UNIVERSITAT DE BARCELONA
Facultat de Matemàtiques i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

FOURIER ANALYSIS AND ITS APPLICATIONS IN IMAGE PROCESSING

Autor: Álvaro Recolons Simón

Director: Dr. Joan Carles Tatjer
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 10 de juny de 2024

Abstract

Fourier Analysis is a theory of pivotal relevance in many fields, as it allows any periodic function in a finite interval to be represented as a sum of sines and cosines. The Fourier Transform extends this concept to non-periodic functions by decomposing them into their frequency components. This paper aims to present the fundamentals of Fourier Analysis, covering the key properties and results of the Fourier Series and the Fourier Transform. Subsequently, the discrete version of the Fourier Transform, known as the Discrete Fourier Transform (DFT), will be discussed. Additionally, we will examine the correct methods for sampling continuous signals, addressing issues of sampling and aliasing. The paper will then introduce the groundbreaking work of Cooley and Tukey (1965) on the Fast Fourier Transform (FFT), an algorithm that reduces the computational cost of the DFT from $O(N^2)$ to $O(N \log N)$. Finally, the application of Fourier Analysis in the field of image processing will be explored, demonstrating its practical significance and versatility.

Agradecimientos

Primero de todo, quiero dar gracias a mi familia. A mis padres, a Cris, a Ale y a Carol. Gracias por su comprensión y su paciencia, especialmente en época de exámenes y los nervios que conllevan.

Gracias también a todos los profesores que me han acompañado durante mi etapa formativa, especialmente a Alberto Delgado y a Pepe Rodríguez. Sin sus consejos, no estaría aquí; de hecho, quizá estaría en la facultad de psicología ahora mismo. Agradezco también a todos mis amigos, en particular a Pepe, gran wingman y aún mejor persona.

Finalmente, quiero expresar mi profundo agradecimiento a Joan Carles Tatjer, quien me ha acompañado durante la elaboración de este trabajo y me ha ayudado con sus revisiones y reuniones semanales.

Contents

Introduction

Fourier Analysis has its roots in the early 19th century. The concept was pioneered by the French mathematician and physicist Joseph Fourier. Fourier's groundbreaking work was published in 1822, in a book titled "Théorie Analytique de la Chaleur". His work, however, had been complete since 1807 but was not published due to a harsh reception. He proposed that any periodic function could be expressed as a sum of sines and cosines, now known as the Fourier series. That is,

$$f(\theta) = \sum_{-\infty}^{\infty} c_n e^{in\theta}.$$

Fourier's ideas were motivated by the search for a solution to the heat equation, a partial differential equation. Before Fourier's work, a general solution to this equation was not known. The particular solution for when the heat source behaved sinusoidally was understood, but Fourier's innovation was to model a complex heat source as a superposition of sine and cosine waves. This superposition is now known as the Fourier series. Fourier's discussion on this topic was rather unrigorous by today's standards. However, later mathematicians Dirichlet and Riemann reformulated Fourier's results with greater precision and formality. Despite the initial skepticism surrounding Fourier's ideas, his theory gradually gained acceptance and has since become a cornerstone of mathematical analysis.

The Fourier series is a powerful tool for solving problems related to heat conduction, vibration, and acoustics. Over time, the theory expanded to include the Fourier transform, which enables the decomposition of non-periodic functions into their frequency components. This generalization has profound applications in various fields, such as signal processing, quantum mechanics, and image processing.

The significance of Fourier's work is immense, as it laid the foundation for modern harmonic analysis. His methods continue to influence contemporary mathematics, physics, and engineering disciplines. Fourier's work exemplifies how innovative thinking can drive remarkable advancements in science and technology.

This work aims to cover the fundamentals of Fourier analysis and provide the reader with an in-depth understanding of its applications in the field of image processing.

In the first chapter, we will explore the construction of the Fourier series and the conditions under which it converges to the target function. We will demonstrate that this extension is possible because the set $\{e^{int}\}_{n\in\mathbb{Z}}$ is dense in $L^2[-\pi, \pi]$. While some results in this chapter may have been covered in previous coursework, this work offers new insights into the Fourier series. We will then introduce the Fourier transform of a function

$$\hat{f}(t) = \int_{-\infty}^{\infty} f(x)e^{-ixt}\, dx,$$

initially defined for functions in L^1. We will discuss its relationship with convolution, the inversion theorem, and its extension to L^2 and multivariable spaces. This section primarily follows the structure of the works of Folland [3] and Rudin [4], [5].

The second chapter focuses on Discrete Fourier Analysis. It will begin by presenting the Nyquist-Shannon Sampling Theorem [11], a crucial result that explains how to sample continuous signals using a discrete set of samples. Following this, we will introduce the Discrete Fourier Transform (DFT) and its properties, drawing parallels to the continuous Fourier Transform. Much of this chapter is based on the works of Cheney [7], Bredies [12], and Brunton [13].

A significant limitation of the DFT is its high computational cost when computed directly, with a complexity of $O(N^2)$. In the final part of the second chapter, we will present the Fast Fourier Transform (FFT) algorithm. This method for efficiently calculating the DFT was first published by James Cooley and John Tukey [8] in 1965 and reduces the computation cost to $O(N \log N)$.

> *"The most important Algorithm of our Lifetime."*
>
> – Gilbert Strang, 1994, on the FFT

However, it was originally used by Gauss in an unpublished work from 1805, where he aimed to interpolate the orbits of the asteroids Pallas and Juno from a sample of observations. Remarkably, this work predates Fourier's breakthrough by almost 20 years!

The last chapter is dedicated to presenting an application of the concepts of Fourier analysis in the field of image processing. We will begin by explaining how an image is defined and exploring its representation in the frequency domain. Following this, we will discuss various filtering methods in the frequency domain. All the code used to generate the images and the filters is self-made and is provided along with this document.

2020 Mathematics Subject Classification. 42A16, 42A20, 42A38, 65T50, 94A08

Chapter 1

Fourier Analysis Fundamentals

1.1 Basic definitions and properties

Definition 1.1. Let X be an arbitrary interval of $\mathbb{R}$. If $1 \leq p < \infty$ and f be a complex measurable function on X, we define

$$\|f\|_p = \left\{ \int_X |f(t)|^p \, dt \right\}^{1/p}.$$

Let $L^p(X)$ consist of all f such that

$$\|f\|_p < \infty.$$

We call $\|f\|_p$ the L^p*-norm* of f. If $X = (-\infty, \infty)$, we will write L^p instead of $L^p(-\infty, \infty)$.

Theorem 1.2. *Let $p \in (1, \infty)$. Let X be an interval of $\mathbb{R}$. Let f and g be measurable, non-negative functions in X. Then,*

$$\left\{ \int_X (f(p) + g(p))^p dt \right\}^{1/p} \leq \left\{ \int_X f^p dt \right\}^{1/p} + \left\{ \int_X g^p dt \right\}^{1/p}. \tag{1.1}$$

This is known as Minkowski's inequality.

Proof. The proof can be found in [4]. □

Theorem 1.3. *Let X be an arbitrary interval of $\mathbb{R}$. If $1 \leq p < \infty$, and $f, g \in L^p(X)$ then $f + g \in L^p(X)$, and*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. Let $p \in (1, \infty)$. We have that,

$$\|f+g\|_p = \left\{ \int_X |f(t)+g(t)|^p \, dt \right\}^{1/p}$$
$$\leq \left\{ \int_X (|f(t)|+|g(t)|)^p \, dt \right\}^{1/p}$$
$$\leq \left\{ \int_X |f(t)|^p \, dt \right\}^{1/p} + \left\{ \int_X |g(t)|^p \, dt \right\}^{1/p} = \|f\|_p + \|g\|_p.$$

The first inequality is obvious and in the second inequality, we have used Minkowski's inequality. For $p = 1$, it is trivial since $|f+g| \leq |f| + |g|$. □

Since we have a norm in L^p, it is natural to assume that we will get a metric space with the distance defined as $d(f,g) = \|f-g\|_p$. This is, however, not entirely true since $d(f,g) = 0$ does not imply that $f = g$, but that $f(t) = g(t)$ for almost all t, i.e., for all t except for a set of Lebesgue measure zero (see Theorem 2.22 of [4]).

Let us write $f \sim g \Leftrightarrow d(f,g) = 0$. This defines an equivalence relation in L^p. When L^p is regarded as a metric space, the space that is really under consideration is not the space whose elements are functions, but a space whose elements are equivalence classes of functions. For the sake of simplicity, although this distinction we will speak of L^p as a space of functions.

Definition 1.4. We say that a metric space X is *complete* if every Cauchy sequence in X converges to an element of X.

Theorem 1.5. *L^p is a complete metric space for every $1 \leq p < \infty$.*

Proof. The proof of this result can be found in Theorem 3.11. of [4]. □

1.2 The Fourier Series

1.2.1 Definition

Definition 1.6. Suppose f is 2π-periodic and integrable over $[-\pi, \pi]$. The numbers c_n defined by

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} \, d\theta \tag{1.2}$$

are called the *Fourier coefficients* of f, and the corresponding formal series

$$S^f = \sum_{-\infty}^{\infty} c_n e^{in\theta} \tag{1.3}$$

is called the *Fourier series* of f.

Remark 1.7. If f is a real function, then clearly $f(\theta) = \overline{f(\theta)}$. We have then that

$$\overline{c_n} = \overline{\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{f(\theta)}e^{-in\theta} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta} = c_{-n}$$

Remark 1.8. By using Euler's formula

$$e^{i\theta} = cos(\theta) + isin(\theta)$$

we obtain an analogous expression for the Fourier series and the Fourier coefficients:

$$S^f = \frac{1}{2}a_0 + \sum_{1}^{\infty}(a_n cos(n\theta) + b_n sin(b\theta)),$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)cos(n\theta)d\theta, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)sin(n\theta)d\theta.$$

The real and complex coefficients are related by:

$$c_n = \frac{1}{2}(a_n - ib_n)$$
$$c_{-n} = \frac{1}{2}(a_n + ib_n)$$

with $n \in \mathbb{N}$.

Remark 1.9. All the definitions in this section could be extended to integrable functions over an arbitrary interval $[a, b]$. If $L = |b - a|$ then

$$c_n = \frac{1}{L}\int_{a}^{b} f(\theta)e^{\frac{-2in\theta}{L}}\, d\theta, \tag{1.4}$$

$$S^f = \sum_{-\infty}^{\infty} c_n e^{\frac{2in\theta}{L}}.$$

We will, however, treat the case where $[a, b] = [-\pi, \pi]$ for simplicity in the notation and for the fact that it is a widely accepted convention in the field of Fourier Analysis.

1.2.2 A basis for 2π-periodic functions

Definition 1.10. We say that a complex vector space H is an *inner product space* if for each pair of vector $x, y \in H$ there is an associated complex number (x, y) which we call the *inner product* of x and y, such that the following rules hold:

1. $(y, x) = \overline{(x, y)}$.

2. $(x+y,z)=(x,z)+(y,z)$, with $x,y,z\in H$.
3. $(\alpha x,y)=\alpha(x,y)$ where α is scalar.
4. $(x,x)\geq 0, \forall x\in H$
5. $(x,x)=0\Leftrightarrow x=0$

It is generally known that every inner product space H is also a *normed space* with the norm of a given vector $u\in H$ defined as

$$\|u\|=\sqrt{(u,u)}.$$

Definition 1.11. A system $\{\varphi_j\}_{j=1}^{\infty}$ with $\varphi_j\in H$, $\forall j$, is said to be *complete* in H if, for every, $u\in H$ and every $\varepsilon>0$, there is a linear combination $\sum_{j=1}^{N}a_j\varphi_j$ such that

$$\left\|u-\sum_{j=1}^{N}a_j\varphi_j\right\|<\varepsilon.$$

Definition 1.12. If an inner product space H is complete, it is called a *Hilbert Space*.

Remark 1.13. Consider L^2 with the inner product

$$(f,g)=\int_{-\infty}^{\infty}f(t)\,\overline{g}(t)\,dt. \tag{1.5}$$

This inner product is well defined as shown in theorem 3.8. of [4] and the completeness of L^2 is proven in theorem 3.11. of [4]. Thus L^2 is a Hilbert space. Note also that,

$$\|f\|=(f,f)^{\frac{1}{2}}=(\int_{-\infty}^{\infty}|f|^2\,dx)^{\frac{1}{2}}=\|f\|_2.$$

Definition 1.14. Given a Hilbert Space H and a set of vectors of this space $\{u_i\}_{i\in I}$ is said to be *orthonormal* if we have

$$(u_i,u_j)=\begin{cases}1, & i=j\\ 0, & \text{otherwise}\end{cases}\qquad \forall\, i,j\in I.$$

If an orthonormal set is complete, we call it an *orthonormal basis*.

Theorem 1.15. *Let $\{u_i\}_{i\in I}$ be an orthonormal set in H. The following are equivalent*

1. *The set S of all finite linear combinations of members of $\{u_i\}_{i\in I}$ is dense in H.*
2. *For every $x\in H$, we have $\|x\|^2=\sum_{i\in I}|(x,u_i)|^2$.*

3. *If $x \in H$ and $y \in H$, then $(x, y) = \sum_{i \in I}(x, u_i)\overline{(y, u_i)}$.*

Proof. The proof can be found in Rudin [4] 4.18. □

Definition 1.16. We define the *Fourier System* as the set $\{e^{int}\}_{n \in \mathbb{Z}}$.

Proposition 1.17. *The Fourier system is an orthonormal set on the interval $[-\pi, \pi]$.*

Proof. Obviously, the exponential function is on $L^2[-\pi, \pi]$. We then can consider the inner product 1.5 with integration limits $-\pi$ and π and multiplied by a factor of $1/2\pi$:

$$(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\overline{g(t)}\, dt$$

If $u_k = e^{ikt}$ for all k, we see that

$$(u_n, u_m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{int} e^{-imt}\, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)t}\, dt = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases} .$$

□

Theorem 1.18. *The Fourier system is complete in $L^2[-\pi, \pi]$.*

Proof. The proof can be found in theorem 5.8. of Vretbald [2]. □

We have shown that the Fourier system is orthonormal and complete in $L^2[-\pi, \pi]$, therefore, by Theorem 1.15 we determine that it is dense in this space. We get the idea now that the Fourier coefficients of a function $f \in L^2[-\pi, \pi]$ are in a way the "coordinates" of f in the base $\{e^{int}\}_{n \in \mathbb{Z}}$. The Fourier series is the expansion of f with respect to the Fourier system.

1.2.3 Convergence of Fourier Series

We want to study the convergence of the Fourier series to our function f. We denote the Nth partial sum of the Fourier series by

$$S_N^f = \sum_{-N}^{N} c_n e^{in\theta}.$$

We want to study under what conditions the succession of partial sums converge $S_N^f \underset{N \to \infty}{\longrightarrow} f$ (pointwise or uniformly).

