Chapter 6

Fourier analysis

(Historical intro: the heat equation on a square plate or interval.)

Fourier's analysis was tremendously successful in the 19th century for formulating series expansions for solutions of some very simple ODE and PDE. This class shows that in the 20th century, Fourier analysis has established itself as a central tool for numerical computations as well, for vastly more general ODE and PDE when explicit formulas are not available.

6.1 The Fourier transform

We will take the Fourier transform of integrable functions of one variable $x \in \mathbb{R}$.

Definition 13. *(Integrability) A function f is called integrable, or absolutely integrable, when*

$$\int_{-\infty}^{\infty} |f(x)| \, dx < \infty,$$

in the sense of Lebesgue integration. One also writes $f \in L^1(\mathbb{R})$ for the space of integrable functions.

We denote the physical variable as x, but it is sometimes denoted by t in contexts in which its role is time, and one wants to emphasize that. The frequency, or wavenumber variable is denoted k. Popular alternatives choices for the frequency variable are ω (engineers) or ξ (mathematicians), or p (physicists).

Definition 14. *The Fourier transform (FT) of an integrable function $f(x)$ is defined as*

$$\hat{f}(k) = \int_{-\infty}^{\infty} e^{-ikx} f(x)\, dx. \tag{6.1}$$

When $\hat{f}(k)$ is also integrable, $f(x)$ can be recovered from $\hat{f}(k)$ by means of the inverse Fourier transform (IFT)

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} \hat{f}(k)\, dk. \tag{6.2}$$

Intuitively, $\hat{f}(k)$ is the amplitude density of f at frequency k. The formula for recovering f is a decomposition of f into constituent waves.

The justification of the inverse FT formula belongs in a real analysis class (where it is linked to the notion of approximate identity.) We will justify the form of (6.2) heuristically when we see Fourier series in the next section.

The precaution of assuming integrability is so that the integrals can be understood in the usual Lebesgue sense. In that context, taking integrals over infinite intervals is perfectly fine. If (6.1) and (6.2) are understood as limits of integrals over finite intervals, it does not matter how the bounds are chosen to tend to $\pm\infty$.

One may in fact understand the formulas for the FT and IFT for much larger function classes than the integrable functions, namely distributions, but this is also beyond the scope of the class. We will generally not overly worry about these issues. It is good to know where to draw the line: the basic case is that of integrable functions, and anything beyond that requires care and adequate generalizations.

Do not be surprised to see alternative formulas for the Fourier transfom in other classes or other contexts. Wikipedia lists them.

Here are some important properties of Fourier transforms:

- (Differentiation)

  $$\widehat{f'}(k) = ik\hat{f}(k).$$

  Justification: integration by parts in the integral for the FT.

- (Translation) If $g(x) = f(x+a)$, then

  $$\hat{g}(k) = e^{ika}\hat{f}(k).$$

  Justification: change of variables in the integral for the FT.

Let's see some examples of FT.

Example 17. *Let*

$$f(x) = \frac{1}{2a}\chi_{[-a,a]}(x) = \begin{cases} \frac{1}{2a} & \text{if } x \in [-a,a]; \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\hat{f}(k) = \frac{1}{2a}\int_{-a}^{a} e^{-ikx}dx = \frac{\sin(ka)}{ka}.$$

This function is a scaled version of the sinc function,

$$sinc(k) = \frac{\sin k}{k}.$$

It is easy to check by L'Hospital's rule that

$$sinc(0) = 1.$$

At $k \to \infty$, $sinc(k)$ decays like $1/k$, but does so by alternating between positive and negative values. It is a good exercise to check that sinc is not absolutely integrable. It turns out that the Fourier transform can still be defined for it, so lack of integrability is not a major worry.

Example 18. *Consider the Gaussian function*

$$f(x) = e^{-x^2/2}.$$

By completing the square and adequately modifying the contour of integration in the complex plane (not part of the material for this class), it can be shown that

$$\hat{f}(k) = \sqrt{2\pi}\, e^{-k^2/2}.$$

Example 19. *The Dirac delta $\delta(x)$ has a FT equal to 1 (why?).*

Another basic property of Fourier transforms is the convolution theorem.

