What Is Uniform Continuity? Neil Falkner, June 13, 2024

Abstract. Every introductory analysis textbook defines uniform continuity of a function on its entire domain. It is also worthwhile to define uniform continuity of a function on a subset of its domain. To my surprise, I have been unable to find a single textbook that defines this in the way that is most useful in applications. In this "What Is?" seminar talk, I explained how this concept should be defined and, to illustrate the value of defining it that way, I sketched the proof that the Fourier series of a function f of bounded variation on $[0, 1]$ converges to f uniformly on each compact subset of the set of points of continuity of f. This is a refinement of the Dirichlet-Jordan theorem. Along the way, I discussed Fejér's theorem and Hardy's Tauberian theorem, and an elementary proof for the rate of decay of the Fourier coefficients of a function of bounded variation on $[0, 1]$.

Remark. Notes covering the contents of this talk follow. They consist of a concatenation of pdfs that I already had written up in the course of the many times that I have taught real analysis, so the page numbers that appear are not consecutive throughout, but are the ones from the original separate pdfs.

Table of Contents.

Let X and Y be topological spaces, let $f\colon X \to Y$, and let $H \subseteq X$. What should it mean to say that f *is continuous on* H? Some textbooks, such as Apostol, *Mathematical Analysis*, Second Edition, or Royden, *Real Analysis*, First, Second, or Third Edition,[1] define it to mean that for each $a \in H$, f is continuous at a. This is the definition I prefer.[2] Unfortunately, some textbooks instead define continuity of f on H to mean that the restriction of f to H is continuous. Note that this is a strictly weaker condition. For instance, the so-called Dirichlet function $1_{\mathbf{Q}}$ is discontinuous at each point of $\mathbf{R}$ but its restriction to $\mathbf{Q}$ is continuous because it is constant.

Now let (X, ρ_X) and (Y, ρ_Y) be metric spaces, let $f\colon X \to Y$, and let $H \subseteq X$. What should it mean to say that f *is uniformly continuous on* H? By analogy with my preferred definition of continuity on H, I would say that it should mean that for each $\varepsilon > 0$, there exists $\delta > 0$ such that for each $a \in H$, for each $x \in X$, if $\rho_X(a,x) < \delta$, then $\rho_Y(f(a), f(x)) < \varepsilon$. Note that here a is required to be in H but x is not. However, I am not aware of any textbook that adopts this definition. Most textbooks either define uniform continuity of f on H only when H is the entire domain of f or define it to mean uniform continuity of the restriction of f to H. I think this is regrettable, because the usual proofs that continuity on a compact set implies uniform continuity establish the stronger notion of uniform continuity on H that I prefer and because this stronger notion is what is needed for certain important applications of uniform continuity, such as Fejér's theorem and the Dirichlet-Jordan theorem. (See below.)

Remark. Since the "right" definition of uniform continuity on H is not the usual definition, it is prudent to formulate the theorem about uniform continuity on a compact set in a way that avoids the possibility of confusion, as follows.

Theorem on Uniform Continuity on a Compact Set. *Let (X, ρ_X) and (Y, ρ_Y) be metric spaces, let $f\colon X \to Y$, let C be the set of points in X at which f is continuous, and let H be a compact subset of C. Then f is continuous at a uniformly for a in H. In other words, for each $\varepsilon > 0$, there exists $\delta > 0$ such that for each $a \in H$, for each $x \in X$, if $\rho_X(a,x) < \delta$, then $\rho_Y(f(a), f(x)) < \varepsilon$.*

Proof. Let $\varepsilon > 0$. Let $E = \{\, (x,r) \in H \times (0,\infty) : f[B_X(x,2r)] \subseteq B_Y(f(x), \varepsilon/2) \,\}$. By assumption, for each $x \in H$, f is continuous at x, so there exists $r \in (0,\infty)$ such that $(x,r) \in E$. Thus the collection $\{\, B_X(x,r) : (x,r) \in E \,\}$ is an open cover of H. Hence, since H is compact, there exists $n \in \mathbf{N}$ and there exist $(x_1, r_1), \ldots, (x_n, r_n) \in E$ such that $H \subseteq \bigcup_{j=1}^n B_X(x_j, r_j)$. Let $\delta = \min\{\, r_1, \ldots, r_n \,\}$. Then $\delta > 0$. Let $a \in H$ and let $x \in X$ with $\rho_X(a,x) < \delta$. Then $a \in B_X(x_j, r_j)$ for some $j \in \{\, 1, \ldots, n \,\}$. Now

$$\rho_X(x, x_j) \le \rho_X(x,a) + \rho_X(a, x_j) < \delta + r_j \le r_j + r_j = 2r_j.$$

Thus a and x both belong to $B_X(x_j, 2r_j)$, so $f(a)$ and $f(x)$ both belong to $B_Y(f(x_j), \varepsilon/2)$, so

$$\rho_Y(f(a), f(x)) \le \rho_Y(f(a), f(x_j)) + \rho_Y(f(x_j), f(x)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This completes the proof. ■

The Dirichlet-Jordan Theorem and Fejér's Theorem.

The Dirichlet-Jordan theorem (Dirichlet, 1829; Jordan, 1881) says in part that for each $f\colon \mathbf{R} \to \mathbf{C}$, if f is 1-periodic and locally of bounded variation and if C is the set of points at which f is continuous, then for each compact set $H \subseteq C$, $S_n f \to f$ uniformly on H, where $S_n f$ is the n-th symmetric partial sum of the Fourier series for f. Note that the set of points at which such a function f is discontinuous is countable but may be dense.

Fejér's theorem (1904) says in part that for each $f\colon \mathbf{R} \to \mathbf{C}$, if f is 1-periodic and locally integrable and if C is the set of points at which f is continuous, then for each compact set $H \subseteq C$, $\sigma_n f \to f$ uniformly on H, where $\sigma_n f$ is the average of the first $n+1$ symmetric partial sums $S_0 f, \ldots, S_n f$ of the Fourier series

[1] but not Royden and Fitzpatrick, *Real Analysis*, Fourth Edition,

[2] Here are two other books that also use the definition I prefer: Bartle and Sherbert, *Introduction to Real Analysis*; Ross, *Elementary Analysis: The Theory of Calculus.* These happen to be the two textbooks currently prescribed for our regular undergraduate analysis sequence Math 4547 and 4548.

for f. In other words, under the stated assumptions, the sequence $(S_n f)$ is uniformly Cesàro-convergent to f on H. The proof of this portion of Fejér's theorem, and the proof of the portion of the Dirichlet-Jordan theorem just stated, both depend on the theorem on uniform continuity on a compact set in the form presented above. It would not be enough to know that the restriction of f to H is uniformly continuous.

Hardy (1910) showed how to derive the Dirichlet-Jordan theorem from Fejér's theorem. He did this by proving a general summability result that is now known as Hardy's Tauberian theorem and which says that if the sequence (s_n) of partial sums of (u_k) is Cesàro-convergent to s and $u_k = O(1/k)$, then (s_n) is convergent to s. Here $s, u_1, u_2, \ldots$ may be elements of any normed linear space.

Appendix: Lebesgue Numbers.

The first half of the proof of the theorem on uniform continuity on a compact set can be adapted into a proof of the theorem on Lebesgue numbers, a result that is useful for many purposes.

Definition: Lebesgue Numbers. Let (X, ρ) be a metric space, let H be a subset of X, and let $\mathscr{U}$ be a collection of subsets of X which covers H. To say that *δ is a Lebesgue number for $\mathscr{U}$ and H* means that $\delta > 0$ and for each $a \in H$, there exists $U \in \mathscr{U}$ such that $B(a, \delta) \subseteq U$.

