mathcal{B}$ existe $V \in \mathcal{B}$ tal que $V + V \subset U$. Entonces $\mathcal{B}$ es base de entornos de cero para una (única) topología vectorial en X.*

Proposición 2.1.1. *Sea X un EVT. Entonces, los operadores traslación $T_a(x) = a + x$, $\forall a, x \in X$ y para cada $\lambda \neq 0$ los operadores multiplicación $M_\lambda(x) = \lambda x$, $\forall x \in X$ son homeomorfismos de X. Esto es, todas las traslaciones, giros y homotecias son homeomorfismos en X.*

Como consecuencia de la proposición anterior, tomando una base de entornos de cero $\mathcal{B}$ en X, $\{x + \mathcal{B}\}$ será una base de entornos para cada $x \in X$. Por tanto, una topología vectorial queda totalmente determinada por una base de entornos de cero.

Teorema 2.2 (Teorema de Tihonov). *Si X e Y son EVT separados de dimensión finita, toda biyección lineal de X sobre Y es un isomorfismo.*

Lema 2.2.1. *Lema de Riesz Sea X un EVT, M un subespacio cerrado y A un subconjunto acotado de X. Supongamos que existe un $\lambda \in \mathbb{K}$ con $|\lambda| < 1$ tal que $A \subset M + \lambda A$. Entonces $A \subset M$.*

Teorema 2.3 (Teorema de Riesz). *Sea X un EVT separado. Entonces equivalen:*

1. *X es localmente compacto.*
2. *Existe en X un entorno de cero compacto.*
3. *Existe en X un entorno de cero precompacto.*
4. *La dimensión de X es finita.*

Demostración.

- 4. ⇒ 1. Como X tiene dimensión finita, como consecuencia del Teorema de Tihonov sabemos que su topología es equivalente a la topología inducida por la norma euclídea en la misma dimensión de X. Como la topología inducida por la norma euclídea es localmente compacta, la topología de X también lo será.
- 1. ⇒ 2. ⇒ 3. Son evidentes.
- 3. ⇒ 4. Sea U un entorno de cero precompacto en X. Entonces, por definición, existirá un subconjunto finito F de X tal que, fijando $\lambda = \frac{1}{2}$, $U \subset F + \lambda U$. Si M es el subespacio de X generado por F, M tiene dimensión finita y, por tanto, es cerrado en X. Como U está contenido en $M + \lambda U$ y es acotado (por tanto, precompacto), el Lema de Riesz nos dice que $U \in M$. Como sabemos además que U es absorbente podemos deducir que $X = M$, por lo que X tiene dimensión finita.

□

Como consecuencia de este teorema, podemos deducir que todo subconjunto precompacto de un EVT de dimensión finita tiene interior vacío. Esto tiene un gran impacto cuando trabajamos en espacios de funciones, por lo que volveremos a ello más adelante.

2.0.2. Convergencia uniforme y complitud

Las topologías vectoriales no son metrizables en general. Por tanto, las sucesiones pueden no ser suficientes para caracterizar la topología. Sin embargo, generalizando el concepto de sucesión, es posible manejar algunos aspectos de topologías generales como si fuesen metrizables.

Definición 2.3.1. Un conjunto dirigido es un conjunto no vacío Λ dotado de un preorden, $\leqslant$, verificando que para cualesquiera $\lambda_1, \lambda_2 \in \Lambda$, existe un $\lambda \in \Lambda$ tal que $\lambda_1 \leqslant \lambda$ y $\lambda_2 \leqslant \lambda$.
Si X es un conjunto no vacío, una red de elementos de X es una aplicación $\phi : \Lambda \to X$, donde Λ es un conjunto dirigido. De manera análoga a la notación para sucesiones, si $\phi : \Lambda \to X$ es una red, escribimos $x_\lambda = \phi(\lambda)$, $(x_\lambda)_{\lambda \in \Lambda} = \phi$.

Definición 2.3.2. Sea X un espacio topológico. Decimos que una red $(x_\lambda) \subset X$ converge a $x \in X$ y escribimos $(x_\lambda) \to x$ si para todo entorno U de x puede encontrarse un índice $\lambda_0 \in \Lambda$ tal que

$$\lambda \in \Lambda, \quad \lambda_0 \leqslant \lambda \qquad \Rightarrow \qquad x_\lambda \in U.$$

Definición 2.3.3. Decimos que x es un valor adherente a la red (x_λ) si para todo entorno U de x y para todo $\lambda \in \Lambda$ puede encontrarse $\mu \in \Lambda$ tal que $\lambda \leqslant \mu$ y $x_\mu \in U$. Es claro que si $(x_\lambda) \to x$, x es un valor adherente a la red (x_λ), no siendo cierto el recíproco.

Sea $\phi = (x_\lambda)_{\lambda\in\Lambda}$ una red en X. Sea $A(\phi) = \{A_\lambda : \lambda \in \Lambda\}$ la familia de subconjuntos de X definida por

$$A_\lambda = \{x_\mu : \mu \in \Lambda, \lambda \leqslant \mu\} \qquad (\lambda \in \Lambda)$$

Entonces A_λ cumple las siguientes condiciones:

1. $A_\lambda \neq \emptyset$, $\forall \lambda \in \Lambda$.
2. $\forall \lambda_1, \lambda_2 \in \Lambda$, existe $\lambda \in \Lambda$ tal que $A_\lambda \subset A_{\lambda_1} \bigcup A_{\lambda 2}$ (Bastará que $\lambda_1 \leqslant \lambda$ y $\lambda_2 \leqslant \lambda$).

Si X es además un espacio topológico, podemos caracterizar la posible convergencia de ϕ en términos de la familia $A(\phi)$:

1. Dado $x \in X$, ϕ converge a x si, y solo si, cada entorno de x contiene a un elemento de $A(\phi)$.
2. x es valor adherente a la red ϕ si, y solo si, cada entorno de x tiene intersección no vacía con cada elemento de $A(\phi)$.
3. El conjunto de los valores adherentes a ϕ coincide con $\bigcup_{A\in A(\phi)} \bar{A}$.

Proposición 2.3.1. *Sea X un espacio topológico, $A \subset X$ y $x \in X$. Las siguientes afirmaciones son equivalentes:.*

1. *$x \in \bar{A}$.*
2. *Existe una red en A que converge a x.*
3. *x es un valor adherente a una red en A.*

Proposición 2.3.2. *Sea X un espacio topológico Hausdorff y (x_λ) una red en X que converge a $x \in X$. Entonces, x es el único valor adherente a la red (x_λ). Recíprocamente, si toda red en X converge, a lo sumo, a un punto de X, entonces X es un espacio de Hausdorff.*

Proposición 2.3.3. *Sean X e Y espacios topológicos, $f : X \to Y$ una función y $x \in X$. Las siguientes afirmaciones son equivalentes:*

1. *f es continua en x.*
2. *$(x_\lambda) \to x \Rightarrow f(x_\lambda) \to f(x)$.*

Definición 2.3.4. Se dice que una red $(x_\lambda)_{\lambda\in\Lambda}$ de elementos de un EVT X es una red de Cauchy si para cada entorno de cero U puede encontrarse un índice $\lambda_0 \in \Lambda$ tal que,

$$\lambda, \mu \in \Lambda, \lambda_0 \leqslant \lambda, \lambda_0 \leqslant \mu \quad \Rightarrow \quad x_\lambda - x_\mu \in U$$

equivalentemente, $B_{\lambda_0} - B_{\lambda_0} \subset U$, donde $B_{\lambda_0} = \{x_\lambda : \lambda \in \Lambda, \quad \lambda_0 \leqslant \lambda\}$

Lema 2.3.1. *Toda red de Cauchy en un EVT que posea un valor adherente, converge a dicho valor adherente.*

Definición 2.3.5. Sea X un EVT y d una semidistancia. Decimos que d es invariante por traslaciones cuando cumple:

$$d(x+z, y+z) = d(x, y) \qquad (x, y, z \in X).$$

Proposición 2.3.4. *Sea X un EVT y d una semidistancia completa que genera una topología de X y es invariante por traslaciones. Entonces, toda red de Cauchy en X es convergente.*

Definición 2.3.6. Se dice que un EVT X es completo cuando toda red de Cauchy en X es convergente. Si ocurre solamente que toda sucesión de Cauchy es convergente, decimos que X es secuencialmente completo. Análogamente se definen la complitud y la complitud secuencial de un subconjunto de un EVT.

Proposición 2.3.5. *Sea X un EVT cuya topología proviene de una semidistancia d invariante por traslaciones. Entonces, son equivalentes:*

1. *X es completo.*
2. *X es secuencialmente completo.*
3. *La distancia d es completa.*

2.0.3. Aplicaciones lineales entre EVT

Definición 2.3.7. Sean X e Y EVT y sea $F : X \to Y$ una aplicación. Decimos que F es uniformemente continua si para cada entorno de cero V en Y existe un entorno de cero U en X tal que $F(x) - F(y) \in V$, para todos $x, y \in X$ tales que $x - y \in U$.

Proposición 2.3.6. *Sean X e Y dos EVT y $T : X \to Y$ una aplicación lineal. Son equivalentes:*

- *T es uniformemente continua.*
- *T es continua .*
- *T es continua en cero, esto es, $T^{-1}(V)$ es entorno de cero en X para cada entorno de cero V en Y.*

2.0.4. Topologías iniciales

Definición 2.3.8. Sea $X \neq 0$ un conjunto, $\{(X_i, \mathcal{T}_i), i \in I\}$ espacios topológicos, $\{f_i : i \in I\}$ una familia de aplicaciones definidas en X, cada una tomando valores en un espacio topológico X_i respectivamente. Llamamos topología inicial para $\{f_i : i \in I\}$ a la mínima topología en X que hace continuas a todas las aplicaciones f_i.

Proposición 2.3.7. *Sea X un espacio vectorial, $\{X_i : i \in I\}$ una familia de EVT y, para cada $i \in I$, sea f_i una aplicación lineal de X en X_i. Entonces, la topología inicial en X para la familia $\{f_i : i \in I\}$ es una topología vectorial en X. Si, para cada $i \in I$, $\mathcal{B}_i$ es una base de entornos de cero en X_i, la familia*

$$\mathcal{B} = \left\{ \bigcup_{j \in J} f_j^{-1}(U_j) : \quad J \subset I, J \text{ finito}, \ U_j \in \mathcal{B}_j \ \forall j \in J \right\}$$

es base de entornos de cero en X para dicha topología inicial.

2.0.5. EVT Localmente Convexos

Definición 2.3.9. Sea X un espacio vectorial. Una topología localmente convexa en X es una topología vectorial que admite una base de entornos de cero convexos. Un espacio localmente convexo (ELC) es un par $(X, \mathcal{T})$ formado por un espacio vectorial X y una topología localmente convexa $\mathcal{T}$ en X. Si no hay lugar a confusión, diremos que X es un ELC.

Teorema 2.4 (Caracterización de los entornos de cero y de las bases de entornos de cero en un ELC)**.** *Todo ELC posee una base de entornos de cero formada por conjuntos equilibrados y convexos.*

Observación 2.4.1. Consecuencias directas:

- En un EVT, el cierre y el interior de un conjunto convexo son conjuntos convexos.
- En cualquier EVT, la envolvente convexa de un conjunto abierto es abierta.

Definición 2.4.1. Sea X espacio vectorial, decimos que una aplicación $\nu : X \to \mathbb{R}$ es una pseudonorma cuando cumple:

1. $\nu(x+y) \leqslant \nu(x) + \nu(y) \quad \forall x, y \in X.$
2. $\lambda \in \mathbb{D} \Rightarrow \nu(\lambda x) \leqslant \nu(x) \quad \forall x \in X.$
3. $\lim_{n\to\infty} \nu(\frac{x}{n}) = 0 \quad \forall x \in X.$

Por otro lado, una aplicación $p : X \to \mathbb{R}$ es una seminorma en X cuando:

1. $p(x+y) \leqslant p(x) + p(y) \quad \forall x, y \in X.$
2. $p(\lambda x) \leqslant |\lambda| p(x) \quad \forall \lambda \in \mathbb{K}, \quad \forall x \in X.$

De la definición podemos ver cómo toda seminorma es una pseudonorma.