Theorem 11. *(The convolution theorem.) Denote convolution as $f \star g(x) = \int_{-\infty}^{\infty} f(y)g(x-y)\,dy$. Then*

$$\widehat{f \star g}(k) = \hat{f}(k)\,\hat{g}(k).$$

Proof. Let $h = f \star g$. We can use Fubini below provided every function is integrable.

$$\begin{aligned}
\hat{h}(k) &= \int e^{-ikx} \int f(y) g(x-y)\, dy dx \\
&= \int \int e^{-iky} f(y) e^{-ik(x-y))} g(x-y)\, dy dx \\
&= \left(\int e^{-iky} f(y)\, dy \right) \left(\int e^{-ikx'} g(x')\, dx' \right) \\
&= \hat{f}(k)\, \hat{g}(k).
\end{aligned}$$

□

The Fourier transform is an important tool in the study of linear differential equations because it turns differential problems into algebraic problems. For instance, consider a polynomial $P(x) = \sum a_n x^n$, and the ODE

$$P\left(\frac{d}{dx}\right) u(x) = f(x), \qquad x \in \mathbb{R},$$

which means $\sum a_n \frac{d^n u}{dx^n} = f$. (Such ODE are not terribly relevant in real life because they are posed over the whole real line.) Upon Fourier transformation, the equation becomes

$$P(ik)\hat{u}(k) = \hat{f}(k),$$

which is simply solved as

$$\hat{u}(k) = \frac{\hat{f}(k)}{P(ik)},$$

and then

$$u(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} \frac{\hat{f}(k)}{P(ik)}\, dk.$$

Beware the zeros of P when applying this formula! They always carry important physical interpretation. For instance, they could be resonances of a mechanical system.

The formula $\hat{u}(k) = \frac{\hat{f}(k)}{P(ik)}$ also lends itself to an application of the convolution theorem. Let $K(x)$ be the inverse Fourier transform of $1/P(ik)$. Then we have

$$u(x) = \int K(x-y) f(y)\, dy.$$

The function K is called *Green's function* for the original ODE.

6.2 Sampling and restriction

We aim to use Fourier transforms as a concept to help understand the accuracy of representing and manipulating functions on a grid, using a finite number of degrees of freedom. We also aim at using a properly discretized Fourier transform as a numerical tool itself.

For this purpose, $x \in \mathbb{R}$ and $k \in \mathbb{R}$ must be replaced by x and k on finite grids. Full discretization consists of *sampling* and *restriction*.

Let us start by sampling $x \in h\mathbb{Z}$, i.e., considering $x_j = jh$ for $j \in \mathbb{Z}$. The important consequence of sampling is that some complex exponential waves e^{ikx} for different k will appear to be the same on the grid x_j. We call *aliases* such functions that identify on the grid.

Definition 15. *(Aliases) The functions* e^{ik_1x} *and* e^{ik_2x} *are aliases on the grid* $x_j = jh$ *if*

$$e^{ik_1x_j} = e^{ik_2x_j}, \qquad \forall j \in \mathbb{Z}.$$

Aliases happen as soon as

$$k_1jh = k_2jh + 2\pi \times \text{integer}(j).$$

Letting $j = 1$, and calling the integer n, we have

$$k_1 - k_2 = \frac{2\pi}{h}n,$$

for some $n \in Z$. To wave numbers k_1, k_2 are indistinguishable on the grid if they differ by an integer multiple of $2\pi/h$.

For this reason, we restrict without loss of generality the wavenumber to the interval

$$k \in [-\pi/h, \pi/h].$$

We also call this interval the fundamental cell in frequency (in reference to a similar concept in crytallography.)

Real-life examples of aliases are rotating wheels looking like they go backwards in a movie, Moiré patterns on jackets on TV, and stroboscopy.

The proper notion of Fourier transform on a grid is the following.

Definition 16. *Let* $x_j = hj, \qquad f_j = f(x_j)$. *Semidiscrete Fourier transform (SFT):*

$$\hat{f}(k) = h \sum_{j=-\infty}^{\infty} e^{-ikx_j} f_j, \qquad k \in [-\pi/h, \pi/h]. \tag{6.3}$$

Inverse semidiscrete Fourier transform (ISFT):

$$f_j = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{ikx} \hat{f}(k)\, dk. \tag{6.4}$$

As we saw, sampling in x corresponds to a restriction in k. If one still wanted to peek outside $[-\pi/h, \pi/h]$ for the SFT, then the SFT would simply repeat by periodicity:

$$\hat{f}(k + \frac{2n\pi}{h}) = \hat{f}(k).$$

(why?). That's why we restrict it to the fundamental cell.

