# Regularity and the Fourier Series:
## How the Smoothness of a Periodic Function Governs Coefficient Decay, Convergence, and the Gibbs Phenomenon

Akshit Khera

July 14, 2026

**Abstract**

The Fourier coefficients of a periodic function are not arbitrary numbers: how quickly they shrink is a precise fingerprint of how smooth the function is. This paper develops that single idea into one explanatory chain,

$$\text{regularity} \implies \text{coefficient decay} \implies \text{convergence behavior.}$$

We begin with the Riemann–Lebesgue lemma, which guarantees that coefficients fade to zero but says nothing about *how fast*. An integration-by-parts argument then upgrades this into a quantitative law: each derivative the function possesses contributes a factor of $1/n$, so that functions with $k$ continuous derivatives have coefficients of order $o(|n|^{-k})$, while real-analytic functions enjoy exponential decay. We show that fast enough decay (absolute summability) forces uniform convergence, and we record the clean sufficient condition that any continuously differentiable function has a uniformly convergent Fourier series. We then study the failure case. Dirichlet's theorem identifies the limit at a jump as the midpoint of the one-sided values, while the Gibbs phenomenon exposes a persistent overshoot near a jump (about 8.95% of the jump height in the model cases we compute, and present at every jump of a piecewise-smooth function) that no number of terms can remove, only narrow. A comparison of a square wave, a triangle wave, and an analytic function makes the trichotomy $1/n$ vs. $1/n^2$ vs. exponential concrete. As a bonus, a single application of Parseval's identity to the sawtooth evaluates the celebrated Basel sum, $\sum_{n\geq 1} 1/n^2 = \pi^2/6$. Finally, Fejér's theorem shows that *averaging* the partial sums restores good behavior, and we explain why: the Fejér kernel is nonnegative, whereas the Dirichlet kernel changes sign and has a logarithmically growing $L^1$ norm. That one distinction is the deep source of both the Gibbs overshoot and its cure.

# 1 Introduction

The first thing anyone learns about Fourier series is that a periodic function can be rebuilt out of sines and cosines. That is true, but it is only the beginning of the story. The interesting question, the one that makes the subject worth studying, is *how well* this rebuilding works, and what decides the answer.

Here is the picture to keep in mind. A Fourier series is like a recipe: it builds a function by mixing together "pure tones," the simple waves $\cos nx$ and $\sin nx$ of frequency $n$. The *Fourier coefficients* are the amounts of each ingredient. Some functions are easy to build: a few ingredients already give an almost perfect copy. Others are stubborn: no matter how many ingredients we add, the reconstruction keeps wobbling, especially near sharp features. The thesis of this paper is that the difference is governed by one property of the original function: its *regularity*, meaning how smooth it is.

The slogan, made precise in the sections below, is that smoothness controls how fast the high-frequency ingredients become negligible:

> *How does the regularity of a periodic function determine the decay of its Fourier coefficients, the convergence of its Fourier series, and the appearance of Gibbs-type oscillations near discontinuities?*

A rough or discontinuous function needs large amounts of high-frequency tones, so its coefficients die away slowly, typically like $1/n$. A smoother function gets by with much less at high frequency, so its coefficients fall off faster, like $1/n^2$, or $1/n^k$, or even exponentially fast for the smoothest (analytic) functions. Slow decay and poor reconstruction go hand in hand, and a jump in the function leaves an unmistakable visible scar. We organize everything around the chain of implications

$$\boxed{\text{regularity of } f \implies \text{decay rate of } \widehat{f}(n) \implies \text{mode of convergence of } S_N f.} \tag{1}$$

Figure 1 previews the whole paper in one image. Each panel shows a function together with an approximation that uses the *same* number of Fourier ingredients. The discontinuous square wave is reconstructed poorly and rings near its jumps; the continuous (but cornered) triangle wave is reconstructed well; the analytic function is reconstructed so well that the approximation cannot be told apart from the target by eye. The only thing that differs between the panels is the smoothness of the function, read through the chain (1).

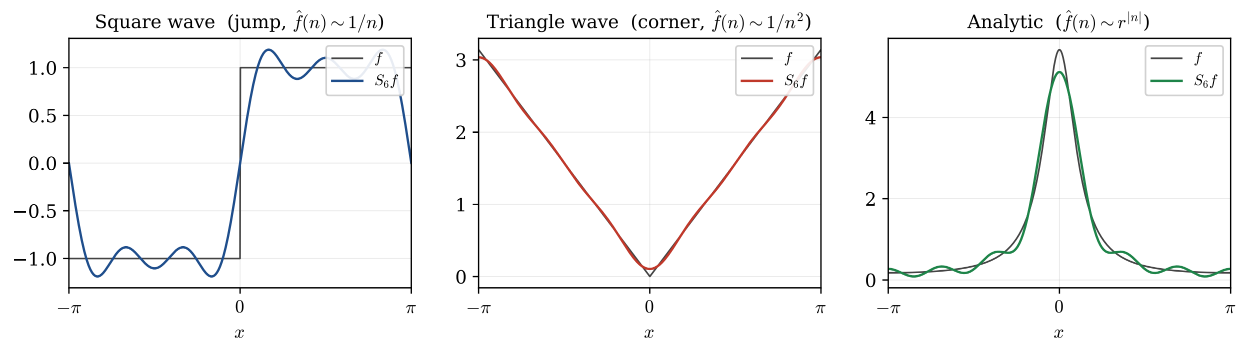


Figure 1: Reconstruction quality at a fixed budget of terms is governed by smoothness. Each partial sum uses frequencies up to $N = 6$; the analytic example is the Poisson kernel with $r = 0.7$. Left: a jump gives slow ($1/n$) decay and a poor, oscillatory fit. Center: a corner gives faster ($1/n^2$) decay and a good fit. Right: an analytic function gives exponential decay and a nearly exact fit.

**How to read this paper.** The argument is a chain, and each link is one short section. Section 2 sets up notation in plain terms. Section 3 shows coefficients always vanish. Section 4 is the heart of the paper: it proves that smoothness controls how *fast* they vanish. Section 5 turns decay into convergence. Section 6 looks closely at what happens at a jump. Section 7 establishes the Gibbs overshoot and computes its size. Section 8 works the three examples in full. Section 9 pauses for a famous application, evaluating the Basel sum $\sum 1/n^2$. Section 10 shows how averaging cures the bad behavior, and explains the reason. Section 11 steps back. Statements of the famous theorems are attributed to standard references such as Stein–Shakarchi [1], Katznelson [2], Körner [3], and Zygmund [6].

