

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



UNA APLICACIÓN DE ONDÍCULAS AL ESTUDIO DE ESTILOS DE ARTISTAS PLÁSTICOS MEXICANOS

TESIS

QUE PARA OBTENER EL TÍTULO DE
LICENCIADO EN MATEMÁTICAS APLICADAS

PRESENTA

MARIANA GRACIELA MARTÍNEZ AGUILAR

ASESOR: CARLOS BOSCH GIRAL

“Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada “**Una aplicación de ondículas al estudio de estilos de artistas plásticos mexicanos**”, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación”.

MARIANA GRACIELA MARTÍNEZ AGUILAR

FECHA

FIRMA

*A mis profesores y profesoras cuyas enseñanzas fueron más allá del
salón de clases.*

Agradecimientos

Estoy segura que el cómo soy y lo que he logrado se lo debo a las personas con las que me he rodeado. En primera instancia, quiero agradecer a aquellos profesores y profesoras con los que este camino empezó: Aurelio y Mariane. También a mi querido asesor, Carlos Bosch, por motivarme a hacer algo diferente y a mis sinodales Claudia, Jorge y Paco. Agradezco haber podido cruzar caminos con Ernesto Barrios, cuyos consejos no sólo me han servido en estadística y probabilidad. Me siento orgullosa de haber podido conocer y tomar cursos con personas tan interesantes, exigentes e inteligentes como Edith Vargas, Guillermo Grabinsky, Pablo Castañeda, Ramón Espinosa y Zeferino Parada.

A mis amigos, con los que he podido crecer durante diferentes etapas: Andrés, Andrea, Armando, Caro, Charles, Dany, Joaquín, Marc, Mariana y Mikel. Muchas gracias a aquellas personas con las que no solo inicié la carrera, sino que se mantuvieron cercanos a lo largo de estos años: Ale, Atanasio, Carina, Emiliano, Javi, Laura, Majo, Marco, Nubia, Roman, Tona y Vale. Asimismo, a aquellas personas con las cuales ha sido un verdadero placer coincidir en diferentes circunstancias: Ana, Chava, Fede, Jerry, Malú, Omar, Pablo y Roberto. Con amor a mis papás, que han dado todo por mí. A mi familia, sea biológica o por elección. En particular a mi mamá, pues no conozco mujer más fuerte y dedicada que ella.

Igualmente quiero agradecerle a todas las personas que me han cerrado las puertas, que me han decepcionado o que me han hecho sentir mal. Sin ustedes no me hubiera dado cuenta de lo que soy capaz de hacer ni hubiera podido aprender tanto sobre mí misma.

Resumen

El análisis de Fourier marca un antes y un después en la historia no solo de las matemáticas, sino también de disciplinas como la ingeniería. Es justamente a partir de esta rama que se empiezan a estudiar otras funciones conocidas como ondículas, las cuales tratan de atacar problemas específicos del análisis de Fourier así como extender el conocimiento de este tipo de transformaciones. Conforme la cantidad de información que recopilamos ha ido creciendo, también nos hemos visto en la necesidad de desarrollar herramientas que nos permitan entenderla. Varias de estas herramientas vienen del área de ciencias de la computación y son aplicadas en gran medida a problemas financieros, de mercadotecnia y económicos. En este trabajo queremos cambiar el objeto de estudio y llevarlo a la historia del arte mexicano.

Este trabajo fue inspirado en estudios hechos principalmente por Ingrid Daubechies para clasificar pinturas de diferentes épocas en la vida de Van Gogh. Sin embargo, no se han aplicado este tipo de técnicas en México, siendo éste el primer trabajo en intentarlo. Queremos combinar la bella teoría matemática junto con herramientas poderosas de aprendizaje de máquina para poder hacer análisis de obras de artistas mexicanos. Para estas obras queremos examinar las formas, los colores, los patrones y las texturas que hacen que cada artista se distinga de los otros.

Le queremos recordar al querido lector que no debe de temerle a la teoría que a continuación se va a mostrar. Así que, ¿cómo leer este trabajo? Si el lector ya tiene una buena base de series de Fourier y la transformada

de Fourier, se puede saltar al capítulo dos. El segundo capítulo hace una transición natural entre la transformada de Fourier y ondículas. El tercer capítulo plantea los fundamentos estadísticos y sobre aprendizaje de máquina que, junto con los dos capítulos anteriores, son necesarios para el cuarto capítulo. En el cuarto capítulo usamos todo lo que fuimos construyendo a lo largo del trabajo para hacer análisis estilístico de obras de arte y la clasificación de las mismas. A lo largo de este trabajo vamos a tener varios tipos de resultados, unos con valor teórico como los diversos teoremas encontrados en los primeros dos capítulos. Otros resultados se seguirán a partir del hecho que sí podemos combinar teoría matemática con cuestiones aplicadas como es el aprendizaje de máquina. El resultado final de este trabajo son diversos métodos computacionales que se sugieren para analizar obras de arte y compararlas con obras de varios artistas mexicanos.

Abstract

In the history of mathematics, Fourier analysis marks a milestone not just for pure mathematics, but also for many other areas such as engineering. Starting from this beautiful theory many famous mathematicians and engineers began to study other type of functions called wavelets. These functions began to be studied in order to solve many of the problems that Fourier analysis had on several applications. As the era of big data started to kick in, we have been in the need to develop new, powerful tools that would help us understand the overwhelming amount of data we are generating. Various of these tools come from computer science and have been applied to several problems in finance, marketing, and economics. In this thesis we want to change the scope of this study to Mexican art history.

This work was inspired in articles written mainly by Ingrid Daubechies, where she talks about classification techniques in order to distinguish Van Gogh's different brushstrokes during his lifetime. The importance of this thesis is that the techniques presented here have never been applied in a Mexican context, there is no such mathematical analysis for Mexican artists. Our objective lies on combining beautiful mathematics theory with powerful machine learning techniques in order to analyze artwork from some Mexican artists. We want to inspect the forms, colors, patterns, and textures that distinguish each of the Mexican artists considered.

We would like to remind our dear reader that he/she should not fear the beautiful theory that we are about to present. The first two chapters

will give the reader all the necessary theory he/she needs to understand the application we are trying to do. So, how to read this thesis? If the reader already has a good understanding on Fourier analysis, you might start reading on the second chapter. This chapter is a natural transition from the Fourier transform to wavelet analysis. The third chapter states the fundamental methods and techniques used in machine learning and computer vision to attack our problem. The fourth chapter combines all the others for the implementation and analysis of our problem.

Índice general

| | |
|---|------------|
| 1. Series y transformadas de Fourier | 1 |
| 1.1. Nociones básicas de análisis y de álgebra lineal | 1 |
| 1.2. Series de Fourier | 11 |
| 1.3. Teorema integral de Fourier | 26 |
| 1.4. Transformada de Fourier | 39 |
| 1.4.1. Transformada de Fourier en 2 dimensiones | 41 |
| 1.5. Transformadas discreta y rápida de Fourier | 42 |
| 1.6. Problemas con la transformada de Fourier | 47 |
| 1.7. Transformada de Fourier de tiempo corto | 51 |
| 2. Transformadas con Ondículas | 57 |
| 2.1. Otro tipo de representaciones | 58 |
| 2.2. Transformada discreta de ondícula | 66 |
| 2.3. Análisis multirresolución | 68 |
| 2.4. Análisis multirresolución y ondículas | 72 |
| 2.5. Transformada de ondícula discreta para 2D | 85 |
| 2.6. Ondículas específicas | 89 |
| 2.6.1. Ondícula de Haar | 89 |
| 2.6.2. Ondícula de Gabor | 96 |
| 2.6.3. Ondículas complejas de árbol dual (DTCWT) | 99 |
| 3. Métodos computacionales | 105 |
| 3.1. Métodos de obtención de datos a partir de imágenes | 106 |
| 3.1.1. Modelos de Markov ocultos | 106 |
| 3.1.2. Autoencoders | 114 |
| 3.1.3. Método GIST | 121 |

| | |
|---|------------|
| 3.1.4. Método SURF | 126 |
| 3.2. Métodos de clasificación | 130 |
| 3.2.1. Bosques aleatorios | 130 |
| 3.2.2. Máquinas de vector soporte (SVM) | 134 |
| 4. Clasificación de obras de arte | 140 |
| 4.1. Introducción a la historia del arte de inicio del siglo XX . | 141 |
| 4.2. Métodos usados | 149 |
| 4.3. Resultados | 151 |
| 4.4. Ejemplo de análisis | 154 |
| 4.5. Conclusiones | 159 |
| Bibliografía | 161 |

Capítulo 1

Series y transformadas de Fourier

Este capítulo presenta las bases del análisis de Fourier necesarias para el análisis de ondículas, del cual se centra gran parte de este trabajo. Discutiremos nociones básicas de álgebra lineal y análisis matemático hasta llegar a una teoría formal de lo que es el análisis de Fourier.

1.1. Nociones básicas de análisis y de álgebra lineal

Para empezar a analizar imágenes primero debemos definir algunas nociones básicas para poder formalizar los conceptos que emplearemos. En matemáticas esto nos sirve para, de una manera lógica, poder justificar todos los resultados que vayamos enunciando.

Tenemos que mencionar que la mayor parte del trabajo es realizado en los números complejos, \mathbb{C} . A pesar de esto, los números reales (\mathbb{R}) y los naturales (\mathbb{N}) también juegan un papel importante.

Empezaremos con la definición de sucesión, la cual es una función cuyo dominio son los naturales \mathbb{N} y su imagen son los complejos \mathbb{C} o los números reales \mathbb{R} .

La mayor parte del tiempo en este trabajo vamos a estar moviéndonos en espacios vectoriales, los cuales denotaremos como \mathcal{V} . Una base es aquél conjunto de elementos de \mathcal{V} que, bajo ciertas condiciones, determinan por completo a todos los elementos de \mathcal{V} .

Definición 1. Una base es un conjunto \mathcal{W} de elementos w_i , (finito o infinito) de \mathcal{V} los cuales:

- son linealmente independientes entre sí, es decir la única solución para $\alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n = 0$ con $\alpha_k \in \mathbb{C}$ es que todos los escalares α_k sean iguales a 0.
- generan por completo a \mathcal{V} , es decir para todo $z \in \mathcal{V}$ existen escalares únicos α_k tales que $z = \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n$, esto lo escribimos como $\langle \mathcal{W} \rangle = \mathcal{V}$. En este contexto, $\langle \mathcal{W} \rangle$ se refiere al conjunto generado por los elementos en \mathcal{W} , es decir todos los elementos que se pueden formar a partir de combinaciones lineales de los w_i .

A lo largo de este trabajo vamos a estar moviéndonos en espacios vectoriales, cuya definición encontramos a continuación.

Definición 2. Un espacio vectorial \mathcal{V} sobre un campo \mathbb{K} (como \mathbb{R} o \mathbb{C}) es un conjunto tal que con las operaciones de adición vectorial y multiplicación por un escalar se satisfacen ciertas condiciones. Estas condiciones son las siguientes para cualesquiera $v, u, w \in \mathcal{V}$, $\alpha, \beta \in \mathbb{K}$:

- $u + (v + w) = (u + v) + w$.
- $u + v = v + u$.
- Existe un elemento, $0 \in \mathcal{V}$ llamado cero tal que $v + 0 = v$. Este elemento es el neutro aditivo.
- Para cualquier $v \in \mathcal{V}$ existe $-v$ llamado inverso aditivo tal que $v + (-v) = 0$.
- $\alpha(\beta v) = (\alpha\beta)v$.

- Existe un elemento, 1 llamado neutro multiplicativo tal que $1v = v$.
- $\alpha(u + v) = \alpha u + \alpha v$.
- $(\alpha + \beta)v = \alpha v + \beta v$.

Nuestra siguiente noción importante para el desarrollo de este trabajo es el de producto interior.

Definición 3. Dado un espacio vectorial \mathcal{V} , un producto interior es una función $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{C}$ (o \mathbb{R} en su caso) la cual, para cualesquiera $w, z \in \mathcal{V}$ y $\alpha \in \mathbb{C}$ entonces tiene las siguientes propiedades:

- positivo definido: $\langle w, w \rangle \geq 0$ y $\langle w, w \rangle = 0 \Leftrightarrow w = 0$
- es simétrico conjugado: $\langle w, z \rangle = \overline{\langle z, w \rangle}$. Es decir, si $\langle z, w \rangle = \alpha + i\beta$, para algunos $\alpha, \beta \in \mathbb{R}$, entonces $\overline{\langle z, w \rangle} = \alpha - i\beta$ es su conjugado.
- linealidad en el primer argumento: $\langle \alpha w, z \rangle = \alpha \langle w, z \rangle$.
- suma: $\langle w + z, v \rangle = \langle w, v \rangle + \langle z, v \rangle$.

Ya tenemos un espacio vectorial dotado con un producto interior, a este par de objetos matemáticos se le llama un espacio pre-Hilbert (como dice su nombre, nos faltan algunas características para poder llamarlo de Hilbert). Así como tenemos esta función, también hay otra importante llamada norma, con la cual podemos medir de diferentes maneras elementos en nuestro espacio vectorial (dependiendo la estructura de dicha norma).

Definición 4. Dado un espacio vectorial \mathcal{V} , una norma es una función que mapea este espacio vectorial en números reales positivos, \mathbb{R}^+ , $\|\cdot\| : \mathcal{V} \rightarrow \mathbb{R}^+$ tal que para todo $w, z \in \mathcal{V}, \alpha \in \mathbb{C}$ cumple con:

- es no negativa $\|w\| \geq 0$ y $\|w\| = 0 \Leftrightarrow w = 0$,
- cumple con la desigualdad del triángulo $\|w + z\| \leq \|w\| + \|z\|$,
- es absolutamente escalable: $\|\alpha w\| = |\alpha| \|w\|$.

Hay normas que podemos construir basándonos en productos internos llamadas normas inducidas por algún producto interior, pero no necesariamente tiene que pasar esto para que una función se pueda llamar norma. Algunas normas con las que vamos a trabajar son:

- Norma del supremo, si D es el dominio de f y $f : D \rightarrow \mathbb{R}$ ó $f : D \rightarrow \mathbb{C}$:

$$\|f\|_{\infty} = \|f\|_{\infty, D} = \sup_x \{|f(x)| : x \in D\},$$

es decir el supremo de $|f(x)|$ entre todas las $x \in D$ posibles. En el caso de que ese número sí se alcance, es decir en el caso que exista \hat{x} tal que $f(x) \leq f(\hat{x})$ para toda $x \in D$ entonces a $f(\hat{x})$ le llamamos máximo. Pero recordemos que no necesariamente se puede alcanzar esta cota superior.

- Normas p , donde $p \in \mathbb{N}$:

$$\|f\|_p = \left(\int_D |f(x)|^p dx \right)^{1/p},$$

donde D es el dominio de f . La norma p más usada es cuando $p = 2$, como el querido lector va a leer a continuación. Si el querido lector está familiarizado con teoría de la medida o análisis funcional, se podrá dar cuenta que a partir de estas normas es como se caracterizan los espacios de Lebesgue \mathcal{L}_p , los cuales son espacios clásicos de Banach, pero no abordaremos este tema a profundidad.

Con la definición de norma y de producto interno podemos clasificar a distintas bases de un mismo espacio vectorial de acuerdo a las características de sus elementos:

- Diremos que una base es normal si la norma de todos sus elementos es igual a uno.
- Diremos que una base es ortogonal si el producto interno de cualesquiera dos elementos diferentes es cero, es decir $\langle w_k, w_j \rangle = 0$ para cualquier $k \neq j$.

- Diremos que una base es ortonormal si es normal y ortogonal al mismo tiempo, a este tipo especial de bases las denotaremos como B.O.N. y se escriben como \mathcal{S} .

Ya que tenemos definido el concepto de norma, podemos medir distancias entre elementos de nuestro espacio vectorial \mathcal{V} . Con esto podemos también estudiar diferentes formas en las que sucesiones dentro de este espacio vectorial se acercan o no a algún elemento. Este concepto es el de convergencia y hay varias formas de convergencia incluso usando una misma norma. Por el tipo de convergencias que vamos a usar, definiremos algunos tipos de convergencias para sucesiones de funciones.

Definición 5. Una sucesión $(f_k(x))_{k \in \mathbb{N}}$ donde para cada k , $f_k : \mathbb{R} \rightarrow \mathbb{R}$ se dice que converge puntualmente a $f(x)$ en norma $\|\cdot\|$ si para todo $\varepsilon > 0$ existe $N_{\varepsilon, x} \in \mathbb{N}$ tal que para todo $k > N_{\varepsilon, x}$ tenemos que $\|f_k(x) - f(x)\| < \varepsilon$. Esto es lo mismo que afirmar que para cada x , existe alguna $N_\varepsilon \in \mathbb{N}$ tal que si $k > N_\varepsilon$ entonces $\|f_k(x) - f(x)\| < \varepsilon$, pero las N_ε para diferentes valores de x no necesariamente son iguales.

Con esta definición tenemos que cuando una sucesión de funciones converge a una función la N correspondiente a cada x puede variar enormemente. Una definición que limita esto es la de convergencia uniforme, la cual es un tipo de convergencia mucho más fuerte que la puntual.

Definición 6. Una sucesión $(f_k(x))_{k \in \mathbb{N}}$ donde para cada k , $f_k : \mathbb{R} \rightarrow \mathbb{R}$, converge uniformemente a f en norma si para todo $\varepsilon > 0$ existe $N_\varepsilon \in \mathbb{N}$ tal que para todo $k > N_\varepsilon$ tenemos que $\|f_k - f\|_\infty < \varepsilon$, para cualquier x en el dominio de las f_k .

Notemos que para convergencia uniforme la N usada no depende de x , sino tan solo del error, ε . Entonces para cada x la función converge con la misma N , la cual solo depende de la ε .

El conjunto de funciones que son tanto periódicas (de periodo $2l$) como continuas, las denotamos con $\mathcal{PC}[-l, l]$. Estas funciones forman un espacio vectorial sobre el campo que vayamos a usar, ya sea \mathbb{C} (la

mayor parte del tiempo) o \mathbb{R} .

En este trabajo, usualmente nuestro dominio y nuestra imagen serán los números reales, entonces para funciones se escribirá que $f : \mathbb{R} \rightarrow \mathbb{R}$. En algunas situaciones usaremos funciones con dominio e imagen en los números complejos, es decir cuando $f : \mathbb{C} \rightarrow \mathbb{C}$. Es decir, el dominio o los valores para los cuales la función está definida es \mathbb{C} y la imagen, o los valores que puede llegar a tomar la función también es \mathbb{C} . El curioso lector se podrá preguntar cómo se define la periodicidad en este conjunto. Se define de manera análoga a funciones reales periódicas de dos variables pues todo número complejo se puede escribir de la forma $z = x + iy$ donde $z \in \mathbb{C}$, $x, y \in \mathbb{R}$. A este tipo de funciones se les conoce como doblemente periódicas pues tiene periodo en la componente real (T_1) y en la componente imaginaria (T_2):

$$f(z) = f(x + iy) = f((x + T_1) + i(z + T_2)),$$

donde $T_1, T_2 \in \mathbb{R}$ y ambos son números positivos.

Las funciones periódicas reales más empleadas y conocidas son funciones como senos, cosenos, tangentes, etc, son las funciones trigonométricas. Así también podemos definir polinomios basándonos en sumas de estas funciones trigonométricas.

Definición 7. *Un polinomio trigonométrico definido en los reales es una función de la forma:*

$$t(x) = a + \sum_{k=0}^n (a_k \cos(kx) + b_k \sin(kx)) \quad (1.1)$$

donde $a, x, a_k, b_k \in \mathbb{R}$, $n \in \mathbb{N}$.

Notemos que para $n \in \mathbb{N}$ el conjunto de polinomios trigonométricos de grado a lo más igual a n forman un campo vectorial sobre \mathbb{R} de grado $2n + 1$.

No solo podemos tener espacios vectoriales de dimensión finita. Notemos que la dimensión de un espacio vectorial viene dado por la

cardinalidad de alguna de sus bases, pero también podemos tener conjuntos con una cantidad infinita de elementos los cuales también sean la base de un espacio vectorial. Estos espacios se llaman espacios vectoriales de dimensión infinita. Un ejemplo de ellos son los polinomios trigonométricos, pero sin estar condicionados a tener cierto grado.

Es natural preguntarse por una base para este espacio vectorial, el cual denotaremos con \mathcal{T} . Una base que surge de manera natural es la siguiente: $\mathcal{B} = \{1, \text{sen}(kx), \cos(kx)\}_{k=1}^{\infty}$.

La siguiente pregunta que el curioso lector seguramente se estará haciendo es sobre la definición del producto interno en este espacio vectorial.

Definición 8. *El producto interno con el que vamos a trabajar dentro del campo vectorial $\mathcal{PC}[-l, l]$ se define como el siguiente:*

$$\langle f, g \rangle = \int_{-l}^l f(x) g(x) dx. \quad (1.2)$$

Con esta definición ahora podemos estructurar normas inducidas por este producto interno.

Definición 9. *Con el producto interno definido anteriormente, la p -norma para f en $\mathcal{PC}[-l, l]$ es:*

$$\|f\|_p = \left(\int_{-l}^l |f(x)|^p dx \right)^{\frac{1}{p}} \quad (1.3)$$

Con esto obtenemos una definición ampliamente usada en teoría de la medida, los espacios \mathcal{L}_p y los espacios l_p . Con esto queremos decir que $f \in \mathcal{L}_p$ si $\|f\|_p^p < \infty$. Y para sucesiones numéricas, $(x_k)_{k \in \mathbb{N}}$, tenemos que está en l_p si $\sum_{k=1}^{\infty} |x_k|^p < \infty$. Si el estimado lector está familiarizado con teoría de la medida entonces se podrá dar cuenta que estas definiciones son análogas pues en la primera la medida usada en la integral es la medida de Lebesgue, mientras que en la segunda la

medida usada es la medida de conteo.

Con estas definiciones y recordando que los polinomios trigonométricos son de periodo 2π , podemos normalizar (con la norma 2) la base que hemos dado anteriormente, $\mathcal{B} = \{1, \operatorname{sen}(kx), \cos(kx)\}_{k=1}^{\infty}$. Esta nueva base es ortonormal, lo cual nos ayudará para futuros procesos en este trabajo.

Teorema 1. *Una base ortonormal para $\mathcal{PC}[-\pi, \pi]$ es la siguiente:*

$$\mathcal{S} = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos(kx)}{\sqrt{\pi}}, \frac{\operatorname{sen}(kx)}{\sqrt{\pi}} \right\}_{k \in \mathbb{N}}.$$

Demostración. Por la definición de base ortonormal debemos de probar que la norma de todos estos elementos es igual a uno, el producto interno de cualesquiera dos elementos diferentes es cero y que genera a todo $\mathcal{PC}[-\pi, \pi]$. Todo esto lo vamos a probar con la norma $p = 2$ o norma 2.

1. Primero demostramos que la norma de estos elementos es igual a uno.

$$\left\| \frac{1}{\sqrt{2\pi}} \right\|_2^2 = \int_{-\pi}^{\pi} \frac{1}{2\pi} dx = 1.$$

$$\begin{aligned} \left\| \frac{\cos(kx)}{\sqrt{\pi}} \right\|_2^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos^2(kx) dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 + \cos(2kx) dx \\ &= \frac{1}{2\pi} \left(x + \frac{\operatorname{sen}(2kx)}{2k} \right) \Big|_{-\pi}^{\pi} = 1. \end{aligned}$$

$$\begin{aligned} \left\| \frac{\operatorname{sen}(kx)}{\sqrt{\pi}} \right\|_2^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{sen}^2(kx) dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 - \cos(2kx) dx \\ &= \frac{1}{2\pi} \left(x - \frac{\operatorname{sen}(2kx)}{2k} \right) \Big|_{-\pi}^{\pi} = 1. \end{aligned}$$

2. Ahora probamos que el producto interno de cualesquiera dos elementos diferentes es cero. A continuación vamos a considerar que $k, j \neq 0$ y que $k \neq j$.

$$\left\langle \frac{1}{\sqrt{2\pi}}, \frac{\cos(kx)}{\sqrt{\pi}} \right\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \cos(kx) dx = 0.$$

$$\left\langle \frac{1}{\sqrt{2\pi}}, \frac{\sin(kx)}{\sqrt{\pi}} \right\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \sin(kx) dx = 0.$$

$$\begin{aligned} \left\langle \frac{\sin(kx)}{\sqrt{\pi}}, \frac{\cos(jx)}{\sqrt{\pi}} \right\rangle &= \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(jx) \sin(kx) dx \\ &= \frac{1}{\pi} \left(\frac{-1}{k} \cos(jx) \cos(kx) \right) \Big|_{-\pi}^{\pi} - \\ &\quad \frac{j}{k} \int_{-\pi}^{\pi} \sin(jx) \sin(kx) dx \\ &= \frac{-j}{\pi k} \int_{-\pi}^{\pi} \sin(jx) \sin(kx) dx \\ &= \frac{-j}{\pi k} \left(\frac{-1}{k} \sin(jx) \sin(kx) \right) \Big|_{-\pi}^{\pi} - \\ &\quad \frac{j}{k} \int_{-\pi}^{\pi} \cos(jx) \sin(kx) dx \\ &= \frac{-j^2}{\pi k^2} \int_{-\pi}^{\pi} \cos(jx) \sin(kx) dx \\ &\Rightarrow \left(\frac{1}{\pi} + \frac{-j^2}{\pi k^2} \right) \int_{-\pi}^{\pi} \cos(jx) \sin(kx) dx = 0 \\ &\Rightarrow \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(jx) \sin(kx) dx = 0. \end{aligned}$$

$$\begin{aligned}
\left\langle \frac{\cos(jx)}{\sqrt{\pi}}, \frac{\cos(jx)}{\sqrt{\pi}} \right\rangle &= \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(jx) \cos(kx) dx \\
&= \frac{1}{\pi} \left(\frac{-1}{k} \cos(jx) \sin(kx) \right) \Big|_{-\pi}^{\pi} + \\
&\quad \frac{j}{k} \int_{-\pi}^{\pi} \sin(jx) \sin(kx) dx \\
&= \frac{j}{\pi k} \int_{-\pi}^{\pi} \sin(jx) \sin(kx) dx \\
&= \frac{j}{\pi k} \left(\frac{-1}{k} \sin(jx) \cos(kx) \right) \Big|_{-\pi}^{\pi} + \\
&\quad \frac{j}{k} \int_{-\pi}^{\pi} \cos(jx) \cos(kx) dx \\
&= \frac{j^2}{\pi k^2} \int_{-\pi}^{\pi} \cos(jx) \cos(kx) dx,
\end{aligned}$$

lo cual implica que:

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \cos(jx) \cos(kx) dx = 0.$$

$$\begin{aligned}
\left\langle \frac{\sin(jx)}{\sqrt{\pi}}, \frac{\sin(jx)}{\sqrt{\pi}} \right\rangle &= \frac{1}{\pi} \int_{-\pi}^{\pi} \sin(jx) \sin(kx) dx \\
&= \frac{1}{\pi} \left(\frac{-1}{k} \sin(jx) \cos(kx) \right) \Big|_{-\pi}^{\pi} + \\
&\quad \frac{j}{k} \int_{-\pi}^{\pi} \cos(jx) \cos(kx) dx = 0.
\end{aligned}$$

3. Para probar que este conjunto genera a todo $\mathcal{PC}[-\pi, \pi]$ requerimos convergencias relacionas a las series de Fourier. A continuación vamos a dar algunos criterios de convergencias, los cuales como puede ver el lector, son criterios para poder escribir funciones en $\mathcal{PC}[-\pi, \pi]$ como combinación lineal de este conjunto. Si el curioso lector lo desea, puede leer más sobre este último punto en [1].

■

1.2. Series de Fourier

Por ahora, regresemos a cualquier espacio vectorial, \mathcal{V} . Ahora supongamos que dicho espacio vectorial es de dimensión finita (para facilitar nuestros cálculos) y tenemos una base ortonormal de dicho espacio vectorial, llamémosla $\mathcal{B} = \{w_1, w_2, \dots, w_n\}$. Dado que \mathcal{B} es un conjunto con elementos que son linealmente independientes entre sí y genera a \mathcal{V} entonces podemos escribir de manera única a cualquier $z \in \mathcal{V}$ como combinación lineal de las w_i , es decir:

$$z = \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n.$$

Pero, ¿cómo sabemos cuáles coeficientes α_k usar? Echemos un vistazo a los productos interiores de los elementos de \mathcal{B} con z :

$$\begin{aligned} \langle w_k, z \rangle &= \langle w_k, \alpha_1 w_1 + \alpha_2 w_2 + \dots + \alpha_n w_n \rangle \\ &= \langle w_k, \alpha_1 w_1 \rangle + \langle w_k, \alpha_2 w_2 \rangle + \dots + \langle w_k, \alpha_n w_n \rangle \\ &= \langle w_k, \alpha_k w_k \rangle = \alpha_k \langle w_k, w_k \rangle = \alpha_k. \end{aligned}$$

Con esto tenemos que $\langle w_k, z \rangle = \alpha_k$ y así:

$$z = \langle w_1, z \rangle w_1 + \langle w_2, z \rangle w_2 + \dots + \langle w_n, z \rangle w_n = \sum_{k=1}^n \langle w_k, z \rangle w_k.$$

Los coeficientes $\langle w_k, z \rangle$ se llaman los coeficientes abstractos de Fourier y la manera de expresar a z de esta forma se llama la expansión de Fourier de z .

Ahora pasemos a un problema de optimización (sencillo). Pensemos que tenemos un espacio pre-Hilbert de dimensión finita (es decir un espacio vectorial \mathcal{V} dotado de un producto interno $\langle \cdot, \cdot \rangle$), una base ortonormal de dicho espacio, $\mathcal{B} = \{w_1, w_2, \dots, w_n\}$ y alguien nos pide resolver el siguiente problema de minimización:

$$\min_{\alpha_1, \alpha_2, \dots, \alpha_n} \left\| z - \sum_{k=1}^n \alpha_k w_k \right\|_2.$$

Lo primero que haríamos es analizar la norma dos de este elemento del espacio vectorial sobre \mathbb{R} , elevamos al cuadrado para facilitar los cálculos:

$$\begin{aligned}
\left\| z - \sum_{k=1}^n \alpha_k w_k \right\|_2^2 &= \left\langle z - \sum_{k=1}^n \alpha_k w_k, z - \sum_{k=1}^n \alpha_k w_k \right\rangle \\
&= \langle z, z \rangle - \sum_{k=1}^n \alpha_k \langle z, w_k \rangle - \sum_{k=1}^n \alpha_k \langle w_k, z \rangle \\
&\quad + \sum_{k=1}^n \sum_{j=1}^n \alpha_k \alpha_j \langle w_k, w_j \rangle \\
&= \langle z, z \rangle - 2 \sum_{k=1}^n \alpha_k \langle z, w_k \rangle + \sum_{k=1}^n \alpha_k^2 \\
&= \langle z, z \rangle + \sum_{k=1}^n (-2\alpha_k \langle z, w_k \rangle + \alpha_k^2) \\
&= \langle z, z \rangle + \sum_{k=1}^n (-\langle z, w_k \rangle^2 + \langle z, w_k \rangle^2 \\
&\quad - 2\alpha_k \langle z, w_k \rangle + \alpha_k^2) \\
&= \langle z, z \rangle + \sum_{k=1}^n (-\langle z, w_k \rangle^2) \\
&\quad + \sum_{k=1}^n (\alpha_k - \langle z, w_k \rangle)^2 \\
&= \underbrace{\langle z, z \rangle - \sum_{k=1}^n \langle z, w_k \rangle^2}_I \\
&\quad + \underbrace{\sum_{k=1}^n (\alpha_k - \langle z, w_k \rangle)^2}_{II}.
\end{aligned}$$

Con esto podemos notar que el primer término, I , no depende de las α_k , mientras que el segundo es una suma de cuadrados que sí depende de las

α_k . Por ende, $\|z - \sum_{k=1}^n \alpha_k w_k\|_2^2$ se minimiza cuando $II = 0$. Volvemos a tener el mismo resultado para las α_k que el problema anterior, es decir que $\alpha_k = \langle z, w_k \rangle$, la solución son los coeficientes de Fourier (lo cual no nos debe sorprender después de trabajar con el problema anterior). Con esto obtenemos que nuestro óptimo tiene el valor de:

$$\min_{\alpha_1, \alpha_2, \dots, \alpha_n} \|z - \sum_{k=1}^n \alpha_k w_k\|_2 = \sqrt{I} = \sqrt{\langle z, z \rangle - \sum_{k=1}^n \langle z, w_k \rangle^2}.$$

Notemos que de este problema de minimización, obtenemos varias propiedades de nuestros coeficientes abstractos de Fourier de manera inmediata.

Corolario (desigualdad de Bessel finita). *Sea $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ un espacio pre-Hilbert y $\mathcal{B} = \{w_1, \dots, w_n\}$ una base ortonormal de \mathcal{V} , entonces tenemos la siguiente desigualdad para los coeficientes de Fourier:*

$$\sum_{k=1}^n \langle z, w_k \rangle^2 \leq \langle z, z \rangle$$

para cualquier $z \in \mathcal{V}$.

Pero esta desigualdad también la podemos aplicar para espacios vectoriales de dimensión infinita, basando nuestra demostración en el teorema de la convergencia monótona.

Corolario (desigualdad de Bessel numerable). *Sea $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ un espacio pre-Hilbert, $\mathcal{B} = \{w_1, \dots\}$ una base ortonormal de \mathcal{V} , entonces tenemos la siguiente desigualdad para los coeficientes de Fourier:*

$$\sum_{k=1}^{\infty} \langle z, w_k \rangle^2 \leq \langle z, z \rangle \quad (1.4)$$

para cualquier $z \in \mathcal{V}$.

El siguiente lema es una de las diversas versiones del lema de Riemann-Lebesgue. La prueba del caso más general se basa en diversas nociones y teoremas de teoría de la medida, pero dicho caso no lo vamos a usar en este trabajo pues no es necesario tal nivel de abstracción, basta con la versión que.

Lema (lema de Riemann-Lebesgue). Sea $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ un espacio pre-Hilbert y $\mathcal{B} = \{w_1, \dots\}$ una base ortonormal de \mathcal{V} , entonces para todo $z \in \mathcal{V}$:

$$\lim_{k \rightarrow \infty} \langle z, w_k \rangle = 0.$$

Demostración. Por la desigualdad de Bessel numerable denotada en 1.4, tenemos que $\sum_{k=1}^{\infty} \langle z, w_k \rangle^2 \leq \langle z, z \rangle < \infty$, es decir que los últimos términos de esta suma tienden a cero cuando k tiende a infinito, $\lim_{k \rightarrow \infty} \langle z, w_k \rangle^2 = 0$. Esto nos dice que $\lim_{k \rightarrow \infty} |\langle z, w_k \rangle| = 0$, por ende $\lim_{k \rightarrow \infty} \langle z, w_k \rangle = 0$ ■

Si el querido lector fue atento, todo este tiempo hemos hablado de coeficientes de Fourier abstractos. ¿Por qué son abstractos? Esto se debe a que no hemos señalado ningún espacio vectorial concreto o producto interno con el que trabajamos. En el caso histórico, dicho espacio vectorial son las funciones periódicas continuas a pedazos y el producto interno en dicho espacio es el de la definición 1.2.

¿Por qué trabajar con polinomios trigonométricos? ¿A quién se le ocurrió todo esto? (Evidentemente la respuesta es Fourier.)

Como se puede leer en el libro de Bell en [2], después de haber acompañado a Napoleón a sus campañas por Egipto, cierto matemático francés llamado Jean-Baptiste Joseph Fourier se empezó a interesar por estudiar cambios del calor sobre superficies. Hoy en día a esto se le conoce como la ecuación del calor. Tal vez tuviera algo que ver con su experiencia en Egipto. ¿Qué es esta famosa ecuación del calor? Es la forma en la que el calor se esparce en un cuerpo a través del tiempo. Sí, esto involucra ecuaciones diferenciales parciales:

$$\frac{\partial^2 u(x, t)}{\partial x^2} = \frac{\partial u(x, t)}{\partial t}, 0 < x < \pi, 0 < t < T$$

$$u(0, t) = u(\pi, t) = 0, 0 \leq t \leq T, u(x, 0) = f(x), 0 \leq x \leq \pi$$

No nos preocupemos si no se han definido a las ecuaciones diferenciales parciales en este trabajo, no las usaremos. Lo único que quiere decir

esto es que Fourier buscaba cómo se distribuía el calor u en el tiempo t que va de cero a alguna T a través de alguna superficie (placa, hoja, algo plano, tan plano que se encuentra en un espacio de dos dimensiones). Esta superficie convenientemente mide π unidades de largo (centímetros, pies, kilómetros, lo que sea). La condición que manejó Fourier fue que en ambos extremos de esta superficie el calor fuera igual a cero.

Antes de Fourier hubo otros personajes que también se interesaron por la llamada ecuación del calor, como d'Alembert y Bernoulli. Pero fue a Fourier a quien se le ocurrió que la solución podría ser la suma de senos y cosenos pues las soluciones debían de ser superposiciones de soluciones previamente dadas. De hecho, Fourier propuso soluciones que tenían la siguiente forma:

$$u(x, t) = \sum_{k=1}^{\infty} f_k \exp(-k^2 t) \operatorname{sen}(kx),$$

donde:

$$f_k = \frac{2}{\pi} \int_0^{\pi} f(x) \operatorname{sen}(kx) dx, \forall k \in \mathbb{N}.$$

De ahí nacen lo que conocemos como los coeficientes de Fourier (históricos) que vinieron a revolucionar el análisis matemático. Cabe mencionar que el trabajo de Fourier no fue matemáticamente formal, como hoy en día se conoce. Este desarrollo se fue haciendo a lo largo de los años.

Ahora vamos a describir y dar algunas propiedades de las llamadas series de Fourier. Como ya lo hemos hecho, vamos a denotar como $P([0, 2\pi])$ al conjunto de funciones periódicas de periodo 2π y a $PC([0, 2\pi])$ como a las funciones que además de ser periódicas son continuas por partes, es decir a las funciones continuas, excepto en un número finito de puntos. Pero con todo lo que ya hemos mencionado y trabajado anteriormente tenemos una buena base (literal y figurativamente) para poder trabajar con las series de Fourier de una

variable real.

Ahora ya tenemos la base ortonormal de los polinomios trigonométricos que vamos a usar, $\mathcal{S} = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos(kx)}{\sqrt{\pi}}, \frac{\sin(kx)}{\sqrt{\pi}} \right\}_{k \in \mathbb{N}}$. Entonces podemos aterrizar la teoría de los coeficientes de Fourier abstractos a los coeficientes de Fourier históricos para una función real $f \in PC([0, 2\pi])$ (tenemos tres tipos, uno ligado a la componente constante, uno al coseno y otro al seno):

$$\begin{aligned} A &= \frac{a_0}{2} = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx, \\ a_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx, \\ b_k &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx. \end{aligned} \tag{1.5}$$

Estos son los $\langle w_i, z \rangle$ que habíamos discutido en el caso abstracto, solo que ya tenemos idea de cómo se ven puesto que tenemos tres tipos distintos de w_i . Y ahora, la siguiente serie se llama la serie de Fourier de la función real f :

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)).$$

Usamos el símbolo \sim en lugar de $=$ porque no sabemos si esta serie converge, mucho menos sabemos si converge a f y cómo es dicha convergencia.