Theorem on Lebesgue Numbers. *Let (X, ρ) be a metric space, let H be a compact subset of X, and let $\mathscr{U}$ be a collection of open subsets of X which covers H. Then there exists a Lebesgue number δ for $\mathscr{U}$ and H.*

Proof. Let $E = \{\, (x, r) \in H \times (0, \infty) : B(x, 2r) \subseteq U \text{ for some } U \in \mathscr{U} \,\}$. For each $x \in H$, we have $x \in U$ for some $U \in \mathscr{U}$, and since U is open, $B(x, 2r) \subseteq U$ for some $r \in (0, \infty)$, so $(x, r) \in E$, and of course $x \in B(x, r)$. Thus $\{\, B(x, r) : (x, r) \in E \,\}$ is an open cover of H. Since H is compact, $H \subseteq \bigcup_{j=1}^{n} B(x_j, r_j)$ for some $n \in \mathbf{N}$ and some $(x_1, r_1), \ldots, (x_n, r_n) \in E$. Let $\delta = \min\{\, r_1, \ldots, r_n \,\}$. Then $\delta > 0$. Let $a \in H$. Then $a \in B(x_j, r_j)$ for some $j \in \{\, 1, \ldots, n \,\}$. By the definition of E, $B(x_j, 2r_j) \subseteq U$ for some $U \in \mathscr{U}$. Let $x \in B(a, \delta)$. Then

$$\rho(x, x_j) \leq \rho(x, a) + \rho(a, x_j) < \delta + r_j \leq r_j + r_j = 2r_j,$$

so $x \in B(x_j, 2r_j)$. Since $B(x_j, 2r_j) \subseteq U$, it follows that $x \in U$. This holds for each $x \in B(a, \delta)$. Therefore $B(a, \delta) \subseteq U$. ■

Remark. Next, we present an alternate proof of the theorem on uniform continuity on a compact set, based on the theorem on Lebesgue numbers.

Alternate Proof of Theorem on Uniform Continuity on a Compact Set. Let $\varepsilon > 0$. For each $b \in H$, let $V_b = B_Y(f(b), \frac{\varepsilon}{2})$, let U_b be the interior of $f^{-1}[V_b]$ in X, and observe that $b \in U_b$ because f is continuous at b. Let $\mathscr{U} = \{\, U_b : b \in H \,\}$. Then $\mathscr{U}$ is a collection of open subsets of X which covers H. Since H is compact, there exists a Lebesgue number δ for $\mathscr{U}$ and A. Then $\delta > 0$. Let $x \in X$ and $a \in H$. Suppose $\rho_X(a, x) < \delta$. Then $x \in B_X(a, \delta)$. Since δ is a Lebesgue number for $\mathscr{U}$ and A, we have $B_X(a, \delta) \subseteq U_b$ for some $b \in H$. Then a and x both belong to U_b, so $f(a)$ and $f(x)$ both belong to $V_b = B_Y(f(b), \frac{\varepsilon}{2})$, so $\rho_Y(f(a), f(x)) \leq \rho_Y(f(a), f(b)) + \rho_Y(f(b), f(x)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. ■

Remark. Finally, we present an alternate proof of the theorem on Lebesgue numbers. It is less elementary than the proof above, but it provides additional insight into why the theorem is true.

Alternate Proof of the Theorem on Lebesgue Numbers. If some $U \in \mathscr{U}$ is equal to X, then we may take $\delta = 1$. Suppose no $U \in \mathscr{U}$ is equal to X. Since H is compact, there exists $n \in \mathbf{N}$ and there exist $U_1, \ldots, U_n \in \mathscr{U}$ such that $H \subseteq \bigcup_{j=1}^{n} U_j$. For $j = 1, \ldots, n$, define f_j on H by $f_j(x) = \inf\{\, \rho(x, x') : x' \in X \setminus U_j \,\}$ and observe that $f_j \geq 0$, that $U_j = \{\, f_j > 0 \,\}$ (because $X \setminus U_j$ is closed), that $\{\, f_j = \infty \,\} = \varnothing$ (because $X \setminus U_j \neq \varnothing$), and that f_j is continuous (because, as is easy to check, $|f_j(x_1) - f_j(x_2)| \leq \rho(x_1, x_2)$ for all $x_1, x_2 \in H$.). Define f on H by $f(x) = \max\{\, f_1(x), \ldots, f_n(x) \,\}$. Then f is continuous, so f achieves a minimum at some point $x_* \in K$. Now $x_* \in U_{j_*}$ for some $j_* \in \{\, 1, \ldots, n \,\}$, so $f(x_*) \geq f_{j_*}(x_*) > 0$. Let $\delta = f(x_*)$. Then $\delta > 0$. For each $x \in H$, there exists $j \in \{\, 1, \ldots, n \,\}$ such that $f(x) = f_j(x)$, so $f_j(x) \geq \delta$, so for each $x' \in X \setminus U_j$, $\rho(x, x') \geq f_j(x) \geq \delta$, so $B(x, \delta) \subseteq U_j$. ■

Fourier Series Motivated Modern Analysis.

This is a course on the general topic of real analysis, so you might well ask "Why begin with the specific topic of Fourier series?" One answer is that while the tools of modern analysis have found widely varied uses, many of these tools were first developed in order to better understand Fourier series.

Preliminary Remarks about Fourier Series.

For each $k \in \mathbf{Z}$, define $e_k \colon \mathbf{R} \to \mathbf{C}$ by $e_k(x) = e^{2\pi ikx}$, where of course $i = \sqrt{-1}$. These functions e_k satisfy the following orthogonality relations:

$$\int_0^1 \overline{e_k(x)} e_\ell(x)\, dx = \begin{cases} 0 & \text{if } k \neq \ell, \\ 1 & \text{if } k = \ell, \end{cases} \tag{1}$$

where $\overline{e_k(x)}$ denotes the complex conjugate of $e_k(x)$. Consider a function $f \colon \mathbf{R} \to \mathbf{C}$ of the form

$$f(x) = \sum_{k \in \mathbf{Z}} c_k e_k(x), \tag{2}$$

where the coefficients c_k are complex numbers. Suppose these coefficients are such that the series (2) converges in some suitable sense.[1] Then, as Euler (1777) observed, it follows from the orthogonality[2] relations (1) that for each $k \in \mathbf{Z}$, we have

$$c_k = \hat{f}(k), \qquad \text{where } \hat{f}(k) = \int_0^1 e^{-2\pi ikx} f(x)\, dx. \tag{3}$$

Thus the coefficients c_k are uniquely determined[3] by the function f. Conversely, it is natural to ask which functions f can be represented in the form (2). Of course f should be 1-periodic,[4] since each e_k is. In the years between 1807 and 1822, Fourier, in his studies of heat conduction, was led to consider series of the form (2) and he made the claim that any 1-periodic function f could be represented in the form of such a series with the coefficients c_k defined by (3). For the next century or so, much of the development of analysis, and especially of the theory of integration, was motivated by efforts to make sense of Fourier's astonishing claim.

Development of the Theory of Integration.

In the years between 1814 and 1823, Augustin Cauchy came to realize the importance of giving a mathematical definition to the notion of an integral.[5] Cauchy used suitable limits of sums to define the integral of a continuous function. Bernhard Riemann (1854) took the next natural step and defined the integral as the limit of such sums whenever the limit existed, whether or not the function was continuous. However, Riemann's definition was still too restricted. Henri Lebesgue (1902) finally discovered the ideal formulation of the theory of integration. While Lebesgue's theory is powerful, it is also difficult. Lebesgue himself, when asked which theory of integration a student should study first, answered "Riemann's, of course." We shall assume that the reader is already familiar with the Riemann integral.