El siguiente resultado refleja lo relacionadas que estan las seminormas con la convexidad local en los EVT:

Teorema 2.5 (Caracterización de las topologías localmente convexas)**.** *Una topología $\mathcal{T}$ en un espacio vectorial X es localmente convexa si, y solo si, es la topología asociada a una familia de seminormas en X. Más concretamente, sea X un ELC y sea $\mathcal{B}$ una familia de entornos de cero convexos y equilibrados. Si para cada $U \in \mathcal{B}$, ϕ_U es el funcional de Minkowski de U, la topología de partida en X es la asociada a la familia de seminormas $\{\phi_U : U \in \mathcal{B}\}$.*

Corolario 2.5.1. *Todo ELC separado es isomorfo a un subespacio de un producto de espacios normados.*

2.0.6. EVT Metrizables

Decimos que un espacio topológico $(X, \mathcal{T})$ es metrizable si es homeomorfo a un espacio métrico, esto es, si existe una distancia en X que genere su topología. Por tanto, en el contexto de los EVT, es natural preguntarse en qué casos dicha topología es vectorial y, análogamente, cuándo podemos metrizar o asociar una distancia a una topología vectorial dada. En esta sección introduciremos formalmente el concepto de metrizabilidad en espacios topológicos y daremos respuesta a las dos cuestiones planteadas.

Definición 2.5.1. Sean X un espacio vectorial y ν una pseudonorma en X. Definiendo para cada $\varepsilon > 0$, $U_\varepsilon = \{x \in X : \nu(x) \leqslant \varepsilon\}$ tenemos que $\{U_\varepsilon : \varepsilon > 0\}$ es una base de entornos de cero para la topología vectorial en X asociada a la pseudonorma ν, con la que X es un EVT pseudonormable (seminormable cuando ν es una seminorma).

Definición 2.5.2. Un espacio topológico X es semimetrizable si existe una semidistancia d en X que genera su topología. Si la semidistancia d es una distancia, decimos que el espacio topológico X es metrizable.

Observación 2.5.1. La topología asociada a una pseudonorma es siempre semimetrizable, esto es, existe una semidistancia que genera la topología. Concretamente, si $\nu : X \to \mathbb{R}$ es una pseudonorma, la aplicación $d : X \times X \to \mathbb{R}$ dada por $d(x, y) = \nu(x - y)$, es una semidistancia que genera la topología.

La primera cuestión es agenciada por el resultado:

Teorema 2.6. *Sea X un espacio vectorial y sea d una semidistancia en X verificando:*

1. *d es invariante por traslaciones.*
2. *Si (λ_n) es una sucesión convergente a cero en $\mathbb{K}$, entonces $d(\lambda_n x, 0) \to 0$ para cada $x \in X$.*
3. *Si (x_n) es una sucesión en X tal que $d(x_n, 0) \to 0$, entonces $d(\lambda x_n, 0) \to 0$ para cada $\lambda \in \mathbb{K}$.*

Entonces la topología $\mathcal{T}$ asociada a d es una topología vectorial. Además, $(X, \mathcal{T})$ es completo si, y sólo si, d es completa.

Demostración. □

La segunda problemática planteada se resuelve en el siguiente resultado:

Teorema 2.7 (Criterio de metrizabilidad de Birkhoff-Kakutani)**.** *Si X es un EVT, equivalen:*

1. *X es pseudonormable.*
2. *X es semimetrizable.*
3. *X tiene una base numerable de entornos de cero.*

Es decir, un EVT es metrizable si, y solo si, es separado y tiene una base numerable de entornos de cero.

Ya tenemos todas las herramientas necesarias para presentar los dos siguientes conceptos, de los que haremos uso durante el trabajo:

Definición 2.7.1. Llamamos F-espacio a un EVT completo metrizable. Si X tiene la topología asociada a una pseudonorma ν, X es un F-espacio cuando toda serie absolutamente convergente es convergente:

$$x_n \in X \ \forall n \in \mathbb{N}, \ \sum_{n=1}^{\infty} \nu(x_n) < \infty \Rightarrow \sum_{n \geqslant 1} x_n < \infty$$

Definición 2.7.2. Llamamos espacio de Fréchet a un F-espacio localmente convexo. Esto es, un espacio localmente convexo que es completo y metrizable.

2.1. Teoría de dualidad

Definición 2.7.3. Dado un espacio vectorial X, denotaremos como $X^{\#}$ a su dual algebraico, esto es, al espacio formado por los funcionales lineales en X. Si (X, τ) es un EVT, definimos su dual topológico $(X, \tau)^*$ como el subespacio de $X^{\#}$ formado por los funcionales lineales y τ-continuos en X. Si τ se sobrentiende, escribimos X^*.

Corolario 2.7.1. *Sea X un espacio vectorial, U un subconjunto absorbente y convexo de X y $x_0 \in X$ tal que $x_0 \notin U$. Existe un funcional lineal f en X tal que*

$$\mathit{Re}\, f(x) \leqslant 1 \quad (x \in U) \qquad y \qquad \mathit{Re}\, f(x_0) \geqslant 1.$$

Demostración. Sea μ el funcional de Minkowski en U. Entonces, μ es sublineal en X y se tiene que

$$\{x \in X : \mu(x) < 1\} \subset U \subset \{x \in X : \mu(x) \leqslant 1\}$$

por lo que $\mu(x_0) \geqslant 1$ y $\lambda\mu(x_0) \leqslant \mu(\lambda x_0)$ se cumple para todo $\lambda > 0$ por la homogeneidad de μ. Por tanto, tomando el subespacio $\mathbb{R}\{x_0\}$ de X y el funcional lineal que a cada $\lambda \in \mathbb{R}$ le asigna $g(\lambda) = \lambda\mu(x_0)$, definido en $\mathbb{R}\{x_0\}$, el Teorema de Hahn-Banach nos proporciona un funcional $f \in X^{\#}$ que cumple

$$f(x) \leqslant \mu(x) \quad (x \in X) \qquad y \qquad f(x_0) = \mu(x_0).$$

□

Corolario 2.7.2. *Sea X un EVT. Las siguientes afirmaciones son equivalentes:*

1. *X^* separa los puntos de X.*
2. *Para cada $x_0 \in X \backslash \{0\}$ existe una seminorma continua ϕ en X tal que $\psi(x_0) \neq 0$.*
3. *La intersección de todos los entornos convexos de cero en X se reduce a $\{0\}$.*

Demostración.

- $(i) \Rightarrow (ii)$
  Si X^* separa los puntos de X, fijando un $x_0 \in X\backslash\{0\}$ y un $f \in X^*$ (el cual, por ser lineal y continuo, cumplirá que $f(x_0) \neq 0$), podemos definir la seminorma continua $\varphi = |f|$, que verifica $\varphi(x_0) \neq 0$.
- $(ii) \Rightarrow (iii)$ Sea ψ la seminorma definida en (ii). Entonces, los entornos de cero convexos asociados a la topología de la seminorma vendrán dados por $U_{x_0} = \{x \in X : \psi(x) < \psi(x_0)\}$ y cumplirán que $x_0 \notin U_{x_0}$, para cada $x_0 \in X\backslash\{0\}$. Por tanto, $\bigcap_{x_0 \in X\backslash\{0\}} U_{x_0} = \{0\}$.
- $(iii) \Rightarrow (i)$ El corolario anterior nos proporciona, para cada $x_0 \in X\backslash\{0\}$, un funcional $f \in X^*$ cumpliendo que $f(x_0) = \mu(x_0) \neq 0$.

□

Corolario 2.7.3. *Sea X un ELC. Entonces equivalen:*

1. *X^* separa los puntos de X.*
2. *X es separado.*

Demostración. En un ELC todos los entornos de cero son localmente convexos, por lo que la equivalencia entre (i) y (iii) del corolario anterior se traduce en este resultado. □

Teorema 2.8. *Sean A y B subconjuntos no vacíos, convexos, disjuntos, de un ELC X tales que A es cerrado y B es compacto. Entonces, existen $f \in X^*$ y $\alpha, \beta \in \mathbb{R}$ tales que*

$$Ref(a) \leqslant \alpha < \beta \leqslant Ref(B) \qquad (a \in A \ ,b \in B).$$

Observación 2.8.1. Si tomamos $B = \{x\}$ donde $x \in X$
A, el teorema anterior nos muestra que f separa puntos de X.

2.2. Topología débil y débil-*

La compacidad es una propiedad muy útil y deseable pero difícil de garantizar en espacios normados de dimensión infinita pues, como consecuencia del Teorema de Riesz, cualquier compacto en tales espacios tiene interior vacío. Por tanto, la topología de la norma presenta pocos compactos.

Por esta razón, en dichos espacios conviene trabajar con topologías más pequeñas que, preservando la continuidad de funcionales del dual, presenten más conjuntos compactos. Estas topologías se conocen como la topología débil y débil-*.

Definición 2.8.1. Sea X un espacio vectorial, Y un subespacio del dual algebraico $X^\#$ que separa los puntos de X. Entonces (X, Y) es un par dual. Escribimos $\langle x, y \rangle$ para denotar la acción de $y \in Y \leqslant X^\#$ sobre $x \in X$.

Como X separa los puntos de Y y $X \leqslant Y^\#, (Y, X)$ es par dual.

Definición 2.8.2. Sea (X, Y) par dual. La topología inicial en X para los elementos de Y se denota por $\sigma(X, Y)$ y se denomina la topología débil en X asociada al par dual (X, Y). Se trata de una topología localmente convexa y separada en X, asociada a la familia de seminormas

$$\varphi_y(x) = |\langle x, y \rangle| \quad (x \in X, y \in Y).$$

Los conjuntos de la forma

$$U(J, \varepsilon) = \{x \in X : |\langle x, y \rangle| \leqslant \varepsilon \quad \forall y \in J\}$$

donde J es un subconjunto finito de Y y $\varepsilon > 0$ forman una base de entornos de cero para $\sigma(X, Y)$.

Proposición 2.8.1. *Sea (X, Y) un par dual.*

- *Un funcional lineal f en X es $\sigma(X, Y)$-continuo si, y sólo si, existe un $y_0 \in Y$ tal que*

$$f(x) = \langle x, y_0 \rangle, \quad x \in X.$$

- *La topología $\sigma(X, Y)$ es compatible con el par dual (X, Y), esto es, $(X, \sigma(X, Y))^* = Y$; y esta es la mínima topología en X con esa propiedad.*

Definición 2.8.3. Sea X un ELC separado, (X, X^*) par dual.

- $\sigma(X, X^*)$ es la topología débil de X.
- $\sigma(X^*, X)$ es la topología débil-* de X^* asociada al par dual (X^*, X)

Observación 2.8.2. Se trata de la menor topología en X^* que hace continuos a los elementos de X. En particular, $(X^*, \sigma(X^*, X))^* = X$.

Observación 2.8.3. Se trata de la topología en X^* de la convergencia puntual sobre los elementos de X. Esto es, la topología en X^* de la convergencia uniforme sobre los subconjuntos finitos de X.

Observación 2.8.4. Sea X ELC. En general, es posible que no tengamos en X^* ninguna topología destacable. En ese caso, sabemos que al menos dispondremos de la topología débil-*.

2.3. Espacio de funciones Test

Definición 2.8.4. Sea Ω un conjunto abierto no vacío de $\mathbb{R}^n$. Llamamos multi-índice a la n-tupla $\alpha = (\alpha_1, ..., \alpha_n)$ de enteros no negativos. A cada multi-índice α, podemos asociar un operador diferencial

$$D^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}$$

cuyo orden es $|\alpha| = \alpha_1 + \cdots + \alpha_n$

Observación 2.8.5. Cuando $|\alpha| = 0$, tenemos que $D^\alpha f = f$.

Definición 2.8.5. Una función f definida en $\Omega \subset \mathbb{R}^n$ pertenece a $C^\infty(\Omega)$ si cumple que $D^\alpha f \in C(\Omega)$ para todo multi-índice α.

Definición 2.8.6. Sea K un compacto en $\mathbb{R}^n$. Llamamos $\mathcal{D}(K)$ al espacio de todas las $f \in C^\infty(\mathbb{R}^n)$ cuyo soporte queda contenido en K. Si $K \in \Omega$, entonces $\mathcal{D}(K)$ se puede identificar con un subespacio de $C^\infty(\Omega)$.

A continuación veremos como, dotado de la topología adecuada, $C^\infty(\Omega)$ es un espacio de Fréchet con la propiedad de Heine-Borel y, por tanto, $\mathcal{D}(K)$ es un subespacio cerrado suyo, siempre que $K \subset \Omega$.