Procedemos a establecer la desigualdad de Bessel y una versión del lema de Riemann-Lebesgue para esta base ortonormal:

Corolario (desigualdad de Bessel numerable).

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty} (a_k^2 + b_k^2) \leq \frac{1}{\pi} \int_0^{2\pi} f^2(x) dx. \tag{1.6}$$

Lema (lema de Riemann-Lebesgue).

$$\lim_{k \rightarrow \infty} a_k = 0, \lim_{k \rightarrow \infty} b_k = 0. \quad (1.7)$$

Ahora juguemos un poco con el álgebra de la k -ésima suma de la serie de Fourier (notemos que, como $f \in \mathcal{PC}[0, 2\pi]$, entonces podemos integrar en el intervalo $[0, 2\pi]$ o en el intervalo $[-\pi, \pi]$ obteniendo resultados idénticos):

$$\begin{aligned} S_n(f)(x) &= \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos(kx) + b_k \sin(kx)) = \frac{1}{2} \left(\frac{1}{\pi} \int_{-\pi}^{\pi} f(t) dt \right) \\ &+ \sum_{k=0}^n \left(\frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(kt) dt \right) \cos(kx) \\ &+ \left(\frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(kt) dt \right) \sin(kx) \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left(\frac{1}{2} + \sum_{k=1}^n (\cos(kt) \cos(kx) + \sin(kt) \sin(kx)) \right) dt \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left(\frac{1}{2} + \sum_{k=1}^n \cos(k(t-x)) \right) dt \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) D_n(t-x) dt = \frac{1}{\pi} \int_{-\pi-x}^{\pi-x} f(s+x) D_n(s) ds \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+t) D_n(t) dt. \end{aligned}$$

Donde $D_n(t-x) = \frac{1}{2} + \sum_{k=1}^n \cos(k(t-x))$, pero también lo podemos escribir como:

$$D_n(t) = \begin{cases} \frac{\sin((n+\frac{1}{2})t)}{2 \sin(\frac{1}{2}t)}, & \text{si } t \neq 2\pi j, \\ n + \frac{1}{2}, & \text{si } t = 2\pi j, \end{cases} \quad (1.8)$$

para $j \in \mathbb{Z}$. D_n es el núcleo de Dirichlet que facilita enormemente muchas demostraciones y que acorta la notación.

Johann Peter Gustav Lejeune Dirichlet fue un matemático alemán relevante en el ámbito de teoría de números y análisis de Fourier, como menciona Bell en [2]. Curiosamente se le atribuye a él la primera definición formal de función. Notemos que él vivió entre 1805 y 1859, recalcando el hecho que la formalización de las matemáticas es un hecho bastante reciente. Dirichlet estuvo en contacto con el mismo Fourier y con Poisson mientras el primero estudiaba la dispersión del calor. Cabe mencionar que también tuvo contacto con figuras como Alexander von Humboldt, Goerg Ohm y fue asesor de Rudolf Lipschitz. Su aportación tuvo impacto en varias áreas (incluso en biología), pero aquí nos delimitaremos al impacto que tiene el núcleo de Dirichlet para el estudio de la convergencia de series de Fourier.

Algunas propiedades que tiene este importante núcleo son:

1. D_n es una función continua, pudiéndose verificar de su definición alterna en 1.8
2. el núcleo de Dirichlet es una función par y también es un polinomio trigonométrico

3.

$$\frac{1}{\pi} \int_{-\pi}^{\pi} D_n(t) dt = 1 \quad (1.9)$$

4. alcanza su máximo cuando $t = 0$, siendo su valor máximo igual a $n + \frac{1}{2}$

Los siguientes dos teoremas son de suma importancia, pues nos dicen bajo qué condiciones y cómo convergen las series de Fourier. Hay que resaltar el hecho de que estos teoremas no son los únicos referentes a la convergencia de estas series, pues bajo diferentes condiciones se pueden llegar a diferentes convergencias.

Teorema 2 (convergencia puntual de Dirichlet). *Sea $f \in \mathcal{PC}[0, 2\pi]$, supongamos que $f'(x_0)$ existe para alguna x_0 entonces $\lim_{n \rightarrow \infty} S_n(f)(x_0) = f(x_0)$, lo cual es equivalente a afirmar que para esta x_0 :*

$$f(x_0) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx_0) + b_k \sin(kx_0)).$$

Como esto se cumple para esta particular x_0 , este teorema es de convergencia puntual.

Demostración. Primero hagamos algunas manipulaciones algebraicas con la siguiente resta (gracias a la propiedad 1.9 podemos hacer esto):

$$\begin{aligned}
S_n(f)(x_0) - f(x_0) &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x_0 + t) D_n(t) dt - f(x_0) \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x_0 + t) D_n(t) dt - f(x_0) 1 \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x_0 + t) D_n(t) dt - f(x_0) \frac{1}{\pi} \int_{-\pi}^{\pi} D_n(t) dt \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x_0 + t) D_n(t) dt - \frac{1}{\pi} \int_{-\pi}^{\pi} f(x_0) D_n(t) dt \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} (f(x_0 + t) - f(x_0)) D_n(t) dt \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} (f(x_0 + t) - f(x_0)) \frac{t}{2 \operatorname{sen}(\frac{1}{2}t)} \operatorname{sen} \left(\left(n + \frac{1}{2} \right) t \right) dt
\end{aligned}$$

Ahora, como la derivada de f en x_0 existe y por la definición de D_n , sea:

$$G(t) = \begin{cases} \frac{f(x_0+t)-f(x_0)}{t} \left(\frac{t}{2 \operatorname{sen}(\frac{1}{2}t)} \right) & \text{si } 0 < |t| \leq \pi \\ f'(x_0) & \text{si } t = 0. \end{cases}$$

Con esto tenemos que $G(t) : [0, 2\pi] \rightarrow \mathbb{R}$ es una función continua y así:

$$S_n(x_0) - f(x_0) = \int_0^{2\pi} G(t) \operatorname{sen} \left(\left(n + \frac{1}{2} \right) t \right) dt.$$

Usando la versión del lema de Riemann - Lebesgue mencionado anteriormente en 1.7:

$$\lim_{n \rightarrow \infty} (S_n(x_0) - f(x_0)) = \lim_{n \rightarrow \infty} \int_0^{2\pi} G(t) \operatorname{sen} \left(\left(n + \frac{1}{2} \right) t \right) dt = 0,$$

quedando demostrado este teorema. ■

El siguiente teorema nos da una convergencia mucho más fuerte con hipótesis más restringidas, la convergencia uniforme.

Teorema 3 (convergencia uniforme y absoluta). *Sea $f \in \mathcal{PC}[0, 2\pi]$, y supongamos también que f es de clase \mathcal{C}^1 (su derivada existe y es continua) entonces*

$$\lim_{n \rightarrow \infty} S_n(f)(x) = f(x) \quad \forall x \in \mathbb{R}$$

uniformemente y

$$\left| \frac{a_0}{2} \right| + \sum_{k=1}^{\infty} |a_k| + |b_k| < \infty.$$

Para poder hacer esta demostración usaremos un teorema de teoría de la medida, derivado del poderoso teorema de la convergencia dominada de Lebesgue (TCDL). No vamos a demostrar este teorema pues se escapa de nuestro objeto de estudio, pero el querido lector puede leer más al respecto en los libros de Rudin o Grabinisky en [3, 4].

Teorema 4. *Sea (X, S, μ) un espacio de medida y $f : X \times [a, b] \rightarrow \mathbb{R}$ tal que:*

1. $f^t : X \rightarrow \mathbb{R}$ (la t -sección de f , t fija, x no fija) es medible para toda t en su dominio y es integrable respecto a μ .
2. $\frac{\partial f}{\partial t} \Big|_{(x,t)}$ existe para cualquier par (x, t)
3. existe $g \in \mathcal{L}_1(\mu)$ tal que $\left| \frac{\partial f}{\partial t} \Big|_{(x,t)} \right| \leq g(x)$ para todo x y para todo t .

Entonces $F(x) = \int f^t d\mu$ existe para toda t y

$$F'(t_0) = \int \frac{\partial f}{\partial t} \Big|_{(x,t_0)} d\mu.$$

En este teorema tomamos a X como \mathbb{R} , a S como los Borelianos de \mathbb{R} y μ como la medida de Lebesgue (que en nuestro caso coincide con la integral de Riemann). Con este teorema podemos “meter” la derivada dentro de la integral, lo cual no es trivial pues ambos son procesos límites.

Usamos a $\hat{f}(x, k) = f(x) \cos(kx)$ o en otro caso a $\hat{f}(x, k) = f(x) \sin(kx)$. Para la g basta con tener a la misma f . Esto se debe a que \hat{f} está “modulada” por el seno o el coseno y dichas funciones toman valores entre 1 y -1 . Tenemos que, como $f \in \mathcal{PC}[0, 2\pi]$ entonces $|f| = g \in \mathcal{L}_1(\lambda)$, donde λ es la medida de Lebesgue. Con esto obtenemos que podemos calcular los coeficientes a partir de su derivada, es decir:

$$\begin{aligned} a'_k &= \frac{\partial}{\partial k} \left(\frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx \right) = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial k} f(x) \cos(kx) dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} -k f(x) \sin(kx) dx = -k b_k \Rightarrow b_k = \frac{-a'_k}{k} \\ b'_k &= \frac{\partial}{\partial k} \left(\frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx \right) = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial k} f(x) \sin(kx) dx \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} k f(x) \cos(kx) dx = k a_k \Rightarrow a_k = \frac{b'_k}{k}. \end{aligned}$$

Con este resultado ya podemos proseguir con la demostración del teorema 3.

Demostración. Vamos a demostrar que S_n es una sucesión de Cauchy uniforme, lo cual sería equivalente a demostrar que S_n converge en $\mathcal{PC}[0, 2\pi]$. Sea $x \in \mathbb{R}$ cualquiera, $n, p \in \mathbb{N}$ con $p > 1$ cualquier natural, entonces:

$$\begin{aligned}
|S_{n+p}(f)(x) - S_n(f)(x)| &= \left| \sum_{k=n+1}^{n+p} (a_k \cos(kx) + b_k \sin(kx)) \right| \\
&\leq \sum_{k=n+1}^{n+p} (|a_k \cos(kx)| + |b_k \sin(kx)|) \leq \sum_{k=n+1}^{n+p} (|a_k| + |b_k|) \\
&= \sum_{k=n+1}^{n+p} \frac{1}{k} (|a'_k| + |b'_k|).
\end{aligned}$$

Pero para a_k y para b_k tenemos las siguientes desigualdades (las cuales también nos recuerdan a la famosa desigualdad de Hölder):

$$\begin{aligned}
\sum_{k=n+1}^{n+p} \frac{1}{k} |a'_k| &= \left(\frac{1}{n+1}, \frac{1}{n+2}, \dots, \frac{1}{n+p} \right) \cdot (|a'_{n+1}|, |a'_{n+2}|, \dots, |a'_{n+p}|) \\
&\leq \left(\sum_{k=n+1}^{n+p} \frac{1}{k^2} \right)^{\frac{1}{2}} \left(\sum_{k=n+1}^{n+p} a'^2_k \right)^{\frac{1}{2}},
\end{aligned}$$

$$\begin{aligned}
\sum_{k=n+1}^{n+p} \frac{1}{k} |b'_k| &= \left(\frac{1}{n+1}, \frac{1}{n+2}, \dots, \frac{1}{n+p} \right) \cdot (|b'_{n+1}|, |b'_{n+2}|, \dots, |b'_{n+p}|) \\
&\leq \left(\sum_{k=n+1}^{n+p} \frac{1}{k^2} \right)^{\frac{1}{2}} \left(\sum_{k=n+1}^{n+p} b'^2_k \right)^{\frac{1}{2}}.
\end{aligned}$$

Notemos que por el teorema de Dirichlet de convergencia puntual (el teorema 2) sabemos que el límite puntual de S_n es precisamente f , entonces tanto $\sum_{k=n+1}^{n+p} a'^2_k$ como $\sum_{k=n+1}^{n+p} b'^2_k$ son las colas de una serie convergente (entonces tienden a cero cuando n tiende a infinito). Recapitulando todo lo que hemos obtenido hasta el momento tenemos que:

$$\begin{aligned}
|S_{n+p}(f)(x) - S_n(f)(x)| &\leq \sum_{k=n+1}^{n+p} \frac{1}{k} (|a'_k| + |b'_k|) \\
&\leq \left(\sum_{k=n+1}^{n+p} \frac{1}{k^2} \right)^{\frac{1}{2}} \left(\sum_{k=n+1}^{n+p} a_k'^2 + b_k'^2 \right)^{\frac{1}{2}} \\
&\leq \left(\sum_{k=n+1}^{\infty} \frac{1}{k^2} \right)^{\frac{1}{2}} \left(\sum_{k=n+1}^{\infty} a_k'^2 + b_k'^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

Lo cual tiende a cero cuando n tiende a infinito, independientemente del valor que tomen p y x . Entonces S_n es una sucesión de Cauchy uniforme, por ende tiende a su límite puntual uniformemente. Es decir, $\lim_{n \rightarrow \infty} S_n(f)(x) = f(x)$ uniformemente y $|\frac{a_0}{2}| + \sum_{k=1}^{\infty} |a_k| + |b_k| < \infty$. ■

El estimado lector se podrá preguntar qué pasa si f es una función compleja (no en el sentido de difícil, sino que toma valores no sólo reales, sino también complejos). Para tales casos nos va a ser de suma utilidad la identidad de Euler.

Proposición. *La fórmula de Euler establece la relación entre la función exponencial y las funciones seno y coseno de la siguiente manera:*

$$e^{i\theta} = \text{sen}(\theta) + i\cos(\theta).$$

Donde $x \in \mathbb{R}$ e i es la unidad imaginaria.

Un resultado inmediato de esta fórmula es su relación con los senos y cosenos:

$$\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2}, \quad \text{sen}(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

Y recordando lo que hemos hecho, si $\theta = kx$ entonces tenemos que:

$$\cos(kx) = \frac{e^{ikx} + e^{-ikx}}{2}, \quad \text{sen}(kx) = \frac{e^{ikx} - e^{-ikx}}{2i}$$

Entonces ya tenemos otra forma de escribir a nuestra serie de Fourier:

$$\begin{aligned}
f(x) &\sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)) \\
&= \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(\frac{a_k - ib_k}{2} e^{ikx} + \frac{a_k + ib_k}{2} e^{-ikx} \right)
\end{aligned}$$

Notemos que agrupamos a los coeficientes dependiendo si estaban asociados con e^{ikx} o con e^{-ikx} y lo único que cambia entre estas dos funciones es el signo del exponente, el cual se lo podemos atribuir a x o a k . Si se lo atribuimos a k entonces podríamos escribir de distinta manera esta suma y definir otros coeficientes a partir de los a_k y b_k :

$$\begin{aligned}
c_0 &= \frac{a_0}{2} \\
c_k &= \frac{a_k - ib_k}{2} \\
c_{-k} &= \frac{a_k + ib_k}{2}.
\end{aligned}$$

Con lo cual podemos escribir a las series de Fourier como:

$$f(x) \sim c_0 + \sum_{k=-\infty}^{\infty} (c_k e^{ikx} + c_{-k} e^{-ikx}) = \sum_{k=-\infty}^{\infty} c_k e^{-ikx}.$$

A esta última forma de escribir a las series de Fourier se le conoce como la forma compleja de series de Fourier y a los coeficientes c_k se les conoce como los coeficientes complejos de Fourier, los cuales se pueden escribir de la siguiente manera:

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx.$$

Es natural preguntarse por la forma en la que una función compleja se puede integrar, pero la manera de calcular dichas integrales es bastante sencilla, pues si f es una función compleja, entonces se puede escribir separando su parte real de su parte imaginaria $f(x) = u(x) + iv(x)$. Y así integramos:

$$\int_a^b f(x) dx = \int_a^b u(x) dx + i \int_a^b v(x) dx.$$

Con esto ya podríamos calcular los coeficientes complejos de la serie de Fourier de una función compleja. Si f es una función real entonces $f(x) = u(x) + iv(x) = u(x)$, pero nos gustaría que su serie de Fourier sea la misma con la que hemos trabajado. Notemos que si f es real, entonces:

$$\begin{aligned} c_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) (\cos(kx) - i \operatorname{sen}(kx)) dx \\ &= \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} f(x) \cos(kx) dx - i \int_{-\pi}^{\pi} f(x) \operatorname{sen}(kx) dx \right) \\ &= \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} u(x) \cos(kx) dx - i \int_{-\pi}^{\pi} u(x) \operatorname{sen}(kx) dx \right) \\ &= \frac{a_k - ib_k}{2}, \end{aligned}$$

$$\begin{aligned} c_{-k} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{ikx} dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) (\cos(kx) + i \operatorname{sen}(kx)) dx \\ &= \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} f(x) \cos(kx) dx + i \int_{-\pi}^{\pi} f(x) \operatorname{sen}(kx) dx \right) \\ &= \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} u(x) \cos(kx) dx + i \int_{-\pi}^{\pi} u(x) \operatorname{sen}(kx) dx \right) \\ &= \frac{a_k + ib_k}{2}. \end{aligned}$$

Y así podemos llegar a la serie de Fourier que ya conocíamos:

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \operatorname{sen}(kx)).$$

Notemos que esto depende de que c_k y c_{-k} sean conjugados y eso no necesariamente pasa si f toma valores complejos.

1.3. Teorema integral de Fourier

Ahora pasemos a funciones reales de dos variables. Siguiendo en la misma línea, tenemos definiciones y propiedades análogas a aquellas que teníamos con funciones de una sola variable real.

También tenemos que definir el producto interno dentro de este sistema bivariado, el cual no debe sorprendernos dada su similitud con el producto interno univariado definido previamente.

Definición 10. *El producto interno en un sistema de funciones continuas bivariadas que vamos a usar es el siguiente:*

$$\langle \varphi_k, \varphi_j \rangle = \int \int \varphi_k(x, t) \varphi_j(x, t) dx dt.$$

Un concepto bastante importante es el de sistemas ortogonales el cual es análogo a su definición en una variable, lo cual nos facilita bastantes cuentas, forman bases bonitas, son “fáciles” de visualizar, etc.

Definición 11. *Sea $\varphi_k : \mathbb{R}^2 \rightarrow \mathbb{R}$ un sistema de funciones continuas, ninguna con raíces iguales (no se anulan al mismo tiempo). El sistema es ortogonal si:*

$$\int \int \varphi_k(x, t) \varphi_j(x, t) dx dt = 0 \quad \text{si } k \neq j$$

De igual manera tenemos una norma 2 en de este sistema.

Definición 12. *La norma de algún miembro de nuestro sistema de funciones, $\varphi_k(x, t)$ es:*

$$\|\varphi_k\|_2 = \sqrt{\int_{\mathbb{R}} \int_{\mathbb{R}} \varphi_k(x, t)^2 dx dt}.$$

Así como sucede con las funciones de una variable, ahora con funciones de dos variables también podemos asociarles su respectiva serie de Fourier. Esto nos sirve para poder aproximar cualquier función cuadrado integrable (\mathcal{L}_2) que sea combinación lineal de nuestro sistema

de funciones φ_n y como sucede en funciones de una sola variable esta aproximación da también el menor error cuadrático medio (norma 2).

Corolario. *Para cualquier función cuadrado integrable, es decir que está en \mathcal{L}_2 , podemos asociar una serie de Fourier:*

$$f(x, t) \sim c_0 \varphi_0(x, t) + c_1 \varphi_1(x, t) + \dots + c_k \varphi_k(x, t) + \dots ,$$

en donde los coeficientes están dados por:

$$c_k = \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \varphi_k(x, t) dx dt}{\int_{\mathbb{R}} \int_{\mathbb{R}} \varphi_k(x, t)^2 dx dt}.$$

Lo que queremos ahora es tener un sistema de funciones trigonométricas, como en el caso univariado, para poder computar estas diferentes series de Fourier. Entonces definimos nuestro sistema trigonométrico bivariado.

Definición 13. *El sistema trigonométrico básico en 2 variables está dado por:*

$$\begin{aligned} T = \{ & 1, \cos(jx), \sin(jx), \cos(kx), \sin(kx), \dots \\ & \cos(jx) \cos(kt), \sin(jx) \cos(kt), \dots \\ & \cos(jx) \sin(kt), \sin(jx) \sin(kt) \}_{k,j=1}^{\infty} \end{aligned}$$

Como es de suponer, cada elemento de T es de periodo 2π en la variable x y en la variable t . Como en el caso del sistema trigonométrico básico univariado, en el caso bivariado T también es un sistema de funciones ortogonales (mas no ortonormales).

Ahora, al tener 2 variables tenemos 9 tipos diferentes de coeficientes para nuestras series de Fourier.

Proposición. *Los coeficientes de la serie de Fourier relacionada con una función bivariada f son:*

$$A_{00} = \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) dx dt,$$

$$\begin{aligned}
A_{j0} &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \cos(jx) dx dt, \\
A_{0k} &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \cos(kt) dx dt, \\
B_{j0} &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \sin(jx) dx dt, \\
B_{0k} &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \sin(kt) dx dt, \\
a_{jk} &= \frac{1}{(\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \cos(jx) \cos(kt) dx dt, \\
b_{jk} &= \frac{1}{(\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \sin(jx) \cos(kt) dx dt, \\
c_{jk} &= \frac{1}{(\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \cos(jx) \sin(kt) dx dt, \\
d_{jk} &= \frac{1}{(\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \sin(jx) \sin(kt) dx dt.
\end{aligned}$$

Claramente no queremos escribir explícitamente estos coeficientes cada vez que queramos trabajar con la serie de Fourier de dos variables. Para facilitar su escritura volveremos a usar la formula de Euler que ya mencionamos en 1.2.

¿De qué nos puede servir esto? La identidad de Euler simplifica de cierta manera la suma de cosenos y senos, lo que tenemos que hacer es pensar en nuestras parejas de números (x, t) como si fueran números complejos. Es decir, pensar que son de la forma $z = x + it$. Dado que \mathbb{R}^2 y \mathbb{C} son isomorfos, claro que podemos pensar de esta manera. Dado esto, podemos escribir la serie de Fourier de dos variables de manera más compacta:

$$f(x, t) \sim \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} c_{jk} e^{i(jx+kt)}.$$

Donde $c_{jk} = \frac{1}{4\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) e^{-i(jx+kt)} dx dt$. Recordándonos esta manera de escribir la serie como la forma compleja de la serie de

Fourier para el caso de una sola variable.

¿Qué pasa si nuestra función f tiene diferentes periodos en x y en y ? La respuesta es sencilla: hacemos un cambio de variable. Supongamos que $f(x, t)$ está definida en un rectángulo $R(-l \leq x \leq l, -h \leq y \leq h)$ y que es de periodo $2l$ en x mientras que es de periodo $2h$ en t . El cambio que proponemos es el siguiente:

$$u = \frac{\pi x}{l}, \quad v = \frac{\pi t}{h}.$$

Y así nuestra nueva función $\gamma(u, v)$ es una función de periodo 2π en u y en v con:

$$f\left(\frac{lu}{\pi}, \frac{hv}{\pi}\right) = \gamma(u, v).$$

Entonces computamos la serie de Fourier de γ para volver a las unidades originales x y t .

Consideremos $f(x)$, una función definida en todo \mathbb{R} , continua y suave a trozos. Esto significa que puede tener una cantidad numerable de discontinuidades, pero donde es continua es suave. En palabras de teoría de la medida decimos que es continua casi donde sea, relativo a la medida de Lebesgue. Entonces, por todo lo que hemos estado argumentando en la sección de convergencias, en cada intervalo finito $[-l, l]$, f puede ser expandida como su serie de Fourier. Es decir:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{\pi k x}{l}\right) + b_k \sen\left(\frac{\pi k x}{l}\right) \right). \quad (1.10)$$

Ahora supongamos que f es absolutamente integrable (está en \mathcal{L}_1), supongamos también que hacemos que l tienda a infinito (es decir, hacemos el intervalo cada vez más grande hasta abarcar todos los reales). Entonces, recordando cómo son los coeficientes de Fourier para funciones de una variable, tenemos:

$$a_0 = \frac{1}{l} \int_{-l}^l |f(x)| dx \implies \lim_{l \rightarrow \infty} a_0 = 0.$$

Recordemos la siguiente identidad trigonométrica:

$$\cos(\alpha \pm \beta) = \cos(\alpha) \cos(\beta) \mp \sin(\alpha) \sin(\beta).$$

Entonces desarrollando 1.10 y sustituyendo por la forma explícita de los coeficiente de Fourier tenemos lo siguiente:

$$\begin{aligned} f(x) &= \lim_{l \rightarrow \infty} \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{\pi kx}{l}\right) + b_k \sin\left(\frac{\pi kx}{l}\right) \right) \\ &= \lim_{l \rightarrow \infty} \sum_{k=1}^{\infty} \left(\left[\frac{1}{l} \int_{-l}^l f(u) \cos\left(\frac{\pi ku}{l}\right) du \right] \cos\left(\frac{\pi kx}{l}\right) \right. \\ &\quad \left. + \left[\frac{1}{l} \int_{-l}^l f(u) \sin\left(\frac{\pi ku}{l}\right) du \right] \sin\left(\frac{\pi kx}{l}\right) \right) \\ &= \lim_{l \rightarrow \infty} \sum_{k=1}^{\infty} \left(\frac{1}{l} \int_{-l}^l f(u) \cos\left(\frac{\pi ku}{l}\right) \cos\left(\frac{\pi kx}{l}\right) du \right. \\ &\quad \left. + \frac{1}{l} \int_{-l}^l f(u) \sin\left(\frac{\pi ku}{l}\right) \sin\left(\frac{\pi kx}{l}\right) du \right) \\ &= \lim_{l \rightarrow \infty} \sum_{k=1}^{\infty} \left(\frac{1}{l} \int_{-l}^l f(u) \left[\cos\left(\frac{\pi ku}{l}\right) \cos\left(\frac{\pi kx}{l}\right) \right. \right. \\ &\quad \left. \left. + \sin\left(\frac{\pi ku}{l}\right) \sin\left(\frac{\pi kx}{l}\right) \right] du \right) \\ &= \lim_{l \rightarrow \infty} \sum_{k=1}^{\infty} \left(\frac{1}{l} \int_{-l}^l f(u) \cos\left(\frac{\pi k}{l}(u-x)\right) du \right). \end{aligned}$$

Si además hacemos el siguiente cambio de variable: $\lambda_1 = \frac{\pi}{l}, \lambda_2 = \frac{\pi^2}{l}, \dots, \lambda_k = \frac{\pi k}{l}, \dots$ y con los cambios en las lambdas definido como $\Delta\lambda_k = \frac{1}{l}$, entonces:

$$f(x) = \lim_{l \rightarrow \infty} \frac{1}{\pi} \sum_{k=1}^{\infty} \left(\Delta\lambda_k \int_{-l}^l f(u) \cos(\lambda_k(u-x)) du \right).$$

Esto nos recuerda mucho a la definición de integral, ¿no? Una suma con una tasa de cambio que cada vez se va haciendo más pequeña.

Efectivamente, a $\int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) du$ se le conoce como la integral de Fourier, mientras que a toda la siguiente expresión se le conoce como el teorema integral de Fourier. Este teorema es el que nos permite construir la famosa transformada de Fourier así como su transformada inversa.

Teorema 5 (Teorema Integral de Fourier). *Sea $f(x)$ una función continua (o continua a trozos) y absolutamente integrable en toda la recta real \mathbb{R} entonces:*

$$f(x) = \frac{1}{\pi} \int_0^{\infty} \int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) du d\lambda$$

Esta no es la demostración formal para el teorema integral de Fourier, pues no podemos intercambiar procesos infinitos (suma e integral) de manera tan fácil. Para poder lograrla y explicarla mejor van a ser necesarios unos teoremas y lemas más.

Así como existen sucesiones de números o de funciones uniformemente convergentes, también tenemos una definición análoga para integrales que dependan de un parámetro. Es decir, integrales que se ven de la siguiente manera:

$$\int_a^{\infty} F(x, \lambda) dx,$$

donde a es cualquier número real arbitrario, F es una función de dos variables y λ es un parámetro (usualmente tendremos un intervalo en donde dicho parámetro pueda vivir). De manera similar a las t-secciones vistas en el corolario del TCDL en 4. De esto tenemos la siguiente definición.

Definición 14. *Sea $F(x, \lambda) : \mathbb{R}^2 \rightarrow \mathbb{R}$ una función tal que $\int_a^{\infty} F(x, \lambda) dx$ es convergente para cualquier valor de λ para algún intervalo $\alpha \leq \lambda \leq \beta$. Decimos que esta integral es absolutamente convergente para este intervalo $\alpha \leq \lambda \leq \beta$ si $\forall \varepsilon > 0, \exists L \in \mathbb{R}$ tal que se cumpla:*

$$\left| \int_L^{\infty} F(x, \lambda) dx \right| \leq \varepsilon, \quad (1.11)$$

$\forall l \geq L$ y $\forall \lambda$ en el intervalo previamente descrito.

Notemos que esta definición es análoga a las funciones en \mathcal{L}_1 , usando un parámetro y cambiando ciertos aspectos. La función no necesariamente tiene que estar definida para valores de λ fuera del intervalo dado. Un criterio para determinar si una integral es absolutamente convergente se basa en sucesiones de números, siendo un criterio análogo al de series de funciones absolutamente convergentes.

Proposición. *Una condición necesaria y suficiente para que la integral definida en la definición 14 sea uniformemente convergente es que para toda sucesión de números reales*

$$x_0 = a < x_1 < x_2 < \dots < x_k < \dots,$$

$$\lim_{k \rightarrow \infty} x_k = \infty,$$

es que los términos de la serie:

$$\begin{aligned} \int_a^\infty F(x, \lambda) dx &= \int_{x_0}^{x_1} F(x, \lambda) dx + \int_{x_1}^{x_2} F(x, \lambda) dx + \dots \\ &+ \int_{x_{k-1}}^{x_k} F(x, \lambda) dx + \dots = \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} F(x, \lambda) dx \end{aligned} \quad (1.12)$$

sean uniformemente convergentes para $\alpha \leq \lambda \leq \beta$. Es decir, de la manera que sea que tomemos trozos de la recta real e integremos en ellos, estos trozos de integral sean uniformemente convergentes.

Demostración. Iniciamos suponiendo que $\int_a^\infty F(x, \lambda) dx$ es uniformemente convergente, es decir se cumple lo establecido en la definición 14. Entonces para todo $\varepsilon > 0$ existe $L \in \mathbb{R}$ tal que si $l \geq L$ tenemos que:

$$\left| \int_l^\infty F(x, \lambda) dx \right| \leq \varepsilon.$$

Ahora, si tenemos una sucesión creciente de números reales que “converge a infinito”, $(x_k)_{k=0}^\infty$ entonces podemos encontrar alguna n tal que $x_n \geq l$

y por ende:

$$\begin{aligned}
 & \left| \int_l^\infty F(x, \lambda) dx - \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} F(x, \lambda) dx \right| \\
 &= \left| \int_l^\infty F(x, \lambda) dx - \int_l^{x_n} F(x, \lambda) dx \right| \\
 &= \left| \int_{x_n}^\infty F(x, \lambda) dx \right| \leq \varepsilon.
 \end{aligned}$$

Pero recordemos que, por hipótesis:

$$\left| \int_{x_n}^\infty F(x, \lambda) dx \right| = \left| \sum_{k=n}^\infty \int_{x_k}^{x_{k+1}} F(x, \lambda) dx \right| \Rightarrow \left| \sum_{k=n}^\infty \int_{x_k}^{x_{k+1}} F(x, \lambda) dx \right| \leq \varepsilon.$$

Entonces los últimos términos de la serie convergen a cero, pero dicha convergencia no dependió nunca del valor de x , entonces son uniformemente convergentes.

Supongamos ahora que, al revés, los términos de la serie son uniformemente convergentes. Supongamos que 1.11 no se cumple, esto quiere decir que $\exists \varepsilon > 0$, tal que $\forall l \in \mathbb{R}$ y tal que $\left| \int_l^\infty F(x, \lambda) dx \right| > \varepsilon$.

Como la serie es uniformemente convergente, esto quiere decir que $\forall \varepsilon > 0 \exists N \in \mathbb{N}$ tal que si $n \geq N$, entonces pasa lo siguiente (volvemos a usar el criterio de Cauchy):

$$\begin{aligned}
 & \left| \sum_{k=0}^\infty \int_{x_k}^{x_{k+1}} F(x, \lambda) dx - \sum_{k=0}^N \int_{x_k}^{x_{k+1}} F(x, \lambda) dx \right| \leq \varepsilon \\
 & \Rightarrow \left| \sum_{k=N+1}^\infty \int_{x_k}^{x_{k+1}} F(x, \lambda) dx \right| \leq \varepsilon \Rightarrow \left| \int_{x_{N+1}}^\infty F(x, \lambda) dx \right| \leq \varepsilon \perp
 \end{aligned}$$

Esto es una contradicción, entonces forzosamente 1.11 se cumple. Es decir, esta integral es absolutamente convergente para este intervalo $\alpha \leq \lambda \leq \beta$. ■

Este teorema nos va a ayudar para demostrar otro, el cual nos da información de cómo es la integral definida en la definición 14.

Teorema 6. Si $F(x, \lambda)$ es continua (considerándola como una función de dos variables) o si tiene un número finito de discontinuidades en x , (pero es integrable con respecto a x y continua en λ) y si la integral mencionada en la definición 14 es uniformemente convergente para dicho intervalo de λ ($\alpha \leq \lambda \leq \beta$) entonces esta integral es una función continua de λ .

Demostración. Tomemos una sucesión de números reales $(x_k)_{k \in \mathbb{N}}$ con $x_0 = a$ y de tal manera que todas las discontinuidades de $F(x, \lambda)$ estén en valores de esta sucesión. Ahora recordemos que:

$$\int_a^\infty F(x, \lambda) dx = \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} F(x, \lambda) dx.$$

Pero cada uno de las integrales de la suma es una función continua (pues en estos pedazos $F(x, \lambda)$ es continua). Entonces la integral inicial es una función continua. ■

Con el hecho de que esta integral es una función continua de λ y que existe convergencia uniforme nos da las herramientas necesarias para poder demostrar fácilmente el siguiente teorema.

Teorema 7. Bajo las hipótesis anteriores sobre $F(x, \lambda)$ y si $\alpha \leq \lambda \leq \beta$ entonces se cumple lo siguiente:

$$\int_a^\beta \int_a^\infty F(x, \lambda) dx d\lambda = \int_a^\infty \int_a^\beta F(x, \lambda) d\lambda dx,$$

es decir podemos invertir el orden de integración.

Demostración. Dada la convergencia uniforme, 1.12 se puede integrar término a término y el orden de integración se puede cambiar (por ser integrales finitas una por una), es decir,

$$\int_a^\infty F(x, \lambda) dx = \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} F(x, \lambda) dx$$

se puede integrar término a término y se puede cambiar su orden. Entonces tenemos:

$$\begin{aligned}
\int_a^\beta \int_a^\infty F(x, \lambda) dx d\lambda &= \int_a^\beta \left(\sum_{k=0}^\infty \int_{x_k}^{x_{k+1}} F(x, \lambda) dx \right) d\lambda \\
&= \int_a^\beta \int_{x_0}^{x_1} F(x, \lambda) dx d\lambda + \int_a^\beta \int_{x_1}^{x_2} F(x, \lambda) dx d\lambda + \dots \\
&= \int_{x_0}^{x_1} \int_a^\beta F(x, \lambda) d\lambda dx + \int_{x_1}^{x_2} \int_a^\beta F(x, \lambda) d\lambda dx + \dots \\
&= \sum_{k=0}^\infty \int_{x_k}^{x_{k+1}} \int_a^\beta F(x, \lambda) d\lambda dx \\
&= \int_a^\infty \int_a^\beta F(x, \lambda) d\lambda dx.
\end{aligned}$$

■

Los siguientes teoremas y lemas no serán demostrados por brevedad, pero servirán para probar el teorema integral de Fourier.

Teorema 8. Sea $F(x, \lambda)$ una función continua con derivada parcial $\frac{\partial F(x, \lambda)}{\partial \lambda}$ continua, que existan las siguientes integrales:

$$\int_a^\infty F(x, \lambda) dx, \quad \int_a^\infty \frac{\partial F(x, \lambda)}{\partial \lambda} dx,$$

y que la segunda sea uniformemente convergente para algún intervalo de λ ($\alpha \leq \lambda \leq \beta$). Entonces:

$$\frac{\partial}{\partial \lambda} \int_a^\infty F(x, \lambda) dx = \int_a^\infty \frac{\partial F(x, \lambda)}{\partial \lambda} dx.$$

Un bosquejo de la prueba sería usando el corolario del TCDL mencionado en 4, el cual nos permite “meter” derivadas a integrales bajo ciertas condiciones (aunque ambos sean procesos límites).

Teorema 9. Si para $\alpha \leq \lambda \leq \beta$, $|F(x, \lambda)| \leq f(x)$ para alguna función continua $f(x)$ y $\int_a^\infty |f(x)| dx$ existe, entonces la integral definida previamente, $\int_a^\infty F(x, \lambda) dx$ es uniformemente convergente.

Si el estimado lector ha leído el TCDL y algunos de sus corolarios nos podemos dar cuenta que el teorema pasado es justamente uno de ellos aplicado a convergencia uniforme, a funciones de variable real, usando la medida de la identidad y a la sigma álgebra de los Borelianos.

El siguiente lema nos dice a dónde va a converger la integral de una función $f(x)$ con seno si ésta tiene límites laterales. Es decir, esta función puede tener una discontinuidad en algún punto, pero los límites laterales deben de existir para que esto se cumpla.

Lema. *Si la función $f(x)$ es absolutamente integrable para todo x en \mathbb{R} y si $f(x)$ tiene límites laterales en un punto x , entonces:*

$$\lim_{l \rightarrow \infty} \frac{1}{\pi} \int_{-\infty}^{\infty} f(x+u) \frac{\sin(lu)}{u} du = \frac{f(x)_+ + f(x)_-}{2},$$

donde:

$$\begin{aligned} f(x)_+ &= \lim_{h \rightarrow 0} f(x+h) \\ f(x)_- &= \lim_{h \rightarrow 0} f(x-h) \end{aligned}$$

Es decir, que esta integral converge al punto medio de la discontinuidad. Notemos que si $f(x)$ es continua en x entonces esta integral simplemente converge al valor puntual de f en x .