Pointwise convergence

Theorem 1.19. ***(Bessel's Inequality)*** *If f is 2π-periodic and Lebesgue integrable on $[-\pi, \pi]$, and the Fourier coefficients c_n are defined like in 1.2 then*

$$\sum_{n=-\infty}^{\infty} |c_n|^2 \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)|^2 \, d\theta.$$

Proof. Since $|z|^2 = z\overline{z}$ for any complex number z,

$$\left| f(\theta) - \sum_{n=-N}^{N} c_n e^{in\theta} \right|^2 = \left(f(\theta) - \sum_{n=-N}^{N} c_n e^{in\theta} \right) \left(\overline{f(\theta)} - \sum_{n=-N}^{N} \overline{c_n} e^{-in\theta} \right)$$
$$= |f(\theta)|^2 - \sum_{n=-N}^{N} \left[c_n \overline{f(\theta)} e^{in\theta} + \overline{c_n} f(\theta) e^{-in\theta} \right] + \sum_{m,n=-N}^{N} c_m \overline{c_n} e^{i(m-n)\theta}.$$

If we divide both sides of the equality by 2π and integrate from $-\pi$ to π we obtain

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| f(\theta) - \sum_{n=-N}^{N} c_n e^{in\theta} \right|^2 d\theta$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)|^2 d\theta - \sum_{n=-N}^{N} [c_n \int_{-\pi}^{\pi} \overline{f(\theta)} e^{in\theta} \, d\theta + \overline{c_n} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} \, d\theta] +$$
$$\sum_{n=-N}^{N} c_n \overline{c_n} \int_{-\pi}^{\pi} e^{i(m-n)\theta} \, d\theta$$

Trivially, if $m = n$ then $\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)\theta} d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 \, d\theta = 1$. Because of the 2π-periodicity of the complex exponential function, we have that if $m \neq n$ then $\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)\theta} d\theta = 0$. We have the formula

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)\theta} d\theta = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases}. \tag{1}$$

Now if we recall,

$$c_n = \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} \, d\theta, \tag{2}$$

we obtain

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|f(\theta)-\sum_{n=-N}^{N}c_ne^{in\theta}\right|^2 d\theta\\
&=\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(\theta)|^2 d\theta-\sum_{n=-N}^{N}[c_n\overline{c_n}+\overline{c_n}c_n]+\sum_{n=-N}^{N}c_n\overline{c_n}\\
&=\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(\theta)|^2 d\theta-\sum_{n=-N}^{N}|c_n|^2.
\end{aligned}$$

But the integral on the left cannot be negative, therefore

$$0\leq\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(\theta)|^2\,d\theta-\sum_{n=-N}^{N}|c_n|^2.$$

If we let $N\to\infty$, we obtain the desired result. □

Remark 1.20. Bessel's inequality can also be deducted from Theorem1.15. It can be proved that Bessel's Inequality is, in fact, an equality. However, our main goal with this demonstration was to see that the Fourier coefficients a_n, b_n, and c_n tend to zero as $n\to\infty$. This is because $|a_n^2|$, $|b_n^2|$, and $|c_n^2|$ are all terms of a convergent series, and therefore tend to zero as $n\to\infty$. Hence so do a_n, b_n, and c_n. This is a special case of Theorem 1.37.

Definition 1.21. A function f defined over a closed interval $[a,b]$ is said to be *piecewise continuous* if f is continuous except for a finite number of jump discontinuities.
We say a function f is *piecewise smooth* if f and its first derivative f' are both piecewise continuous on $[a,b]$.

Lemma 1.22. *Given a continuous, derivable, and P-periodic function f, then $\int_a^{a+P} f(x)\,dx$ is independent of a.*

Proof. Let

$$g(a)=\int_a^{a+P}f(x)\,dx=\int_0^{a+P}f(x)\,dx-\int_0^a f(x)\,dx.$$

By the fundamental theorem of calculus, $g'(a)=f(a+P)-f(a)$, and given the periodicity of f we see that g is constant. □

Definition 1.23. We call the function

$$D_n(\theta)=\frac{1}{2\pi}\sum_{-N}^{N}e^{in\theta} \tag{1.6}$$

the *Nth Dirichlet Kernel.*

Lemma 1.24. *Consider the Nth Dirichlet kernel defined as in 1.6. For any N we have,*

$$\int_{-\pi}^{0} D_N(\theta)\,d\theta = \int_{0}^{\pi} D_N(\theta)\,d\theta = \frac{1}{2}.$$

Proof. We have that

$$D_n(\theta) = \frac{1}{2\pi}\sum_{n=-N}^{N} e^{in\theta} = \frac{1}{2\pi}\left(1 + \sum_{n=1}^{N}(e^{in\theta} + e^{-in\theta})\right)$$
$$= \frac{1}{2\pi}\left(1 + \sum_{n=1}^{N} 2\cos(n\theta)\right) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{N}\cos(n\theta)$$

Where we have used that $e^{in\theta} = cos(n\theta) + isin(n\theta)$ and $e^{-in\theta} = \cos(n\theta) - i\sin(n\theta)$. We have then,

$$\int_{0}^{\pi} D_N(\theta)\,d\theta = \left[\frac{\theta}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{N}\frac{\sin n\theta}{n}\right]_0^{\pi} = \frac{1}{2}.$$

The case where the integral goes from $-\pi$ to 0 is analogous. □

Remark 1.25. We have that

$$D_n(\theta) = \frac{1}{2\pi}\sum_{n=-N}^{N} e^{in\theta} = \frac{1}{2\pi}e^{-iN\theta}\sum_{n=0}^{2N} e^{in\theta}.$$

Applying the geometric series formula $\sum_{n=0}^{N} r^n = \frac{r^{N+1}-1}{r-1}$ we have that

$$D_n(\theta) = \frac{1}{2\pi}e^{-iN\theta}\frac{e^{i(2N+1)\theta}-1}{e^{i\theta}-1} = \frac{1}{2\pi}\frac{e^{i(N+1)\theta}-1}{e^{i\theta}-1}.$$

Finally, if we multiply the top and bottom by $e^{-i\theta/2}$, we obtain the expression

$$D_n(\theta) = \frac{1}{2\pi}\frac{e^{i(N+\frac{1}{2})\theta} - e^{-i(N+\frac{1}{2})\theta}}{e^{i\frac{1}{2}\theta} - e^{-i\frac{1}{2}\theta}} = \frac{1}{2\pi}\frac{sin(N+\frac{1}{2})\theta}{sin(\frac{1}{2}\theta)} \tag{1.7}$$

Theorem 1.26. ***(Pointwise convergence)****. If f is a 2π-periodic function piecewise smooth on $\mathbb{R}$, and S_N^f is defined as above, then*

$$\lim_{N\to\infty} S_N^f(\theta) = \frac{1}{2}\left[f(\theta-) + f(\theta+)\right]$$

for every θ. In particular, $\lim_{N\to\infty} S_N^f(\theta) = f(\theta)$ for all θ where f is continuous.

Proof. We have that

$$S_N^f(\theta) = \frac{1}{2\pi}\sum_{-N}^{N} c_n e^{in\theta} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\psi)e^{in(\theta-\psi)}\,d\psi = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\psi)e^{in(\psi-\theta)}\,d\psi$$

where we have plugged the formula of c_n into the formula of S_N^f and replaced n with $-n$ in the last equality (it doesn't affect the result since the sum goes from $-N$ to N). Now if we make the change of variable $\tau = \psi - \theta$ and use the Lemma 1.22 we get

$$S_N^f(\theta) = \frac{1}{2\pi}\sum_{-N}^{N}\int_{-\pi+\theta}^{\pi+\theta} f(\theta+\tau)e^{in\tau}\,d\tau = \frac{1}{2\pi}\sum_{-N}^{N}\int_{-\pi}^{\pi} f(\theta+\tau)e^{in\tau}\,d\tau$$

In short,

$$S_n^f(\theta) = \int_{-\pi}^{\pi} f(\theta+\tau)D_N(\tau)d\tau \tag{1}$$

By Lemma 1.24, we have

$$\frac{1}{2}f(\theta-) = f(\theta-)\int_{-\pi}^{0} D_N(\tau)\,d\tau, \qquad \frac{1}{2}f(\theta+) = f(\theta+)\int_{0}^{\pi} D_N(\tau)\,d\tau, \tag{2}$$

therefore by (1) and (2) we have that

$$S_N^f(\theta) - \frac{1}{2}\left[f(\theta-)+f(\theta+)\right] =$$
$$\int_{-\pi}^{0}\left[f(\theta+\tau)-f(\theta-)\right]D_N(\tau)\,d\tau + \int_{0}^{\pi}\left[f(\theta+\tau)-f(\theta+)\right]D_N(\tau)\,d\tau.$$

We wish to show that for each fixed θ, this quantity approaches zero as $N \to \infty$. By the formula 1.7, we can write it as

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\theta)(e^{i(N+1)\theta} - e^{-iN\theta})\,d\theta \tag{3}$$

where g is defined as

$$g(\theta) = \begin{cases} \frac{f(\theta-\tau)-f(\theta-)}{e^{i\tau}-1} & \text{if } -\pi < \tau < 0, \\ \frac{f(\theta-\tau)-f(\theta+)}{e^{i\tau}-1} & \text{if } 0 < \tau < \pi. \end{cases}$$

g behaves well on the interval $[-\pi,\pi]$ except near $\tau = 0$. However, it is easy to check that by l'Hôpital rule that $\lim_{\tau\to 0+} g(\tau) = \frac{f'(\theta+)}{i}$. Similarly, $g(\tau)$ approaches the limit $f'(\theta-)/i$ as τ approaches zero from the left. Therefore g is piecewise continuous on $[-\pi,\pi]$, and its Fourier coefficients

$$C_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\tau)e^{-un\tau}\,d\tau$$

tend to zero as $n \to \pm\infty$ (as shown by Bessel's inequality). But expression (3) is nothing but $C_{-(N+1)} - C_N$, thus it vanishes at $N \to \infty$. This is all we need to show. □

Uniform convergence

Theorem 1.27. *Suppose f is 2π-periodic, continuous, and piecewise smooth. Let a_n, b_n, and c_n be the Fourier coefficients of f, and let a'_n, b'_n and c'_n be the corresponding Fourier coefficients of f'. Then*

$$a'_n = nb_n, \qquad b'_n = -na_n, \qquad c'_n = inc_n.$$

Proof. We have that $a'_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)cos(nx)dx$. If we integrate by parts taking $u = cos(nx)$ and $dv = f'(x)$, then $du = -nsin(nx)dx$ and $v = f(x)$. We have then,

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\cos(nx)dx = \frac{1}{\pi}\left[f(x)\cos(nx)\Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f(x)(-n\sin(nx))\,dx\right].$$

Since f is 2π-periodic and so is $cos(n\pi)$ we have that the first term cancels out and then

$$a'_n = \frac{n}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\,dx = nb_n.$$

The second equality is analogous if we choose $u = sin(nx)$ and $dv = f'(x)$. For the third equality,

$$\begin{aligned} c'_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)e^{-inx\,dx} \\ &= \frac{1}{2\pi}\left[f(x)e^{-inx}\Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f(x)(-ine^{inx})\,dx\right] \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)e^{-inx}\,dx = \frac{in}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx = inc_n. \end{aligned}$$

□

Theorem 1.28. *If f is 2π-periodic, continuous, and piecewise smooth, then the Fourier series converges to f absolutely and uniformly in $\mathbb{R}$.*

Proof. By theorem 1.26, we only have to show that $\sum_{-\infty}^{\infty} |c_n|$ converges. Let c'_n denote the Fourier coefficients of f'. In theorem 1.27 we have shown that $c_n = (in)^{-1}c'_n$ for $n \neq 0$, and applying Bessel's inequality to f',

$$\sum_{n=-\infty}^{\infty} |c'_n|^2 \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f'(\theta)|^2 d\theta < \infty.$$

We see then that $\sum_{n=-\infty}^{\infty} |c'_n|^2$ is a convergent series; in other words, the sequence $\{c'_n\}_n$ is square-summable.

Consider the Euclidean space $\mathbb{R}^{2N+1}$. A sequence can be written as a vector:

$$x = (x_{-N}, x_{-N+1}, \ldots, x_0, \ldots, x_{N-1}, x_N)$$

and the inner product can be defined as

$$x \cdot y = \sum_{k=-N}^{N} x_k y_k.$$

Given that $\mathbb{R}^{2N+1}$ is an inner product space, the Cauchy-Schwartz inequality holds i.e. $|x \cdot y| \leq \|x\|^{1/2}\|y\|^{1/2}$. This argument can be extended to the case of square summable series. Suppose that $x = \{x_n\}_{n\in\mathbb{Z}}$ and $y = \{y_n\}_{n\in\mathbb{Z}}$ are both infinite sequences such that $\sum_{n\in\mathbb{Z}} x_n^2 < \infty$ and $\sum_{n\in\mathbb{Z}} y_n^2 < \infty$. Then we have that

$$\sum_{n=-N}^{N} |x_n y_n| \leq \left(\sum_{n=-N}^{N} x_n^2\right)^{1/2} \left(\sum_{n=-N}^{N} y_n^2\right)^{1/2} \leq \left(\sum_{-\infty}^{\infty} x_n^2\right)^{1/2} \left(\sum_{-\infty}^{\infty} y_n^2\right)^{1/2} < \infty$$

Then if we make $n \to \infty$, we see that $\sum_{-\infty}^{\infty} |x_n y_n| < \infty$.

Consider now the sum $\sum_{-\infty}^{\infty} |c_n|$. We have that

$$\sum_{-\infty}^{\infty} |c_n| = |c_0| + \sum_{n\neq 0} \left|\frac{c_n'}{n}\right| \leq |c_0| + \left(\sum_{n\neq 0} \frac{1}{n^2}\right)^{1/2} \left(\sum_{n\neq 0} |c_n'|^2\right)^{1/2} < \infty.$$

We have used the fact that $\{c_n'\}$ and $\{\frac{1}{n}\}_{n\neq 0}$ are both square summable series. □

Remark 1.29. Fourier series have some intrinsic limitations when dealing with discontinuous functions. Consider a periodic function f that exhibits a discontinuous jump at a point x_0. In such cases, the Fourier series for f cannot achieve uniform convergence across any interval containing x_0. The lack of uniformity manifests itself in a dramatic way known as the **Gibbs phenomenon** which is characterized by the oscillatory nature of the series near the point of discontinuity. As one adds terms to the series, the partial sums overshoot and undershoot f near the discontinuity, and thus develop 'spikes' that tend to zero in width but not in length. For more detailed information on the Gibbs phenomenon see Plonka [6].

1.3 The Fourier Transform

We will begin this section by introducing the concept of convolution between two functions. This operation is deeply intertwined with the Fourier transform, as the Fourier transform of a convolution of two functions is the pointwise product of their Fourier transforms. This powerful relationship underlies many important results and applications in Fourier analysis.

Definition 1.30. If f and g are functions on $\mathbb{R}$, their *convolution* is the function $f * g$ defined by

$$f * g(x) = \int_{-\infty}^{\infty} f(x-y)g(y)\,dy \tag{1.8}$$

provided that this integral exists.

Theorem 1.31. *Consider $f, g \in L^1$. The convolution of these two functions, assuming that it exists, obeys the same algebraic laws as ordinary multiplication:*

1. $f * (ag + bh) = a(f * g) + b(f * h)$ *for any* $a, b \in \mathbb{R}$
2. $f * g = g * f$
3. $f * (g * h) = (f * g) * h$

Proof. If f and g are both in L^1, then the convolution $f * g(x)$ exists for almost every x as shown in section 7.1. of Folland [3].
(1) is obvious given that integration is a linear operation. For (2) if we make the change of variable $z = x - y$:

$$f * g(x) = \int_{\mathbb{R}} f(x-y)g(y)\,dy = \int_{\mathbb{R}} f(z)g(x-z)\,dz$$

For (3) we use (2) and interchange the order of integration:

$$(f * g) * h(x) = \int_{\mathbb{R}} f * g(x-y)h(y)\,dy = \int_{\mathbb{R}}\int_{\mathbb{R}} f(z)g(x-y-z)h(y)\,dz\,dy$$
$$= \int_{\mathbb{R}}\int_{\mathbb{R}} f(z)g(x-z-y)h(y)\,dy\,dz = \int_{\mathbb{R}} f(z)g * h(x-z)\,dz = f * (g * h)(x).$$

□

Theorem 1.32. *Suppose that f is differentiable and the convolutions $f * g$ and $f' * g$ are well-defined. Then $f * g$ is differentiable and $(f * g)' = f' * g$. Likewise, if g is differentiable, then $(f * g)' = f * g'$.*

Proof. We can think of f' as a limit of functions $\{f_n\}_n$ and apply the dominated convergence theorem to differentiate under the integral sign (see Rudin [4], [5] or Folland [3]):

$$(f * g)'(x) = \frac{d}{dx}\int_{\mathbb{R}} f(x-y)g(y)\,dy = \int_{\mathbb{R}} \frac{d}{dx} f(x-y)g(y)\,dy = f' * g(x).$$

Since $f * g = g * f$, the same argument works with f and g interchanged. □

Definition 1.33. Given $f \in L^1$, the function

$$\hat{f}(t) = \int_{-\infty}^{\infty} f(x)e^{-ixt}\,dx$$

is well defined $\forall t \in \mathbb{R}$. We call $\hat{f}$ the *Fourier transform* of f.