Para cada $N \in \mathbb{R}$, definimos la seminorma

$$p_N(f) = \max\{|D^\alpha f(x)| : x \in K_N, |\alpha| \leqslant N\}$$

siendo K_i conjuntos compactos que cumplen

- $K_i \subset int(K_{i+1})$.
- $\Omega = \bigcup K_i$.

Lema 2.8.1. *La familia de seminormas p_N definen una topología metrizable y localmente convexa en $C^\infty(\Omega)$. Para esta topología, los conjuntos*

$$V_N = \left\{f \in C^\infty(\Omega) : p_N(f) < \frac{1}{N}\right\}$$

con $N \in \mathbb{N}$ forman una base local.

Observación 2.8.6. Para cada $x \in \Omega$, el funcional $f \to f(x)$ es continuo para dicha topología.

Lema 2.8.2. *$C^\infty(\Omega)$ es un espacio de Fréchet.*

Lema 2.8.3. *$C^\infty(\Omega)$ tiene la propiedad de Heine-Borel.*

Definición 2.8.7. Consideramos ahora el espacio vectorial $\mathcal{D}(\Omega) = \bigcup \mathcal{D}_K$. Entonces, $\phi \in \mathcal{D}(\Omega)$ si, y solo si, $\phi \in C^\infty(\Omega)$ y $\mathrm{supp}(\phi) = K$, para algún subconjunto compacto K de Ω.

La topología τ_K heredada del espacio de Fréchet $C^\infty(\Omega)$ para cada $\mathcal{D}_K$ no es completa y por tanto tampoco lo será en $\mathcal{D}(\Omega)$.Por tanto, buscaremos definir una topología τ en $\mathcal{D}(\Omega)$ que sí lo sea.

Definición 2.8.8.

- Llamaremos β al conjunto de $W \subset \mathcal{D}(\Omega)$ tales que $\mathcal{D}_K \bigcap W \in \tau_K$ para todo compacto $K \in \Omega$.
- $\tau = \bigcup \phi + W$ tales que $\phi \in \mathcal{D}(\Omega)$ y $W \in \beta$.

Teorema 2.9.

1. *τ es una topología en $\mathcal{D}(\Omega)$ y β es una base local para esta topología.*
2. *$\mathcal{D}(\Omega)$ dotado de la topología τ es un EVT localmente convexo.*

Demostración.

1. Supongamos que $V_1, V_2 \in \tau$, $\phi \in V_1 \bigcup V_2$. Entonces, atendiendo a la definición de β, para demostrar 1. bastará con probar que existe $W \in \beta$ tal que $\phi + W \subseteq V_1 \bigcup V_2$.

   Como $V_1, V_2 \in \tau$ y atendiendo a la construcción de τ, sabemos que existen $\phi_i \in \mathcal{D}\Omega$ y $W_i \in \beta$ tales que $\phi \in \phi_i + W_i \subset V_i$, para $i = 1, 2$.

   Tomando un K lo suficiéntemente grande como para que $\phi_1, \phi_2 \in \mathcal{D}_K$, tendremos que $\mathcal{D}_K \bigcup W_i$ es abierto para la topología τ_K y, por tanto $\phi - \phi_i \in (1 - \delta_i)W_i$ para algún $\delta_i > 0$. Como escogimos W_i convexo, tenemos que $\phi - \phi_i + \delta_i W_i \subset (1 - \delta_i)W_i + \delta_i W_i = W_i$ y por tanto $\phi + \delta_i W_i \subset \phi_i + W_i \subset V_i$, para $i = 1, 2$.

   Por tanto, hemos encontrado el $W = (\delta_1 W_1) \bigcup (\delta_2 W_2) \in \beta$ que necesitábamos.

2. Tomemos ahora ϕ_1, $\phi_2 \in \mathcal{D}(\Omega)$, y sea

   $$W = \{\phi \in \mathcal{D}(\Omega) : \quad \| \phi \|_0 < \| \phi_1 - \phi 2 \|_0\}$$

   donde $\| \cdot \|_0$ es la norma definida en esta sección. Entonces $W \in \beta$ y $\phi_1 \notin \phi_2 + W$. Por tanto, $\{\phi_1\}$ es un conjunto cerrado para la topología τ.
   Como estamos trabajando sobre un EVT, sabemos que la suma es τ-continua y, como cada $W \in \beta$ es convexo, tenemos que

   $$(\psi_1 + \frac{1}{2}W) + (\psi_2 + \frac{1}{2}W) = (\psi_1 + \psi_2) + W$$

para todas $\psi_1, \psi_2 \in \mathcal{D}(\Omega)$ siempre que $\psi_1 \neq \psi_2$. Tomemos ahora un escalar α_0 y un $\phi_0 \in \mathcal{D}(\Omega)$. Entonces, para toda $\phi \in \mathcal{D}(\Omega)$ y todo escalar α se cumplirá:

$$\alpha\phi - \alpha_0\phi_0 = \alpha(\phi - \phi_0) + (\alpha - \alpha_0)\phi_0.$$

Además, si $W \in \beta$ existirá un $\delta > 0$ tal que $\delta\phi_0 \in \frac{1}{2}W$ y podremos escoger una constante c que cumpla $3c(|\alpha_0 + \delta) = 1$. Como W es balanceado y convexo, tenemos que

$$\alpha\phi - \alpha_0\phi_0 \in W$$

siempre que $\phi - \phi_0 \in cW$ y $|\alpha - \alpha_0| < \delta$.

□

Teorema 2.10.

1. *Un conjunto convexo y equilibrado en $\mathcal{D}(\Omega)$ es abierto si, y solo si, pertenece a β.*
2. *La topología τ_K de cada $\mathcal{D}_K \subset \mathcal{D}(\Omega)$ coincide con la topología τ restringida a $\mathcal{D}_K$.*
3. *Sea E un subconjunto acotado de $\mathcal{D}(\Omega)$. Entonces existe un $K \in \Omega$ tal que $E \in \mathcal{D}_K$. Además, existen constantes $M_N < \infty$ tales que $\| \phi \|_N \leqslant M_N$, $\forall \phi \in E$, $\forall N \in \mathbb{N}$.*
4. *$\mathcal{D}(\Omega)$ tiene la propiedad de Heine-Borel.*
5. *En $\mathcal{D}(\Omega)$, toda sucesión de Cauchy es convergente.*

3. Teoría de Distribuciones

3.1. Distribuciones

La Teoría de Distribuciones, desarrollada por L. Schwartz durante las décadas de 1940 y 1950, surgió como respuesta a la necesidad de manejar "funciones generalizadas"desde varias ramas de la Física, donde el concepto clásico de función como observable físico presentaba limitaciones de precisión. En aquel momento ya se utilizaban formalismos como la δ de Dirac que funcionaban y daban respuesta a esta problemática sin haber sido aún integrados dentro de una teoría matemática que les diese respaldo. Planteada esta necesidad, se inició la búsqueda de una clase de objetos que extendiesen la noción de función y sobre la que se pudiese generalizar el cálculo diferencial clásico.

Dicha clase de objetos tendrá como ojetivo albergar a todas las funciones continuas, permitir que cada objeto perteneciente a ella tenga derivadas parciales, y que estas de nuevo sean un objeto de la misma clase y que dicho concepto de derivada case con el concepto de derivada y sus propiedades existentes para funciones diferenciables.

Una primera aproximación intuitiva puede exponerse tomando una función $f : \mathbb{R} \to \mathbb{K}$ a la que únicamente le exigimos ser localmente integrable. El salto a la visión distribucional consiste en ver esta función como un operador $\phi \to \int f\phi$, que asocia a cada "función test"ϕ de soporte compacto el valor $\int f\phi$, esto es, que nos permita asomarnos a los valores que toma f en el soporte de ϕ.

Durante las siguientes secciones expondremos los contenidos necesarios para llegar a una definición rigurosa del concepto de distribución y posteriormente presentaremos los resultados principales de esta teoría.

3.1.1. Definición

Definición 3.0.1. Una distribución es un funcional lineal en $\mathcal{D}(\Omega)$ que es continuo con respecto a la topología τ. Denotaremos $\mathcal{D}'(\Omega)$ al espacio de las distribuciones.

Teorema 3.1. *Sea Λ un funcional lineal sobre $\mathcal{D}(\Omega)$. Equivalen:*

1. $\Lambda \in \mathcal{D}'(\Omega)$.

2. *Para cada compacto $K \subset \Omega$, podemos contrar $N \in \mathbb{N}$, $C < \infty$ tales que $|\Lambda\phi| \leqslant C \parallel \phi \parallel_N$, $\forall \phi \in \mathcal{D}_K$.*

Observación 3.1.1. Cada $x \in \Omega$ determina una distribución $\delta_x \in \mathcal{D}'(\Omega)$ dada por $\delta_x(\phi) = \phi(x)$.

3.1.2. Funciones vistas como distribuciones

Sea $f : \Omega \to \mathbb{K}$ localmente integrable, K compacto en Ω. Definimos

$$\Lambda_f(\phi) = \int_\Omega \phi(x) f(x) dx \qquad (\phi \in \mathcal{D}(\Omega)).$$

Entonces, se cumple que

$$|\Lambda_f(\phi)| = \left| \int_\Omega \phi(x) f(x) dx \right| \leqslant \left(\int_K |f| \right) \cdot \| \phi \|_0$$

siempre que $\phi \in \mathcal{D}_K$. Por tanto, $\Lambda_f \in \mathcal{D}'(\Omega)$. A la distribución Λ_f la llamaremos distribución asociada a f y diremos que este tipo de distribuciones son "funciones".

De manera similar, si μ es una medida de Borel en Ω, para cada compacto $K \in \Omega$, la igualdad

$$\Lambda_\mu(\phi) = \int_\Omega \phi d\mu \qquad (\phi \in \mathcal{D}(\Omega))$$

define una distribución Λ_μ a la que llamaremos distribución asociada a μ.

3.2. Cálculo con distribuciones

En esta sección introduciremos algunos de los resultados que definen y demuestran la compatibilidad de las distribuciones con el cálculo diferencial clásico.

3.2.1. Multiplicación por funciones

Definición 3.1.1. Sean $\Lambda \in \mathcal{D}'(\Omega)$ y $f \in C^\infty(\Omega)$. Definimos la distribución $f\Lambda \in \mathcal{D}'(\Omega)$ dada por

$$(f\Lambda)(\phi) = \Lambda(f\phi)$$

para cada $\phi \in \mathcal{D}(\Omega)$.

3.2.2. Derivadas de una distribución

Definición 3.1.2. Para cada multi-índice α, de aquí en adelante denotaremos:

$$D^\alpha = \left(\frac{\partial}{\partial x_1} \right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n} \right)^{\alpha_n}.$$

Definición 3.1.3. Sea α un multi-índice y sea $\Lambda \in \mathcal{D}'(\Omega)$. Para cada $\phi \in \mathcal{D}(\Omega)$ definimos la distribución derivada α-ésima de Λ mediante

$$(D^\alpha \Lambda)(\phi) = (-1)^{|\alpha|} \Lambda(D^\alpha \phi).$$

Para este nuevo objeto definido, es inmediato comprobar que:

Definición 3.1.4. En particular, si f es una función localmente integrable en Ω, se define la distribución derivada α-ésima de f mediante $D^\alpha f = D^\alpha \Lambda_f$, esto es,

$$\langle D^\alpha f, \phi \rangle = (-1)^{|\alpha|} \int_\Omega f D^\alpha \phi d\lambda \qquad (\phi \in \mathcal{D}(\Omega)).$$

Observación 3.1.2. Si la función f tiene derivadas parciales hasta orden k en Ω, la derivada distribucional $D^\alpha f$ coincide con la distribución asociada a la derivada parcial clásica de orden α de f, para cada multi-índice α con $|\alpha| \leqslant k$. Si las derivadas parciales de f hasta orden k son continuas, se llega a la igualdad como consecuencia del Teorema de Fubini y de la fórmula de integración por partes. En otro caso, la igualdad no podrá darse.

Proposición 3.1.1. *Para cada multi-índice α y β:*

1. $D^\alpha : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ *es un operador lineal y continuo.*
2. $D^\alpha D^\beta \Lambda = D^{\alpha+\beta}\Lambda = D^\beta D^\alpha \Lambda$, $\forall \Lambda \in \mathcal{D}'(\Omega)$.