Procedemos a la prueba formal del teorema integral de Fourier (Teorema 5):

Demostración. Sea $f(x)$ absolutamente integrable para todo $x \in \mathbb{R}$, entonces por definición:

$$\begin{aligned} & \frac{1}{\pi} \int_0^{\infty} \int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) du d\lambda \\ &= \lim_{t \rightarrow \infty} \frac{1}{\pi} \int_0^t \int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) du d\lambda. \end{aligned} \tag{1.13}$$

Pero tenemos que:

$$|f(u) \cos(\lambda(x-u))| \leq |f(u)|,$$

porque $|\cos(\theta)| \leq 1 \quad \forall \theta \in \mathbb{R}$. Entonces por el teorema 9:

$$\int_{-\infty}^{\infty} f(u) \cos(\lambda(x-u)) du$$

es uniformemente convergente. Por ende, por el teorema 7 podemos invertir el orden de integración:

$$\int_0^t \int_{-\infty}^{\infty} f(u) \cos(\lambda(x-u)) d\lambda du = \int_{-\infty}^{\infty} \int_0^t f(u) \cos(\lambda(x-u)) d\lambda du.$$

Procedemos a hacer un cambio de variable, con $u-x=:u$, por lo que lo anterior desarrollándolo se convierte en:

$$\begin{aligned} \int_{-\infty}^{\infty} \int_0^t f(u) \cos(\lambda(x-u)) d\lambda du &= \int_{-\infty}^{\infty} f(u) \int_0^t \cos(\lambda(x-u)) d\lambda du \\ &= \int_{-\infty}^{\infty} f(u) \frac{\text{sen}(t(u-x))}{u-x} du \\ &= \int_{-\infty}^{\infty} f(x+u) \frac{\text{sen}(tu)}{u} du \end{aligned}$$

Pero por la definición que establecimos al inicio del límite en 1.13,

$$\frac{1}{\pi} \int_0^{\infty} \int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) d\lambda du = \frac{1}{\pi} \lim_{t \rightarrow \infty} \int_{-\infty}^{\infty} f(x+u) \frac{\text{sen}(tu)}{u} du.$$

Ahora, si $f(x)$ tiene límites laterales en x , por el lema 1.3 tenemos lo siguiente:

$$\frac{1}{\pi} \int_0^{\infty} \int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) d\lambda du = \frac{f(x)_+ + f(x)_-}{2},$$

pero si $f(x)$ es continua en x , como habíamos comentado:

$$\frac{1}{\pi} \int_0^{\infty} \int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) d\lambda du = f(x).$$

Y por ende queda demostrado nuestro teorema. ■

Notemos que la siguiente integral (la “integral de adentro”):

$$\int_{-\infty}^{\infty} f(u) \cos(\lambda(u-x)) du$$

es una función par de λ . Este teorema puede tomar diferentes formas. Como sabemos, la función coseno es par, mientras que la función seno es impar. Entonces, ¿habrá una forma análoga de representar la función con funciones impares usando este teorema? Si pensamos que el seno es una función trigonométrica impar, entonces de seguro sí se puede. Considerando la siguiente integral:

$$\int_{-\infty}^{\infty} f(u) \sin(\lambda(u-x)) du$$

notemos que se parece mucho a la integral anterior, solo que esta vez se trata de una función, también de λ , pero que es impar. También, dado que tanto coseno como seno son funciones acotadas por uno y menos uno,

$$|f(u) \sin(\lambda(x-u))| \leq |f(u)|.$$

Podemos aplicar también el teorema 9, el cual nos asegura que esta integral es uniformemente convergente para cierto intervalo de λ . Entonces podríamos repetir la demostración del teorema integral de Fourier (5), pero para el caso de seno.

El estimado lector se puede volver a preguntar sobre el dominio de f , como nos hemos preguntado anteriormente, ¿qué pasa si f es una función evaluada en los complejos? Este teorema en versión compleja, uniendo todo lo mencionado anteriormente, recordando la identidad de Euler puede ser escrito como:

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u) [\cos(\lambda(u-x)) + \sin(\lambda(u-x))] dud\lambda \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u) e^{i\lambda(u-x)} dud\lambda. \end{aligned} \tag{1.14}$$

De esta manera, 1.14 se le conoce como la forma compleja del teorema integral de Fourier.

¿Qué podemos concluir de esto? ¿Para qué nos sirve saber el teorema integral de Fourier? Recordemos que la integral es “una suma infinita de partes pequeñas”. Notemos también que f aparece de ambos lados de la ecuación. Entonces este teorema nos dice que, de cierta manera y bajo ciertas hipótesis, f se puede descomponer por algún factor que involucre a las funciones seno y coseno.

A este teorema también se le conoce como teorema de inversión de Fourier. Esto se debe a que tenemos a f de ambos lados de la ecuación, por ende podríamos creer que podemos rescatar f de cierta transformación. Esto da lugar a la conocida transformación de Fourier.

1.4. Transformada de Fourier

Esta transformación tiene varias aplicaciones, por lo cual es bastante conocida. La transformación está definida como:

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi xu} dx. \quad (1.15)$$

Cabe mencionar que F no es la misma función que f , pues F es una nueva función que queda definida basándonos en f . Bajo las hipótesis del teorema integral de Fourier, podemos obtener fácilmente la inversa de dicha transformación:

$$f(x) = \int_{-\infty}^{\infty} F(u)e^{i2\pi xu} du. \quad (1.16)$$

Notemos la aparición del signo menos dentro del exponente. Esto se debe a la aplicación del teorema integral de Fourier. ¿A qué se refiere esta transformación? Podemos hacer una analogía a una receta. Si f es un platillo, nos gustaría tener la receta de éste para poder analizarlo mejor, ver si hay algún problema al respecto, poder ver qué ingredientes podemos tratar de omitir, etc. La receta es F , la integral es la descomposición de f en sus elementos base. Con la transformada

inversa simplemente volvemos a combinar todos los ingredientes de la receta (F) para poder obtener el platillo de regreso (f).

En lenguaje más matemático, recordemos que e^{-ixu} representa un círculo de radio unitario en el plano complejo. Así, al multiplicarlo con $f(x)$ cambiamos su radio al módulo de $f(x)$. Si cambiamos x (es lo que al fin de cuentas significa dx), entonces cambiamos continuamente el radio de nuestro círculo. Con esto podemos llegar a la conclusión que lo que estamos haciendo es midiendo el promedio de los radios del círculo a lo largo del tiempo.

Usualmente decimos que $f(x)$ está en el dominio de tiempo (es una señal) y $F(u)$ está en el dominio de las frecuencias, la transformada descompone a $f(x)$ en sus respectivas frecuencias. Podemos ver la transformada intuitivamente en el siguiente esquema:

Así como podemos descomponer a f en sus respectivas frecuencias, también podemos “ver” a los componentes f desde distintos puntos de vista (como si lo estuviéramos viendo a través de ventanas de diferentes tamaños). Entonces si el coeficiente $a_k = 0$ significa que la k -ésima frecuencia no es empleada en la señal, que ni la tomemos en consideración.

El operador $f(x)e^{ixu}$ se llama la modulación de f , mientras que la exponencial es la función de modulación, es la técnica que se usa para transportar información.

Notemos que la transformada de Fourier es lineal (por propiedades de la integral y de la función exponencial). Así mismo, se cumplen dos cualidades muy importantes: la identidad de Parseval y la ecuación de Plancherel. Es decir, si F es la transformada de Fourier de f y G es la de g :

$$\text{Identidad de Parseval : } \langle f, g \rangle = \langle F, G \rangle. \quad (1.17)$$

$$\text{Ecuación de Plancherel : } \|f\|^2 = \|F\|^2. \quad (1.18)$$

1.4.1. Transformada de Fourier en 2 dimensiones

El querido lector seguramente se estará preguntando qué pasa si en lugar de tener una $x \in \mathbb{R}$ o en \mathbb{C} tenemos una variable en dimensiones más altas. Este sería el caso para poder analizar imágenes, ¿no? En el capítulo 3 vamos a ver explícitamente qué hacer con imágenes, pero una respuesta corta sería que sí, una imagen se representa con variables de dimensiones mayores a 1.

Si recordamos la definición de la transformada de Fourier (y la transformada inversa) lo podemos ver con el producto interno que definimos en 1.2:

$$\begin{aligned} F(u) &= \langle f(x), e^{-i2\pi xu} \rangle \\ f(x) &= \langle F(u), e^{i2\pi xu} \rangle \end{aligned} \quad (1.19)$$

Si recordamos, también podemos tener productos internos de vectores que justamente son variables x en “dimensiones mayores”. Entonces la transformada de Fourier en estas dimensiones mayores se ve exactamente igual si la escribimos como producto interno:

$$\begin{aligned} F(u) &= \langle f(x), e^{-i2\pi xu} \rangle \\ f(x) &= \langle F(u), e^{i2\pi xu} \rangle \\ x, u &\in \mathbb{C}^n \end{aligned} \quad (1.20)$$

Como podemos ver, esta definición es un producto interno, por ende la podemos generalizar a n dimensiones. Esto se debe a que para funciones $g, h : \mathbb{C}^n \rightarrow \mathbb{C}$, el producto interno está definido como:

$$\begin{aligned} \langle h(x), g(x) \rangle &= \int_{\mathbb{C}^n} h(x) \overline{g(x)} dx \\ &= \int_{\mathbb{C}} \cdots \int_{\mathbb{C}} h(x_1, \dots, x_n) \overline{g(x_1, \dots, x_n)} dx_1 \cdots dx_n \end{aligned} \quad (1.21)$$

Con esto ahora tenemos que la transformada de Fourier y la inversa en \mathbb{C}^n es de la siguiente manera:

$$\begin{aligned}
F(u) &= \int_{\mathbb{C}^n} f(x) e^{-i2\pi x^T u} dx. \\
f(x) &= \int_{\mathbb{C}^n} F(u) e^{i2\pi x^T u} du. \\
u, x, &\in \mathbb{C}^n
\end{aligned}$$

Toda la teoría que ya compartimos y seguiremos desarrollando se sigue cumpliendo con este pequeño “ajuste” de dimensiones.

Pero recordemos que una computadora no puede calcular integrales como procesos límite (procesos infinitos) dada la limitación en memoria. Podemos aproximar dichos valores bastante bien de tal manera que llegamos a las transformadas discreta y rápida de Fourier.

1.5. Transformadas discreta y rápida de Fourier

Usando toda la información de la función que tenemos (sean f_0, f_1, \dots, f_{N-1} , con $f(x_k) = f_k$), la transformada discreta de Fourier es:

$$F(x_n) = \sum_{k=0}^{N-1} f_k e^{-ink \frac{2\pi}{N}} \quad (1.22)$$

Si pensamos desde punto de vista de teoría de la medida no tenemos la medida de Lebesgue, sino la medida de conteo y por eso podemos pasar de una integral a una suma tan naturalmente.

Análogamente a lo que hicimos para el caso continuo, si tenemos que las x_k que pertenecen a \mathbb{C}^n , entonces tendríamos n sumas (como teníamos n integrales en el caso pasado). Para poder facilitar las demostraciones y notación haremos el caso “unidimensional”, aunque su generalización a dimensiones mayores no es mucho más difícil a lo que haremos.

Aquí es donde podemos ver una de las muchas fusiones del álgebra lineal con el análisis. Notemos que $e^{ik \frac{2\pi}{N}}$ son las N raíces complejas de

la unidad. Estas raíces, como sabemos son cíclicas, es decir si $k \geq N$ entonces si $\omega = e^{i\frac{2\pi}{N}}$ tenemos que $\omega^k = \omega^j$ si $k = j \bmod(N)$, del mismo modo $e^{-ik\frac{2\pi}{N}}$ son también las N raíces de la unidad, pero en orden contrario a las manecillas del reloj.

Podemos notar que la transformada rápida de Fourier tiene aire de ser un producto matricial, de esta forma es intuitivo definir la matriz de Fourier.

Definición 15 (Matriz de Fourier). *La matriz de $N \times N$ cuya (k, n) -ésima entrada es $\xi^{kn} = \omega^{-kn}$ para $0 \leq k, n \leq N-1$ se llama la matriz de Fourier de orden N y tiene la siguiente forma:*

$$FR_N = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \xi & \xi^2 & \dots & \xi^{N-1} \\ 1 & \xi^2 & \xi^4 & \dots & \xi^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi^{N-1} & \xi^{2(N-2)} & \dots & \xi^{(N-1)^2} \end{bmatrix}$$

Bajo esta forma tenemos que el producto punto entre cualquier par de columnas es:

$$\sum_{k=0}^{N-1} \overline{\xi^{kr}} \xi^{ks} = \sum_{k=0}^{N-1} \xi^{-kr} \xi^{ks} = \sum_{k=0}^{N-1} \xi^{k(s-r)} = 0.$$

Es decir, la matriz de Fourier FR_N es una matriz ortogonal. Cada columna de FR_N es de norma \sqrt{N} (norma 2), pues:

$$\|FR_{*,k}\|_2^2 = \sum_{k=0}^{N-1} |\xi^{kn}|^2 = \sum_{k=0}^{N-1} 1 = N.$$

Con esto obtenemos que la matriz $\frac{1}{\sqrt{N}}FR_N$ es una matriz unitaria, por ende su inversa es su transpuesta conjugada, pero FR_N es una matriz simétrica entonces la inversa de $\frac{1}{\sqrt{N}}FR_N$ es simplemente su conjugada. De esto podemos deducir simplemente que:

$$FR_N^{-1} = \frac{1}{N} \overline{FR_N},$$

donde:

$$\overline{FR_N} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \bar{\xi} & \bar{\xi}^2 & \dots & \bar{\xi}^{N-1} \\ 1 & \bar{\xi}^2 & \bar{\xi}^4 & \dots & \bar{\xi}^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \bar{\xi}^{N-1} & \bar{\xi}^{2(N-2)} & \dots & \bar{\xi}^{(N-1)^2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-2)} & \dots & \omega^{(N-1)^2} \end{bmatrix}.$$

Y así podemos escribir la transformada discreta de Fourier del vector $f(x)$ como un bello producto matricial:

$$F(x) = FR_N f(x),$$

y cuya transformada inversa es:

$$f(x) = FR_N^{-1} F(x).$$

Como la inversa de la matriz FR es tan fácil de conjugar tenemos una gran ventaja al tratar de hacer cálculos computacionales, pues es más barato conjugar y multiplicar por un escalar que invertir toda una matriz.

Consideramos a la transformada rápida de Fourier como un método bonito, pero ¿qué tan eficiente es usarlo de esta manera? Consideramos que el producto matricial hace N^2 operaciones (multiplicación por renglón y por columna), entonces si la cantidad de información que tenemos es grande, es decir N es grande, entonces este método se vuelve poco eficiente. De seguro podemos aprovechar el hecho de que la matriz de Fourier es simétrica para hacer un método más eficiente. Si además consideramos que la computadora comete errores de redondeo (y errores de cancelaciones catastróficas) entonces no es muy eficiente hacer N^2 operaciones, donde cada una puede tener error de redondeo.

Lo lógico sería partir nuestra secuencia de datos $\{f_0, \dots, f_{N-1}\}$ en secuencias más cortas y calcular la transformada discreta de Fourier con estas secuencias más cortas para luego combinarlas de algún modo

y obtener la transformada de todos los datos. El siguiente teorema nos puede ayudar con lo que queremos hacer.

Teorema 10. *Sea $a \times b$ el producto entrada a entrada de los vectores a y b , es decir:*

$$a \times b = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{N-1} \end{bmatrix} = \begin{bmatrix} \alpha_0 \beta_0 \\ \alpha_1 \beta_1 \\ \vdots \\ \alpha_{N-1} \beta_{N-1} \end{bmatrix}.$$

Ahora definimos a los vectores \hat{a} y \hat{b} en \mathbb{C}^{2N} como siguen:

$$\hat{a} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \hat{b} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{N-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Si $FR = FR_{2N}$ es la matriz de Fourier de orden $2N$, entonces:

$$FR(a \odot b) = (FR\hat{a}) \times (FR\hat{b}),$$

y

$$a \odot b = FR^{-1} \left[(FR\hat{a}) \times (FR\hat{b}) \right],$$

donde $a \odot b \in \mathbb{C}^{2N}$ se define como:

$$a \odot b = \begin{bmatrix} \alpha_0 \beta_0 \\ \alpha_0 \beta_1 + \alpha_1 \beta_0 \\ \alpha_0 \beta_2 + \alpha_1 \beta_1 + \alpha_2 \beta_0 \\ \vdots \\ \alpha_{N-2} \beta_{N-1} + \alpha_{N-1} \beta_{N-2} \\ \alpha_{N-1} \beta_{N-1} \\ 0 \end{bmatrix}.$$

Demostración. Podemos notar que la t -ésima entrada en $FR_{*,j} \times FR_{*,k}$ está dada por:

$$[FR_{*,j} \times FR_{*,k}]_t = \xi^{tj} \xi^{tk} = \xi^{t(j+k)} = [FR_{*,j+k}]_t,$$

entonces las columnas de FR son de tal forma que se cumple que para todo j, k :

$$FR_{*,j} \times FR_{*,k} = FR_{*,j+k}.$$

Esto nos dice que $FR\hat{a}$, $FR\hat{b}$ y $FR(a \odot b)$ se pueden expresar como combinaciones lineales de las columnas de FR ,

$$FR\hat{a} = \sum_{j=0}^{N-1} \alpha_j FR_{*,j} \quad FR\hat{b} = \sum_{j=0}^{N-1} \beta_j FR_{*,j},$$

$$FR(a \odot b) = \sum_{j=0}^{2N-2} (a \odot b)_j FR_{*,j},$$

entonces el producto $(FR\hat{a}) \times (FR\hat{b})$ se puede expresar de la siguiente manera:

$$\begin{aligned} (FR\hat{a}) \times (FR\hat{b}) &= \left(\sum_{j=0}^{N-1} \alpha_j FR_{*,j} \right) \left(\sum_{j=0}^{N-1} \beta_j FR_{*,j} \right) \\ &= \sum_{j=0}^{2N-2} \left[\sum_{k=0}^j \alpha_k \beta_{j-k} \right] FR_{*,j} \\ &= \sum_{j=0}^{2N-2} [a \odot b]_j FR_{*,j} = FR(a \odot b). \end{aligned}$$

■

Usamos este teorema junto con la transformada discreta de Fourier para poder hacer un algoritmo más rápido llamado la transformada rápida de Fourier. Si $N = 2^r$, entonces podemos descomponer a la matriz de Fourier como:

$$FR_N = \begin{bmatrix} FR_{N/2} & D_{N/2}FR_{N/2} \\ FR_{N/2} & -D_{N/2}FR_{N/2} \end{bmatrix} P_N,$$

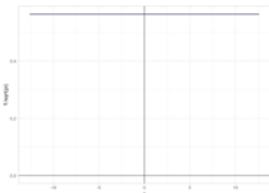
donde $D_{N/2}$ es la matriz diagonal con $\{1, \xi, \xi^2, \dots, \xi^{\frac{N}{2}-1}\}$ en sus entradas y P_N es la matriz par-impar de las permutaciones de la identidad, es decir:

$$P_N = [e_0 \ e_2 \ \dots \ e_{N-2} \mid e_1 \ e_3 \ \dots \ e_{N-1}].$$

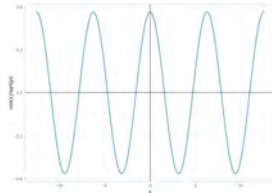
Y con esto tenemos lo que en un inicio queríamos, la transformada discreta de Fourier de orden $N = 2^r$ se puede lograr mediante dos transformaciones discretas de Fourier de orden $\frac{N}{2}$. También se puede separar la matriz de Fourier de diferentes formas, dependiendo del tamaño y de la divisibilidad de N , pero esta es la base para el famoso algoritmo de la transformada rápida de Fourier (FFT). Si el querido lector se encuentra interesado en esto lo puede leer más a fondo en el libro de Meyer en [5].

1.6. Problemas con la transformada de Fourier

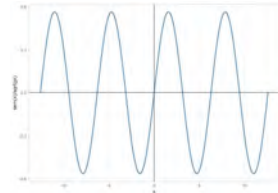
Veamos ahora una gráfica de alguno de los elementos de nuestra base ortonormal, los senos, los cosenos y las constantes:



(a) $\frac{1}{\sqrt{\pi}}$



(b) $\frac{\cos(x)}{\sqrt{\pi}}$



(c) $\frac{\text{sen}(x)}{\sqrt{\pi}}$

Podemos notar que en los tres tipos de elementos en la base no se distinguen “en el tiempo”. Es decir, que si notamos solo un intervalo que comience en algún punto α , se va a ver igual que otro intervalo que comience en $\alpha + 2k\pi$ para cualquier $k \in \mathbb{N}$. No podemos distinguir

cambios en el tiempo pues son funciones periódicas.

¿Por qué esto es un problema? Esto se debe a que si nuestra señal tiene impulsos solo en cierto tiempo, entonces no los vamos a poder distinguir. Si nuestra imagen tiene un área muy brillante o un área que queremos analizar detalladamente, entonces dicha área no se va a distinguir de las demás.

Lo lógico sería entonces enfocarnos en un área de nuestra función para aplicar la transformada de Fourier. Claramente hay muchas formas de hacerlo, dependiendo de qué se quiera resaltar o para qué se esté haciendo. Precisamente estamos “viendo” a la función con diferentes filtros (sus componentes vistas desde diferentes ventanas). Así llegamos a que un filtro es un operador en el espacio de funciones de “energía finita”, $L : \mathcal{L}^2 \rightarrow \mathcal{L}^2$.

Un tipo de filtros importante son los invariantes en el tiempo, los cuales, como su nombre indica, tienen la siguiente propiedad:

$$L(f(x - t_0)) = (Lf)(x - t_0).$$

Con esta definición llegamos naturalmente a la convolución de funciones.

Definición 16. Se llama convolución de dos funciones, f , h a la siguiente integral:

$$(h * f)(x) = \int_{-\infty}^{\infty} h(u) f(x - u) du.$$

A h se le llama el kernel del filtro o también se le llega a llamar el impulso respuesta del filtro, pues queda determinado por el impulso de la señal $h = L\delta(x)$.

Si ahora aplicamos el filtro L a la función de modulación de la transformada de Fourier, e^{-ixu} tenemos:

$$\begin{aligned}
L(e^{-i2\pi xu}) &= \int_{-\infty}^{\infty} h(u) e^{-i2\pi x(t-u)} du \\
&= \int_{-\infty}^{\infty} h(u) e^{-i2\pi xt} e^{i2\pi xu} du \\
&= e^{-i2\pi xt} \int_{-\infty}^{\infty} h(u) e^{i2\pi xu} du \\
&= e^{-i2\pi xt} H(x),
\end{aligned}$$

donde H es la transformada de Fourier de h . Aplicando el filtro lineal e invariante en el tiempo a f de su expresión en la transformada inversa de Fourier obtenemos:

$$\begin{aligned}
L(f(x)) &= L\left(\int_{-\infty}^{\infty} F(u) e^{ixu} du\right) \\
&= \int_{-\infty}^{\infty} L(F(u) e^{ixu}) du \\
&= \int_{-\infty}^{\infty} F(u) L(e^{-i2\pi xu}) du \\
&= \int_{-\infty}^{\infty} F(u) e^{-i2\pi xt} H(x) du.
\end{aligned}$$

Esta última igualdad nos muestra cómo el comportamiento repetitivo de suma de senos y cosenos es modulado por H , con el poder de enfocar solo algunas áreas que nos interesan o características que queramos resaltar. También con esta última igualdad podemos llegar a:

$$F(h * f) = F(h) F(f).$$

Podemos tener diferentes tipos de filtros, dependiendo de lo que queramos hacer:

- Filtros de bajo paso: tratan de enfocarse más en las frecuencias bajas que en las altas. En la figura 1.2 tenemos un ejemplo, vemos cómo el filtro solo toma valores positivos para frecuencias bajas, mientras que toma el valor de 0 para frecuencias más altas.

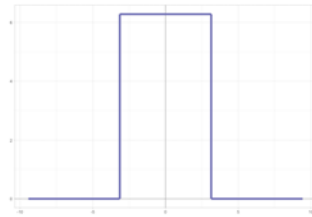


Figura 1.2. Ejemplo filtro de bajo paso.

- Filtros de alto paso: al contrario de los filtros de bajo paso, éstos se enfocan en frecuencias altas. En la figura 1.3 vemos cómo el filtro toma el valor de 0 solo para frecuencias bajas.

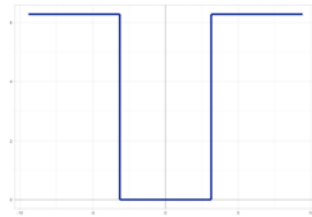


Figura 1.3. Ejemplo filtro de alto paso.

- Filtros de paso de banda: estos filtros se enfocan en ciertas frecuencias únicamente (en ciertas “bandas” de filtros). En la figura 1.4 vemos cómo el filtro toma el valor de 0 para todas las frecuencias menos en dos secciones, las dos bandas que queremos resaltar.

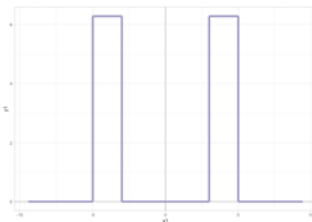


Figura 1.4. Ejemplo filtro de paso de banda.

- Filtros de detención de banda: al contrario de los filtros de paso de banda, éstos dejan pasar casi todas las frecuencias, menos algunas específicas. En la figura 1.5 vemos que el filtro toma el valor de 0 solo en dos bandas, dos secciones de frecuencias.

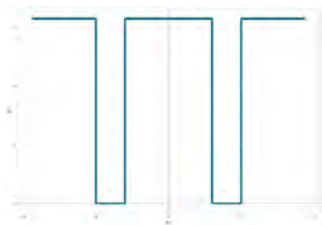


Figura 1.5. Ejemplo filtro de detención de banda.

Estos son algunos ejemplos de filtros lineales e invariantes en el tiempo, mas el estimado lector también se podrá preguntar sobre otro tipo de filtros. Claramente podemos filtrar información dependiendo de las necesidades que tengamos. Algunos filtros llegan a involucrar a la famosa distribución normal (también conocida como Gaussiana), pero éstos son más complicados que los aquí mencionados.

Ahora, si ya estamos pensando en filtrar la información, ¿por qué no aplicar la transformada de Fourier zona por zona? Hacer esto sería un poco análogo a la idea de la transformada rápida de Fourier, pero esta vez lo que buscamos es ver qué frecuencias resaltan en cada zona para así poder comparar las distintas zonas en una sola imagen. Esto nos lleva a estudiar la transformada de Fourier de tiempo corto o también conocida como la transformada de Fourier de ventana (en inglés *Short Time Fourier Transform STFT* o *Windowed Fourier Transform WFT*). Hemos de hacer notar que la mayoría de los algoritmos computacionales que la calculan usan las siglas en inglés STFT o WFT.

1.7. Transformada de Fourier de tiempo corto

Además de lo expresado en el párrafo anterior, ¿por qué es natural pensar en una transformada de Fourier de tiempo corto? Pensemos en

nuestra forma de procesar imágenes o sonidos. Nuestra memoria es limitada (algunos más que otros), no tenemos una cantidad infinita de información del pasado y tampoco tenemos información del futuro. Es decir que tenemos tiempo y espacio limitado para que nuestro cerebro procese toda esta información del exterior.

Una función es una “ventana”, si como las ventanas verdaderas, solo nos deja ver una parte del espacio. Es decir, una función es una “ventana” si toma el valor de cero fuera de cierto intervalo. Usualmente es simétrica alrededor de cierto valor y también llega a asociarse con funciones de probabilidad cortadas (para que tomen el valor de cero fuera de cierto rango, usualmente fuera de donde acumulan cierta masa de probabilidad). Si llamamos g a dicha función “ventana”, entonces aplicar dicho filtro a una función f resulta en una nueva función, $g_{x,t}$:

$$g_{x,t}(u) = g(u - t) e^{-2\pi i x u},$$

Lo cual no accidentalmente nos recuerda al kernel de una convolución o a los filtros que ya vimos. Como ejemplo, supongamos que g es la distribución normal modificada (pues tiene que valer cero fuera de nuestro intervalo) y que solo vamos a graficar el coseno de la función exponencial compleja. Podemos notar que en el centro, en este caso es 0 (la media de la función de probabilidad) tenemos la mayor amplitud, pero mientras nos vamos acercando a las orillas del intervalo la función va tendiendo a cero. Por ende solo nos concentramos en los valores cercanos a la media de la función de probabilidad, lo cual se expresa con la t .

En la figura 1.6 podemos apreciar la acción de aplicar una ventana de tipo Gaussiano (por la distribución Gaussiana) a la función coseno. Podemos ver que, en lugar de que el coseno oscile a lo largo de la recta real, dichas oscilaciones tienden a “aplanarse” en las orillas de la ventana.

De esta manera se define la transformada de Fourier de tiempo corto como:

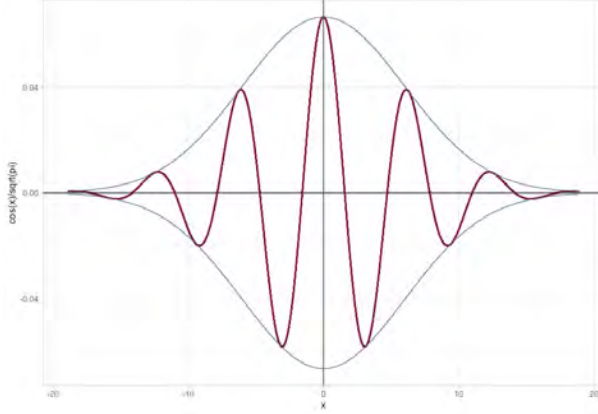


Figura 1.6. Función coseno visto con una ventana

$$\begin{aligned}
 \tilde{F}(x, t) &= \int_{-\infty}^{\infty} g(u - t) f(u) e^{-2\pi i x u} du \\
 &= \int_{-\infty}^{\infty} g_{x,t}(u) f(u) du = \langle g_{x,t}, f \rangle.
 \end{aligned} \tag{1.23}$$

Si queremos pasar a “dimensiones mayores”, entonces podemos comparar este caso con lo que hicimos en 1.21:

$$\begin{aligned}
 \tilde{F}(x, t) &= \int_{\mathbb{C}^n} g(u - t) f(u) e^{-2\pi i x^T u} du \\
 u, t, x &\in \mathbb{C}^n
 \end{aligned}$$

Donde tendríamos n “retrasos” del estilo $u - t$. También debemos notar que ahora esta transformada tiene “una alta dimensión” en el sentido en el que si $u, t, x \in \mathbb{C}^n$, entonces \tilde{F} tiene como dominio \mathbb{C}^{2n} . Volveremos a seguir desarrollando nuestra teoría en el caso “unidimensional”, más el querido lector ya puede tener claro que esta transformación no es única del caso univariado.

Recordemos que una de las mejores características de la transformada de Fourier es lo fácil que es lograr su inversión. Fácil en el sentido

computacional, pues para demostrar esto se requirieron varios lemas y teoremas como el teorema integral de Fourier visto en el teorema 5. Si usamos la transformada inversa de Fourier en 1.23 obtenemos lo siguiente:

$$\begin{aligned} g(u-t)f(u) &= f_t(u) = \int_{-\infty}^{\infty} F_t(\omega) e^{2\pi i u \omega} d\omega \\ &= \int_{-\infty}^{\infty} \tilde{F}(\omega, t) e^{2\pi i u \omega} d\omega. \end{aligned}$$

Ahora el problema que tenemos es que no podemos despejar a $f(u)$ de la expresión anterior, porque no sabemos si $g(u-t)$ se anula en alguna parte de su dominio. Sabemos que $g(u-t)$ no es la función cero pues esto sería equivalente a poner una ventana tal que se anule f , con lo que ya no tendríamos ningún tipo de información. Esto sería inútil. Lo que sí podemos hacer es lo siguiente:

$$\begin{aligned} g(u-t)f(u) &= f_t(u) = \int_{-\infty}^{\infty} \tilde{F}(x, t) e^{2\pi i u \omega} d\omega \\ \iff (g(u-t))^2 f(u) &= \int_{-\infty}^{\infty} g(u-t) \tilde{F}(x, t) e^{2\pi i u \omega} d\omega \\ \iff \int_{-\infty}^{\infty} (g(u-t))^2 f(u) dt &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u-t) \tilde{F}(x, t) e^{2\pi i u \omega} d\omega dt \\ \iff f(u) \int_{-\infty}^{\infty} (g(u-t))^2 dt &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u-t) \tilde{F}(x, t) e^{2\pi i u \omega} d\omega dt \\ \iff f(u) &= \frac{1}{\|g\|^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u-t) \tilde{F}(x, t) e^{2\pi i u \omega} d\omega dt. \end{aligned}$$

Esta es la manera de obtener la inversa de la transformada de Fourier de tiempo corto. Notemos que esta expresión también se puede interpretar como la descomposición de f en una suma infinita de ondas localizadas, las cuales tienen la siguiente forma:

$$g(u-t) e^{2\pi i u \omega} \tag{1.24}$$

A estas ondas localizadas se les llama átomos de tiempo-frecuencia. Pero notemos que es justo lo que se integra al hacer la convolución (definida en 16) entre g y la exponencial. Tal vez podríamos hacer una convolución con funciones diferentes para lograr diferentes representaciones de f , ¿no?

Notemos que si la función ventana g está bien localizada en el dominio del tiempo, entonces $\tilde{F}(x, t)$ también lo estará. Por la identidad de Parseval enunciada en 1.17 tenemos que:

$$\tilde{F}(x, t) = \langle g_{x,t}, f \rangle = \langle G_{x,t}, F \rangle,$$

donde G es la transformada de Fourier de g y F es la de f . Entonces, \tilde{F} también está localizada en el dominio de la frecuencia. Aún así tenemos un problema llamado el principio de incertidumbre. Para poder localizar frecuencias (incluso las frecuencias altas), debemos de medir la señal por un lapso considerado de tiempo. Esto hace que no podamos localizar bien la señal en el dominio del tiempo y en el dominio de frecuencias al mismo tiempo.

Si recordamos a la transformada discreta de Fourier, definida en 1.22 tenemos lo siguiente:

$$F(x_n) = \sum_{k=0}^{N-1} f_k e^{-ink \frac{2\pi}{N}},$$

entonces podemos decir que los elementos $\{1, e^{-in \frac{2\pi}{N}}, \dots, e^{-ink \frac{2\pi}{N}}\}$ son los átomos de dicha transformada. En general lo que queremos es descomponer a f en átomos de la siguiente manera:

$$f = \sum_{k \in \Omega} a_k g_k,$$

donde Ω es un conjunto contable para los átomos de tiempo y frecuencia, g_k . Pero dichos átomos no necesariamente tienen que ser de la forma $e^{-ink \frac{2\pi}{N}}$.

Como hicimos con la transformada de Fourier, también para el caso de la transformada de Fourier de tiempo corto podemos pensarla de forma discreta. Esto nos facilita entender cómo las computadoras hacen dichos cálculos. Como la transformada de Fourier de tiempo corto es de la siguiente manera,

$$\tilde{F}(x, t) = \int_{-\infty}^{\infty} g(u - t) f(u) e^{-2\pi i x u} du,$$

entonces podemos apreciar que los átomos son de la siguiente forma:

$$g_k = g_{k,j}(u) = g(u - kt_0) e^{-i2\pi j\omega_0 u}.$$

Esto nos dice que tenemos rectángulos en el dominio de tiempo-frecuencia, los t_0 y ω_0 nos dice en dónde se empiezan a formar dichos rectángulos y simplemente nos vamos desplazando con j y con k a través de ellos.

Esa figura nos compara las propiedades de localización de ambas transformadas en los dominios de frecuencia y tiempo. La transformada de Fourier nos sirve bastante bien para localizar frecuencias, sus átomos tienen la forma de $e^{-ink\frac{2\pi}{N}}$. La transformada de Fourier de tiempo corto nos sirve para localizar frecuencias y el tiempo donde se localizan dichas frecuencias, a pesar del principio de incertidumbre, sus átomos son de la forma $g(u - kt_0) e^{-i2\pi j\omega_0 u}$.

Ambas transformadas tienen como núcleo la exponencial, $e^{-i2\pi j\omega_0 u}$. Si recordamos ésto nos mapea a un círculo. ¿Podríamos pensar en tener otro tipo de función que no necesariamente sea un círculo?

Capítulo 2

Transformadas con Ondículas

La primera vez que se habló de ondículas fue en 1909 en la tesis doctoral de Alfredo Haar, la cual se puede revisar en [6]. Su nombre en inglés, “wavelet” es usado con mayor frecuencia. Haar fue un matemático húngaro cuyo trabajo doctoral fue supervisado por David Hilbert (sí, el matemático alemán cuyo apellido es nombre de ciertos espacios que antes mencionamos). Lo que hizo Haar en su tesis doctoral fue examinar varios sistemas ortonormales de funciones trigonométricas y otro tipo de sistemas ortonormales para relacionarlos con los sistemas de Sturm-Liouville. Entre estos tipos de sistemas ortonormales se encontró por primera vez el sistema ortonormal de ondículas de Haar. Intuitivamente podemos entender a las ondículas como una oscilación de tipo onda cuya amplitud empieza en cero, aumenta y luego disminuye de nuevo a cero. Son consideradas como “breves oscilaciones” y “ondas pequeñas”, de ahí su nombre.

Tuvo un gran impacto esta nueva manera de pensar. Varios matemáticos se interesaron en formar sistemas ortonormales de funciones para diferentes usos, como fue el caso de Jean Morlet y su famosa ondícula de Morlet. A pesar de esto, la mayor parte del análisis de ondículas fue desarrollado por Yves F. Meyer. Después de que se empezaran a crear algoritmos computacionales, las ondículas tuvieron un auge aún mayor para el procesamiento de señales, series de tiempo y mucho más. Sus aplicaciones las podemos ver en áreas sumamente

diferentes que van desde la ingeniería en audio, finanzas, procesamiento de imágenes, compresión, transformación de datos para regresión, etc.

2.1. Otro tipo de representaciones

Lo que hemos estado tratando de hacer con las series de Fourier y sus transformadas es poder obtener una representación de nuestra función de tal manera que podamos identificar las características importantes de esta función. Una representación del espacio de funciones \mathcal{F} es un operador entre dicho espacio y el espacio de sucesiones \mathcal{J} . Entonces, dada una función f su representación, denotada como $R(f)$ es una sucesión:

$$R(f) = (f_k)_{k \in \mathbb{Z}} \in \mathcal{J}.$$

Suponemos que en este espacio de funciones, \mathcal{F} hay una cierta norma ($\|\cdot\|_{\mathcal{F}}$) y que también el espacio de sucesiones tiene una norma ($\|\cdot\|_{\mathcal{J}}$). Entonces la representación de f tiene la siguiente propiedad:

$$\|R(f)\|_{\mathcal{J}} = \|f\|_{\mathcal{F}},$$

o al menos que se tiene cierta condición de estabilidad:

$$\|R(f)\|_{\mathcal{J}} \leq \|f\|_{\mathcal{F}}.$$

Notemos que estas relaciones son entre dos normas de dos espacios diferentes. Como estamos trabajando con funciones de energía constante, \mathcal{L}_2 y con su análogo para sucesiones, l_2 , entonces tenemos que el operador representación es tal que $R : \mathcal{L}_2 \rightarrow l_2$. Entonces de cierta manera lo que buscamos es “discretizar” las funciones en \mathcal{L}_2 .

