The Discrete Fourier Transform:
From Hilbert Spaces to the FFT

Hayden Borg

22 April 2021

1 Introduction

Fourier analysis, broadly, is the study of representing or approximating functions with sums of trigonometric functions or complex exponential functions. Fourier analysis has a wide range of applications including: solving partial differential equations, probability theory and statistics, and signal processing. Its applicability in signal processing gives it a vast number of use cases including: X-ray crystallography, infrared spectroscopy, nuclear magnetic resonance spectroscopy, MRIs, image processing, and audio editing.

This paper seeks to give readers an brief introduction to Fourier analysis beginning with Fourier series and the Fourier transform. In particular, this paper seeks to introduce readers to the Discrete Fourier Transform and the Fast Fourier Transform, as motivated by Hilbert spaces and Fourier series, using the 'tools' of linear algebra.

2 Hilbert Spaces

We begin with Hilbert spaces. Hilbert spaces can be intuitively understood as an extension or generalization of Euclidean spaces, such as $\mathbb{R}^2$ or $\mathbb{R}^3$, to $n-$dimensions for $n \in \mathbb{Z}^+$.

We define a Hilbert space to be an inner product space which is also a complete metric space. [1] For the purposes of this paper we will consider only real and complex inner product spaces.

2.1 Inner Product Spaces

An inner product is bifunction, linear in the first argument and conjugate linear in the second, acting on elements of a vector space which is:

1. Conjugate symmetric: $\langle \vec{x}, \vec{y} \rangle = \overline{\langle \vec{y}, \vec{x} \rangle}$
2. Linear in its first argument: $\langle a\vec{x_1} + b\vec{x_2}, \vec{y} \rangle = a\langle \vec{x_1}, \vec{y} \rangle + b\langle \vec{x_2}, \vec{y} \rangle$
3. Positive definite: $\langle \vec{x}, \vec{x} \rangle > 0$ for all $\vec{x} \neq \vec{0}$ and $\langle \vec{x}, \vec{x} \rangle = 0$ if and only if $\vec{x} = \vec{0}$

A vector space equipped with an inner product is an inner product space. We define the norm in an inner product space to be $||\vec{x}||$ where $||\vec{x}||^2 = \langle \vec{x}, \vec{x} \rangle$.

We then have some familiar results. [1]

Let V be an inner product space.

Theorem 1. *(Parallelogram Law.) For $\vec{x}, \vec{y} \in V$, $||\vec{x} + \vec{y}||^2 + ||\vec{x} - \vec{y}||^2 = 2||\vec{x}||^2 + 2||\vec{y}||^2$.*

Theorem 2. *(Pythagorean Theorem.) For $\vec{x}, \vec{y} \in V$ such that $\langle \vec{x}, \vec{y} \rangle = 0$, $||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2$.*

Theorem 3. *(Bessel's inequality.) If $\{\vec{v_i}\}$ is a finite orthonormal family of vectors then, $\sum_k |\langle \vec{v}, \vec{v_k} \rangle|^2 \leq ||\vec{v}||^2$ for every $\vec{v} \in \{\vec{v_i}\}$.*

Theorem 4. *(Schwartz's inequality.) For $\vec{x}, \vec{y} \in V$, $|\langle \vec{x}, \vec{y} \rangle|^2 \leq ||\vec{x}||||\vec{y}||$.*

2.2 Metric Spaces

A metric space is a set S with a metric μ which defines a distance between points in S. [2]

μ must satisfy the following for all $s_1, s_2, s_3 \in S$:

1. $\mu(s_1, s_2) = 0$ implies $s_1 = s_2$
2. Symmetric: $\mu(s_1, s_2) = \mu(s_2, s_1)$
3. Triangle Inequality: $\mu(s_1, s_3) \leq \mu(s_1, s_2) + \mu(s_2, s_3)$

A Cauchy sequence, $\{s_i\}$ for $i \geq 1$, in a metric space (S, μ) is a sequence in which for every positive, real number $\epsilon > 0$ there is an integer N for which $\mu(s_n, s_m) < \epsilon$ for $n, m > N$. That is to say the sequence converges.

A metric space (S, μ) is complete if every Cauchy sequence in S converges to a point in S. This is all, more or less, to say that we can 'do' calculus in a complete metric space.

The closure of a proper or improper subset of a metric space $S \subset (M, \mu)$ is defined $\overline{S} = S \cup \{\lim_{n\to\infty} s_n | s_n \in S \text{ for all } n \in \mathbb{N}\}$ A subset of a metric space (S, μ) is dense if $S = \overline{S}$. Informally, we may say that a set is dense if every point arbitrarily close to an element of that set is also in that set.