## 2 Setup and conventions

We work with $2\pi$-periodic functions, which we think of as functions on the circle $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ (going once around the circle is the same as moving by $2\pi$). For an integrable function $f$,[1] the *Fourier coefficients* are the numbers

$$\widehat{f}(n) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\, e^{-inx}\, dx, \qquad n \in \mathbb{Z}, \tag{2}$$

and the *Fourier series* of $f$, together with its symmetric partial sums, is

$$f \;\sim\; \sum_{n=-\infty}^{\infty} \widehat{f}(n)\, e^{inx}, \qquad S_N f(x) \;=\; \sum_{n=-N}^{N} \widehat{f}(n)\, e^{inx}. \tag{3}$$

Using the complex exponential $e^{inx} = \cos nx + i \sin nx$ keeps the bookkeeping short; the equivalent "real" form is

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}\bigl(a_n \cos nx + b_n \sin nx\bigr), \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f \cos nx\, dx, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f \sin nx\, dx,$$

where $\widehat{f}(0) = a_0/2$ is just the average value of $f$, and for $n > 0$ one has $\widehat{f}(\pm n) = \frac{1}{2}(a_n \mp i b_n)$. The partial sum $S_N f$ is exactly the reconstruction "using ingredients up to frequency $N$."

It is convenient to measure sizes and overlaps with the *normalized* inner product and norms

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,\overline{g(x)}\, dx, \qquad \|f\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2\, dx, \qquad \|f\|_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|\, dx.$$

(The complex conjugate on $g$ is essential.) The point of this choice is that the pure tones $e_n(x) = e^{inx}$ then form an *orthonormal* family: $\langle e_m, e_n\rangle = 1$ if $m = n$ and 0 otherwise. In this language the coefficient $\widehat{f}(n) = \langle f, e_n\rangle$ literally measures "how much of the tone $e^{inx}$ is present in $f$." Two standard facts will be used freely [1]: *Bessel's inequality* $\sum_n |\widehat{f}(n)|^2 \le \|f\|_2^2$, and, for finite-energy $f$, *Parseval's identity*

$$\sum_{n=-\infty}^{\infty} \left|\widehat{f}(n)\right|^2 \;=\; \|f\|_2^2\,, \tag{4}$$

which says the total energy of $f$ equals the total energy of its coefficients, a conservation law for the recipe.

## 3 Coefficients must vanish: the Riemann–Lebesgue lemma

The first structural fact costs us nothing in smoothness: for *any* reasonable function, the high-frequency ingredients become negligible.

**Theorem 3.1** (Riemann–Lebesgue [1, 2]). *If $f \in L^1(\mathbb{T})$, then $\widehat{f}(n) \to 0$ as $|n| \to \infty$.*

[1] $L^1(\mathbb{T})$ is the class of *integrable* functions, those with finite total area $\frac{1}{2\pi}\int_{-\pi}^{\pi}|f| < \infty$; $L^2(\mathbb{T})$ is the smaller class of *finite-energy* functions, with $\frac{1}{2\pi}\int_{-\pi}^{\pi}|f|^2 < \infty$. Every continuous function on the circle belongs to both, so a first-time reader may safely picture "reasonable" functions throughout.

*Proof.* From (2) we have the crude bound $\left|\widehat{f}(n)\right| \le \|f\|_1$, valid for every $n$. Fix a tolerance $\varepsilon > 0$. Smooth functions are dense in $L^1(\mathbb{T})$, so we may pick $g \in C^1(\mathbb{T})$ with $\|f-g\|_1 < \varepsilon$. For the smooth $g$, one integration by parts (the calculation is Lemma 4.1 below) gives $\widehat{g}(n) = \widehat{g'}(n)/(in)$ for $n \neq 0$, hence

$$|\widehat{g}(n)| \;\le\; \frac{\|g'\|_1}{|n|} \;\xrightarrow[|n|\to\infty]{}\; 0.$$

Now compare $f$ with $g$: since coefficients depend linearly and continuously on the function, $\left|\widehat{f}(n)\right| \le \left|\widehat{f}(n) - \widehat{g}(n)\right| + |\widehat{g}(n)| \le \|f-g\|_1 + |\widehat{g}(n)|$. Letting $|n| \to \infty$ gives $\limsup_{|n|\to\infty} \left|\widehat{f}(n)\right| \le \varepsilon$, and since $\varepsilon$ was arbitrary the limit is 0. ☐

*Remark* 3.2. The lemma is purely *qualitative*: it says the coefficients vanish, but not how quickly. The rate is exactly what regularity controls, and recovering it is the business of the next section. Notice that the proof already contains the key mechanism in miniature: borrowing one derivative from $g$ produced one factor of $1/n$.

## 4 Smoothness controls the rate of decay

Everything in this paper runs on a single identity, which says that differentiating a function multiplies its $n$-th coefficient by $in$.

**Lemma 4.1** (Differentiation rule). *Let $f$ be $2\pi$-periodic and continuously differentiable (more generally, absolutely continuous with $f' \in L^1$). Then*

$$\widehat{f'}(n) \;=\; in\,\widehat{f}(n), \qquad n \in \mathbb{Z}. \tag{5}$$

*Proof.* Integrate by parts in the definition of $\widehat{f'}(n)$:

$$\widehat{f'}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x)e^{-inx}\,dx = \frac{1}{2\pi}\Big[f(x)e^{-inx}\Big]_{-\pi}^{\pi} + \frac{in}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx.$$

The boundary term is $\frac{1}{2\pi}\big(f(\pi) - f(-\pi)\big)e^{-in\pi}$, and it vanishes because $f(\pi) = f(-\pi)$ by periodicity. What remains is exactly $in\,\widehat{f}(n)$. ☐

Read (5) backwards: if $f$ is the derivative of something smoother, then *its* coefficients are the smoother function's coefficients *divided* by $n$, so they are smaller. Iterating this turns derivatives directly into powers of $1/n$.

