MATH 171 WIM

Michael Sun (SUNET: msun415)

November 2020

1 Introduction

When Fourier published his informal results on using a trigonometric series to solve a heat equation in his *Treatise on the propagation of heat in solid bodies*, little did he realize his work would become the foundation of a wide array of mathematical and physics problems and a beautiful branch of mathematics known as harmonic analysis. In this paper, we trace the progression of these results, beginning by motivating its definition, building on that with later formalization by famous successors like Riemann and Dirichlet, and finishing with an exploration of its wide-ranging applications.

2 Definitions

Consider a complicated 2π-periodic function $f(x)$ such that $f(x+2\pi) = f(x)$ for all x. Our goal is to decompose it into periodic sinusoids whose behavior and solutions are easier for us to study. Similar to how we studied Taylor Series, the composition of polynomials to approximate a function on any interval, our goal is to approximate f with the summation of trigonometric terms while maintaining the periodicity of f. We know $\cos x$ has period 2π, but in the same way we involved monomials of all powers to achieve the ability to approximate any continuous function, we'll compose terms with shorter periods but varying amplitudes (coefficients) to attain a similar ability. Specifically, we'll compose a weighted summation of the form

$$\sum_{n=0}^{\infty} C_n \cdot \cos(nx - \delta_n),$$

where each sinusoid $\cos(nx - \delta_n)$ has periodicity $\frac{2\pi}{n}$ and offset δ_n. The goal of Fourier analysis is the discovery of these coefficients. To simplify this to terms not involving δ_n, we can invoke the trigonometric identify for $\cos(x-y)$ to write $C_n \cdot \cos(nx - \delta_n) = C_n \cos\delta_n \cos(nx) + C_n \sin\delta_i n \sin(nx)$. Defining $a_n = C_n \cos\delta_n, b_n = C_n \sin\delta_n$ for $n \geq 0$, we arrive at our intended form, $\sum_{n=0}^{\infty}(a_n \cos(nx) + b_n \sin(nx))$ while preserving convenient identities at our disposal for analysis like $a_n^2 + b_n^2 = C_n^2$.

3 Convergence of Fourier Series

3.1 Dirichlet Kernel

To prove our composed summation for our complicated 2π-periodic function $f(x)$ converges asymptotically to what we want, we define:

$$S_N(x) = \sum_{n=0}^{N}(a_n \cos(nx) + b_n \sin(nx))$$
$$(a_n, b_n) = \left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx, \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx\right)$$

and study the limit of this sequence of partial sums. An expression that will come handy to evaluate is known as the Dirichlet Kernel, defined as $D_N(x) = \frac{1}{\pi}\left(\frac{1}{2} + \sum_{n=1}^{N}\cos(nx)\right)$. We proceed to prove some lemmas about it.

Lemma 1: $\int_{-\pi}^{\pi} D_N(x)dx = 1$
Proof:

$$\begin{aligned}
\int_{-\pi}^{\pi} D_N(x)dx &= \int_{-\pi}^{\pi} \frac{1}{\pi}\left(\frac{1}{2} + \sum_{n=1}^{N}\cos(nx)\right)dx \\
&= 1 + \sum_{n=1}^{N}\int_{-\pi}^{\pi}\cos(nx)dx \\
&= 1 + \sum_{n=1}^{N}\frac{\sin(nx)}{n}|_{-\pi}^{\pi} \\
&= 1 + \sum_{n=1}^{N}\frac{\sin(n\pi) - \sin(n\pi)}{n} = 1
\end{aligned}$$