---

[1] Uniform convergence would suffice though, as we shall see in due time, much less than this can still be good enough.

[2] For Euler, the equations (1) were just convenient algebraic relations. He did not call them orthogonality relations. The geometrical interpretation of such equations as expressing the orthogonality of the functions e_k, $k \in \mathbf{Z}$, was popularized by Erhard Schmidt, a student of David Hilbert, beginning around 1905, and can be found in papers published by Schmidt in 1907 and 1908, by F. Riesz in 1906 and 1907, and by Ernst Fischer in 1907, for instance. Schmidt (1908) thanks a certain Kowalewski, presumably Gerhard Kowalewski of Bonn, where Schmidt was at the time, for this idea. See page 46 in Michael Bernkopf, The development of function spaces with particular reference to their origins in integral equation theory, *Arch. Hist. Exact Sci.* 3 (1966), 1-96. See also page 352 in Thomas Hawkins, *Emergence of the theory of Lie groups: an essay in the history of mathematics, 1869-1926*, Springer, 2000.

[3] Earlier, Clairaut (1757) had shown this by a more involved argument.

[4] To say that f *is 1-periodic* means that for each $x \in \mathbf{R}$, we have $f(x+1) = f(x)$. More generally, if $T \in (0, \infty)$, then to say that f *is T-periodic* means that for each $x \in \mathbf{R}$, we have $f(x+T) = f(x)$.

[5] By the way, it was in 1818 that Cauchy became aware of Fourier's work.

Fourier Series from the Perspective of the Riemann Integral.

Although Lebesgue's theory of integration is essential for a truly satisfactory treatment of Fourier series, it is valuable to study Fourier series first using just Riemann's theory of integration. By doing so, we shall develop a better appreciation for the ways in which Lebesgues's theory of integration is an improvement over Riemann's.

A Remark on 1-Periodic Riemann-Integrable Functions.

Let $f\colon \mathbf{R} \to \mathbf{C}$ be 1-periodic. Let $u, v \in \mathbf{R}$. Suppose f is Riemann-integrable over $[u, u+1]$. Then, since f is 1-periodic, f is also Riemann-integrable over $[v, v+1]$ and

$$\int_u^{u+1} f(x)\,dx = \int_v^{v+1} f(x)\,dx.$$

Let's see why this is so. First, it is obvious by 1-periodicity that for all $a, b \in \mathbf{R}$ with $a \le b$ and for each $n \in \mathbf{Z}$, if f is Riemann-integrable over $[a, b]$, then f is Riemann-integrable over $[a+n, b+n]$ and $\int_a^b f(x)\,dx = \int_{a+n}^{b+n} f(x)\,dx$. Let n be an integer such that $u \le v-n \le u+1$. Note that $v \le u+1+n \le v+1$. Since f is Riemann-integrable over $[u, u+1]$, f is also Riemann-integrable over any subinterval of $[u, u+1]$. Thus f is Riemann-integrable over $[u, v-n]$ and also over $[v-n, u+1]$. Hence f is Riemann-integrable over $[u+1+n, v+1]$ and also over $[v, u+1+n]$ and we have

$$\int_u^{v-n} f(x)\,dx = \int_{u+1+n}^{v+1} f(x)\,dx \qquad \text{and} \qquad \int_{v-n}^{u+1} f(x)\,dx = \int_v^{u+1+n} f(x)\,dx.$$

Adding, we find that f is Riemann-integrable over $[v, v+1]$ and $\int_u^{u+1} f(x)\,dx = \int_v^{v+1} f(x)\,dx$, as desired. It follows in particular that for each $a \in \mathbf{R}$, f is Riemann-integrable over $[a, a+1]$ iff f is Riemann-integrable over $[0, 1]$ and that in this case,

$$\int_a^{a+1} f(x)\,dx = \int_0^1 f(x)\,dx.$$

Two Spaces of 1-Periodic Functions.

We shall write $C(\mathbf{T})$ for the vector space of continuous, complex-valued, 1-periodic functions on $\mathbf{R}$. Also, we shall write $\mathscr{R}(\mathbf{T})$ for the vector space of complex-valued, 1-periodic functions f on $\mathbf{R}$ such that f is Riemann-integrable over $[0, 1]$. Note that $C(\mathbf{T})$ is a proper linear subspace of $\mathscr{R}(\mathbf{T})$.

Some Notation Concerning Fourier Series.

Let the functions e_k, $k \in \mathbf{Z}$, be as defined in our preliminary remarks about Fourier series. For each $f \in \mathscr{R}(\mathbf{T})$, we define $\hat{f}\colon \mathbf{Z} \to \mathbf{C}$ by

$$\hat{f}(k) = \int_0^1 e^{-2\pi i k x} f(x)\,dx$$

for all $k \in \mathbf{Z}$. For each integer k, the number $\hat{f}(k)$, which was already mentioned in (3), is called *the k-th Fourier coefficient of f*. For each $f \in \mathscr{R}(\mathbf{T})$, *the Fourier series for f* is the series

$$\sum_{k=-\infty}^{\infty} \hat{f}(k) e_k.$$

In merely writing this series, we make no claims about whether or not it converges. We shall write $\boldsymbol{\omega}$ for the set $\{0, 1, 2, 3, \dots\}$ of nonnegative integers. For each $f \in \mathscr{R}(\mathbf{T})$ and each $n \in \boldsymbol{\omega}$, *the n-th symmetric partial sum of the Fourier series for f* is the function $S_n f$ defined on $\mathbf{R}$ by

$$S_n f = \sum_{k=-n}^{n} \hat{f}(k) e_k.$$

The Dirichlet-Jordan Theorem.

The first rigorous, reasonably general result on convergence of Fourier series was proved by Dirichlet (1829). He showed that if $f\colon \mathbf{R} \to \mathbf{C}$ is 1-periodic, piecewise continuous, and piecewise monotone, and if $x \in \mathbf{R}$, then as $n \to \infty$, we have $S_n f(x) \to f(x)$ if f is continuous at x, while $S_n f(x) \to \frac{1}{2}\big(f(x-) + f(x+)\big)$ otherwise. Jordan (1881) introduced functions of bounded variation and, observing that a function of bounded variation on an interval is a difference of two increasing functions on that interval, he extended Dirichlet's result to the case where $f \in \mathscr{R}(\mathbf{T})$ and f is of bounded variation in a neighborhood of x. Though Jordan's extension of Dirichlet's theorem was easy, it is quite significant, because a function of bounded variation need not be monotone on any nondegenerate interval and even a monotone function need not be continuous on any nondegenerate interval.

Convolution of Functions on a Finite Group.

Let $(G, \cdot)$ be a finite group. Let $f, g\colon G \to \mathbf{C}$. Then we may consider the corresponding formal linear combinations $\sum_{z \in G} f(z)z$ and $\sum_{y \in G} g(y)y$ of elements of G. Formally multiplying these two expressions and grouping like terms, we get

$$\left(\sum_{z \in G} f(z)z\right)\left(\sum_{y \in G} g(y)y\right) = \sum_{(z,y) \in G \times G} f(z)g(y)zy = \sum_{x \in G}\left(\sum_{zy = x} f(z)g(y)\right)x = \sum_{x \in G} h(x)x,$$

where

$$h(x) = (f * g)(x) = \sum_{y \in G} f(xy^{-1})g(y)$$

for all $x \in G$. The function $h = f * g$ is called *the convolution of f and g*. Thus convolution is how we would naturally multiply functions on G if we think of such functions as formal linear combinations of the group elements. In this sense, the operation of convolution is the natural "extension" of the group operation to the vector space of complex-valued functions on G. Since the group operation is associative, it is always the case that the operation of convolution of functions on G is associative. The operation of convolution of functions on G is commutative iff G is abelian.