**Theorem 4.2** (Smoothness ⇒ polynomial decay [1]). *If $f$ is $2\pi$-periodic and has $k$ continuous derivatives ($f \in C^k(\mathbb{T})$), then*

$$\widehat{f}(n) \;=\; \frac{\widehat{f^{(k)}}(n)}{(in)^k}, \qquad n \neq 0, \tag{6}$$

*and consequently*

$$|n|^k \left|\widehat{f}(n)\right| = \left|\widehat{f^{(k)}}(n)\right| \xrightarrow[|n|\to\infty]{} 0, \qquad i.e. \qquad \widehat{f}(n) = o\big(|n|^{-k}\big).$$

*If in addition $f^{(k)}$ is bounded, then* $\left|\widehat{f}(n)\right| \le \left\|f^{(k)}\right\|_\infty |n|^{-k} = O\big(|n|^{-k}\big)$.

*Proof.* Apply Lemma 4.1 repeatedly: $\widehat{f}(n) = \widehat{f'}(n)/(in) = \widehat{f''}(n)/(in)^2 = \cdots = \widehat{f^{(k)}}(n)/(in)^k$, which is (6). The decay statement is the Riemann–Lebesgue lemma (Theorem 3.1) applied to the integrable function $f^{(k)}$. The bounded case follows from $\left|\widehat{f^{(k)}}(n)\right| \le \left\|f^{(k)}\right\|_\infty$. □

*Remark* 4.3 (Exactly how much smoothness is needed). The hypothesis "$f \in C^k$" is the cleanest, but slightly more than necessary. What the proof actually uses is that one may integrate by parts $k$ times: it is enough that $f^{(k-1)}$ be *absolutely continuous* with $f^{(k)} \in L^1(\mathbb{T})$ (in the language of Sobolev spaces, $f \in W^{k,1}(\mathbb{T})$), with the periodic boundary values matching. Absolute continuity is precisely the condition that makes the boundary terms behave and the fundamental theorem of calculus apply. We will lean on this sharper form in the triangle-wave example, where the function has a corner and so is *not* $C^1$.

The slogan is exact: *each derivative the function possesses buys one factor of* $1/n$ *in its coefficients.* A function that is infinitely differentiable therefore has coefficients decaying faster than every power of $1/n$. At the very top of the smoothness hierarchy sit the *analytic* functions (those given locally by a convergent power series), and they are detected by exponential decay.

**Proposition 4.4** (Analyticity ⇔ exponential decay [2]). *Let $f$ be $2\pi$-periodic.*

(i) *If $f$ extends to a bounded holomorphic (complex-differentiable) function on the strip $\{z \in \mathbb{C} : |\operatorname{Im} z| < a\}$, then for every $a' < a$ there is a constant $C$ with $\left|\widehat{f}(n)\right| \le Ce^{-a'|n|}$.*

(ii) *Conversely, if $\left|\widehat{f}(n)\right| \le Ce^{-a|n|}$ for some $a > 0$, then the series* (3) *converges to a function $g$ that extends holomorphically (hence is real-analytic) to the strip $|\operatorname{Im} z| < a$, and $g = f$ almost everywhere; if $f$ is continuous, then $f = g$ everywhere.*

*Sketch.* For (i), the integrand in (2) is holomorphic and $2\pi$-periodic, so Cauchy's theorem lets us slide the contour from the real axis to the line $\operatorname{Im} z = -\sigma \operatorname{sgn}(n)$ for any $\sigma < a$ without changing the integral. On that line $\left|e^{-inz}\right| = e^{-|n|\sigma}$, which gives $\left|\widehat{f}(n)\right| \le (\sup_{\text{strip}} |f|)\, e^{-\sigma|n|}$. For (ii), the exponential bound makes $\sum_n \widehat{f}(n)e^{inz}$ converge uniformly on every smaller strip $|\operatorname{Im} z| \le a' < a$ (since $\left|\widehat{f}(n)e^{inz}\right| \le Ce^{-(a-a')|n|}$ is summable there) to a holomorphic limit $g$; its restriction to the real axis has the same coefficients as $f$, so $g = f$ almost everywhere by uniqueness (Corollary 10.2). □

Theorem 4.2 and Proposition 4.4 together establish the *first* arrow of the chain (1): regularity determines decay, on a scale that runs from slow polynomial decay (few derivatives) all the way up to exponential decay (analyticity).

# 5 Decay controls convergence

Now the *second* arrow: from decay to convergence. First a word on what "convergence" should mean. The weakest useful notion is *pointwise* convergence: $S_N f(x) \to f(x)$ at each individual point $x$. A much stronger and more desirable notion is *uniform* convergence: the worst-case error $\sup_x |S_N f(x) - f(x)|$ over the *whole* circle tends to 0, so there are no stubborn bad spots anywhere. The basic principle of this section is that fast decay buys uniform convergence.

**Proposition 5.1** (Absolute summability ⇒ uniform convergence). *If $\sum_n \left|\widehat{f}(n)\right| < \infty$, then $S_N f$ converges uniformly on $\mathbb{T}$ to a continuous function. If moreover $f$ is continuous, the limit is $f$ itself.*

*Proof.* Each term satisfies $\sup_x \left|\widehat{f}(n)e^{inx}\right| = \left|\widehat{f}(n)\right|$, so by the Weierstrass $M$-test the series (3) converges uniformly to some continuous function $g$. Because uniform convergence lets us integrate term by term, $g$ has the same Fourier coefficients as $f$; and a continuous function is determined by its coefficients (uniqueness, Corollary 10.2), so $g = f$ whenever $f$ is continuous. □

Combined with the decay theorem this already settles smooth functions: if $f \in C^2(\mathbb{T})$ then $\left|\widehat{f}(n)\right| = O(n^{-2})$, which is summable (because $\sum 1/n^2$ converges), so the Fourier series converges uniformly. But we can do better with less: a *single* derivative already suffices.