Lemma 2: $S_N(x) = \int_{-\pi}^{\pi} f(y)D_N(x-y)dy$
Proof: We substitute in the expression from the definition of the Dirichlet Kernel and invoke the trigonometric identity $\cos(a-b) = \cos a\cos b + \sin a \sin b$:

$$\begin{aligned}
\int_{-\pi}^{\pi} f(y)D_N(x-y)dy &= \int_{-\pi}^{\pi} f(y)\cdot\frac{1}{\pi}\left(\frac{1}{2} + \sum_{n=1}^{N}\cos(nx-ny)\right)dy \\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} f(y)\left(\frac{1}{2} + \sum_{n=1}^{N}\cos(nx-ny)\right)dy \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)dy + \frac{1}{\pi}\sum_{n=1}^{N}\int_{-\pi}^{\pi}(\cos(nx)\cos(ny) + \sin(nx)\sin(ny))dy)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)dy \\
&+ \frac{1}{\pi}\sum_{n=1}^{N}\left(\cos(nx)\int_{-\pi}^{\pi} f(y)\cos(ny)dy + \sin(nx)\int_{-\pi}^{\pi} f(y)\sin(ny)dy\right) \\
&= \sum_{n=1}^{N}(a_n\cos(nx) + b_n\sin(nx)) \\
&= \sum_{n=0}^{N}(a_n\cos(nx) + b_n\sin(nx)) = S_N(x).
\end{aligned}
$$

Lemma 3: $D_N(x) = \frac{1}{2\pi}\frac{\sin((N+1)x)}{\sin(x/2)}$
Proof: Once again, we substitute in the expression from the definition of the Dirichlet Kernel, so

$$
\begin{aligned}
2\pi\sin(x/2)D_N(x) &= 2\pi\sin(x/2)\frac{1}{\pi}\left(\frac{1}{2} + \sum_{n=1}^{N}\cos(nx)\right) \\
&= \sum_{n=0}^{N}\sin(x/2)\cos(nx).
\end{aligned}
$$

The simplification of sequentially related summation terms and the resulting expression motivate telescoping! We can make use of the most relevant identity $2\sin(x/2)\cos(nx) = \sin((n+1/2)x) - \sin((n-1/2)x)$, so that:

$$
\begin{aligned}
2\pi\sin(x/2)D_N(x) &= \sin(x/2) + 2\sum_{n=1}^{N}\sin(x/2)\cos(nx) \\
&= \sin(x/2) + \sum_{n=1}^{N}(\sin((n+1/2)x) - \sin((n-1/2)x)) \\
&= \sin(x/2) + (\sin((2N+1)x/2) - \sin(x/2)) \\
&= \sin((2N+1)x/2)
\end{aligned}
$$

A natural curious question to ask is the asymptotic behavior of these coefficients. Take a_n. Intuitively, we visualize that as n increases, the $f(x)$ and $\cos(nx)$ terms are in a bevel-gear-like relationship, except the frictionless $\cos(nx)$ gear is rotating at faster and faster rates against the $f(x)$ gear whose constant rate independent of n. When n becomes really large, each arbitrarily small region, in which we know $\{f(x) : x \in [x_0 \pm \epsilon]\}$ is approximately uniform, is dotted against an arbitrarily large number of cycles of $\cos(nx)$. Because the integral of $\cos(nx)$ for one cycle is 0, we conjecture $a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx \to 0$, with a similar intuition for b_n. In fact, this was proven by Riemann and Lebesgue as a step to proving convergence.

Riemann-Lebesgue Lemma: If f is a 2π-periodic continuous function on $[-\pi, \pi]$, then $\lim_{n\to\infty} |a_n| = \lim_{n\to\infty} |b_n| = 0$.
Finally, armed with these lemmas, we return to prove our main goal.

Result: If f is continuous and differentiable on $[-\pi, \pi]$, then for all $x \in [-\pi, \pi], \{S_N(x)\}_{N=1}^{\infty} \to f(x)$.

Proof: Our goal is to show

$$\lim_{N\to\infty} S_N(x) - f(x) = \lim_{N\to\infty} \int_{-\pi}^{\pi} (f(y) - f(x)) D_N(x-y) dy = 0,$$

where $f(x)$ has been integrated (get it?) into the integral due to Lemma 1. As we'll have to manipulate $D_N(x-y)$, it's convenient to do a change of variables with $z = y - x$, picking off the fact $D_N(-x) = \frac{1}{2\pi} \frac{\sin(-(N+1)x)}{\sin(-x/2)} = \frac{1}{2\pi} \frac{-\sin((N+1)x)}{-\sin(x/2)} = D_N(x)$ from Lemma 2, so

$$\begin{aligned} S_N(x) - f(x) &= \int_{-\pi}^{\pi} (f(x+z) - f(x)) D_N(z) dz \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{f(x+z) - f(x)}{2\sin(z/2)} \cdot \sin((N+1/2)z) dz. \end{aligned}$$