La idea fundamental para representar funciones en bases, como las que hemos visto, es descomponer a dichas funciones como elementos de estas bases. Otra forma de descomponer funciones es con lo que se conoce como marcos.

Definición 17. *Teniendo un espacio de Hilbert \mathcal{H} con un producto interior, una colección de funciones $(\varphi_k)_{k \in \mathbb{Z}}$ se llama marco si existen*

las constantes α, β en \mathbb{R} tales que $0 < \alpha \leq \beta < \infty$ y tales que para cualquier $f \in \mathcal{H}$ se satisface que:

$$\alpha \|f\|^2 \leq \sum_{k=-\infty}^{\infty} |\langle f, \varphi_k \rangle|^2 \leq \beta \|f\|^2.$$

Claramente el sistema trigonométrico en el espacio de las funciones cuadrado integrables forman un marco, el cual hemos estado usando para toda esta parte de series y transformada de Fourier. Pero notemos que las funciones que forman un marco no necesariamente tienen que ser un conjunto ortonormal ni linealmente independiente. Entonces, ¿podremos formar otra base que sea un marco para poder representar mejor cierto tipo de funciones? De esta manera no tendríamos que usar senos y cosenos.

Si $\alpha = \beta$ entonces decimos que tenemos un marco ajustado. Por la identidad de Plancherel en 1.17 tenemos que, para una base ortonormal $\alpha = \beta = 1$. Lo que es impresionante es que las bases ortonormales son un caso particular de marcos, especialmente marcos ajustados.

Teorema 11. Si $\mathcal{B} = (\varphi_k)_{k \in \mathbb{Z}}$ es un marco ajustado tal que $\alpha = 1$ y $\|\varphi_k\| = 1 \quad \forall k$ entonces \mathcal{B} es una base ortonormal.

Demostración. “ \Leftarrow ” Ya lo hemos comentado, es un resultado inmediato. “ \Rightarrow ” Si $\mathcal{B} = (\varphi_k)_{k \in \mathbb{Z}}$ es un marco ajustado, entonces se cumple que:

$$\sum_{k=-\infty}^{\infty} |\langle f, \varphi_k \rangle|^2 = \|f\|^2.$$

Tenemos que probar 3 cosas:

- $\langle \varphi_k, \varphi_j \rangle = 0$ si $j \neq k$. Como la propiedad de arriba se cumple para cualquier f en el espacio de Hilbert, sea esta f particularmente φ_j ,

entonces tenemos lo siguiente:

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} |\langle \varphi_j, \varphi_k \rangle|^2 &= \|\varphi_j\|^2 \Leftrightarrow \sum_{k=-\infty}^{\infty} |\langle \varphi_j, \varphi_k \rangle|^2 = \langle \varphi_j, \varphi_j \rangle \\
&\Leftrightarrow \dots + \langle \varphi_j, \varphi_{j-1} \rangle^2 + \langle \varphi_j, \varphi_j \rangle^2 + \langle \varphi_j, \varphi_{j+1} \rangle^2 + \dots = \langle \varphi_j, \varphi_j \rangle \\
&\Leftrightarrow \dots + \langle \varphi_j, \varphi_{j-1} \rangle^2 + 1 + \langle \varphi_j, \varphi_{j+1} \rangle^2 + \dots = 1 \\
&\Leftrightarrow \dots + \langle \varphi_j, \varphi_{j-1} \rangle^2 + \langle \varphi_j, \varphi_{j+1} \rangle^2 + \dots = 0 \\
&\Leftrightarrow \langle \varphi_j, \varphi_k \rangle = 0 \quad \forall k \neq j
\end{aligned}$$

- $\|\varphi_k\| = 1 \quad \forall k$ es directo por la hipótesis
- El conjunto \mathcal{B} genera a todas las funciones del espacio de Hilbert. Por las propiedades anteriores tenemos que \mathcal{B} contiene funciones linealmente independientes y de norma uno, falta para cualquier f existen escalares α_k tales que $f = \sum_{k=-\infty}^{\infty} \alpha_k \varphi_k$. Pero, por las hipótesis tenemos que $\sum_{k=-\infty}^{\infty} \|\langle f, \varphi_k \rangle\|^2 = \|f\|^2$ y $\|\varphi_k\| = 1$. Si el querido lector se acuerda, $\langle f, \varphi_k \rangle$ son los coeficientes de Fourier. Dado que se cumple la igualdad en este caso, por ser un marco ajustado, entonces \mathcal{B} sí es una base ortonormal.

■

Con lo que acabamos de ver más el hecho que probablemente podamos hacer convoluciones con funciones que no necesariamente tengan que ver con la exponencial tenemos una buena base para hablar sobre las ondículas. Ya habíamos mencionado que una ondícula puede ser considerada como una breve oscilación, pero ¿qué las hace distinguirse? Para poder analizar una señal (que en nuestro caso son imágenes) necesitamos un conjunto de ondículas. Nos gustaría descomponer esta señal usando estos conjuntos de ondículas, que dicha descomposición sea, preferentemente, reversible y con una mínima pérdida de información. Formalmente, esta representación usando series de ondículas es una representación de funciones cuadrado integrables con respecto a un conjunto ortonormal de funciones base o respecto a un marco en un espacio vectorial.

Lo que tenemos que destacar antes es el hecho de que la transformada de ventana de Fourier usa una función de n variables y nos regresa un conjunto de funciones de $2n$ variables. ¿Qué significa esto? No solo tenemos los valores transformados para cada x sino también su localización t . Algo similar pasa con la transformada de ondículas, solo que para dicha transformada necesitamos una función ψ la cual va a ser la “madre ondícula”, la base de nuestra transformación. Con $\psi(t)$ tenemos un candidato a función de modulación (como la exponencial en el caso de Fourier). Es evidente que después de la discusión que hemos tenido, queremos que ψ tenga una buena localización en el tiempo. Ahora lo que tenemos que hacer es variar su escala para así obtener una familia de funciones (no necesariamente ortogonales). Una condición que se le pone a esta “madre ondícula”, como lo menciona Daubechies en [7] es la siguiente:

$$\int_{-\infty}^{\infty} \psi(u) du = 0.$$

Definición 18. *Dada una “madre ondícula” ψ obtenemos una familia de ondículas dado un valor $p \geq 0$ y para cualquier $s \in \mathbb{R} \setminus \{0\}$ de la siguiente manera:*

$$\psi_s(u) = |s|^{-p} \psi\left(\frac{u}{s}\right) = \frac{1}{|s|^p} \psi\left(\frac{u}{s}\right).$$

Usamos $|\cdot|$ pues estas variables podrían estar en \mathbb{R} , representando valor absoluto, o en \mathbb{C} , representando el módulo. Notemos que de esta manera cambiamos la “amplitud” de las ondículas en esta familia con $\frac{1}{|s|^p}$ y al poner $\frac{u}{s}$ en el argumento de ψ cambiamos las frecuencias de la “madre ondícula”. Ahora, como lo hicimos para la transformada de ventana de Fourier queremos localizar a ψ en el tiempo (o espacio, dependiendo de la interpretación que se le de a la variable t). Para esto hacemos lo siguiente:

$$\psi_{s,t} = \psi_s(u - t) = \frac{1}{|s|^p} \psi\left(\frac{u - t}{s}\right). \quad (2.1)$$

Como fue en el caso de Fourier, si $\psi \in \mathcal{L}_2$ entonces las $\psi_{s,t}$ también lo estarán. Adicionalmente obtenemos una versión similar a la identidad

de Parseval (pero claramente escalada):

$$\|\psi_{s,t}\|^2 = |s|^{1-2p} \|\psi\|^2$$

Con todos estos elementos podemos definir una transformada usando las ideas que ya hemos mencionado en el capítulo anterior:

Definición 19 (Transformada de ondícula abstracta). *Dada una madre ondícula ψ tal que esta función esté en \mathcal{L}_2 y usando la familia de ondículas $\psi_{s,t}$ como previamente las definimos, obtenemos la transformada de ondículas:*

$$\tilde{f}(s, t) = \int_{-\infty}^{\infty} f(u) \psi_{s,t}(u) du = \langle \psi_{s,t}, f \rangle.$$

Pausemos un momento para apreciar la flexibilidad de esta transformada. No estamos diciendo quién es ψ , s , t o p . Podemos elegir estos parámetros como mejor nos convenga dependiendo de la situación.

Es de esperarse, como en el caso de la transformada de Fourier de ventana, que la transformada inversa no va a ser tan directa como en el caso de la transformada de Fourier “original”. Usando la identidad de Parseval como en 1.17 y la definición anterior tenemos:

$$\tilde{f}(s, t) = \langle \psi_{s,t}, f \rangle = \langle \Psi_{s,t}, F \rangle,$$

donde $\Psi_{s,t}$ y F son las transformadas de Fourier de $\psi_{s,t}$ y f respectivamente. Por la definición misma de $\psi_{s,t}$ con la madre ondícula, ψ tenemos que:

$$\begin{aligned}
\Psi_{s,t}(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_{s,t}(u) e^{-2\pi i x u} du \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\frac{1}{|s|^p} \psi \left(\frac{u-t}{s} \right) \right) e^{-2\pi i x u} du \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\frac{1}{|s|^p} \psi(\hat{u}) \right) e^{-2\pi i x (s\hat{u}+t)} s d\hat{u} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |s|^{1-p} \psi(\hat{u}) e^{-2\pi i (sx)\hat{u}} e^{-2\pi i x t} d\hat{u} \\
&= |s|^{1-p} e^{-2\pi i x t} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(\hat{u}) e^{-2\pi i (sx)\hat{u}} d\hat{u} \right) \\
&= |s|^{1-p} e^{-2\pi i x t} \Psi(sx) \Rightarrow \Psi(sx) = |s|^{p-1} e^{2\pi i x t} \Psi_{s,t}(x)
\end{aligned}$$

Ahora, si consideramos a $g(x) = \psi(sx)f(x)$, entonces podemos sustituir los valores que hemos encontrado para obtener lo siguiente

$$\begin{aligned}
\tilde{f}(s,t) &= \langle \Psi_{s,t}, F \rangle = \int_{-\infty}^{\infty} \Psi_{s,t}(u) F(u) du \\
&= \int_{-\infty}^{\infty} |s|^{p-1} e^{2\pi i u t} \Psi_{s,t}(u) F(u) du \\
&= |s|^{p-1} \int_{-\infty}^{\infty} e^{2\pi i u t} \Psi_{s,t}(u) F(u) du \\
&= |s|^{p-1} \int_{-\infty}^{\infty} e^{2\pi i u t} G(u) du = |s|^{p-1} g.
\end{aligned}$$

donde G es la transformada de Fourier de g . Ahora, si aplicamos la inversa de Fourier a este resultado obtenemos

$$\int_{-\infty}^{\infty} e^{-2\pi i t x} \tilde{f}(s,t) dt = |s|^{p-1} G(x) = |s|^{p-1} \Psi(sx) F(x).$$

No podemos despejar F de la expresión anterior, pues Ψ se podría anular en alguna parte del dominio. Para esto necesitamos una condición adicional para ψ

$$C = \int_{-\infty}^{\infty} \frac{|\Psi(u)|^2}{|u|} du < \infty,$$

llamada condición de admisibilidad. También tenemos que contar con el famoso teorema de Fubini que establece la condición necesaria para poder cambiar el orden de integración. Suponer esta condición es bastante razonable puesto que lo que queremos es trabajar con ondículas localizadas y funciones en \mathcal{L}_2 , entonces las integrales con las que vayamos a usar el teorema de Fubini las podemos suponer como absolutamente finitas. Con estas condiciones ahora podemos “rescatar” a f de la transformada de ondícula:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-2\pi itx} \tilde{f}(s, t) dt &= |s|^{p-1} \Psi(sx) F(x) \\ \Rightarrow \Psi(sx) |s|^{2-p} \int_{-\infty}^{\infty} e^{-2\pi itx} \tilde{f}(s, t) dt &= \frac{(\Psi(sx))}{|s|} F(x) \\ \Rightarrow \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(sx) |s|^{2-p} e^{-2\pi itx} \tilde{f}(s, t) dt ds &= F(x) \left(\int_{-\infty}^{\infty} \frac{(\Psi(sx))}{|s|} ds \right) \\ \Rightarrow \frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s|^{2-p} e^{-2\pi itx} \Psi(sx) e^{-2\pi itx} \tilde{f}(s, t) dt ds &= F(x) \\ \Rightarrow e^{2\pi iux} \frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s|^{2-p} e^{-2\pi itx} \Psi(sx) e^{-2\pi itx} \tilde{f}(s, t) dt ds &= e^{2\pi iux} F(x) \\ \Rightarrow \frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s|^{2-p} e^{-2\pi itx} \Psi(sx) e^{2\pi i(t-u)x} \tilde{f}(s, t) dt ds &= e^{2\pi iux} F(x) \\ \Rightarrow \frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s|^{2-p} \Psi(sx) e^{2\pi i(t-u)x} \tilde{f}(s, t) dt ds dx &= \\ \int_{-\infty}^{\infty} e^{2\pi iux} F(x) dx & \\ \Rightarrow \frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \Psi(sx) e^{2\pi i(t-u)x} dx \right) |s|^{2-p} \tilde{f}(s, t) dt ds &= \\ \int_{-\infty}^{\infty} e^{2\pi iux} F(x) dx & \end{aligned}$$

ahora, si hacemos el cambio de variable $\hat{x} = sx$ obtenemos:

$$\Rightarrow \frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \Psi(\hat{x}) e^{2\pi i \left(\frac{t-u}{s}\right) \hat{x}} d\hat{x} \right) |s|^{2-p} \tilde{f}(s, t) dt ds = \int_{-\infty}^{\infty} e^{2\pi i u x} F(x) dx,$$

de manera que:

$$\frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s|^{2-p} \psi\left(\frac{u-t}{s}\right) \tilde{f}(s, t) dt ds = \int_{-\infty}^{\infty} e^{2\pi i u x} F(x) dx,$$

lo cual implica:

$$\frac{1}{C} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s|^{2p-3} \psi_{s,t}(u) \tilde{f}(s, t) dt ds = f(u). \quad (2.2)$$

Entonces esta es la forma de “recuperar” a nuestra función original f de su transformada de ondícula continua bajo ciertas condiciones. También podemos interpretar este resultado como que f es la superposición de ondículas localizadas. Tenemos que notar que aquellas funciones que satisfacen la condición de admisibilidad que citamos arriba son llamadas ondículas. Sin dar detalles podemos considerar una ondícula como una función localizada en el tiempo y en el espacio que puede o no tener varias frecuencias asociadas a ellas. Por la condición de admisibilidad y para que dicha función pueda satisfacer las condiciones del teorema de Fubini necesitamos que cumpla lo siguiente:

$$\lim_{u \rightarrow 0} \Psi(u) = 0,$$

pues por dicha condición tenemos que $\int_{-\infty}^{\infty} \frac{|\Psi(u)|^2}{|u|} du < \infty$ y como $\frac{|\Psi(u)|^2}{|u|} \geq 0 \quad \forall u$ lo único que nos queda es que se cumpla el límite anterior. Ahora, como la función exponencial es positiva en todo su dominio, forzosamente tenemos que la ondícula también cumple que:

$$\int_{-\infty}^{\infty} \psi(u) du = 0.$$

Esto nos da una bonita característica: ψ debe de oscilar alrededor del 0. Entonces ψ fuera de cierto intervalo toma valores muy chicos, tal vez

cero y oscila alrededor del cero. Por estas características a estas funciones se les llama ondículas, o en inglés “wavelets”, son ondas pequeñas.

A esta transformada y forma de “recuperar” a nuestra función original f se le conoce como la transformada continua de ondícula de f . Esto se debe a que los parámetros de escala y traslación (s, p) varían continuamente en el tiempo. Hemos asumido que ψ es una función de variable real. Si consideramos a ψ como una función compleja (en \mathbb{C}) entonces en 2.2 tendríamos que escribir $\overline{\psi}$ (conjugado de ψ) en lugar de simplemente escribir ψ . Análogamente, en la definición de la transformada de ondícula abstracta (en 19) también tenemos que escribir $\overline{\psi_{s,t}}$ en lugar de simplemente $\psi_{s,t}$.

2.2. Transformada discreta de ondícula

Ya vimos que la transformada de Fourier y la transformada de Fourier de ventana tienen versiones discretas (son las que se usan en aplicaciones para computadoras). Nuestra intuición nos dirá que seguramente existe la versión discreta de la transformada de ondícula. Esta intuición está en lo correcto.

En el caso de transformadas de ondículas, lo que estamos discretizando son los parámetros de escala y traslación t, s . Para este caso no tenemos una transformada inversa (como la teníamos con la transformada continua), la reconstruimos con 2.2. Si queremos reconstruir nuestra función de esta transformada debemos de hacerlo de otra manera. Esto ya dependerá de la madre ondícula que queramos usar, nuestra ψ . Sabiendo esto podríamos aplicar incluso diferentes métodos numéricos

Si tenemos una partición en el dominio de tiempo-escala de tal manera que:

$$\Delta_{t_0, \omega_0} = \{(mt_0, n\omega_0); m, n, \in \mathbb{Z}\}.$$

Sabemos que al escalar las ondículas actuamos de forma multiplicativa,

así que pasar a la versión discreta de la escala sería:

$$s_m = s_0^m.$$

Nos falta hacer la discretización del tiempo. Para esto debemos de notar que la ondícula es invariante a cambios de escala. Es decir, si cambiamos la escala de f y cambiamos la escala del espacio donde estemos trabajando, entonces la transformada de ondícula no cambia, esto es:

$$\begin{aligned} f_{s_0} &= s_0^{-1/2} f\left(\frac{t}{s_0}\right) \\ \Rightarrow \tilde{f}_{s_0}(s_0 s, s_0 t) &= \tilde{f}(s, t). \end{aligned}$$

Entonces para cada escala s_0^m la discretización de la “malla” del dominio espacio-escala es:

$$t_{m,n} = n s_0^m t_0, \quad n \in \mathbb{N}.$$

Esto último solo nos indica que cuando nos movemos con los índices m, n uno de ellos actúa moviéndose a lo largo de la ventana del tiempo y otro amplía la del espacio (escala). Con esto, la partición del dominio tiempo-escala es:

$$\Delta = \{(s_0^m, n s_0^m t_0), \quad m, n \in \mathbb{Z}\}.$$

Por ende tenemos diferentes opciones para realizar esta partición, dependiendo de lo que queramos hacer con la transformada y la función que vayamos a transformar.

Ahora, como en el caso de la transformada de ventana de Fourier tenemos que definir los átomos que vamos a usar en cada una de las “nuevas ventanas”. Lo más difícil ya lo hemos hecho, que es construir la partición del espacio tiempo-escala. El problema que teníamos con la transformada de Fourier es que no era localizable en el tiempo, pero este problema ya no lo tenemos con las ondículas. Entonces simplemente podemos realizar la discretización de las ondículas de esta manera:

$$\begin{aligned}
\psi_{m,n}(u) &= \psi_{s_0^m, nt_0 s_0^m}(u) = s_0^{-\frac{m}{2}} \psi\left(\frac{u - nt_0 s_0^m}{s_0^m}\right) \\
&= s_0^{-\frac{m}{2}} \psi(s_0^{-m}u - nt_0).
\end{aligned}$$

Y así obtenemos:

$$\tilde{f}_{m,n}(s, t) = \langle f, \psi_{m,n}(u) \rangle.$$

Notemos que lo que hicimos aquí fue la discretización del dominio de tiempo-escala. Tenemos una forma ahora de “contar” las escalas y tiempos para hacer la transformada de ondícula. Con esto tenemos una sucesión, $\langle f, \psi_{m,n}(u) \rangle$ $m, n \in \mathbb{Z}$ para representar a f . Recordemos que las ondículas toman valores diferentes a cero en una cierta ventana del dominio de tiempo-escala entonces esta sucesión nos da información sobre dichas ventanas.

Para poder determinar si esta es una buena representación de la función f nos gustaría dar algunas características de las ondículas y del dominio tiempo-escala.

2.3. Análisis multirresolución

En un sentido muy formal y abstracto, un análisis multirresolución es uno hecho en un espacio de Lebesgue L_2 y consiste en una sucesión de subespacios anidados que se ven de la siguiente manera:

$$\{\emptyset\} \subset \dots \subset V_1 \subset V_0 \subset V_{-1} \subset \dots \subset V_{-n} \subset \dots \subset L_2.$$

Estos subespacios satisfacen ciertas condiciones de similitud (en sentido del dominio de espacio-tiempo o escala-frecuencia).

El lector ya tiene todas las herramientas para entender lo que es un análisis multirresolución. Todo esto en el sentido de las ondículas que acabamos de describir en la sección anterior. El propósito de esta sección es justamente explorar este tipo de análisis que nos permiten

las ondículas.

Empezamos por entender cómo es que nuestro cerebro percibe nuestro ambiente (sean pinturas, personas, sentimientos, lenguaje, etc.), como se lee en el capítulo de Charles D. Gilbert en [8] sobre la neurobiología detrás del procesamiento de información visual. La información que recibimos de un estímulo (por ejemplo, la información visual al ver una pintura) no es la única información que usamos para entender el mundo, lo que estamos viendo. Más información es involucrada en este proceso, como la expectativa, nuestra memoria, etc. Es por esto que necesitamos muchas herramientas para poder tratar de analizar el mundo exterior.

Cuando nuestro ojo recibe información, lo que está recibiendo es luz con distintas longitudes de onda, como lo explica Edmund T. Rolls en [9]. Estas longitudes de onda son interpretadas por nuestro cerebro como color, mas la luz por sí misma no tiene color. Cuando la luz llega a nuestro ojo, llega a un conjunto de capas llamadas retina. Aquí tenemos fotorreceptores, los cuales son células capaces de convertir esta energía óptica (luz) en energía eléctrica. Tenemos dos tipos de fotorreceptores: bastones y conos. Los bastones sirven para captar cambios de intensidad, nos sirven para detectar cambios entre oscuro y claro. Los conos nos sirven para distinguir colores. En el centro de nuestra retina captamos información de donde estamos enfocando nuestra vista. En esta sección tenemos solamente conos. En la periferia tenemos bastones. Nuestro cerebro automáticamente hace la distinción entre lo que enfocamos y lo que no.

Por esto es muy natural analizar funciones (información) bajo diferentes escalas. Podríamos decir que la cantidad de conos y bastones en cada parte de la retina simbolizan diferentes escalas (de color, de intensidad). Es también por esta justificación biológica que la idea de escala está relacionada con problema de muestreo puntual. ¿En qué nos tenemos que fijar para tener una buena idea de lo que estamos percibiendo? ¿En qué tenemos que enfocar nuestra atención? El querido lector se podrá dar cuenta que estas preguntas se pueden hacer

dentro del contexto matemático (funcional) como dentro del contexto neurobiológico (visual).

Llamamos frecuencia de muestreo al número de muestras dentro de una unidad de tiempo. El intervalo de tiempo en el que estamos recolectando dicha información se llama periodo de muestreo. Si fijamos una escala, digamos $|s|^p$ entonces estamos fijando una escala para representar nuestra información (función), entonces todos aquellos detalles fuera de esta escala estarán perdidos. Pero si tenemos $\hat{p} > p$ entonces dichos detalles los podríamos rescatar con una escala $|s|^{\hat{p}}$.

¿Qué tiene que ver esto con la definición que dimos al inicio de esta sección? Para algunos casos particulares de la madre ondícula ψ y de los parámetros de escala y traslación, s, t , las ondículas $\psi_{m,n}$ constituyen una base ortonormal de \mathcal{L}_2 , siendo esto asociado al análisis multirresolución del que vinimos hablando. También lo podemos interpretar de la siguiente manera: para el entero p estamos viendo aquellas funciones que tienen escala $|s|^p$. Es decir, estamos en un subespacio de L_2 donde la escala es justamente $|s|^p$. Llamémosle a este subespacio V_p . Así, conforme cambiemos escalas vamos incluyendo diferentes funciones. Si p es negativa entonces estamos viendo detalles específicos, si p es positiva y grande estamos viendo rasgos generales. Es por esto que tenemos la sucesión $\{\emptyset\} \subset \dots \subset V_1 \subset V_0 \subset V_{-1} \subset \dots \subset V_{-n} \subset \dots \subset L_2$.

Lo que ahora nos hace falta es un operador que pueda representar las características de nuestra función en estas diferentes escalas. Ahora, queremos hacer estos cambios de escala. En la sección anterior nos dimos cuenta que las ondículas son particularmente buenas para esta tarea. También nos dimos cuenta que una familia de ondículas puede o no ser ortonormal. Como se podrá haber dado cuenta el lector, las bases más bonitas son las ortonormales (en el sentido que cumplen con ciertas propiedades que consideramos deseables). Entonces podemos suponer que existe una función $\phi \in \mathcal{L}_2$ tal que la familia definida por

$$\phi_{j,k}(u) = 2^{-j/2} \phi(2^{-j}u - k) \quad , \quad j, k \in \mathbb{Z}$$

es ortonormal y forman una base de V_j . Esta definición nos recuerda mucho a la familia que definimos en 2.1 a partir de una ondícula madre, solo que en este caso esa ondícula madre es ϕ . Siguiendo con la analogía que ya hemos desarrollado, con esta ϕ podemos definir diferentes escalas:

$$\phi_s(u) = |s|^{-1/2} \phi\left(\frac{u}{s}\right).$$

Con esto tenemos que el “ancho” de la “madre ondícula”, ϕ , es s veces el ancho de la ondícula ϕ_s . Así, conforme la escala aumenta (o decrece), la “anchura” de ϕ_s también lo hace. Es como si estuviéramos cambiando el tamaño de la foto de una ondícula con el curso en una computadora. Aquí, V_j se llama el espacio de escala j , o simplemente espacio de escala. Los detalles a escala j también aparecen a escala $j - 1$, pues la ondícula con escala $j - 1$ es menos ancha, capta los detalles de j , pero también detalles más finos. Con esto ya tenemos nuestra sucesión de subespacios anidados de \mathcal{L}_2 , $\{\emptyset\} \subset \dots \subset V_1 \subset V_0 \subset V_{-1} \subset \dots \subset V_{-n} \subset \dots \subset L_2$. Con esta propiedad tenemos que:

$$f \in V_j \Leftrightarrow f(2^j u) \in V_0.$$

De hecho, el conjunto $\tau = \{\emptyset, L_2, V_j, j \in \mathbb{Z}\}$ es una topología en \mathcal{L}_2 . Es decir, con estos espacios podemos describir la relación “espacial” entre cualesquiera elementos de L_2 , que es justamente como nuestro cerebro interpreta estímulos visuales. Esto solo lo usaremos como dato para convencer al lector que lo que estamos haciendo es precisamente tratar con cambios espaciales en funciones y tratar de analizarlos con funciones especiales (ondículas).

Con todo lo que hemos desarrollado podemos definir lo que es una representación multirresolución:

Definición 20 (Representación multirresolución). *Definimos una representación multirresolución en L_2 como un análisis multirresolución, en el sentido que ya hemos definido y que además cuenta con las siguientes propiedades:*

- $V_j \subset V_{j-1}, \forall j \in \mathbb{Z}$

- $f \in V_j \Rightarrow f(2u) \in V_{j-1}$
- $\bigcap_{j \in \mathbb{Z}} V_j = \emptyset$
- $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = \mathcal{L}_2$ (la cerradura de esta unión)
- Existe una función, $\phi \in V_0$, tal que el conjunto $\{\phi(u - k); k \in \mathbb{Z}\}$ es una base ortonormal de V_0

A la función ϕ se le llama función de escala para la representación multirresolución.

Podemos considerar a ϕ como la “base” o “generadora” (no en el sentido algebraico) de las escalas. En un sentido intuitivo, esta definición nos quiere decir que todos los espacios V_k son versiones “escaladas” del espacio “central” V_0 . En sí, todos los espacios se pueden reescalar para ser versiones de cualquiera que nos gustara. Por ejemplo, si decidimos que el “centro” va a ser el V_m entonces podemos ver a todos los demás espacios como un escalamiento de nuestro “centro”.

2.4. Análisis multirresolución y ondículas

Una pregunta muy válida que el estimado lector se podrá estar haciendo es, ¿qué pasa con $V_{j-1} \setminus V_j$? Es decir, con los elementos que pertenecen a V_{j-1} , pero que no pertenecen a V_j . Esto sería muy útil si, por ejemplo, desde un punto de vista de teoría de la medida, quisiéramos definir una σ -álgebra. Para hacer esto y que sigamos teniendo bases ortonormales necesitamos unas definiciones adicionales.

Definición 21 (Complemento ortonormal). Sea V un espacio vectorial con producto interior $\langle \cdot, \cdot \rangle$ y sea U un subespacio de V . El complemento ortonormal de U es el conjunto:

$$U^\perp = \{w \in V : \langle w, v \rangle = 0 \forall w \in U\}$$

Es decir, son aquellos elementos que le “faltan” a U para ser V y además son ortogonales a todos los elementos de U . También tenemos que definir lo que es la suma directa en espacios vectoriales, la cual es distinta a la suma de elementos en espacios vectoriales que ya conocemos.

Definición 22 (Suma directa). *La suma directa entre dos subespacios vectoriales, W , U tales que $W \cap U = \{\emptyset\}$, se define de la siguiente manera:*

$$W \oplus U = \{w + u : w \in W, u \in U\}$$

Es decir, la suma directa entre U y W es el conjunto de elementos tales que se pueden descomponer de manera única como la suma entre algún elemento en W y algún otro en U .

Ahora nos falta definir un operador llamado proyección, pero vamos a definir una proyección especial, la cual va a ser una proyección ortogonal.

Definición 23 (Proyección ortogonal). *Supongamos que tenemos la descomposición en suma directa de algún espacio V , tal que $V = M \oplus N$, es decir que para cualquier $v \in V$ lo podemos escribir como $v = m + n$ de manera única, donde $m \in M$ y $n \in N$. Ahora, el mapeo definido como $Ez = m$ se llama la proyección de M sobre N .*

Lo especial de una proyección ortogonal es que en este caso descomponemos a V en espacios ortogonales mediante el uso de complementos ortogonales: $V = M \oplus M^\perp$. Así tenemos que para cualquier $v \in V$ lo descomponemos en partes ortogonales: $v = m + \tilde{m}$ donde $m \perp \tilde{m}$. Y así nuestra proyección ortogonal es $P(v) = m$.

Con lo que acabamos de definir podemos responder a la pregunta que nos planteamos al inicio de esta sección. ¿Qué pasa con $V_{j-1} \setminus V_j$? Una forma de construir esta respuesta es mediante complementos ortogonales. Recordemos que $V_j \subset V_{j-1}$ podemos definir a W_j como el complemento ortogonal de V_j en V_{j-1} , es decir:

$$W_j = \{w \in V_{j-1} : \langle w, v \rangle = 0 \forall v \in V_j\},$$

Ahora para poder escribir explícitamente la relación que buscamos entre V_{j-1} y V_j usaremos el siguiente lema:

Lema. *Sea V un espacio vectorial con producto interior y sea U un subespacio de V , entonces $V = U \oplus U^\perp$.*

Con esto podemos tener la siguiente relación entre V_{j-1} y V_j :

$$V_{j-1} = V_j \bigoplus W_j.$$

Es decir, W_j es lo que “le falta” a V_j para ser V_{j-1} . Notemos que los elementos de W_j son ortogonales a W_k para $k \neq j$. Entonces podemos escribir a V_j como la suma directa de todas estas partes que le faltan a algún V_{j_0} , donde $j_0 > j$:

$$V_j = V_{j_0} \bigoplus_{k=0}^{j_0-j} W_k.$$

Y así podemos “completar” V_{j_0} para hacerlo V_j . Otra manera de interpretar esto es decir que aquellas señales cuyo espectro esté en la banda de frecuencias de V_j son la suma de señales que están en la banda de frecuencias de V_{j_0} y las que están en la banda de frecuencias de las W_k . Nos dice un aspecto de esta suma: es una suma entre elementos ortogonales. Por la cadena de subespacios anidados que teníamos, ahora podemos escribir a todo L_2 como la suma directa entre estas partes de subespacios, es decir:

$$L_2 = \bigoplus_{k \in \mathbb{Z}} W_k. \quad (2.3)$$

Entonces ya tenemos una manera de descomponer a L_2 mediante la suma directa de subespacios ortogonales entre sí. Lo que no sabemos es qué tipo de bases tienen las W_k . Nos gustaría que estas bases fueran ondículas, ¿no?

Teorema 12. *Dada una representación multirresolución en \mathcal{L}_2 (cumple las condiciones de la definición 20), entonces existe una base ortonormal de ondículas $\{\psi_{m,n}; m, n \in \mathbb{Z}\}$ estando en \mathcal{L}_2 construidas a partir de una madre ondícula, como en 2.1 tal que para toda función f en \mathcal{L}_2 tenemos que:*

$$P_{j-1}(f) = P_j(f) + \sum_{k \in \mathbb{Z}} \langle f, \psi_{m,k} \rangle \psi_{m,k},$$

donde el operador P_i (como en la definición 23) es la proyección ortogonal de f en el i -ésimo espacio, en V_i .

Demostración. Esta demostración es sumamente bella e intuitiva para el querido lector, pues le ofrece una construcción de nuestras ondículas de manera explícita. Veamos cómo.

Notemos que los espacios W_k son ortogonales entre sí, es decir $W_k \perp W_j$ para todo $k \neq j$. Ahora, por la ecuación 2.3 tenemos una manera de descomponer a \mathcal{L}_2 en subespacios ortogonales. También tenemos que estos espacios W_k “heredan” las propiedades de escalamiento de los espacios V_k , es decir siguen cumpliendo con el punto 2 de la definición de representación multirresolución:

$$f(u) \in W_k \iff f(2^k u) \in W_0.$$

Esta propiedad es bastante linda porque si para alguna k fija obtenemos una base ortonormal con ondículas de W_k , entonces por la propiedad de escala de las ondículas ya tenemos una base ortonormal para cualquier otro W_j y por ende para \mathcal{L}_2 completo (por 2.3).

Ahora tenemos que poner todo nuestro esfuerzo en construir esta ψ para algún W_k . Primero vamos a enlistar unas propiedades de ϕ y de W_0 , ϕ en el sentido de la quinta propiedad de la definición de representación multirresolución.

1. Por definición tenemos que existe una $\phi \in V_0$ tal que el conjunto $\{\phi(u - k); k \in \mathbb{Z}\}$ es una base ortonormal de V_0 . Como $V_0 \subset V_{-1}$, si la base ortonormal de V_{-1} es $\{\phi_{-1,n}\}_{n \in \mathbb{Z}}$, entonces:

$$\phi = \sum_{n \in \mathbb{Z}} \langle \phi, \phi_{-1,n} \rangle \phi_{-1,n} = \sum_{n \in \mathbb{Z}} h_n \phi_{-1,n}.$$

Además también tenemos que $\sum_{n \in \mathbb{Z}} |h_n|^2 = 1$. Como tenemos la propiedad 2 y las $\phi(x - k)$ son una base ortonormal de V_0 , entonces las podemos reescalar para que las $\phi_{-1,n}$ sean estos reescalamientos: $\phi_{-1,n}(x) = \phi(2x - n)$. Pero queremos que sea una base ortonormal y por el reescalamiento que hicimos ahora necesitamos multiplicar por un factor de $\sqrt{2}$. Entonces tenemos lo siguiente:

$$\phi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \phi(2x - n).$$

Ahora, usando la transformada de Fourier esto es:

$$\Phi(u) = \frac{1}{\sqrt{2}} \sum_n h_n e^{-inu/2} \Phi(u/2),$$

donde Φ es la transformada de Fourier de ϕ . Ahora, para “acortar” lo que acabamos de escribir, si $m_0(u) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n e^{-inu}$, entonces lo de arriba lo podemos escribir como:

$$\Phi(u) = m_0(u/2) \Phi(u/2).$$

El estimado lector ahora puede notar que m_0 es una función periódica (con periodo 2π por el “efecto” que tiene e^{-inu}). También se podrá dar cuenta que, como estamos trabajando solo con funciones en \mathcal{L}_2 entonces m_0 también estará en \mathcal{L}_2 .