**Theorem 5.2** ($C^1 \Rightarrow$ uniform convergence [1])**.** *Let $f$ be $2\pi$-periodic and continuously differentiable (more generally, absolutely continuous with $f' \in L^2(\mathbb{T})$). Then $\sum_n \left|\widehat{f}(n)\right| < \infty$, so the Fourier series of $f$ converges to $f$ absolutely and uniformly.*

*Proof.* Use the differentiation rule (5) to write $\left|\widehat{f}(n)\right| = \left|\widehat{f'}(n)\right| / |n|$ for $n \neq 0$, then apply the Cauchy–Schwarz inequality:

$$\sum_{n\neq 0} \left|\widehat{f}(n)\right| = \sum_{n\neq 0} \frac{1}{|n|} \left|\widehat{f'}(n)\right| \leq \left(\sum_{n\neq 0} \frac{1}{n^2}\right)^{1/2} \left(\sum_{n\neq 0} \left|\widehat{f'}(n)\right|^2\right)^{1/2}.$$

The first factor is the finite number $(\pi^2/3)^{1/2}$, since $\sum_{n\neq 0} n^{-2} = 2\zeta(2) = \pi^2/3$. By Parseval's identity (4) the second factor equals $\|f'\|_2 < \infty$. Hence the coefficients are absolutely summable, and Proposition 5.1 finishes the proof. □

*Remark* 5.3 (A clean sufficient condition)*.* Theorem 5.2 is a *sufficient* condition for uniform convergence, and a particularly clean one, though it is not a sharp dividing line (there are non-differentiable functions whose Fourier series still converge uniformly). What *is* decisive for the absolute method is summability of the coefficients. Decay like $1/n^2$ is summable, so a function with that rate (a triangle wave, say) converges uniformly; decay like $1/n$ is *not* summable, so this route fails for a square wave. This is no accident: a uniform limit of continuous functions must be continuous, so the Fourier series of a *discontinuous* function can *never* converge uniformly. The slow $1/n$ tail and the discontinuity are two faces of the same coin, and their shared signature is the Gibbs phenomenon of Section 7.

# 6 A close look at a jump

What happens exactly at a discontinuity, where uniform convergence is off the table? To answer this we rewrite the partial sum as an averaging operation. Summing the geometric series in (3) gives

$$S_N f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t)\, D_N(t)\, dt, \qquad D_N(t) = \sum_{n=-N}^{N} e^{int} = \frac{\sin\big((N+\frac{1}{2})t\big)}{\sin(t/2)},$$

where $D_N$ is the *Dirichlet kernel.* Think of $D_N$ as the "lens" through which the partial sum views $f$: $S_N f(x)$ is a weighted average of nearby values $f(x-t)$, with weights $D_N(t)$. The lens has total weight 1, that is, $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N\, dt = 1$.

**Theorem 6.1** (Dirichlet [1, 6]). *Let $f \in L^1(\mathbb{T})$, and suppose that at a point $x$ the one-sided limits $f(x^+), f(x^-)$ exist and $f$ is well-behaved there (for instance, it has a left-hand and a right-hand derivative). Then*

$$S_N f(x) \xrightarrow[N\to\infty]{} \frac{f(x^+)+f(x^-)}{2}.$$

*In particular, at a point of continuity the series converges to $f(x)$; at a jump, it converges to the* midpoint *of the two one-sided values.*

*Sketch.* Because $\frac{1}{2\pi}\int_0^\pi D_N = \frac{1}{2}$ and $D_N$ is even, the difference between $S_N f(x)$ and the proposed midpoint limit equals

$$\frac{1}{2\pi}\int_0^\pi \Big(\big[f(x+t)-f(x^+)\big]+\big[f(x-t)-f(x^-)\big]\Big)\,\frac{\sin\big((N+\frac{1}{2})t\big)}{\sin(t/2)}\,dt.$$

The one-sided derivative hypothesis means each bracketed difference shrinks at least linearly in $t$, so dividing by $\sin(t/2)$ leaves an *integrable* function of $t$ near 0. The Riemann–Lebesgue lemma applied to that function sends the whole integral to 0. □

*Remark* 6.2 (Why the partial sums can misbehave). The Dirichlet lens is a treacherous one. Unlike an honest averaging weight it *takes negative values* (Figure 4, left), and its total size (the *Lebesgue constant*) grows without bound [6]:

$$L_N = \frac{1}{2\pi}\int_{-\pi}^{\pi}|D_N(t)|\,dt = \frac{4}{\pi^2}\log N + O(1).$$

Negative weights let the average overshoot the data, and the growing size lets the overshoot persist. This is the analytic reason partial sums can converge slowly and ring near jumps. As we will see, it is exactly what Fejér's averaging repairs.

# 7 The Gibbs phenomenon

Dirichlet's theorem tells us the partial sums converge to the midpoint *at* a jump. It says nothing about their behavior in a shrinking neighborhood *near* the jump, and there something remarkable and permanent appears.

The cleanest example is the sawtooth

$$g(x) = \sum_{n=1}^{\infty}\frac{\sin nx}{n} = \frac{\pi - x}{2}, \qquad 0 < x < 2\pi, \tag{7}$$

extended $2\pi$-periodically. It has a jump of height $\pi$ at $x=0$, with $g(0^+) = \pi/2$ and $g(0^-) = -\pi/2$.

**Theorem 7.1** (Gibbs, model case [3, 6]). *Let $S_N g(x) = \sum_{n=1}^{N}\frac{\sin nx}{n}$. Then*

$$\lim_{N\to\infty} S_N g\Big(\frac{\pi}{N}\Big) = \int_0^\pi \frac{\sin t}{t}\,dt = \mathrm{Si}(\pi) \approx 1.851937 > \frac{\pi}{2} \approx 1.570796.$$

*So the partial sums overshoot the upper value $g(0^+) = \pi/2$ by the fixed amount* $\mathrm{Si}(\pi) - \frac{\pi}{2} \approx 0.2811$*, which is a fraction*

$$\frac{\mathrm{Si}(\pi)}{\pi} - \frac{1}{2} \approx 0.0895$$

*of the jump height $\pi$. This overshoot of about* 8.95% *does* not *depend on $N$; only its location $x_N = \pi/N \to 0$ slides toward the discontinuity.*