Remark 1.34. Sometimes we will refer to the Fourier transform with the notation

$$\mathcal{F}(f) = \hat{f}.$$

Remark 1.35. Since e^{-itx} has module value 1, the integral converges absolutely for all t and defines a bounded function of t:

$$|\hat{f}(t)| \leq \int_{\mathbb{R}} |f(x)|\,dx$$

Moreover, since $|e^{-i\eta x}f(x) - e^{-itx}f(x)| \leq 2|f(x)|$, the dominated convergence theorem implies that $\hat{f}(\eta) - \hat{f}(t) \to 0$ as $\eta \to t$, therefore $\hat{f}$ is continuous.

Theorem 1.36. *Suppose $f \in L^1$ and α and λ are real numbers.*

1. *If $g(x) = f(x)e^{i\alpha x}$, then $\hat{g}(t) = \hat{f}(t-\alpha)$.*
2. *If $g(x) = f(x-\alpha)$, then $\hat{g}(t) = \hat{f}(t)e^{-i\alpha t}$.*
3. *If $g \in L^1$ and $h = f * g$, then $\hat{h}(t) = \hat{f}(t)\hat{g}(t)$ (Convolution Theorem).*
4. *If f is continuous and piecewise smooth and $f' \in L^1$ then $\mathcal{F}(f')(x) = ix\hat{f}(x)$. On the other hand, if $xf(x)$ is integrable then $\widehat{xf}(x) = i\hat{f}'(x)$.*

Proof. For 1. we have that,

$$\hat{g}(t) = \int_{-\infty}^{\infty} f(x)e^{i\alpha x}e^{-ixt}\,dx = \int_{-\infty}^{\infty} f(x)e^{ix(\alpha-t)}\,dx = \hat{f}(t-\alpha).$$

Regarding 2., if we consider the change of variable $y = x - \alpha$ we have,

$$\hat{g}(t) = \int_{-\infty}^{\infty} f(x-\alpha)e^{-ixt}dx = \int_{-\infty}^{\infty} f(y)e^{-i(y-\alpha)t}\,dy = e^{-i\alpha t}\int_{-\infty}^{\infty} f(y)e^{-iyt}\,dy = \hat{f}(t)e^{-i\alpha t}.$$

Now for the proof of 3., if we consider the change of variable $z = x - y$ then:

$$\begin{aligned}
\hat{h}(t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-itx}f(x-y)g(y)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-it(x-y)}f(x-y)e^{-ity}g(y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-itz}f(z)e^{-ity}g(y)dzdy \\
&= \hat{g}(t)\hat{f}(t).
\end{aligned}$$

Where in the last equality we have separated the two integrals (this move is legitimate. In Rudin, [4] theorem 7.14 it is shown that if f and g are functions of L^1, we can apply Fubini's theorem. Furthermore, theorem 7.8 in Rudin [4] provides the proof of a rather general version of Fubini's theorem).
In regards to 4., since f is derivable in $\mathbb{R}$, by the Fundamental Theorem of calculus we have that $f(x) - f(0) = \int_0^x f'(y)\,dy$. If we make $x \to \infty$ we obtain $\lim_{x\to\infty} f(x) - f(0) = \int_0^\infty f'(x)\,dx$. Thus we have that

$$\lim_{x\to\infty} f(x) = f(0) + \int_0^\infty f'(x)\,dx$$

Since $f' \in L^1$ this limit exists. And since $f \in L^1$ (and therefore vanishes at infinity) the limit must be zero. Likewise, $\lim_{x\to-\infty} f(x) = 0$. Hence we can integrate by parts, and the boundary terms vanish:

$$\mathcal{F}(f')(t) = \int_{-\infty}^{\infty} e^{-ixt} f'(x)\,dx = -\int_{-\infty}^{\infty} (-it)e^{ixt} f(x)\,dx = it\hat{f}(t).$$

On the other hand, if $xf(x)$ is integrable, since $xe^{ixt} = i(d/dt)e^{it}$ we have

$$\widehat{xf}(x) = \int_{-\infty}^{\infty} e^{-ixt} xf(x)\,dx = i\frac{d}{dt}\int_{-\infty}^{\infty} e^{-ixt} f(x)\,dx = i\hat{f}'(x).$$

□

Lemma 1.37. ***(Riemann-Lebesgue Lemma)***. *If $f \in L^1$, then $\hat{f}(x) \to 0$ as $x \to \pm\infty$.*

Proof. First, consider the simple function

$$\phi_{[a,b]}(x) = \begin{cases} 1 & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases} .$$

We compute its Fourier transform

$$\hat{\phi}_{[a,b]}(t) = \int_a^b e^{-ixt}\,dx = \frac{-e^{-iat} + e^{ibt}}{-it}, \tag{1.9}$$

and see that

$$\left|\frac{-e^{-iat} + e^{-ibt}}{-it}\right| \le \frac{|e^{-itb}| + e^{-ita}}{t} = \frac{2}{|t|} \xrightarrow{|t|\to\infty} 0.$$

Now, suppose that f is the linear combination of step functions, i.e. $f = \sum_{N=1}^{M} c_N \phi_{A_N}$ with $A_N = [a_N, b_N]$. Thus, by the linearity of the Fourier transform, we have that

$$\hat{f}(t) = \sum_{N=1}^{M} c_N \hat{\phi}_{A_N}(t) \xrightarrow{|t|\to\infty} 0.$$

Let $\epsilon > 0$. Then since $f \in L^1$ and simple functions are dense in L^1 (see Chapter 11 in Rudin [5]), there exists a simple function g such that $\|f - g\| < \epsilon/2$, or what is the same,

$$\int_{-\infty}^{\infty} |f(x) - g(x)| dx < \frac{\epsilon}{2}.$$

Since g is a step function, we have already determined that $\hat{g}(t) \to 0$ as $|t| \to \infty$, meaning that there exists M such that if $|t| > M$ then $|\hat{g}(t)| < \epsilon/2$. With the same M, if $|t| > M$ then

$$|\hat{f}(t)| = |\hat{f}(t) + \hat{g}(t) - \hat{g}(t)| \leq |\hat{f}(t) - \hat{g}(t)| + |\hat{g}(t)| < |\hat{f}(t) - \hat{g}(t)| + \frac{\epsilon}{2}.$$

But we have

$$\begin{aligned} |\hat{f}(t) - \hat{g}(t)| &= \left| \int_{-\infty}^{\infty} f(x)e^{-itx} - g(x)e^{-itx}\, dx \right| \leq \int_{-\infty}^{\infty} |f(x) - g(x)||e^{-itx}|\, dx \\ &= \int_{-\infty}^{\infty} |f(x) - g(x)|\, dx = \|f - g\| < \frac{\epsilon}{2}. \end{aligned}$$

Putting all this together, we conclude that

$$|\hat{f}(t)| < |\hat{f}(t) - \hat{g}(t)| + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

Theorem 1.38. ***(Fourier Inversion Theorem)***. *Suppose that f is integrable and piecewise continuous on $\mathbb{R}$, defined at its points of discontinuity to satisfy $f(x) = \frac{1}{2}[f(x-) + f(x+)]$ for all x. Then*

$$f(x) = \lim_{\epsilon \to 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} e^{-\epsilon^2 t^2/2} \hat{f}(t)\, dt, \qquad x \in \mathbb{R}. \tag{1.10}$$

Moreover, if $\hat{f} \in L^1$, then f is continuous and

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \hat{f}(t)\, dt, \qquad x \in \mathbb{R}. \tag{1.11}$$

The first difficulty that we encounter while proving the theorem is that $\hat{f}$ is not necessarily in L^1, and we, therefore, can't establish 1.11 by simply substituting it in the defining formula for $\hat{f}$,

$$\int_{-\infty}^{\infty} e^{itx} \hat{f}(t)\, dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it(x-y)} f(y)\, dy\, dt$$

and interchanging the order of integration, because the integral $\int_{-\infty}^{\infty} e^{it(x-y)}\, dt$ is divergent. We solve this by multiplying $\hat{f}$ by a cutoff function $e^{-\epsilon^2 t^2/2}$ and we remove it by letting $\epsilon \to 0$. This leads to a more general version of the inversion formula shown in equation 1.10. A detailed proof of the Inversion Theorem can be found in section 7.2 of Folland [3].

Corollary 1.39. *If $\hat{f} = \hat{g}$, then $f = g$.*

Proof. If $\hat{f} = \hat{g}$, then $\widehat{(f-g)} = 0$, so $f - g = 0$ by the inversion equation 1.11. □

Remark 1.40. There are many functions for which both the function f and its Fourier transform $\hat{f}$ are in L^1. For instance, if f is twice differentiable and f' and f'' are also integrable, then $\mathcal{F}(f'')(t) = -t^2\hat{f}(t)$ is bounded, so $|\hat{f}(t)| \leq \frac{C}{1+t^2}$, whence $\hat{f} \in L^1$. Such functions have the property that f and $\hat{f}$ are bounded and continuous as well as integrable, and hence f and $\hat{f}$ are also in L^2.

1.3.1 L^2-Theory

Let's discuss Fourier theory in the L^2 space. It is generally known that L^2 is not a subset of L^1, and therefore the definition of the Fourier transform 1.33 is not directly applicable to every $f \in L^2$.

Theorem 1.41. ***(The Plancharel Theorem)***. *The Fourier transform, defined originally on $L^1 \cap L^2$, extends uniquely to a map from L^2 to itself that satisfies*

$$(\hat{f}, \hat{g}) = 2\pi(f, g) \quad \textit{and} \quad \|\hat{f}\|^2 = 2\pi\|f\|^2 \quad \textit{for all} \;\; f, g \in L^2. \tag{1.12}$$

Moreover, the formulas of theorem1.38 still hold for L^2 functions.

Proof. Suppose that f and g are L^1 functions such that $\hat{f}$ and $\hat{g}$ are also in L^1. Then by the formula 1.11 we have that

$$\begin{aligned} 2\pi(f,g) = 2\pi \int_{-\infty}^{\infty} f(x)\overline{g(x)}\,dx &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\overline{e^{itx}\hat{g}(t)}\,dt\,dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)e^{-itx}\overline{\hat{g}(t)}\,dx\,dt = \int_{-\infty}^{\infty} \hat{f}(t)\overline{\hat{g}(t)}\,dt = (\hat{f},\hat{g}). \end{aligned}$$

We see then that the Fourier transform preserves inner products up to a factor of 2π. In particular, if we take $f = g$, we obtain

$$\|\hat{f}\|^2 = 2\pi\|f\|^2.$$

Now, if f is an arbitrary L^2 function, we can find a sequence $\{f_n\}$ such that f_n and $\hat{f}_n$ are in L^1 and $f_n \to f$ in the L^2 norm (see chapter 2 of [3]). Then

$$\|\hat{f}_n - \hat{f}_m\|^2 = 2\pi\|f_n - f_m\|^2 \to 0 \qquad \text{as} \qquad m, n \to \infty$$

Hence $\{\hat{f}_n\}$ is a Cauchy sequence in L^2, and since this is a complete space, it has a limit. We can see that this limit depends only on f and not on the approximating sequence $\{f_n\}$. We define this limit to be $\hat{f}$. Thus, we have extended the domain of the Fourier transform to all L^2, and a simple limiting argument shows that this extended Fourier transform still preserves the norm and inner product up to a factor of 2π and that it still satisfies the properties stated in theorem 1.36. □

Remark 1.42. The Plancharel Theorem shows that the Fourier transform operator defines an isometry in L^2, which implies that if $f \in L^2$ then we also have that $\hat{f} \in L^2$. Not only this but also that $\hat{f}$ has the same "amount of energy" as f, scaled by a constant factor of 2π. This isometry of $L^1 \cap L^2$ into L^2 extends to an isometry of L^2 into L^2, and this extension defines the Fourier transform of every $f \in L^2$. As a result, L^2-theory has much more symmetry than L^1 because f and $\hat{f}$ play exactly the same role.

Theorem 1.43. *If $f \in L^2$ and $\hat{f} \in L^1$, then*

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(t)e^{ixt}\, dt \quad a.e.$$

Remark 1.44. In 1.43 lies the essential difference between the theory of the Fourier transforms in L^1 and L^2. In L^1, $\hat{f}(t)$ is defined unambiguously for all t. However, in L^2, $\hat{f}(t)$ is defined almost everywhere. For the proof, see chapter 9 of Rudin [4].

1.3.2 Multidimensional Fourier transform

In this section, we consider functions in $\mathbb{R}^n$. We will denote by $x \cdot y$ and $|x|$ the usual dot product and norm on $\mathbb{R}^n$:

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n \quad , \quad |x| = (x \cdot x)^{1/2}.$$

Definition 1.45. The *convolution* between two functions f and g on the space $\mathbb{R}^n$ is defined as:

$$f * g(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\, dy.$$

Remark 1.46. The basic algebraic properties of convolution stated in Theorem 1.31 remain true in the n-variable case, with the same proofs.

Definition 1.47. Given $f \in L^1(\mathbb{R}^n)$, the function

$$\hat{f}(t) = \int_{\mathbb{R}^n} f(x)e^{-it\cdot x}\, dx$$

is well defined for all $\mathbf{t} \in \mathbb{R}^n$. We call $\hat{f}$ the Fourier transform of f.

Theorem 1.48. *Suppose $f \in L^1(\mathbb{R}^n)$, and $\alpha \in \mathbb{R}^n$.*

1. *If $g(x) = f(x)e^{i\alpha x}$, then $\hat{g}(t) = \hat{f}(t - \alpha)$.*
2. *If $g(x) = f(x - \alpha)$, then $\hat{g}(t) = \hat{f}(t)e^{-i\alpha t}$.*
3. *If $g \in L^1$ and $h = f * g$, then $\hat{h}(t) = \hat{f}(t)\hat{g}(t)$.*
4. *If $\partial f/\partial x_j$ exists and is in L^1, then*
$$\mathcal{F}(\partial f/\partial x_j)(t) = it_j\hat{f}(t)$$
whereas if $x_j f(x)$ is integrable, then
$$\mathcal{F}\left[x_j f(x)\right] = i\partial\hat{f}/\partial t_j.$$
5. *The Fourier transform commutes with rotations: If R is a rotation of $\mathbb{R}^n$, then*
$$\mathcal{F}(f(Rx)) = \hat{f}(Rt).$$

Proof. The proofs of 1 to 4 are done as in the one-variable case. For 5, we need the fact that rotations preserve the dot products and volumes (for the rotation properties see section 6 of [1]). In other words, $t \cdot x = Rt \cdot Rx$ and $d(Rx) = dx$:

$$\begin{aligned}\widehat{f(Rx)} &= \int_{\mathbb{R}^n} e^{-it\cdot x} f(Rx)\,dx = \int_{\mathbb{R}^n} e^{-iRt\cdot x} f(Rx)\,d(Rx)\\ &= \int_{\mathbb{R}^n} e^{-iRt\cdot y} f(Ry)\,d(Ry) = \hat{f}(Rt).\end{aligned}$$

□

The formulas of the Fourier Inversion theorem 1.10 and 1.11 have the n-dimensional analogous:

Theorem 1.49. *If f is integrable and continuous in $\mathbb{R}$, then*

$$f(x) = \lim_{\epsilon\to 0} \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} e^{itx} e^{-\epsilon^2 t^2/2} \hat{f}(t)\,dt, \qquad x \in \mathbb{R}^n.$$

And if $\hat{f}$ is integrable, then

$$f(x) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} e^{itx} \hat{f}(t)\,dt, \qquad x \in \mathbb{R}^n.$$

The Plancharel theorem also remains true in n dimensions, except that the factor 2π is replaced by $(2\pi)^n$:

$$\|\hat{f}\|^2 = (2\pi)^n \|f\|^2.$$

Chapter 2

The Discrete Fourier Transform

2.1 Sampling periodic functions

We now raise the natural question of whether it is possible to reconstruct a periodic function $f : \mathbb{R} \to \mathbb{C}$ with a set of sample points of this function. Usually, periodic functions are sampled with a constant *sampling rate* $T > 0$. Thus, we are trying to recover a continuous function given that we have the set of points $\{(nT, f(nT))\}_{n\in\mathbb{Z}}$.