Demostración.

1. Dada una distribución Λ, para cada compacto K existen $n \in \mathbb{N}$ y $C > 0$ tales que

$$|\langle \Lambda, \phi \rangle| \leqslant C \parallel \phi \parallel_n \qquad (\phi \in \mathcal{D}(K)).$$

   Entonces, tenemos:

$$|\langle D^\alpha \Lambda, \phi \rangle| = |\langle \Lambda, D^\alpha \phi \rangle| \leqslant C \parallel \phi \parallel_{n+|\alpha|} \qquad (\phi \in \mathcal{D}(K))$$

   y, por tanto, $D^\alpha \Lambda$ es un funcional continuo sobre todo $\mathcal{D}(K)$ y como consecuencia sobre $\mathcal{D}(\Omega)$.

2. cambiar notación

$$\begin{aligned}(D^\alpha D^\beta \Lambda)(\phi) &= (-1)^{|\alpha|}(D^\beta \Lambda)(D^\alpha \phi) = (-1)^{|\alpha|+|\beta|}\Lambda(D^\beta D^\alpha \phi) \\ &= (-1)^{|\alpha|+|\beta|}\Lambda(D^{\alpha+\beta}\phi) = D^{\alpha+\beta}\Lambda(\phi)\end{aligned}$$

□

Teorema 3.2. *Sean $\Lambda \in \mathcal{D}'(\Omega)$ y K un compacto de Ω. Entonces, existen una función contínua f definida en Ω y un multi-índice α tales que*

$$\langle \Lambda, \phi \rangle = (-1)^{|\alpha|} \int_\Omega f(x)(D^\alpha \phi)(x) dx$$

para cada $\phi \in \mathcal{D}(K)$.

Definición 3.2.1. Dada una distribución $\Lambda \in \mathcal{D}'(\Omega)$ definimos su restricción a U como la distribución $\Lambda|_U \in \mathcal{D}'(U)$ definida por

$$\langle \Lambda|_U, \phi \rangle = \langle \Lambda, \phi \rangle \qquad (\phi \in \mathcal{D}(U)).$$

Diremos que Λ es cero en U si $\Lambda|_U = 0$.

Definición 3.2.2. Sea $\Lambda \in \mathcal{D}'(\Omega)$ y sea U el mayor subconjunto abierto de Ω donde Λ es cero, entonces llamaremos soporte de Λ y denotaremos por $\operatorname{supp}(\Lambda)$ al conjunto $\Omega \backslash U$.

Teorema 3.3. *Sea $\Lambda \in \mathcal{D}'(\Omega)$ de orden N y $p \in \Omega$ tal que* $\operatorname{supp}(\Lambda) = \{p\}$. *Entonces, existen constantes c_α tales que*

$$\Lambda = \sum_{|\alpha| \leqslant N} c_\alpha D^\alpha \delta_p.$$

3.2.3. Convolución

Definición 3.3.1. Sean $u : \mathbb{R}^d \to \mathbb{K}$, $x \in \mathbb{R}^d$. Se definen las funciones $\tau_x u, \check{u} : \mathbb{R}^d \to \mathbb{K}$ por

$$\begin{aligned} [\tau_x u] &= u(y-x) \\ \check{u}(y) &= u(-y) \end{aligned}$$

para cada $y \in \mathbb{R}^d$.

Definición 3.3.2. Sean $u, v, : \mathbb{R}^d \to \mathbb{K}$. Definimos su convolución como la función $u * v : \mathbb{R}^d \to \mathbb{K}$ dada por

$$\begin{aligned} (u * v)(x) &= \int_{\mathbb{R}^d} u(y) v(x-y) dy = \int_{\mathbb{R}^d} u(y) [\tau_x \check{v}](y) dy \\ &= \langle \Lambda_u, \tau_x \check{v} \rangle \end{aligned}$$

para todo $x \in \mathbb{R}^d$, suponiendo que dicha integral exista. Por tanto, para $\Lambda \in \mathcal{D}'(\mathbb{R}^d)$ y $\phi \in \mathcal{D}(\mathbb{R}^d)$ podemos definir la convolución de Λ y ϕ como la distribución dada por

$$[\Lambda * \phi](x) = \langle \Lambda, \tau_x \check{\phi} \rangle$$

para cada $x \in \mathbb{R}^d$.

3.3. Transformada de Fourier. Distribuciones Temperadas

3.3.1. Motivación

Definición 3.3.3. Dada $f \in L_1(\mathbb{R}^d)$, se define su transformada de Fourier como la función $\mathcal{F} : \mathbb{R}^d \to \mathbb{K}$ dada por

$$[\mathcal{F}(f)](\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-ix\xi} f(x) dx, \xi \in \mathbb{R}^d$$

donde $x\xi = x_1\xi_1 + \cdots + x_d\xi_d$.Con esta definición, también es natural ver la transofrmada de Fourier como un operador que lleva cada función $f \in L_1(\mathbb{R}^d)$ en la función $\mathcal{F}(f)$.

Sabemos, por el lema de Riemann-Lebesgue, que $\mathcal{F}(f)$ pertenece a $C_0(\mathbb{R}^d)$ y es, por tanto, localmente integrable. Así, $\mathcal{F}(f)$ se identifica con una distribución $\Lambda_{\mathcal{F}(f)}$ dada por :

$$\begin{aligned}\langle \mathcal{F}(f), \phi\rangle &= \int_{\mathbb{R}^d} [\mathcal{F}(f)](\xi)\phi(\xi)d\xi \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-ix\cdot\xi} f(x)\phi(\xi)dxd\xi \\ &= \int_{\mathbb{R}^d} [\mathcal{F}(\phi)](x)f(x)dx = \langle f, \mathcal{F}(\phi)\rangle\end{aligned}$$

para cada $\phi \in \mathcal{D}(\mathbb{R}^d)$.

Esto nos puede llevar a querer generalizar la Transformada de Fourier de una distribución Λ de la siguiente manera:

$$\langle \mathcal{F}(\Lambda), \phi\rangle = \langle \Lambda, \mathcal{F}(\phi)\rangle \qquad (\phi \in \mathcal{D}(\mathbb{R}^d)).$$

Sin embargo, hay dos cuestiones que impiden dicha generalización. Por un lado, no es posible asegurar que $\mathcal{F}(\phi) \in \mathcal{D}(\mathbb{R}^d)$ siempre que $\phi \in \mathcal{D}(\mathbb{R}^d)$. Esto es, $\mathcal{D}(\mathbb{R}^d)$ es demasiado pequeño como espacio de funciones test para extender la transformada de Fourier.

Por otro lado, $C^\infty(\mathbb{R}^d)$ no está conteniedo en $L_1(\mathbb{R}^d)$, por lo que la transformada de Fourier de una función de $C^\infty(\mathbb{R}^d)$ puede no existir. Esto es, $C^\infty(\mathbb{R}^d)$ es demasiado grande como espacio de funciones test.

3.3.2. Notación

Definición 3.3.4. Definimos la medida normalizada de Lebesgue en $\mathbb{R}^n$ como

$$dm_n(x) = (2\pi)^{-\frac{n}{2}}dx.$$

Definición 3.3.5. Para cada $t \in \mathbb{R}^n$, e_t denotará:

$$e_t(x) = e^{it\cdot x} \quad (x \in \mathbb{R}^n).$$

Es claro que para cada $t \in \mathbb{R}^n$, e_t satisface la propiedad $e_t(x+y) = e_t(x)e_t(y)$. Además, podemos reescribir la definición de transformada de Fourier en términos de e_t:

$$\mathcal{F}(f)(t) = (f * e_t)(0) = \int_{\mathbb{R}^n} fe_{-t}dm_n \qquad (t \in \mathbb{R}^n).$$

Definición 3.3.6. Sea α un multi-índice. De aquí en adelante vamos a llamar D_α al operador

$$D_\alpha = i^{-|\alpha|}D^\alpha = \left(\frac{1}{i}\frac{\partial}{\partial x_1}\right)^{\alpha_1}\cdots\left(\frac{1}{i}\frac{\partial}{\partial x_n}\right)^{\alpha_n}.$$

Sea P un polinomio de n variables y coeficientes complejos c_α. Definimos los operadores diferenciales

$$P(D) = \sum c_\alpha D_\alpha \quad \text{y} \quad P(-D) = \sum (-1)^{|\alpha|}c_\alpha D_\alpha.$$

3.3.3. Funciones de decrecimiento rápido.

Definición 3.3.7 (Funciones de decrecimiento rápido)**.** Sea $f \in C^{\infty}(\mathbb{R}^n)$ cumpliendo que

$$\sup_{|\alpha| \leqslant N} \sup_{x \in \mathbb{R}^n} \left\{ (1 + |x|^2)^N |(D^\alpha f)(x)| < \infty \right\} \qquad (N \in \mathbb{N}).$$

Esto es, $PD_\alpha f$ es una función acotada en $\mathbb{R}^n$, para todo polinomio P. Entonces, decimos que f es una función de decrecimiento rápido en infinito. Las funciones con decrecimiento rápido en infinito forman un espacio vectorial, denotado por $\mathcal{S}(\mathbb{R}^d)$, al que llamaremos la clase de Schwartz en $\mathbb{R}^d$. En $\mathcal{S}(\mathbb{R}^d)$ consideraremos la topología (localmente convexa y metrizable) asociada a la familia de seminormas

$$s_{k,N}(\phi) = \sup_{|\alpha| \leqslant N} \sup_{x \in \mathbb{R}^k} \left\{ (1 + |x|^2)^N |(D^\alpha \phi)(x)| < \infty \right\} \qquad (\phi \in \mathcal{S}(\mathbb{R}^d) \quad k, N \in \mathbb{N}).$$

Observación 3.3.1. $\mathcal{D}(\mathbb{R}^d) \subset \mathcal{S}(\mathbb{R}^d) \subset L_1(\mathbb{R}^d)$ y $\mathcal{S}(\mathbb{R}^d)$ es denso en $L_1(\mathbb{R}^d)$ (pues lo es $\mathcal{D}(\mathbb{R}^d)$).

Proposición 3.3.1. *$\mathcal{D}(\mathbb{R}^d)$ es denso en $\mathcal{S}(\mathbb{R}^d)$ y la inclusión $I : \mathcal{D}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ es continua.*

Teorema 3.4.

1. *$\mathcal{S}(\mathbb{R}^d)$ es un espacio de Fréchet.*
2. *Sean P un polinomio, $\phi, \psi \in \mathcal{S}(\mathbb{R}^d)$ y α un multi-índice. Entonces las aplicaciones*

   $$\phi \to P\phi, \quad \phi \to \psi\phi, \quad \phi \to D^\alpha \phi \qquad (\phi \in \mathcal{S}(\mathbb{R}^d))$$

   son aplicaciones continuas de $\mathcal{S}(\mathbb{R}^d)$ en $\mathcal{S}(\mathbb{R}^d)$.
3. *Sea $\phi \in \mathcal{S}(\mathbb{R}^d)$ y sea α un multi-índice. Entonces, $\mathcal{F}(D^\alpha \phi) = i^{|\alpha|}\xi^\alpha \mathcal{F}(\phi)$ y $\mathcal{F}(x^\alpha \phi) = i^{|\alpha|} D^\alpha \mathcal{F}(\phi)$.*
4. *Sean $f \in \mathcal{S}(\mathbb{R}^d)$ y P un polinomio. Entonces $\mathcal{F}(P(D)f) = P\mathcal{F}(f)$ y $\mathcal{F}(Pf) = P(-D)\mathcal{F}(f)$.*
5. *La transformada de Fourier es un operador lineal y continuo de $\mathcal{S}(\mathbb{R}^d)$ en sí mismo.*

Demostración.