Convolution of 1-Periodic Functions on R.

Consider the group $(\mathbf{R}, +)$. Let $f, g \in \mathscr{R}(\mathbf{T})$. Since f and g are 1-periodic, they may be thought of as functions on the quotient group $(\mathbf{R}/\mathbf{Z}, +)$. Since this is a continuous group, not a finite one, for all $f, g \in \mathscr{R}(\mathbf{T})$, we define *the convolution of f and g* not by a sum but by an integral, specifically

$$(f * g)(x) = \int_0^1 f(x-y)g(y)\,dy$$

for all $x \in \mathbf{R}$. For all $f, g \in \mathscr{R}(\mathbf{T})$, it is clear that $f * g$ is 1-periodic and, by approximating f and g by step functions, it is not too hard to show that $f * g$ is continuous (though we don't need this for now). It is not hard to check that for all $f, g \in \mathscr{R}(\mathbf{T})$, we have $f * g = g * f$ (and we shall use this soon). And of course, one can also check that for all $f, g, h \in \mathscr{R}(\mathbf{T})$, we have $f * (g * h) = (f * g) * h$ (though we don't need this for now either).

Translations of Functions on R.

Let $f\colon \mathbf{R} \to \mathbf{C}$ and let $a \in \mathbf{R}$. Then the *translation of f by a* is the function $\tau_a f$ defined on $\mathbf{R}$ by $\tau_a f(x) = f(x-a)$. Notice that if $E \subseteq \mathbf{R}$, then $\tau_a 1_E = 1_{E+a}$, where $E + a = \{\, x + a : x \in E \,\}$.

Remark. It is clear that if $f \in \mathscr{R}(\mathbf{T})$, then so is $\tau_a f$, for each $a \in \mathbf{R}$.

Exercise 1. Let $f, g \in \mathscr{R}(\mathbf{T})$ and let $a \in \mathbf{R}$. Prove that $(\tau_a f) * g = \tau_a(f * g) = f * (\tau_a g)$.

Remark. It may be enlightening to explain Exercise 1 in another way, even though we are not yet in a position to make rigorous sense of this other way. Let δ_a denote a (1-periodic) Dirac delta function centered at a. (If you have no idea what that means, feel free to skip this remark.) Then by the associativity of convolution, $(\delta_a * f) * g = \delta_a * (f * g)$. But $\delta_a * f = \tau_a f$ and $\delta_a * (f * g) = \tau_a(f * g)$, as one may easily check formally. This shows that $(\tau_a f) * g = \tau_a(f * g)$. Similarly, $(f * g) * \delta_a = f * (g * \delta_a)$. But since $(\mathbf{R}/\mathbf{Z}, +)$ is an abelian group, $(f * g) * \delta_a = \delta_a * (f * g) = \tau_a(f * g)$ and $g * \delta_a = \delta_a * g = \tau_a g$. Hence $\tau_a(f * g) = f * (\tau_a g)$. (In a non-abelian group, we should distinguish between left translation L_a and right translation R_a. Then $(L_a f) * g = (\delta_a f) * g = \delta_a * (f * g) = L_a(f * g)$ and $f * (R_a g) = f * (g * \delta_a) = (f * g) * \delta_a = R_a(f * g)$.) As we said, this explanation must be considered nonrigorous for now because there is actually no such function as δ_a. Eventually, though, we shall be able to give this explanation a rigorous interpretation.

The Dirichlet Kernel.

For each $n \in \boldsymbol{\omega}$, *the Dirichlet kernel of order* n is the 1-periodic function D_n defined on $\mathbf{R}$ by

$$D_n = \sum_{k=-n}^{n} e_k. \tag{4}$$

The observation in the next exercise is one of the ingredients in the proof of the Dirichlet-Jordan theorem.

Exercise 2. Let $f \in \mathscr{R}(\mathbf{T})$.

- **(a)** Verify that for each $k \in \mathbf{Z}$, we have $e_k * f = \hat{f}(k) e_k$.
- **(b)** Deduce that for each $n \in \boldsymbol{\omega}$, we have $S_n f = D_n * f$.

Remark. It follows from Exercise 2(b) and Exercise 1 that $S_n(\tau_a f) = \tau_a(S_n f)$ for all $a \in \mathbf{R}$, all $n \in \boldsymbol{\omega}$, and all $f \in \mathscr{R}(\mathbf{T})$.

Remark. For each $x \in \mathbf{R}$, we have $e_0(x) = 1$. Hence $\int_{-1/2}^{1/2} e_0(x)\,dx = 1$. For each nonzero integer k, we have $\int_{-1/2}^{1/2} e_k(x)\,dx = 0$. It follows that for each $n \in \boldsymbol{\omega}$,

$$\int_{-1/2}^{1/2} D_n(x)\,dx = 1. \tag{5}$$

Remark. From (5) and from Exercise 2(b), it follows that for each $f \in \mathscr{R}(\mathbf{T})$, for each $n \in \boldsymbol{\omega}$, and for each constant function $C \colon \mathbf{R} \to \mathbf{C}$, we have $S_n(f + C) = (S_n f) + C$.

A Closed-Form Expression for the Dirichlet Kernel.

Let $n \in \boldsymbol{\omega}$. For each $k \in \mathbf{Z}$ and for each $x \in \mathbf{Z}$, we have $e_k(x) = 1$. From this, it is clear that at each integer point, D_n takes the value $2n + 1$. Thus to find a closed-form expression for D_n, the main thing is to investigate the value of D_n at noninteger points.

Exercise 3. Let $n \in \boldsymbol{\omega}$ and let $x \in \mathbf{R} \setminus \mathbf{Z}$. Prove that

$$D_n(x) = \frac{\sin(2n+1)\pi x}{\sin \pi x}.$$

Suggestion: To save writing, let $w = e^{\pi i x}$. Note that $D_n(x) = \sum_{k=-n}^{n} (w^2)^k$. Verify that

$$(w - w^{-1}) D_n(x) = w^{2n+1} - w^{-(2n+1)}. \tag{6}$$

Explain why $w - w^{-1} \neq 0$.

Warning: Fourier Series Can Behave Badly.

As we shall see eventually, there exist functions $f \in C(\mathbf{T})$ such that for certain points $x \in \mathbf{R}$, the sequence $(S_n f(x))$ does not converge to $f(x)$. It is even possible for the set of such bad points x to be dense in $\mathbf{R}$. The fact that this can happen is related to the fact that the graph of a general continuous function can "wiggle" a lot.

In fact, we shall see eventually that for "most" functions $f \in C(\mathbf{T})$, for "most" points $x \in \mathbf{R}$, the sequence $(S_n f(x))$ is unbounded. Here "most" means "for all but a meager set of." A meager set is a set of the first category, in Baire's terminology. This notion will be discussed in detail later.

Fejér's Theorem and Some of Its Consequences.