We can now define the proper notion of Fourier analysis for functions that are restricted to x in some interval, namely $[-\pi, \pi]$ for convention. Unsurprisingly, the frequency is sampled as a result. the following formulas are dual to those for the SFT.

Definition 17. *Fourier series (FS):*

$$\hat{f}_k = \int_{-\pi}^{\pi} e^{-ikx} f(x)\, dx. \tag{6.5}$$

Inverse Fourier series (IFS)

$$f(x) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{ikx} \hat{f}_k, \qquad x \in [-\pi, \pi]. \tag{6.6}$$

If one uses the Fourier series inversion formula (6.6) for x outside of its intended interval $[-\pi, \pi]$, then the function simply repeats by periodicity:

$$f(x + 2n\pi) = f(x).$$

(again, why?)

The two formulas (6.5) and (6.6) can be justified quite intuitively. The expression $\int f(x)\overline{g(x)}\, dx$ is an inner product on functions. It is easy to see that the complex exponentials

$$\sqrt{\frac{h}{2\pi}} e^{-ikx_j} = v_j(k)$$

form an orthonormal set of functions on $[-\pi/h, \pi/h]$, for this inner product. Hence, up to normalization constants, (6.5) is simply calculation of the coefficients in an orthobasis (analysis), and (6.6) is the synthesis of the function back from those coefficients. We'd have to understand more about the peculiarities of infinite-dimensional linear algebra to make this fully rigorous, but this is typically done in a real analysis class.

Here's an example of SFT.

Example 20.

$$f_j = \frac{1}{2a}\chi_{[-a,a]}(x_j)$$

Then

$$\begin{aligned}
\hat{f}(k) &= \frac{h}{2a}\sum_{j=-a}^{a} e^{-ikjh} \\
&= \frac{h}{2a}e^{ikah}\sum_{j=0}^{2a} e^{-ikjh} \\
&= \frac{h}{2a}e^{ikah}\frac{(e^{-ikh})^{2a+1}-1}{e^{-ikh}-1} \qquad \textit{(geometric series)} \\
&= \frac{h}{2a}\frac{\sin(kh(a+1/2))}{\sin(kh/2)}.
\end{aligned}$$

This function is called the discrete sinc. It looks like a sinc, but it periodizes smoothly when $k = -\pi/h$ and $k = \pi/h$ are identified.

Our first slogan is therefore:

Sampling in x corresponds to restriction/periodization in k, and restriction/periodization in k corresponds to sampling in x.

6.3 The DFT and its algorithm, the FFT

The discrete Fourier transform is what is left of the Fourier transfom when both space and frequency are sampled and restricted to some interval.

Consider

$$x_j = jh, \qquad j = 1, \ldots, N.$$

The point $j = 0$ is identified with $j = N$ by periodicity, so it is not part of the grid. If the endpoints are $x_0 = 0$ and $x_N = 2\pi$, then N and h relate as

$$h = \frac{2\pi}{N} \quad \Rightarrow \quad \frac{\pi}{h} = \frac{N}{2}.$$

For the dual grid in frequency, consider that N points should be equispaced between the bounds $[-\pi/h, \pi/h]$. The resulting grid is

$$k, \qquad k = -\frac{N}{2} + 1, \ldots, \frac{N}{2}.$$

We have the following definition.

Definition 18. *Discrete Fourier transform (DFT):*

$$\hat{f}_k = h \sum_{j=1}^{N} e^{-ikjh} f_j, \qquad k = -\frac{N}{2}, \ldots, \frac{N}{2}. \tag{6.7}$$

Inverse discrete Fourier transform (IDFT)

$$f_j = \frac{1}{2\pi} \sum_{k=-N/2+1}^{N/2} e^{ikjh} \hat{f}_k, \qquad j = 1, \ldots, N. \tag{6.8}$$

The DFT can be computed as is, by implementing the formula (6.7) directly on a computer. The complexity of this calculation is a $O(N^2)$, since there are N values of j, and there are N values of k over which the computation must be repeated.

There is, however, a smart algorithm that allows to group the computation of all the f_k in complexity $O(N \log N)$. It is called the fast Fourier transform (FFT). It is traditionally due to Tukey and Cooley (1965), but the algorithm had been discovered a few times before that by people who are not usually credited as much: Danielson and Lanczos in 1942, and well as Gauss in 1805.