2.3 Hilbert Spaces

Theorem 5. *The norm in an inner product space is positive definite, positive homogeneous, and subadditive.*

Proof. The inner product is positive definite and so the norm is positive definite.

Grab $\alpha \in \mathbb{C}$ and $\vec{x} \in \mathcal{H}$. Then,

$$||\alpha\vec{x}||^2 = \langle \alpha\vec{x}, \alpha\vec{x}\rangle = \langle \overline{\alpha}\alpha\vec{x}, \vec{x}\rangle = \overline{\alpha}\alpha\langle \vec{x}, \vec{x}\rangle = |\alpha|^2||\vec{x}||^2.$$

So, the norm is positive homogeneous.

$$\begin{aligned}
||\vec{x}+\vec{y}||^2 &= \langle \vec{x}+\vec{y}, \vec{x}+\vec{y}\rangle \\
&\leq \langle \vec{x}, \vec{x}\rangle + |\langle \vec{x}, \vec{y}\rangle| + |\langle \vec{y}, \vec{x}\rangle| + \langle \vec{y}, \vec{y}\rangle \\
&= ||\vec{x}||^2 + |\langle \vec{x}, \vec{y}\rangle| + |\langle \vec{y}, \vec{x}\rangle| + ||\vec{y}||^2 \\
&\leq |\vec{x}||^2 + 2||\vec{x}||||\vec{y}|| + ||\vec{y}||^2 && \text{Schwartz's inequality} \\
&= ||\vec{x}||^2 + ||\vec{y}||^2
\end{aligned}$$

So, $||\vec{x}+\vec{y}|| \leq ||\vec{x}||^2 + ||\vec{y}||^2$ and the norm is subadditive. □

It then quickly follows:

Theorem 6. *If we define the distance between two elements of an inner product space, V, to be $||\vec{x}-\vec{y}||$, then V is a metric space with respect to $\mu(\vec{x}, \vec{y}) = ||\vec{x}-\vec{y}||$.*

Proof. The norm is positive definite so $||\vec{x}-\vec{y}|| = 0$ implies that $\vec{x} = \vec{y}$.

Observe: $\vec{x} - \vec{y} = -1(\vec{y} - \vec{x})$.
So using the positive homogeneity of the norm,

$$||\vec{x} - \vec{y}|| = || - 1(\vec{y} - \vec{x})|| = | - 1|||\vec{y} - \vec{x}|| = ||\vec{y} - \vec{x}||.$$

So μ is symmetric.

Then, observe: $\vec{x} - \vec{y} = (\vec{x} - \vec{z}) + (\vec{z} - \vec{y})$.
So,

$$||\vec{x} - \vec{y}|| = ||(\vec{x} - \vec{z}) + (\vec{z} - \vec{y})|| \leq ||(\vec{x} - \vec{z})|| + ||(\vec{z} - \vec{y})||.$$

Then, the triangle inequality holds for μ. $\square$

Given two metric spaces (M, μ) and (S, σ), a function $f : M \to S$ is uniformly continuous if for any real real $\epsilon > 0$ there exists a $\delta > 0$ such that for any $m_1, m_2 \in M$ with $\mu(m_1, m_2) < \delta$, $\sigma(f(m_1), f(m_2)) < \epsilon$.

We may consider vector addition, scalar multiplication, and the norm as defined for an inner product space, V, as functions $\Phi : V \to V$ or $\Phi : V \to \mathbb{R}$

- Vector addition with a fixed $\vec{y}$: $\Phi_{+,\vec{y}}(\vec{x}) = \vec{x} + \vec{y}$
- Scalar multiplication with a fixed scalar α: $\Phi_{\cdot,\alpha}(\vec{x}) = \alpha\vec{x}$
- Inner product with a fixed vector $\vec{y}$: $\Phi_{\cdot,\vec{y}}(\vec{x}) = \langle\vec{x}, \vec{y}\rangle$
- The norm: $\Phi_{|| \, ||}(\vec{x}) = ||\vec{x}||$

Theorem 7. *Vector addition ($\Phi_{+,\vec{y}}$), scalar multiplication ($\Phi_{\cdot,\alpha}$), the inner product ($\Phi_{\cdot,\vec{y}}$), and the norm ($\Phi_{|| \, ||}$) in a Hilbert space are uniformly continuous. [1]*

To summarize, a Hilbert space is a vector space equipped with an inner product which can be used to define a norm. This norm is then the metric for the vector space which is also a complete metric space.