The preceding warning sounds like very bad news for convergence of Fourier series of continuous functions. Happily, there is something positive we can say. Recall that a sequence of numbers (s_n) is said to be Cesàro-convergent to a number s if the average of the first n terms of the sequence converges to s in the ordinary sense as n tends to infinity. Recall also that while ordinary convergence implies Cesàro convergence, a sequence can be Cesàro convergent without being convergent in the ordinary sense.[6]

Fejér (1904) showed that for each $f \in C(\mathbf{T})$, the partial sums $(S_n f)$ of the Fourier series for f are uniformly Cesàro-convergent to f. Let us formulate this precisely. For each $f \in \mathscr{R}(\mathbf{T})$, for each $n \in \boldsymbol{\omega}$, *the n-th Fejér sum for f* is the function $\sigma_n f$ defined on $\mathbf{R}$ by

$$\sigma_n f = \frac{1}{n+1} \sum_{m=0}^{n} S_m f,$$

so that for each $x \in \mathbf{R}$, $\sigma_n f(x)$ is the average of the $n+1$ numbers $S_0 f(x), \ldots, S_n f(x)$. What Fejér proved is that for each $f \in C(\mathbf{T})$, we have $\sigma_n f \to f$ uniformly on $\mathbf{R}$ as $n \to \infty$.

In fact, Fejér showed more. For instance, let $f \in \mathscr{R}(\mathbf{T})$ and let Γ be the set of all $x \in \mathbf{R}$ such that f is continuous at x. Let K be a compact subset of Γ. Then $\sigma_n f \to f$ uniformly on K. And for instance, let $f \in \mathscr{R}(\mathbf{T})$ and let $x \in \mathbf{R}$ such that the one-sided limits $f(x-)$ and $f(x+)$ both exist in $\mathbf{C}$. Then $\sigma_n f(x) \to \frac{1}{2}\big(f(x-) + f(x+)\big)$.

Exercise 4. Here are some simple but important consequences of Fejér's theorem.

- **(a)** Let $f, g \in C(\mathbf{T})$. Suppose $\hat{f} = \hat{g}$. Prove that $f = g$.
- **(b)** Let $f \in \mathscr{R}(\mathbf{T})$ and let $x \in \mathbf{R}$ such that f is continuous at x. Suppose $a \in \mathbf{C}$ and $S_n f(x) \to a$. Prove that $a = f(x)$.
- **(c)** Let $f \in \mathscr{R}(\mathbf{T})$ and let $x \in \mathbf{R}$ such that the one-sided limits $f(x-)$ and $f(x+)$ both exist in $\mathbf{C}$. Suppose $a \in \mathbf{C}$ and $S_n f(x) \to a$. Then $a = \frac{1}{2}\big(f(x-) + f(x+)\big)$.

The Fejér Kernel.

For each $n \in \boldsymbol{\omega}$, *the Fejér kernel of order n* is the 1-periodic function K_n defined on $\mathbf{R}$ by

$$K_n = \frac{1}{n+1} \sum_{m=0}^{n} D_m. \tag{7}$$

Obviously, for each $n \in \boldsymbol{\omega}$ and for each $x \in \mathbf{R}$, $K_n(x)$ is the average of the $n+1$ numbers $D_0(x), \ldots, D_n(x)$. The observation in the next exercise is one of the ingredients in the proof of Fejér's theorem.

Exercise 5. Let $f \in \mathscr{R}(\mathbf{T})$ and let $n \in \boldsymbol{\omega}$. Verify that $\sigma_n f = K_n * f$.

Remark. It follows from Exercise 5 and Exercise 1 that $\sigma_n(\tau_a f) = \tau_a(\sigma_n f)$ for all $a \in \mathbf{R}$, all $n \in \boldsymbol{\omega}$, and all $f \in \mathscr{R}(\mathbf{T})$.

Remark. From (5) and (7), it is clear that for each $n \in \boldsymbol{\omega}$,

$$\int_{-1/2}^{1/2} K_n(x)\, dx = 1. \tag{8}$$

Remark. From (8) and from Exercise 5, we see that for each $f \in \mathscr{R}(\mathbf{T})$, for each $n \in \boldsymbol{\omega}$, and for each constant function $C\colon \mathbf{R} \to \mathbf{C}$, we have $\sigma_n(f + C) = (\sigma_n f) + C$

[6] For instance, $\frac{1}{n+1} \sum_{m=0}^{n} (-1)^m \to 0$ as $n \to \infty$.

A Closed-Form Expression for the Fejér Kernel.

Let $n \in \boldsymbol{\omega}$. We know for each $m \in \mathbf{Z}$, for each $x \in \mathbf{Z}$, $D_m(x) = 2m+1$. From this, it is easy to see that for each $x \in \mathbf{Z}$, $K_n(x) = n+1$. Thus to find a closed-form expression for K_n, the main thing is to investigate the value of K_n at noninteger points.

Exercise 6. Let $n \in \boldsymbol{\omega}$ and let $x \in \mathbf{R} \setminus \mathbf{Z}$. Prove that

$$K_n(x) = \frac{\sin^2(n+1)\pi x}{(n+1)\sin^2 \pi x}. \tag{9}$$

Suggestion: Let $w = e^{\pi i x}$, as in Exercise 3. Begin by using (6) to show that

$$(n+1)(w - w^{-1})K_n(x) = (w + w^3 + \cdots + w^{2n+1}) - (w^{-1} + w^{-3} + \cdots + w^{-(2n+1)}).$$

Then multiply both sides of this equation by $w - w^{-1}$ and simplify.

Two Nice Properties of the Fejér Kernel.

It follows from Exercise 6 that $K_n \geq 0$ for each $n \in \boldsymbol{\omega}$ and that $K_n \to 0$ uniformly on compact subsets of $\mathbf{R} \setminus \mathbf{Z}$ as $n \to \infty$, These two properties are crucial for the proof of Fejér's theorem. Neither of these properties is obvious from the definition of K_n and neither of them is shared by D_n.

Another Way to Look at the Fejér Kernel and at Fejér's Theorem.

From (4), we see that e_0 occurs in D_m for each m, e_1 and e_{-1} occur in D_m for each $m \geq 1$, e_2 and e_{-2} occur in D_m for each $m \geq 2$, and so on. From this we see that for each $n \in \boldsymbol{\omega}$,

$$\sum_{m=0}^{n} D_m = \sum_{k=-n}^{n} \big((n+1) - |k|\big)e_k, \tag{10}$$

so

$$K_n = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) e_k,$$

so for each $f \in \mathscr{R}(\mathbf{T})$,

$$\sigma_n f = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n+1}\right) \hat{f}(k) e_k,$$

so $\sigma_n f$ is like $S_n f$ except that it is "regularized" by making the terms with larger values of $|k|$ less heavily weighted, while for each fixed k, these weights tend to 1 as $n \to \infty$. Fejér's theorem tells us that for each $f \in C(\mathbf{T})$, this "regularized" version of $S_n f$ converges uniformly to f.

Yet Another Way to Look at the Fejér Kernel.

Let $n \in \boldsymbol{\omega}$. Notice that $\left|\sum_{j=0}^{n} e_j\right|^2 = \overline{\left(\sum_{j=0}^{n} e_j\right)}\left(\sum_{j'=0}^{n} e_{j'}\right) = \left(\sum_{j=0}^{n} e_{-j}\right)\left(\sum_{j'=0}^{n} e_{j'}\right) = \sum_{j,j'=0}^{n} e_{j'-j} = \sum_{k=-n}^{n}\big((n+1) - |k|\big)e_k = \sum_{m=0}^{n} D_m$, where we have used (10) in the last step. Thus $K_n = \frac{1}{n+1}\left|\sum_{j=0}^{n} e_j\right|^2$, which already shows that $K_n \geq 0$, and which can also be used to obtain (9), because for each $x \in \mathbf{R} \setminus \mathbf{Z}$, $\sum_{j=0}^{n} e_j(x)$ is easy to evaluate, since it is the sum of a finite geometric progression.

