Fourier Analysis

Edward Zeng

April 2, 2020

Contents

1 Least Squares

Suppose $T : V \to \mathbb{R}^n$ is linear and $y \in \mathbb{R}^n$. As the solution to the equation $Tx = y$ may not exist, we are concerned with the best ℓ^2 approximation of y from $\operatorname{ran}(T)$. This is the least squares problem.

To correctly formulate the theorem, we require the following exercise.

Exercise 1. *Show that T is injective on $\operatorname{ran}(T^*)$. Deduce that T^* is injective on $\operatorname{ran}(T)$. Conclude that T^*T can be restricted to an automorphism on $\operatorname{ran}(T^*)$.*

Theorem 1.1. *The least squares approximation of y is $T(T^*T)^{-1}T^*y$. Consequently, we best approximate y when we set $x = (T^*T)^{-1}T^*y$.*

Remark. Above, T^*T denotes the map T^*T restricted to $\operatorname{ran}(T^*)$. See Exercise 1.

Proof gist. Observe $T(T^*T)^{-1}T^*$ is the projection onto $\operatorname{ran}(T)$.

Key trick. Suppose W is a subspace of $\mathbb{R}^n$. Then P is the orthogonal projection to W iff $P^* = P$, $P^2 = P$, and $\operatorname{ran}(P) = W$.

2 Pointwise Convergence of Fourier Series

When does a Fourier series converge pointwise to the original function? To answer this question, we need the following tool:

Lemma 2.1 (Riemann-Lebesgue)**.** *Let f be a piecewise continuous function defined on $[-\pi, \pi]$. Then*

$$\int_{-\pi}^{\pi} f(x)e^{ikx}dx \to 0$$

as $k \to \infty$ or $k \to -\infty$.

Proof gist. Prove for $f \in C^1$ using integration by parts. Then extend to continuous functions via uniform approximation by C^1 functions.

Remark. The lemma still holds when we replace e^{ikx} with $\sin(kx)$ or $\cos(kx)$.

Now we are ready to state the result on pointwise convergence.

Theorem 2.2. *The Fourier series of f converges pointwise to itself if f is 2π-periodic and $C^1(\mathbb{R})$.*

Proof gist. Use Riemann-Lebesgue to prove

$$\left|f(x) - \frac{1}{2\pi}\Big(\sum_{n=-k}^{k}\int_{-\pi}^{\pi} f(t)e^{-int}dt\Big)e^{inx}\right| \to 0$$

as $k \to \infty$.

Key trick. To evaluate the above expression, we introduce the Dirichlet kernel

$$D_k(x) = \frac{1}{2\pi}\sum_{n=-k}^{k} e^{inx} = \frac{\sin(k+\frac{1}{2})x}{\sin\frac{1}{2}x}$$

and rewrite the expression as

$$\left|\int_{-\pi}^{\pi}\big(f(x) - f(t)\big)D_k(x-t)dt\right|.$$

Remark. We require f to be 2π-periodic and $f \in C^1(\mathbb{R})$ instead of just $f \in C^1([-\pi,\pi])$ to avoid some technical issues on the boundary.

Remark. Because of our formulation of Riemann-Lebesgue, Theorem 2.2 can be extended to continuous 2π-periodic functions that are piecewise C^1 with no extra effort.

3 Uniform Convergence of Fourier Series

Theorem 3.1. *The Fourier series of f converges uniformly to f if f is 2π-periodic and $f \in C^2(\mathbb{R})$.*

Proof gist. Use double integration by parts to show the k-th Fourier coefficent $c_k \sim \Theta(\frac{1}{k^2})$. Then invoke pointwise convergence.

Remark. The proof can be generalized to show that smoother functions have faster decaying Fourier coefficents. A more rigorous proof can be used to generalize the result to continuous 2π-periodic functions that are piecewise C^1.

4 The Fourier Transform

Definition 4.1. The Fourier transform of f is

$$\hat{f}(w) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(y)e^{-iwy}dy.$$

The Fourier transform comes about when extending the notion of Fourier series from $L^2([-\pi,\pi])$

to $L^2(\mathbb{R})$. We do this extension in 2 steps. First, note that for $f \in L^2([-m, m])$, we have

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2m} \Big(\int_{-m}^{m} f(y) e^{\frac{-in\pi y}{m}} dy \Big) e^{\frac{in\pi x}{m}},$$

where the equality is in the sense of L^2. If $f \in L^2(\mathbb{R})$, we can (non-rigorously) conclude

$$f(x) = \lim_{m\to\infty} \sum_{n=-\infty}^{\infty} \frac{1}{2m} \Big(\int_{-m}^{m} f(y) e^{\frac{-in\pi y}{m}} dy \Big) e^{\frac{in\pi x}{m}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} f(y) e^{-iwy} dy \Big) e^{iwx} dw.$$

Here, the last equality follows by turning a Riemann sum into an integral. When we compare the equations:

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi} \Big(\int_{-\infty}^{\infty} f(y) e^{-iwy} dy \Big) e^{iwx} dw, \quad f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2\pi} \Big(\int_{-\pi}^{\pi} f(y) e^{-iny} dy \Big) e^{inx},$$

we see that the Fourier transform is a generalization of Fourier coefficents (up to a scaling factor).