The trick of the FFT is to split the samples of f into even samples (j even) and odd samples (j odd). Assume that N is a power of 2. Then

$$\begin{aligned}
\hat{f}_k &= h \sum_{j=1}^{N} e^{-ikjh} f_j \\
&= h \sum_{j=1}^{N/2} e^{-ik(2j)h} f_{2j} + h \sum_{j=1}^{N/2} e^{-ik(2j+1)h} f_{2j+1} \\
&= h \sum_{j=1}^{N/2} e^{-ikj(2h)} f_{2j} + h e^{ikh} \sum_{j=1}^{N/2} e^{-ikj(2h)} f_{2j+1}.
\end{aligned}$$

The first term on the last line is simply the DFT of length $N/2$, on a grid of spacing $2h$, of the even samples of f. The second term is, besides the multiplicative factor, the DFT of length $N/2$, on a grid of spacing $2h$, of the odd samples of f.

Note that those smaller DFTs would normally be only calculated for $k = -\frac{N}{4} + 1, \ldots, \frac{N}{4}$, but we need them for $k = -\frac{N}{2} + 1, \ldots, \frac{N}{2}$. This is not a big problem: we know that the DFT extends by periodicity outside the standard bounds for k, so all there is to do is copy $\hat{f}_k$ by periodicity outside of $k = -\frac{N}{4} + 1, \ldots, \frac{N}{4}$.

Already, we can see the advantage in this reduction: solving one problem of size N is essentially reduced to solving two problems of size $N/2$. Even better, the splitting into even and odd samples can be repeated recursively until the DFT are of size 1. When $N = 1$, the DFT is simply multiplication by a scalar.

At each stage, there are $O(N)$ operations to do to put together the summands in the equation of the last line above. Since there are $O(\log N)$ levels, the overall complexity is $O(N \log N)$.

There are variants of the FFT when N is not a power of 2.

6.4 Smoothness and truncation

In this section, we study the accuracy of truncation of Fourier transforms to finite intervals. This is an important question not only because real-life numerical Fourier transforms are restricted in k, but also because, as we know, restriction in k serves as a proxy for sampling in x. It will be apparent in Chapter 2, section 2.1 that every claim that we make concerning truncation of Fourier transforms will have an implication in terms of accuracy of sampling a function on a grid, i.e., how much information is lost in the process of sampling a function $f(x)$ at points $x_j = jh$.

We will manipulate functions in the spaces L^1, L^2, and L^∞. We have already encountered L^1.

Definition 19. *Let $1 \le p < \infty$. A function f of $x \in \mathbb{R}$ is said to belong to the space $L^p(\mathbb{R})$ when*

$$\int_{-\infty}^{\infty} |f(x)|^p \, dx < \infty.$$

Then the norm of f in $L^p(\mathbb{R})$ is $\|f\|_p = \left(\int_{-\infty}^{\infty} |f(x)|^p \, dx \right)^{1/p}$.

A function f of $x \in \mathbb{R}$ is said to belong to L^∞ when

$$\operatorname{ess\,sup} |f(x)| < \infty.$$

Then the norm of f in $L^\infty(\mathbb{R})$ is $\operatorname{ess\,sup} |f(x)|$.

In the definition above, "ess sup" refers to the essential supremum, i.e., the infimum over all dense sets $X \subset \mathbb{R}$ of the supremum of f over X. A set X is dense when $\mathbb{R} \backslash X$ has measure zero. The notions of supremum and infimum correspond to maximum and minimum respectively, when they are not necessarily attained. All these concepts are covered in a real analysis class. For us, it suffices to heuristically understand the L^∞ norm as the maximum value of the modulus of the function, except possibly for isolated points of discontinuity which don't count in calculating the maximum.

It is an interesting exercise to relate the L^∞ norm to the sequence of L^p norms as $p \to \infty$.

We will need the very important Parseval and Plancherel identitites. They express "conservation of energy" from the physical domain to the frequency domain.