*Proof.* Evaluate the partial sum at the special point $x = \pi/N$ and recognize a Riemann sum:

$$S_N g\Big(\frac{\pi}{N}\Big) = \sum_{n=1}^{N} \frac{\sin(n\pi/N)}{n} = \sum_{n=1}^{N} \frac{\sin(n\pi/N)}{n\pi/N} \cdot \frac{\pi}{N}.$$

With sample points $t_n = n\pi/N$ and mesh $\Delta t = \pi/N$, the right-hand side is a Riemann sum for $\int_0^\pi \frac{\sin t}{t}\, dt$ (the integrand extends continuously by $\frac{\sin t}{t} \to 1$ as $t \to 0$). Therefore $S_N g(\pi/N) \to \mathrm{Si}(\pi)$. Since $g$ approaches $\pi/2$ from the right, the stated overshoot follows by subtraction. □

For a square wave of amplitude $\pm 1$ the same computation gives a peak of $\frac{2}{\pi}\,\mathrm{Si}(\pi) \approx 1.1790$, i.e. an overshoot of about 17.9% of the half-height, equivalently 8.95% of the full jump on each side. Figure 2 shows this directly: as $N$ increases through $11, 41, 161$, the ringing crowds ever more tightly against the jump, yet the height of the first spike refuses to shrink.

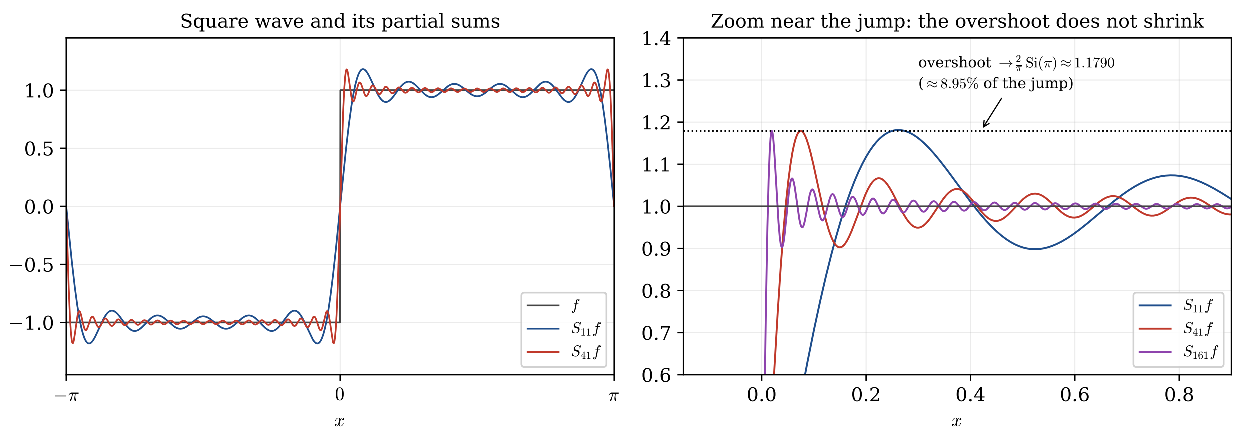


Figure 2: The Gibbs phenomenon for the square wave (amplitude $\pm 1$). Left: the function and two partial sums. Right: a zoom on the jump at $x = 0$. For every $N$ the first overshoot approaches $\frac{2}{\pi}\,\mathrm{Si}(\pi) \approx 1.1790$; raising $N$ narrows the spike but never lowers it.

*Remark* 7.2 (The same overshoot at every jump). The constant above is not special to the sawtooth. At *any* jump discontinuity of a piecewise-smooth function, the partial sums overshoot by the same universal fraction of the local jump, because near the jump the function looks like a scaled, shifted copy of the sawtooth model; see [6]. The 8.95% figure is therefore a genuine feature of Fourier reconstruction, not an artifact of one example.

*Remark* 7.3 (Overshoot versus energy). The Gibbs overshoot is the visible face of Remark 5.3: the non-summable $1/n$ tail of a jump function blocks uniform convergence in every neighborhood of the discontinuity. It does *not* block convergence in *energy*, however. Because the spike's width shrinks like $1/N$ while its height stays bounded, its contribution to the total energy tends to 0, and indeed $S_N f \to f$ in $L^2(\mathbb{T})$ for every finite-energy $f$.

## 8 Three case studies

We now run the chain (1) on three functions of increasing smoothness. Their coefficients realize the trichotomy $1/n$ vs. $1/n^2$ vs. exponential, and Figure 3 displays all three rates on one set of axes.

**Example 8.1** (Square wave: a jump, decay $1/n$)**.** Let $f(x) = \operatorname{sgn}(\sin x)$, which equals $+1$ on $(0, \pi)$ and $-1$ on $(-\pi, 0)$. It is an odd function, so all $a_n = 0$, and we compute the sine coefficients once, in full. Since $f(x)\sin nx$ is even and $f = 1$ on $(0, \pi)$,

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = \frac{2}{\pi}\int_0^{\pi}\sin nx\,dx = \frac{2}{\pi}\left[\frac{-\cos nx}{n}\right]_0^{\pi} = \frac{2}{\pi n}\big(1-(-1)^n\big).$$

Thus

$$b_n = \begin{cases} \dfrac{4}{\pi n}, & n \text{ odd}, \\ 0, & n \text{ even}, \end{cases} \qquad \text{so} \qquad f(x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin\big((2k+1)x\big)}{2k+1}.$$

Only the odd harmonics appear, and the nonzero ones satisfy $|b_{2k+1}| = \frac{4}{\pi(2k+1)}$, i.e. they decay at the rate $1/n$, the slow rate forced by the jumps at $0, \pm\pi, \ldots$. The coefficients are not absolutely summable, the convergence is not uniform, and the Gibbs overshoot is present. At each jump the partial sums converge to the midpoint 0, exactly as Theorem 6.1 predicts.