2. El querido lector podrá recordar que justo $\phi(x-k)$ forma una base ortonormal. Esto evidentemente va a tener consecuencias para la m_0 que definimos en el inciso anterior. Si $\delta_{k,0}$ es la delta de Kronecker entonces por la ortogonalidad de las ϕ tenemos que $\delta_{k,0} = \langle \phi(x), \phi(x-k) \rangle$. Nos gustaría demostrar que $|m_0(\xi)|^2 = 1$ pues es una propiedad que siempre resulta útil (más abajo se hará evidente porqué). Usando otra vez la identidad de Parseval (en 1.17), el lema de Riemann-Lebesgue en 1.7 y haciendo un par de cambios de variable tenemos lo siguiente:

$$\begin{aligned} \delta_{k,0} &= \langle \phi(x), \phi(x-k) \rangle = \langle \Phi(u), \Phi(u-k) \rangle \\ &= \langle \Phi(u), \Phi(u) e^{iku} \rangle = \int_{-\infty}^{\infty} |\Phi(u)|^2 e^{iku} du \\ &= \int_0^{2\pi} e^{iku} \sum_{w \in \mathbb{Z}} |\Phi(u+2\pi w)|^2 du \end{aligned}$$

Entonces para que $\delta_{k,0} \neq 0$ necesitamos que $k=0$, y como estamos

trabajando en L_2 :

$$\begin{aligned}
 1 &= \int_0^{2\pi} \sum_{w \in \mathbb{Z}} |\Phi(u + 2\pi w)|^2 du \\
 1 &= \sum_{w \in \mathbb{Z}} \int_0^{2\pi} |\Phi(u + 2\pi w)|^2 du \\
 1 &= 2\pi \sum_{w \in \mathbb{Z}} |\Phi(u + 2\pi w)|^2 \\
 \frac{1}{2\pi} &= \sum_{w \in \mathbb{Z}} |\Phi(u + 2\pi w)|^2
 \end{aligned} \tag{2.4}$$

Recordemos que ya habíamos visto que

$$\Phi(u) = m_0(u/2) \Phi(u/2),$$

entonces

$$\Phi(u + 2\pi w) = m_0(u/2 + \pi w) \Phi(u/2 + \pi w)$$

y si sustituimos esto en lo que hemos encontrado obtenemos lo siguiente:

$$\frac{1}{2\pi} = \sum_{w \in \mathbb{Z}} |m_0(u/2 + \pi w)|^2 |\Phi(u/2 + \pi w)|^2.$$

Ahora podemos separar esta suma en los pares e impares de w y

por la misma periodicidad de m_0 :

$$\begin{aligned}
\frac{1}{2\pi} &= \sum_{w \in \mathbb{Z}} |m_0(u/2 + \pi w)|^2 |\Phi(u/2 + \pi w)|^2 \\
\frac{1}{2\pi} &= \sum_{w \in \mathbb{Z}_{par}} |m_0(u/2 + \pi w)|^2 |\Phi(u/2 + \pi w)|^2 \\
&\quad + \sum_{w \in \mathbb{Z}_{impar}} |m_0(u/2 + \pi w)|^2 |\Phi(u/2 + \pi w)|^2 \\
\frac{1}{2\pi} &= \sum_{w \in \mathbb{Z}_{par}} |m_0(u/2)|^2 |\Phi(u/2 + \pi w)|^2 \\
&\quad + \sum_{w \in \mathbb{Z}_{impar}} |m_0(u/2 + \pi)|^2 |\Phi(u/2 + \pi w)|^2 \\
\frac{1}{2\pi} &= (|m_0(u/2)|^2 + |m_0(u/2 + \pi)|^2) \sum_{w \in \mathbb{Z}} |\Phi(u/2 + \pi w)|^2 \\
\frac{1}{2\pi} &= (|m_0(u/2)|^2 + |m_0(u/2 + \pi)|^2) \frac{1}{2\pi} \\
&\Rightarrow 1 = |m_0(u/2)|^2 + |m_0(u/2 + \pi)|^2 \quad (2.5)
\end{aligned}$$

3. Ahora pasemos a ver qué propiedades podemos obtener de las $f \in W_0$. Como sabemos W_0 es el complemento ortogonal de V_0 para “formar” V_{-1} . Entonces si $f \in W_0$ tenemos que $f \in V_{-1}$ y que $f \perp V_0$. Como $f \in V_{-1}$ y V_{-1} sabemos que tiene como base ortonormal $\phi_{-1,n}$ entonces evidentemente podemos escribir a f como combinación lineal de estas $\phi_{-1,n}$:

$$f = \sum_{n \in \mathbb{Z}} \langle f, \phi_{-1,n} \rangle \phi_{-1,n} = \sum_{n \in \mathbb{Z}} f_n \phi_{-1,n},$$

donde $f_n = \langle f, \phi_{-1,n} \rangle$. Ahora, usando la transformada de Fourier obtenemos:

$$F(u) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} f_n e^{-inu/2} \Phi(u/2) = m_f(u/2) \Phi(u/2), \quad (2.6)$$

donde $m_f(u/2) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} f_n e^{-inu}$. El estimado lector se podrá dar cuenta que en este inciso estamos haciendo cosas análogas al inciso anterior. De hecho, m_f también es periódica y está en \mathcal{L}_2 (como su análogo al inciso anterior, m_0 . Como $f \perp V_0$, f es ortogonal a cualquier elemento en V_0 , particularmente a cualquier base ortogonal de V_0 , en particular a las $\phi_{0,k}$. Así obtenemos lo siguiente (ya no tenemos una delta de Kronecker como en inciso pasado):

$$\begin{aligned}
0 &= \langle f(x), \phi(x-k) \rangle = \langle F(u), \Phi(u-k) \rangle \\
&= \langle F(u), \Phi(u) e^{iku} \rangle = \int_{-\infty}^{\infty} F(u) \Phi(u) e^{iku} du \\
&= \int_0^{2\pi} e^{iku} \sum_{w \in \mathbb{Z}} F(u+2\pi w) \Phi(u+2\pi w) du \\
&= \int_0^{2\pi} e^{iku} \sum_{w \in \mathbb{Z}} \langle F(u+2\pi w), \Phi(u+2\pi w) \rangle du \\
&= 2\pi \sum_{w \in \mathbb{Z}} F(u+2\pi w) \Phi(u+2\pi w) \\
\Rightarrow 0 &= \sum_{w \in \mathbb{Z}} F(u+2\pi w) \Phi(u+2\pi w)
\end{aligned}$$

Si volvemos a hacer lo que hicimos en el inciso pasado de agrupar la suma entre las w pares y las impares, para luego sustituir lo que encontramos en 2.5 obtenemos lo siguiente:

$$m_f(u/2) m_0(u/2) + m_f(u/2 + \pi) m_0(u/2 + \pi) = 0 \quad (2.7)$$

Pero por 2.5 tenemos que $1 = |m_0(u/2)|^2 + |m_0(u/2 + \pi)|^2$ entonces $m_0(u/2)$ y $m_0(u/2 + \pi)$ no pueden anularse en el mismo punto en un intervalo (cuya medida no es cero, es decir que no consta de solo un punto). Entonces, como ya vimos que forzosamente 2.7 se cumple, m_f y m_0 son periódicas de periodo 2π :

$$m_f(u/2) m_0(u/2) = -m_f(u/2 + \pi) m_0(u/2 + \pi)$$

Entonces de alguna manera tenemos una igualdad entre $m_f m_0$ (con signo contrario) al “movernos” π unidades. Esto implica que existe alguna función “escaladora” que relaciona a $m_f(u/2)$ con $m_0(u/2 + \pi)$ la cual vamos a denotar con $g(u/2)$ y debe de cumplir la siguiente relación:

$$m_f(u/2) = g(u/2) m_0(u/2 + \pi) \quad (2.8)$$

Porque si esto pasa (dado que m_0 y m_f son periódicas de periodo 2π) usando 2.7 tenemos que $m_f(u/2 + \pi) = g(u/2 + \pi) m_0(u/2 + 2\pi) = g(u/2 + \pi) m_0(u/2)$ y así:

$$\begin{aligned} 0 &= (g(u/2) m_0(u/2 + \pi)) m_0(u/2) \\ &\quad + (g(u/2 + \pi) m_0(u/2)) m_0(u/2 + \pi) \\ \iff 0 &= m_0(u/2) m_0(u/2 + \pi) (g(u/2) + g(u/2 + \pi)) \end{aligned} \quad (2.9)$$

Entonces por esta última igualdad tenemos que nuestra g también tiene que cumplir lo siguiente:

$$g(u/2) + g(u/2 + \pi) = 0.$$

Es decir que $g(u/2) = -g(u/2 + \pi)$, lo cual también nos da un indicio de algún tipo de periodicidad (con cambio de signo). Justo por este cambio de signo cuando pasamos en g de $u/2$ a $u/2 + \pi$ podemos escribir a g como:

$$g(u/2) = e^{iu/2} \nu(u) \quad (2.10)$$

donde ν es una función de periodo 2π , la cual es la causante de la relación entre $u/2$ y $u/2 + \pi$ en nuestra g . Mientras que el factor $e^{iu/2}$ nos da los cambios de signo. Sustituyendo esta “construcción” de g en 2.8 obtenemos lo siguiente:

$$m_f(u/2) = \left(e^{iu/2} \nu(u) \right) m_0(u/2 + \pi). \quad (2.11)$$

Si ahora recordamos la expresión que tenía la transformada de Fourier de nuestra f usando justamente m_f podemos sustituir esta última igualdad en 2.6 para obtener:

$$F(u) = e^{iu/2} m_0(u/2 + \pi) \nu(u) \Phi(u/2) \quad (2.12)$$

4. Con la última relación de igualdad obtenida en el inciso anterior (en 2.12) podemos sugerir la forma de nuestra madre ondícula (o por lo menos la forma de su transformada de Fourier). Podemos sugerir que sea de la siguiente manera sin que suene descabellado:

$$\Psi(u) = e^{iu/2} m_0(u/2 + \pi) \Phi(u/2) \quad (2.13)$$

Con esto podemos escribir 2.12 usando esta transformada de nuestra madre ondícula (en este punto solo queremos ver la posible forma de ψ , no nos preocuparemos por ahora de detalles de convergencia):

$$\begin{aligned} F(u) &= \left(\sum_{k \in \mathbb{Z}} \nu_k e^{-iku} \right) \Psi(u) \\ \Rightarrow f(x) &= \sum_{k \in \mathbb{Z}} \nu_k(x) \psi_0(x - k). \end{aligned} \quad (2.14)$$

Que es justo lo que queremos, una manera de escribir a las f como combinación de las $\psi_{0,k}(x) = \psi_0(x - k)$ para que justo estas $\psi_{n,k}(x) = \psi_n(x - k)$ sean la base ortogonal de nuestros espacios W_n . Esta forma de escribir a las ψ nos dice que nuestra base de ondículas se forma trasladando a una función ψ y para “extenderla” a otros V_n la “estiramos”, cambiamos su escala. Justamente esto es lo que queremos, unas funciones que se representen con cambios de escala y con traslaciones. Por los incisos anteriores tenemos que los m_0 y Φ son funciones en \mathcal{L}_2 estando particularmente en V_{-1} y son ortogonales a V_0 (es decir están en W_0). Por ende tenemos que esta propuesta de ψ está también en W_0 . Vamos a revisar la ortogonalidad de esta propuesta de ondícula usando la identidad de Parseval (en 1.17):

$$\begin{aligned}
\langle \psi_{0,j}(x), \psi_{0,k+j}(x-k) \rangle &= \langle \psi_{0,0}(x), \psi_{0,k}(x-k) \rangle \\
&= \langle \psi(x), \psi(x-k) \rangle \\
&= \langle \Psi(u), \Psi(u-k) \rangle
\end{aligned}$$

Pero si $k = 0$:

$$\begin{aligned}
\langle \psi_{0,0}(x), \psi_{0,0}(x) \rangle &= \int_{-\infty}^{\infty} e^{iku} |\Psi(u)|^2 du \\
&= \int_0^{2\pi} e^{iku} \left(\sum_{w \in \mathbb{Z}} |\Psi(u + 2\pi w)|^2 \right) du
\end{aligned}$$

Pero por lo que hemos hecho en los incisos pasados podemos simplificar $\sum_{w \in \mathbb{Z}} |\Psi(u + 2\pi w)|^2$:

$$\begin{aligned}
\sum_{w \in \mathbb{Z}} |\Psi(u + 2\pi w)|^2 &= \sum_{w \in \mathbb{Z}} |(m_0(u/2 + w\pi + \pi)\Phi(u/2 + w\pi))|^2 \\
&= \sum_{w \in \mathbb{Z}} |(m_0(u/2 + w\pi + \pi)|^2 |\Phi(u/2 + w\pi)|^2 \\
&= \sum_{w \in \text{impar}} |(m_0(u/2 + w\pi + \pi)|^2 |\Phi(u/2 + w\pi)|^2 \\
&\quad + \sum_{w \in \text{par}} |(m_0(u/2)|^2 |\Phi(u/2 + w\pi)|^2 \\
&= \sum_{w \in \text{impar}} |(m_0(u/2 + \pi)|^2 |\Phi(u/2 + w\pi)|^2 \\
&= \frac{1}{2\pi}
\end{aligned}$$

Entonces:

$$\langle \psi_{0,k}(x), \psi_{0,k}(x-k) \rangle = \int_0^{2\pi} e^{iku} \frac{1}{2\pi} du = 1$$

Si $j \neq 0$ entonces tenemos que $\langle \psi_{0,k}(x), \psi_{0,k+j}(x-k) \rangle = 0$ por la definición misma de

$$m_0(u/2) = \sum_{n \in \mathbb{Z}} h_n e^{-inu} = \sum_{n \in \mathbb{Z}} \langle \phi, \phi_{-1,n} \rangle e^{-inu}$$

y recordando que las ϕ son una base ortonormal. De esta manera tenemos que nuestras $\psi_{0,k}$ forman un conjunto ortonormal en W_0 :

$$\langle \psi_{0,k}(x), \psi_{0,k+j}(x) \rangle = \delta_{k,0},$$

donde $\delta_{k,0}$ es la delta de Kronecker. Ahora lo único que nos falta es verificar que estas ψ generan a todo W_0 , es decir que para cualquier $f \in W_0$ podemos escribirlo de manera única de la siguiente manera:

$$f(x) = \sum_{k \in \mathbb{Z}} \gamma_k \psi_{0,k}(x).$$

donde $\sum_{k \in \mathbb{Z}} |\gamma_k|^2 < \infty$. Lo anterior es lo mismo que escribir:

$$F(u) = \gamma(u) \Psi(u).$$

En este caso queremos que γ sea periódica con periodo 2π y que esté en \mathcal{L}_2 . Pero de 2.14 tenemos que:

$$F(u) = \left(\sum_{k \in \mathbb{Z}} \nu_k e^{-iku} \right) \Psi(u) = \nu(u) \Psi(u).$$

Que es la forma en la que nos gustaría escribir a $F(u)$. Ahora, por 2.10 tenemos que:

$$g(u/2) = e^{iu/2} \nu(u)$$

Entonces:

$$\int_0^\pi |g(u)|^2 du = \int_0^\pi |e^{iu} \nu(2u)|^2 du = \frac{1}{2} \int_0^{2\pi} |\nu(u)|^2 du.$$

Entonces por 2.11 y por 2.10:

$$\begin{aligned}
 m_f(u) &= e^{iu} \nu(2u) m_0(u + \pi) \\
 \Rightarrow |m_f(u)|^2 &= |\nu(2u) m_0(u + \pi)|^2 \\
 \Rightarrow \int_0^{2\pi} |m_f(u)|^2 du &= \int_0^{2\pi} |\nu(2u)|^2 |m_0(u + \pi)|^2 du \\
 &= \int_0^{2\pi} |g(u)|^2 |m_0(u + \pi)|^2 du \\
 &= \int_0^\pi |g(u)|^2 (|m_0(u + \pi)|^2 + |m_0(u)|^2) du \\
 &= \int_0^\pi |g(u)|^2 du = \frac{1}{2} \int_0^{2\pi} |\nu(u)|^2 du
 \end{aligned}$$

Pero queremos ver que esto no es igual a ∞ , entonces volvemos a la definición de $m_f(u)$ y recordamos que $\langle \phi, \phi \rangle = \langle \Phi, \Phi \rangle = 1$ tenemos que:

$$\begin{aligned}
 \int_0^{2\pi} |m_f(u)|^2 du &= \int_0^{2\pi} |m_f(u)|^2 |\Phi(u)|^2 du \\
 &= \int_0^{2\pi} |m_f(u) \Phi(u)|^2 du \\
 &= \int_0^{2\pi} |F(u)|^2 du \\
 &= \int_0^{2\pi} |f(x)|^2 dx < \infty
 \end{aligned}$$

por 2.6 y porque $f \in \mathcal{L}_2$. Y así tenemos que $\int_0^{2\pi} |\nu(u)| du = 2\pi \|f\|^2 < \infty$. Y así ya tenemos que, en efecto $\sum_{k \in \mathbb{Z}} |\gamma_k|^2 < \infty$.

Con todos estos incisos hemos construido unas $\psi_{0,k}$ a partir de las ϕ tales que son ondículas que forman una base de W_0 . Como estas W_m “heredan” propiedades de escala de los V_k entonces con estas $\psi_{m,k}$ tenemos una base ortonormal de ondículas para \mathcal{L}_2 y de tal manera que:

$$P_{j-1}(f) = P_j(f) + \sum_{k \in \mathbb{Z}} \langle f, \psi_{m,k} \rangle \psi_{m,k}.$$

■

Gracias a este gran teorema ahora tenemos una forma de construir ondículas tales que formen un conjunto ortonormal y generen a todo \mathcal{L}_2 . La siguiente propiedad es resultado de este gran teorema y es un buen ejemplo de porqué nos gusta trabajar con este tipo de bases.

Corolario. *Para toda $f \in \mathcal{L}_2$ tenemos que:*

$$\lim_{k \rightarrow -\infty} \|P_k(f) - f\|_2 = 0$$

Es decir, podemos aproximar a f como proyecciones ortogonales en nuestros espacios V_k y esta aproximación se irá “mejorando” cuando k tienda a $-\infty$.

Demostración. Como $f(x) \in \mathcal{L}_2$ entonces, como $\overline{\bigcup_{k \in \mathbb{Z}} V_k} = \mathcal{L}_2$ entonces $f(x) \in \overline{\bigcup_{k \in \mathbb{Z}} V_k}$. Ahora tenemos la propiedad de que $V_k \subset V_{k-1}$ y que para cada una de estas V_{k-1} existe una base ortonormal, ϕ_{k-1} . Con esto, $f(x) \in \langle \phi_k \rangle_{k \in \mathbb{Z}}$, donde $\langle \rangle$ es el span o generado por las ϕ_k . Entonces para toda ε podemos elegir un espacio V_k tal que exista una función, llamémosle g formada a partir de estas ϕ_k y que $\|f - g\|_2 < \frac{\varepsilon}{2}$. Pero justamente como esta $g \in V_k$ entonces $P(g) = g$ y así:

$$\begin{aligned} \|f - P(f)\|_2 &= \|f - P(f) - g + g\|_2 \\ &= \|f - g + P(g) - P(f)\|_2 \\ &\leq \|f - g\|_2 + \|P(f - g)\|_2 \\ &\leq \|f - g\|_2 + \|f - g\|_2 \leq \varepsilon. \end{aligned}$$

■

2.5. Transformada de ondícula discreta para 2D

Hasta ahora solo hemos mencionado a las transformadas de ondícula en una sola dimensión, pero ¿se puede hacer para funciones bidimensionales? Si se puede hacer esto para la transformada de Fourier, entonces muy posiblemente también se pueda para las

ondículas. Lo que es particular del tipo de ondículas con las que hemos estado trabajando es su separabilidad.

Para una función $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, como lo es una imagen, las filas y las columnas de la misma se tratan como señales unidimensionales por separado, como lo hicimos para la transformada de Fourier en 2 dimensiones. A partir de una madre ondícula bidimensional, ψ podemos obtener una familia de ondículas bidimensionales de la siguiente manera:

$$\psi_{s_1 s_2}(u_1, u_2) = \frac{1}{\sqrt{s_1 s_2}} \psi\left(\frac{u_1}{s_1}, \frac{u_2}{s_2}\right).$$

Por lo cual tenemos que también la podemos trasladar en t_1 y t_2 , como lo hicimos en el caso unidimensional en 2.1:

$$\psi_{a_1 a_2 t_1 t_2}(u_1, u_2) = \frac{1}{\sqrt{s_1 s_2}} \psi\left(\frac{u_1 - t_1}{s_1}, \frac{u_2 - t_2}{s_2}\right). \quad (2.15)$$

Pero podemos “separar” esta forma de escribir las ondículas si el análisis multirresolución que estamos considerando es también separable. Es decir, podemos escribir a todos sus elementos como dos partes de dos análisis multirresolución unidimensionales. Si esto pasa, entonces en particular podemos escribir de esta manera a su base, la base que hemos mencionado en el quinto punto de la definición 20. Así una ondícula bidimensional es separable si podemos a 2.15 escribir como lo siguiente:

$$\psi_{a_1 a_2 t_1 t_2}(u_1, u_2) = \frac{1}{\sqrt{a_1}} \psi\left(\frac{u_1 - t_1}{a_1}\right) \frac{1}{\sqrt{a_2}} \psi\left(\frac{u_2 - t_2}{a_2}\right),$$

donde ψ es una ondícula unidimensional. Con esto, y recordando lo que vimos en la sección 2.2 de la transformación de ondícula discreta obtenemos que la transformación de ondícula discreta en dos dimensiones de f viene dada como:

$$\tilde{f}(s_1, s_2, t_1, t_2) = \frac{1}{s_1 s_2} \sum_j \sum_k f(j, k) \psi\left(\frac{j - t_1}{s_1}\right) \psi\left(\frac{k - t_2}{s_2}\right). \quad (2.16)$$

Gracias a que tenemos esta forma de escribir la transformada podemos aplicar una transformada discreta de ondícula unidimensional (descrita en la sección 2.2) a las columnas y a las filas de la imagen de manera separada. A continuación se describe brevemente el algoritmo para poder hacer dicha transformación.

1. Primero se hace una transformación horizontal, es decir cada fila de la imagen se escala (con un promedio ponderado) y se aplica la transformación de ondícula discreta. El resultado de este primer paso son dos mitades de imagen, una con los coeficientes relacionados a la escala y la otra mitad con los coeficientes de ondícula.
2. Luego se hace una transformación vertical, a cada columna de ambas mitades de imagen que nos sacó el paso número 1.

El resultado de esto es la transformada discreta de ondícula de nuestra imagen en un nivel. Si se requieren más niveles se siguen aplicando los pasos 1 y 2 hasta que ya no se pueda, recordemos que en cada paso nos quedamos con la mitad de píxeles que teníamos. La siguiente figura ilustra este proceso. En morado se resalta el primer paso, la transformación horizontal. En azul se destaca el segundo paso, la transformación vertical. Notamos que el resultado son 4 imágenes. A la parte HH se le conoce como los detalles diagonales, a la HL como los detalles horizontales y a LH los detalles verticales. Si quisiéramos hacer un segundo nivel lo haríamos extendiendo este árbol en LL. La pintura aquí transformada es el Paricutín de Gerardo Murillo (Dr. Atl).

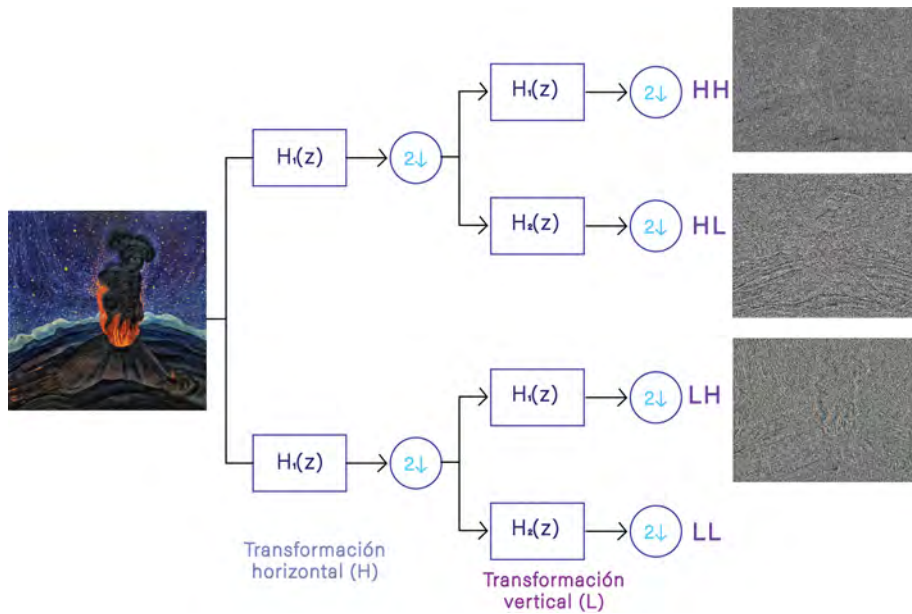


Figura 2.1. Transformación con 1 nivel

La forma usual de mostrar las transformaciones en diferentes niveles de una imagen es la siguiente:

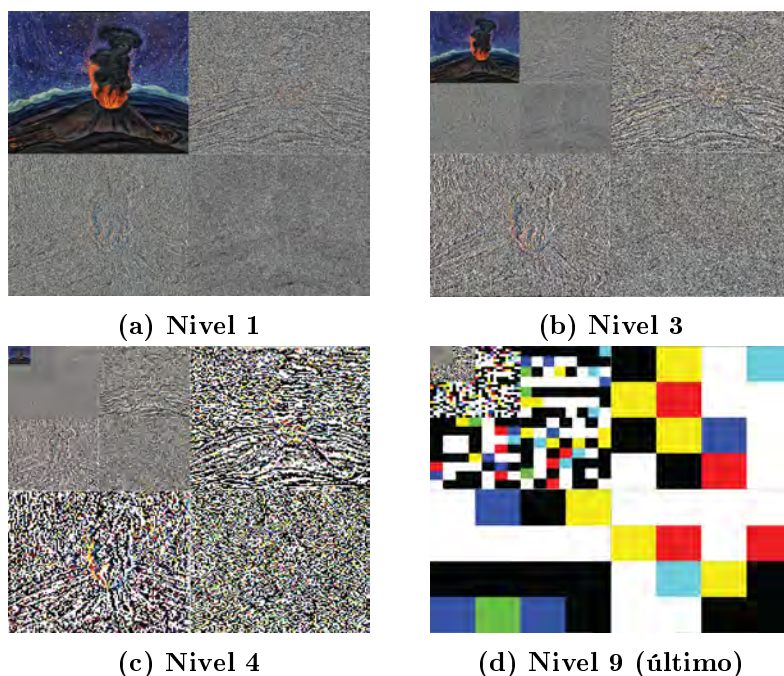


Figura 2.2. Diferentes niveles transformación de Haar

Como podemos ver, en el cuadrante izquierdo superior se van guardando los niveles previos.

2.6. Ondículas específicas

El querido lector de seguro se estará preguntando cómo son estas ondículas de las cuales hemos estado hablando. ¿Qué forma tienen para lograr ciertos fines? ¿Qué tan diferentes son entre sí? El propósito de esta sección es dar a conocer ciertas ondículas que serán importantes para nuestro propósito de identificar el estilo de pintores mexicanos.

2.6.1. Ondícula de Haar

Como mencionamos al inicio de este capítulo, la primera vez que se habló de ondículas fue en la tesis doctoral de Alfredo Haar. A pesar de

esto, él solo trabajó con un tipo de ondículas, las ondículas de Haar que aquí vamos a presentar. La primera diferencia notable entre las funciones usadas en la transformada de Fourier y las ondículas de Haar es que las funciones trigonométricas son continuas y diferenciables en todo su dominio (de hecho están en \mathcal{C}^∞). Las ondículas de Haar son funciones discontinuas y son escalonadas. Haar no usó la noción de frecuencia, como se hace en el análisis de Fourier, sino que usó escala y localización, como ya lo hemos estado manejando en este capítulo. Para hablar de escala y localización específicos de esta ondícula, vamos a definir lo que es un intervalo diádico.

Definición 24. *Para los enteros $j, k \in \mathbb{Z}$, un intervalo diádico es de la siguiente forma:*

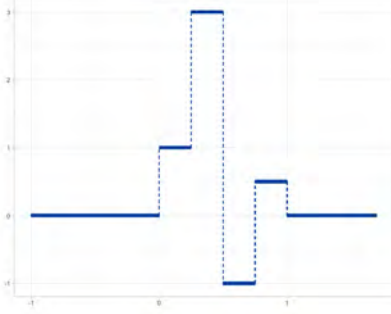
$$I_{j,k} = [2^{-j}k, 2^{-j}(k+1)].$$

Si el querido lector nota alguna similitud con los espacios V_k de la sección pasada, no es coincidencia alguna. De hecho, los intervalos diádicos cumplen con una y solo una de las siguientes características si $j_0 \neq j_1$ ó $k_0 \neq k_1$:

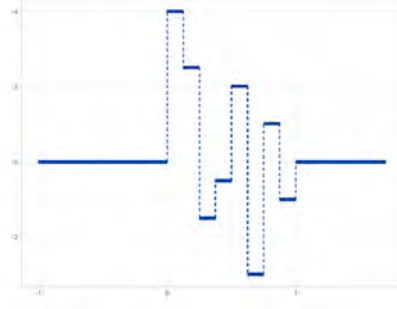
- $I_{j_1,k_1} \cap I_{j_0,k_0} = \emptyset$
- $I_{j_1,k_1} \subset I_{j_0,k_0}$ ó
- $I_{j_0,k_0} \subset I_{j_1,k_1}$

Definición 25. *Una función diádica es precisamente aquella función escalonada $f(x)$ tal que para algún $j \in \mathbb{Z}$, $f(x)$ es constante en todos los intervalos diádicos $I_{j,k}$, con $k \in \mathbb{Z}$. En este caso, f es una función escalonada diádica de escala j .*

Esta definición puede sonar algo complicada, pero simplemente describe a una función escalonada cuyos escalones están en intervalos de longitud 2^{-j} . Algunos ejemplos de esto:



(a) Función escalonada
diádica de escala 2



(b) Función escalonada
diádica de escala 3

Si el querido lector está familiarizado con teoría de la medida, puede notar que este tipo de funciones son las llamadas simples, es decir, toman una cantidad finita de valores a lo largo de su dominio.

Para definir el sistema de funciones escaladoras de Haar necesitamos definir la función indicadora.

Definición 26 (Función indicadora). *La función indicadora se define como:*

$$\chi_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Con esto ya podemos definir las funciones escaladoras de Haar.

Definición 27. Sea $p(x) = \chi_{[0,1)}(x)$ definimos para cada $j, k \in \mathbb{Z}$:

$$p_{j,k} = 2^{j/2} p(2^j x - k) = D_{2^j} T_k p(x)$$

Donde D_{2^j} se refiere a la escala y T_k a la traslación. A este conjunto $\{p_{j,k}\}_{j,k \in \mathbb{Z}}$ se le llama el sistema de funciones escaladoras de Haar.

De nuevo, la similitud con lo que hemos estado comentando de funciones escalonadas no es una coincidencia. Ahora solo nos falta una definición más para poder llegar a lo que estamos buscando: el sistema de ondículas de Haar.

Definición 28. Sea $h(x) = \chi_{[0,1/2)}(x) - \chi_{[1/2,1)}(x)$ y para cada $j, k \in \mathbb{Z}$ definimos:

$$h_{j,k} = 2^{j/2} h(2^j x - k) = D_{2^j} T_k h(x)$$

Al conjunto $\{h_{j,k}\}_{j,k \in \mathbb{Z}}$ se le conoce como el sistema de Haar en \mathbb{R} .

Por las propiedades que mencionamos de los intervalos diádicos y lo que mencionamos en la sección pasada de análisis multirresolución, el lector puede darse cuenta que el sistema de Haar forma un sistema ortonormal en \mathbb{R} . Ahora tenemos todo lo necesario para definir a las bases de Haar dentro del intervalo $[0, 1]$, aunque se puede generalizar para cualquier intervalo cerrado $[a, b]$.

Definición 29 (Bases de Haar). Para cualquier entero $j \geq 0$, la escala j en el sistema de Haar en el intervalo $[0, 1]$ es el conjunto:

$$\{p_{j,k}(x) : 0 \leq k \leq 2^j - 1\} \cup \{h_{j,k}(x) : 0 \leq k \leq 2^j - 1\}.$$

Cuando $j = 0$ entonces este conjunto se llama el sistema de Haar en el intervalo $[0, 1]$.

De nuevo consideramos que es útil ilustrar algunas funciones de Haar $h_{j,k}$ para que el querido lector pueda interpretar este sistema más fácilmente.

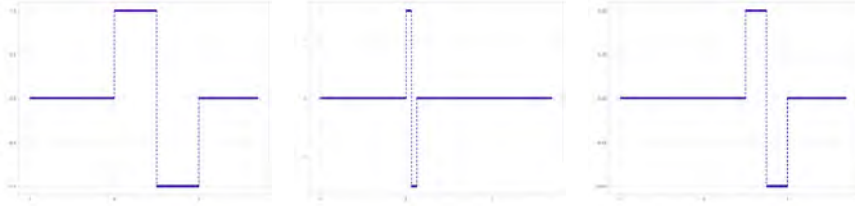


Figura 2.4. Ejemplos de funciones de Haar

Como podemos ver, una de las características esenciales de estas funciones es su simetría y que sólo son diferentes a cero en un subintervalo de $[0, 1]$. Por los teoremas que desarrollamos en la sección pasada sabemos que podemos aproximar cualquier función f que sea continua y esté en el intervalo $[0, 1]$ con este tipo de funciones. Como tenemos dos tipos de funciones en este sistema, las h y las p entonces la

transformada discreta de ondícula de Haar tiene que involucrar ambos tipos. Primero, notemos que por las propiedades del sistema de Haar, cualquier función f en el intervalo $[0, 1]$ se puede representar como :

$$f(x) = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} \langle f, h_{j,k} \rangle h_{j,k}(x) + \sum_{k=0}^{2^j-1} \langle f, p_{j,k} \rangle p_{j,k}(x)$$

Pero queremos tener sumas finitas y una manera fácil para hacer nuestra transformada. Entonces definimos unos coeficientes para facilitar el cálculo de los productos interiores usados arriba como:

$$c_j(k) = \langle f, p_{N-j,k} \rangle \quad d_k(k) = \langle f, h_{N-j,k} \rangle$$

Ahora, recordando la manera en la que definimos las p y las h tenemos una manera iterativa de calcular estos productos interiores si se sabe algún valor, pues:

$$\begin{aligned} p_{w,k}(x) &= \frac{1}{\sqrt{2}} (p_{w+1,2k}(x) + p_{w+1,2k+1}(x)) \\ h_{w,k}(x) &= \frac{1}{\sqrt{2}} (p_{w+1,2k}(x) - p_{w+1,2k+1}(x)) \end{aligned}$$

Y, dada una sucesión finita de valores a transformar, obtenemos nuestros coeficientes $c_j(k)$ y $d_j(k)$ como:

$$\begin{aligned} c_j(k) &= \frac{1}{\sqrt{2}} (c_{j-1}(2k) + c_{j-1}(2k+1)) \\ d_j(k) &= \frac{1}{\sqrt{2}} (c_{j-1}(2k) - c_{j-1}(2k+1)) \end{aligned}$$

A pesar de esto, no podemos deducir muchas propiedades a raíz de esta definición, por lo cual preferiremos trabajar de manera matricial, como en el caso de la transformada discreta de Fourier. Notemos que a partir de estos coeficientes podemos definir las siguientes matrices normalizadas, ambas de dimensiones $(n/2) \times (n)$:

$$H_n = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \end{bmatrix}$$

$$G_n = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

Usualmente a la matriz H_n se le llama la matriz de aproximación y a G_n la matriz de detalles. Con estas matrices ya podemos definir la transformada de ondícula de Haar de una manera más “sencilla”.

Definición 30 (Transformada discreta de ondícula de Haar). *La matriz de Haar queda definida usando las dos matrices descritas anteriormente como:*

$$W_n = \begin{bmatrix} H_n \\ G_n \end{bmatrix}$$

Para una muestra de tamaño n de la función f , digamos $f_n(x)$ la transformada discreta de Haar queda definida como:

$$y_n = W_n f_n$$

Notemos que la matriz de Haar, W_n es una matriz ortonormal, por ende su inversa es simplemente su transpuesta, $W_n^{-1} = W_n^T$. Gracias a esta gran cualidad es fácil asegurar que la transformación discreta inversa de Haar está dada como $W_n^T y_n = f_n$. Por estos cálculos rápidos y sencillos la transformada de Haar ha tenido una gran gama de aplicaciones.

Claramente para poder hacer transformaciones de Haar para imágenes tenemos que transformar una matriz, no un vector. Para esto nos va a servir la distinción de las matrices H_n y G_n . Para poder transformar

una imagen necesitamos su representación matricial, digamos $X \in \mathcal{M}_{n \times n}(\mathbb{R})$.

Si $X_{(i)}$ son los renglones de X , $X_{(i)} \in \mathcal{M}_{1 \times n}(\mathbb{R})$ y $X^{(i)}$ son los renglones de X , $X^{(i)} \in \mathcal{M}_{n \times 1}(\mathbb{R})$ entonces X tiene la siguiente estructura:

$$X = \begin{bmatrix} X_{(1)} \\ \vdots \\ X_{(n)} \end{bmatrix} = [X^{(1)} \quad \dots \quad X^{(n)}]$$

Ahora debemos definir las transformaciones por columnas y por renglones usando a las matrices H_n y G_n .

$$\begin{aligned} H^{ren} X &= \begin{bmatrix} HX_{(1)} \\ \vdots \\ HX_{(n)} \end{bmatrix} \\ G^{ren} X &= \begin{bmatrix} GX_{(1)} \\ \vdots \\ GX_{(n)} \end{bmatrix} \\ H^{col} X &= [HX^{(1)} \quad \dots \quad HX^{(n)}] \\ G^{col} X &= [GX^{(1)} \quad \dots \quad GX^{(n)}] \end{aligned}$$

Con esto tenemos que $H^{ren} X \in \mathcal{M}_{n \times n/2}$, $H^{col} X \in \mathcal{M}_{n/2 \times n}$, $G^{ren} X \in \mathcal{M}_{n \times n/2}$ y $G^{col} X \in \mathcal{M}_{n/2 \times n}$.

Definición 31 (Transformación de ondícula de Haar en dos dimensiones). *Sea $j \in \mathbb{N}$, la transformación de ondícula de Haar de la matriz $X \in \mathcal{M}_{2^j}$ es el conjunto de matrices X_k, D_k^1, D_k^2, D_k^3 que se definen como:*

$$\begin{aligned} X_k &= H^{col} H^{ren} X_{k-1} \\ D_k^1 &= G^{col} H^{ren} X_{k-1} \\ D_k^2 &= H^{col} G^{ren} X_{k-1} \\ D_k^3 &= G^{col} G^{ren} X_{k-1} \end{aligned}$$

con $k = 1, \dots, j$ y $X_0 = X$.

Es decir, la manera de aplicar la transformación de Haar a matrices es bastante sencilla pues se hacen por columnas y por renglones por separado. Esto también es uno de los grandes atractivos de esta ondícula.

Usualmente a esta transformación se le identifica con un árbol, como el de la siguiente imagen. En particular esta imagen muestra la transformada de Haar de una función univariada en 3 niveles.

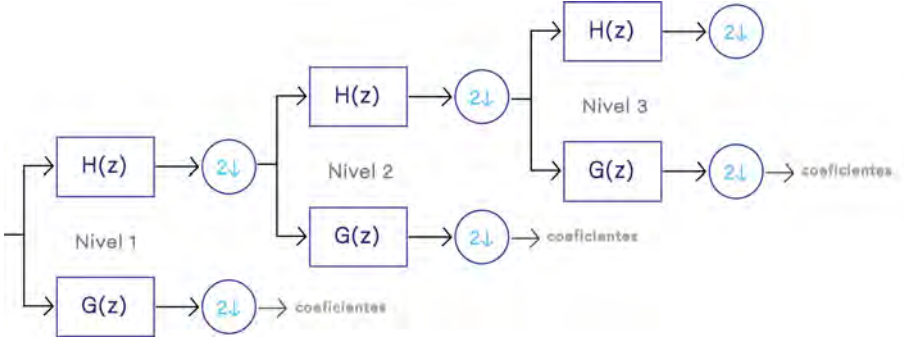


Figura 2.5. Árbol para una transformada de ondícula discreta

Como ya vimos, con este tipo de ondículas es fácil hacer su representación con un árbol como en la figura 2.1. De hecho, dicha figura muestra la transformación del *Paricutín* de Gerardo Murillo usando una ondícula de Haar en dos dimensiones. Y la figura 2.2 muestra sus diferentes niveles.

2.6.2. Ondícula de Gabor

Pensando en la transformada de Fourier de tiempo corto, una opción para nuestra g sería una función de tipo Gaussiana. En esta sección sí vamos a considerar que estamos en \mathbb{C}^2 . Vamos a pensar que x_1, x_2 siguen una distribución normal, es decir:

$$X = (x_1, x_2)^T \sim N(\mu, \Sigma), \quad \mu = (0, 0)^T, \quad \Sigma = \begin{bmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{bmatrix},$$

por ende su función masa de probabilidad es:

$$g(x_1, x_2) = \frac{1}{2\pi\sigma_{x_1}\sigma_{x_2}} \exp \left\{ -\frac{1}{2} \left(\frac{x_1^2}{\sigma_{x_1}^2} + \frac{x_2^2}{\sigma_{x_2}^2} \right) \right\}.$$

Por las propiedades de la distribución normal lo que podríamos hacer es tomar ventanas donde los tamaños estén dados por las desviaciones estándar marginales de x_1 y x_2 , es decir por σ_{x_1} y por σ_{x_2} .