Trigonometric Polynomials.

A *trigonometric polynomial* is a function on $\mathbf{R}$ which is a linear combination of finitely many of the functions e_k. For each $f \in C(\mathbf{T})$ and each $n \in \boldsymbol{\omega}$, it is clear that the functions $S_n f$ and $\sigma_n f$ are trigonometric polynomials. Thus a particular consequence of Fejér's theorem is that each function $f \in C(\mathbf{T})$ is the limit of a uniformly convergent sequence of trigonometric polynomials.

Theorem (Leopold Fejér, 1904)**.**

(a) *Let* $f \in C(\mathbf{T})$. *Then as* $n \to \infty$, $\sigma_n f \to f$ *uniformly on* $\mathbf{R}$.

(b) *Let* $f \in \mathscr{R}(\mathbf{T})$, *let* C *be the set of all* $x \in \mathbf{R}$ *such* f *is continuous at* x, *and let* H *be a compact subset of* C. *Then as* $n \to \infty$, $\sigma_n f \to f$ *uniformly on* H.

(c) *Let* $x \in \mathbf{R}$ *such that the one-sided limits* $f(x-)$ *and* $f(x+)$ *exist in* $\mathbf{C}$. *Then as* $n \to \infty$, $\sigma_n f(x) \to \frac{1}{2}\big(f(x-) + f(x+)\big)$.

Proof. By periodicity, to prove (a), it suffices to show that $\sigma_n f \to f$ uniformly on $[0,1]$. Thus (a) follows from (b) with $H = [0,1]$. Before proving (b), we make some observations. For each $n \in \boldsymbol{\omega}$, $\sigma_n f = f * K_n$, where K_n is the Fejér kernel of order n. Now $K_n \in C(\mathbf{T})$, $\int_{-1/2}^{1/2} K_n(x)\,dx = 1$, $K_n \geq 0$, and for each $\delta \in (0, 1/2)$, letting $E(\delta) = [-1/2, -\delta] \cup [\delta, 1/2]$, we have $K_n \to 0$ uniformly on $E(\delta)$ as $n \to \infty$. Consider any $x \in \mathbf{R}$, any $\delta \in (0,1/2)$, and any $n \in \boldsymbol{\omega}$. Then $f(x) = \int_{-1/2}^{1/2} f(x) K_n(y)\,dy$ and $(f * K_n)(x) = \int_{-1/2}^{1/2} f(x-y) K_n(y)\,dy$, so

$$\big|f(x) - (f * K_n)(x)\big| \leq \int_{-1/2}^{1/2} |f(x) - f(x-y)| K_n(y)\,dy = I_1(x,\delta,n) + I_2(x,\delta,n),$$

where

$$I_1(x,\delta,n) = \int_{|y|<\delta} |f(x) - f(x-y)| K_n(y)\,dy \qquad \text{and} \qquad I_2(x,\delta,n) = \int_{E(\delta)} |f(x) - f(x-y)| K_n(y)\,dy.$$

Now let us prove (b). Let $\varepsilon > 0$. Since f is continuous at each point of the compact set H, f is continuous at x uniformly for $x \in H$. Hence there exists $\delta \in (0, 1/2)$ such that for each $x \in H$, for each $t \in \mathbf{R}$, if $|x - t| < \delta$, then $|f(x) - f(t)| < \frac{\varepsilon}{2}$. Hence for each $x \in H$ and for each $n \in \boldsymbol{\omega}$, we have

$$I_1(x,\delta,n) \leq \frac{\varepsilon}{2} \int_{|y|<\delta} K_n(y)\,dy \leq \frac{\varepsilon}{2}.$$

Now let us consider $I_2(x,\delta,n)$. Since f is 1-periodic and locally Riemann-integrable, f is bounded on $\mathbf{R}$. However, in anticipation of our study of the Lebesgue integral, we wish to arrange our proof so that it would also apply if f were 1-periodic and locally Lebesgue-integrable. Accordingly, let us not rely on the fact that f is bounded on $\mathbf{R}$. It is nevertheless true that f is bounded on H, since f is continuous at each point of the compact set H. Let $M = \sup_H |f|$. Then $M < \infty$. Since $K_n \to 0$ uniformly on $E(\delta)$, there exists $N \in \boldsymbol{\omega}$ such that for each $n \geq N$, for each $y \in E(\delta)$, we have

$$K_n(y) \leq \frac{\varepsilon}{1 + 2M + 2\|f\|_1},$$

where $\|f\|_1 = \int_0^1 |f(t)|\,dt$. Then for each $n \geq N$ and for each $x \in H$, we have

$$I_2(x,\delta,n) \leq \int_{E(\delta)} \big(M + |f(x-y)|\big) K_n(y)\,dy \leq \big(M + \|f\|_1\big) \frac{\varepsilon}{1 + 2M + 2\|f\|_1} < \frac{\varepsilon}{2},$$

so $\big|f(x) - (f * K_n)(x)\big| \leq I_1(x,\delta,n) + I_2(x,\delta,n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. This completes the proof of (b). As for (c), by an argument similar to but a bit simpler than the one we just gave for (b), if x is as in (c), then as $n \to \infty$, we have

$$\int_{-1/2}^{0} f(x-y) K_n(y)\,dy \to \frac{f(x+)}{2} \qquad \text{and} \qquad \int_{0}^{1/2} f(x-y) K_n(y)\,dy \to \frac{f(x-)}{2},$$

so $(f * K_n)(x) \to \frac{1}{2}\big(f(x-) + f(x+)\big)$, as desired. ∎

Remark. In the proof of (b), we were careful to avoid using the fact that if $f \in \mathscr{R}(\mathbf{T})$, then f is bounded. Hence, once we have the Lebesgue integral, the argument we gave for (b) can be used to establish the same conclusion under the weaker assumption that $f \in L^1(\mathbf{T})$.

Biographical Note. Fejér (1880–1959) obtained his doctorate under the direction of Hermann Schwarz. Fejér's doctoral students included Paul Erdős, John von Neumann, Pál Turán, George Pólya, Tibor Radó, Marcel Riesz, and Gábor Szegő. According to the Mathematics Genealogy Project, he has 9914 mathematical descendants (as of 13 June 2024).

Let $E = \mathbf{C}$, or more generally, let E be a normed linear space over $\mathbf{C}$. Let (u_k) be a sequence in E. For each $n \in \mathbf{N}$, let

$$S_n = \sum_{k=1}^{n} u_k$$

and let

$$\sigma_n = \frac{1}{n} \sum_{\ell=1}^{n} S_\ell.$$

Notice that σ_n is the average of $S_1, S_2, \ldots, S_n$. Let $L \in E$. To say that $\sum_{k=1}^{\infty} u_k = L$ means that $S_n \to L$. To say that (u_k) *is Cesàro summable to* L means that $\sigma_n \to L$. To say that (u_k) *is Abel summable to* L means that $\sum_{k=1}^{\infty} t^k u_k$ converges for $0 < t < 1$ and that this sum tends to L as $t \uparrow 1$. It is well known that ordinary convergence of a sequence implies Cesàro convergence of that sequence to the same limit. Applying this to the sequence (S_n), we see that if $\sum_{k=1}^{\infty} u_k = L$, then (u_k) is Cesàro summable to L. Abel[1] showed that if $\sum_{k=1}^{\infty} u_k = L$, then (u_k) is what we have just defined as Abel summable to L. More generally, it can be shown[2] that if (u_k) is Cesàro summable to L, then (u_k) is Abel summable to L. In summary,

$$\sum_{k=1}^{\infty} u_k = L \qquad \Rightarrow \qquad (u_k) \text{ is Cesàro summable to } L \qquad \Rightarrow \qquad (u_k) \text{ is Abel summable to } L.$$