1. Sea $\{f_i\}$ una sucesión de Cauchy en $\mathcal{S}(\mathbb{R}^d)$. Por ser las funciones f_i de decrecimiento rápido, para cada par de multi-índices α y β, $\{x^\beta D^\alpha f_i(x)\} \to g_{\alpha\beta}$, donde $g_{\alpha\beta}$ es una función acotada. Por tanto, $\{f_i\} \to g \in \mathcal{S}(\mathbb{R}^d)$, lo que demuestra que $\mathcal{S}(\mathbb{R}^d)$ es completo.
2. Si $\phi \in \mathcal{S}(\mathbb{R}^d)$ sabemos que $D^\alpha \phi \in \mathcal{S}(\mathbb{R}^d) \in$. En el caso de $\phi\psi$, sabemos que $D^\alpha(\phi\psi) = (D^\alpha \phi)\psi + \phi D^\alpha \psi \mathcal{S}(\mathbb{R}^d)$. Análogamente, $D^\alpha(Pf) = (D^\alpha P)f + PD^\alpha f$, siendo ambos sumandos de crecimiento lento por serlo ϕ. Hemos demostrado que las aplicaciones anteriores están bien definidas. Para demostrar su continuidad, basta con utilizar el teorema de la gráfica cerrada.
3. Fijado un multi-índice α, tenemos que

   $$[\mathcal{F}(D_\alpha \phi)](\xi) = \int_{\mathbb{R}^n}$$

4. Si $f \in \mathcal{S}(\mathbb{R}^d)$, por 2. sabemos que $P(D)f \in \mathcal{S}(\mathbb{R}^d)$ y, además, se cumple

$$(P(D)f) * e_t = \int_{\mathbb{R}^n} (P(D)f)(x-y)e_t(y)dm_n = \int_{\mathbb{R}^n} (\sum c_\alpha (i)^{-|\alpha|} D^\alpha f)(x-y)e^{it\cdot y}dm_n$$
$$= \int_{\mathbb{R}^n} f(x-y)(\sum c_\alpha (i)^{-|\alpha|} D^\alpha e_t)(y)dm_n = \int_{\mathbb{R}^n} f(x-y)P(t)e_t(y)dm_n = P(t)[f * e_t].$$

   Evaluando en el origen, concluimos que $\mathcal{F}(P(D)f) = P\mathcal{F}(f)$.

5. De nuevo demostrar que está bien definido y usar el teorema de la gráfica cerrada.

□

Teorema 3.5 (Teorema de inversión). *La transformada de Fourier es un isomorfismo de $\mathcal{S}(\mathbb{R}^d)$ en sí mismo. Si $\phi \in \mathcal{S}(\mathbb{R}^d)$, entonces*

$$\phi(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{ix\cdot\xi}[\mathcal{F}(\phi)](\xi)d\xi \qquad (x \in \mathbb{R}^d)$$

esto es, $\mathcal{F}^{-1} \circ \mathcal{F} = Id_{\mathcal{S}(\mathbb{R}^d)}$. Además, este isomorfismo cumple que $\mathcal{F}^4(\phi) = \phi$.

3.3.4. Distribuciones Temperadas

Definición 3.5.1. Una distribución temperada o función generalizada de crecimiento lento es un funcional lineal y continuo en $\mathcal{S}(\mathbb{R}^d)$. Se denota por $\mathcal{S}'(\mathbb{R}^d)$ al espacio de las distribuciones temperadas, esto es, al dual de $\mathcal{S}(\mathbb{R}^d)$. La topología que consideramos en $\mathcal{S}'(\mathbb{R}^d)$ es la débil-*, esto es, $\sigma(\mathcal{S}'(\mathbb{R}^d), \mathcal{S}(\mathbb{R}^d))$.

En el Teorema 3.4 veíamos cómo la transformada de Fourier está bien definida en el espacio $S(\mathbb{R}^n)$. Hemos encontrado, por tanto, el espacio de funciones test de tamaño adecuado que nos permite, siguiendo el espíritu natural de las distribuciones, definir la transformada de Fourier en $\mathcal{S}'(\mathbb{R}^d)$.

Definición 3.5.2. Para cada $\Lambda \in \mathcal{S}'(\mathbb{R}^d)$ definimos su transformada de Fourier:

$$\langle \mathcal{F}(\Lambda), \phi \rangle = \langle \Lambda, \mathcal{F}(\phi) \rangle.$$

Como la transformada de Fourier es un isomorfismo de $\mathcal{S}(\mathbb{R}^d)$ en sí mismo, $\mathcal{F}(\Lambda)$ está bien definida y es una distribución temperada.

Teorema 3.6. *Sean $\Lambda \in \mathcal{S}'(\mathbb{R}^n)$ y P un polinomio. Entonces $\mathcal{F}(P(D)\Lambda) = P\mathcal{F}(\Lambda)$ y $\mathcal{F}(P\Lambda) = P(-D)\mathcal{F}(\Lambda)$. Para los operadores $P(D)$ y $P(-D)$ definidos en términos de D_α.*

Demostración.

$$\langle \mathcal{F}(P(D)\Lambda), \phi \rangle = \langle P(D)\Lambda, \mathcal{F}(\phi) \rangle = \langle \Lambda, P(-D)\mathcal{F}(\phi) \rangle$$
$$= \langle \Lambda, \mathcal{F}(P\phi) \rangle = \langle \mathcal{F}(\Lambda), P\phi \rangle = \langle P\mathcal{F}(\Lambda), \phi \rangle.$$

□

Lema 3.6.1. *Si una distribución tiene como transformada de Fourier una combinación lineal de deltas de Dirac y sus derivadas, esa distribución vista como función es un polinomio.*

Demostración. Sea Λ una distribución tal que $\mathcal{F}(\Lambda) = \sum_{k=0}^{N} c_k D^k \delta$. Ya hemos visto que los polinomios son distribuciones temperadas. Podemos por tanto calcular sus transformadas de Fourier.

- Supongamos que $N = 0$. Entonces $\mathcal{F}(\Lambda) = c_0\delta$. Utilizando el Teorema de Inversión, deducimos que

$$\begin{aligned}\langle \mathcal{F}(\Lambda), \phi \rangle &= \int_{\mathbb{R}^n} c_0\delta(\phi) = c_0\phi(0) = \int_{\mathbb{R}^n} c_0(2\pi)^{-\frac{n}{2}} e^{i0\cdot\xi}[\mathcal{F}(\phi)](\xi)d\xi \\ &= \int_{\mathbb{R}^n} P_0(2\pi)^{-\frac{n}{2}} e^{i0\cdot\xi}[\mathcal{F}(\phi)](\xi)d\xi = \langle \Lambda_{P_0}, \mathcal{F}(\phi) \rangle\end{aligned}$$

para cualquier función test ϕ. $P_0(x) = c_0$ es, por tanto, el polinomio que buscamos. En particular, tenemos que $\mathcal{F}(\Lambda_1) = \delta$. Análogamente, es fácil comprobar que:

$$\langle \mathcal{F}(\delta), \phi \rangle = \langle \delta, \mathcal{F}(\phi) \rangle = \mathcal{F}(\phi)(0) = \int_{\mathbb{R}^n} \phi dm_n = \langle \Lambda_1, \phi \rangle.$$

- Supongamos ahora $N \geqslant 1$. Por Teorema 3.6 sabemos que

$$\mathcal{F}(P(D)\delta) = P\mathcal{F}(\delta) = P.$$

$\square$

4. Aprendizaje profundo. Fundamenos

4.1. Descripción de un modelo formal de Aprendizaje Automático

El Aprendizaje Automático comprende una serie de prácticas y algoritmos que comparten como objetivo común la tarea de aprender información a partir de unos datos de entrada. La descripción formal para un modelo de Aprendizaje Automático suele tener los siguientes elementos:

- Dominio del problema: se trata de un conjunto X que contendrá todos los objetos que tenemos a nuestra disposición para realizar la tarea de aprendizaje.
- Conjunto de entrenamiento: Un conjunto finito $S = \{(x_1, y_1), \ldots, (x_n, y_n\}\} \subset X \times Y$, donde X e Y son respectivamente el espacio del dominio del problema y su conjunto de etiquetas asociado.
- Una hipótesis de aprendizaje que el algoritmo A construye tras haber recibido el conjunto de entrenamiento S. La denotaremos por $A(S)$.
- Un predictor $h : X \to Y$ devuelto por el algoritmo tras la tarea de aprendizaje que describe la hipótesis $A(S)$.
- Una medida de bondad de la predicción realizada por el algoritmo. Esta medida tiene el nombre de función de coste y está asociada a la probabilidad de que, dado un elemento arbitrario $x \in X$, el algoritmo comenta un error en la predicción.

Este modelo tan general se ha concretado en varios modelos de aprendizaje. En este trabajo nos centraremos en describir las principales características del Aprendizaje Profundo, esto es, las Redes Neuronales.

4.2. Redes Neuronales

Una Red Neuronal Artificial (RNA) es un modelo de Aprendizaje Automático basado en la idea de que muchas unidades de cómputo pequeñas pueden unirse y comunicarse para construir modelos complejos. Para ello, imitan estructura y el comportamineto de las neuronas biológicas.

Podemos describirlo como un grafo dirigido acíclico $G = (V, E)$ organizado por capas ($V = \{V_0, V_1, \ldots, V_n\}$) cuyos vértices representan neuronas artificiales (unidades de cómputo simple) y cuyas aristas representan las conexiones entre ellas.

5. Aproximación de operadores no lineales mediante redes neuronales

5.1. Notación y definiciones

Comenzaremos aclarando algunos conceptos clave para seguir las demostraciones de esta sección:

Definición 5.0.1. Una función $\sigma : \mathbb{R} \to \mathbb{R}$ es sigmoidal si cumple

$$\sigma(x) = \begin{cases} \lim_{x\to-\infty} \sigma(x) = 0 \\ \lim_{x\to\infty} \sigma(x) = 1 \end{cases}$$

Observación 5.0.1. A diferencia de otras definiciones, cuando nos refiramos a una función sigmoidal no vamos a presuponer continuidad ni monotonía.

Definición 5.0.2. Una función $g : \mathbb{R} \to \mathbb{R}$ se llama de Tauber-Wiener (TW) si satisface que todas las combinaciones lineales $\sum_{i=1}^{N} c_i g(\lambda_i x + \theta_i), \lambda_i, \theta_i, c_i \in \mathbb{R}, i = 1, \cdots, N$ son densas en cualquier espacio $C[a, b]$.

Definición 5.0.3. Denotaremos por $C_p[-1, 1]^n$ al espacio de las funciones periódicas con respecto a cada variable $x_1, \cdots, x_n$ cuyo periodo es 2.

5.2. Características de las funciones de activación

En esa sección trabajaremos en debilitar las condiciones bajo las cuales una función puede ser considerada función de activación, lo que nos dará más libertad a la hora de trabajar con operadores no lineales en la próxima sección. El resultado más importante es el Teorema 5.2, que nos dice que pertenecer a la clase (TW) es condición suficiente para que una función sea de activación. Como las funciones (TW) toman valores en $\mathbb{R}$, vamos a poder trabajar con problemas simplificados. La convergencia uniforme que nos da este teorema va a ser muy importante para los resultados de la sección siguiente.

Teorema 5.1. *Supongamos que g es una función continua y $g \in S'(\mathbb{R}^1)$ vista como distribución. Entonces $g \in (TW)$ si g no es un polinomio.*

Demostración. Si las combinaciones $\sum_{i=1}^{N} c_i g(\lambda_i x + \theta_i)$, vistas como subespacio, no son densas en $C[a, b]$ (es decir, si su cierre forma un subespacio cerrado propio en $C[a, b]$) entonces por el Teorema de Extensión de Hahn-Banach existe un funcional lineal y continuo F cuyo núcleo se traga al subespacio. Por otra parte, el Teorema de Representación de Riesz asegura que hay una medida de Borel $d\mu$ con $\mathrm{supp}(d\mu) \subseteq [a, b]$ cumpliendo tal que $F(f) = \int_{\mathbb{R}} f d\mu$.

Como $g(\lambda x + \theta)$ está en dicho subespacio para todo $\lambda \neq 0$ y $\theta \in \mathbb{R}$, por los dos resultados anteriores, sabemos que:

$$F(g(\lambda x+\theta)) = \int_{\mathbb{R}} g(\lambda x+\theta)d\mu = 0$$

para todo $\lambda \neq 0$ y $\theta \in \mathbb{R}$. Tomando ahora cualquier $w \in S(\mathbb{R})$, tenemos que

$$\int_{\mathbb{R}} w(\theta)d\theta \int_{\mathbb{R}} g(\lambda x+\theta)d\mu = 0$$

para todo $\lambda \neq 0$ y $\theta \in \mathbb{R}$.