At the risk of being too vague, we may say that a Hilbert space is a vector space with 'euclidean' geometry in which we can 'do' calculus.

Many of the concepts important to vector spaces may be generalized to Hilbert spaces.

A set in a Hilbert space is orthogonal if each element is orthogonal to every other element. That is $\{\vec{y_i}\}$ is orthogonal if $\langle\vec{y_i}, \vec{y_j}\rangle = 0$ for all $i \neq j$. If each $\vec{y} \in \{\vec{y_i}\}$ has $||\vec{y}|| = 1$, $\{\vec{y_i}\}$ is orthonormal.

Theorem 8. *Let $\{\vec{y_i}\}$ be an orthonormal set in Hilbert space $\mathcal{H}$ and $\vec{v} \in \mathcal{H}$ such that $\vec{v} = \sum_k a_k\vec{y_k}$. Then, $a_i = \langle\vec{v}, \vec{y_k}\rangle$ [3]*

Proof. $\langle\vec{v}, \vec{y_j}\rangle = \langle\sum_k a_k\vec{y_k}, \vec{y_j}\rangle = \sum_k a_k\langle\vec{y_k}, \vec{y_j}\rangle = a_k$ $\square$

The set $\{\vec{e_i}\}$ in a Hilbert space, $\mathcal{H}$, is an orthonormal basis if:

1. Orthogonality: $\langle\vec{e_j}, \vec{e_k}\rangle = 0$ for all $j \neq k$.

2. Normal: $||\vec{e_i}|| = 1$ for all $\vec{e_i}$.

3. Completeness: The linear span of $\{\vec{e_i}\}$ is dense in $\mathcal{H}$.

Note: the orthogonality of $\vec{e_i}$ guarantees linear independence.

2.4 Example 1

Define ℓ^2 to be the set of all sequences of complex numbers $a_0, a_1, a_2, \dots$, which we note $\{a_i\}$, such that $\sum\limits_{k=0}^{\infty} |a_k|^2$ converges.

Scalar multiplication and vector addition are defined in the obvious way.

Define the inner product: $\langle \{a_i\}, \{b_j\} \rangle = \sum\limits_{k=0}^{\infty} a_k \overline{b_k}$.

Define the norm $||\{a_i\}||^2 = \langle \{a_i\}, \{a_i\} \rangle$.

With these definitions, ℓ^2 is a Hilbert space. For the purpose of brevity, the proof is omitted.

Define $\vec{e_i}$ to be the sequence $\{\delta_{ij}\}$ where δ is the Kronecker delta. For example, $\vec{e_1} = \{1, 0, 0, 0, \dots\}$.

Proposition 1. *$B = \{\vec{e_i}$ for $i \in \mathbb{N}\}$ is an orthonormal basis of ℓ^2.*

Proof. First, $\sum\limits_{k=0}^{\infty} |[\vec{e_i}]_k|^2 = 1$ so $\vec{e_i} \in \ell^2$.

Then, for all $i \neq j$, $\langle \vec{e_i}, \vec{e_j} \rangle = \sum\limits_{k=0}^{\infty} [\vec{e_i}]_k \overline{[\vec{e_j}]_k} = 0$.

Then, $||\vec{e_i}||^2 = \sum\limits_{k=0}^{\infty} [\vec{e_i}]_k \overline{[\vec{e_i}]_k} = 1$.

So, B is an orthonormal set.

Let $\vec{a} = \{a_i\}$ be an element of ℓ^2.

Then, $\lim\limits_{n \to \infty} \sum\limits_{k=1}^{\infty} a_k \vec{e_k} = \{a_i\} = \vec{a}$. So, B is dense in ℓ^2.

Then, B is an orthonormal basis of ℓ^2. $\square$

2.5 Example 2

$L^2(\mathbb{R})$ is the set of all square-integrable functions with real input. That is the set of all functions with real input such that $\int\limits_{-\infty}^{\infty} |f(x)|^2 dx < \infty$.

We may also consider square-integrable functions over bounded intervals. Suppose $a \leq b$, $L^2([a, b])$ is the set of square integrable functions over the interval $[a, b]$. That is $\int\limits_{a}^{b} |f(x)|^2 dx < \infty$.

Let $f, g \in L^2(\mathbb{R})$. Then define $\langle f, g \rangle = \int\limits_{\mathbb{R}} f \overline{g} dx$.