Note that in our derivation, we have also proved the Fourier inversion formula.

Theorem 4.1. *If $\hat{f}$ is the Fourier transform of f, then*

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(w) e^{iwx} dw.$$

Remark. As we have followed a nonrigorous derivation, we are not in the place to discuss the conditions to apply FT or IFT. We thus assume (but do not prove) that FT makes sense when $f \in L^1(\mathbb{R})$ and IFT makes sense when $f, \hat{f} \in L^1(\mathbb{R})$. These conditions will be implicit whenever we are working with FT or IFT.

5 Properties of the Fourier Transform

Of the many properties of the Fourier Transform, the following two are the most imporant. The proofs are straightforward.

Theorem 5.1. *Let $f, g \in L^1(\mathbb{R})$ where $f * g$ makes sense. Then*

$$\widehat{f * g} = \sqrt{2\pi} \hat{f} \hat{g}.$$

Theorem 5.2. *Let $f \in L^2(\mathbb{R})$ such that FT, IFT make sense. Then $\|f\|_2 = \|\hat{f}\|_2$.*

Proof gist. Show $\mathcal{F}^* = \mathcal{F}^{-1}$.

Remark. The Fourier transform can be linearly extended to all of $L^2(\mathbb{R})$ (although for $f \in L^2 - L^1$, the Definition 4.1 does not hold). Then by the above theorem, the Fourier transform is a unitary operator on $L^2(\mathbb{R})$.

6 FT Application I: Compact Support

Compact support is a powerful property that allows us to avoid difficult technicalities. The following is an important example.

Theorem 6.1 (Shannon-Whittaker). *Let f we defined so FT, IFT make sense. Then if $\hat{f}$ is continuous, piecewise C^1, and supported on $[-\Omega, \Omega]$, we have*

$$f(t) = \sum_{j=-\infty}^{\infty} f(\frac{j\pi}{\Omega}) \operatorname{sinc}(\Omega t - j\pi)$$

where the series converges uniformly.

Proof gist. Represent $\hat{f}$ with its Fourier series (which converges uniformly to $\hat{f}$) and then apply IFT. Key trick is to replace $\int_{-\infty}^{\infty}$ with $\int_{-\Omega}^{\Omega}$.

However, compact support is so powerful that it may be too good to always wish for. For instance, if both f and $\hat{f}$ are compactly supported, one can show using Leibniz's rule that $\hat{f}$ is infinitely differentiable. In fact, with some complex analysis, we can show that $\hat{f}$ is holomorphic! However, the only holomorphic function with compact support is the trivial map.

7 FT Application II: An Uncertainty Principle

Definition 7.1. Define the dispersion of f at a to be

$$\Delta_a f = \frac{\int_{-\infty}^{\infty} |t-a|^2 |f(t)|^2 dt}{\int_{-\infty}^{\infty} |f(t)|^2 dt}$$

whenever the RHS makes sense.

Note a small dispersion means that f is "localized" around a. We note state an uncertainty principle, which intuitively says that both f and $\hat{f}$ cannot both be localized.

Theorem 7.1. *For f such that $\Delta_a f$, $\Delta_\alpha \hat{f}$ make sense,*

$$\Delta_a f \Delta_\alpha \hat{f} \geq \frac{1}{4}.$$

The proof of this statement is rather slick, so we before we present the proof, we first provide a historial note.

Heisenberg's uncertainty principle states that momentum and position cannot be simultaneously well determined. Physicists in the 20th century generalized the ideas of momentum and position as the following two linear operators:

$$\text{Momentum: } \frac{d}{dx} : f(x) \to f'(x), \quad \text{Position: } x\mathbb{1} : f(x) \to xf(x).$$

The key mathematical reason behind Heisenberg's uncertainty principle is the fact that the two linear operators do not commute! This is our motivation.

Proof gist. Observe that the operators $\frac{d}{dx}$ and $x\mathbb{1}$ do not commute.

Proof. By translation and multiplication by a phase factor, we may assume $a = \alpha = 0$. For simplicity of this proof, we also assume f vanishes at $\pm\infty$. Then note

$$\begin{aligned}
\|f\|\|\hat{f}\| &= \|f\|^2 \\
&= \langle f, f\rangle \\
&= \langle (\frac{d}{dx})(x\mathbb{1})f - (x\mathbb{1})(\frac{d}{dx})f, f\rangle \\
&= \langle \frac{d}{dx}(x\mathbb{1})f, f\rangle - \langle (x\mathbb{1})(\frac{d}{dx})f, f\rangle \\
&= \langle (x\mathbb{1})f, (-\frac{d}{dx})f\rangle - \langle (\frac{d}{dx})f, xf\rangle \\
&= -2\Re(\langle xf, f'\rangle) \\
&\le 2\|xf\|\|f'\| \\
&= 2\|xf(x)\|\|w\hat{f}(w)\|
\end{aligned}$$

where the fifth equality comes from the fact that $\frac{d}{dx}$ is anti-Hermitian (use integration by parts and f vanishing at $\pm\infty$). $\square$

8 Filters

A filter takes an input signal and outputs an output signal. More formally, a filter is a map between function spaces, often L^1, L^2, or piecewise continuous functions. We are purposely vague on the precise domain and codomain, as they may differ depending on context.