De esta manera si comparamos a la transformada de Fourier, F y a la transformada de Fourier cuando le hemos “aplicado” g nos podemos dar cuenta que están asociadas vía una función, llamémosle G que también es una función de probabilidad asociada a una normal. El problema que ahora tenemos con nuestra función g es que su soporte es en todos los reales. Y justamente no queremos esto (por los problemas que ya vimos). Entonces tenemos que “cortarla” de alguna manera. Para esto nos sirven las varianzas, para poder quedarnos con la parte donde tenemos más información relevante.

De esta manera podemos definir el kernel de Gabor que es el resultado de tratar de usar una distribución normal junto con senos y cosenos como función moduladora. Lo que busca es minimizar la incertidumbre de la información que contiene dicha ondícula.

Definición 32 (Kernel de Gabor). *Dados los parámetros $\sigma, \theta, \lambda, \phi, \gamma$ definimos al kernel o filtro de Gabor como:*

$$G(x, y) = \exp \left\{ -\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right\} \exp \left\{ i \left(2\pi \frac{x'}{\lambda} + \phi \right) \right\} \quad (2.17)$$

donde:

$$\begin{aligned} x' &= (x - x_0) \cos(\theta) + (y - y_0) \sin(\theta) \\ y' &= -(x - x_0) \sin(\theta) + (y - y_0) \cos(\theta) \end{aligned}$$

El querido lector se podrá haber dado cuenta que este filtro es complejo, también existen las versiones puramente reales o imaginarias de él:

$$G_{real}(x, y) = \exp\left\{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right\} \cos\left(2\pi\frac{x'}{\lambda} + \phi\right) \quad (2.18)$$

$$G_{imag}(x, y) = \exp\left\{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right\} \sin\left(2\pi\frac{x'}{\lambda} + \phi\right) \quad (2.19)$$

Lo que hicimos con la transformada de Fourier es considerar a nuestra función f como una superposición de ondas de cosenos y senos. Cada uno de los valores que obtuvimos al hacer la transformada nos indicaba la “intensidad” de esta onda en esa posición. Pero solo queremos seleccionar algunas ondas con cierta frecuencia y cierta orientación, por ejemplo, para poder clasificar ciertas texturas o patrones. Para esto introducimos a la transformada de Fourier de tiempo corto. Pero no especificamos nada de la g a usar hasta ahora.

La transformada con el filtro de Gabor es un filtro de paso de banda, como los hemos definido anteriormente. Nos sirve para “cortar” la transformada de Fourier y aislar solo cierto tipo de información. Cada elemento de salida de esta transformada es un valor complejo (es decir consta de una parte imaginaria y de una real). Como nos dimos cuenta el kernel o filtro de Gabor consta de varios parámetros, los cuales nos sirven para filtrar la información que nos interese:

- λ es el largo de la onda en el cual nos queremos enfocar
- θ es la orientación de la distribución normal respecto a los ejes
- ϕ es el desfase para la onda
- σ es la desviación estándar de la normal en la que nos estamos basando
- γ es la razón para el aspecto del soporte de la elipse, la cual va a ser el soporte para el filtro de Gabor

El querido lector se podrá dar cuenta que, dado que tenemos diferentes orientaciones θ posibles para este filtro, este tipo de funciones no forman una base ortonormal, a diferencia de las ondículas de las cuales hemos

estado hablando. Como λ es el largo de la onda, entonces tenemos que $\frac{1}{\lambda}$ es la frecuencia de dicha onda. Los parámetros γ y ϕ nos sirven para poder definir a las ondículas que ya vimos, las cuales son de la forma:

$$\psi_{s,t} = \psi_s(u - t) = \frac{1}{|s|^p} \psi\left(\frac{u - t}{s}\right).$$

Se pueden diseñar ondículas de Gabor eligiendo las rotaciones y dilataciones adecuadas, pero esto es muy costoso computacionalmente. En la práctica lo que se hace es elegir varias escalas y rotaciones para así formar un “banco” de filtros de Gabor. Con este filtro se pueden analizar diferentes orientaciones y escalas en las cuales la información se podría acumular.

2.6.3. Ondículas complejas de árbol dual (DTCWT)

A pesar de que con la transformada de ondículas ya tenemos una buena localización en el dominio de espacio-tiempo, esta transformada no soluciona todos nuestros posibles problemas. Estos problemas son:

- Oscilan: las ondículas son funciones de pase de banda, tienden a oscilar alrededor del cero. Esto complica el procesamiento de ondículas. Si la función a transformar tiene alguna raíz, es posible que la ondícula que está superpuesta a esta parte tenga coeficiente de cero.
- *Aliasing* o solapamiento: cuando dos funciones distintas se vuelvan indistinguibles cuando se muestrean. Así la función original no puede ser reconstruida de manera única. Esto se puede deber a que cualquier coeficiente de ondícula podría ser muy importante y altera el balance que hay entre la transformada y la transformada inversa, causando problemas.
- Falta de direccionalidad: lo que hemos estado construyendo con ondículas ortonormales es un patrón de tipo tablero, a diferencia de Fourier. Los senos y cosenos en dimensiones mayores de la transformada de Fourier corresponden a diferentes direcciones. Por esta característica de ondículas ortonormales es difícil

distinguir objetos geométricos, nos faltan direcciones para distinguir sus orillas.

Por ende, si para la transformada de Fourier tomamos como kernel una combinación de senos y cosenos, solo que la parte correspondiente a los senos estaba multiplicada por la unidad compleja:

$$e^{iku} = \cos(ku) + i \operatorname{sen}(ku).$$

Tal vez podríamos pensar en una ondícula que se expresará de la misma manera, con parte real y parte compleja:

$$\psi_c(u) = \psi_r(u) + i\psi_{im}(u). \quad (2.20)$$

A raíz de esto, podemos usar ψ_c en lugar de ψ en todo lo que hemos estado haciendo en las secciones 2.1 y 2.2 de este trabajo. Así obtenemos la familia de ondículas como previamente lo hicimos:

$$\psi_{s,c}(u) = |s|^{-p} \psi_c\left(\frac{u}{s}\right)$$

Como lo hicimos en la sección 2.4 o con las ondículas de Haar para conseguir los coeficientes de ondículas lo que debemos hacer es tomar el producto interior de nuestra función f con $2^{j/2}\psi_c(2^ju - k)$:

$$\begin{aligned} d_r(j, k) &= \langle f, 2^{j/2}\psi_r(2^ju - k) \rangle \\ d_{im}(j, k) &= \langle f, 2^{j/2}\psi_{im}(2^ju - k) \rangle \\ d_c(j, k) &= d_r(j, k) + d_{im}(j, k) \end{aligned} \quad (2.21)$$

Para diseñar ondículas complejas hay dos acercamientos: tratar de formar una base ortonormal para todas las ψ_c o que solo cada ψ_r y ψ_{im} sean ortogonales entre sí. Este segundo acercamiento es el que nos gustaría hacer porque, a pesar de tener un marco ajustado con redundancia (no todas las ψ_c necesariamente son ortonormales) vamos a ganar ciertas propiedades. ¿Qué podemos ganar con esto? Las transformadas usando ondículas ortogonales nos permiten analizar una imagen en dos direcciones ortogonales. Por ejemplo, en la de Haar lo que hicimos fue obtener transformaciones en columnas y en filas, con

los HH, HL y LH como en la figura 2.1. Tener redundancia nos va a permitir analizar las obras en más direcciones (usualmente se analizan en 6 direcciones diferentes), permitiéndonos obtener un mejor análisis de los detalles en dicha obra.

La transformada de ondícula compleja de árbol dual usa dos árboles, es decir dos tipos de bases. Ambas bases se diseñan de tal manera que resaltan detalles específicos y distintos. Esta transformada tiene varios aspectos que también tiene la transformada de Fourier: suave, no oscila tanto, es casi invariante a traslaciones, reduce *aliasing* y podemos explotar varias direcciones con ella.

Cada árbol contiene filtros reales, pero los dos árboles en conjunto producen las partes reales e imaginarias respectivamente de los coeficientes complejos de ondículas. Gran parte de esta teoría y de las ondículas complejas de árbol dual fue desarrollada por Nick Kingsbury, actual profesor de la universidad de Cambridge. A los filtros de los árboles les llama filtros de q desplazamientos, donde q es el periodo muestral. Algunas propiedades de dichos filtros son:

- Hay un retraso de tamaño igual a medio periodo de muestreo. Esto es porque los filtros tienen retraso de $\frac{1}{4}$ y $\frac{3}{4}$ del periodo de muestreo.
- El punto anterior se consigue con un filtro $H(z)$ asimétrico de tamaño par y con su análogo en tiempo inverso o reverso, $H(z^{-1})$.
- Gracias a la asimetría del punto anterior, los filtros se pueden diseñar de tal manera que se tenga una reconstrucción perfecta ortonormal a partir de su transformada de ondícula. Esto es una característica muy atractiva, la misma que la de la transformada de Fourier.
- Ambos árboles están diseñados de tal manera que sus filtros de reconstrucción son el reverso de sus filtros de análisis (en el sentido anteriormente mencionado). Con esta propiedad tenemos que todos los filtros vienen del mismo conjunto ortonormal.

Ahora, para cada uno de los árboles lo que se requiere es tener filtros de pase de banda bajos que estén retrasados uno del otro por medio periodo de muestreo. Una característica derivada de esto es que vamos a tener poco *aliasing* y por ende una buena invarianza de desplazamiento. Como las bases son ortonormales entonces estamos en presencia de una transformación que usa marcos ajustados, los cuales han sido mencionados anteriormente en este trabajo. La figura a continuación es una forma de ver gráficamente a este tipo de transformada, como lo hemos hecho con las otras transformadas. En morado/rosa se muestra el árbol A y en azul/verde el árbol B. Los números que se encuentran entre paréntesis (como 1, 0, 3q y q) son los retrasos que se consideran en cada caso. Notemos que para el primer nivel son diferentes que para el resto del árbol.

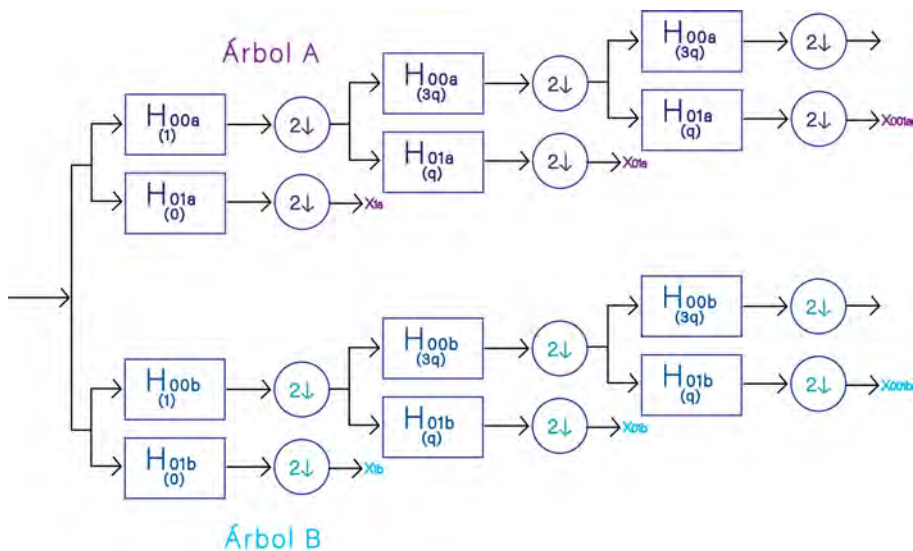


Figura 2.6. DT \mathbb{C} WT

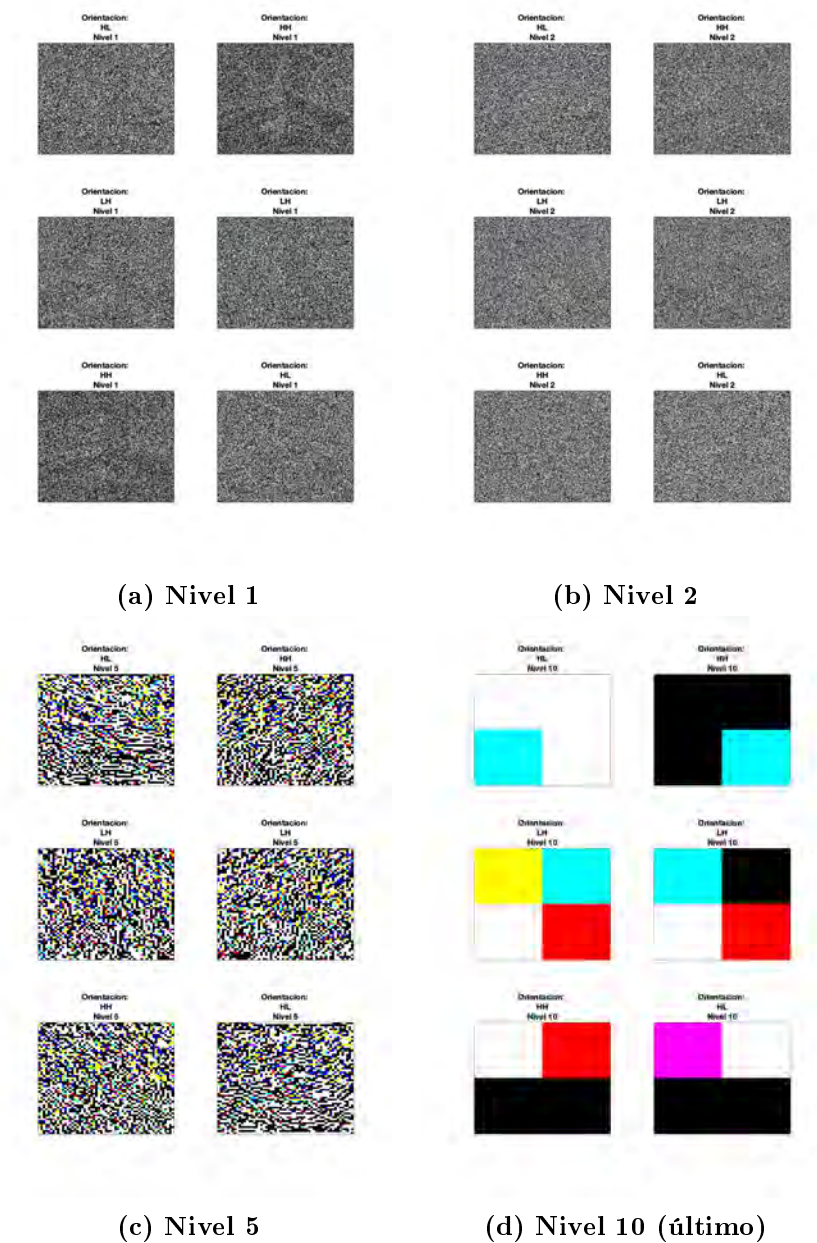


Figura 2.7. Diferentes niveles transformación DTCWT

Aplicando la transformada de ondícula compleja de árbol dual de q-traslaciones de Kingsbury a la obra del *Paricutín* del Dr. Atl obtenemos la figura 2.7 (a diferentes niveles, caso análogo a la figura 2.2). Entonces por cada nivel tenemos 6 direcciones distintas en las cuales la transformación es hecha. Gracias a estas diferentes direcciones (redundancia) podemos obtener más información de la imagen y baja sensibilidad a traslaciones.

A diferencia de las transformaciones que hemos mencionado hasta ahora, los filtros empleados en la transformada de ondícula compleja de árbol dual se consiguen solucionando problemas de optimización. Dichos problemas y el porqué de su implementación los podemos encontrar en el trabajo de Kingsbury. Dado que el objetivo de este trabajo no es describir y probar diferentes problemas de optimización y por brevedad se omite detallarlos, más se pueden consultar en [10].

Capítulo 3

Métodos computacionales

Ya tenemos la teoría de análisis de Fourier y de ondículas que vamos a requerir para este trabajo. Lo que nos falta es describir cómo las vamos a usar para poder analizar el estilo de los pintores mexicanos elegidos. Dado que este trabajo no se centra en estadística o “machine learning” (o en su caso, “deep learning”), en este capítulo pasaremos a describir brevemente los métodos que vamos a emplear. Si el curioso lector desea saber más sobre éstos, los puede consultar en los libros de Aggarwal, Zhang o Izenman en [11, 12, 13].

Queremos empezar mencionando que cada uno de estos métodos tiene un sustento teórico computacional, pero también neurobiológico. Dicho sustento o interpretación se irá mencionando para cada uno de los métodos. Si el querido lector desea saber más sobre cómo nuestro cerebro procesa estímulos visuales le recomendamos leer el capítulo Charles D. Gilbert en “Introduction to Neural Science” en [8]. Debemos de aclarar que es imposible hacer un análisis de estilo 100 % pues hay muchos factores a considerar:

- El análisis de obras de arte tiene cierto carácter subjetivo, para cada uno el estilo de cierto artista tiene cualidades diferentes.
- Intentar modelar la forma en la que el cerebro humano interpreta obras de arte también es muy complejo dado que el cerebro humano mismo es sumamente complejo.

3.1. Métodos de obtención de datos a partir de imágenes

Nosotros vamos a partir de imágenes, de las obras de los diferentes artistas que estamos considerando: Gerardo Murillo (Dr. Atl), Leonora Carrington, Frida Kahlo, María Izquierdo, Carlos Mérida, José Clemente Orozco, Alfredo Ramos Martínez, Diego Rivera, Rufino Tamayo y Remedios Varo. Necesitamos métodos para obtener información de estas imágenes y a partir de dicha información, también conocida como descriptores de imágenes, poder hacer nuestra clasificación. Lo que queremos lograr es que la computadora aprenda los estilos de los artistas con muestras de estos descriptores para luego clasificarlas dependiendo de quién las pintó. En esta primera sección del capítulo daremos una breve introducción a los métodos que usamos para la obtención de dichos descriptores.

Antes de detallar los métodos computacionales que usaremos, es importante destacar que una imagen en blanco y negro (o escala de grises) es una matriz cuyas entradas definen la intensidad de cada uno de los píxeles de esta imagen. Los píxeles de las imágenes a color se forman a partir de los colores primarios (rojo, azul y verde). Un canal para las imágenes a color es una matriz, como las matrices que se usan para describir imágenes en blanco y negro, pero esta matriz simboliza uno de los colores primarios. Hay diferentes maneras de describir estos canales, diferentes escalas que se pueden usar, pero en este trabajo vamos a usar exclusivamente el RGB, por sus siglas en inglés *Red Green Blue*. Para este tipo de modelo de canales, cada matriz simboliza los colores primarios rojo, verde y azul. Este modelo trata de parecerse a la manera en la que el ojo humano interpreta colores.

3.1.1. Modelos de Markov ocultos

Cadenas de Markov

El querido lector probablemente esté familiarizado con las cadenas de Markov. Si no lo está, primero daremos una breve explicación de éstas

para después pasar a la descripción de los modelos de Markov ocultos, como lo sugieren Choi et al en [14].

Una cadena de Markov es un modelo estocástico. Es una secuencia de variables aleatorias X_0, X_1, \dots, X_n tal que cumple la propiedad de Markov, la cual es:

$$\begin{aligned} P(X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_0 = x_0) \\ = P(X_t = x_t | X_{t-1} = x_{t-1}), \end{aligned}$$

donde $P(X_t = x_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots, X_0 = x_0)$ es la probabilidad de que en el tiempo t se esté en el estado x_t dado que en los tiempos $t-1, t-2, \dots, 0$ se estuvo en los estados $x_{t-1}, x_{t-2}, \dots, x_0$ respectivamente y $P(X_t = x_t | X_{t-1} = x_{t-1})$ es la probabilidad de estar en el estado x_t en el tiempo t dado que se estuvo en el estado x_{t-1} en el tiempo anterior, es decir en $t-1$. El dominio de las variables aleatorias se llama espacio de estados. Esto nos dice que nuestro modelo solo tiene memoria del evento anterior. La probabilidad de que suceda tal evento solo está influida por el evento que sucedió justo antes. La probabilidad de ir de un estado (i) a otro (j) se expresa de la siguiente manera:

$$p_{i,j} = P(X_t = j | X_{t-1} = i).$$

Gracias a esta probabilidad podemos construir una matriz cuadrada donde en la entrada i, j se escribe la probabilidad de ir del estado i al estado j . Dicha matriz es llamada matriz de transición. También necesitamos una distribución inicial, usualmente denotada $\pi(x) = P(X_0 = x)$. Podemos hacer un grafo a partir de una cadena de Markov, donde los nodos son los estados y los pesos de los arcos son las probabilidades de pasar de un estado a otro (los $p_{i,j}$). Un ejemplo se muestra en la figura 3.1:

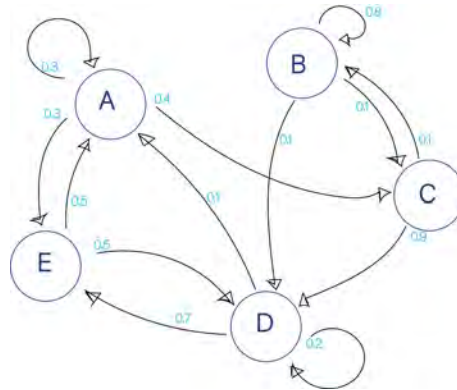


Figura 3.1. Ejemplo cadena de Markov

Cadenas de Markov ocultas

La diferencia de las cadenas de Markov con las cadenas de Markov ocultas es que en las últimas no sabemos en cuál estado estamos (está oculto), solo vemos realizaciones de éste. Al analizar un modelo de Markov oculto (sea una cadena o un árbol), nos podemos plantear las siguientes preguntas:

- ¿Cuál es la secuencia de los estados dados los datos observados?
- ¿Cómo se pueden estimar las probabilidades de transición y emisiones del modelo?
- ¿Cuál es la probabilidad de avance?
- ¿Cuál es la probabilidad posterior?

Por ejemplo, una cadena de Markov puede relacionarse a elegir diferentes frascos conteniendo diferentes proporciones de canicas de colores. En este caso sabemos en qué frasco estamos en cada momento. Si lo viéramos como una cadena de Markov oculta, como en la figura 3.2, no sabríamos en qué frasco estamos, solo sabríamos las canicas que sacamos en cada “turno” y las proporciones de los diferentes colores de canicas en cada frasco. Lo que buscaríamos haciendo un análisis en este ejemplo es encontrar la secuencia de los estados dadas las canicas que

sí podemos ver. El ejemplo de las canicas y los frascos lo podemos ver como sigue (solo que en este caso desconocemos la estructura de la figura 3.1 o de la distribución inicial $\pi(x)$):

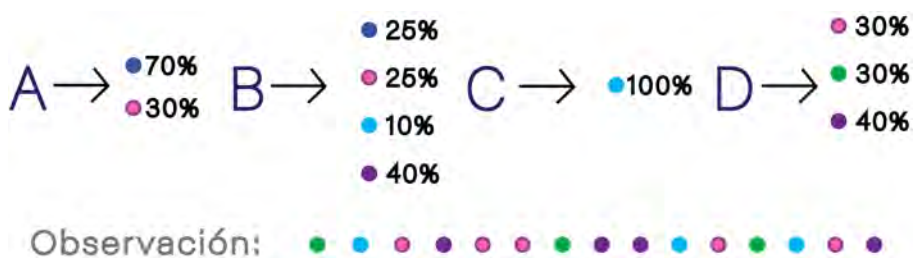


Figura 3.2. Ejemplo cadena oculta de Markov

Campos aleatorios de Markov

Lo que vamos a emplear en este trabajo son campos aleatorios ocultos de Markov. Esto se debe a que las imágenes que vamos a usar no son configuraciones unidimensionales (como las cadenas de Markov) y que, además, queremos distinguir entre coeficientes “pequeños” de los coeficientes “grandes”. Cada imagen puede ser representada como una matriz (si es blanco y negro o escala de grises) o como un tensor (para el caso de imágenes a color). Lo importante es que cada píxel es representada como una entrada en una matriz y que ésta tiene un valor numérico. Lo importante de implementar campos aleatorios de Markov es que las imágenes (o pinturas) tienden a ser suaves en la mayoría de sus píxeles. Esto quiere decir que la mayoría de las zonas de una imagen no cambian drásticamente de color de un píxel a otro. La figura 3.3 en la página 109, ilustra la propiedad de Markov en un campo aleatorio de Markov. Otra distinción entre este modelo y las tradicionales cadenas de Markov es que cada nodo es una variable aleatoria.

Para aportar la característica de qué nodos o píxeles adjuntos usualmente se comportan similarmente, se usa la definición de clique o clan. Un clique (o clan) es un conjunto de nodos n en N tal que todo par de vértices, n_i, n_j distintos, $i \neq j$, son adyacentes (están

conectados). Un clique es máximo cuando éste no forma parte de un clique de tamaño mayor (en cuestión a la cantidad de nodos que contiene). Cuando usamos campos aleatorios de Markov para imágenes los cliques maximales dado un píxel son aquellos nodos o píxeles que se encuentren en su vecindad: el píxel de arriba, el de abajo y los dos de los lados. Hay ocasiones en los que también se consideran los píxeles en las diagonales, pero éste no será el caso en este trabajo.

La configuración de un modelo de campo aleatorio de Markov depende de los cliques máximos. Esto se debe a que a cada uno de los cliques maximales se le atribuye una función de energía. Esta función de energía sirve para evaluar los nodos en cada clique y devolvernos un escalar, el cual es “bajo” si esa combinación de valores de los nodos es “buena”. Con esto, la energía global del grafo o imagen es la suma de la energía de los cliques maximales. Esta función de energía la podemos pensar como una función auxiliar, lo que hace es calcular algún tipo de distancia entre nodos adjuntos. Como suponemos que las zonas de una imagen no cambian drásticamente entonces dicha distancia debe de ser pequeña.

Campos aleatorios ocultos de Markov con distribución Gaussiana y su algoritmo

Vamos a suponer que tenemos dos tipos de distribuciones Gaussianas observadas dependiendo si la intensidad del píxel en cuestión es grande o no. Asumir esto tiene bastante sentido pues podemos decir que en cada imagen hay distintos tipos de píxeles: los que corresponden a zonas oscuras y los que no. Pero no podemos asumir que cada uno es independiente (por lo expresado arriba). Tampoco tenemos información previa de cuáles son los parámetros $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ de dichas distribuciones normales o Gaussianas. Para esto le agregamos la característica de oculto al campo aleatorio de Markov previamente descrito. Si nos referimos a la analogía de los frascos con canicas, ahora tenemos dos frascos y cada uno de ellos nos produce observaciones de dos distribuciones normales diferentes. Ya que tampoco tenemos una cadena como en la figura 3.2, sino algo como en la figura 3.3. Los cuadrados simbolizan las distintas distribuciones que consideramos,

están ocultas. Lo que podemos ver son variables aleatorias que tienen esa distribución, que son los círculos.

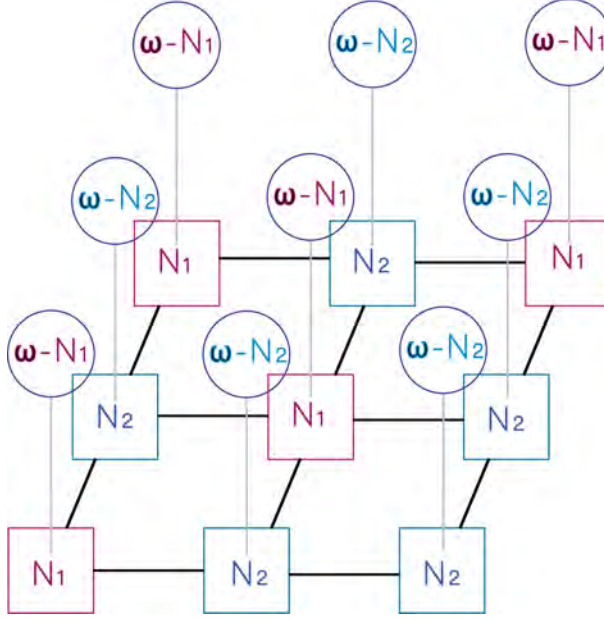


Figura 3.3. Ejemplo campo aleatorio de Markov

Ahora vamos a describir el algoritmo usado para extraer las características de los cuadros de los pintores seleccionados. Vamos a iniciar explicando cómo obtenemos la estructura estadística de una imagen asumiendo el modelo de campos aleatorios ocultos de Markov con distribución Gaussiana.

Dada una imagen $y = \{y_1, \dots, y_N\}$ con píxeles y_k buscamos inferir cuáles etiquetas $x = \{x_1, \dots, x_N\}$ podemos atribuirle. Las etiquetas x_k vienen de un conjunto, L de todas las posibles etiquetas. En nuestro caso dicho conjunto consta de dos elementos, uno para cada distribución normal considerada. Queremos poner estas etiquetas lo mejor que podamos, es decir queremos:

$$x^* = \underset{x}{\operatorname{argmax}} \{P(y|x, \Theta) P(x)\},$$

donde Θ son los posibles vectores de parámetros de las distintas etiquetas θ_{x_k} y éste es el vector de parámetros de las normales, $\theta_{x_k} = (\mu_{x_k}, \sigma_{x_k})$. La probabilidad a priori, $P(x)$ es la función de energía, en este modelo se asume una distribución de Gibbs, de la cual no vamos a hablar. De la misma manera, $P(y|x, \Theta)$ es la probabilidad conjunta de verosimilitud, es decir:

$$P(y|x, \Theta) = \prod_k P(y_k|x, \Theta) = \prod_k P(y_k|x_k, \theta_{x_k}),$$

donde $P(y_k|x_k, \theta_{x_k})$ es una distribución normal con vector de parámetros θ_{x_k} . Para calcular este vector de parámetros para ambas distribuciones normales asumidas usamos el algoritmo esperanza-maximización (o EM). Dicho algoritmo calcula los estimadores de máxima verosimilitud del vector de parámetros, el cual es un método usual para calcular parámetros en estadística. Cabe mencionar que el modelo que vamos a implementar es de índole Bayesiano. Esto se debe a que se va ajustando conforme le damos los datos (los píxeles de la imagen). Entonces los vectores de parámetros que obtenemos son los estimadores de máxima probabilidad a posteriori (o MAP). Si el querido lector quiere saber más sobre el método y el paquete computacional que vamos a usar puede ver en [15].

Entonces para extraer las características de cada pintura de cada artista seleccionado lo que hacemos es lo siguiente:

1. Leer cada pintura, escalar cada una a un tamaño de 1536×2048 píxeles. Decidimos usar este tamaño en particular por ser un tamaño en el que podíamos escalar todas las imágenes sin que perdieran su calidad pues es cercano al tamaño de la imagen más pequeña considerada en este trabajo.
2. Dividir cada pintura en áreas de 512×512 píxeles. Este tamaño lo elegimos por dos razones. La primera es que $512 = 2^9$, es una buena factorización para la transformación DTCWT. La segunda razón es porque queríamos obtener divisiones que fueran cuadrados, para la transformación DTCWT y dividirla de esta

manera nos dejaba con 12 cuadros, una buena cantidad para obtener suficiente información, pero que nuestro método no tarde demasiado tiempo.

3. A cada uno de estos cuadros aplicar la transformación DTCWT. Usamos los tres primeros niveles de dicha transformada, es decir, para cada sección de cada obra obtenemos tres transformadas con seis orientaciones diferentes a color de tamaños: 256×256 , 128×128 y 64×64 . Este es el paso que más tiempo consumió.
4. Para cada nivel calcular la norma euclidiana de cada pixel (pues para cada nivel tenemos 3 matrices, una para el canal azul, otro para el verde y otro para el rojo).
5. Para cada uno de los 6 niveles de la transformada obtener los vectores de parámetros de las distribuciones normal asociados a los diferentes estados ocultos del campo de Markov, $\theta_{1,k} = (\mu_{1,k}, \sigma_{1,k})$ y $\theta_2 = (\mu_{2,k}, \sigma_{2,k})$, esto se hace con el paquete aportado por Wang en [15].
6. El vector de características de cada subregión de 512×512 píxeles es $c_{a,b} = (\mu_{1,1}, \mu_{2,1}, \dots, \mu_{1,6}, \mu_{2,6}, \sigma_{1,1}, \sigma_{2,1}, \dots, \sigma_{1,6}, \sigma_{2,6},)$, es decir $c_{a,b} \in \mathbb{R}^{24}$ y a se refiere al número de cuadro dentro de los 12 posibles, b el número de pintura. Por ende, el descriptor de cada cuadro (de los 12 posibles) para cada pintura es de tamaño 24. En este caso, las $\mu_{1,i}, \mu_{2,i}$ se refieren a las dos medias de las distribuciones normales que estamos distinguiendo para cada dirección i , es decir 12 medias en total. Las $\sigma_{1,i}, \sigma_{2,i}$ son las dos desviaciones estándar de las distribuciones normales por cada dirección i , es decir 12 en total.
7. Para cada cuadro a utilizar algún método de clasificación para los $c_{a,b}$ de todos los artistas. En este modelo vamos a hacer la clasificación por sección, es decir, para cada imagen vamos a tener 12 votos (clasificaciones) del posible autor de esa obra, uno voto por subregión.
8. La clasificación final de cada pintura está dada como artista que

más frecuentemente fue identificado dentro de los 12 cuadros o el que más votos tenga.

3.1.2. Autoencoders

Introducción a redes neuronales

Como lo menciona Aggarwal en [11], los autoencoders son un tipo especial de redes neuronales, las cuales han tenido un gran auge en los últimos años. Se han implementado redes neuronales para clasificar imágenes, para predecir texto, para crear sistemas de recomendación, etc. Las redes neuronales son técnicas computacionales que tratan de simular lo mejor posible lo que pasa en sistemas biológicos, como lo menciona Aggarwal. En el cerebro humano, la información fluye de neurona a neurona, las cuales se encuentran conectadas por axones. Lo que tratamos de hacer con redes neuronales artificiales es recrear dicha estructura para poder procesar información. Cada axón artificial o puente tiene un peso, lo cual afecta la función que es evaluada en la neurona de entrada que conecta. Entonces los valores de entrada a nuestra red neuronal artificial se propaga de manera diferente a través de estos arcos con pesos y mediante las funciones en cada neurona hasta conseguir uno o varios valores de salida. Usualmente las redes neuronales artificiales con más de una capa son consideradas como parte de una rama de aprendizaje de máquina llamado aprendizaje profundo. Esto se debe a que podemos combinar varias capas de neuronas artificiales en una sola red neuronal. Estas técnicas se diferencian de otras, como los modelos de Markov ocultos, porque son más eficientes cuando tenemos una gran cantidad de datos a que cuando tenemos una cantidad pequeña o mediana de datos. En este contexto, la palabra “eficiencia” la asociamos a que son métodos rápidos, capaces de procesar grandes cantidades de información y han dado resultados útiles en la práctica.

La arquitectura más simple de una red neuronal empieza con un perceptrón. En 3.4a se muestra un diagrama del mismo. Esta red neuronal artificial contiene varios nodos de entrada conectados a una neurona mediante arcos con pesos diferentes. La capa de entrada no

realiza ninguna operación. La Σ es la operación de entrada, el producto punto entre el vector de entrada x y el vector de pesos w , mientras que la Υ es la función de activación no lineal, es decir la función que transforma al vector de entrada ponderado por los pesos en los arcos. Esta red neuronal se puede entrenar (es decir, optimizar sus pesos) con un conjunto de datos para obtener algún valor de salida en particular (clasificación, reducción de dimensión, etc). Tenemos una gran flexibilidad con las redes neuronales puesto que podemos usar diferentes funciones en las neuronas, podemos conectarlas de manera distinta y las podemos entrenar para nuestro objetivo. En la figura 3.4b ilustramos solo una neurona y una capa, pero como vamos a ver a continuación, podemos hacer redes neuronales con varias capas y con distintas formas. Cuando tenemos más de dos capas es cuando ya se considera aprendizaje profundo. Todo esto lo menciona Aggarwal en su libro en [11], por si el querido lector quisiera leer sobre esto a más profundidad.

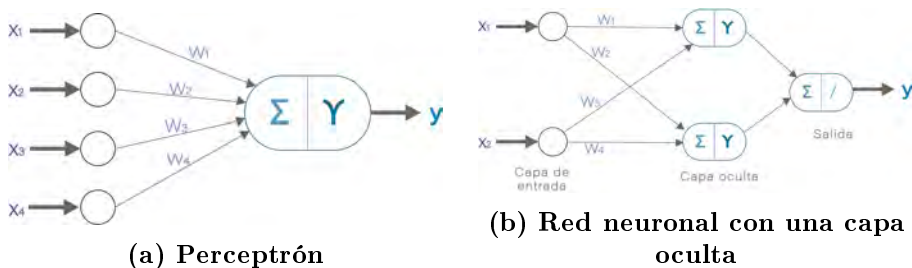


Figura 3.4. Ejemplos básicos redes neuronales

Para entrenarlas debemos de optimizar sus parámetros. En el aprendizaje supervisado conocemos el tipo de valor de salida que queremos para cada valor de entrada. Por ejemplo, si queremos clasificar artistas ya tenemos el autor de cada cuadro. Es evidente que, en nuestro contexto, ya conocemos el autor de cada uno de los cuadros que vamos a usar. El valor de salida que queremos que nos indique la red neuronal es el autor de la obra. Por ende queremos minimizar el error cometido que clasifique erróneamente y le debemos de dar varios ejemplos de los artistas que estamos considerando. La red neuronal itera sobre todas las obras de arte junto con sus etiquetas (sus autores)

y ajusta sus parámetros (los pesos en los arcos) hasta que converja o hasta que otro criterio de paro es alcanzado (por ejemplo, cantidad de iteraciones). En cualquier tipo de modelo computacional debemos de separar entre el conjunto de datos que vamos a usar para ajustar los parámetros y otro conjunto para probar este ajuste. Lo interesante de este modelo es en el conjunto que vamos a usar para probar el ajuste. ¿De cuáles cuadros podemos obtener definitivamente su autor? ¿De cuáles el autor predicho por el modelo es erróneo?

Cuando tenemos distintas capas con varias neuronas, como en la figura 3.5, optimizar la función objetivo y obtener los mejores parámetros posibles se vuelve una tarea complicada. Hay diversos métodos numéricos para optimizar estas funciones. Uno de los más usados se llama descenso por gradiente estocástico (en inglés, *stochastic gradient-descent*), para este trabajo usaremos SGD y vamos a tratar de minimizar el error al cuadrado de la predicción basándose en puntos de entrenamiento elegidos al azar.

Como el querido lector puede leer en ambos libros de Aggarwal citados en este trabajo en [16, 11], el proceso de obtención de parámetros vía optimización para redes neuronales de más de una capa no es un proceso trivial. El problema con esto es que la función objetivo es una composición de funciones que viene dada por los pesos de las capas anteriores. Para resolver este problema usamos un método de programación dinámica llamado propagación hacia atrás (o mejor conocido en inglés como *backpropagation*). Contiene dos fases: paso para adelante y paso para atrás. Es muy importante este método porque calcula el gradiente de la función objetivo (también conocida como función de pérdida) con respecto a los pesos de la red para cada valor de entrada. Necesitamos dicho gradiente porque SGD se basa en la aproximación de dicho gradiente para poder minimizar la función objetivo.