Fijados un $\lambda \neq 0$ y un $\theta \in \mathbb{R}$, definimos la medida $\hat{d\mu}(A) = d\mu(\frac{A-\theta}{\lambda})$ y efectuamos el cambio de variable $y = \lambda x + \theta$. Esto nos va a permitir ver a g como distribución:

$$\int_{\mathbb{R}} g(y) \int_{\mathbb{R}} w(\theta)d\mu(\frac{y-\theta}{\lambda}) = \langle \Lambda_g, w(\cdot) * d\mu(\lambda\cdot)\rangle = 0. \tag{5.1}$$

Por el Teorema de Inversión, tenemos:

$$\begin{aligned}\langle \Lambda_g, w(\cdot) * d\mu(\lambda\cdot)\rangle &= \langle \Lambda_g, \mathcal{F}(\mathcal{F}(w(\cdot) * d\mu(\lambda\cdot)))\rangle = \langle \mathcal{F}(\Lambda_g), \mathcal{F}(w(\cdot) * d\mu(\lambda\cdot))\rangle \\ &= \langle \mathcal{F}(\Lambda_g), \mathcal{F}(w(\cdot)) \cdot \mathcal{F}(d\mu(\lambda\cdot))\rangle = 0\end{aligned}$$

Para que la parte derecha de (5.1) tenga sentido, tenemos que probar que $[\mathcal{F}(w)](t) \cdot [\mathcal{F}(d\mu)](\lambda t) \in \mathcal{S}(\mathbb{R}^d)$. Como $\operatorname{supp}(d\mu) \subseteq [a,b]$, podemos escribir

$$\begin{aligned}\langle \mathcal{F}(d\mu), \phi\rangle = \langle d\mu, \mathcal{F}(\phi)\rangle &= \int_{\mathbb{R}} d\mu(t)\mathcal{F}(\phi)(t)dt = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \phi(x)e^{-ixt}dx\right] d\mu(t)dt \\ &= \int_{\mathbb{R}} \phi(x) \left[\int_{\mathbb{R}} e^{-ixt}d\mu(t)dt\right] dx = \int_{\mathbb{R}} \phi(x)h(x)dx = \langle \Lambda_h, \phi\rangle\end{aligned}$$

donde $h(x) = \int_{\mathbb{R}} e^{-ixt}d\mu(t)$.

Además, por el Teorema de derivación bajo el signo integral sabemos que $h \in C^\infty(\mathbb{R})$, lo cual nos dice que $\mathcal{F}(d\mu)$ es, vista como una función, de clase $C^\infty(\mathbb{R})$.
Veamos ahora que para cada $k = 1, 2, \ldots$, existe una constante c_k tal que:

$$\left|\frac{\partial^k}{\partial t^k}\mathcal{F}(d\mu)(t)\right| \leqslant c_k.$$

Efectivamente, tenemos que

$$|h(x)| = \left|\int_{\mathbb{R}} e^{-ixt}d\mu(t)\right| \leqslant \int_{\mathbb{R}} |e^{-ixt}||d\mu(t)| \leqslant \int_{\mathbb{R}} |d\mu(t)| \leqslant 1, \forall x \in \mathbb{R}.$$

De forma análoga, podemos ver que

$$\begin{aligned}\langle \frac{\partial^k}{\partial t^k}\mathcal{F}(d\mu),\phi\rangle &= \langle \mathcal{F}(t^k d\mu),\phi\rangle = \langle t^k d\mu, \mathcal{F}(\phi)\rangle = \int_{\mathbb{R}} \mathcal{F}(\phi)(t)t^k d\mu(t)dt \\ &= \int_{\mathbb{R}}\left[\int_{\mathbb{R}}\phi(x)e^{-ixt}t^k d\mu(t)dx\right]dt = \int_{\mathbb{R}}\phi(x)\left[\int_{\mathbb{R}} e^{-ixt}t^k d\mu(t)dt\right]dx \\ &= \langle \Lambda_{h_k},\phi\rangle\end{aligned}$$

Donde estamos considerando $\Lambda_{h_k} = \mathcal{F}(t^k d\mu)$, lo cual nos permite ver a h_k como función. Desde aquí, teniendo en cuenta que $\mathrm{supp}(d\mu) \subset [a,b]$, es fácil ver que

$$|h_k(x)| \leqslant \int_{\mathbb{R}} |t^k| d\mu(t) = c_k$$

para alguna constante $c_k \in \mathbb{R}$.

Como consecuencia, por ser $\mathcal{F}(w)$ de crecimiento lento y tener $\mathcal{F}(d\mu)$ todas sus derivadas acotadas, $\mathcal{F}(w)\cdot\mathcal{F}(d\mu) \in \mathcal{S}(\mathbb{R})$.
Por el Teorema de Representación de Riesz sabíamos que $d\mu \not\equiv 0$ y, además, hemos visto que $\mathcal{F}(d\mu) \in C^\infty(\mathbb{R})$. Por tanto, podemos afirmar que existe algún $t_0 \neq 0$ con un entorno $(t_0-\delta, t_0+\delta)$ tal que $\mathcal{F}(d\mu)(t) \neq 0$ para todo $t \in (t_0-\delta, t_0+\delta)$. Si tomamos $\lambda = \frac{t_0}{t_1}$, $t_1 \neq 0$, entonces $\mathcal{F}(d\mu)(\lambda t) \neq 0$, $\forall t \in (t_1 - \frac{\delta}{\lambda}, t_1 + \frac{\delta}{\lambda})$. Tomando un $\mathcal{F}(w) \in C_c^\infty(t_0 - \frac{\delta}{2\lambda}, t_0 + \frac{\delta}{2\lambda})$ cualquiera, entonces $\frac{\mathcal{F}(w)}{\mathcal{F}(d\mu)(\lambda t)} \in \mathcal{S}(\mathbb{R})$ y, por **(??)**:

$$\mathcal{F}(g)(\mathcal{F}(w)(\cdot)) = \mathcal{F}(g)\left(\frac{\mathcal{F}(w)(\cdot)}{\mathcal{F}(d\mu)(\lambda\cdot)}\mathcal{F}(d\mu)(\lambda\cdot)\right) = 0$$

Acabamos de ver que para cualquier punto $t^* \in \mathbb{R}$ fijo, existe un entorno $[t^*-\eta, t^*+\eta]$ tal que $\mathcal{F}(g)(\mathcal{F}(w)(\cdot)) = 0$ se cumple para todo $\mathcal{F}(w)$ con soporte compacto en $[t^*-\eta, t^*+\eta]$, esto es, $\mathrm{supp}(\mathcal{F}(g)) \subseteq 0$. Por el Teorema 3.3, $\mathcal{F}(g)$ es una combinación de deltas de Dirac y sus derivadas. Lo que equivale, como vimos en el Lema 3.6.1 a que g sea un polinomio. Con lo que el Teorema 1 queda probado. □

Teorema 5.2. *Supongamos que K es un conjunto compacto en $\mathbb{R}^n$, U es un conjunto compacto en $C(K)$ y $g \in (TW)$. Entonces, para cada $\varepsilon > 0$, existe un entero positivo N, números reales θ_i, vectores $w_i \in \mathbb{R}^n$, $i = 1, ..., N$, con independencia de $f \in U$ y constantes $c_i(f), i = 1, ..., N$ dependientes de f tal que*

$$|f(x) - \sum_{i=1}^{N} c_i(f)g(w_i x + \theta_i)| < \varepsilon$$

se cumple para todo $x \in K$ y para toda $f \in U$.

Antes de demostrar el Teorema 5.2, veamos un resultado previo que nos va a ser necesario:

Lema 5.2.1. *Supongamos que K es un conjunto compacto en $I^n = [0,1]^n$ y V es un conjunto compacto en $C(K)$. Entonces V puede extenderse a un conjunto compacto en $C_p[-1,1]^n$.*

Demostración. Para toda $f \in V$ sabemos, por el Lema 1.6.1, que existe una extensión continua $E(f) \in C([0,1]^n)$ tal que $sup_{x\in[0,1]^n}\{|E(f)(x)|\} \leqslant sup_{x\in K}\{|f(x)|\}$. Por el Elemento 5.3, sabemos también que existe una constante M tal que $\|f(x)\|_{C(K)} \leqslant M$, por lo que $sup_{x\in[0,1]^n}\{|E(f)(x)|\} \leqslant sup_{x\in K}\{|f(x)|\} < M$.

Llamemos V_1 al conjunto de las extensiones $E(f)$ tales que $f \in V$. Concluimos que V_1 es compacto en $C([0,1])$, pues $V_1 = E(V)$ y E es un operador continuo.

Para cada $f \in V_1$, podemos construir una extentión par de f:

$$f^*(x_1, ..., x_k, ..., x_n) = \begin{cases} f(x_1, ..., -x_k, ..., x_n) & \text{si } x_k \in [-1, 0[, \\ f(x_1, ..., x_k, ..., x_n) & \text{si } x \in [0,1]^n. \end{cases}$$

Entonces para cada $f \in V_1$, $f^* \in C_p([-1,1]^n)$ y por tanto $U = \{f^* : f \in V_1\}$ es el compacto en $C_p[-1,1]^n$ que buscamos. □

Lema 5.2.2. *Supongamos que U es un compacto en $C_p([-1,1])$,*

$$B_R(f;x) = \sum_{\|m\| \leqslant R} (1 - \frac{\|m\|^2}{|R|^2})^\alpha c_m(f) e^{i\pi m x}$$

es la medida de Bochner-Riesz de la serie de Fourier de f, donde $m = (m_1, ..., m_n)$, $\|m\|^2 = \sum_{i=1}^n |m_i|^2$ y $c_m(f)$ son los coeficientes de Fourier de f. Entonces, para todo $\varepsilon > 0$ existe un $R > 0$ tal que

$$|B_R(f;x) - f(x)| < \varepsilon$$

para toda $f \in U$ y $x \in [-1,1]^n$, dado un $\alpha > \frac{n-1}{2}$.

Demostración del Teorema 5.2. Sin perder generalidad, podemos asumir que $K \subseteq [0,1]^n$. Por el Lema 5.2.1, sabemos que podemos extenter U a un $U' \subseteq C_p[-1,1]^n$. Además, por el Lema 5.2.2, sabemos que para todo $\varepsilon > 0$ existe un $R > 0$ tal que se cumple

$$|B_R(f^*;x) - f^*(x)| = |\sum_{i=1}^n (1 - \frac{\|m\|^2}{|R|^2}) c_m(f^*) e^{i\pi(m \cdot x)} - f^*(x)| < \frac{\varepsilon}{2} \tag{5.2}$$

para $x \in [-1,1]^n$ y $f \in U$, donde f^* sigue la misma construcción expuesta en el Lema 5.2.1.

Por la paridad de f^*, simplificando los coeficientes de Fourier podemos reescribir (5.2)

$$|\sum_{\|m\| \leqslant R} d_m cos(\pi(m \cdot x)) - f^*(x)| < \frac{\varepsilon}{2} \tag{5.3}$$

donde $d_m = (1 - \frac{\|m\|^2}{|R|^2}) c_m(f^*)$.

Podemos ahora realizar el cambio de variable $u = \pi m \cdot x$ para $x \in [-1,1]^n$, por lo que $u \in [-\sqrt{n}\pi R, \sqrt{n}\pi R]$, ya que $\|x\| \leqslant \sqrt{n}$, $\|m\| \leqslant R$. Como $cos(u)$ es una función continua en $[-\sqrt{n}\pi R, \sqrt{n}\pi R]$ y $g \in (TW)$, existen un entero M y números reales s_j, η_j, ξ_j con $j = 1, \ldots, M$ tales que

$$|\sum_{j=1}^M s_j g(u\xi_j + \eta_j) - cos(u)| < \frac{\varepsilon}{2L}$$

se cumple para todo $u \in [-\sqrt{n}\pi R, \sqrt{n}\pi R]$, donde $L = \sum_{\|m\| \leqslant R} |d_m|$. Esto equivale a que

$$|\sum_{j=1}^{M} s_j g(\xi_j \pi(m \cdot x) + \eta_j) - cos(\pi m \cdot x)| < \frac{\varepsilon}{2L} \tag{5.4}$$

se cumpla para todo $x \in [-1,1]^n$.