For that purpose, let us consider the Fourier transform of f

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt.$$

As discussed earlier, this exhibits f as the superposition of simple periodic waves $e^{i\omega t}$ as ω ranges over all possible frequencies. We ask what is the best sampling rate that allows us to recover f with the minimum number of sample points possible. Common sense tells us that if we choose a large number of points (oversample), we will be able to reconstruct the function. However, we could recover another function if we chose fewer points than necessary. This phenomenon is called *aliasing* and is shown in 2.1.

Definition 2.1. A function f is called *band limited* if its Fourier transform $\hat{f}$ is 0 outside of a finite interval $[-\Omega, \Omega]$. This Ω is called the *bandwidth* of f.

Theorem 2.2. ***(The Sampling Theorem).*** *Suppose that $f \in L^2$ is a band limited function with bandwidth Ω. Then f is completely determined by its values a t the points $\{t_n\}_{n\in\mathbb{Z}}$ with $t_n = n\pi/\Omega$. We also have that,*

$$f(t) = \sum_{-\infty}^{\infty} f(\frac{n\pi}{\Omega})\frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi} = \sum_{-\infty}^{\infty} f(\frac{n\pi}{\Omega})\,\text{sinc}\,(\Omega t - n\pi)\,. \tag{2.1}$$

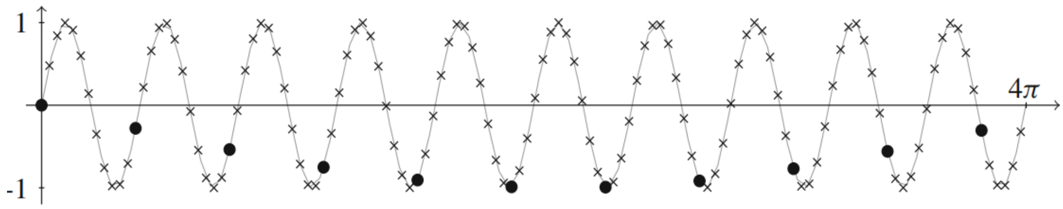


Figure 2.1: Undersampling of the function $f(x) = \sin 5x$. While the sampling rate $T = 0.1$ (crosses) works correctly, the sampling rate $T = 1.2$ (dots) gives a bad interpretation. This example is borrowed from [12] chapter 4.

Proof. First, we expand $\hat{f}$ in a Fourier series on the interval $[-\Omega, \Omega]$, writing $-n$ in place of n because it will be convenient later:

$$\hat{f}(\omega) = \sum_{-\infty}^{\infty} c_{-n} e^{-in\pi\omega/\Omega} \quad , \quad |\omega| \leq \Omega.$$

By the formula 1.4 the Fourier coefficients are given by

$$c_{-n} = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \hat{f}(\omega) e^{in\pi\omega/\Omega} d\omega = \frac{1}{2\Omega} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{in\pi\omega/\Omega} d\omega = \frac{\pi}{\Omega} f\left(\frac{n\pi}{\Omega}\right).$$

Here we have used that $\hat{f}$ vanishes outside the interval $[-\Omega, \Omega]$ and the Fourier inversion formula 1.11. Using these two results, we have that

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \hat{f}(\omega) e^{i\omega t}\, d\omega = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \sum_{-\infty}^{\infty} c_{-n} e^{-in\pi\omega/\Omega} e^{i\omega t}\, d\omega \\ &= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \sum_{-\infty}^{\infty} \frac{\pi}{\Omega} f\left(\frac{n\pi}{\Omega}\right) e^{-in\pi\omega/\Omega} e^{i\omega t}\, d\omega = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \sum_{-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) e^{i(\Omega t - n\pi)\omega/\Omega}\, d\omega \\ &= \frac{1}{2\Omega} \sum_{-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \frac{e^{i(\Omega t - n\pi)\omega/\Omega}}{i(\Omega t - n\pi)\omega/\Omega} \Bigg|_{-\Omega}^{\Omega} = \sum_{-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi}. \end{aligned} \tag{1}$$

Where in the last equality we have used the fact that $e^{i\omega} = \cos\omega + i \sin\omega$. Given that, by the Plancharel theorem 1.41, $\hat{f} \in L^2$, we have that by applied Bessel's inequality

$$\sum_{n=-\infty}^{\infty} |c_{-n}|^2 \leq \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} |\hat{f}(\omega)|^2\, d\omega < \infty,$$

or in other words, the Fourier series of $\hat{f}$ converges in $L^2(-\Omega, \Omega)$ in norm. Therefore the termwise integration that we have done in (1) is valid. since we are essentially taking the inner product of this series with $e^{i\omega t}$. □

This theorem is also known as the *Nyquist-Shannon Sampling Theorem* or the *Shannon-Whittaker Sampling Theorem* (see Shannon [11]). Both results are essentially the same, but there are subtle differences: the Nyquist-Shannon theorem is slightly more general, while the Shannon-Whittaker theorem also provides a specific formula for reconstructing the sampled function.

The theorem states that "a signal with bandwidth Ω must be sampled at a rate of at least $\frac{\pi}{\Omega}$ to capture all the information in the signal." In other words, "we must sample a periodic function at a rate of at least twice its highest frequency". This minimum sampling rate is known as the *Nyquist sampling rate*.

Remark 2.3. It is worth noting that the functions

$$\operatorname{sinc}(\Omega t - n\pi) = \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi} \qquad n \in \mathbb{Z}$$

form an orthogonal basis for the space of L^2 functions with bandwidth Ω. The sampling formula 2.1 is just the expansion of f with respect to this basis. Indeed, we can check that $\operatorname{sinc}(\Omega t - n\pi)$ is the inverse Fourier transform of the function

$$s_n(\omega) = \begin{cases} (\pi/\Omega)e^{-in\pi\omega/\Omega} & |\omega| < \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

The assertion, therefore, follows from the Plancharel theorem and the fact that the functions $e^{-in\pi\omega/\Omega}$ is an orthogonal basis for $L^2(-\Omega, \Omega)$.

We have shown that aliasing cannot occur if we sample a function at a sufficiently high rate. We want to understand how aliasing appears and how to eliminate it.

Lemma 2.4. ***(Poisson Formula)*** *Let $f \in L^2(\mathbb{R})$ and $\Omega > 0$ be such that either the function $\sum_{k\in\mathbb{Z}} f(\cdot + 2\Omega k) \in L^2([-\Omega, \Omega])$ or the series $\sum_{k\in\mathbb{Z}} |f\left(\frac{k\pi}{\Omega}\right)|^2$ converges. Then, for almost all $t \in \mathbb{R}$,*

$$\sum_{k\in\mathbb{Z}} \hat{f}(t + 2\Omega k) = \frac{\pi}{\Omega} \sum_{k\in\mathbb{Z}} f\left(\frac{k\pi}{\Omega}\right) e^{-i\frac{k\pi}{\Omega}t}. \tag{2.2}$$

Proof. See Bredies [12] Lemma 4.37. □

Remark 2.5. For the special case in the Possion formula where $t = 0$ we obtain the interesting expression

$$\sum_{k\in\mathbb{Z}} \hat{f}(2\Omega k) = \frac{\pi}{\Omega} \sum_{k\in\mathbb{Z}} f\left(\frac{\pi}{\Omega}k\right)$$

that relates the values of f and its Fourier transform.

We can represent our function, sampled discretely with a sample rate π/Ω as

$$f_d = \sum_{k\in\mathbb{Z}} f\left(\frac{k\pi}{\Omega}\right)\delta_{k\frac{\pi}{\Omega}}. \tag{2.3}$$

Remark 2.6. Let's comment now that the delta defined above is the *Dirac measure in x* and is defined as

$$\delta_x(A) = \begin{cases} 1 & , \quad x \in A \\ 0 & , \quad \text{otherwise} \end{cases}$$

where $A \subset \mathbb{R}$. The distribution induced by the Dirac measure is the *delta distribution*, denoted also as δ_x:

$$\delta_x(\phi) = \int_{-\infty}^{\infty} \phi \, d\delta_x = \phi(x).$$

Its Fourier transform is given by

$$\hat{\delta_x}(\phi) = \delta_x(\hat{\phi}) = \hat{\phi}(x) = \int_{-\infty}^{\infty} e^{-ixy}\phi(y)\, dy.$$

Therefore, its Fourier transform is represented by the function $y \mapsto e^{-ixy}$. It is worth noting, thus, that f_d is not a function, but rather a distribution, however, its Fourier transform is a function. This is an informal explanation of this subject, for references in distributions and the Fourier transform of distributions see Folland [3], Plonka [6], or Bredies [12].

What we want to express in 2.3 is that we are *discretizing* the continuous signal f in a sampling rate of $k\pi/\Omega$. It is worth noting that this notation is somewhat an abuse of formalism because the Dirac delta is technically a measure, not a function.

The following lemma establishes a connection via the Fourier transform:

Lemma 2.7. *For almost all $t \in \mathbb{R}$,*

$$\hat{f_d}(t) = \frac{\Omega}{\pi}\sum_{k\in\mathbb{Z}} \hat{f}(t + 2\Omega k).$$

Proof. As exposed in remark 2.6, the Fourier transform of $\delta_{k\frac{\pi}{\Omega}}$ is given by

$$\mathcal{F}(\delta_{k\frac{\pi}{\Omega}})(t) = e^{-i\frac{k\pi}{\Omega}t}.$$

If we now take the Fourier transform of f_d, we obtain:

$$\hat{f_d}(t) = \sum_{k\in\mathbb{Z}} f\left(\frac{k\pi}{\Omega}\right) e^{-i\frac{k\pi}{\Omega}t} = \frac{\Omega}{\pi}\sum_{k\in\mathbb{Z}} \hat{f}(t + 2\Omega k),$$

where in the last equality we used the formula of lemma 2.4. □

Expressed in words, this lemma says that the Fourier transform of the sampled signal corresponds to a periodization of the Fourier transform of the original signal with a period of 2Ω. In this way of speaking, we can interpret the reconstruction formula 2.1 as a convolution as well:

$$f(x) = \sum_{-\infty}^{\infty} f(\frac{n\pi}{\Omega}) \frac{\sin(\Omega x - n\pi)}{\Omega t - n\pi} = f_d(x) * \operatorname{sinc}(\Omega x - n\pi).$$

In the Fourier realm, this means that

$$\hat{f}(t) = (\pi/\Omega)\hat{f}_d(t)e^{-in\pi t/\Omega}\phi_{[-\Omega,\Omega]} = \begin{cases} (\pi/\Omega)\hat{f}_d(t)e^{-in\pi t/\Omega} & |\omega| < \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

Where $\phi_{[-\Omega,\Omega]}$ is the characteristic function of the simple $[-\Omega, \Omega]$, and we have used the convolution property of the Fourier transform (Theorem 1.36) and the Fourier transform of $\operatorname{sinc}(\Omega x - n\pi)$ which we have already seen in 2.3.

If the support of $\hat{f}$ is contained within the interval $[-\Omega, \Omega]$, then no overlap occurs during periodization. In this case, the function $(\pi/\Omega)\hat{f}_d(t)e^{-in\pi t/\Omega}\phi_{[-\Omega,\Omega]}$ corresponds exactly to $\hat{f}$.

However, if $\hat{f}$ has larger support, then the support of $\hat{f}(\cdot + 2\Omega k)$ intersects with $[-\Omega, \Omega]$ for multiple values of k. This "folding" in the frequency domain is responsible for aliasing.

Remark 2.8. We can obtain an analog of the Sampling theorem for two dimensions using the tensor product. Let $f : \mathbb{R}^2 \to \mathbb{C}$ be such that its Fourier transform has its support in the rectangle $[-\Omega_1, \Omega_1] \times [-\Omega_2, \Omega_2]$. In this case, f is determined by the values $f(k_1\pi/\Omega_1, k_2\pi/\Omega_2)$, and

$$f(x_1, x_2) = \sum_{k \in \mathbb{Z}^2} f\left(\frac{k_1\pi}{\Omega_1}, \frac{k_2\pi}{\Omega_2}\right) \operatorname{sinc}\left(\frac{\Omega_1}{\pi}(x_1 - \frac{k_1\pi}{\Omega_1})\right) \operatorname{sinc}\left(\frac{\Omega_2}{\pi}(x_2 - \frac{k_2\pi}{\Omega_2})\right)$$

Where $\operatorname{sinc}(\pi\omega) = \frac{\sin \pi\omega}{\pi\omega}$. We denote by

$$f_d = \sum_{k \in \mathbb{Z}^2} u(k_1T_1, k_2T_2)\delta(k_1T_1, k_2T_2)$$

an image that is discretely sampled on a rectangular grid with sampling rates T_1 and T_2. Using the Fourier transform the connection to the continuous image f can be expressed as

$$\hat{f}_d(t) = \sum_{k \in \mathbb{Z}^2} \hat{f}(t_1 + 2\Omega_1 k_1, t_2 + 2\Omega_2 k_2).$$

Notice that the condition of the function f being band-limited is quite restrictive as it implies that the function's frequency spectrum has a finite support. However, in practical scenarios, band-limited functions are prevalent, given that natural signals originate from physical processes that cannot produce infinite sequences. Walker [10] provides an extensive discussion on how to sample and how aliasing appears in non-band-limited functions.

2.2 Discrete Fourier Transform (DFT)

Consider the problem of representing a finite set of data points using sinusoidal functions. In other words, given a set $X_N = \{x_k \mid k \in \{0, 1, ..., N-1\}\}$. and a function

$$\begin{aligned} f \ : X_N &\to \mathbb{C} \\ x_k &\mapsto f_k := f(x_k) \end{aligned}$$

we aim to express this data in the frequency domain. It is important to note that the term "Discrete Fourier Transform" can be somewhat misleading, as we are computing the discrete analog of the Fourier series rather than the Fourier transform.

From now on, we will denote

$$X_N = \{x_k : k = 0, 1, \dots, N-1\} = \{\frac{2\pi k}{N} : k = 0, 1, \dots, N-1\}.$$

Let us define the set $l_N = l^2(X_N)$ of functions $f : X_N \to \mathbb{C}$. The set l_N is a complex vector space of dimension N. Consider the inner product:

$$\langle f, g \rangle = \frac{1}{N} \sum_{k=0}^{N-1} f(x_k)\overline{g(x_k)}. \tag{2.4}$$

Note that this is not an inner product in the continuous case, because $\langle f, f \rangle = 0$ does not imply that $f = 0$. However, this is valid since we are only in a discrete set of points. The rest of the inner product properties follow easily. With this, we see that l_N is a Hilbert space.