Definition 8.1. A filter L is

(a) time-invariant if $(L[f])(x-a) = (L[f(\_ - a)])(x)$.

(b) causal if $f(x) = 0$ for $x < 0$ implies $L[f](x) = 0$ for $x < 0$.

In practice, we only care about linear time-invariant causal filters. Here, we attempt an informal characterization of such filters.

Exercise 2. *Consider the convolutional filter $L : L^2 \to L^2$ defined as $L[f] = f * g$ for some fixed $g \in L^2$. Show L is linear and time-invariant.*

Theorem 8.1. *Under some mild assumptions, all time-invariant linear filters are convolution filters.*

Remark. We add "mild assumptions" because we present an informal argument. We will not delve into the conditions that will make this argument formal.

The motivation for this proof is to note that for any convolutional filter L, we have $g = L[\delta]$, where δ is the Dirac delta function.

Proof gist. Show $L[f] = f * L[\delta]$.

Proof.

$$\begin{aligned}(f * L[\delta])(t) &= \int_{-\infty}^{\infty} f(x)L[\delta](t-x)dx \\ &= \int_{-\infty}^{\infty} f(x)L[\delta_x](t)dx \\ &= L[\int_{-\infty}^{\infty} f(x)\delta_x dx](t) \\ &= L[f](t)\end{aligned}$$

Here, the second equality comes from time invariance, and the third equality comes from linearity (intuitively, the integral can be seen as a sum). $\square$

Exercise 3. *Show that a convolutional filter L is causal iff $g(x) = 0$ a.e. when $x < 0$.*

9 The Discrete Fourier Transform

We now develop Fourier series in a different direction. Computers often compute integrals as Riemann sums, so the coefficents of Fourier series are often approximated as

$$c_k \approx a_k := \frac{1}{n}\sum_{j=0}^{n-1} f(\frac{2\pi j}{n})e^{-ik\frac{2\pi j}{n}}.$$

Thus given these n sample points, we can approximately calculate all Fourier coefficents. This is the discrete Fourier transform. We express our finding more compactly in the following definition.

Definition 9.1. DFT is the following matrix multiplication:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \frac{1}{n}\overline{F_n} \begin{pmatrix} f(0) \\ f(\frac{2\pi}{n}) \\ \vdots \\ f(\frac{2\pi(n-1)}{n}) \end{pmatrix}, \quad F_n = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots & w^{(n-1)^2} \end{pmatrix}, \quad w = e^{\frac{2\pi i}{n}}.$$

Remark. As the a_i's are cyclic, we need not calculate more than $a_1, \dots, a_n$.

One key property of DFT is that it can be computed in $\mathcal{O}(n \log n)$ time (using a divide-and-conquer technique called the FFT). This provides another reason why DFT is often preferable in practical situations.

Remark. There are many FT analogs in DFT, but the proofs are much easier in the discrete case (as many technicalities disappear).

10 Haar Decomposition

We now begin our discussion on signal processing. Our first aim is to develop a new orthogonal decomposition for $L^2(\mathbb{R})$ of $\mathbb{R}$-valued functions.

Definition 10.1. Let

(a) $\phi = \chi_{[0,1]}$, the Haar scaling function.

(b) $\psi = \frac{1}{2}\chi_{[0,\frac{1}{2}]} - \frac{1}{2}\chi_{[\frac{1}{2},1]}$, the Haar wavelet.

(c) $V_j = \operatorname{span}\{\phi(2^j x - k) : k \in \mathbb{Z}\}$.

(d) $W_j = \operatorname{span}\{\psi(2^j x - k) : k \in \mathbb{Z}\}$.

Using translation properties, it is easy to check that $V_j \oplus W_j = V_{j+1}$, an orthogonal direct sum. Then it is not hard to see the following orthogonal decomposition:

$$L^2(\mathbb{R}) = \overline{\operatorname{span}\{V_j\}} = \overline{V_0 \oplus W_0 \oplus W_1 \oplus \dots}.$$

Thus by Hilbert space theory, we conclude:

Theorem 10.1 (Haar Decomposition)**.** *For $f \in L^2(\mathbb{R})$, we have*

$$f = P_{V_0} f + \sum_{i=0}^{\infty} P_{W_i} f,$$

where $P_U f$ denotes the orthogonal projection of f onto U.

Intuitively, this means we can breakdown a signal f as $v_0 + \sum_{j \geq 0} w_j$ where v_0 is a baseline approximation, and each w_j adds finer and finer detail.

In practice, computing these projections is (relatively) slow. However, by sampling f at the right points, we can produce an approximation $\tilde{f} \in V_n$ for some big n, and then iteratively apply a fast decomposition algorithm that gives us $V_j = V_{j-1} \oplus W_{j-1}$. This gives a good approximation.