Sean $A = \{m \in \mathbb{N}^n : \|m\| \leqslant R\}$, $A = \{m^1, \dots, m^N\}$ una indexación de A y sea $N = |A|M$, donde $|A|$ es el cardinal de A. Considerando ahora el índice $i = 1, \dots, N$, que despliega $|A|$ M veces, denotamos $i \mod |A|$ al resto de dividir i por $|A|$ ($i \leqslant N$). Así, conforme vayamos avanzando en i, $m^{i \mod |A|}$ denotará al elemento correspondiente de A dentro de uno de los M despliegues sobre el conjunto y podemos definir:

- $c_i(f) = z_i d_{m^{i \mod |A|}} = z_i(1 - \frac{\|m^{i \mod |A|}\|^2}{|R|^2}) c_{m^{i \mod |A|}}(f^*)$, donde:

$$z_i = \begin{cases} s_1 & \text{si } 1 \leqslant i \leqslant |A| \\ s_2 & \text{si } |A| < i \leqslant 2|A| \\ & \dots \\ s_M & \text{si } (M-1)|A| < i \leqslant M|A| \end{cases}$$

- $\omega_i = \begin{cases} \xi_1 \pi m^{i \mod |A|} & \text{si } 1 \leqslant i \leqslant |A| \\ \xi_2 \pi m^{i \mod |A|} & \text{si } |A| < i \leqslant 2|A| \\ & \dots \\ \xi_M \pi m^{i \mod |A|} & \text{si } (M-1)|A| < i \leqslant M|A| \end{cases}$

- $\theta_i = \begin{cases} \eta_1 & \text{si } 1 \leqslant i \leqslant |A| \\ \eta_2 & \text{si } |A| < i \leqslant 2|A| \\ & \dots \\ \eta_M & \text{si } (M-1)|A| < i \leqslant M|A|. \end{cases}$

Usando (5.3) y (5.4), podemos concluir que los elementos $N \in \mathbb{N}, \theta_i \in \mathbb{R}, w_i \in \mathbb{R}^n, i = 1, \dots, N$ definidos cumplen:

$$\begin{aligned} |f^*(x) - \sum_{i=1}^{N} c_i(f^*) g(w_i x + \theta_i)| &= |f^*(x) - \sum_{\|m\| \leqslant R} d_m cos(\pi(m \cdot x)) \\ &+ \sum_{\|m\| \leqslant R} d_m cos(\pi(m \cdot x)) - \sum_{i=1}^{N} c_i(f^*) g(w_i x + \theta_i)| \\ \leqslant |f^*(x) - \sum_{\|m\| \leqslant R} d_m cos(\pi(m \cdot x))| &+ |\sum_{\|m\| \leqslant R} d_m cos(\pi(m \cdot x)) - \sum_{i=1}^{N} c_i(f^*) g(w_i x + \theta_i)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

para todo $x \in [-1, 1]^n$. Restringiéndonos ahora a $[0, 1]^n$ tenemos que

$$|f(x) - \sum_{i=1}^{N} c_i(f) g(w_i x + \theta_i)| < \varepsilon$$

se cumple para todo $x \in [0, 1]^n$ y $f \in U$.

Con respecto al paso de $c_i(f^*)$ a $c_i(f)$, para cada $m = (m_1, \ldots, m_n)$, el coeficiente de Fourier

$$c_m(f^*) = \int_{-1}^{1} \cdots \int_{-1}^{1} e^{i\pi(m \cdot x)} f^*(x) dx$$

es un funcional continuo definido en U'. Por la definición de f^* y cambiando los límites de integración, $c_i(f)$ está bien definido en U. Así, el Teorema 5.2 queda demostrado. □

Observación 5.2.1. Por ser los $c_i(f)$ múltiplos de los coeficientes de Fourier de f^* con la restricción a U mencionada, cada $c_i(f)$, $i = 1, \ldots N$ actúa de forma continua sobre U

Teniendo ya en mente, por el teorema anterior, que basta que una función sea de clase (TW) para considerarla función de activación, el siguiente resultado nos permite relajar las condiciones clásicas de continuidad o monotonía exigidas a una función de activación sigmoidal:

Teorema 5.3. *Si σ es una función sigmoidal acotada, entonces $\sigma \in (TW)$.*

Demostración. El enunciado de este teorema es equivalente a afirmar que si $\sigma(x)$ es una función sigmoidal acotada y tomamos $f(x) \in C(\mathbb{R})$ tal que $\lim_{x \to -\infty} f(x) = A$ y $\lim_{x \to \infty} f(x) = B$, con $A, B \in \mathbb{R}$, entonces para cada $\varepsilon > 0$ existirán $N \in \mathbb{N}$, $c_i, y_i, \theta_i \in \mathbb{R}$ tales que

$$|f(x) - \sum_{i=1}^{N} c_i \sigma(y_i x + \theta_i)| < \varepsilon$$

se cumple para todo $x \in (-\infty, \infty)$.

Partiendo de la función f que hemos fijado, para cada $\varepsilon > 0$ podemos encontrar una constante real $M > 0$ tal que

1. $|f(x) - A| < \frac{\varepsilon}{4}$ si $x < -M$.
2. $|f(x) - B| < \frac{\varepsilon}{4}$ si $x > M$.
3. $|f(x') - f(x'')| < \frac{\varepsilon}{4}$ si $|x'| \leqslant M$, $|x''| \leqslant M$ y $|x' - x''| \leqslant \frac{1}{M}$.

Tomemos una partición de $[-M, M]$ en $2M^2$ segmentos iguales, cada uno de longitud $\frac{1}{M}$.

$$-M = x_0 < x1 < \cdots < x_M = 0 < x_{M+1} < \cdots < x_{2M} = M.$$

Por ser σ una función sigmoidal, existe un $W > 0$ tal que

1. $|\sigma(u) - 1| < \frac{1}{M^2}$ si $u > W$.
2. $|\sigma(u)| < \frac{1}{M^2}$ si $u < -W$.

Tomando una constante $K > 0$ que cumpla $K\frac{1}{2M} > W$ podemos entonces construir una función

$$g(x) = f(-M) + \sum_{i=1}^{N} [f(x_i) - f(x_{i-1})]\sigma(K(x - t_{i-1})) \tag{5.5}$$

siendo $t_i = \frac{1}{2}(x_i + x_{i+1})$, $|x - t_i| \leqslant \frac{1}{2M}$, $\forall x \in (x_i, x_{i+1})$ y $N = 2M^2$. Veamos que se cumple que $|f(x) - g(x)| < \varepsilon$ para cada $x \in \mathbb{R}$.

- Si $x < -M$, entonces

  $$|f(x) - f(-M)| = |f(x) - A + A - f(-M)| < |f(x) - A| + |A - f(-M)| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}.$$

  Además, $(x - t_{i-1}) < (-M - (-M + \frac{1}{2M})) = -\frac{1}{2M}$, por lo que usando (5.5) obtenemos:

  $$|g(x) - f(-M)| \leqslant \sum_{i=1}^{N} |f(x_i) - f(x_{i-1})||\sigma(K(x - t_{i-1}))| < \sum_{i=1}^{N} \frac{\varepsilon}{4}\frac{1}{M^2} = \frac{\varepsilon}{2}$$

  por ser $K(x - t_{i-1}) < -\frac{K}{2M} < -W$.

  Por tanto,

  $$|f(x) - g(x)| \leqslant |f(x) - f(-M)| + |f(-M) - g(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

- La demostración es análoga para $x > M$.

- Consideremos ahora el caso en el que $x \in [-M, M]$. Entonces, para algún $k = 0, \ldots, 2M$, tendremos que $x \in [x_{k-1}, x_k]$. Por cómo hemos definido la partición sabemos que $|x - t_{i-1}| \leqslant \frac{1}{2M}$ si $i = k$ y $|x - t_{i-1}| > \frac{1}{2M}$ si $i \neq k$. Además, si $i < k$, $K(x - t_{i-1}) > W$ y por tanto $|\sigma(K(x - t_{i-1})) - 1| < \frac{1}{M^2}$. Análogamente, si $i > k$ entonces $K(x - t_{i-1}) < -W$ y por tanto $|\sigma(K(x - t_{i-1}))| < \frac{1}{M^2}$. En este caso, (5.5) nos permite escribir

  $$|g(x) - f(-M) - [f(x_k) - f(x_{k-1})]\sigma(K(x - t_{k-1})) - \sum_{i=1}^{k-1} [f(x_i) - f(x_{i-1})]|$$

  $$\leqslant \sum_{i=1}^{k-1} |f(x_i) - f(x_{i-1})||\sigma(K(x - t_{k-1})) - 1| + \sum_{i=k+1}^{2M^2} |f(x_i) - f(x_{i-1})||\sigma(K(x - t_{i-1}))|$$

  $$\leqslant \sum_{i=1}^{k-1} \frac{\varepsilon}{4}\frac{1}{M^2} + \sum_{i=k+1}^{2M^2} \frac{\varepsilon}{4}\frac{1}{M^2} \leqslant \frac{\varepsilon}{2}.$$

Por otro lado,

$$
\begin{aligned}
f(-M) + [f(x_k) - f(x_{k-1})]\sigma(K(x - t_{k-1})) - \sum_{i=1}^{k-1}[f(x_i) - f(x_{i-1})] \\
= f(x_{k-1}) + [f(x_k) - f(x_{k-1})]\sigma(K(x - t_{k-1}).
\end{aligned}
$$

De las desigualdades anteriores, deducimos que

$$
|g(x) - f(x)| \leqslant \frac{\varepsilon}{2} + |f(x) - f(x_{k-1})| + |f(x_k) - f(x_{k-1})||\sigma(K(x - t_{k-1}))| < \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon
$$

lo que concluye la prueba.

□

5.3. Aproximación de funcionales y operadores continuos no lineales

Teorema 5.4. *Supongamos que $g \in (TW)$, X es un espacio de Banach, $K \subseteq X$ es compacto, V es un conjunto compacto de $C(K)$ y f es un funcional definido en V. Entonces, para cada $\varepsilon > 0$ existen un entero positivo N, m puntos $x_1, ..., x_m \in K$ y constantes $c_i, \xi_{ij}, \theta_i \in \mathbb{R}$ tales que*

$$
|f(u) - \sum_{i=1}^{N} c_i g(\sum_{j=1}^{m} \xi_{ij} u(x_j) + \theta_i)| < \varepsilon \qquad (\forall u \in V).
$$

Lema 5.4.1. *Sea X un espacio de Banach y $K \subseteq X$, entonces K es compacto si y solo si se cumplen simultáneamente:*

1. *K es cerrado en X.*
2. *Para todo $\delta > 0$, existe una red $N(\delta) = \{x_1, ..., x_{n(\delta)}\}$. Esto es, para cada $x \in K$, existe $x_k \in N(\delta)$ tal que $\| x - x_k \|_X < \delta$.*

Observación 5.4.1. Para cada δ, el número $n(\delta)$ de elementos de $N(\delta)$ es finito, lo que equivale a decir que K está totalmente acotado.

Lema 5.4.2 (Otra lectura del Teorema de Ascoli-Azela)**.** *Si $V \subseteq C(K)$ es compacto en $C(K)$, entonces V es uniformemente acotado y equicontinuo. Esto es:*

1. *Existe $A > 0$ tal que $\| u(x) \|_{C(K)} \leqslant A, \forall u \in V$.*
2. *$\forall \varepsilon > 0, \exists \delta > 0$ tal que $|u(x') - u(x'')| < \varepsilon, \forall u \in V$, siempre que $\| x' - x'' \|_X < \delta$.*

Definición 5.4.1. Construiremos ahora varios elementos que vamos a utilizar en los resultados siguientes.

Sea f un funcional continuo definido en un compacto $V \subseteq C(K)$. Entonces, escogiendo una sucesión $\varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_n \to 0$, podemos encontrar otra sucesión $\delta_1 > \delta_2 > \cdots > \delta_n \to 0$ tal que $|f(u) - f(v)| < \varepsilon_k$ siempre que los $u, v \in V$ escogidos cumplan $\| u - v \|_{C(k)} < 2\delta_k$, $\forall f \in V$. Por el lema Lema 5.4.2 V es acotado y equicontinuo, por lo que además podemos encontrar una sucesión $\eta_1 > \eta 2 > \cdots > \eta_n \to 0$ tales que $|u(x') - u(x'')| < \delta_k$ para todo $u \in V$

y $\| x' - x'' \|_K < \eta_k$.

Usando el Lema 5.4.1, podemos reordenar los elementos de K para conseguir una sucesión $\{x_i\}_{i=1}^{\infty}$, con $x_i \in K$, y una sucesión de enteros positivos $n(\eta_1) < n(\eta_2) < \cdots < n(\eta_k) \to \infty$ tal que los $n(\eta_k)$ primeros elementos $N(\eta_k) = \{x_1, \dots, x_{n(\eta_k)}\}$ formen una η_k-red en K.