We note a striking resemblance of this expression to a Fourier coefficient with $g_x(z) = \frac{f(x+z)-f(x)}{2\sin(z/2)}$, a function we know is also continuous on $[-\pi, \pi]$ as $f(x), 2\sin(z/2)$ are both continuous. The only difference is $\sin((N+1/2)z)$ is not of form $\sin(N'y), N' \in \mathbb{N}$, so we expand it out as $\sin((N+1/2)y) = \cos(z/2) \cdot \sin(Nz) + \sin(z/2) \cdot \cos(Nz)$ and we can coalesce $\cos(z/2), \sin(z/2)$ into two separated, refined functions $g_{1,x}(z) = g_x(z)\cos(z/2), g_{2,x}(z) + g_x(z)\sin(z/2)$, so to pick up where we left off

$$\begin{aligned} &\frac{1}{\pi} \int_{-\pi}^{\pi} \frac{f(x+z) - f(x)}{2\sin(z/2)} \cdot \sin((N+1/2)z) dz \\ =&\frac{1}{\pi} \int_{-\pi}^{\pi} g_x(z) \cdot \sin((N+1/2)z) dz \\ =&\frac{1}{\pi} \int_{-\pi}^{\pi} (g_x(z)\cos(z/2) \cdot \sin(Nz) + g_x(z)\sin(z/2) \cdot \cos(Nz)) dz \\ =&\frac{1}{\pi} \int_{-\pi}^{\pi} g_{1,x}(z) \sin(Nz) + \frac{1}{\pi} \int_{-\pi}^{\pi} g_{2,x}(z) \cos(Nz) \end{aligned}$$

as the first and second terms can be interpreted as coefficients of the Fourier series for $g_{1,x}$ and $g_{2,x}$ respectively, which retain continuity over $[-\pi, \pi]$ (as it's just $g(x)$ multiplied by continuous $\cos(z/2), \sin(z/2)$), so by the Riemann-Lebesgue Lemma, these coefficients have limit 0 as $N \to \infty$.

The hanging question is: is f's differentiability a necessary condition, or could f being continuous be sufficient to show the convergence of Fourier series? To motivate the latter, we'll employ an improvement on the Dirichlet kernel - the Fejer kernel, defined $K_N(t) = \frac{1}{N+1} \sum_{n=0}^{N} D_n(t)$. We first prove some basic lemmas that hold under this new kernel which will help us employ an analogous argument to show convergence.

Lemma 4: $\int_{-\pi}^{\pi} K_N(x)dx = 1$
Proof: $\int_{-\pi}^{\pi} K_N(x)dx = \frac{1}{N+1} \sum_{n=0}^{N} \int_{-\pi}^{\pi} D_n(t) = \frac{1}{N+1} \sum_{n=0}^{N} 1 = 1$

Like the Dirichlet kernel, there's a nice closed form for the Fejer kernel.
Lemma 5: $K_N(t) = \frac{1}{2(N+1)} \left[\frac{\sin((N+1)t/2)}{\sin(t/2)}\right]^2$.
Proof: We can invoke an analogous identity as before: $2 \sin x \sin y = \cos(x - y) - \cos(x + y)$ and its special case $2 \sin^2 x = 1 - \cos 2x$:

$$\begin{aligned}
K_N(t) &= \frac{1}{N+1} \sum_{k=0}^{N} D_k(t) \\
&= \frac{1}{N+1} \sum_{k=0}^{N} \frac{\sin((k+1/2)t)}{2\sin(t/2)} \\
&= \frac{1}{2(N+1)\sin^2(t/2)} \sum_{k=0}^{N} \sin((k+1/2)t)\sin(t/2) \\
&= \frac{1}{2(N+1)\sin^2(t/2)} \frac{\sum_{k=0}^{N}(\cos(kt) - \cos((k+1)t))}{2} \\
&= \frac{1}{2(N+1)\sin^2(t/2)} \frac{\cos 0 - \cos((n+1)t)}{2} \\
&= \frac{1}{2(N+1)\sin^2(t/2)} \sin((n+1)t/2)^2 \\
&= \frac{1}{2(N+1)} \left[\frac{\sin((N+1)t/2)}{\sin(t/2)}\right]^2.
\end{aligned}$$