**Example 8.2** (Triangle wave: a corner, decay $1/n^2$)**.** Let $f(x) = |x|$ on $[-\pi, \pi]$, extended periodically. It is *continuous* but has corners at $0, \pm\pi, \ldots$, so it is *not* $C^1$, and we cannot apply Theorem 4.2 with $k = 2$ directly. Instead we use the differentiation rule once. The derivative is $f'(x) = \operatorname{sgn}(x)$ on $(-\pi, \pi)$, essentially the square wave of Example 8.1: a bounded, piecewise constant function (of bounded variation) whose coefficients are of size $O(1/n)$. By the differentiation rule (5), $\widehat{f}(n) = \widehat{f'}(n)/(in)$, so

$$\left|\widehat{f}(n)\right| = O\left(\frac{1}{n^2}\right).$$

A direct computation gives the explicit series

$$f(x) = \frac{\pi}{2} - \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\cos\big((2k+1)x\big)}{(2k+1)^2},$$

with constant term $\pi/2$ (the average of $|x|$) and only odd cosine coefficients, of size $\frac{4}{\pi(2k+1)^2}$. The single corner costs exactly one power of $n$ compared with the discontinuous square wave. Being summable, the series converges absolutely and uniformly (Proposition 5.1); there is no Gibbs phenomenon.

**Example 8.3** (Analytic: exponential decay)**.** The *Poisson kernel*

$$P_r(x) = \frac{1-r^2}{1-2r\cos x + r^2} = \sum_{n=-\infty}^{\infty} r^{|n|}e^{inx}, \qquad 0 < r < 1,$$

has coefficients $\widehat{f}(n) = r^{|n|}$ that decay *exponentially*. By Proposition 4.4 this is the signature of a real-analytic function; indeed $P_r$ extends holomorphically to the strip $|\operatorname{Im} z| < \log(1/r)$. Its Fourier series converges geometrically fast and uniformly, so a handful of terms already gives an essentially exact reconstruction; this is the right panel of Figure 1, drawn with $r = 0.7$.

| Function | Regularity | Coefficient decay | Convergence |
|---|---|---|---|
| Square wave | jump discontinuity | $\asymp 1/n$ (odd $n$) | pointwise, to midpoints; Gibbs |
| Triangle wave | continuous, corners | $\asymp 1/n^2$ (odd $n$) | absolute, uniform; no Gibbs |
| Analytic ($P_r$) | real-analytic | $\asymp r^{\lvert n \rvert}$ | uniform, geometrically fast |

Table 1: The chain (1) across three regularity classes (showing the nonzero modes). More smoothness yields faster decay yields stronger convergence.

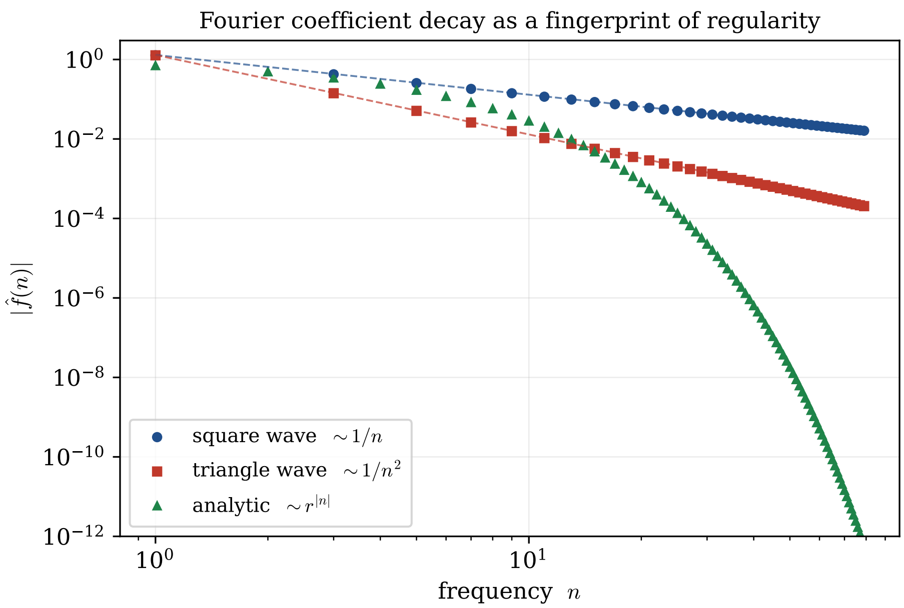


Figure 3: Coefficient magnitudes on a log–log scale, plotting only the *nonzero* modes: odd $n$ for the square and triangle waves, and all $n$ for the Poisson kernel ($r = 0.7$). The square wave's coefficients lie along a line of slope $-1$, the triangle wave's along a line of slope $-2$, and the analytic function's plunge below every power law, reflecting exponential decay. Fourier-coefficient decay is a quantitative signature of regularity and, under Sobolev or analytic hypotheses, an exact characterization of it.

# 9 An application: the Basel problem

So far we have used Fourier coefficients to *describe* functions: how smooth they are, how well their series converge. But the same machinery can also *compute*. As a striking illustration, we evaluate one of the most famous infinite sums in mathematics,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots ,$$

whose value Euler first found in 1735; the challenge of summing it is known as the *Basel problem*. It is not obvious that this has anything to do with waves at all, and yet a single application of Parseval's identity (4) delivers the answer. Better still, the number we obtain is exactly the constant the paper already leaned on in Theorem 5.2.

The tool is the sawtooth once more. Let $f$ be the $2\pi$-periodic function with

$$f(x) = x, \qquad -\pi < x < \pi$$

(the same shape, up to a shift, as the sawtooth of Section 7); it has a jump at $x = \pm\pi$, and so, in keeping with the theme of the paper, its coefficients decay only like $1/n$.