Theorem 12. *(Parseval's identity). Let $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Then*

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)}\, dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(k)\overline{\hat{g}(k)}\, dk.$$

Proof. Let h be the convolution $f \star \tilde{g}$, where $\tilde{g}(x) = \overline{g}(-x)$. It is easy to see that the Fourier transform of $\tilde{g}$ is $\overline{\hat{g}(k)}$ (why?). By the convolution theorem (Section 1.1), we have

$$\hat{h}(k) = \hat{f}(k)\, \overline{\hat{g}(k)}.$$

If we integrate this relation over $k \in \mathbb{R}$, and divide by 2π, we get the IFT at $x = 0$:

$$h(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(k)\overline{\hat{g}(k)}\, dk.$$

On the other hand,

$$h(0) = \int_{-\infty}^{\infty} f(x)\overline{g(-(0-x))}\, dx = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\, dx.$$

□

Theorem 13. *(Plancherel's identity). Let $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Then*

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}(k)|^2 \, dk.$$

Proof. Apply Parseval's identity with $g = f$. □

(With the help of these formulas, it is in fact possible to extend their validity and the validity of the FT to $f, g \in L^2(\mathbb{R})$, and not simply $f, g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. This is a classical density argument covered in many good analysis texts.)

We need one more concept before we get to the study of truncation of Fourier transforms. It is the notion of total variation. We assume that the reader is familiar with the spaces $C^k(\mathbb{R})$ of bounded functions which are k times continuously differentiable.

Definition 20. *(Total variation) Let $f \in C^1(\mathbb{R})$. The total variation of f is the quantity*

$$\|f\|_{TV} = \int_{-\infty}^{\infty} |f'(x)| \, dx. \tag{6.9}$$

For functions that are not C^1, the notion of total variation is given by either expression

$$\|f\|_{TV} = \lim_{h \to 0} \int_{-\infty}^{\infty} \frac{|f(x) - f(x-h)|}{|h|} \, dx = \sup_{\{x_p\} \text{ finite subset of } \mathbb{R}} \sum_p |f(x_{p+1}) - f(x_p)|, \tag{6.10}$$

These more general expressions reduce to $\int_{-\infty}^{\infty} |f'(x)| \, dx$ when $f \in C^1(\mathbb{R})$. When a function has finite total variation, we say it is in the space of functions of bounded variation, or $BV(\mathbb{R})$.

The total variation of a piecewise constant function is simply the sum of the absolute value of the jumps it undergoes. This property translates to a useful intuition about the total variation of more general functions if we view them as limits of piecewise constant functions.

The important meta-property of the Fourier transform is that

decay for large $|k|$ corresponds to smoothness in x.

There are various degrees to which a function can be smooth or rates at which it can decay, so therefore there are several ways that this assertion can be made precise. Let us go over a few of them. Each assertion either expresses a decay (in k) to smoothness (in x) implication, or the converse implication.

- *Let $\hat{f} \in L^1(\mathbb{R})$ (decay), then $f \in L^\infty(\mathbb{R})$ and f is continuous (smoothness).* That's because $|e^{ikx}| = 1$, so

  $$|f(x)| \leq \frac{1}{2\pi} \int |e^{ikx} \hat{f}(k)| \, dk = \frac{1}{2\pi} \int |\hat{f}(k)| \, dk,$$

  which proves boundedness. As for continuity, consider a sequence $y_n \to 0$ and the formula

  $$f(x - y_n) = \frac{1}{2\pi} \int e^{ik(x-y_n)} \hat{f}(k) \, dk.$$

  The integrand converges pointwise to $e^{ikx}\hat{f}(k)$, and is uniformly bounded in modulus by the integrable function $|\hat{f}(k)|$. Hence Lebesgue's dominated convergence theorem applies and yields $f(x - y_n) \to f(x)$, i.e., continuity in x.

- *Let $\hat{f}(k)(1 + |k|^p) \in L^1(\mathbb{R})$ (decay). Then $f \in C^p$ (smoothness).* We saw the case $p = 0$ above; the justification is analogous in the general case. We write

  $$|f^{(n)}(x)| \leq \frac{1}{2\pi} \int |e^{ikx}(ik)^n \hat{f}(k)| \, dk \leq \int |k|^n |\hat{f}(k)| \, dk,$$

  which needs to be bounded for all $0 \leq n \leq p$. This is obviously the case if $\hat{f}(k)(1 + |k|^p) \in L^1(\mathbb{R})$. Continuity of $f^{(p)}$ is proved like before.