Corollary: $K_N(t) \geq 0$

Beyond the fact $K_N(t)$, by virtue of Lemma 4, retains the same property as $D_N(t)$, the crucial difference is the concentration of $K_N(t)$ around 0, which first motivates the following lemma.

Lemma 6: Let $0 < \delta \leq |t| \leq \pi$. For all $\epsilon > 0$, there exists N such that if $n > N$ then $K_n(t) < \epsilon$.

Proof: For fixed t, pick N so $N > \frac{1}{2\epsilon \sin^2(t/2)} - 1$ so $K_N(t) = \frac{1}{2(N+1)} \left[\frac{\sin((N+1)t/2)}{\sin(t/2)}\right]^2 \leq \frac{1}{2(N+1)\sin^2(t/2)} = \frac{1}{\epsilon^{-1}\sin^2(t/2)} \sin^2(t/2) = \epsilon$.

Theorem: Let f be a continuous function. Define $\sigma_N(x) = \int_{-\pi}^{\pi} f(y)K_N(x-y)dy$. Show $\sigma_N(x)$ is a trigonometric polynomial. Show that $\sigma_N(x)$ converges to f uniformly on $[-\pi, \pi]$.

Proof:

$$\begin{aligned}
\int_{-\pi}^{\pi} f(y)K_N(x-y)dy &= \int_{-\pi}^{\pi} f(y)\frac{1}{N+1}\sum_{n=0}^{N} D_n(x-y)dy \\
&= \frac{1}{N+1}\sum_{n=0}^{N} f(y)D_n(x-y)dy \\
&= \frac{1}{N+1}\sum_{n_0=0}^{N}\sum_{n=0}^{N}(a_n\cos(nx) + b_n\sin(nx)) \\
&= \sum_{n=0}^{N}(a_n\cos(nx) + b_n\sin(nx))
\end{aligned}$$

where the substitution in line 3 follows from Lemma 2. Thus $\sigma_N(x)$ is trigonometric. Our approach for showing convergence is again to show that for all $x \in \mathbb{R}$ and arbitrarily small $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $|\sigma_N(x) - f(x)| = \int_{-\pi}^{\pi}(f(y) - f(x))K_N(x-y)dy < \epsilon$. We can again do a change of variables with $z = y - x$ exploiting the fact $K_N(-t) = \frac{1}{2(N+1)}\left[\frac{\sin(-(N+1)t/2)}{-\sin(t/2)}\right]^2 = K_N(t)$ and thus,

$$\sigma_N(x) - f(x) = \int_{-\pi}^{\pi}(f(x+z) - f(x))K_N(z)dz$$

Our strategy is to use the key difference between K_N and D_N as highlighted by Lemma 6. As alluded to from Lemma 6, we pick δ so whenever $x, y \in [-\pi, \pi]$, $|x-y| \leq \delta$, $|f(x) - f(y)| < \frac{\epsilon}{2}$, which we can do as f is uniform continuous on $[-\pi, \pi]$. Additionally, let $|f(x)| \leq M$. Let

$$\begin{aligned}
\int_{-\pi}^{\pi}(f(x+z) - f(x))K_N(z)dz &= \left(\int_{-\pi}^{-\delta} + \int_{-\delta}^{\delta} + \int_{\delta}^{\pi}\right)(f(x+z) - f(x))K_N(z)dz \\
&= \int_{t:\delta\leq|t|\leq\pi}(f(x+z) - f(x))K_N(z)dz \\
&\quad + \int_{-\delta}^{\delta}(f(x+z) - f(x))K_N(z)dz
\end{aligned}$$