**Lemma 9.1.** *The Fourier coefficients of $f$ are $\widehat{f}(0) = 0$ and*

$$\widehat{f}(n) = \frac{(-1)^{n+1}}{in} \quad (n \neq 0), \qquad \textit{so that} \qquad \left|\widehat{f}(n)\right|^2 = \frac{1}{n^2}.$$

*Proof.* Since $f$ is odd, its average value is $\widehat{f}(0) = 0$. For $n \neq 0$, integrate by parts with $u = x$ and $dv = e^{-inx}\,dx$:

$$\widehat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} x\,e^{-inx}\,dx = \frac{1}{2\pi}\left(\left[\frac{x\,e^{-inx}}{-in}\right]_{-\pi}^{\pi} + \frac{1}{in}\int_{-\pi}^{\pi} e^{-inx}\,dx\right).$$

The remaining integral is 0, and using $e^{\pm in\pi} = (-1)^n$ the boundary term equals $\frac{1}{2\pi} \cdot \frac{2\pi(-1)^{n+1}}{in}$. Hence $\widehat{f}(n) = (-1)^{n+1}/(in)$, and taking absolute values (with $|in| = n$) gives $\left|\widehat{f}(n)\right|^2 = 1/n^2$. $\square$

Now we feed this into Parseval. On one side, the energy of $f$ is

$$\|f\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} x^2\,dx = \frac{1}{2\pi} \cdot \frac{2\pi^3}{3} = \frac{\pi^2}{3}.$$

On the other side, since $\left|\widehat{f}(n)\right|^2 = 1/n^2$ for $n \neq 0$ and the coefficients are symmetric in $\pm n$,

$$\sum_{n=-\infty}^{\infty} \left|\widehat{f}(n)\right|^2 = \sum_{n\neq 0} \frac{1}{n^2} = 2\sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Parseval's identity (4) equates these two sides, so $2\sum_{n\geq 1} n^{-2} = \pi^2/3$, which is the celebrated evaluation.

**Theorem 9.2** (Basel problem, Euler 1735)**.**

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

*Remark* 9.3 (The constant we already used)*.* This is precisely the value invoked inside the proof of Theorem 5.2, where we needed $\sum_{n\neq 0} n^{-2} = 2\zeta(2) = \pi^2/3$. There the sum was the "price" of trading one derivative for faster decay; here Parseval pins down its exact value. It is worth pausing over what happened: a fact about the *energy* of a simple straight-line ramp turned into the exact value of a sum over every positive integer.

*Remark* 9.4 (The same idea with a smoother function)*.* A similar computation with the *triangle* wave $f(x) = |x|$ of Example 8.2, whose coefficients decay like $1/n^2$, squares those coefficients to $1/n^4$ and yields the companion evaluation

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}.$$

This is no coincidence: the smoother the function, the faster its coefficients decay, and the higher the power of $n$ whose reciprocal sum Parseval can reach. Euler's values of $\sum 1/n^{2k}$ all arise in this way, another echo of the paper's central theme that regularity is stored in the size of the coefficients.

## 10 Taming the oscillations: Fejér's theorem

The Gibbs overshoot and the slow convergence of rough functions are defects of the *ordinary* partial sums $S_N$. A classical and beautiful remedy, due to Fejér, is simply to *average* them. Define the *Cesàro means*

$$\sigma_N f = \frac{1}{N+1}\sum_{n=0}^{N} S_n f = F_N * f, \qquad F_N(t) = \frac{1}{N+1}\left(\frac{\sin\big((N+1)t/2\big)}{\sin(t/2)}\right)^2.$$

Here $\sigma_N f$ is the running average of the first $N+1$ partial sums, and it is again an averaging of $f$ against a lens, the *Fejér kernel* $F_N$. The decisive contrast with the Dirichlet kernel is visible in Figure 4: whereas $D_N$ oscillates and dips below zero, $F_N$ is *nonnegative*. Together with $\frac{1}{2\pi}\int_{-\pi}^{\pi} F_N = 1$ and the concentration of its weight near $t = 0$, this makes $\{F_N\}$ an honest averaging family (a *good kernel*, or approximate identity).

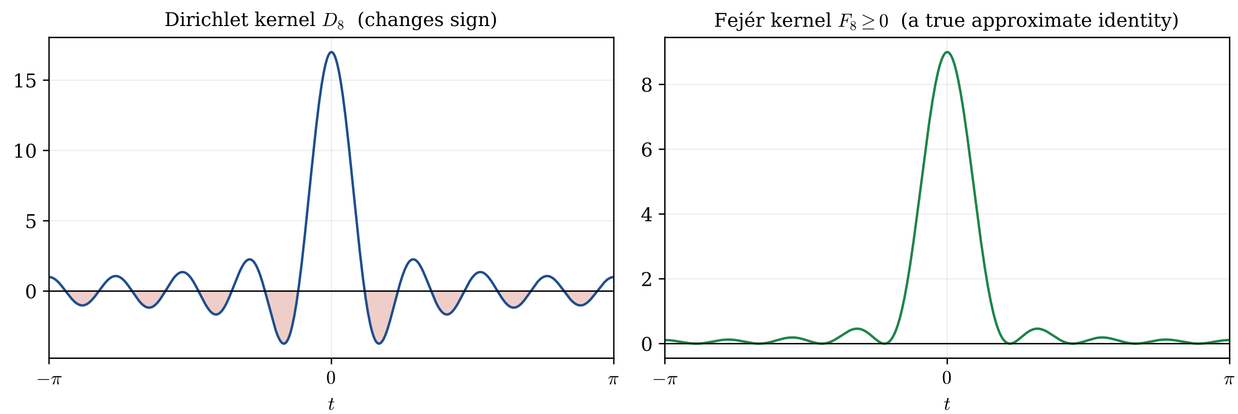


Figure 4: The two lenses, drawn at $N = 8$. The Dirichlet kernel $D_8$ (left) changes sign, with its negative lobes shaded, and its total size grows like $\log N$. The Fejér kernel $F_8$ (right) is nonnegative. This single difference accounts for both the Gibbs overshoot and its disappearance under averaging.

**Theorem 10.1** (Fejér [1, 2])**.** *If $f \in C(\mathbb{T})$ then $\sigma_N f \to f$ uniformly on $\mathbb{T}$. More generally, if $f \in L^1(\mathbb{T})$ and the one-sided limits $f(x^{\pm})$ exist at a point $x$, then $\sigma_N f(x) \to \frac{1}{2}\big(f(x^+) + f(x^-)\big)$; and for $1 \leq p < \infty$, $\sigma_N f \to f$ in $L^p(\mathbb{T})$ whenever $f \in L^p(\mathbb{T})$.*

*Sketch.* Because $F_N \geq 0$, has total weight $\frac{1}{2\pi}\int F_N = 1$, and for each $\delta > 0$ has vanishing weight away from the origin ($\frac{1}{2\pi}\int_{\delta\leq|t|\leq\pi} F_N \to 0$), the family $\{F_N\}$ is a good kernel. For continuous $f$, split $\sigma_N f(x) - f(x) = \frac{1}{2\pi}\int F_N(t)\big(f(x-t) - f(x)\big)\,dt$ into the piece $|t| < \delta$, where uniform continuity makes the integrand small, and the piece $|t| \geq \delta$, where the vanishing tail weight controls it. Uniformity comes from uniform continuity of $f$ on the compact circle. The remaining statements follow from the same good-kernel argument in the relevant norm. $\square$

Fejér's theorem repays several debts at once.