- *Let $f \in BV(\mathbb{R})$ (smoothness). Then $\hat{f}(k) \leq \|f\|_{TV} \, |k|^{-1}$ (decay).* If $f \in C^1 \cap BV(\mathbb{R})$, then this is justified very simply from (6.9), and

  $$ik\hat{f}(k) = \int e^{-ikx} f'(x) \, dx.$$

  Take a modulus on both sides, and get the desired relation

  $$|k||\hat{f}(k)| \leq \int |f'(x)| dx = \|f\|_{TV} < \infty.$$

When $f \in BV(\mathbb{R})$, but $f \notin C^1$, either of the more general formulas (6.10) must be used instead. It is a great practice exercise to articulate a modified proof using the $\lim_{h\to 0}$ formula, and properly pass to the limit.

- *Let f such that $f^{(k)} \in L^2(\mathbb{R})$ for $0 \leq k < p$, and assume that $f^{(p)} \in BV(\mathbb{R})$ (smoothness). Then there exists $C > 0$ such that $|\hat{f}(k)| \leq |k|^{-p-1}$ (decay).* This claim is also formulated as point (a) in Theorem 1 on page 30 of Trefethen's "Spectral methods in Matlab". The justification is very simple when $f \in C^{p+1}$: we then get

$$(ik)^{p+1}\hat{f}(k) = \int e^{-ikx} f^{(p+1)}(x)\, dx,$$

so

$$|k|^{p+1}|\hat{f}(k)| \leq \int |f^{(p+1)}(x)|\, dx = \|f^{(p)}\|_{TV} < \infty.$$

Again, it is a good exercise to try and extend this result to functions not in C^{p+1}.

- (This is the one we'll use later). (Same proof as above.)

In summary, let f have p derivatives in L^1. Then $|\hat{f}(k)| \leq C|k|^{-p}$. This is the form we'll make the most use of in what follows.

Example 21. *The canonical illustrative example for the two statements involving bounded variation is that of the B-splines. Consider*

$$s(x) = \frac{1}{2}\chi_{[-1,1]}(x),$$

the rectangle-shaped indicator of $[-1, 1]$ (times one half). It is a function in $BV(\mathbb{R})$, but it has no derivative in $L^2(\mathbb{R})$. Accordingly, its Fourier transform $\hat{s}(k)$ is predicted to decay at a rate $\sim |k|^{-1}$. This is precisely the case, for we know

$$\hat{s}(k) = \frac{\sin k}{k}.$$

Functions with higher regularity can be obtained by auto-convolution of s; for instance $s_2 = s \star s$ is a triangle-shaped function which has one derivative in

L^2, and such that $s_2' \in BV(\mathbb{R})$. We anticipate that $\hat{s_2}(k)$ would decay like $|k|^{-2}$, and this the case since by the convolution theorem

$$\hat{s_2}(k) = (\hat{s}(k))^2 = \left(\frac{\sin k}{k}\right)^2.$$

Any number of autoconvolutions $s \star s \ldots \star s$ can thus be considered: that's the family of B-splines.

The parallel between smoothness in x and decay in k goes much further. We have shown that p derivatives in x very roughly corresponds to an inverse-polynomial decay $|k|^{-p}$ in k. So C^∞ functions in x have so called super-algebraic decay in k: faster than $C_p|k|^{-p}$ for all $p \geq 0$.

The gradation of very smooth functions goes beyond C^∞ functions to include, in order:

- analytic functions that can be extended to a strip in the complex plane (like $f(x) = 1/(1+x^2)$), corresponding to a Fourier transform decaying exponentially (in this case $\hat{f}(k) = \pi e^{-|k|}$). That's Paley-Wiener theory, the specifics of which is not material for this course.

- analytic functions that can be extended to the whole complex plane with super-exponential growth (like $f(x) = e^{-x^2/2}$), whose Fourier transform decays faster than exponentially (in this case $\hat{f}(k) = \sqrt{\pi/2}e^{-k^2/2}$).

- analytic functions that can be extended to the whole complex plane with exponential growth at infinity (like $f(x) = \frac{\sin x}{x}$), whose Fourier transform is compactly supported (in this case $\hat{f}(k) = 2\pi\chi_{[-1,1]}(k)$). That's Paley-Wiener theory in reverse. Such functions are also called bandlimited.

More on this in Chapter 4 of Trefethen's book.

An important consequence of a Fourier transform having fast decay is that it can be truncated to some large interval $k \in [-N, N]$ at the expense of an error that decays fast as a function of N. The smoother f, the faster $\hat{f}$ decays, and the more accurate the truncation to large N in frequency. On the other hand, there may be convergence issues if $f(x)$ is not smooth.