Then, $||f||^2 = \langle f, f \rangle = \int\limits_{-\infty}^{\infty} f \overline{f} dx \int\limits_{-\infty}^{\infty} |f|^2 dx$.

We then have the following results which are presented without proofs for the sake of brevity.

Proposition 2. *$L^2(\mathbb{R})$ is a Hilbert space.*

Proposition 3. *$L^2([a, b]$ is a Hilbert space*

Now let's consider a basis.

Proposition 4. *The set $B = \{\frac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}kx} | k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2([0, T])$*

Proof. First, we must show that B is a subset of $L^2([0, T])$.

$$\begin{aligned} \int_0^T |\frac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}kx}|^2 dx &= \int_0^T |\frac{1}{\sqrt{T}}|^2 |e^{i\frac{2\pi}{T}kx}|^2 dx \\ &= \int_0^T \frac{1}{T} dx \\ &= 1 \end{aligned}$$

So our set, B, is in $L^2([0, T])$.

Now we must show our set is orthonormal.

$$\begin{aligned} \left\langle \frac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}nx}, \frac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}mx} \right\rangle &= \frac{1}{T} \int_0^T e^{i\frac{2\pi}{T}nx} \overline{e^{i\frac{2\pi}{T}mx}} dx \\ &= \frac{1}{T} \int_0^T e^{i\frac{2\pi}{T}(n-m)x} dx \\ &= \begin{cases} \frac{1}{T} \int_0^T 1 dx \text{ for } n = m \\ \frac{1}{i2\pi(n-m)} (e^{2\pi(n-m)} - e^0) \text{ for } n \neq m \end{cases} \\ &= \begin{cases} 1 \text{ for } n = m \\ 0 \text{ for } n \neq m \end{cases} \end{aligned}$$

So, our set is orthogonal
Then,

$$1 = \langle \tfrac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}nx}, \tfrac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}nx} \rangle = || \tfrac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}nx} ||^2$$

So, our set is orthonormal.

At this point we must prove that the span of B is dense in $L^2([0, T])$. However, this goes well beyond the scope of this course so stating this result will have to suffice. The proof uses the Stone-Weierstrass Theorem, so that may be a good stepping stone if one desires. [4] □

3 Fourier Series

The Fourier series is more or less a consequence of Proposition 4.

Theorem 9. *A complex valued function of a real variable, $f(x)$ which is square-integrable on the interval $[0,T]$ may be represented by the series $f(x) = \sum\limits_{n=-\infty}^{\infty} c_n e^{i\frac{2\pi}{T}nx}$ where $c_n = \frac{1}{T}\int\limits_0^T f(x)e^{-i\frac{2\pi}{T}nx}dx$*

Proof. $f(x) \in L^2([0,T])$ and $\{\frac{1}{\sqrt{T}}e^{i\frac{2\pi}{T}kx} | k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2([0,1])$.

We can rewrite Theorem 8: $\vec{v} = \sum\limits_k \langle \vec{v}, \vec{y_k}\rangle \vec{y_k}$

Then we can rewrite $f(x)$:

$$\begin{aligned} f(x) &= \sum_{k\in\mathbb{Z}} \left\langle f(x), \frac{1}{\sqrt{T}}e^{i\frac{2\pi}{T}kx} \right\rangle \frac{1}{\sqrt{T}}e^{i\frac{2\pi}{T}kx} \\ &= \sum_{k\in\mathbb{Z}} \left(\frac{1}{\sqrt{T}} \int_0^T f(x)e^{-i\frac{2\pi}{T}kx}dx \right) \frac{1}{\sqrt{T}}e^{i\frac{2\pi}{T}kx} \\ &= \sum_{k\in\mathbb{Z}} e^{i\frac{2\pi}{T}kx} \frac{1}{T} \int_0^T f(x)e^{-i\frac{2\pi}{T}kx}dx \end{aligned}$$

Defining $c_n = \frac{1}{T}\int\limits_0^T f(x)e^{-i\frac{2\pi}{T}nx}dx$ and modifying the index we have

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{2\pi}{T}nx}$$

□

By limiting the complex value function to the interval $[0,T]$ we assume that the function over $\mathbb{R}$ is periodic in T. The extension of Fourier series to aperiodic functions requires us to expand the interval to $\mathbb{R}$. This extension goes beyond the scope of this paper, but for the sake of completeness will be quickly presented. [4]

Given a continuous function $f(t)$ with real input, the Fourier Transform of $f(t)$ is $\hat{f}(\omega) = \int\limits_{-\infty}^{\infty} f(t)e^{i2\pi\omega t}dt$ (provided the integral converges.)