**Corollary 10.2** (Weierstrass approximation; uniqueness)**.** *Trigonometric polynomials are dense in $C(\mathbb{T})$. Consequently, if $f \in L^1(\mathbb{T})$ has $\widehat{f}(n) = 0$ for all $n$, then $f = 0$ almost everywhere.*

*Proof.* The first statement is immediate: $\sigma_N f$ is a trigonometric polynomial that converges uniformly to $f$. For uniqueness, if every coefficient vanishes then $\sigma_N f \equiv 0$, so the $L^1$ case of Theorem 10.1 gives $f = \lim \sigma_N f = 0$ in $L^1$, hence almost everywhere. (This is the uniqueness fact used in Propositions 4.4 and 5.1.) □

The conceptual payoff, though, is the *explanation* of why averaging works.

*Remark* 10.3 (Positivity kills the overshoot). For a real-valued bounded $f$, the fact that $F_N \geq 0$ and $\frac{1}{2\pi}\int F_N = 1$ makes $\sigma_N f$ a genuine weighted average of the values of $f$:

$$\inf_x f \;\leq\; \sigma_N f(x) \;\leq\; \sup_x f \qquad \text{for all } x, N.$$

An average can never exceed the data it averages, so $\sigma_N f$ *cannot overshoot*: the Gibbs phenomenon simply does not occur for Cesàro means. By contrast, the Dirichlet kernel's negative parts (Remark 6.2) let $S_N f$ overshoot, and its growing Lebesgue constants let the overshoot persist. Figure 5 shows the two side by side on the square wave: $S_N$ rings and overshoots, while $\sigma_N$ stays within the range of $f$ and never overshoots. The *same* structural property, the sign of the lens, explains both the disease of Section 7 and its cure here.

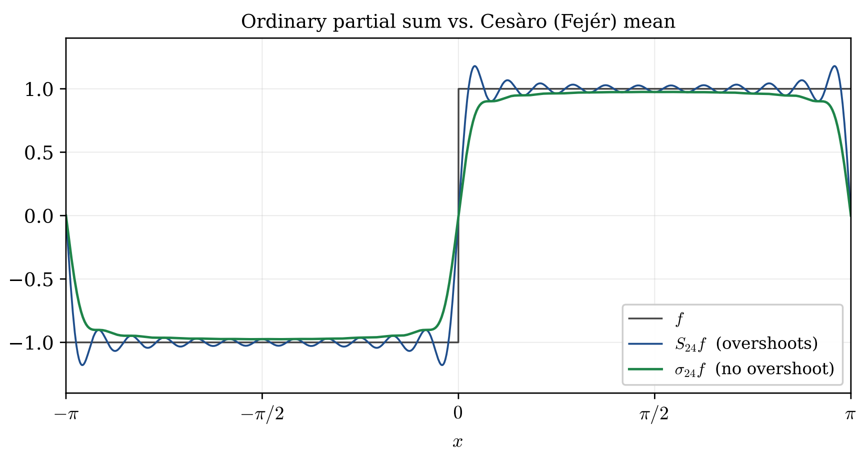


Figure 5: Square wave: the ordinary partial sum $S_{24}f$ overshoots near the jumps, while the Cesàro (Fejér) mean $\sigma_{24}f$ stays within the range of $f$ and shows no overshoot, illustrating Remark 10.3.

## 11 Conclusion

We have followed a single thread from smoothness to convergence. The Riemann–Lebesgue lemma guarantees that Fourier coefficients fade to zero; integration by parts sharpens this into a quantitative dictionary in which each derivative contributes a factor of $1/n$, so that a function's smoothness can be read straight off the decay rate of its coefficients: polynomial for finitely differentiable functions, exponential for analytic ones. Decay in turn controls convergence: absolute summability of the coefficients yields uniform convergence, and a single continuous derivative already secures it. At a jump, where uniform convergence is impossible, Dirichlet's theorem locates the limit at the midpoint while the Gibbs phenomenon exposes a permanent overshoot of about 8.95% of the jump. The three case studies (square, triangle, and analytic) make the $1/n$, $1/n^2$, exponential trichotomy concrete. Finally, Fejér's theorem shows that averaging the partial sums restores uniform convergence and erases the overshoot, for a reason that is structural rather than accidental: the Fejér kernel is nonnegative where the Dirichlet kernel is not.

One caution worth stating plainly: Fejér's averaging removes the overshoot, but at a jump it still converges to the *midpoint* of the one-sided values, just as the ordinary partial sums do, not to whatever value the function may have been assigned at the jump itself. Averaging improves *how* the series converges, not the limit it is allowed to reach at a discontinuity.

Several natural directions extend this circle of ideas. The finite-energy ($L^2$) theory turns Parseval's identity into a perfect correspondence and, in the form of *Sobolev spaces*, makes "decay $\leftrightarrow$ regularity" an exact equivalence: $f$ lies in the Sobolev space $H^s$ precisely when $\sum_n (1+n^2)^s \left|\widehat{f}(n)\right|^2 < \infty$ [4]. The far subtler question of *pointwise* convergence for general finite-energy functions was settled by Carleson's theorem on almost-everywhere convergence [7]. And the same philosophy (local regularity governs the size of transform coefficients) reappears, in sharper and more localized form, in the theory of wavelets.

## References


[1] E. M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, 2003.

[2] Y. Katznelson, *An Introduction to Harmonic Analysis*, 3rd ed., Cambridge University Press, 2004.

[3] T. W. Körner, *Fourier Analysis*, Cambridge University Press, 1988.

[4] L. Grafakos, *Classical Fourier Analysis*, 3rd ed., Springer, 2014.

[5] H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, 1972.

[6] A. Zygmund, *Trigonometric Series*, 3rd ed., Cambridge University Press, 2002.

[7] L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.