Estos métodos son parte de un curso de optimización numérica, lo cual se escapa del estudio de este trabajo, por ende no lo describiremos a fondo. Si el querido lector se encuentra interesado en estos temas puede

tomar como referencia a Aggarwaval o a Nocedal en [11, 17].

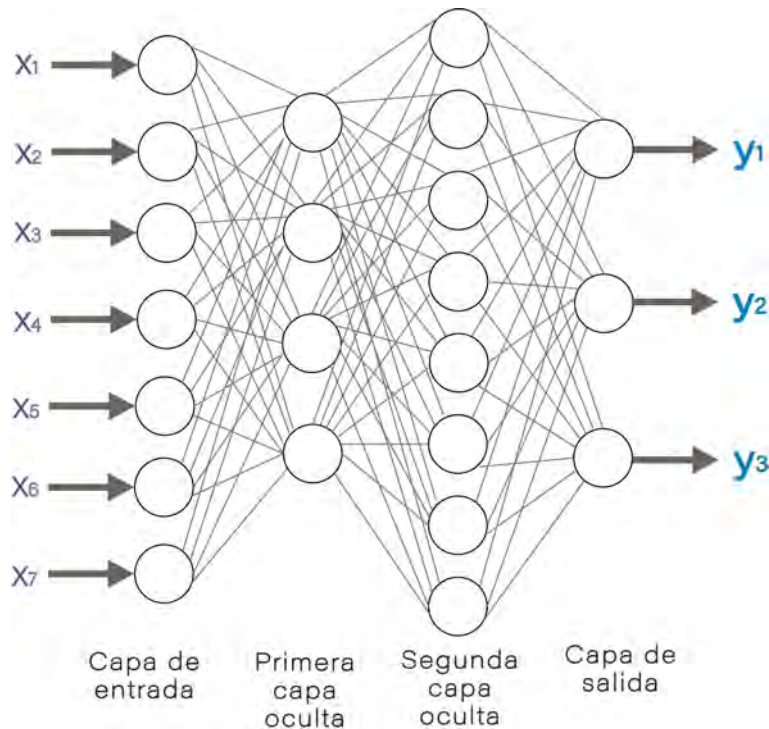


Figura 3.5. Ejemplo red neuronal

Hay varios tipos de funciones de activación Υ , dependiendo del propósito que vaya a servir esta red neuronal. En este trabajo usaremos dos funciones de activación:

- ReLU: por su nombre en inglés *Rectified Linear Unit* es una función lineal a trozos, por ende es “fácil” de entrenar a nuestra red neuronal con este tipo de funciones. Esta función es de la siguiente forma:

$$\Upsilon(v) = \max\{v, 0\}.$$



Figura 3.6. Gráfica de la función ReLU

- Sigmoid: históricamente esta función se usa para regresión logarítmica. Este tipo de regresión trata de clasificar un conjunto en dos grupos, usualmente denotados con 0 y 1. La forma de la sigmoide nos ayuda a indicar la probabilidad de pertenecer a cada uno de estos grupos. Como podemos ver en su expresión, conforme $v \rightarrow -\infty$ entonces $\Upsilon(v) \rightarrow 0$, mientras que $v \rightarrow \infty$ entonces $\Upsilon(v) \rightarrow 1$. Esta función es de la siguiente forma:

$$\Upsilon(v) = \frac{1}{1 + e^{-v}}.$$

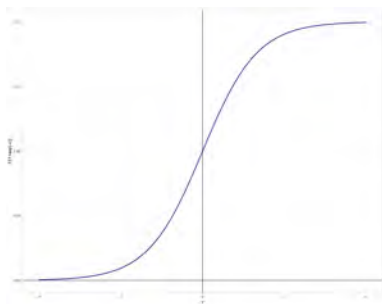


Figura 3.7. Gráfica de la función sigmoide

Autoencoders como redes neuronales artificiales

Como hemos mencionado, las redes neuronales artificiales nos proporcionan una gran flexibilidad para poder crear diferentes modelos

para propósitos variados. A pesar de que existe un tipo de red neuronal llamado convolucional y que dicho tipo de redes neuronales son aptas para clasificar imágenes, en este trabajo no las vamos a emplear. Esto se debe a que dichas redes neuronales, si se quieren entrenar desde cero, requieren de conjuntos de entrenamiento enormes y, a pesar que nosotros contamos con cerca de 900 imágenes, debemos tener cuidado con la cantidad de parámetros que queremos estimar.

Un autoencoder es una red neuronal con una forma muy peculiar, como se muestra en la figura 3.8. Como su nombre en inglés lo indica, el objetivo de esta red neuronal es cifrar una imagen para luego descifrarla, queremos obtener una reducción de dimensión con este modelo. Las primeras capas del modelo van reduciendo de dimensión, a esto se le llama el codificador. La capa de la mitad es lo que vamos a usar como descriptor de imagen, es el cuello de botella y la representación de menor dimensión de la imagen, la representación latente. La última parte se le llama el decodificador y trata de reconstruir la imagen original a partir del cuello de botella. La función objetivo por ende es la diferencia entre la imagen original y la reconstrucción, queremos que sean lo más parecidas posibles. Como la parte de la mitad del modelo tiene menos unidades que los dos extremos, este modelo es de compresión con pérdida.

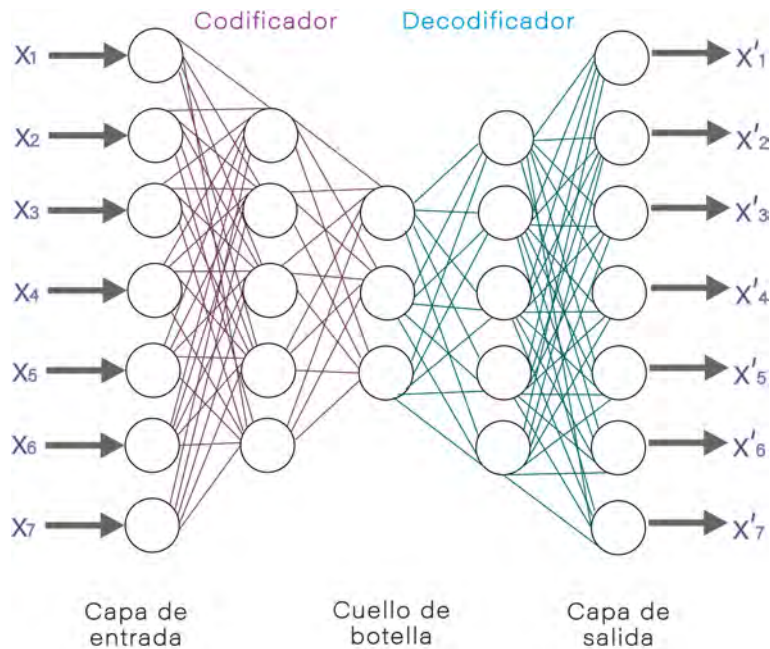


Figura 3.8. Ejemplo de un autoencoder

Vamos a usar dos variaciones de métodos que usan autoencoders para hacer nuestra clasificación de obras de arte. Vamos a contrastar el uso de autoencoders por sí solos y usando filtros de Gabor en las imágenes antes de pasarlas por el autoencoder, esto con el fin de investigar si la teoría de análisis matemático presentada en este trabajo aporta alguna ventaja o si es preferible trabajar sin dichos filtros.

1. Leemos todas las imágenes y las escalamos a 128×128 para poder usar menos capas para llegar a un cuello de botella de dimensión 64.
 - a) Para el modelo en el cual filtramos la imagen primero con filtros de Gabor, usamos un banco de filtros con 3 orientaciones y 3 escalas diferentes. De esta manera cada imagen tiene 9 filtros aplicados y dichos resultados son los que usamos para construir nuestro autoencoder. La decisión

del número de orientaciones y escalas se sigue de lo sugerido por Chao en [18].

2. Para poder trabajar con imágenes lo primero que hacemos es “aplanarla” de tal manera que la matriz de representación de la imagen en blanco y negro de tamaño $n \times m$ sea un vector de dimensión nm . Como el querido lector ya se habrá dado cuenta, en este modelo vamos a usar imágenes en escalas de grises para que la dimensión del vector de entrada del autoencoder, nm no sea tan grande y podamos usar menos capas.
3. En total usamos 3 capas del codificador y otras 3 para el decodificador, pues en el libro de Zhang (en [12]) se lee como guía dado el tamaño de imágenes que tenemos.
4. Después dividimos al conjunto de las obras de arte entre el que vamos a usar para entrenar a nuestro modelo y con el que lo vamos a validar.
5. Entrenamos nuestro modelo y lo validamos.
6. Después obtenemos los valores de la capa de en medio una vez que evaluamos el modelo en cada obra de arte. Estos son los descriptores que usaremos para la clasificación.
7. Usando diversos métodos de clasificación como máquinas de vector soporte o bosques aleatorios hacemos la clasificación de estos descriptores de imágenes.
8. Comparamos aquellas imágenes que fueron clasificadas erróneamente y con qué artistas fueron confundidas.

3.1.3. Método GIST

Originalmente el descriptor GIST fue desarrollado por Aude Oliva y Antonio Torralba en [19] para clasificar imágenes y poder atribuirles algún lugar (escena de bosque, playa, desierto, etc). A partir de su trabajo dicho descriptor se ha vuelto popular para otros propósitos. En este trabajo lo implementaremos para nuestro objetivo de identificación

de estilos de artistas mexicanos. Consideramos que este descriptor es útil porque, según los autores, una escena es un lugar en donde nos podemos mover. No están tratando de categorizar objetos individuales sino lugares, vistas. Según Friedman, en inglés un “gist” es una representación abstracta de una escena que espontáneamente activa la memoria de representación de categorías de escenas o paisajes. Esto tiene cierta similitud con lo que tratamos de hacer en este trabajo.

Con este método vamos a trabajar con envolventes espaciales, los cuales son conjuntos de orillas. Estos bordes se le atribuyen a los diferentes elementos de una imagen. Lo interesante del trabajo de Torralba y Oliva es que primero hicieron experimentos con personas para ver cómo ellos separaban las imágenes que se parecieran más no por categorías fijas (como coches contra no coches). Esto nos dice que este descriptor fue hecho para clasificar imágenes globalmente e imitando (en lo que se pueda) a lo que una persona haría. La base de este método es a partir del análisis de componentes principales (o PCA) y la transformada de Fourier, específicamente la transformada de ventana de Fourier.

Idea detrás del descriptor de GIST

Primero describimos los tres niveles de descripción implementados por Torralba y Oliva:

- Subordinado, el cual se refiere a objetos en particular (como frutas, personas, joyería, etc.). Requiere de analizar estructuras locales y etiquetar regiones específicas de la imagen o pintura.
- Nivel básico, el cual es una combinación de objetos particulares para formar elementos mayores (como una ventana, un mueble, etc). Este es el nivel común en el cual se hace clasificación de imágenes. Se analizan áreas con forma similar, con componentes parecidos. Usualmente es un acercamiento geométrico.
- Nivel superordinado, el cual es la escena como un todo. Este es el nivel mayor de abstracción y por ende el nivel con menor semejanza en categoría visual pues cada imagen o pintura usada es única.

Como este método se basa en la transformada discreta de Fourier, tiene un elemento de no localización innato a él, como lo hemos visto en el capítulo 1. La magnitud al cuadrado de esta transformación es la distribución de la energía de la señal (imagen) a lo largo de diferentes frecuencias mas no están localizadas en la señal. Esto puede ser una ventaja porque el objetivo de este método es describir globalmente a la escena, no localmente. Para obtener la distribución espacial de la información espectral sí vamos a usar la transformada de Fourier de ventana, que como vimos en el capítulo 1 sirve para localizar las frecuencias detectadas. Esta localización de las frecuencias nos sirve para obtener un análisis más detallado de la escena, a nivel básico. En este método se calcula la transformada de Fourier de ventana en 8×8 localizaciones en la imagen con superposición de 64 píxeles cada uno. Esto se hace porque no queremos separar completamente a la imagen, como vimos con los modelos de Markov ocultos, distintos pedazos de la imagen dependen de sus vecinos. Para este caso la transformada de ventana de Fourier se escribe de la siguiente manera:

$$I(x, y, f_x, f_y) = \sum_{x', y'=0}^{N-1} i(x', y') h_r(x' - x, y' - y) e^{-i2\pi(f_x x' + f_y y')}.$$

A diferencia de lo que hicimos en el capítulo 1, este es un cambio de notación para que sea igual al del artículo de Torralba-Oliva. En este caso h_r es la ventana que estamos considerando, $i(x', y')$ es la imagen. Como habíamos mencionado, el espectro de energía localizado es el cuadrado de la magnitud de esta transformación, es decir:

$$A(x, y, f_x, f_y) = |I(x, y, f_x, f_y)|^2.$$

Para poder reducir la dimensión de los espectros obtenidos tras realizar las transformaciones usamos la descomposición de Karhunen-Loeve (KL). Esta descomposición es parte de la teoría de procesos estocásticos. Análogamente a la descomposición en series de Fourier, KL descompone un proceso en funciones ortogonales pero estas funciones son variables aleatorias. De esta manera, el espectro de la

energía de la señal usando la transformada de Fourier, se puede escribir de la siguiente manera:

$$A(f_x, f_y)^2 \sim \sum_{k=1}^{N_G} v_k \psi_k(f_x, f_y). \quad (3.1)$$

El querido lector, después de haber leído los capítulos 1 y 2 ya estará familiarizado con este tipo de descomposiciones. En el caso de la transformada de Fourier de ventana obtenemos:

$$A(x, y, f_x, f_y)^2 \sim \sum_{k=1}^{N_G} w_k \Psi_k(x, y, f_x, f_y). \quad (3.2)$$

Los coeficiente del tipo v_k aportan información estructural sin localizar, contienen descriptores de baja resolución del espectro de energía de la imagen. Los coeficientes del tipo w_k aportan información espacial, dado que estamos reduciendo la dimensión, tienen baja resolución en los dominios de frecuencia y espacio. Estas representaciones son redundantes pues la información de los v_k está contenida y más completa en los coeficientes w_k . A pesar de esto queremos ambas representaciones pues una nos da un contexto global, ligado al nivel superordinario y la otra al nivel básico.

Para estimar los envolventes espaciales usamos diferentes tipos de regresión, en particular Torralba y Oliva usan regresión lineal por la facilidad de lograr una buena interpretación. De esta manera el envolvente espacial para los v_k lo estima como el producto punto entre la amplitud del espectro y algo que llama DST (el modelo de discriminante espacial), el cual está asociado con las funciones que forman la base de la descomposición de KL con los pesos adecuados para la regresión. Para los w_k hacemos algo parecido, pero la transformada de ventana de Fourier, usando los Ψ_k como en 3.2, solo que en este caso a esa parte del producto punto se la llama WDST (modelo de discriminante espacial de ventana).

Cálculo del descriptor GIST

Después de varios experimentos con un banco de imágenes que contenía diversos paisajes, Torralba y Oliva introdujeron lo que después se llamó el descriptor GIST. La base de KL que terminaron usando fue un banco de filtros de Gabor con kernel real, como lo vimos en el capítulo 2, específicamente en 2.19.

Este banco de filtros consta de 32 filtros de Gabor con 4 escalas y 8 orientaciones. Cuando hemos definido estos filtros los aplicamos (vía convolución) a la imagen en cuestión, esto es para poder capturar información a nivel superordinado. Luego dividimos a la imagen en 4×4 secciones. En cada una de estas secciones, para cada uno de los filtros del banco calculamos su promedio de respuesta. Es decir, calculamos el promedio a lo ancho y a lo largo del resultado de aplicar el filtro a la imagen en cada sección. El descriptor de GIST es el vector donde se concatenan todos estos promedios para todas las secciones. Es decir, es un vector de dimensión $4 \cdot 4 \cdot 32 = 512$.

Lo que hace este descriptor es reducir la dimensión aportada en diferentes escalas y rotaciones en diferentes partes de la imagen. Esto da una descripción burda de la representación abstracta de una escena.

Para la implementación de este método seguimos la hecha por Oliva y Torralba en [20].

1. Para el banco de filtros de Gabor usamos 4 escalas y 8 orientaciones.
2. Al igual que en el artículo de Oliva y Torralba, dividimos a la imagen en 4×4 secciones, aplicamos los filtros y calculamos el promedio de su respuesta.
3. Para cada imagen entonces obtenemos un vector de dimensión 512, el cual usamos para hacer la clasificación.

3.1.4. Método SURF

El método SURF (por sus siglas en inglés *Speeded Up Robust Features*) fue presentado por primera vez por Herbert Bay, Tinne Tuytelaars y Luc Van Gool de la Escuela Politécnica de Zúrich (ETH) en [21]. La gran novedad de este descriptor que, como su nombre lo indica, se basa en cómputos rápidos usando filtros de caja. Este método se divide en dos pasos: extraer las características, describir dichas características.

Extracción de características

Lo primero que debemos de calcular es la integral de una imagen, lo cual es, dados $\mathbf{x} = (x, y)$ la suma de todos los píxeles de la imagen que caen dentro de la región rectangular conformada por el origen y este punto, es decir:

$$I_{\Sigma} = \sum_{k=0}^{k \leq x} \sum_{j=0}^{j \leq y} I(k, j).$$

Como en varios métodos numéricos modernos, el algoritmo de SURF usa a la matriz Hessiana por su rápido cálculo y precisión. Primero filtramos la imagen con un filtro Gaussiano con escala σ :

$$g(\mathbf{x}) = g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Entonces la matriz Hessiana $H(\mathbf{x}, \sigma)$ de la imagen bajo este filtro es:

$$H(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{x,y}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix},$$

donde $L_{xx}(\mathbf{x}, \sigma)$ es la convolución del filtro Gaussiano de la segunda derivada de la imagen I en el punto x . Esto nos dice que sí estamos perdiendo precisión, pues no nos estamos quedando con la Hessiana exacta, sino con la Hessiana que le ha sido aplicado un filtro, esta es una gran diferencia del método SURF contra otros métodos y hace que sea sumamente rápido. Gracias a esto podemos calcular el determinante de la matriz Hessiana aproximada bastante rápido, en contraste si tuviéramos una matriz con valores bastante diferentes en

todas sus entradas sin haber aplicado el filtro. De esta manera podemos trabajar con esta aproximación como una matriz por bloques.

Para obtener los puntos de interés debemos de analizar el espacio de escalas, como se describe en [22]. Esto se hace escalando hacia arriba el tamaño del filtro usado, a diferencia de otros métodos similares (como SIFT en [23]). Es decir, si empezamos con un filtro de 9×9 , luego usamos uno de tamaño 15×15 , luego uno de 21×21 y así hasta llegar al tamaño de la imagen original. Así los intervalos posibles para la extracción de puntos de interés son agrandados al doble de tamaño. Para poder localizar estos puntos de interés de la imagen en varias escalas diferentes, encontramos el máximo local en vecindades de tamaño 3×3 . Aquellos puntos de interés que prevalescan en diferentes escalas son considerados como los más importantes.

Descripción de características

Esta parte se vuelve a dividir en dos pasos: fijar una orientación y construir una región cuadrada a partir de esta orientación de la cual podamos extraer el descriptor SURF.

El primer paso de esta parte es sumamente importante porque queremos que nuestro descriptor local sea invariante a rotaciones. Esto es natural pedirlo pues queremos que el descriptor de una imagen sea el mismo aunque rotemos dicha imagen. Para esta parte se usan las ondículas de Haar vistas en el capítulo 2. Queremos encontrar la orientación principal de la imagen. Para esto calculamos la transformada de Haar en las direcciones x y y en una vecindad circular alrededor de un punto clave. Sumamos las respuestas de dicha transformada vertical y horizontalmente. Cambiamos la orientación en la cual estamos haciendo todo esto y volvemos a repetir los pasos anteriores. Nos vamos a quedar con la orientación que tenga la mayor suma, pues esa es la dirección principal.

Una vez que tenemos la dirección que vamos a usar, construimos un cuadrado centrado en el punto clave que ya usamos. Este cuadrado lo

dividimos en 4×4 regiones. Para cada una de estas subregiones vamos a tomar 5×5 puntos o píxeles y calculamos la transformada de Haar en las direcciones horizontales y verticales (respecto al cuadrado). Como lo hicimos en el primer paso, también vamos a filtrar estas respuestas con un filtro Gaussiano centrado en el punto de interés (los mejores resultados se han logrado con $\sigma = 3,3$). Luego estas transformadas en ambas direcciones se suman para cada subregión. También hacemos otro cálculo con estas transformadas: sumamos sus valores absolutos. Esto es para poder obtener información sobre cambios de intensidad, si no usamos valores absolutos puede ser que tengan signos opuestos en diferentes direcciones y por ende no captemos los cambios en intensidad. A las transformadas se les abrevia con un dx y dy para las distintas direcciones. Con lo que acabamos de hacer ahora tenemos un vector de dimensión 4 para cada subregión $V = (\Sigma dx, \Sigma dy, \Sigma |dx|, \Sigma |dy|)$. Por ende, el descriptor para el cuadrado original es un vector de dimensión 64. Este es el vector que vamos a usar como descriptor local del punto de interés. El estimado lector se podrá haber dado cuenta que nos centramos en un punto para hacer todo esto. Esta es la flexibilidad del método SURF. Podemos elegir que nos de un descriptor (vector) para “el punto más representativo” de la imagen o podemos decirle que nos de los n vectores de los n puntos más representativos de la imagen. El código que usamos en MATLAB también nos deja visualizar estos puntos junto con los círculos. La figura 3.9 es un ejemplo de esto. El querido lector puede tener acceso a la librería usada en MATLAB por medio de [24].



Figura 3.9. Los 10 puntos y descriptores SURF para la obra *Birth Bath* de Leonora Carrington

Como lo hemos descrito, al implementar este método, podemos elegir la cantidad de puntos representativos a considerar y obtener sus descriptores. Para este trabajo usaremos los 10 puntos más importantes para cada obra de arte, es decir los 10 puntos que fueron reconocidos como “más importantes” en varias escalas diferentes. Luego de calcular dichos puntos extrajimos el vector de dimensión 64 asociado a éstas.

Estos vectores son los que usamos para la clasificación final.

3.2. Métodos de clasificación

Una vez que obtenemos los descriptores de imágenes podemos clasificarlos con diferentes técnicas. El objetivo es poder tener distintas técnicas para los distintos descriptores pues fueron obtenidos de diferentes maneras.

3.2.1. Bosques aleatorios

Como sucede con un bosque de árboles reales, los bosques aleatorios están formados por varios árboles, solo que en este caso estos árboles son árboles de decisión, como lo explica Izenman en [13]. Para poder explicar el método de clasificación con bosques aleatorios primero debemos dar una breve introducción sobre árboles de clasificación.

Árboles de clasificación

Suponga el querido lector que le preguntan que adivine lo que el escritor está pensando, puede hacer preguntas, pero sólo obtendrá “sí” o “no” como respuestas. El lector empezará haciendo preguntas y basándose en estas respuestas hará preguntas más específicas hasta llegar a un punto en el que haya adivinado o esté lo suficientemente cerca como para asumir que sí adivinó. En esta técnica se basan los árboles de clasificación: en secuencias ordenadas de preguntas para clasificar.

En cada nodo del árbol se le hace una pregunta a los datos que tengamos y las ramas simbolizan las respuestas, como se aprecia en la figura 3.10. Al primer nodo se le conoce como la raíz, de ahí partimos. Este algoritmo no puede seguir indefinidamente, por ende debe de haber nodos terminales, es decir donde ya no podamos hacer más preguntas. En esos nodos vamos a suponer que ya estamos lo “suficientemente cerca” de la clasificación verdadera. Podemos tener

más de un nodo terminal con la misma clasificación, es decir llegaron a la misma conclusión, pero con preguntas diferentes.

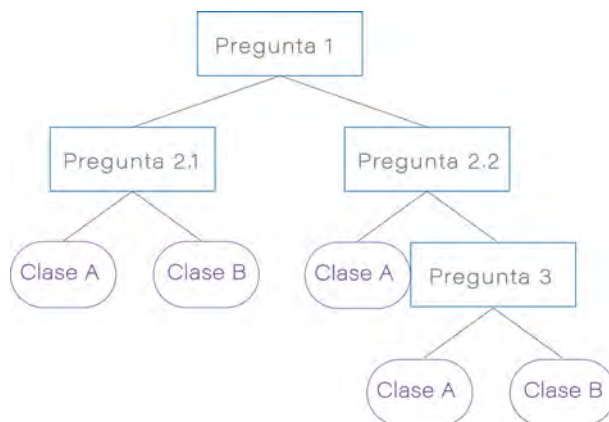


Figura 3.10. Ejemplo árbol de clasificación

El querido lector seguramente se estará cuestionando sobre las preguntas que debe de hacer el árbol de clasificación, cómo decidir cuando un nodo ya es terminal y cómo asignarle una clase a cada nodo terminal. Este es el arte de “crecer” árboles de decisión. En cada nodo, el algoritmo para crear árboles de decisión debe de decidir en cuál variable es mejor hacer la división. Debe de considerar todas las divisiones posibles y decidir cuál es la mejor. Si tenemos M categorías posibles, entonces hay $2^{M-1} - 1$ distintos posibles cortes para esa variable.

Ahora debemos de elegir el criterio para decir cuál división es mejor que otra. De manera similar a las redes neuronales, necesitamos una función objetivo. En este caso dicha función se llama función de impureza de los nodos y se escribe de la siguiente manera:

$$i(\tau) = \phi(p(1|\tau), p(2|\tau), \dots, p(M|\tau)),$$

donde $p(k|\tau)$ es la probabilidad condicional de la observación k dado que está en el nodo τ . Hay diferentes formas de definir a ϕ , las maneras más populares es con la función de entropía:

$$i(\tau) = - \sum_{k=1}^M p(k|\tau) \log(p(k|\tau))$$

y el índice de diversidad de Gini:

$$i(\tau) = \sum_{k \neq k'} p(k|\tau) p(k'|\tau) = 1 - \sum_{k=1}^M (p(k|\tau))^2.$$

La bondad de la división hecha en el nodo τ es la reducción en la impureza que se logró al pasar de τ a alguno de sus nodos hijos. La mejor división por ende, es aquella que maximice la bondad de la división. El proceso en el cual construimos las capas del árbol se llama partición recursiva. El problema de ir creciendo árboles de decisión es que podemos crecerlos tanto que abarquen todas las posibilidades. En tal caso decimos que el árbol está saturado, esto nos puede traer problemas de sobreajuste del modelo a los datos. Es decir, que el modelo describe exactamente los datos de entrenamiento que se le da, mas no es capaz de generalizar a situaciones o datos similares. No queremos este tipo de problemas. Hay varias técnicas usadas para evitar este tipo de cuestiones, como definir una profundidad máxima del árbol, saturarlo y luego podarlo, entre otras.

Análogo a lo que mencionamos en la sección sobre redes neuronales, queremos tener una función para calcular los errores de clasificación errónea y por ende minimizar esta función. Sea T un árbol de clasificación con $\tilde{T} = \{\tau_1, \tau_2, \dots, \tau_L\}$ sus nodos terminales. La tasa de clasificación errónea para algún nodo τ en particular es:

$$r(\tau) = 1 - \max_k \{p(k|\tau)\}.$$

De esta manera la tasa de clasificación errónea del árbol es:

$$R(T) = \sum_{l=1}^L R(\tau_l) P(\tau_l),$$

donde $P(\tau_l)$ es la probabilidad de que la observación caiga en el nodo τ .

Una técnica ampliamente usada para evitar sobreajustar los árboles de clasificación se basa en saturarlo y luego podarlo. Lo que primero hacemos es crecer un árbol “grande” (depende de la cantidad de clases que nuestro problema tenga), llamémosle T_{max} . Luego calculamos $R(\tau)$ para cada nodo del árbol. Vamos a ir podando los nodos de los nodos terminales a la raíz conforme $R(\tau)$ se vaya minimizando. Para poder hacer esto definimos una función de costo-complejidad al ir podando el árbol:

$$\begin{aligned} R_\alpha(T) &= \sum_{l=1}^L r(\tau_l) p(\tau_l) + \alpha \\ &= \sum_{l=1}^L R^{re}(\tau_l) + \alpha = \sum_{l=1}^L R_\alpha(\tau_l) \\ &= R^{re}(T) + \alpha|\tilde{T}|, \end{aligned}$$

donde $|\tilde{T}| = L$ es la cantidad de nodos terminales en el subárbol T de T_{max} . Podemos pensar a $\alpha|\tilde{T}|$ como la penalización de tener un árbol “demasiado” grande. Lo que hacemos es que para cada α elegimos aquél subárbol $T(\alpha)$ de T_{max} tal que $R_\alpha(T(\alpha))$ se minimice. Al fin de cuentas α es el parámetro que nos ayuda a definir la profundidad del árbol.

Bosques aleatorios

Una vez que hemos entendido las bases detrás de los árboles de clasificación podemos entender cómo funcionan los bosques aleatorios. Un bosque aleatorio, como ya hemos mencionado, consiste de una “gran” cantidad de árboles de decisión que funcionan como un conjunto. Cada uno de los árboles de decisión da un “voto” de cada una de nuestras clases. Al final la clase con la mayor cantidad de votos es la clase a la que se le asigna nuestra predicción, es una democracia de árboles de decisión. Esta idea es importante porque, al tener varios

árboles, podemos “proteger” a nuestra predicción de errores que pueden cometer dichos árboles de manera individual pues, en teoría los árboles tienen una baja correlación entre sí. Como el querido lector seguramente ya notó, conforme más clases a elegir tengamos, necesitaremos una mayor cantidad de árboles de decisión.

Para poder garantizar que tengamos una “adecuada representación” de árboles, o que cada árbol haga distintas preguntas hay dos métodos que se suelen usar:

- Muestreo tipo bootstrap: dado que cada árbol es altamente sensible al conjunto de datos con el que es entrenado, podemos tomar ventaja de esto y dejar que cada uno tome distintas muestras con reemplazo del conjunto de datos de entrenamiento.
- Aleatoriedad en las características para hacer divisiones: para forzar a que haya una mayor variedad en el tipo de preguntas que cada árbol hace podemos restringir a que cada uno pueda seleccionar las divisiones de un subconjunto de todas las divisiones posibles.

En conclusión, los bosques aleatorios han tomado un gran auge recientemente porque los podemos considerar como una democracia que evita sus errores propios al tomar una gran diversidad de árboles de decisión individuales. En el libro de Izenmann, en [13] se profundiza más sobre estas técnicas de clasificación.

3.2.2. Máquinas de vector soporte (SVM)

El desarrollo e implementación de algoritmos para máquinas de vector soporte (SVM por sus siglas en inglés) tienen valor tanto teórico en probabilidad y optimización como práctico en aprendizaje de máquina, bioinformática y ciberseguridad, como lo menciona Izenman en [13]. Es por esto que varias ramas de las matemáticas y estadística han dedicado grandes esfuerzos por estudiarlas e implementarlas. Hay máquinas de vector soporte lineales y no lineales, con diferentes núcleos o kernels lo que las convierte en mecanismos sumamente flexibles para distintas tareas de clasificación. En este trabajo usamos una lineal y otra no lineal.

Lineales

La idea detrás de las máquinas de vector soporte empieza con las que son lineales y no es complicada de entender. Supongamos que tenemos el siguiente conjunto de entrenamiento de datos: $\mathcal{L} = \{(x_k, y_k) : k = 1, \dots, n\}$. Donde $x_k \in \mathbb{R}^N$ y $y_k \in \{-1, +1\}$. Podemos interpretar esto como que las x_k son datos que pertenecen a alguna de dos categorías, indexadas por y_k . Queremos construir un clasificador a partir de \mathcal{L} basándonos en estas dos categorías, es decir:

$$C(\mathbf{x}) = \text{signo}(f(\mathbf{x})),$$

es decir, $C(x)$ es igual a -1 si $f(x)$ es negativo, 1 si $f(x)$ es positivo y 0 si $f(x) = 0$. Vamos a suponer que estamos trabajando con un conjunto linealmente separable, es decir que existe algún o algunos hiperplanos de tal manera que dividan a \mathbb{R}^N en dos y que en cada uno de estas separaciones se encuentren todos los elementos de cada clase sin mezclarse. Esto quiere decir que los puntos de \mathcal{L} pueden ser separados por un hiperplano

$$\{\mathbf{x} : f(\mathbf{x}) = \beta_0 + \mathbf{x}^T \beta = 0\},$$

donde β es un vector de pesos. Si un hiperplano sí logra separar al conjunto de datos \mathcal{L} entonces lo llamaremos hiperplano separador. Como el curioso lector podrá notar en el ejemplo mostrado en la figura , pueden existir diversos hiperplanos separadores para un mismo conjunto de datos. El algoritmo de la máquina de vector soporte trata de encontrar el mejor hiperplano que logre dicha clasificación.

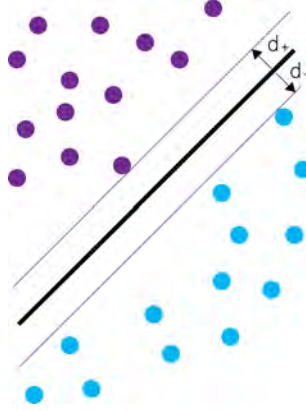


Figura 3.11. Ejemplo máquina de vector soporte lineal

Como podemos ver en la figura 3.11, sea d_- la distancia más corta del hiperplano separador al dato negativo más cercano y sea d_+ la distancia más corta del hiperplano al dato positivo más cercano, entonces el margen de separación de dicho hiperplano es $d = d_- + d_+$. Aquél hiperplano separador que maximiza d es el hiperplano separador óptimo. Esto es en el caso de tener dos categorías, si tenemos w categorías, como es nuestro caso, entonces vamos a tener $d = d_1 + d_2 + \dots + d_{w-1}$. Para ilustrar la teoría detrás de las máquinas de vector soporte vamos a suponer que tenemos solo 2 categorías. Si los datos son linealmente separables entonces:

$$\beta_0 + x_k^T \beta \geq 1 \quad , \quad \text{si } y_k = 1 \quad (3.3)$$

$$\beta_0 + x_k^T \beta \leq -1 \quad , \quad \text{si } y_k = -1 \quad (3.4)$$

Si la igualdad se cumple en 3.3 para algunos puntos en \mathcal{L} , entonces dichos puntos se encuentran sobre el hiperplano $H_{+1} = (\beta_0 - 1) + \mathbf{x}^T \beta = 0$. Aquellos puntos de \mathcal{L} que satisfagan la igualdad en 3.4 se encuentran sobre el hiperplano $H_{-1} = (\beta_0 + 1) + \mathbf{x}^T \beta = 0$. Los puntos que se encuentran sobre H_{+1} o sobre H_{-1} se llaman vectores soporte, de ahí el nombre de este método de clasificación. En este caso, el margen de separación está dado por $d = \frac{2}{\|\beta\|}$. De esta manera, el problema a resolver es uno de optimización convexa:

$$\begin{aligned} & \text{minimizar} \quad \frac{1}{2} \|\beta\|, \\ & \text{tal que} \quad y_k (\beta_0 + x_k^T \beta) \geq 1, \quad k = 1, \dots, n. \end{aligned}$$

Este problema se puede resolver de diferentes maneras, por multiplicadores de Lagrange, programación cuadrática sucesiva u otros métodos vistos en el libro de Nocedal en [17].

En aplicaciones con datos reales, usualmente no nos podemos topar con separaciones lineales de manera tan fácil. Podemos esperar que algún punto pase la barrera del hiperplano, pero no debemos de desechar las máquinas de vector soporte lineales tan sencillamente. Para poder atacar este tipo de problemas usamos márgenes suaves de soluciones, es decir agregamos una variable de holgura no negativa para cada dato en \mathcal{L} , llamémosles ξ_k . La función objetivo en el problema de minimización también la cambiamos, hay varias técnicas para hacer esto. En [17] se sugieren varias técnicas para suavizar este tipo de problemas de optimización, como al usar la penalización de norma 1:

$$\begin{aligned} & \text{minimizar} \quad \frac{1}{2} \|\beta\|^2 + C \sum_{k=1}^n \xi_k, \\ & \text{tal que} \quad y_k (\beta_0 + x_k^T \beta) + \xi_k \geq 1, \\ & \quad \quad \xi_k \geq 0, \quad k = 1, \dots, n, \end{aligned} \tag{3.5}$$

donde C es un parámetro. Hay varios métodos de optimización numérica que relajan el problema original y buscan soluciones por diferentes vías, pero esto se escapa del estudio de este trabajo.

Radiales

Este tipo de máquinas de vector soporte son no lineales, es decir involucran transformaciones no lineales de los vectores de entrada y usa núcleos para simplificar el trabajo computacional lo más que se pueda. La clave para construir máquinas de vector soporte viene dado por el problema dual de optimización, el cual depende de los productos

internos $\langle x_k, x_k \rangle$.

Supongamos que a cada observación x_k de \mathcal{L} le aplicamos una transformación no lineal $\Phi : \mathbb{R}^N \rightarrow \mathcal{H}$, donde \mathcal{H} es un espacio de características $N_{\mathcal{H}}$ dimensional. Lo que queremos es que bajo esta transformación podamos hacer la clasificación de manera más “sencilla”. Para poder facilitar las propiedades y cálculos de dicha transformación, vamos a suponer que \mathcal{H} es un espacio de Hilbert de funciones reales con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ y norma $\| \cdot \|_{\mathcal{H}}$. De esta manera:

$$\Phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_{N_{\mathcal{H}}}(\mathbf{x}))^T.$$

Por ende, ahora tenemos como muestras $\{\Phi(x_k), y_k\}$, con $y_k \in \{-1, 1\}$. Como el querido lector recuerda, queremos encontrar un hiperplano de tal manera que podamos separar las observaciones de las distintas categorías. Si $N_{\mathcal{H}}$ es muy grande entonces vamos a tener problemas computacionales al tratar de optimizar problemas de la forma de 3.5. Lo que hacemos entonces para tratar de evitar problemas con esta dimensión es que, en lugar de calcular los productos internos en \mathcal{H} , los calculamos usando un kernel no lineal, es decir:

$$K(x_k, x_j) = \langle \Phi(x_k), \Phi(x_j) \rangle. \quad (3.6)$$

Pero este producto interno es el que se usa en \mathbb{R}^N , es decir el dominio de los valores de entrada. Esto nos ayuda con el problema de tener $N_{\mathcal{H}}$ muy grandes y evitamos conflictos con errores computacionales. Después de hacer esto ya podemos calcular el clasificador óptimo de tipo máquina de vector soporte.