Alfred Tauber[3] proved a partial converse to Abel's theorem, namely that if (u_k) is Abel summable to L, and if in addition $u_k = o(1/k)$, then $\sum_{k=1}^{\infty} u_k = L$. In particular, if (u_k) is Cesàro summable to L, and if in addition $u_k = o(1/k)$, then $\sum_{k=1}^{\infty} u_k = L$. This special case of Tauber's theorem is very easy to prove. Just notice that

$$\sigma_n = \frac{1}{n} \sum_{\ell=1}^{n} \sum_{k=1}^{\ell} u_k = \frac{1}{n} \sum_{k=1}^{n} \sum_{\ell=k}^{n} u_k = \frac{1}{n} \sum_{k=1}^{n} (n-k+1) u_k = \frac{n+1}{n} S_n - \frac{1}{n} \sum_{k=1}^{n} k u_k$$

and recall that to say $u_k = o(1/k)$ means $k u_k \to 0$, so since ordinary convergence implies Cesàro convergence,

$$\frac{1}{n} \sum_{k=1}^{n} k u_k \to 0.$$

G. H. Hardy[4] gave the general name *Tauberian theorems* to such converse results and showed that if (u_k) is Cesàro summable to L, and if in addition $u_k = O(1/k)$, then $\sum_{k=1}^{\infty} u_k = L$. Thus Hardy improved a special case of Tauber's theorem. J. E. Littlewood[5] improved Tauber's theorem itself in the way pointed to by Hardy's result. In other words, he showed that if (u_k) is Abel summable to L, and if in addition $u_k = O(1/k)$, then $\sum_{k=1}^{\infty} u_k = L$. Littlewood's proof of this was quite difficult. Almost twenty years later, Karamata[6] found a much simpler proof. But even Karamata's proof is more involved than we wish to present here. Instead we shall give a proof of Hardy's result. So suppose that $\sigma_n \to L$ and that $u_k = O(1/k)$. Recall that to say $u_k = O(1/k)$ means there exists $M \in (0, \infty)$ such that for each $k \in \mathbf{N}$, we have $\|u_k\| \le M/k$. We wish to show that $S_n \to L$. Since $S_n = (S_n - \sigma_n) + \sigma_n$, it suffices to show that $S_n - \sigma_n \to 0$. Our strategy will be to express $S_n - \sigma_n$ in terms of $\sigma_n - \sigma_m$ and the average of the quantities $S_n - S_k$, $k = m+1, \ldots, n$, where m is just a little smaller than n. Consider any natural numbers m and n with $m < n$. We have

$$\sum_{k=m+1}^{n} S_k = \sum_{k=1}^{n} S_k - \sum_{k=1}^{m} S_k = n\sigma_n - m\sigma_m,$$

[1] in his famous paper on the binomial series, *Journal für die reine und angewandte Mathematik* 1 (1826), 311–339.

[2] See pages 484 and 489 in Konrad Knopp, *Theory and Application of Infinite Series*, Blackie & Son, 1952, reprinted by Dover Publications in 1990.

[3] *Monatshefte für Mathematik* 8 (1897), 273–277. See also Knopp, *loc. cit.*, page 500.

[4] *Proceedings of the London Mathematical Society* (2) 8 (1910), 301–320.

[5] *Proceedings of the London Mathematical Society* (2) 9 (1911), 434–448.

[6] *Mathematische Zeitschrift* 32 (1930), 319–320. See also Knopp, *loc. cit.*, pp. 501–505.

so

$$\sum_{k=m+1}^{n}(S_n - S_k) = (n-m)S_n - (n\sigma_n - m\sigma_m) = (n-m)(S_n - \sigma_n) - m(\sigma_n - \sigma_m),$$

so

$$S_n - \sigma_n = \frac{m}{n-m}(\sigma_n - \sigma_m) + \frac{1}{n-m}\sum_{k=m+1}^{n}(S_n - S_k). \tag{1}$$

Now for $k = 1, \ldots, n$, we have

$$\|S_n - S_k\| \leq \sum_{\ell=k+1}^{n}\|u_\ell\| \leq \sum_{\ell=k+1}^{n}\frac{M}{\ell} \leq \frac{(n-k)M}{k+1}.$$

Hence for $k = m+1, \ldots, n$, we have $\|S_n - S_k\| \leq \dfrac{(n-m-1)M}{m+2}$. Thus

$$\left\|\frac{1}{n-m}\sum_{k=m+1}^{n}(S_n - S_k)\right\| \leq \frac{1}{n-m}(n-m)\frac{(n-m-1)M}{m+2} = \frac{(n-m-1)M}{m+2}.$$

Therefore by (1),

$$\|S_n - \sigma_n\| \leq \frac{m}{n-m}\|\sigma_n - \sigma_m\| + \frac{(n-m-1)M}{m+2}.$$

This holds for all $m, n \in \mathbf{N}$ with $m < n$. Consider any $\varepsilon > 0$. Let $n \in \mathbf{N}$ with $n \geq 1 + \varepsilon$. Then

$$\frac{n}{1+\varepsilon} \geq 1.$$

Let $m = \left\lfloor \dfrac{n}{1+\varepsilon} \right\rfloor$, the greatest integer less than or equal to $\dfrac{n}{1+\varepsilon}$. Then

$$1 \leq m \leq \frac{n}{1+\varepsilon} < m+1.$$

Since $m \leq \dfrac{n}{1+\varepsilon}$, we have $(1+\varepsilon)m \leq n$, so $\varepsilon m \leq n - m$, so

$$\frac{m}{n-m} \leq \frac{1}{\varepsilon}.$$

Also, since $m + 1 > \dfrac{n}{1+\varepsilon}$, we have

$$\frac{n-m-1}{m+2} < \frac{n - \frac{n}{1+\varepsilon}}{\frac{n}{1+\varepsilon}+1} = \frac{(1+\varepsilon)n - n}{n+1+\varepsilon} = \frac{\varepsilon n}{n+1+\varepsilon} < \frac{\varepsilon n}{n} = \varepsilon.$$

Therefore

$$\|S_n - \sigma_n\| < \frac{1}{\varepsilon}\|\sigma_n - \sigma_m\| + M\varepsilon.$$

This holds for all natural numbers $n \geq 1 + \varepsilon$ and for $m = \lfloor n/(1+\varepsilon) \rfloor$. Now as $n \to \infty$, we have $m \to \infty$ too, so $\|\sigma_n - \sigma_m\| \to 0$. Thus there exists $N \geq 1 + \varepsilon$ such that for all $n > N$, we have $\|\sigma_n - \sigma_m\| < \varepsilon^2$. Then for all $n > N$, we have $\|S_n - \sigma_n\| < (1+M)\varepsilon$. Since M does not depend on ε, this shows that

$$S_n - \sigma_n \to 0,$$

as desired.

Exercise. Let the notation be as above but do not suppose that (σ_n) converges. Instead, just suppose that (σ_n) is bounded. Prove that (S_n) is bounded too. (Hint: This follows easily from part of the proof above.)