Definition 2.9. An *exponential polynomial* of degree at most $N-1$ is any function of the form

$$P(x) = \sum_{k=0}^{N-1} c_k e^{ikx}.$$

Theorem 2.10. *Let us define the functions* φ_n

$$\varphi_n(x) = e^{inx}. \tag{2.5}$$

The set $\Phi = \{\varphi_0, ..., \varphi_{N-1}\}$, *is an orthonormal basis with respect to the inner product 2.4.*

Proof. Given that the number of vectors is N, the only thing left to see is that they are orthogonal. Let $\omega = \omega_N = e^{-2\pi i/N}$. We have

$$\langle \varphi_n, \varphi_m \rangle = \frac{1}{N} \sum_{k=0}^{N-1} \omega^{-nk} \omega^{mk} = \frac{1}{N} \sum_{k=0}^{N-1} \omega^{(m-n)k}.$$

If $n = m$, all the terms in the sum are equal to 1. Which gives $\langle \omega_n, \omega_n \rangle = 1$. If $n \neq m$, we have a finite geometric sum with the ratio ω^{m-n}. By the formula of the geometric sum we have that

$$\langle \varphi_n, \varphi_m \rangle = \frac{1 - \omega^{(m-n)N}}{1 - \omega^{m-n}}.$$

But given that $n \neq m$ we have that $\omega^{(m-n)N} = e^{(-2\pi i(m-n)/N)\cdot N} = e^{-2\pi i(m-m)} = 1$, which implies that $\langle \varphi_n, \varphi_m \rangle = 0$. □

Corollary 2.11. *The exponential polynomial that interpolates a prescribed function f at nodes $x_n \in X_N$ is given by the equations*

$$P = \sum_{k=0}^{N-1} c_k \varphi_k \quad \textit{with} \quad c_k = \langle f, \varphi_k \rangle.$$

Proof. Using the given formula of c_k, let us compute the value of the exponential polynomial at an arbitrary $x_m \in X_N$. We obtain:

$$\begin{aligned}
\sum_{k=0}^{N-1} c_k \varphi_k(x_m) &= \sum_{k=0}^{N-1} \langle f, \varphi_k \rangle \varphi_k(x_m) \\
&= \sum_{k=0}^{N-1} \frac{1}{N} \sum_{j=0}^{N-1} f(x_j) \overline{\varphi_k(x_j)} \varphi_k(x_m) \\
&= \sum_{j=0}^{N-1} f(x_j) \frac{1}{N} \sum_{k=0}^{N-1} \overline{\varphi_k(x_j)} \varphi_m(x_m) \\
&= \sum_{j=0}^{N-1} f(x_j) \langle \varphi_m, \varphi_j \rangle \\
&= f(x_m).
\end{aligned}$$

□

Corollary 2.12. *The polynomial described in corollary 2.11 is the only exponential polynomial of degree at most $N - 1$ that interpolates the nodes described.*

Proof. Suppose $\sum_{k=0}^{N-1} a_k \varphi^k$ is an exponential polynomial that interpolates f at $x_0, x_1, \ldots, x_{N-1}$. Then

$$\sum_{k=0}^{N-1} a_k \varphi_k(x_j) = f(x_j)\ , \qquad 0 \leq j \leq N-1.$$

Multiplying both sides of the equation by $\varphi_n(-x_j)$ and sum with respect to j we obtain

$$\sum_{k=0}^{N-1} a_k \sum_{j=0}^{N-1} \varphi_k(x_j)\varphi_n(-x_j) = \sum_{j=0}^{N-1} f(x_j)\varphi_n(-x_j).$$

But using the definition of the inner product 2.4 we get

$$\sum_{k=0}^{N-1} a_k \langle \varphi_k, \varphi_n \rangle = \langle f, \varphi_n \rangle.$$

Since $\langle \varphi_k, \varphi_n \rangle = 1$ only if $k = n$ then we conclude that

$$a_n = \langle f, \varphi_n \rangle = c_n.$$

□

Definition 2.13. The one-dimensional *discrete Fourier transform* of the vector $f \in \mathbb{C}^n$ is defined by

$$\hat{f}_n = \frac{1}{N} \sum_{k=0}^{N-1} f_k e^{-2\pi i n k/N} = \frac{1}{N} \sum_{k=0}^{N-1} f_k \omega^{nk}.$$

By corollary 2.11, we see that the discrete Fourier transform of a vector $f \in \mathbb{C}^N$ is the unique exponential polynomial that passes through the N points $\{(x_j, f(x_j)) : x_j \in X_N\}$. f can be regarded as N samples of a continuous function f that we wish to interpolate.

Remark 2.14. Let us denote

$$b^n = (e^{-2\pi i k n/N})_{k=0,\ldots,N-1} = (w^{nk})_{k=0,\ldots,N-1} \quad \text{and} \quad f = (f_k)_{k=0,..,N-1}.$$

Note that for each n, the nth term of the DFT can be expressed as $\hat{f}_n = \frac{1}{N} b_n \cdot f$. This leads us to the next definition.

Definition 2.15. The N-by-N *Fourier matrix* is defined by

$$F_N = \frac{1}{N} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{bmatrix} \tag{2.6}$$

It is clear that $\hat{f} = F_N f$, or what is the same:

$$\begin{bmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \hat{f}_2 \\ \vdots \\ \hat{f}_{N-1} \end{bmatrix} = \frac{1}{N} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{bmatrix}$$

Theorem 2.16. *The Fourier transform is inverted by*

$$f_k = \sum_{k=0}^{N-1} \hat{f}_k e^{2\pi i n k/N}$$

Proof. The Fourier matrix 2.6 can be written as

$$F_N = \frac{1}{N}[\varphi_0(x_k), \varphi_1(x_k), ..., \varphi_{N-1}(x_k)]_{k\in\{0,1,\cdots,N-1\}},$$

where the φ functions are defined as in 2.5 and $x_j = \frac{-2\pi j}{N}$ (we interpret $\varphi_j(x_k)$ as a column vector). Given that, as shown in theorem 2.10, these functions are orthonormal, we have that $F_N F_N^* = \frac{1}{N} Id$, where Id is the $N \times N$ identity matrix, and F_N^* is the Hermitian adjoint of F_N. We see then $F_N^{-1} = N F_N^*$. Then we have

$$f = F_N^{-1} \hat{f} = N F_N^* \hat{f}.$$

By definition of the Hermitian adjoint, we have

$$(F_N^*)_{k,l} = e^{2\pi i k l/N}.$$

In particular, we obtain

$$F_N^* = [\overline{\varphi_0(x_k)}, .., \overline{\varphi_{N-1}(x_k)}]_{k\in\{0,\cdots,N-1\}}.$$

$\square$

Remark 2.17. In other words, the discrete Fourier transform expresses a vector $\hat{f}$ in terms of the orthogonal basis $\Phi = \{\varphi_0, ..., \varphi_{N-1}\}$. We have that for all $j \in \{0, ..., N-1\}$, φ_j is a N-periodic function

$$\varphi_j(x_{k+N}) = e^{2\pi i j(k+N)/N} = e^{2\pi i j k/N} = \varphi_j(x_k).$$

Given that $\hat{f}$ is a linear combination of N-periodic functions, it also has period N. Hence, the information in it is completely contained in the finite sequence $\hat{f}_0, ..., \hat{f}_{N-1}$.

Definition 2.18. The *two-dimensional Fourier transform* $\overline{f} \in \mathbb{C}^{N\times M}$ of $f \in \mathbb{C}^{N\times M}$ is defined by

$$\hat{f}_{k,l} = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f_{n,m}e^{-2\pi ink/N}e^{-2\pi iml/M}.$$

Remark 2.19. The two-dimensional Fourier transform is inverted by the expression

$$f_{n,m} = \sum_{k=0}^{N-1}\sum_{l=0}^{M-1} \hat{f}_{k,l}e^{2\pi ink/N}e^{2\pi iml/M}.$$

The proof is similar to the one in theorem 2.16.

Definition 2.20. Let $f, g \in \mathbb{C}^N$. The *discrete periodic convolution* of f and g is defined by

$$(f * g)_n = \sum_{k=0}^{N-1} g_k f_{(n-k) mod\, N}.$$

Theorem 2.21. *For $f, g \in \mathbb{C}^N$,*

$$\widehat{(f * g)}_n = N\hat{f}\hat{g}.$$

Proof. Using the periodicity of the complex exponential function,

$$\begin{aligned}
\widehat{(f * g)}_n &= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1} g_l f_{(k-l) mod\, N} e^{-2\pi ink/N} \\
&= \frac{1}{N}\sum_{l=0}^{N-1} g_l e^{-2\pi inl/N} \sum_{k=0}^{N-1} u_{(k-l) mod\, N} e^{-2\pi in(k-l)/N} \\
&= N\hat{g}_n\hat{f}_n.
\end{aligned}$$

□

2.3 The Fast Fourier Transform (FFT)

The Sampling Theorem states that a continuous signal with limited bandwidth can be completely reconstructed from its samples if the sampling rate meets the Nyquist criterion, namely, it is at least twice the maximum frequency present in the signal. Building on this principle, the DFT is used to transform these sampled signals from the time domain to the frequency domain, providing a powerful tool for signal analysis. Nonetheless, DFT's computational cost is a significant limitation, since it carries a complexity of $O(N^2)$, where N is the number of samples.

Developed initially by James Cooley and John Tukey in 1965, the FFT addresses this issue by recursively breaking down the DFT of an N-point sequence into smaller operations, reducing its computational complexity to $O(N \log N)$.

We will first expose Cooley and Tukey's [**?**] idea of breaking down the computation of the DFT into smaller computations, only to see that if the number of samples is a power of 2 then it is feasible to use this method. The algorithm, however, has been derived from the works of Walker [10] and Cheney [7].

Let f be a signal sampled $N-1$ times, and let f_j denote the jth sampling of f. Consider the problem of computing the DFT,

$$\hat{f}_j = \frac{1}{N}\sum_{k=0}^{N-1} f_k \omega^{jk}. \tag{1}$$

As stated before, a straightforward computation of $\hat{f}$ would take N^2 operations. By operation we mean a complex multiplication followed by a complex addition.

To derive the algorithm, suppose that N is composite i.e., $N = n_1 \cdot n_2$. We can express the indices in (1) as

$$\begin{aligned} j &= j_1 n_1 + j_0, \qquad j_0 = 0,1,\ldots,n_1-1, \qquad j_1 = 0,1,\ldots,n_2-1 \\ k &= k_1 n_2 + k_0, \qquad k_0 = 0,1,\ldots,n_2-1, \qquad k_1 = 0,1,\ldots,n_1-1. \end{aligned}$$

Then if we put $\hat{f}_{j_1,j_0} = \hat{f}_j$ and $f_{k_1,k_0} = f_k$, we can write

$$\hat{f}_{j_1,j_0} = \frac{1}{N}\sum_{k_0}\sum_{k_1} f_{k_1,k_0} \cdot \omega^{jk_1n_2}\omega^{jk_0}. \tag{2}$$

Since

$$\omega^{jk_1n_2} = \omega^{(j_1n_1+j_0)k_1n_2} = \omega^{Nj_1k_1}\omega^{j_0k_nn_2} = \omega^{j_0k_1n_2}, \tag{3}$$

the inner sum, over k_1 depends only on j_0 and k_0 and we can define a new array f^1 as,

$$f^1_{j_0,k_0} = \frac{1}{N}\sum_{k_1} f_{k_1,k_0}\omega^{j_0k_1n_2}.$$

We can write the result as

$$\hat{f}_{j_1,j_0} = \frac{1}{N}\sum_{k_0} f^1_{j_0,k_0}\omega^{(j_1n_1+j_0)k_0}.$$

Clearly, the array f^1 has N elements, each requiring n_1 operations, thus it takes a total of Nn_1 operations to calculate $\hat{f}$ from f^1. Similarly, it takes Nn_2 operations to calculate f from f^1. We conclude that this two-step algorithm defined by (2) and (3) takes a total amount of

$$T = N(n_1 + n_2).$$

It is easy to see that if $N = n_1 \cdot n_2 \cdot \ldots \cdot n_m$ then we have a m-step algorithm requiring

$$T = N(n_1 + n_2 + \ldots + n_m). \tag{4}$$

If some $n_j = s_j t_j$ with $s_j, t_j > 2$, then $s_j + t_j < n_j$ (if $s_j = t_j = 2$ then clearly $s_j + t_j = n_j$). We can see then that a high number of factors provides a minimum of (4). If we choose N to be highly composite, we would really reduce the number of operations.

If all n_j were equal to n then we would have

$$m = \log_n N$$

and the total number of operations would be

$$T(n) = nN \log_n N.$$

If $N = r^m s^n t^p \ldots$, then from (4) we find that

$$\frac{T}{N} = m \cdot r + s \cdot n + t \cdot p \cdots$$

if we combine this with the fact that

$$\log_2 N = m \cdot \log_2 r + n \cdot \log_2 s + p \cdot \log_2 t + \cdots,$$

we obtain that

$$\frac{T}{N \log_2 N}$$

is a weighted mean of the quantities

$$\frac{r}{\log_2 r}, \frac{s}{\log_2 s}, \frac{t}{\log_2 t}, \ldots$$

whose values run as follows

r	$\frac{r}{\log_2 r}$
2	2.00
3	1.88
4	2.00
5	2.15
6	2.31
7	2.49
8	2.67
9	2.82
10	3.01

We see that if the number of points is a power of 3 the decomposition is the most efficient. Still, we only reduce 6% of the operations computed using 2 or 4 which have important computing advantages with binary arithmetic. We could even use, if necessary, $r_j = 10$ and increase the number of operations not over 50%. However, whenever possible, we use $N = 2^m$ (this is known as the Radix-2 FFT and it is the most common application). We will develop the algorithm considering that N is a power of 2. For further information on other types of FFT see Plonka [6], Brigham [9] or Walker [10].

Theorem 2.22. *Let p and q be exponential polynomials of degree $\leq n-1$ such that for the points $x_j = \pi j/n$, we have*

$$p(x_{2j}) = f(x_{2j}) \qquad q(x_{2j}) = f(x_{2j+1}) \qquad 0 \leq j \leq n-1.$$

Then the exponential polynomial of degree $\leq 2n-1$ that interpolates f at the points $x_0, x_1, \ldots x_{2n-1}$ is given by

$$P(x) = \frac{1}{2}(1 + e^{inx})p(x) + \frac{1}{2}(1 - e^{inx})q(x - \pi/n) \tag{2.7}$$

Proof. It is clear that P has degree $\leq 2n-1$ since p and q both have degree $\leq n-1$ and e^{inx} has degree n. Let us see that P interpolates f at all its nodes. We have, for $0 \leq j \leq 2n-1$,

$$P(x_j) = \frac{1}{2}(1 + e^{\pi inj/n})p(x_j) + \frac{1}{2}(1 - e^{\pi inj/n})q(x_j - \pi/n)$$

Notice that

$$e^{\pi inj/n} = e^{\pi ij} = \begin{cases} +1 & \text{j even} \\ -1 & \text{j odd} \end{cases}$$

Thus for even j, we see that $P(x_j) = p(x_j) = f(x_j)$, while for odd j, we have

$$P(x_j) = q(x_j - \pi/n) = q(x_{j-1}) = f(x_j).$$

□

Theorem 2.23. *Let the coefficients of the polynomials described in Theorem 2.22 be as follows*

$$p(x) = \sum_{j=0}^{n-1} \alpha_j e^{ijx} \qquad q(x) = \sum_{j=0}^{n-1} \beta_j e^{ijx} \qquad P(x) = \sum_{j=0}^{2n-1} \gamma_j e^{ijx}$$

Then, for $0 \leq j \leq n-1$ we have

$$\gamma_j = \frac{1}{2}\alpha_j + \frac{1}{2}e^{ij\pi/n}\beta_j \gamma_{j+n} = \frac{1}{2}\alpha_j - \frac{1}{2}e^{ij\pi/n}\beta_j$$

Proof. Notice first that

$$q\left(x-\frac{\pi}{n}\right)=\sum_{j=0}^{n-1}\beta_j e^{ij(x-\pi/n)}=\sum_{j=0}^{n-1}\beta_j e^{i\pi j/n}e^{ijx}$$

Now, from equation 2.7 we have

$$\begin{aligned}P(x)&=\frac{1}{2}(1+e^{inx})p(x)+\frac{1}{2}(1-e^{inx})q(x-\pi/n)\\&=\frac{1}{2}\sum_{j=0}^{n-1}\left\{(1+e^{inx})\alpha_j e^{ijx}+(1-e^{inx})\beta_j e^{-i\pi j/n}e^{ijx}\right\}\\&=\frac{1}{2}\sum_{j=0}^{n-1}\left\{(\alpha_j+\beta_j e^{-i\pi j/n})e^{ijx}+(\alpha_j+\beta_j e^{-i\pi j/n})e^{i(n+j)x}\right\}\end{aligned}$$

The formulas for γ_j and γ_{j+n} can be read from this equation. □

From now on, let $R(N)$ denote the minimum number of multiplications needed to compute the coefficients of an interpolating polynomial on the set of points $X_N=\{2\pi j/N : 0\le j\le N-1\}$.

Theorem 2.24. $R(2^m)\le m2^m$.