Hay kerneles para diferentes aplicaciones, los más usados son los polinomiales y los radiales. Para el caso particular de kerneles o núcleos radiales, esta función es de la forma:

$$K(x_k, x_j) = \exp \{ -\gamma \|x_k - x_j\|^2 \},$$

donde $\| \cdot \|$ es la norma que usamos en \mathbb{R}^N (podemos emplear distintas normas dependiendo de la situación, pero usualmente se usa la norma euclidiana) y γ es un parámetro para evitar que el modelo se sobre

ajuste. Intuitivamente este kernel es una función cuyo valor depende de la distancia desde el origen a un punto.

Capítulo 4

Clasificación de obras de arte

Una vez que tenemos toda la teoría y las técnicas necesarias, podemos proceder a atacar nuestro problema: hacer análisis estilísticos de los artistas mexicanos elegidos basándonos en ondículas y técnicas de aprendizaje de máquina. Vamos a clasificar las pinturas con su autor usando como criterio diferentes métodos. Nos interesa saber cuáles artistas son “fáciles” de distinguir de los demás, es decir cuyas obras no fueron confundidas con otros autores. También nos interesan aquellas pinturas que fueron clasificadas erróneamente pues significa que dicha pintura tiene elementos estilísticos de otro artista y por ende el autor original comparte características con el pintor con el cual fue confundido.

Independientemente de la complejidad matemática y computacional que un problema de clasificación de imágenes representa, clasificar obras de arte es por sí mismo un problema al que muchos historiadores del arte se enfrentan pues hay varios criterios estilísticos (como el formato, la composición, texturas, etc) e históricos. En el caso de la historia del arte mexicano, las líneas fronterizas entre diversas corrientes no están claramente marcadas. Nos vamos a centrar en artistas destacados de inicio del siglo XX: Alfredo Ramos Martínez, Gerardo Murillo (Dr. Atl), José Clemente Orozco, Frida Kahlo, Diego Rivera, Rufino Tamayo, María Izquierdo, Leonora Carrington, Remedios Varo y Carlos Mérida. Varios de ellos son catalogados como

artistas patrimoniales de México y su obra no sólo se cotiza en millones de dólares, sino que es obra protegida como lo explica el Instituto Nacional de Bellas Artes y Literatura (INBAL) en [25]. Elegimos trabajar con este período de la historia del arte mexicano pues fue el inicio del camino propio que tomó México en cuestiones artísticas y cuando empezó a destacar a nivel internacional de manera única, como lo explica Leticia Orozco en [26].

Iniciaremos este capítulo dando una breve introducción a cada uno de los artistas que aquí consideramos para que el lector pueda tener una buena idea de la diferencia estilística con la que estamos trabajando. Proseguiremos a dar los resultados de dichos métodos y hacer un análisis crítico de los resultados. Al final daremos unas conclusiones englobando los resultados de los cuatro capítulos presentes en este trabajo.

4.1. Introducción a la historia del arte de inicio del siglo XX

Como explica Teresa del Conde en [27], no podemos considerar que el siglo XX empieza en 1900 para el arte. Decidir cuándo empieza depende del corte que se considere. Podríamos pensar que el arte del siglo XX empieza con los brotes nacionalistas o que empiece con las vanguardias europeas. Como menciona Oles en su libro en [28], durante el Porfiriato, la arquitectura, pintura y escultura fueron los resultados de la élite para transformar la imagen de México en el extranjero. Los profesores de la Escuela Nacional de Bellas Artes (ENBA) seguían promoviendo corrientes “anticuadas” como el realismo académico y efectos barroco, la base de la pintura descansaba en el dibujo realista. La pintura académica romantizaba la idea del indígena mexicano en el extranjero. Durante la prosperidad económica del Porfiriato, la idea de “lo moderno” se importó de Europa, principalmente de Francia.

Al mismo tiempo que la ENBA promovía las corrientes anticuadas, otros artistas se empezaron a interesar por nuevas técnicas de pintura, nuevas maneras de representar “lo mexicano” y nuevas temáticas.

Como menciona Oles, entre 1900 y 1920 se encuentra una mayor diversidad estilística y temática tanto dentro de la academia como fuera. El muralismo mexicano es aclamado mundialmente, en libros de arte universal siempre es mencionado como un parteaguas en la historia mundial. A pesar de que los artistas empezaron a pintar murales después de la lucha armada de 1910, todas estas ideas, técnicas y modos de ver a la pintura empezaron desde antes con personajes como Alfredo Ramos Martínez y Gerardo Murillo (Dr. Atl). A partir del movimiento muralista y de la situación mundial del momento, fue cuando varios artistas y movimientos llegaron a México, como el surrealismo con Remedios Varo y Leonora Carrington. A continuación damos una breve introducción a la historia de cada uno de los artistas que consideramos para este trabajo, como en [29] se describen.

Alfredo Ramos Martínez (1871-1946)

Es un artista muy importante para la historia del arte mexicano porque funda las Escuelas al Aire Libre de Santa Anita en 1920, parecidas a las escuelas de Barbizon que dieron inicio al movimiento del impresionismo. La idea era salir de la academia, pintar lo que estaba pasando en realidad en México. Los mismos artistas también daban clases a la gente que quería participar, también la gente empezó a entrar a la academia. Fue la primera vez que en México sucedía algo de este estilo, dio pie a que los artistas se interesaran por el “verdadero” México, no solo por la idea romántica que solían representar.



Abrazo Materno, Alfredo Ramos Martínez

Gerardo Murillo, Dr Atl (1875-1964)

Nació en Guadalajara en 1875. Como la mayoría de los artistas de esta época, viajó a Europa para seguir sus estudios, de donde regresó con varias ideas nuevas. Es conocido por haber organizado una exposición de artistas mexicanos durante los festejos del centenario de la Independencia. Como se menciona en el catálogo de dicha exposición en [30], fue un evento importante pues el gobierno de la época no había considerado a los artistas mexicanos para dichos festejos, pero con esta exposición dieron a conocer sus obras, ideas y estilos particulares. Incitó a los artistas de esta época a hacer varios cambios y exigirlos en la academia. También es conocido por sus series de escenas volcánicas, en especial del Parícutín usando materiales que él mismo innovó: los *Atl Colors*.



El Parícutín, Dr. Atl

José Clemente Orozco (1883-1949)

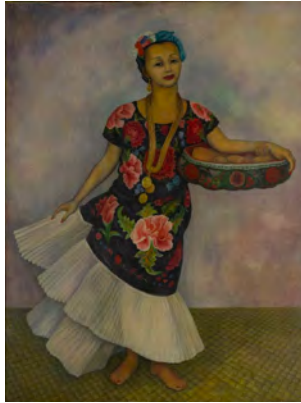
Nació en Jalisco en 1883. Junto con Siqueiros y Rivera es considerado uno de los “3 grandes” del muralismo mexicano. Obtuvo sus primeras lecciones en el taller del famoso grabador José Guadalupe Posada. Trabajó en varios periódicos haciendo caricatura política. Sus murales se encuentran no sólo en México, sino también en Estados Unidos y Europa. Su autobiografía nos relata gran parte de lo que pasaba en el ámbito artístico en México así como una parte crucial del período revolucionario.



La Conquista, José Clemente Orozco

Diego Rivera (1886-1957)

Nació en Guanajuato en 1886. Sin duda es, sino él más famoso, uno de los artistas mexicanos más destacados a nivel internacional. Pasó gran parte de su juventud en Europa, donde se desarrolló como un gran pintor cubista. Regresó por encargo del secretario de educación, José Vasconcelos para iniciar el proyecto muralista educativo post revolucionario. Pintó el primer mural en el anfiteatro de la Escuela Nacional Preparatoria sin haber pintado nada de esas dimensiones ni con esa técnica antes. Pintó varios murales en la Secretaría de Educación Pública, en Estados Unidos y en varios estados de México. Es uno de los pocos artistas mexicanos que experimentó con diversos movimientos (cubismo, muralismo, surrealismo, etc) y destaca en la mayoría de ellos. Durante su vida tuvo una fuerte influencia en lo que pasaba dentro de la historia del arte mexicano, así como de la monetización de la pintura gracias a su poder en la Galería de Arte Mexicano (GAM).



Retrato de Dolores Olmedo, Diego Rivera

Carlos Mérida (1891-1984)

A pesar de haber nacido en Guatemala, obtuvo la nacionalidad mexicana después y es considerado un artista mexicano. Fundó y dirigió la Escuela de Danza de la SEP. Gracias a sus raíces, fue muy apegado a la cultura maya-quiché, por eso en su obra se destaca la grandeza de los pueblos americanos, a diferencia de otros artistas. Estudió en París y tuvo una corta estancia en Nueva York, por lo que es de los pocos artistas que logran fusionar elementos occidentales con formas representativas mesoamericanas en su obra (pintura y escultura). Es de los artistas mexicanos que más destaca en lo figurativo, toma referencias del cubismo y la abstracción. Forma parte del muralismo (pero otro período) y destacan sus murales abstractos.



Pájaro Alucinado, Carlos Mérida

Rufino Tamayo (1899-1991)

El querido lector muy probablemente ubique obras de Tamayo, mas la obra temprana de este artista es completamente distinta al Tamayo que todos conocemos. Estudió bajo el método de dibujo de Best Maugard, el cual se basaba en dibujar con formas simples como líneas, círculos y espirales. Su obra poco a poco se fue alejando del discurso político del cual otros artistas de su época tomaban, en cambio se dedicó a pintar elementos de las culturas mesoamericanas. Su estilo tan particular involucra características derivadas del cubismo, surrealismo y expresionismo.



Las sandías, Rufino Tamayo

María Izquierdo (1902-1955)

Fue la mujer pintora más influyente de su época, más que Frida Kahlo. Fue la primera artista mexicana en exponer en Nueva York y París. Desarrolló una pintura nueva, con un corte primitivo. Pintó sobre la vida cotidiana mexicana, las ferias, circos, mesas, etc. Todo para poder plasmar los sentimientos de lo “mexicano” desde lo cotidiano, sin romantizarlo. Retomó varios elementos relacionados con la nostalgia, fantasía y alegorías. Varios historiadores del arte la consideran dentro del movimiento surrealista.



El Domador, María Izquierdo

Frida Kahlo (1907-1954)

Es sin dudarla la artista mexicana más destacada en la actualidad. Su vida fue bastante trágica, llena de problemas médicos como el querido lector puede leer en su expediente médico en [31]. Se enfermó de poliometitis, tuvo un accidente en el cual se fracturó la columna en tres partes, tuvo varios abortos, a raíz de su accidente tuvo varios problemas para moverse que incluso a veces pintaba con la boca. Estos problemas médicos y su sufrimiento lo plasmó en su pintura, gran parte por la cual se volvió tan famosa. Es considerada por varios historiadores del arte y artistas como surrealista, a pesar de que ella nunca se proclamó como tal.



Las Dos Fridas, Frida Kahlo

Remedios Varo (1908-1963)

A pesar de haber nacido en España, huyó de la Guerra Civil, vivió los problemas de la guerra en Europa y llegó a México, donde terminó teniendo la nacionalidad. En México encontró refugio con varios otros artistas extranjeros como Leonora Carrington, con la que tuvo una gran amistad y ambas se identificaban con el movimiento surrealista en México. Una de las características que más destacan en su pintura es la gran atención al detalle y su inspiración en sus propios sueños. En su obra se plasma su gran inteligencia así como curiosidad hacia la ciencia por los temas y objetos que pinta.



La Creación de las Aves, Remedios Varo

Leonora Carrington (1917-2011)

Como Varo, ella también proviene de Europa, pero de Inglaterra. Por su carácter rebelde e inteligente fue acusada varias veces de estar loca, esto aunado con los problemas de la guerra Europea hizo que llegara a México. En México logró destacar como pintora, escultora y escritora surrealista. Su pintura es impactante, pero también los otros objetos que realizó como botellas de tequila.



El Nigromante, Leonora Carrington

4.2. Métodos usados

¿Cómo podemos clasificar artistas? ¿Cómo podemos distinguir algunos de los rasgos de sus estilos tan particulares usando matemáticas? En esta sección vamos a explicar cómo, a partir de los elementos descritos en el capítulo 3, vamos a clasificar las obras y estilos de estos artistas. Primero queremos obtener información de la base de datos de las imágenes que tenemos de estos artistas, para eso se usan los métodos descritos en la primera sección del capítulo 3. Después queremos clasificar estos descriptores de imágenes, las etiquetas o categorías en este caso son los artistas mismos. Queremos métodos que logran una buena clasificación de las obras por artistas, pero también nos interesan aquellas obras que no fueron clasificadas correctamente. ¿Por

qué? Porque esa clasificación errónea significa que nuestros algoritmos confundieron el estilo de un artista por el de otro, a nivel superordinado. Con esto podemos analizar las pinturas de manera individual que el algoritmo confundió con otro artista, pero también podemos indicar qué tanto confunden nuestros algoritmos un artista con otro. Eso significa que los estilos de estos artistas son similares, mientras que artistas que no fueron confundidos tienen estilos diferentes.

La siguiente tabla muestra la cantidad de obras por artista que estamos considerando. El querido lector debe de notar la diferencia entre la cantidad de obras usadas en este trabajo por artista. Esto se debe a que necesitamos imágenes con buena resolución y calidad de color (recordemos que son imágenes de obras, no son las obras en sí). Aunado a esto, es difícil obtener una buena base de datos de pintura mexicana puesto que varias de las obras son murales (no consideramos murales para este trabajo) o se encuentran en colecciones privadas. Esta base de datos fue obtenida de diversas fuentes como las que se encuentran en [32, 33, 34, 35, 36, 37, 38, 39, 40].

| Artista | Cantidad de obras |
|--------------------------|-------------------|
| Gerardo Murillo (Dr Atl) | 44 |
| Leonora Carrington | 85 |
| Frida Kahlo | 52 |
| María Izquierdo | 57 |
| Carlos Mérida | 99 |
| José Clemente Orozco | 46 |
| Alfredo Ramos Martínez | 144 |
| Diego Rivera | 113 |
| Rufino Tamayo | 94 |
| Remedios Varo | 75 |

En total tenemos 5 modelos para obtener los descriptores de las imágenes y 3 maneras de clasificar dichos descriptores. Estos modelos son los que describimos en el capítulo 3 en la sección titulada “Métodos de obtención de datos a partir de imágenes”. Las maneras de clasificar dichos descriptores son las que se encuentran también en el capítulo 3,

pero en la sección titulada “Métodos de clasificación”. Para todos estos modelos usamos el 65 % de las obras para entrenar al modelo y el restante 35 % para validarlo. Esta separación la hicimos de manera aleatoria, pero usamos una semilla para que fuera igual en todos los casos.

4.3. Resultados

Después de implementar estos métodos y hacer las clasificaciones con bosques aleatorios, máquinas de vector soporte lineal y máquinas de vector soporte radial obtuvimos las siguientes eficiencias (respecto al conjunto de validación):

| | Bosque aleatorio | SVM Lineal | SVM Radial |
|------------|------------------|------------|------------|
| HMM | 94.7 | 46.6 | 41.2 |
| AE | 25.7 | 26.9 | 25.2 |
| AE y Gabor | 15.4 | 19.7 | 58.7 |
| GIST | 41.7 | 70.1 | 59.9 |
| SURF | 23.8 | 51.0 | 43.2 |

Cuadro 4.1. Porcentaje de obras clasificadas correctamente por método usado en el conjunto de validación

Como podemos ver, el modelo con la menor eficiencia fue cuando usamos al autoencoder sin usar los filtros de Gabor. Este modelo parece que no hubo aprendizaje de las diferencias entre las obras de cada artista. La eficiencia aumentó cuando filtramos las obras con los filtros de Gabor en el caso de clasificarlos con máquinas de vector soporte radiales, al pasar de 25,2 % a 58,7 %. Esto probablemente se deba a que, al usar los filtros de Gabor tenemos más información de las obras, para cada imagen tenemos ahora 9 datos a pasar por el autoencoder, no solo uno. El método GIST parece funcionar “bastante bien” al hacer la clasificación de estos descriptores con máquinas de vector soporte lineal. Esto es bastante agradable puesto que el los métodos GIST y SVM lineal son de los más rápidos que logramos implementar y alcanzaron una buena eficiencia. El modelo con campos aleatorios de Markov es otro modelo con una buena eficiencia al ser

clasificado con bosques aleatorios, alcanzando una clasificación correcta del 94,7 %.

A continuación vamos a ver tablas de la eficiencia alcanzada por cada método de clasificación por artista. El primer cuadro en 4.2 nos muestra la eficiencia alcanzada por artista para el método de clasificación de bosque aleatorio. Como podemos notar, en todos los métodos de obtención de datos, Alfredo Ramos Martínez obtuvo una eficiencia consistentemente alta. Esto nos puede indicar que entre todos los artistas considerados aquí, Ramos Martínez es aquél cuyo estilo es más definido. También debemos de tomar otro punto en consideración para llegar a esta conclusión: Ramos Martínez es el artista con la mayor cantidad de obras (145) que utilizamos. Es decir, todos los métodos tuvieron más material para diferenciar el estilo de Ramos Martínez.

| | Atl | Carr | Frida | Izq | Mérida | Orozco | Ramos Mtz | Rivera | Tam | Varo |
|---------------|-------|-------|-------|-------|--------|--------|--------------|--------|-------|-------|
| HMM | 93.18 | 91.76 | 92.31 | 96.49 | 100 | 89.13 | 97.92 | 96.46 | 89.36 | 93.33 |
| AE | 0.00 | 26.19 | 0.00 | 1.72 | 56.57 | 0.00 | 85.21 | 4.55 | 2.11 | 1.33 |
| AE y Gabor | 5.88 | 6.85 | 0.00 | 0.00 | 18.29 | 0.00 | 66.67 | 29.25 | 1.32 | 1.45 |
| GIST | 29.55 | 10.59 | 0.00 | 47.37 | 69.70 | 4.35 | 58.56 | 78.47 | 29.79 | 13.33 |
| SURF | 2.38 | 10.59 | 0.00 | 1.72 | 42.86 | 0.00 | 63.38 | 31.53 | 10.00 | 2.70 |

Cuadro 4.2. Eficiencia por artista para el bosque aleatorio

En el cuadro 4.3 mostramos la eficiencia por artista usando la clasificación de máquinas de vector soporte con kernel lineal. Podemos notar que para el modelo de campos aleatorios de Markov, los dos artistas con una alta tasa de eficiencia en su clasificación fueron Carlos Mérida y Alfredo Ramos Martínez. De nuevo, esto puede indicar que éstos dos son los artistas que más se diferencian del resto. El querido lector podrá apreciar obras de ambos artistas y notar que, efectivamente la abstracción lograda por Carlos Mérida y los materiales empleados por Alfredo Ramos Martínez tienen muy poco en común con los estilos de los otros artistas.

Finalmente, en la tabla 4.4 apreciamos la eficiencia para el modelo de máquina de vector soporte radial. Nuestra hipótesis que Alfredo Ramos Martínez es el artista más diferente dentro de nuestra selección se fortalece al observar que, de nuevo, es el artista que alcanza una mejor

| | Atl | Carr | Frida | Izq | Mérida | Orozco | Ramos Mtz | Rivera | Tam | Varo |
|---------------|-------|-------|-------|-------|--------|--------|--------------|--------|-------|-------|
| HMM | 9.09 | 24.71 | 5.77 | 26.32 | 83.84 | 15.22 | 79.86 | 54.87 | 42.55 | 36.00 |
| AE | 0.00 | 10.71 | 0.00 | 24.14 | 60.61 | 0.00 | 66.20 | 3.64 | 32.63 | 8.00 |
| AE y Gabor | 2.89 | 30.51 | 0.38 | 9.48 | 13.01 | 5.82 | 34.99 | 13.35 | 42.24 | 9.33 |
| GIST | 29.55 | 10.59 | 0.00 | 47.37 | 69.70 | 4.35 | 58.56 | 78.47 | 29.79 | 13.33 |
| SURF | 59.09 | 50.59 | 37.25 | 50.88 | 57.58 | 45.65 | 56.25 | 54.05 | 41.49 | 48.00 |

Cuadro 4.3. Eficiencia por artista para la máquina de vector soporte lineal

clasificación en la mayoría de los métodos de obtención de datos. Los artistas que en estas tres tablas tienen una menor eficiencia de clasificación son Frida Kahlo, María Izquierdo y José Clemente Orozco. El hecho que Kahlo e Izquierdo se confundan no es una novedad dentro del estudio del arte mexicano. Lo que sí es novedad es que estas dos artistas se confundan con Orozco. Aquí se vuelve a abrir otra oportunidad de hacer un estudio artístico a profundidad.

| | Atl | Carr | Frida | Izq | Mérida | Orozco | Ramos Mtz | Rivera | Tam | Varo |
|---------------|-------|-------|-------|-------|--------|--------|--------------|--------|-------|-------|
| HMM | 2.27 | 14.12 | 1.92 | 0.00 | 82.83 | 13.04 | 80.56 | 70.80 | 18.09 | 24.00 |
| AE | 0.00 | 0.00 | 0.00 | 0.00 | 57.58 | 0.00 | 88.03 | 1.82 | 20.00 | 1.33 |
| AE y Gabor | 45.95 | 50.00 | 49.11 | 49.35 | 49.69 | 53.41 | 100.00 | 50.23 | 49.74 | 51.75 |
| GIST | 40.91 | 36.47 | 11.76 | 61.40 | 82.83 | 19.57 | 79.28 | 88.89 | 51.06 | 50.67 |
| SURF | 15.91 | 34.12 | 5.88 | 19.30 | 58.59 | 19.57 | 81.94 | 63.06 | 32.98 | 16.00 |

Cuadro 4.4. Eficiencia por artista para la máquina de vector soporte radial

A continuación, en el cuadro 4.5, mostramos la matriz de confusión para el modelo que usa campos aleatorios de Markov para obtener los descriptores de imágenes y bosques aleatorios para hacer la clasificación. Mostramos esta matriz de confusión porque a continuación haremos un breve análisis de una obra usando este modelo. Con esta matriz nos podemos dar cuenta que la es diagonal dominante, por ende podemos decir que este modelo logra extraer cualidades estilísticas de cada artista. Con esta matriz podemos hacer otro tipo de lectura: podemos notar la cercanía de estilos entre los artistas. ¿Cómo realizar esta comparación? Si la entrada $a_{i,j}$ del artista i con el artista j es igual a cero entonces ambos artistas no comparten características similares según este modelo. Entre mayor sea el valor de

esta entrada, más similares son los estilos del artista i y del artista j .

| | Atl | Carr | Frida | Izq | Mérida | Orozco | Ramos Mtz | Rivera | Tam | Varo |
|--------------|-----|------|-------|-----|--------|--------|--------------|--------|-----|------|
| Atl | 41 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Carr | 0 | 78 | 0 | 0 | 0 | 0 | 2 | 4 | 1 | 0 |
| Frida | 0 | 1 | 48 | 0 | 0 | 0 | 0 | 2 | 0 | 1 |
| Izq | 0 | 0 | 0 | 55 | 1 | 0 | 0 | 1 | 0 | 0 |
| Mérida | 0 | 0 | 0 | 0 | 99 | 0 | 0 | 0 | 0 | 0 |
| Orozco | 0 | 1 | 0 | 1 | 1 | 41 | 0 | 1 | 1 | 0 |
| Ramos Mtz | 0 | 0 | 0 | 0 | 0 | 0 | 141 | 1 | 1 | 1 |
| Rivera | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 109 | 0 | 0 |
| Tam | 0 | 1 | 1 | 1 | 4 | 0 | 2 | 1 | 84 | 0 |
| Varo | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 70 |

Cuadro 4.5. Matriz de confusión para modelo con campos aleatorios de Markov junto con bosques aleatorios

4.4. Ejemplo de análisis

El objetivo de esta sección es dar un ejemplo de cómo podemos usar las obras de arte clasificadas erróneamente para hacer una comparación artística. Esto lo haremos viendo la obra que fue clasificada erróneamente y compararla con algunos ejemplos de los autores con la cual fue confundida. Analizaremos los colores empleados, la composición y temática. Vamos a usar el método de campos aleatorios de Markov junto con bosques aleatorios para realizar este análisis. Esto se debe a que el hecho de haber dividido las imágenes en secciones y “contar votos” nos brinda una oportunidad de análisis más profundo además que es el único método que trabaja con imágenes a color.

La imagen que fue clasificada mal con este método es la figura 4.11, titulada “Martes” por Leonora Carrington. Nuestro método obtuvo 13 votos a favor de Carrington, 2 a favor de Rivera y 25 a favor de Tamayo.



Figura 4.11. Martes, Leonora Carrington

A continuación hay una serie de obras de cada uno de los artistas en cuestión para que podamos hacer una comparación entre ellos y “Martes”, esto se muestra en las figuras 4.12, 4.13 y 4.14. Como podemos ver, “Martes” es una obra bastante más saturada que otras obras de Carrington, su composición es más rica. Así mismo, por los colores que usó no hay una clara separación de todos los elementos que se muestran en la imagen, como usualmente pasa con sus obras. En sus obras, como podemos ver, resaltan los personajes principales. Otro elemento diferente son los colores que usó en “Martes”, una paleta verde-rojiza. Las figuras en la obra de Tamayo constan de mucha textura, haciendo complicada su identificación. Tamayo también usa bastantes rojos, como sucede en “Martes”. Las figuras de “Martes” son alargadas, al igual que muchas de las que se muestran en la obra de Tamayo. Si comparamos “Martes” con la obra de Rivera podemos notar que también goza de una rica composición, sí hay elementos que destacan, pero no es tan minimalista como el resto de la obra de Carrington.

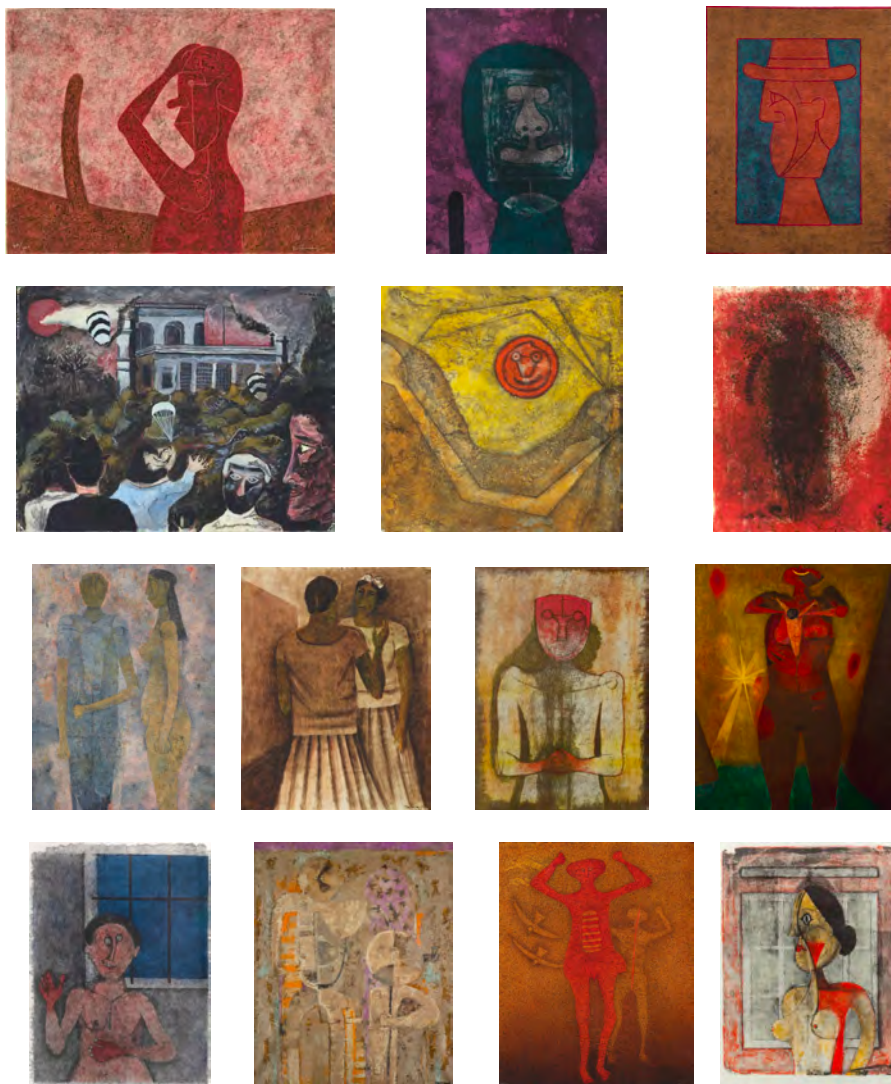


Figura 4.13. Rufino Tamayo

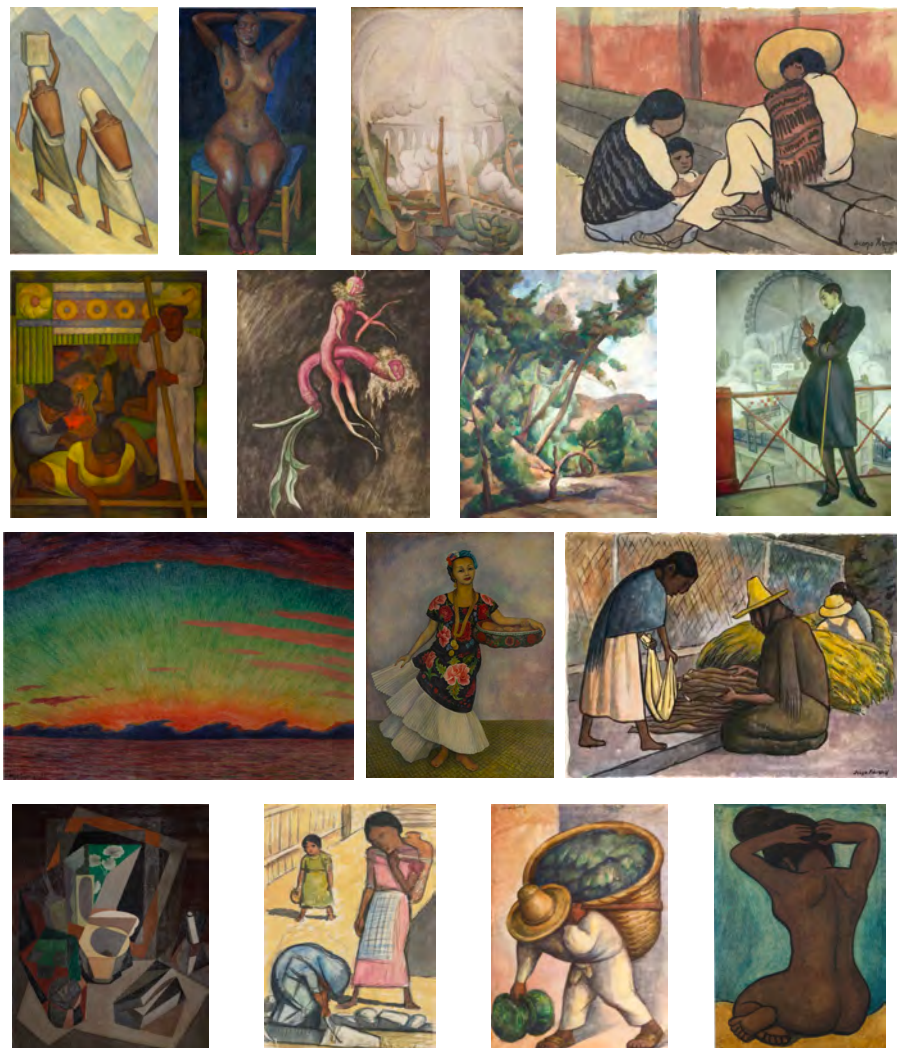


Figura 4.14. Diego Rivera

4.5. Conclusiones

A lo largo de este trabajo trabajamos con varios enfoques de las matemáticas: las puras, estadística, aprendizaje de máquina, visión por computadora. Logramos unir todas estas ramas con historia del arte. Esta solo es una muestra más de la bondad de las matemáticas y la gran diversidad de aplicaciones que puede tener. La separación entre matemáticas “aplicadas” y “puras” no está tan bien delimitada, siempre ha habido un ir y venir de ideas para que ambas sigan creciendo.

En particular, podemos notar que, al unir análisis de ondículas y de Fourier con técnicas de inteligencia artificial podemos lograr buenas clasificaciones de artistas. Más aún, aquellas clasificaciones erróneas hechas por los algoritmos nos pueden indicar similitud entre estilos de diferentes artistas. Esto tiene un sustento histórico así como estilístico y técnico del arte. El objetivo general de este trabajo se logra, al producir varios métodos para clasificar obras de arte respecto a su autor.

La eficiencia de cada uno de los métodos por artista nos da una medida sobre qué tan fácil es distinguir a un artista en particular de los otros. Esto es útil para la historia del arte mexicano pues es una forma de decir si el estilo de un artista destaca o es único dentro de su corriente. Por ejemplo, en el cuadro 4.2, la columna con mayor eficiencia en la mayoría de los métodos es la de Ramos Martínez. Esto nos quiere decir que el estilo de Ramos Martínez es fácil de distinguir del de los demás artistas aquí considerados. A diferencia de Orozco o Frida Kahlo, cuyos porcentajes de eficiencia en varios métodos es igual a cero. Esto nos quiere decir que, para esos métodos, no es posible identificar el estilo de estos artistas y por ende no podemos identificar sus cuadros adecuadamente.

También podemos llegar a usar dichos métodos no sólo para clasificar obras de arte por artista, sino para contrastar la similitud entre los autores aquí estudiados. Un ejemplo de esto es el cuadro 4.5. El número de clasificaciones erróneas por artistas (aquellos números

diferentes de cero que no se encuentran en la diagonal), indican un nivel de proximidad entre artistas. Podemos afirmar, por ejemplo, que las obras de Rufino Tamayo y Remedios Varo comparten características similares (pues la entrada del renglón de Varo y la columna de Tamayo es igual a dos). Al mismo tiempo podemos decir que las obras de Leonora Carrington y Remedios Varo también tienen cualidades en común, pero dicha similitud es menor que la de Rufino Tamayo y Remedios Varo (dicha entrada es igual a uno).

Tal vez se puedan diseñar ondículas específicamente con el propósito de clasificar obras de arte. Después de todo, el diseño de estas funciones no es aleatorio, tienen un propósito inicial y más adelante se usan en otras aplicaciones. A partir de este trabajo se pueden hacer varios análisis desde el punto de vista de la historia del arte mexicano. Dejamos abiertas estas cuestiones.

Bibliografía

- [1] G. P. Tolstov, *Fourier Series*. Dover Publications.
- [2] E. Bell, *Los grandes matemáticos*. Losada, 2009.
- [3] G. Grabinsky, *Teoría de la medida*. las prensas de ciencias, 2000.
- [4] W. Rudin, *Principles of Mathematical Analysis*. McGraw-Hill.
- [5] C. Meyer, *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM), 2000.
- [6] A. Haar, “On the theory of orthogonal function systems,” Ph.D. dissertation.
- [7] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics.
- [8] E. R. Kandel, J. H. Schwartz, J. T. M, S. A. Siegelbaum, and A. J. Hudspeth, *Principles of Neural Science*. McGraw Hill, 2012.
- [9] E. T. Rolls, *Computational Neuroscience of Vision*. Oxford University Press.
- [10] N. G. Kingsbury and I. W. Selesnick, “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine*, 2005.
- [11] C. C. Aggarwal, *Neural Networks and Deep Learning*. Springer, 2018.
- [12] D. Zhang, *Fundamentals of Image Data Mining*. Springer, 2019.

- [13] A. J. Izenman, *Modern Multivariate Statistical Techniques*, G. G. Casella, S. S. Fienberg, and I. Olkin, Eds. Springer.
- [14] H. Choi, J. Romberg, and R. Baraniuk, “Hidden markov tree modeling for complex wavelet transform.”
- [15] Q. Wang, “Hmrf-em-image,” MATLAB Central File Exchange. [Online]. Available: https://la.mathworks.com/matlabcentral/fileexchange/37530-hmrf-em-image?s_tid=srchtitle
- [16] C. C. Aggarwal, *Linear Algebra and Optimization for Machine Learning*. Springer, 2020.
- [17] J. Nocedal and S. J. Wright, *Numerical Optimization*, segunda ed., T. V. Mikosch, Sidney, I. Resnick, and S. M. Robinson, Eds. Springer.
- [18] W.-l. Chao, “Gabor wavelet transform and its application.”
- [19] A. Oliva and A. Torralba, “Modeling the shape of the scene: a holistic representation of the spatial envelope,” *International Journal of Computer Vision*, 2001.
- [20] —, “Modeling the shape of the scene: a holistic representation of the spatial envelope.” [Online]. Available: <https://people.csail.mit.edu/torralba/code/spatialenvelope/>
- [21] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features.”
- [22] D. Tyagi, “Introduction to surf (speeded-up robust features),” 2019. [Online]. Available: <https://medium.com/data-breach/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e>
- [23] —, “Introduction to sift(scale invariant feature transform),” 2019. [Online]. Available: <https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40>
- [24] “detectsurffeatures,” MATLAB Central File Exchange. [Online]. Available: <https://www.mathworks.com/help/vision/ref/detectsurffeatures.html>

- [25] “Obras con declaratoria de monumento artístico.” [Online]. Available: <https://inba.gob.mx/transparencia/obras>
- [26] L. López Orozco, “The revolution, vanguard artists and mural painting,” p. 14, 2014. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/09528822.2014.935566>
- [27] T. del Conde, *Historia mínima del arte mexicano en el siglo XX*. ATTAME.
- [28] J. Oles, *Arte y arquitectura en México*. Taurus, ch. Del Porfiriato a la Revolución (1880-1920).
- [29] L. L. Elizondo, Ed., *250 Artistas Mexicanos Siglos XIX, XX y XXI*. Promoción de Arte Mexicano.
- [30] “Catálogo de exposición: 1910, el arte en un año decisivo,” p. 27, 1989.
- [31] R. Tibol, *Frida Kahlo. Una vida abierta*. Oasis, 2007, ch. II. Aproximaciones.
- [32] Artsy. [Online]. Available: <https://www.artsy.net/>
- [33] Christie’s Auctions. [Online]. Available: https://www.christies.com/?sc_lang=en&lid=1
- [34] Casa de subastas Morton. [Online]. Available: <https://www.mortonsubastas.com/>
- [35] M. Gaxiola, “Fantásticas obras para recordar a dr. atl,” MXCity. [Online]. Available: <https://mxcity.mx/2017/10/obras-para-recordar-dr-atl/>
- [36] “Dr atl,” Inverarte. [Online]. Available: <https://inverarteartgallery.com/artist/dr-atl/>
- [37] “Dr atl el hechicero y pintor amante de los volcanes,” Fahrenheit Magazine. [Online]. Available: <https://fahrenheitmagazine.com/arte/dr-atl-el-hechicero-y-pintor-amante-de-los-volcanes>

- [38] “Selected paintings and drawings 1930s and 1940s,” The Alfredo Ramos Martínez Research Project. [Online]. Available: <https://www.alfredoramosmartinez.com/paintings-and-drawings>
- [39] Philadelphia Museum of Art. [Online]. Available: <https://philamuseum.org/>
- [40] Sotheby’s. [Online]. Available: <https://www.sothebys.com/en/>