Then, $f(t) = \int\limits_{-\infty}^{\infty} \hat{f}(\omega)e^{i2\pi\omega t}dw$.

3.1 Example: Square Wave [5]

Define our square wave function with period $T = 1$ on the interval $[0,1]$:

$$f(x) = \begin{cases} 1 \text{ for } 0 \leq x < \frac{1}{2} \\ -1 \text{ for } \frac{1}{2} \leq x < 1 \end{cases}$$

Then,

$$\begin{aligned} c_n &= \int\limits_0^1 f(x) e^{-i2\pi nx} dx \\ &= \int\limits_0^{\frac{1}{2}} e^{-i2\pi nx} dx - \int\limits_{\frac{1}{2}}^1 e^{-i2\pi nx} dx \\ &= \frac{-1}{i2\pi n}(e^{-i\pi n} - 1) - \frac{-1}{i2\pi n}(1 - e^{-i\pi n}) \\ &= \frac{-1}{in\pi}(e^{-in\pi} - 1) \\ &= \frac{i}{n\pi}((-1)^n - 1) \\ &= \begin{cases} 0 \text{ if } n \text{ is even} \\ -\frac{2i}{n\pi} \text{ if } n \text{ is odd} \end{cases} \end{aligned}$$

So, $f(x) = \sum\limits_{n \text{ is odd}} -\frac{2i}{n\pi} e^{i2\pi nx} = -\frac{2i}{\pi} \sum\limits_{n \text{ is odd}} \frac{1}{n} e^{i2\pi nx}$

4 The Discrete Fourier Transform

Almost always measurements, incoming signals, image, etc. are not continuous functions, but discrete data. But, the Fourier series is limited to continuous, periodic functions.

The Discrete Fourier Transform (DFT) is an extension of the Fourier series and Fourier Transform to discrete data. At this point it is pertinent to note that the DFT may refer to a discretized Fourier series or discretized Fourier Transform. Because the data is inherently limited in these situations, the computations involved in extending the Fourier series and Fourier Transform are identical although they may be considered mathematically distinct. So, this paper will focus on the DFT as motivated by Fourier series.

The Fourier series is more or less an application of Theorem 8: $\vec{v} = \sum\limits_k \langle \vec{v}, \vec{y_k} \rangle \vec{y_k}$ where $\{\vec{y_i}\}$ is the special basis $\{\frac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}kx} | k \in \mathbb{Z}\}$.

We can treat N complex data points as a vector $\vec{v} \in \mathbb{C}^N$.

Consider this data to be sampled points from a continuous function f.

Then, sample the 'basis' functions $\{\frac{1}{\sqrt{T}} e^{i\frac{2\pi}{T}kx} | k \in \mathbb{Z}\}$ at N points:

$$e^{i\frac{2\pi}{N}0x} \rightarrow \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$e^{i\frac{2\pi}{N}x} \rightarrow \begin{bmatrix} 1 \\ e^{i\frac{2\pi}{N}} \\ e^{i2\frac{2\pi}{N}} \\ \vdots \\ e^{i(N-1)\frac{2\pi}{N}} \end{bmatrix}$$

We can more elegantly write these vectors using the Nth root of unity, $\overline{\omega_N} = e^{i\frac{2\pi}{N}}$:

$$e^{i\frac{2\pi}{N}x} \rightarrow \begin{bmatrix} \overline{\omega_N}^0 \\ \overline{\omega_N}^1 \\ \overline{\omega_N}^2 \\ \vdots \\ \overline{\omega_N}^{(N-1)} \end{bmatrix}$$

In general,

$$e^{i\frac{2\pi}{N}kx} \rightarrow \begin{bmatrix} \overline{\omega_N}^{0\cdot k} \\ \overline{\omega_N}^{1\cdot k} \\ \overline{\omega_N}^{2\cdot k} \\ \vdots \\ \overline{\omega_N}^{(N-1)\cdot k} \end{bmatrix}$$

Lemma 10. *While computing the DFT, we only need to consider the N basis vectors $\{\frac{1}{\sqrt{T}}e^{i\frac{2\pi}{T}kx} | 0 \leq k \leq N-1\}$*

Proof. Consider $k = N$:

$$e^{i\frac{2\pi}{N}Nx} \rightarrow \begin{bmatrix} \overline{\omega_N}^{0\cdot N} \\ \overline{\omega_N}^{1\cdot N} \\ \overline{\omega_N}^{2\cdot N} \\ \vdots \\ \overline{\omega_N}^{(N-1)\cdot N} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