Proof. By theorem 2.23 we have that the coefficients γ_j of P can be obtained from the coefficients in p and in q at the cost of $2n$ multiplications. Indeed, we need n multiplications to compute every $\frac{1}{2}\alpha_j$, and another n multiplications to compute $(\frac{1}{2}e^{-ij\pi/n})\beta_j$ (assuming that we already have $\frac{1}{2}e^{-ij\pi/n}$). Since by the definition of $R(n)$, the coefficients $\alpha_0,\ldots,\alpha_{n-1}$ cost $R(n)$ multiplications, and the same is true for the coefficients $\beta_0,\ldots,\beta_{n-1}$, we obtain that the total cost of computing the coefficients of P is at most $2R(n)+2n$ multiplications. We have then that

$$R(2n)\le 2R(n)+2n. \tag{2.8}$$

We shall prove the inequality of the theorem by induction. Consider the case where $m=0$. We wish to interpolate the point $x_0=0$ by an exponential polynomial of degree zero. The solution is the constant $f(0)$; no multiplications are required. Thus, the assertion of the theorem is true for $m=0$. Using inequality 2.8, the calculation for the inductive step from m to $m+1$ is

$$\begin{aligned}R(2^{m+1})&=R(2\cdot 2^m)\le 2R(2^m)+2\cdot 2^m\\&\le 2(m2^m)+2^{m+1}=(m+1)2^{m+1}.\end{aligned}$$

□

As a result of Theorem 2.24, we can deduce that if N is a power of 2, say 2^m, then the computational cost of finding the interpolating exponential polynomial adheres to the inequality $R(N) \leq N \log_2 N$. The algorithm that implements the procedure described in Theorem 2.7 is the Fast Fourier Transform.

The content of Theorem 2.22 can be understood in terms of two linear operators, E_n and T_h. For any function f, let $E_n f$ represent the exponential polynomial of degree $n-1$ that interpolates f at the nodes $\frac{2\pi j}{n}$ for $0 \leq j \leq n-1$. Let T_h be a translation operator defined by

$$(T_h f)(x) = f(x+h).$$

We know from Corollary 2.11 that

$$E_n f = \sum_{k=0}^{n-1} \langle f, \varphi_k \rangle \varphi_k,$$

where the φ functions are defined in the DFT section. Moreover, from Theorem 2.22 we infer that

$$P = E_{2n} f \qquad p = E_n f \qquad q = E_n T_{\pi/n} f$$

The conclusion of Theorems 2.22 and 2.23 is that $E_{2n} f$ can be obtained efficiently from $E_n f$ and $E_n T_{\pi/n} f$. This holds for $n = 1, 2, \ldots$
Note that 2.22 can be expressed as

$$E_{2n} f(x) = \frac{1}{2}(1 + e^{inx}) E_n f(x) + \frac{1}{2}(1 - e^{inx}) T_{-\pi/n} E_n T_{\pi_n} f(x).$$

We want now to establish one version of the fast Fourier transform that allows us to compute $E_N f$ where $N = 2^m$.

We define the exponential polynomials $P_k^{(n)}$ as the one of degree $2^n - 1$ that interpolates f in the following way

$$P_k^{(n)} \left(\frac{2\pi j}{2^n} \right) = f \left(\frac{2\pi k}{N} + \frac{2\pi j}{2^n} \right), \qquad 0 \leq j \leq 2^n - 1$$

We could check that the set of nodes $\frac{2\pi k}{N} + \frac{2\pi j}{2^n}$ corresponding to one value of k is disjoint from the same set with a different k. A straightforward application of Theorem 2.22 shows that

$$P_k^{(n+1)}(x) = \frac{1}{2}(1 + e^{i2^n x}) P_k^{(n)}(x) + \frac{1}{2}(1 - e^{i2^n x}) P^n_{k+2^{m-n-1}} \left(x - \frac{\pi}{2^n} \right).$$

We can illustrate how the exponential polynomials $P_k^{(n)}$ are related. Suppose we want to compute $P_0^{(4)}$. The tree of the relations with the polynomials of lower order is depicted in the following tree diagram:

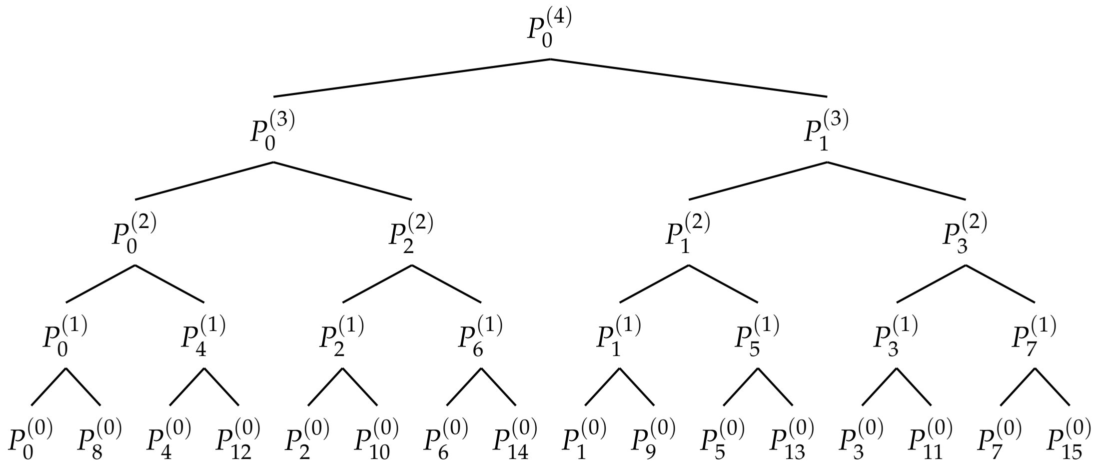


Figure 2.2: Polynomial separation of $P_0^{(4)}$.

2.3.1 Algorithm

We denote the Pk(n) coefficients by $A_{kj}^{(n)}$. Here $0 \leq n \leq m$, $0 \leq k \leq 2^{m-n} - 1$, and $0 \leq j \leq 2^n - 1$. We have that

$$P_k^{(n)}(x) = \sum_{j=0}^{2^n-1} A_{kj}^{(n)} e^{ijx},$$

and by Theorem 2.23, the following equalities hold:

$$A_{k,j}^{(n+1)} = \frac{1}{2}\left[A_{k,j}^{(n)} + e^{-ij\pi/2^n} A_{k+2^{m-n}-1,j}^{(n)}\right]$$

$$A_{k,j+2^n}^{(n+1)} = \frac{1}{2}\left[A_{k,j}^{(n)} - e^{-ij\pi/2^n} A_{k+2^{m-n}-1,j}^{(n)}\right]$$

Given that $0 \leq k \leq 2^{m-n} - 1$ and $0 \leq j \leq 2^n - 1$, the array $A^{(n)}$ requires $2^{m-n} \cdot 2^n = 2^m = N$ storage locations. One way to carry out the computations is to use two linear arrays of size N to hold successive A(n) computations. We call these two arrays C and D. If C stores $A^{(n)}$, D will store $A^{(n+1)}$. In the next iteration, C will store $A^{(n+1)}$ and D will store A^{n+2}.
$A^{(n)}$ is stored in C by the rule

$$C(2^n k + j) \leftarrow A_{kj}^{(n)}, \qquad 0 \leq k \leq 2^{m-n} - 1\,, 0 \leq j \leq 2^n - 1.$$

Similarly, $A^{(n+1)}$ is stored in D by the rule

$$D(2^{n+1} k + j) \leftarrow A_{kj}^{(n+1)}, \qquad 0 \leq k \leq 2^{m-n-1} - 1\,, 0 \leq j \leq 2^{n+1} - 1.$$

$A^{(n)}_{0,0}$	$A^{(n)}_{0,1}$	$\cdots$	$A^{(n)}_{0,2^n-1}$	$A^{(n)}_{1,0}$	$A^{(n)}_{1,1}$	$\cdots$	$A^{(n)}_{2^{m-n}-1,0}$	$\cdots$	$A^{(n)}_{2^{m-n}-1,2^n-1}$

Figure 2.3: Example of the array C storing $A^{(n)}$.

The factors $Z(j) = e^{-2\pi ij/N} = \omega^j$ are computed at the beginning and stored. We use also the fact that $e^{-ij\pi/2^n} = Z(j2^{m-n-1})$. Finally, here is the **fast Fourier transform algorithm**:

```
input m
N ← 2^m
ω ← e^{-2πi/N}
for k = 0,1,2,...,N-1 do
    Z(k) ← ω^k
    C(k) ← f(2πk/N)
end for
for n = 0,1,2,...,m-1 do
    for k = 0,1,2,...,2^{m-n-1}-1 do
        for j = 0,1,2,...,2^n-1 do
            u ← C(2^n k + j)
            v ← Z(j2^{m-n-1})C(2^n k + 2^{m-1} + j)
            D(2^{n+1}k + j) ← (u + v)/2
            D(2^{n+1}k + j + 2^n) ← (u - v)/2
        end for
    end for
    for j = 0,1,2,...,N-1 do
        C(j) ← D(j)
    end for
end for
output C(0),C(1),...,C(N)
```

Let's analyze the number of operations needed:

- The innermost loop runs 2^n times and involves:
    - 2 assignments to u and v.
    - 1 complex multiplication.
    - 2 complex subtractions and additions.

- The next loop runs 2^{m-n-1} times.
- The outside loop runs m times.
- Computing ω and $Z(k)$: involves N multiplications and N assignments.
- Computing $C(k)$ involves N function evaluations (assuming f to be an $O(1)$ operation).
- Copying $D(j)$ to $C(j)$ takes N assignments.

We have a total of

$$\sum_{n=0}^{m-1} \sum_{k=0}^{2^{m-n-1}} \sum_{j=0}^{2^n-1} 5 = 5 \cdot m \cdot 2^{m-n-1} \cdot 2^n = 5 \cdot 2^{m-1}.$$

Since $m = \log_2 N$, the expression simplifies to

$$5 \log_2 N \cdot \frac{N}{2} = \frac{5}{2} N \log_2 N.$$

Combining the results adds a total of N operations. The number of operations is therefore dominated by the $N \log N$ of the main loop. Hence, the algorithm has a complexity of

$$O(N \log_2 N).$$

Remark 2.25. We have an algorithm that reverts the FFT, the Inverse Fast Fourier Transform (IFFT). Like the FFT, the IFFT has a time complexity of $O(N \log_2 N)$, and its steps are exactly the same, except now we want to compute the inverse DFT

$$f_k = \sum_{k=0}^{N-1} \hat{f}_k e^{2\pi i n k/N}$$

It is used when we want to recover the original function once we have computed its DFT.

Remark 2.26. We also have an analog of the FFT in multiple dimensions. For an N-dimensional signal, the multidimensional FFT can be computed by applying the one-dimensional FFT sequentially along each dimension. For example, a 2D FFT is computed by performing a 1D FFT on each row of the matrix, followed by a 1D FFT in each column of the resulting matrix. If the original signal f is of size $M \times N$, the resulting matrix is of the form

$$f_{u,v} = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f_{(x,y)} e^{-2i\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)},$$

and its computation yields a complexity of $O(MN \log N + MN \log M)$, compared to the complexity $O(M^2 N^2)$ of the naive computation. For more information on the two-dimensional FFT see chapter 12 of Brigham [9].

Chapter 3

Applications in Image Processing

3.1 Image processing fundamentals

Mathematical representation of images

Mathematically, an image is a function that maps every point in some domain of definition to a certain color value. In other words, an image u is a map from an image domain Ω to a color space F:

$$u : \Omega \to F.$$

We distinguish between discrete and continuous image domains, in continuous image domains usually $\Omega \subset \mathbb{R}^d$, and in the case of discrete domains the image is a discrete grid of pixels and the color space can be discrete (e.g., grayscale values from 0 to 255).
We also have different color spaces:

- Binary images, where $F = \{0, 1\}$.
- Grayscale images, where $F = [0, 1]$ in the continuous case, and $F = \{0, 1, \ldots, 255\}$ in the discrete case.
- Color images, where $F = [0, 1]^N$ if continuous, and $F = \{0, 1, \cdots, 255\}^N$ with N number of color channels.

Image processing often deals with discrete color sets, which is reasonable given that images are generated in discrete form or need to be transformed to a discrete domain before further processing. However, most methods used in discrete images are motivated by continuous considerations. We will deal mostly with grayscale images.

A continuous image u takes x and y coordinates of the image domain Ω and maps them into the color space. We call the value $u(x,y) \in F$ as the amplitude of the image at this point. Converting a continuous image into a discrete one is based on two processes known as *sampling* and *quantification*. Sampling refers to the process of digitizing (or discretizing) the coordinates of the image domain, and quantifying to digitize the values of the color space F.

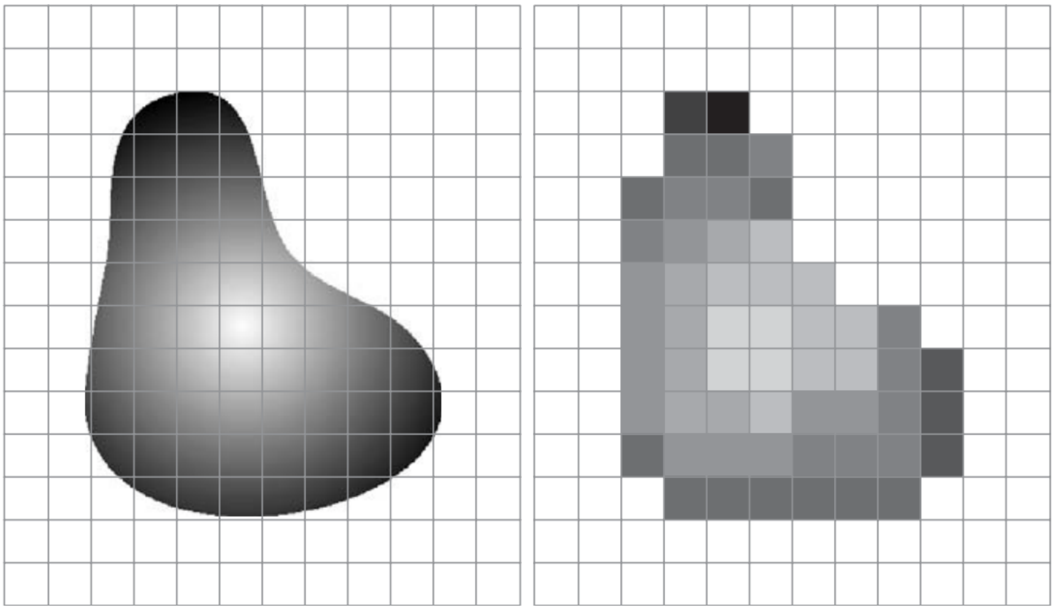


Figure 3.1: Process of converting a continuous into a discrete image. This example is borrowed from Gonzalez [14].

The sampling process may be viewed as partitioning the continuous xy-plane into a grid of coordinates, with each cell of the grid (a pixel) being an element of $\mathbb{Z}^2$. Quantification often divides F into 2^k intervals because of hardware considerations. This process is illustrated in 3.1.
A usual way to represent discrete images is with a matrix. Where each element represents the amplitude of the image at that point:

$$u(x,y) = \begin{bmatrix} u(0,0) & u(0,1) & \cdots & u(0,N-1) \\ u(1,0) & u(1,1) & \cdots & u(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ u(M-1,0) & u(M-1,1) & \cdots & u(M-1,N-1) \end{bmatrix}$$

Image Transforms

The *spatial domain* of an image refers to the representation of the image that we have already introduced: coordinates and amplitude. However, when we want to perform some image processing tasks, it is better to formulate the problem we are tackling in a *transform domain*. The idea is that we take the spatial domain of an image, and apply a transform, once we have the image in the transform domain, we do the specific task, and then, use the inverse transform to return to the spatial domain.

Definition 3.1. Consider the 2-D linear transform of the form

$$T(k,l) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} u(x,y) r(x,y,k,l)$$

where u is the input image, r is called a *forward transformation kernel*, and M and N are the row and column dimensions of u. The function T is called a *forward transform* of u.