Our goal is to show both integrals are less than $\frac{\epsilon}{2}$. We can easily do the first noting by Lemma 6, we can pick N so $K_N(z) < \frac{\epsilon}{8M\pi}$ for all $n > N$, so

$$\begin{aligned} \left| \int_{t:\delta \leq |t| \leq \pi} (f(x+z) - f(x))K_N(z)dz \right| &\leq \int_{t:\delta \leq |t| \leq \pi} |f(x+z) - f(x)|K_N(z)dz \\ &< 2\pi \cdot 2M \cdot \frac{\epsilon}{8M\pi} = \frac{\epsilon}{2}. \end{aligned}$$

where we applied the triangle inequality in the first line by virtue of $K_N(z) \geq 0$. We can do the second from the uniform convergence setup, noting

$$\begin{aligned} \left| \int_{-\delta}^{\delta} (f(x+z) - f(x))K_N(z)dz \right| &\leq \int_{-\delta}^{\delta} |(f(x+z) - f(x))K_N(z)dz| \\ &< \frac{\epsilon}{2} \int_{-\delta}^{\delta} K_N(z)dz \\ &\leq \frac{\epsilon}{2} \cdot 1 = \frac{\epsilon}{2} \end{aligned}$$

where the triangle inequality comes again from the corollary and the last line follows from Lemma 4. Thus,

$$\int_{t:\delta \leq |t| \leq \pi} (f(x+z) - f(x))K_N(z)dz + \int_{-\delta}^{\delta} (f(x+z) - f(x))K_N(z)dz < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and thus $\forall \epsilon$ there exists N so if $n > N, |\sigma_n(x) - f(x)| < \epsilon$ and σ_N converges uniformly to f on $[-\pi, \pi]$.

4 An Exploration

Though the Fourier series was invented to solve a heat equation, it has evolved to encompass much broader applications across many fields. One consequence of the theory of Fourier series is the Fourier Transform and I'll discuss an exciting application of it in machine learning, one of the most rapidly advancing fields in technology.

Given a complicated, periodic function g on a finite interval $[t_1, t_2]$ we imagine to be a (periodic) noisy signal, the theory of Fourier series allows us to decompose it into the summation of finite N terms of the form $\sum_{n=0}^{N} a_n \cos(nx - \delta_n)$. However, every "pure" signal has a fixed amplitude, shift, and frequency derived from a_n, δ_n and n respectively, with higher n corresponding to a higher frequency.

The Fourier Transform of g on $[-\pi, \pi]$, defined as $\hat{g}(f) = \int_{-\pi}^{\pi} g(t)e^{-2\pi i f t}dt$ where f stands for frequency produces potent spikes at the frequencies in which the decomposition produces high amplitudes. The result which allows us to apply this to real-life problems is the bijection between g and $\hat{g}$, which allows us to perform basic surgery on $\hat{g}$ and still recover the original g which in context could be the edited sound signal, void of an annoying frequency.

In practice, g would be incurred at a fixed sampling rate and produce a vector $[x_1, \ldots, x_N].T$, and the Discrete Fourier Transform is $\hat{g} : \mathbb{R}^N \to \mathbb{R}^N, \hat{g}(x)_k = \sum_{j=0}^{N-1} x_i e^{-i2\pi jk/N}$ which can be compactly summarized by the matrix multiplication $\hat{g}(x) = Wx$ such that $W \in \mathbb{R}^{N\times N}$, where $W_{m,n} = e^{-2\pi i mn/N}$, $0 \leq m, n \leq N-1$. We can denote W by W_N and $w_N = e^{-2\pi i/N}$.