And $k = N + 1$:

$$e^{i\frac{2\pi}{N}(N+l)x} \rightarrow \begin{bmatrix} \overline{\omega_N}^{0\cdot (N+1)} \\ \overline{\omega_N}^{1\cdot (N+1)} \\ \overline{\omega_N}^{2\cdot (N+1)} \\ \vdots \\ \overline{\omega_N}^{(N-1)\cdot (N+1)} \end{bmatrix} = \begin{bmatrix} \overline{\omega_N}^0 \\ \overline{\omega_N}^1 \\ \overline{\omega_N}^2 \\ \vdots \\ \overline{\omega_N}^{(N-1)} \end{bmatrix}$$

And generally $k = N + l$:

$$e^{i\frac{2\pi}{N}(N+l)x} \rightarrow \begin{bmatrix} \overline{\omega_N}^{0\cdot (N+l)} \\ \overline{\omega_N}^{1\cdot (N+l)} \\ \overline{\omega_N}^{2\cdot (N+l)} \\ \vdots \\ \overline{\omega_N}^{(N-l)\cdot (N+l)} \end{bmatrix} = \begin{bmatrix} \overline{\omega_N}^0 \\ \overline{\omega_N}^1 \\ \overline{\omega_N}^2 \\ \vdots \\ \overline{\omega_N}^{(N-1)} \end{bmatrix} = \begin{bmatrix} \overline{\omega_N}^{0\cdot l} \\ \overline{\omega_N}^{1\cdot l} \\ \overline{\omega_N}^{2\cdot l} \\ \vdots \\ \overline{\omega_N}^{(N-1)\cdot l} \end{bmatrix}$$

So, we only need to consider the first N basis vectors. Sampling any basis vectors outside of $\{\frac{1}{\sqrt{T}}e^{i\frac{2\pi}{T}kx}|0 \leq k \leq N-1\}$ will only introduce repeated vectors. $\square$

For notational convenience define:

$$\vec{e_i} = \begin{bmatrix} \overline{\omega_N}^{0\cdot i} \\ \overline{\omega_N}^{1\cdot i} \\ \overline{\omega_N}^{2\cdot i} \\ \vdots \\ \overline{\omega_N}^{(N-1)\cdot i} \end{bmatrix}$$

Then using our results from Theorem 8 and Lemma 10, $\vec{v} = \sum_{k=0}^{N-1} \langle \vec{v}, \vec{e_k} \rangle \vec{e_k}$.

We may encode this linear combination of basis vectors as a new vector in $\mathbb{C}^N$ where the entry is the scalar and the index is the index of the basis vector. This vector is the transformed vector.

That is, the transform of $\vec{v}$ is $\hat{\vec{v}}$ where $[\hat{\vec{v}}]_i = \langle \vec{v}, \vec{e_i} \rangle = \langle \vec{e_i}, \vec{v} \rangle$. Note that then, $[\vec{v}]_i = [\hat{\vec{v}}]_i \vec{e_i}$ is equivalent to $\vec{v} = \sum_{k=0}^{N-1} \langle \vec{v}, \vec{e_k} \rangle \vec{e_k}$.

So, $[\hat{\vec{v}}]_i = \vec{e_i}^* \vec{v}$. It then becomes natural to write $\hat{\vec{v}}$ as a matrix vector product:

$$\hat{\vec{v}} = \left[\vec{e_0}|\vec{e_1}| \dots |\vec{e_{N-1}}\right]^* \vec{v}$$

We then define the DFT to be the linear transformation $T : \mathbb{C}^N \to \mathbb{C}^N$ defined by the matrix vector product:

$$T(\vec{v}) = \begin{bmatrix} \omega_N^{0\cdot 0} & \omega_N^{0\cdot 1} & \dots & \omega_N^{0\cdot (N-1)} \\ \omega_N^{1\cdot 0} & \omega_N^{1\cdot 1} & \dots & \omega_N^{1\cdot (N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_N^{N\cdot 0} & \omega_N^{N\cdot 1} & \dots & \omega_N^{N\cdot (N-1)} \end{bmatrix} \vec{v}$$

where $\omega_N = e^{-i\frac{2\pi}{N}}$ is the Nth root of unity. In this paper, we will write this matrix $\mathcal{F}$.

Re-introducing the normalizing factor $\frac{1}{\sqrt{N}}$ can be useful and is mathematically interesting.

Define $\mathcal{U} = \frac{1}{\sqrt{N}}\mathcal{F}$.