Para cada η_k-red, definimos

$$T^*_{\eta_k,j}(x) = \begin{cases} 1 - \frac{\|x-x_j\|_X}{\eta_k} & \text{si } \| x - x_j \|_X \leqslant \eta_k \\ 0 & \text{en otro caso} \end{cases} \quad \text{y} \quad T_{n(\eta_k),j} = \frac{T^*_{\eta_k,j}(x)}{\sum_{j=1}^{n(\eta_k)} T^*_{\eta_k,j}(x)}$$

para $j = 1, \dots, n(\eta_k)$.

Con esta definición, se puede ver que $\{T_{n(\eta_k),j}\}$ es una partición de la unidad, ya que

$$0 \leqslant T_{n(\eta_k),j}(x) \leqslant 1, \quad \sum_{j=1}^{n(\eta_k)} T_{n(\eta_k),j}(x) = 1 \quad \text{y} \quad T_{n(\eta_k),j}(x) = 0 \quad \text{si } \| x - x_j \|_X > \eta_k.$$

Para cada $u \in V$, podemos definir ahora funciones

$$u_{\eta_k}(x) = \sum_{j=1}^{n(\eta_k)} u(x_j) T_{n(\eta_k),j}(x)$$

y conjuntos $V_{\eta_k} = \{u_{\eta_k} : u \in V\}$, $V^* = V \bigcup (\bigcup_{k=1}^{\infty} V_{\eta_k})$.

Lema 5.4.3.

1. *Para cada k fijo, V_{η_k} es compacto en un subespacio de dimensión $n(\eta_k)$ en $C(K)$.*
2. *Para cada $u \in V$, se cumple*
$$\| u - u_{\eta_k} \|_{C(K)} < \delta_k.$$
3. *V^* es un conjunto compacto en $C(K)$.*

Demostración del Teorema 5.4. Por el Teorema 5.1, podemos definir un funcional contínuo en V^* tal que
$$f^*(x) = f(x) \qquad \forall x \in V.$$

Como f^* es un funcional continuo definido en el compacto V^*, por el sabemos que f^* es equicontinua; esto es, para cada $\varepsilon > 0$ podremos encontrar un $\delta > 0$ tal que $|f^*(u) - f^*(v)| < \frac{\varepsilon}{2}$, dados cualesquiera $u, v \in V^*$ tales que $\| u - v \|_{C(K)} < \delta$. Fijando k tal que $\delta_k < \delta$, entonces para cada $u \in V$, tenemos que
$$\| u - u_{\eta_k} \|_X < \delta_k.$$
Lo cual implica que
$$|f^*(u) - f^*(u_{\eta_k})| < \frac{\varepsilon}{2} \qquad \forall u \in V. \tag{5.6}$$

Por el Lema 5.4.3 V_{η_k} es compacto en un subespacio de dimensión finita lo cual equivale, por el Teorema de Hausdorff, a ser isométricamente isomorfo a un compacto del espacio $\mathbb{R}^{n(\eta_k)}$, al que vamos a denotar como $U_{\eta_k} \subset \mathbb{R}^{n(\eta_k)}$. Puesto que f^* es continua en V^* lo será, en concreto, en $V_{\eta_k} \subset V^*$ y, por el isomorfismo indicado, también podremos verla como una función contínua $f^* : U_{\eta_k} \to \mathbb{R}$. Por el Teorema 5.2, tomando $K = U_{\eta_k}$ y teniendo en mente para f^* la preservación de la compacidad por funciónes contínuas, podemos encontrar $N, c_i, \xi_{ij}, \theta_i \in \mathbb{R}$, $i = 1, ..., N$, $j = 1, ..., n(\eta_k)$ tales que

$$|f^*(u_{\eta_k}) - \sum_{i=1}^{N} c_i g(\sum_{j=1}^{n(\eta_k)} \xi_{ij} u(x_j) + \theta_i| < \varepsilon/2$$

Combinando esto con (5.6), concluimos que

$$\begin{aligned} |f(u) - \sum_{i=1}^{N} c_i(\sum_{j=1}^{n(\eta_k)} \xi_{ij} u(x_j) + \theta_i| &= |f^*(u) - \sum_{i=1}^{N} c_i(\sum_{j=1}^{n(\eta_k)} \xi_{ij} u(x_j) + \theta_i| \\ &= |f^*(u) - f^*(u_{\eta_k}) + f^*(u_{\eta_k}) - \sum_{i=1}^{N} c_i(\sum_{j=1}^{n(\eta_k)} \xi_{ij} u(x_j) + \theta_i| \\ &\leqslant |f^*(u) - f^*(u_{\eta_k})| + |f^*(u_{\eta_k}) - \sum_{i=1}^{N} c_i(\sum_{j=1}^{n(\eta_k)} \xi_{ij} u(x_j) + \theta_i| < \varepsilon \end{aligned}$$

Llamando m a $n(\eta_k)$ queda demostrado el Teorema 4. □

Teorema 5.5. *Sean $g \in (TW)$, X un espacio de Banach, $K_1 \subseteq X$, $K_2 \subseteq \mathbb{R}^n$ compactos en X y $\mathbb{R}^n$ respectivamente. Sea G un operador continuo no lienal que lleva V en $C(K_2)$. Entonces, para cada $\varepsilon > 0$ existen enteros $M, N, m > 0$, constantes $c_i^k, \zeta_k, \xi_{ij}^k \in \mathbb{R}$ y puntos $\Omega_k \in \mathbb{R}^n, x_j \in K_1$, con $i = 1, ..., M$, $k = 1, ..., , N$, $j = 1, ..., m$ tales que*

$$|G(u)(y) - \sum_{k=1}^{N} \sum_{i=1}^{M} c_i^k g(\sum_{j=1}^{m} \xi_{ij} u(x_j) + \theta_i^k) g(w_k \cdot y + \zeta_k)| < \varepsilon, \quad \forall u \in V, \forall y \in K_2$$

Demostración. De la hipótesis de que G es un operador continuo que lleva a $V \subset C(K_1)$ en $C(K_2)$, se comprueba directamente que la imagen $G(V) = \{G(u) : u \in V\}$ también es un conjunto compacto en $C(K_2)$. Por el Teorema 5.2, para cada $\varepsilon > 0$ existen $N \in \mathbb{N}$, $c_k(G(u)), \zeta_k \in \mathbb{R}$ y vectores $w_k \in \mathbb{R}^n$, $k = 1, ..., N$ tales que

$$|G(u)(y) - \sum_{k}^{N} c_k(G(u)) g(w_k y + \zeta_k)| < \frac{\varepsilon}{2} \tag{5.7}$$

se cumple para cualesquiera $y \in K_2$ y $u \in V$.

El teorema también nos dice que para cada $k = 1, .., N$, $c_k(G(u))$ es un funcional continuo definido en V. Aplicando repetidamente el Teorema 5.4, para cada $k = 1, ..., N$, podemos encontrar N_k, $m_k \in \mathbb{N}$, constantes $c_i^k, \xi_{ij}^k, \theta_i^k \in \mathbb{R}$ y $x_j \in K_1$, $i = 1, ..., N_k$, $j = 1, ..., m_k$ tales que

$$|c_k(G(u)) - \sum_{i=1}^{N_k} c_i^k g(\sum_{j=1}^{m_k} \xi_{ij}^k u(x_j) + \theta_i^k)| < \frac{\varepsilon}{2L} \tag{5.8}$$

se cumple para cualquier $k = 1, ..., N$ y $u \in V$, donde vamos a tomar

$$L = \sum_{k=1}^{N} \sup |g(w_k y + \zeta_k)|.$$

Usando (5.8) y (5.7), obtenemos que

$$|G(u)(y) - \sum_{k=1}^{N} \sum_{i=1}^{N_k} c_i^k g(\sum_{j=1}^{m_k} \xi_{ij}^k u(x_j) + \theta_i^k) g(w_k y + \zeta_k)| = \tag{5.9}$$

$$|G(u)(y) - \sum_{k=1}^{N} c_k(G(u)) g(w_k y + \zeta_k) + \sum_{k=1}^{N} c_k(G(u)) g(w_k y + \zeta_k)$$
$$- \sum_{k=1}^{N} \sum_{i=1}^{N_k} c_i^k g(\sum_{j=1}^{m_k} \xi_{ij}^k u(x_j) + \theta_i^k) g(w_k y + \zeta_k)| \leqslant$$
$$|G(u)(y) - \sum_{k=1}^{N} c_k(G(u)) g(w_k y + \zeta_k)| + |\sum_{k=1}^{N} c_k(G(u)) g(w_k y + \zeta_k)$$
$$- \sum_{k=1}^{N} \sum_{i=1}^{N_k} c_i^k g(\sum_{j=1}^{m_k} \xi_{ij}^k u(x_j) + \theta_i^k) g(w_k y + \zeta_k)| < \varepsilon$$

se cumple para cualesquiera $u \in V$ e $y \in K_2$.
Conveniendo ahora que $M = \max_k\{N_k\}$, $m = \max_k\{m_k\}$, que para todo $N_k < i \leqslant M$ se tenga $c_i^k = 0$ y que para todo $m_k < j \leqslant m$, $\xi_{ij}^k = 0$, (5.9) se puede reescribir como

$$|G(u)(y) - \sum_{k=1}^{N} \sum_{i=1}^{M} c_i^k g(\sum_{j=1}^{m} \xi_{ij}^k u(x_j) + \theta_i^k) g(w_k y + \zeta_k)| < \varepsilon$$

$\forall u \in V$, $\forall y \in K_2$. Esto completa la prueba del Teorema 5.

□

Parte II.

Aplicación a Sistemas Dinámicos No Lineales

6. Marco Teórico

6.1. Conclusiones sobre los resultados teóricos presentados

Como aplicación directa del Teorema 5.5, podemos utilizar redes neuronales para aproximar la salida de un sistema dinámico no lineal.

La importancia de los resultados anteriores recae en que podemos utilizar redes neuronales para identificar cualquier sistema de este tipo, sea linear. El procedimiento se muestra a continuación:

Sea un sistema $V = KU$, donde U es el input, V es el output y K es sistema que queremos identificar.

Supongamos que conocemos varias relaciones input-output $V_1 = KU_1, ..., V_n = KU_n$. Generalmente, podrán ser expresadas por conjuntos discretos $\{u_s(x_j),\ s = 1, ..., n, j = 1, ..., m\}$ $\{v_s(y_l), s = 1, ..., n, l = 1, ..., L\}$, con K de dimensiones $l \times m$. Usando estos puntos, por elTeorema 5.5 sabemos que podemos construir un funcional

$$E = \sum_{l=1}^{L} \sum_{s=1}^{n} |V_s(y_l) - \sum_{k=1}^{N} \sum_{i=1}^{M} C_i^k g(\sum_{j=1}^{m} \xi_{i,j}^k u_s(x_j) + \theta_i^k) g(w_k y_l + \zeta_k)|^2$$

Donde los parámetros $C_i^k, \xi_{i,j}^k, \theta_i^k, w_k, \zeta$ pueden elegirse de modo que minimicen E. Entonces, la ecuación

$$v(y) = \sum_{k=1}^{N} \sum_{i=1}^{M} C_i^k g(\sum_{j=1}^{m} \xi_{i,j}^k u(x_j) + \theta_i^k) g(w_k) y + \zeta_k$$

se puede ver como una aproximación de $V(y) = (KU)(y)$ y, por tanto, identifica el sistema K.

Si el sistema es lineal, entonces E y V(y) se pueden simplificar como:

$$E = \sum_{l=1}^{L} \sum_{s=1}^{n} |V_s(y_l) - \sum_{k=1}^{N} \sum_{i=1}^{M} \sum_{j=1}^{m} \xi_{i,j}^k u_s(x_j) g(w_k y_l + \zeta_k)|2$$

$$v(y) = \sum_{k=1}^{N} \sum_{i=1}^{M} \sum_{j=1}^{m} \xi_{i,j}^k u(x_j) g(w_k) y + \zeta_k$$

Cuanto mayores sean los valores de n, L y m, más precisión tendrá esta aproximación.

Por tanto, tenemos una forma de construir modelos de redes neuronales que se identifiquen con sistemas dinámicos.