In practice, performing the matrix multiplication has complexity $O(N^2)$, an expensive process, but it can be vastly reduced by the Fast Fourier Transform to $O(N \log N)$ (legend has that it was first discovered by Gauss for mental calculations but not published some hundred years later). I'll show how! Let $N = 2N'$. Suppose we want to compute $\hat{g}(x)_k = (W_N g(x))_k : 0 \leq k \leq N-1$. Observe

$$
\begin{aligned}
(W_N g(x)_k &= \sum_{n=0}^{N'-1} g(x))_{2n} w_N^{2nk} + \sum_{n=0}^{N'-1} g(x)_{2n+1} w_N^{(2n+1)k} \\
&= \sum_{n=0}^{N'-1} g(x)_{2n} w_{N'}^{nk} + w_N^k \sum_{n=0}^{N'-1} g(x)_{2n+1} w_{N'}^{nk} \\
&= (W_{N'} g(x)_{\text{even}})_k + w_N^k (W_{N'} g(x)_{\text{odd}})_k (*)
\end{aligned}
$$

where we split $g(x)_{\text{even}} = [g(x)_0, \ldots, g(x)_{2N'-2}].T, g(x)_{\text{odd}} = [g(x)_1, \ldots, g(x)_{2N'-1}].T$. In fact, this operation over all k can be computed in matrix block form,

$$
\begin{bmatrix} I_{N'} & D_{N'} \\ I_{N'} & D_{N'} \end{bmatrix} \begin{bmatrix} W_{N'} & 0 \\ 0 & W_{N'} \end{bmatrix} \begin{bmatrix} I_{N'\text{even}} \\ I_{N'\text{even}} \end{bmatrix} g(x) = W_{2N'} g(x)
$$

where $(I_{N'\text{even}})_{i,j} = 1$ if $2j = i$ else $0, (I_{N'\text{odd}})_{i,j} = 1$ if $2j+1 = i$ else 0 so

$$
\begin{bmatrix} I_{N'\text{even}} \\ I_{N'\text{even}} \end{bmatrix} g(x) = \begin{bmatrix} g(x)_{\text{even}} \\ g(x)_{\text{even}} \end{bmatrix}.
$$

Letting the above be written as $AB \begin{bmatrix} g(x)_{\text{even}} \\ g(x)_{\text{even}} \end{bmatrix} = W_{2N'} g(x)$, we can indeed verify, WLOG $0 \leq i \leq N'-1$ that

$$
(W_{2N'} g(x))_i = \sum_{j=0}^{N'-1} (AB)_{i,j} g(x)_{2j} + \sum_{j=N'}^{2N'-1} (AB)_{i,j} g(x)_{2j+1}
$$

$$
\begin{aligned}
&= \sum_{j=0}^{N'-1} \left(\sum_{k=0}^{N'-1} A_{i,k} B_{k,j} + \sum_{k=N'}^{2N'-1} A_{i,k} B_{k,j} \right) g(x)_{2j} \\
&+ \sum_{j=N'}^{2N'-1} \left(\sum_{k=0}^{N'-1} A_{i,k} B_{k,j} + \sum_{k=N'}^{2N'-1} A_{i,k} B_{k,j} \right) g(x)_{2j+1} \\
&= \sum_{j=0}^{N'-1} W_{N'i,j} g(x)_{2j} + \sum_{j=N'}^{2N'-1} w_{N'}{}^{i} W_{N'i,j-n} g(x)_{2j+1} \\
&= (*)(\text{except } i, k \text{ are swapped}).
\end{aligned}
$$

Recall that diagonal matrix - vector multiplication only takes $O(N)$, so denoting $T(N)$ as the amount of computation needed to compute $W_N g(x)$, we can set up the recursive relation $T(N) = O(N) + 2T(N') + O(N) = O(N) + 2T(N')$ and by the Master Theorem, this is sufficient to conclude $T(N) = O(N \log N)$. One amazing feature of the FFT operation is its near-synonymous inverse, $\frac{1}{N} \sum_{j=0}^{N-1} \hat{g}(x)_j w^{-jk}$, and we can show

$$
\begin{aligned}
\frac{1}{N} \sum_{j=0}^{N-1} \hat{g}(x)_j w^{-jk} &= \frac{1}{N} \sum_{j=0}^{N-1} \sum_{l=0}^{N-1} g(x)_l w^{lj} w^{-jk} \\
&= \frac{1}{N} \sum_{j=0}^{N-1} \sum_{l=0}^{N-1} w^{(l-k)j} \\
&= \frac{1}{N} \sum_{l=0}^{N-1} g(x)_l \sum_{j=0}^{N-1} w^{j(l-k)} \\
&= \frac{1}{N} \sum_{l=0}^{N-1} g(x)_k 1 = g(x)_k
\end{aligned}
$$