Definition 3.2. *Given a forward transform T, we can recover u using the inverse transform of T:*

$$u(x,y) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} T(k,l) s(x,y,k,l),$$

where s is called the inverse transformation kernel.

The nature of a transform is determined by its kernel. The transform we will be looking at is the Fourier transform of an image, which is basically computing the DFT of the image. If we want to be more precise, we can define the Fourier transform of an image as the transform with the forward and inverse kernels given by

$$r(x,y,k,l) = \frac{1}{MN} e^{-2\pi i (\frac{kx}{M} + \frac{ly}{N})} \quad , \quad s(x,y,k,l) = e^{2\pi i (\frac{kx}{M} + \frac{ly}{N})}$$

It is important to note that we can't directly represent images in the frequency domain because complex numbers now represent the pixels. What we see is the logarithm of the module of the complex number. We use the logarithm because it covers a wider range of numbers. The phase of these complex numbers also plays an important role, and if we get rid of it, applying the IDFT would result in an illegible image. However, representing the phase graphically does not give us much insight into the image's composition.

In addition to the Fourier transform, there are a number of other important transforms, including the *Walsh*, *Hadamard*, *discrete cosine*, *Haar* and *slant* transforms.

Filtering images

"Filtering" refers to passing, modifying, or rejecting specified components of an image. We can filter an image in the spatial domain or its transform domain. In the spatial domain, we typically use convolution with a kernel, while in the transform domain, we use frequency filters.

When filtering an image in the Fourier domain, we focus on *frequency filters*. These filters modify the frequency content of an image, allowing us to enhance or suppress specific frequency components. If the operation performed on the pixels of the image is linear, we refer to them as *linear filters*. Conversely, operations that are not linear are termed *nonlinear filters*.

A *kernel* is an array whose size defines the neighborhood of operation, and whose coefficients determine the nature of the filter. Examples of kernels include those used for smoothing (e.g., Gaussian kernel) and edge detection (e.g., Sobel kernel).

Definition 3.3. The *convolution* of a kernel w of size $m \times n$ with an image u is defined as

$$(w * f)(x, y) = \sum_{s=-a}^{a} \sum_{r=-b}^{b} w(s, t) f(x - s, y - t),$$

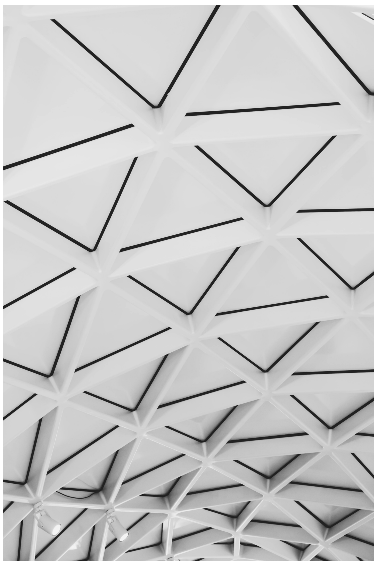
where $m = 2a + 1$ and $n = 2b + 1$.

When we use the term *linear spatial filtering*, we mean *convolving a kernel with an image*.

The link between spatial and frequency image processing is the Fourier transform. We use the Fourier transform to go from the spatial domain to the frequency domain and to return to the spatial domain we use the inverse Fourier transform. The key property is that convolution in the spatial domain is equivalent to multiplication in the frequency domain, as seen in Theorem 2.21. Linear filtering involves finding suitable ways to modify the frequency content of an image. In the spatial domain, we achieve this through convolution filtering, while in the frequency domain, we use multiplicative filters.

When performing convolution, especially at the borders of an image, we often use *padding*. Padding involves adding a border of pixels around the original image to ensure that the convolution operation can be applied uniformly across the entire image. There are several padding strategies:

- *Zero Padding*: Adding a border of zero-valued pixels.
- *Replicate Padding*: Extending the border pixels of the image.
- *Reflect Padding*: Mirroring the border pixels of the image.



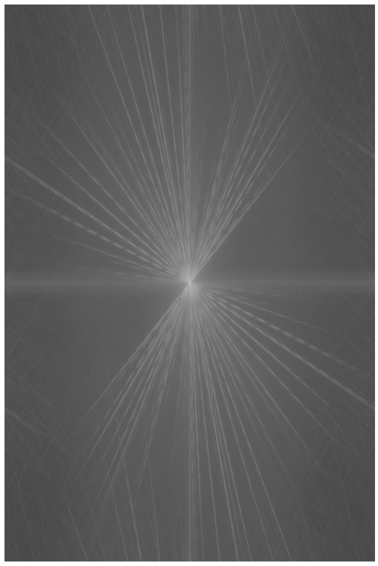


Figure 3.2: Image in the spatial domain (left) and the same image in the frequency domain (right)

Padding helps prevent the reduction in image size and avoids artifacts that can occur at the edges of the image during convolution.

Here are some types of filters in the frequency domain:

- *Low-Pass Filters*: Allow low-frequency components to pass through while attenuating high-frequency components, typically used for smoothing and noise reduction.
- *High-Pass Filters*: Allow high-frequency components to pass through while attenuating low-frequency components, used for edge detection and sharpening.
- *Band-Pass Filters*: Allow a specific range of frequencies to pass through, used for specific feature extraction.

3.2 Smoothing and Denoising Images

Smoothing in the spatial domain is used to reduce sharp transitions in intensity. Edges and other sharp-intensity transitions (such as noise) in an image contribute significantly to the high-frequency content of its Fourier transform.

Definition 3.4. Consider a filter that passes without attenuation all frequencies within a circle of radius from the origin, and cuts off all frequencies outside. This circle is called an *ideal lowpass filter* (ILPF). It is specified by the transfer function

$$H(k,l) = \begin{cases} 1 & \text{if } D(k,k) \leq D_0 \\ 0 & \text{if } D(k,k) > D_0 \end{cases}$$

where D_0 is a positive constant. And $D(k,l)$ is the distance between a point (k,l) in the frequency domain and the center of the $P \times Q$ frequency rectangle, where $P = \lfloor \frac{M}{2} \rfloor$ and $Q = \lfloor \frac{N}{2} \rfloor$. That is

$$D(k,l) = \left[(k - P/2)^2 + (l - Q/2)^2\right]^{1/2}. \tag{3.1}$$

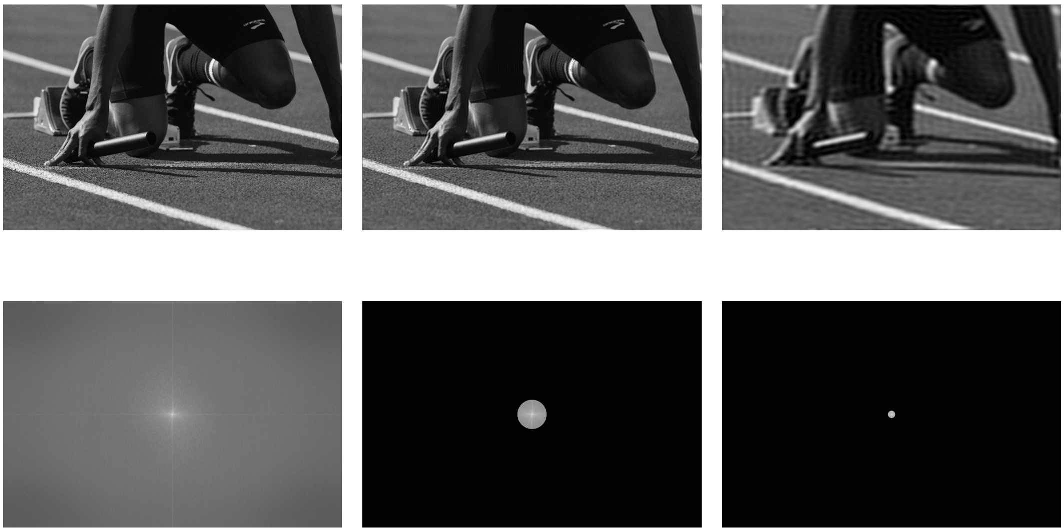


Figure 3.3: Applying ILPF with different cutoff frequencies.

In Figure 3.3, we observe the effect of an ideal lowpass filter. The top-left image is the original, and directly below it is its Fourier transform. The center column shows the result of applying an ideal lowpass filter with a cutoff frequency of 100. On the right, we see the outcome with a cutoff frequency of 25. Notice how reducing the number of high frequencies results in a progressively smoother (and blurrier) image.

Definition 3.5. The *Gaussian lowpass filters* (GLPF) are the transfer functions that have the form

$$H(k,l) = e^{-D^2(k,l)/2D_0^2}$$

where D_0 is the cutoff frequency and D is the distance defined as in definition 3.4.

We can see the differences between the ILPF and the GLPF in Figure 3.4. We observe in the bottom right corner of 3.3 that a noticeable ringing effect appears. This is caused because the IDFT of the ILPF transfer function is a sinc-like function, and by the convolution theorem, we are convolving the spatial domain by an oscillatory function, and therefore it is natural for some kind of ringing pattern to appear. This makes the ILPFs quite useless for practical purposes, but they give us much insight into how lowpass filtering smoothing works. The IDFT of a Gaussian function, however, is also Gaussian. This means that a spatial Gaussian filter kernel computed using the IDFT, will not have this ringing effect.

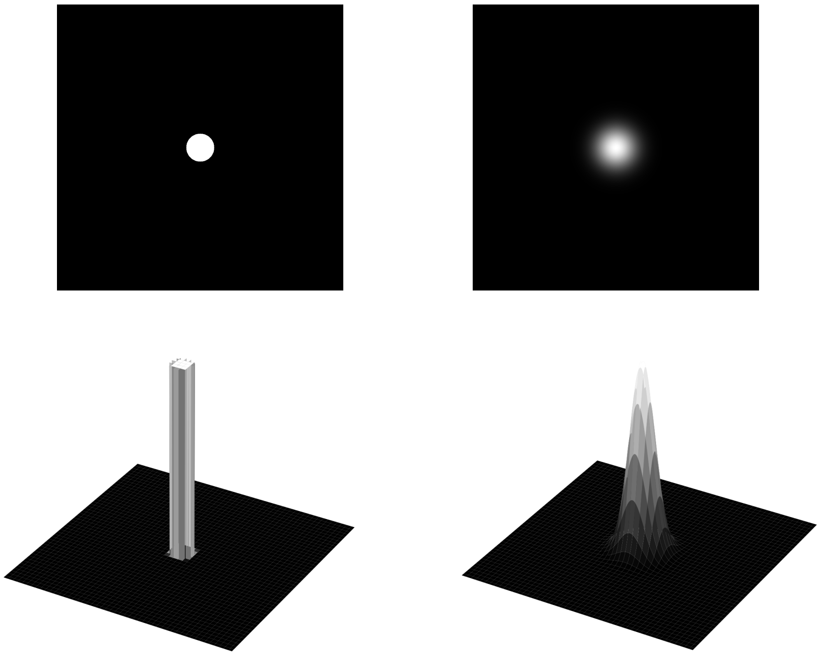


Figure 3.4: Comparison of ILPF and GLPF, both with $D_0 = 50$. Top row: 2D representations. Bottom row: 3D models.

Definition 3.6. *The transfer function of a Butterworth lowpass filter (BLPF) of order n, with a cutoff frequency at a distance D_0 from the center of the frequency rectangle, is defined as*

$$H(k,l) = \frac{1}{1 + [D(k,l)/D_0]^{2n}}$$

where D is defined as in 3.1.

BLPFs are used when we want to have more control over how we want to treat high frequencies. When n is high, the BLPF approaches an ILPF. When n has a low value, it is similar to a GLPF.

Sometimes, texts have low resolution as shown in the left side of Figure 3.5. We humans can fill the gaps without any difficulties, however, machine recognition systems and machine learning models have real problems reading these broken characters. One approach to solve this is to fill the gaps blurring the image as we see on the right side of Figure 3.5. These characters have been repaired with a GLPF with $D_0 = 50$.

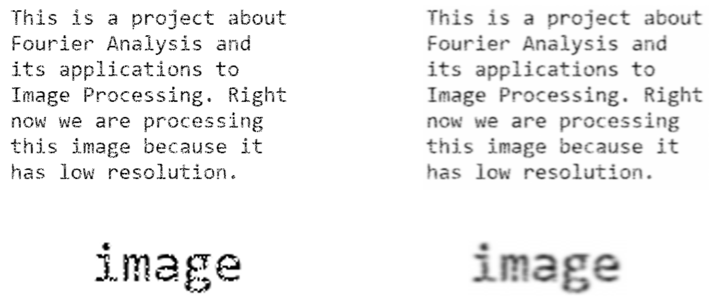


Figure 3.5: In the top row the same text before and after applying a GLPF with $D_0 = 50$. In the bottom row a comparison of the same word before and after the filter.

Lowpass filtering is also widely used in the printing and publishing industry to produce smoother, softer-looking images rather than the sharp original. For human faces, the typical use is to reduce the sharpness of face wrinkles and small blemishes. We see an application of this in Figure 3.6, where we see a reduction of Federer's skin lines and spots noticeably.

Figure 3.6: Before (left) and after (right) of applying a GLPF with $D_0 = 130$. In the bottom left corner of the images there is a detail of Federer's eye.

3.3 Image sharpening

Because edges and other changes in intensities are associated with high-frequency components of images, image sharpening can be achieved by filtering the frequency domain with highpass filters.

We can easily define the analog highpass filters from the lowpass ones. The ideal highpass filter transform function (IHPF) is given by

$$H(k,l) = \begin{cases} 0 & \text{if } D(k,l) \leq D_0 \\ 1 & \text{if } D(k,l) > D_0 \end{cases}.$$

Similarly, it follows that the transfer function of a Gaussian highpass filter (GHPF) transfer function is given by

$$H(k,l) = 1 - e^{-D^2(k,l)/2D_0^2}$$

and the transfer function of a Butterworth highpass filter is

$$H(k,l) = \frac{1}{1 + [D_0/D(k,l)]^{2n}}$$

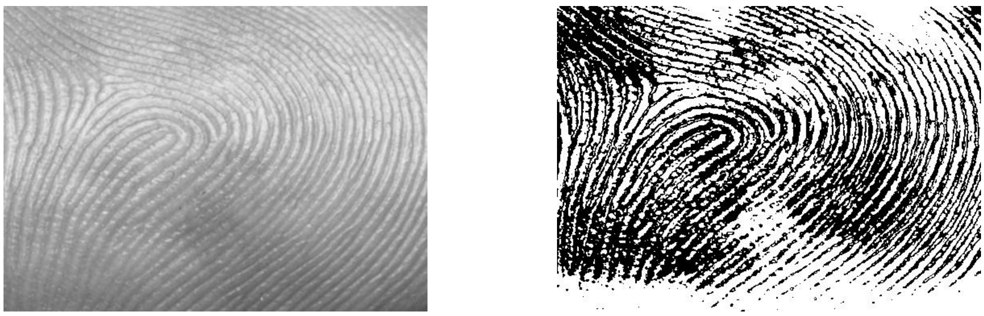


Figure 3.7: Fingerprint image and the same image Filtered using a BHPF with $D_0 = 30$ and $n = 4$. A threshold filter has also been used to convert the grayscale image to a binary one.

In 3.7 we can see the image of the fingerprint with smudges. For automated fingerprint recognition systems to work properly, it is necessary to enhance the ridges and reduce the smudges. We can use highpass filters to minimize this smudging. Enhancement of these ridges is accomplished by the fact that high frequencies characterize their boundaries. On the other hand, the filter reduces low-frequency components, which correspond to slowly varying intensities in the image.

Remember that the *Laplacian* of a function $f(x,y)$ of two variables is defined as

$$\nabla^2 f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y}.$$

As a derivative operator, the Laplacian accentuates sharp intensity changes in an image while reducing the emphasis on regions with gradually varying intensities. To retain the background details and enhance the sharpening effect of the Laplacian, we can add the Laplacian image back to the original image. However, for optimal sharpening results, we subtract the Laplacian from the original image:

$$g(x,y) = f(x,y) - [\nabla^2 f(x,y)] \tag{3.2}$$

where $f(x,y)$ and $g(x,y)$ are the input and the sharpened images, respectively.