To make this quantitative, put

$$\hat{f}_N(k) = \chi_{[-N,N]}(k)\hat{f}(k).$$

Recall that $\frac{1}{2\pi}\int_{-N}^{N} e^{ikx} dx = \frac{\sin(Nx)}{\pi x}$. By the convolution theorem, we therefore have

$$f_N(x) = \frac{\sin Nx}{\pi x} \star f(x).$$

In the presence of $f \in L^2(\mathbb{R})$, letting $N \to \infty$ always gives rise to convergence $f_N \to f$ in $L^2(\mathbb{R})$. This is because by Plancherel's identity,

$$\|f - f_N\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{f}_N(k) - \hat{f}(k)|^2 \, dk = \frac{1}{2\pi}\int_{|k|>N} |\hat{f}(k)|^2 \, dk.$$

This quantity tends to zero as $N \to \infty$ since the integral $\int_{\mathbb{R}} |\hat{f}(k)|^2$ over the whole line is bounded.

The story is quite different if we measure convergence in L^∞ instead of L^2, when f has a discontinuity. Generically, L^∞ (called uniform) convergence fails in this setting. This phenomenon is called *Gibb's effect*, and manifests itself through ripples in reconstructions from truncated Fourier transforms. Let us illustrate this on the Heaviside step function

$$u(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ 0 & \text{if } x < 0. \end{cases}$$

(It is not a function in $L^2(\mathbb{R})$ but this is of no consequence to the argument. It could be modified into a function in $L^2(\mathbb{R})$ by some adequate windowing.) Since $u(x)$ is discontinuous, we expect the Fourier transform to decay quite slowly. Consider the truncated FT

$$\hat{u}_N(k) = \chi_{[-N,N]}(k)\hat{u}(k).$$

Back in the x domain, this is

$$\begin{aligned} u_N(x) &= \frac{\sin Nx}{\pi x} \star u(x) \\ &= \int_0^\infty \frac{\sin N(x-y)}{\pi(x-y)} \, dy \\ &= \int_{-\infty}^{Nx} \frac{\sin y}{\pi y} \, dy \\ &\equiv s(Nx). \end{aligned}$$

(Draw picture)

The function $s(Nx)$ is a sigmoid function going from 0 at $-\infty$ to 1 at ∞, going through $s(0) = 1/2$, and taking on min. value $\simeq -.045$ at $x = -\pi/N$ and max. value $\simeq 1.045$ at $x = \pi/N$. The oscillations have near-period π/N. The parameter N is just a dilation index for the sigmoid, it does not change the min. and max. values of the function. Even as $N \to \infty$, there is no way that the sigmoid would converge to the Heaviside step in a uniform manner: we always have

$$\|u - u_N\|_\infty \gtrsim .045.$$

This example of Gibb's effect goes to show that truncating Fourier expansions could be a poor numerical approximation if the function is nonsmooth.

However, if the function is smooth, then everything goes well. Let us study the decay of the approximation error

$$\epsilon_N^2 = \|f - f_N\|_2^2$$

for truncation of the FT to $[-N, N]$. Again, we use Plancherel to get

$$\epsilon_N^2 = \int_{|k|>N} |\hat{f}(k)|^2 \, dk$$

Now assume that $\hat{f}(k) \leq C|k|^{-p}$, a scenario already considered earlier. Then

$$\epsilon_N^2 \leq C \int_{|k|>N} |k|^{-2p} \leq C' \, N^{-2p+1},$$

so, for some constant C'' (dependent on p but independent of N, hence called constant),

$$\epsilon_N \leq C'' \, N^{-p+1/2}.$$

The larger p, the faster the decay as $N \to \infty$.

When a function is C^∞, the above decay rate of the approximation error is valid for all $p > 0$. When the function is analytic and the Fourier transform decays exponentially, the decay rate of the approximation error is even faster, itself exponential (a good exercise). In numerical analysis, either such behavior is called *spectral accuracy*.

6.5 Chebyshev expansions

Chebyshev expansions, interpolation, differentiation, and quadrature rules are not part of the material for 18.330 in 2012.

In the previous sections, we have seen that smooth functions on the real line have fast decaying Fourier transforms.