Proposition. *Let* $f\colon \mathbf{R} \to \mathbf{C}$ *be* 1*-periodic. Suppose* f *is of bounded variation on* $[0,1]$*. Let* V *be the variation of* f *on* $[0,1]$*. Let* $k \in \mathbf{Z} \setminus \{0\}$*. Then*

$$|\hat{f}(k)| \le \frac{V}{|k|}.$$

Proof. For $0 \le a < b \le 1$, let $V(f,[a,b])$ be the variation of f on $[a,b]$. Recall that for $0 \le a < b < c \le 1$, we have

$$V(f,[a,c]) = V(f,[a,b]) + V(f,[b,c]). \tag{1}$$

Let $\kappa = |k|$. For $j = 0, 1, \ldots, \kappa$, let

$$x_j = \frac{j}{\kappa}.$$

Then $0 = x_0 < x_1 < x_2 < \cdots < x_\kappa = 1$ and for $j = 1, \ldots, \kappa$, we have

$$x_j - x_{j-1} = \frac{1}{\kappa}.$$

For $j = 1, \ldots, \kappa$, let $V_j = V(f,[x_{j-1},x_j])$. It follows from (1) that $V = \sum_{j=1}^{\kappa} V_j$. For $j = 1, \ldots, \kappa$, let $I_j = (x_{j-1}, x_j]$, let $g_j = 1_{I_j}$, and observe that $\hat{g}_j(k) = 0$ by inspection, because for each $x \in \mathbf{R}$,

$$\exp\left(-2\pi i k\left(x + \tfrac{1}{2\kappa}\right)\right) = \exp(-2\pi i k x)\exp\left(-\pi i \tfrac{k}{\kappa}\right) = -\exp(-2\pi i k x).$$

Let $g = \sum_{j=1}^{\kappa} f(x_{j-1}) g_j$. Then $\hat{g}(k) = 0$. Let $h = f - g$. Since $\hat{g}(k) = 0$, we have $\hat{f}(k) = \hat{h}(k)$. For $j = 1, \ldots, \kappa$, for $x_{j-1} < x \le x_j$, we have $|h(x)| = |f(x) - g(x)| = |f(x) - f(x_{j-1})| \le V_j$. Therefore

$$\|h\|_1 = \int_0^1 |h(x)|\,dx = \sum_{j=1}^{\kappa} \int_{x_{j-1}}^{x_j} |h(x)|\,dx \le \sum_{j=1}^{\kappa} \frac{V_j}{\kappa} = \frac{V}{\kappa},$$

so $|\hat{f}(k)| = |\hat{h}(k)| \le \|h\|_1 \le \dfrac{V}{\kappa} = \dfrac{V}{|k|}$. ■

Hardy's Tauberian Theorem.

Let (v_k) be a sequence in a normed linear space V. Let $S_m = \sum_{k=1}^m v_k$ for each $m \in \mathbf{N}$, let $S \in V$, and let $\sigma_n = \frac{1}{n}\sum_{m=1}^n S_m$ for each $n \in \mathbf{N}$. If $S_m \to S$ as $m \to \infty$, then as we know,[17] $\sigma_n \to S$ as $n \to \infty$. The converse is not true in general but it is very easy to prove that if $\sigma_n \to S$ as $n \to \infty$ and if in addition $v_k = o(1/k)$ as $k \to \infty$, then $S_m \to S$ as $m \to \infty$. This is a special case of a theorem of Tauber (1897). Such converse results have come to be known as *Tauberian theorems.* Hardy (1910)[18] proved the following strengthening of the result that we just stated: If $\sigma_n \to S$ as $n \to \infty$ and if in addition $v_k = O(1/k)$ as $k \to \infty$, then $S_m \to S$ as $m \to \infty$. This is called *Hardy's Tauberian theorem.*

Exercise 38. Let (v_k) be a sequence in a normed linear space V. Let $S_m = \sum_{k=1}^m v_k$ for each $m \in \mathbf{N}$, and let $\sigma_n = \frac{1}{n}\sum_{m=1}^n S_m$ for each $n \in \mathbf{N}$. Suppose (σ_n) is bounded in V and $v_k = O(1/k)$ as $k \to \infty$. Prove that (S_m) is bounded in V. (Hint: Adapt part of the proof of Hardy's Tauberian theorem.)

Convergence of Fourier Series for Functions of Bounded Variation.

Exercise 39.

(a) Let $f \in BV(\mathbf{T})$. Let V be the variation of f on $[0,1]$ and let $k \in \mathbf{Z} \setminus \{0\}$. Prove that

$$|\hat{f}(k)| \leq \frac{V}{|k|}.$$

(Hint: Let $\kappa = |k|$. For $j = 0, \ldots, \kappa$, let $x_j = j/\kappa$ and if $j \geq 1$, let $I_j = (x_{j-1}, x_j]$, let $g_j = 1_{I_j}$, and notice that $\hat{g}_j(k) = 0$. Let $g = \sum_{j=1}^{\kappa} f(x_{j-1}) g_j$. Then $\hat{g}(k) = 0$. Let $h = f - g$. Then $|\hat{f}(k)| = |\hat{h}(k)|$. Use (23) to show that $\int_0^1 |h(x)|\,dx \leq V/\kappa$. Remember that $|\hat{h}(k)| \leq \int_0^1 |h(x)|\,dx$.)

(b) (*The Dirichlet-Jordan Theorem.*) Let $f \in BV(\mathbf{T})$. Prove that for each $x \in \mathbf{R}$,

$$S_n f(x) \to \frac{f(x-) + f(x+)}{2}.$$

In particular, for each $x \in \mathbf{R}$, if f is continuous at x, then $S_n f(x) \to f(x)$. Also, let C be the set of all x in $\mathbf{R}$ such that f is continuous at x and let H be a compact subset of C. Prove that $S_n f \to f$ uniformly on H. (Hint: By Fejér's theorem, for each $x \in \mathbf{R}$, $\sigma_n f(x) \to \frac{1}{2}\big(f(x-) + f(x+)\big)$. By part (a), $\hat{f}(k) = O(1/|k|)$ as $k \to \pm\infty$. To obtain the first desired conclusion, apply Hardy's Tauberian theorem to the series $\sum_{k=0}^{\infty} g_k(x)$, where $g_0 = \hat{f}(0)e_0$ and $g_k = \hat{f}(k)e_k + \hat{f}(-k)e_{-k}$ for $k \geq 1$. Again by Fejér's theorem, $\sigma_n f \to f$ uniformly on H. To obtain the second desired conclusion, consider the space $C(H)$ equipped with the uniform norm and apply Hardy's Tauberian theorem in this space to the series $\sum_{k=0}^{\infty} h_k$, where h_k is the restriction of g_k to H.)

Warning. In Exercise 39(b), the series $\sum_{k=-\infty}^{\infty} \hat{f}(k)e_k$ may only be conditionally convergent and only the convergence of the symmetric partial sums $S_n f = \sum_{k=-n}^{n} \hat{f}(k)e_k$ is asserted. Exercise 39(b) does not tell us whether the asymmetric partial sums $S_{mn} f = \sum_{k=-m}^{n} \hat{f}(k)e_k$ converge as $m, n \to \infty$.

Remark. Hardy (1910) himself pointed out that his Tauberian theorem can be applied to obtain the Dirichlet-Jordan theorem as a simple corollary of Fejér's theorem, as in Exercise 39.

Exercise 40. Let $f \in BV(\mathbf{T})$. Prove that the sequence of functions $(S_n f)_{n \in \mathbf{N}}$ is uniformly bounded. (Hint: Use Exercise 38 together with Exercise 39(a).)

---

[17] Remember that ordinary convergence implies Cesàro convergence. The proof that this is so in a normed linear space is essentially the same as the proof that it holds for numbers.

[18] G. H. Hardy, Theorems relating to the summability and convergence of slowly oscillating series, *Proceedings of the London Mathematical Society* (2) 8 (1910), 301–320.