Proposition 5. *$\mathcal{U}$ is a unitary matrix.*

Proof. $\mathcal{U} = \frac{1}{\sqrt{N}} \begin{bmatrix} \omega_N^{0\cdot 0} & \omega_N^{0\cdot 1} & \dots & \omega_N^{0\cdot (N-1)} \\ \omega_N^{1\cdot 0} & \omega_N^{1\cdot 1} & \dots & \omega_N^{1\cdot (N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_N^{(N-1)\cdot 0} & \omega_N^{(N-1)\cdot 1} & \dots & \omega_N^{(N-1)\cdot (N-1)} \end{bmatrix}$

So, $[\mathcal{U}]_{ij} = \frac{1}{\sqrt{N}}\omega_N^{(i-1)\cdot(j-1)}$.

Then, $[\mathcal{U}^*]_{ij} = \overline{[\mathcal{U}]_{ji}} = \frac{1}{\sqrt{N}}\overline{\omega_N}^{(j-1)\cdot(i-1)}$.

So,

$$
\begin{aligned}
[\mathcal{U}\mathcal{U}^*]_{ij} &= \sum_{k=0}^{N-1} [\mathcal{U}]_{ik}[\mathcal{U}^*]_{kj} \\
&= \sum_{k=0}^{N-1} [\mathcal{U}]_{ik}\overline{[\mathcal{U}]}_{jk} \\
&= \frac{1}{N}\sum_{k=0}^{N-1} \omega_N^{(i-1)(k-1)}\overline{\omega_N}^{(j-1)(k-1)} \\
&= \frac{1}{N}\sum_{k=0}^{N-1} (\omega_N\overline{\omega_N})^{(i-1)(k-1)}\omega_N^{(i-j)(k-1)} \\
&= \frac{1}{N}\sum_{k=0}^{N-1} 1^{(i-1)(k-1)}\omega_N^{(i-j)(k-1)} \\
&= \frac{1}{N}\sum_{k=0}^{N-1} \omega_N^{(i-j)(k-1)} \\
&= \begin{cases} \frac{1}{N}\sum_{k=0}^{N-1} 1^{(k-1)} & \text{for } i = j \\ \frac{1}{N}\sum_{k=0}^{N-1} \omega_N^{(i-j)(k-1)} & \end{cases} \\
&= \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}
\end{aligned}
$$

Then, $\mathcal{U}\mathcal{U}^* = I_N$.

Direct computation shows: $\frac{1}{N}\sum_{k=0}^{N-1} \omega_N^{(i-j)(k-1)} = 0$ for $i \neq j$. This can intuitively be understood as taking the average of N, Nth roots of unity. Geometrically, its is easy to see this is 0.

□

5 Cooley-Tukey Algorithm and Matrix Decomposition

This matrix-vector product is useful for defining the DFT, but it is computationally inefficient. Directly computing the entries of the DFT matrix and the matrix vector product is $\mathcal{O}(n^2)$, so the naive algorithm for computing the DFT $2\mathcal{O}(n^2)$. While this may not seem very computationally complex, data sets can frequently be millions of points long. The symmetries in the DFT matrix as well as the symmetries in calculating its entries can be exploited to give faster algorithms for computing the DFT. These algorithms are called Fast Fourier transform (FFT) algorithms

This paper will focus on perhaps the first FFT algorithm, the Cooley-Tukey algorithm from the perspective of matrix decomposition.

We first need some notation.

Given a vector $\vec{v}$, let $\vec{v}_{\text{even}}$ to be the vector containing the entries of $\vec{v}$ with even index and $\vec{v}_{\text{odd}}$ to be the vector containing the entries of $\vec{v}$ with odd index.

Note: for this section we've adopted the convention of the first entry having index 0

That is

$$[\vec{v}_{\text{even}}]_k = [\vec{v}]_{2k} \text{ and } [\vec{v}_{\text{odd}}]_k = [\vec{v}]_{2k+1}.$$

For example,

$$\vec{v} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 1 \\ 3 \\ 1 \\ 5 \end{bmatrix} \to \vec{v}_{\text{even}} = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 5 \end{bmatrix}, \vec{v}_{\text{odd}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Let the even-odd permutation matrix be the matrix which permutes the entries of a vector so that the entries with even index, in order, are first followed by the entries with odd index, in order.