by virtue of the fact $w^n = 1$, and thus $w^{n(l-k)} - 1 = (w^{l-k} - 1)(w^{(n-1)(l-k)} + w^{(n-2)(l-k)} + \ldots + 1) = 0$, with $w^{l-k} = 1$ if and only if $l = k$, with all other l summing to 0, a common trick in Fourier series. Denoting this inverse $\mathcal{F}_N(x)_k = \frac{1}{N} \sum_{j=0}^{N-1} x_j w^{-jk}$ as the "Fourier operator" we then have $\mathcal{F}_N(W_N g(x)) = g(x)$. Here're a few more tricks we can do with the Fourier operator.

$$
\mathcal{F}_{N-k} x = \frac{1}{N} \sum_{j=0}^{N-1} x_j w^{-j(N-k)} = \frac{1}{N} \sum_{j=0}^{N-1} x_j w^{kj} = \frac{1}{N} \sum_{j=0}^{N-1} X_{N-j} w^{-kj}
$$

equating to

$$
(\mathcal{F}x)_{N-k} = (\mathcal{F}^{-1} x)_k = \mathcal{F}(Rx)
$$

where $(Rx)_k = x_{N-k}$. Now, we'll see an application of the FFT in machine learning.

If we let $\bullet, *$ represent the dot and convolution ($(x * y)_i = \sum_{j=0}^{i} x_j y_{i-j}$) operations while denoting $a = F_N(x), b = F_N(y)$, we have

$$\begin{aligned}
\mathcal{F}_N(x \bullet y)_i &= \frac{1}{N} \sum_{k=0}^{N-1} w^{-ik} x_k y_k \\
&= \frac{1}{N} \sum_{k=0}^{N-1} w^{-ik} x_k \sum_{j=0}^{N-1} w^{jk} b_j \\
&= \sum_{j=0}^{N-1} b_j \sum_{k=0}^{N-1} \left(\frac{1}{N} w^{(j-i)k} x_k \right) \\
&= \sum_{j=0}^{N-1} b_j a_{i-j} = (\mathcal{F}_N x * \mathcal{F}_N y)_i.
\end{aligned}$$

This manipulation, seemingly just a gimmick, shows **the dot product of x and y within the Fourier domain is the convolution of the Fourier domains of x and y**. This property, not shockingly, holds in 2D as well when $x, y \in \mathbb{R}^{N \times N}$ represent the sample image and filter to perform the convolution step $x * y$, a significant computation step in training a CNN, the model architecture behind recent advancements in deep learning. In 2D, the Fourier operation costs $O(N^2 \log N^2) = O(N^2 \log N)$. Usually, the filter $w \in \mathbb{R}^{K \times K}$ is a sliding window over the image, for a total of $N^2(K, K)$ convolutions that cost $O(N^2 K^2)$. However, by padding w to be size $N \times N$ and computing $x * y = F_N(F_N^{-1}(x) \bullet F_N^{-1}(y))$ where $\bullet$ has cost $O(N^2)$, **the whole operation now takes just** $O(N^2 \log N)$! Thanks to this, machine learning researchers have reported significant reductions in training time with an improved $O(N^2 K^2) \to O(N^2 \log N)$ convolution step, and we have the Fourier transform to thanks for it, alongside countless other performance improvements brought by it in the real world.

5 Citations

1. Johnsonbaugh, Richard, and W. E. Pfaffenberger. Foundations of Mathematical Analysis. Dover Publications, 2010.

2. `https://sites.math.washington.edu/~morrow/464_12/fft.pdf`

3. `http://www.math.utk.edu/~freire/m435f07/m435f07fourierconvergence.pdf`

4. `https://arxiv.org/pdf/2003.12621.pdf`