On a finite interval, a very similar property hold for Fourier series: if a function is smooth in $[-\pi, \pi]$ and connects smoothly by periodicity at $x = -\pi$ and $x = \pi$, then its Fourier series (6.5) decays fast. Periodicity is essential, because the points $x = -\pi$ and $x = \pi$ play no particular role in (6.5). They might as well be replaced by $x = 0$ and $x = 2\pi$ for the integration bounds. So if a function is to qualify as smooth, it has to be equally smooth at $x = 0$ as it is at $x = -\pi$, identified with $x = \pi$ by periodicity.

For instance if a function f is smooth inside $[-\pi, \pi]$, but has $f(-\pi) \neq f(\pi)$, then for the purpose of convergence of Fourier series, f is considered discontinuous. We know what happens in this case: the Gibbs effect takes place, partial inverse Fourier series have unwelcome ripples, and convergence does not occur in L^∞.

If f is smooth and periodic, then it is a good exercise to generalize the convergence results of the previous section, from the Fourier transform to Fourier series.

How shall we handle smooth functions in intervals $[a, b]$, which do not connect smoothly by periodicity? The answer is not unique, but the most standard tool for this in numerical analysis are the *Chebyshev polynomials*.

For simplicity consider $x \in [-1, 1]$, otherwise rescale the problem. Take a C^k nonperiodic $f(x)$ in $[-1, 1]$. The trick is to view it as $g(\theta) = f(\cos\theta)$. Since $\cos[0, \pi] = \cos[\pi, 2\pi] = [-1, 1]$, all the values of $x \in [-1, 1]$ are covered twice by $\theta \in [0, 2\pi]$. Obviously, at any point θ, g inherits the smoothness of f by the chain rule. Furthermore, g is periodic since $\cos\theta$ is periodic. So g is exactly the kind of function which we expect should have a fast converging Fourier series:

$$\hat{g}_k = \int_0^{2\pi} e^{-ik\theta} g(\theta)\, d\theta, \qquad g(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{ik\theta} \hat{g}_k, \qquad k \in \mathbb{Z}.$$

Since $g(\theta)$ is even in θ, we may drop the $\sin\theta$ terms in the expansion, as well as the negative k:

$$\hat{g}_k = \int_0^{2\pi} \cos(k\theta) g(\theta)\, d\theta, \qquad g(\theta) = \frac{1}{2\pi}\hat{g}_0 + \frac{1}{\pi} \sum_{k=1}^{\infty} \cos(k\theta) \hat{g}_k, \qquad k \in \mathbb{Z}^+.$$

Back to f, we find

$$\hat{g}_k = 2\int_{-1}^{1} \cos(k \arccos x) f(x) \frac{dx}{\sqrt{1-x^2}}, \qquad f(x) = \frac{1}{2\pi}\hat{g}_0 + \frac{1}{\pi}\sum_{k=1}^{\infty} \cos(k \arccos x)\hat{g}_k, \qquad k \in \mathbb{Z}^+$$

The function $\cos(k \arccos x)$ happens to be a polynomial in x, of order k, called the *Chebyshev polynomial* of order k. Switch to the letter n as is usually done:

$$T_n(x) = \cos(n \arccos x), \qquad x \in [-1, 1].$$

The first few Chebyshev polynomials are

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1.$$

It is a good exercise (involving trigonometric identities) to show that they obey the recurrence relation

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

The orthogonality properties of $T_n(x)$ over $[-1, 1]$ follows from those of $\cos(n\theta)$ over $[0, \pi]$, but watch the special integration weight:

$$\int_{-1}^{1} T_m(x)T_n(x) \frac{dx}{\sqrt{1-x^2}} = c_n \delta_{mn},$$

with $c_n = \pi/2$ if $n = 0$ and $c_n = \pi$ otherwise. The Chebyshev expansion of f is therefore

$$\langle f, T_n \rangle = \int_{-1}^{1} T_n(x) f(x) \frac{dx}{\sqrt{1-x^2}}, \qquad f(x) = \frac{1}{\pi}\langle f, T_0\rangle + \frac{2}{\pi}\sum_{n=1}^{\infty} T_n(x)\langle f, T_n\rangle, \qquad k \in \mathbb{Z}^+.$$

Under the hood, this expansion is the FS of a periodic function, so we can apply results pertaining to Fourier series and Fourier transforms to obtain fast decay of Chebyshev expansions. This way, spectral accuracy is restored for nonperiodic functions defined on an interval.