It can be shown that the Laplacian can be implemented in the frequency domain using the filter transfer function (see Gonzalez [14])

$$H(k,l) = -4\pi^2 D^2(k,l).$$

Using this transfer function, we see that the Laplacian of an image $u(x,y)$, is obtained with the expression

$$\nabla^2 u(x,y) = \mathcal{F}^{-1}[H(k,l)U(k,l)]$$

where $\mathcal{F}^{-1}$ denotes the IDFT and U is the DFT of u. Using these facts, we can write equation 3.2 in the frequency domain

$$\begin{aligned} v(x,y) &= \mathcal{F}^{-1}\{U(k,l) - H(k,l)U(k,l)\} \\ &= \mathcal{F}^{-1}\left\{[1 - H(k,l)]U(k,l)\right\} \\ &= \mathcal{F}^{-1}\left\{\left[1 + 4\pi^2 D^2(k,l)\right] U(k,l)\right\}. \end{aligned}$$

Figure 3.8 demonstrates the effect of applying the Laplacian filter to an image of Jupiter. In the filtered image on the right, the round shape of Jupiter is more clearly defined, and the details of the inner lines are more pronounced. This enhancement gives the impression of improved image quality, though the image has simply been sharpened.

3.4 Selective filtering

The filters we've discussed so far operate across the entire frequency spectrum. However, there are situations where we might want to process only specific frequency bands or small regions within the frequency spectrum. Filters that target

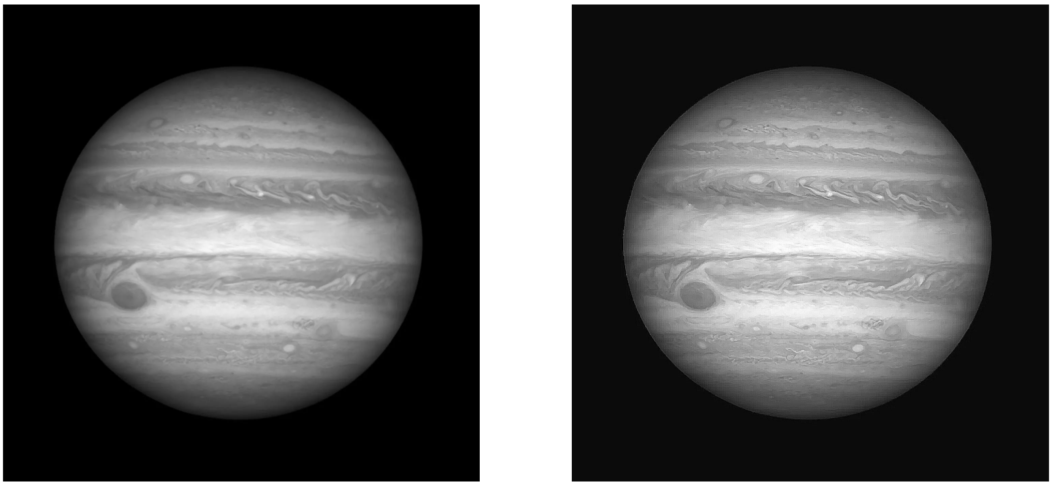


Figure 3.8: Telescope image of Jupiter (left) and the same image with a Laplacian sharpening (right).

particular frequency bands are known as *band filters*. If these filters allow certain frequencies to pass through, they are called *bandpass filters*. Conversely, they are referred to as bandreject filters if they block specific frequencies. When filters are designed to affect particular regions within the frequency spectrum, they are known as *notch filters*. Depending on their function, these can be further classified as *notch reject filters* or *notch pass filters*.

Bandpass and bandreject functions can be constructed by combining lowpass and highpass filter transfer functions. Let us see the analog expressions of the lowpass and highpass filter transfer functions that we have already seen. Here C_0 is the center of the band, W is the width of the band and $D(k,l)$ is the distance from the center of the transfer function to a point (k,l) in the frequency rectangle. The ideal bandreject filter (IBRF) is

$$H(k,l) = \begin{cases} 0 & \text{if } C_0 - \frac{W}{2} \leq D(k,l) \leq C_0 + \frac{W}{2} \\ 1 & \text{otherwise} \end{cases}$$

The Gaussian bandreject filer (GBRF) is

$$H(k,l) = 1 - e^{-\left[\frac{D^2(k,l) - C_0^2}{D(k,l)W}\right]^2}$$

The Butterworth bandreject filter (BBRF) is

$$H(k,l) = \frac{1}{1 + \left[\frac{D(k,l)W}{D^2(k,l) - C_0^2}\right]^{2n}}$$

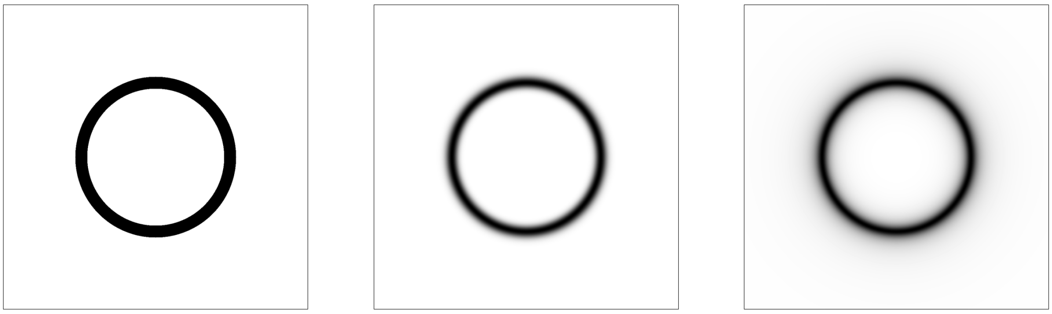


Figure 3.9: Comparison of Bandreject Filters. IBRF (left), GBRF (center), BBRF (right) with $C_0 = 250$, $W = 40$, and $n = 1$ in the BBRF.

Notch filters are among the most effective selective filters. These filters allow or block frequencies within a predefined neighborhood of the frequency domain. A notch filter transfer function centered at (k_0, l_0) must have a corresponding notch at the location $(-k_0, l_0)$. *Notch reject* transfer functions are created by multiplying highpass filter transfer functions whose centers have been shifted to the locations of the notches. The general form is

$$H_{NR}(k,l) = \prod_{j=0}^{Q} H_j(k,l) H_{-j}(k,l)$$

where $H_j(k,l)$ and $H_{-j}(k,l)$ are highpass filter transfer functions centered at (k_j, l_j) and (k_{-j}, l_{-j}) respectively. These centers are specified relative to the center of the frequency rectangle. The distance calculations for each filter transfer function are expressed as:

$$D_{\pm j}(k,l) = \left[(u - \frac{M}{2} \mp k_j)^2 + (v - \frac{N}{2} \mp l_j)^2 \right].$$

To derive a notch pass filter transfer function from a notch reject filter, we use the following relationship:

$$H_{NP}(k,l) = 1 - H_{NR}(u,v).$$

A primary use of notch filtering is to selectively adjust specific regions within the Discrete Fourier Transform (DFT). This technique is frequently applied interactively.

Figure 3.10 shows an example of the application of notch filtering. The top-right image shows a newspaper image of a soldier that exhibits a moiré pattern. The top left image is its spectrum. We observe that this transform exhibits some "energy bursts" that are the result of the periodicity of the moiré pattern. We attenuate this bursts by using notch filtering. The bottom-left image shows the result of multiplying the DFT by a Gaussian notch reject function with $D_0 = 20$. The location of the notches and the radius have been selected by interacting with the spectrum. The bottom-right image shows the result of applying these filters, observing a significant improvement.

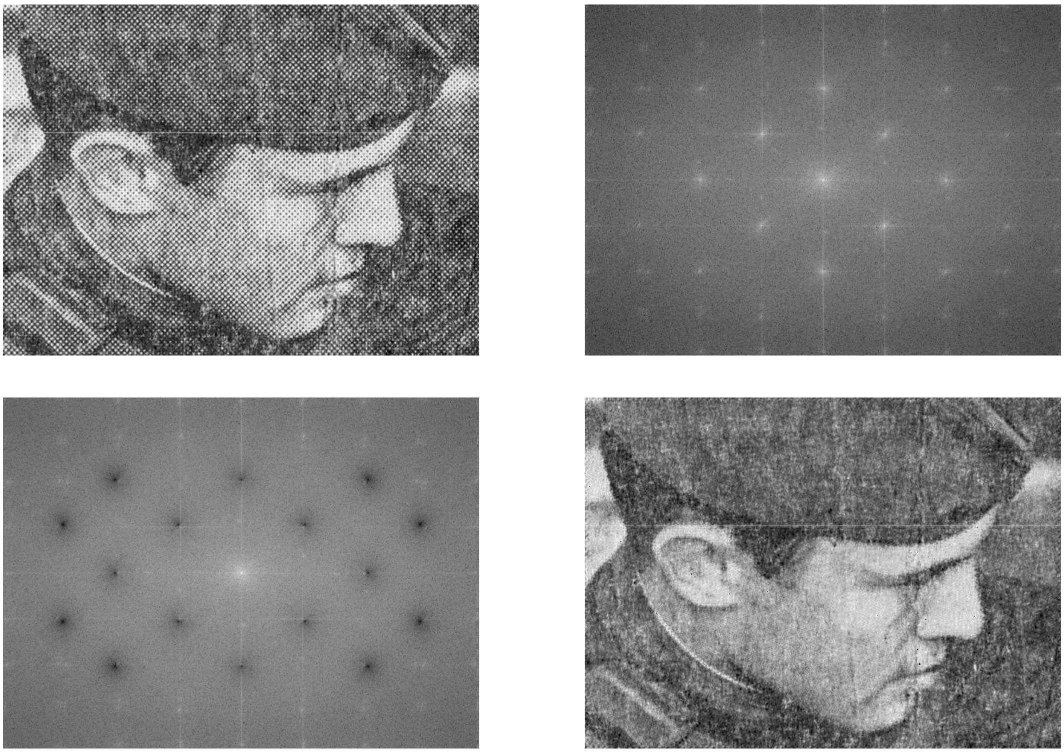


Figure 3.10: Process of filtering an image with a moiré pattern with a notch filter.

Figure 3.11 shows us another application of notch filtering. In the top-left of the figure, we see a picture presenting periodic noise. Its DFT (top-right) presents some energy bursts in the center vertical axis. We "cover" them with a notch filter using a Gaussian reject transfer function with $D_0 = 20$. We observe in the bottom right that this noise has been clearly mitigated.

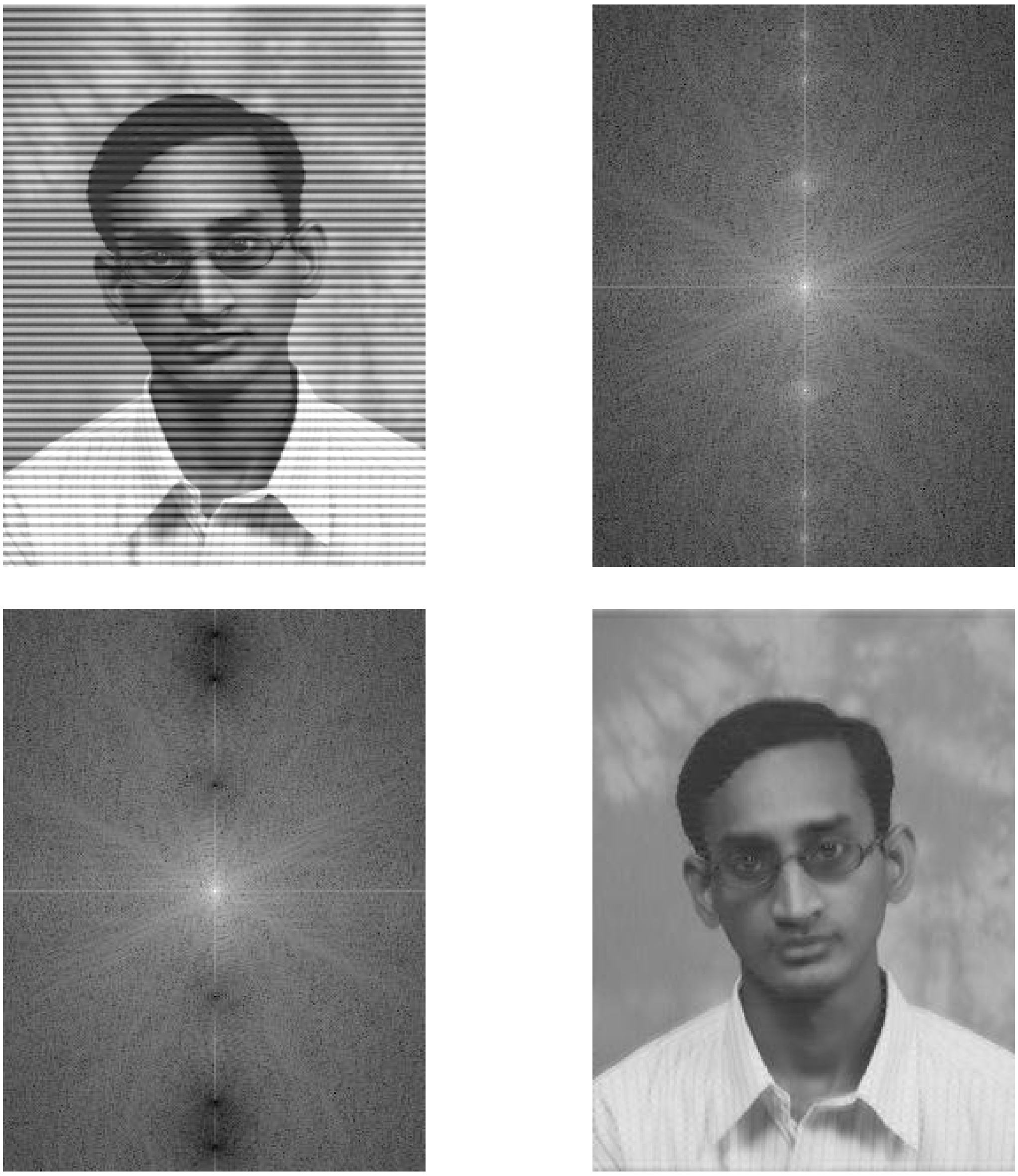


Figure 3.11: Process of filtering an image with periodic noise.

Bibliography

[1] Stein, E. M., & Shakarchi, R. (2011). *Fourier analysis: an introduction* (Vol. 1). Princeton University Press.

[2] Vretblad, A. (2003). *Fourier Analysis and Its Applications*. Springer.

[3] Folland, G. B. (2009). *Fourier analysis and its applications* (Vol. 4). American Mathematical Soc..

[4] Rudin, W. (1966). *Real and Complex Analysis*. McGraw-Hill.

[5] Rudin, W. (1964). *Principles of mathematical analysis* (Vol. 3). New York: McGraw-hill.

[6] Plonka, G., Potts, D., Steidl, G., & Tasche, M. (2023). *Numerical Fourier Analysis*. Springer Nature.

[7] Kincaid, D., & Cheney, E. (1991). *Numerical Analysis: Mathematics of Scientific Computing*. Brooks/Cole.

[8] Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, *19*(90), 297-301.

[9] Brigham, E. O. (1988). *The fast Fourier transform and its applications*. Prentice-Hall, Inc..

[10] Walker, J. S. (1996). *Fast Fourier Transforms* (Vol. 24). CRC Press.

[11] Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, *37*(1), 10-21.

[12] Bredies, K., & Lorenz, D. (2018). *Mathematical image processing* (pp. 1-469). Cham: Springer International Publishing.

[13] Brunton, S. L., & Kutz, J. N. (2022). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.

[14] Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (Fourth edition). Pearson.

[15] Gonzalez, R., Woods, R., & Eddins, S. (2003). *Digital Image Processing Using MATLAB*. Prentice-Hall, Inc..