That is,

$$\begin{bmatrix} \text{even-odd} \\ \text{permutation} \end{bmatrix} \vec{v} = \begin{bmatrix} \vec{v}_{\text{even}} \\ \vec{v}_{\text{odd}} \end{bmatrix}.$$

For example, the even-odd permutation matrix of size 4 is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Given a DFT on an $N-$ dimensional vector, which we may call an $N-$points DFT, define D_n to be the $n \times n$ diagonal matrix with entries $\omega_N^0, \omega_N, \ldots, \omega_N^{n-1}$.

For example, given an $8-$point DFT,

$$D_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \omega_8 & 0 & 0 \\ 0 & 0 & \omega_8^2 & 0 \\ 0 & 0 & 0 & \omega_8^3 \end{bmatrix}$$

Lemma 11. *(Danielson-Lanczos Lemma) [7] The DFT for $N = 2^m$ for some $m \in \mathbb{N}$, $[\hat{\vec{v}}]_i = \sum_{k=0}^{N-1} [\vec{v}]_k [\mathcal{F}_N]_{ik}$, may be rewritten*

$$[\hat{\vec{v}}]_i = \mathcal{F}_{\frac{N}{2}} \vec{v}_{\text{even}} + D_{\frac{N}{2}} \mathcal{F}_{\frac{N}{2}} \vec{v}_{\text{odd}} \textit{ for } 0 \le i \le \tfrac{N}{2} - 1$$

$$[\hat{\vec{v}}]_i = \mathcal{F}_{\frac{N}{2}}\vec{v}_{\text{even}} - D_{\frac{N}{2}}\mathcal{F}_{\frac{N}{2}}\vec{v}_{\text{odd}} \textit{ for } \tfrac{N}{2} < i < N-1$$

Proof.

$$\begin{aligned}
[\hat{\vec{v}}]_l &= [\mathcal{U}_N\vec{v}]_l \\
&= \sum_{k=0}^{N-1}[\vec{v}]_k\omega_N^{lk} \\
&= \sum_{k=0}^{\frac{N}{2}-1}[\vec{v}]_{2k}\omega_N^{l(2k)} + \sum_{k=0}^{\frac{N}{2}-1}[\vec{v}]_{2k+1}\omega_N^{l(2k+1)} \\
&= \sum_{k=0}^{\frac{N}{2}-1}[\vec{v}]_{2k}\omega_{\frac{N}{2}}^{l(k)} + \omega_N^l\sum_{k=0}^{\frac{N}{2}-1}[\vec{v}]_{2k+1}\omega_{\frac{N}{2}}^{lk}
\end{aligned}$$

If we let $0 \le l \le \frac{N}{2}-1$ this simply becomes:

$$[\hat{\vec{v}}]_l = [\mathcal{F}_{\frac{N}{2}}\vec{v}_{\text{even}}]_l + \omega_N^l[\mathcal{F}_{\frac{N}{2}}\vec{v}_{\text{odd}}]_l.$$

For $\frac{N}{2} < l \le N$, define $\hat{l} = l - \frac{N}{2}$. Then we may rewrite:

$$\begin{aligned}
[\hat{\vec{v}}]_l &= \sum_{k=0}^{\frac{N}{2}-1}\omega_{\frac{N}{2}}^{\frac{N}{2}}[\vec{v}]_{2k}\omega_{\frac{N}{2}}^{\hat{l}(2k)} + \omega_N^{\hat{l}}\omega_N^{\frac{N}{2}}\sum_{k=0}^{\frac{N}{2}-1}\omega_{\frac{N}{2}}^{\frac{N}{2}}[\vec{v}]_{2k+1}\omega_{\frac{N}{2}}^{\hat{l}(2k)} \\
&= \sum_{k=0}^{\frac{N}{2}-1}[\vec{v}]_{2k}\omega_{\frac{N}{2}}^{\hat{l}(2k)} - \omega_N^{\hat{l}}\sum_{k=0}^{\frac{N}{2}-1}[\vec{v}]_{2k+1}\omega_{\frac{N}{2}}^{\hat{l}(2k)} \\
&= [\mathcal{F}_{\frac{N}{2}}\vec{v}_{\text{even}}]_{\hat{l}} - \omega_N^{\hat{l}}[\mathcal{F}_{\frac{N}{2}}\vec{v}_{\text{odd}}]_{\hat{l}}
\end{aligned}$$

□

Using this Lemma, we can decompose DFT matrices with size 2 to an integer power. [8]

Theorem 12. *Let $N = 2^k$ where $k \in \mathbb{N}$. Then,*

$$\mathcal{F}_N = \begin{bmatrix} I_{\frac{N}{2}} & D_{\frac{N}{2}} \\ I_{\frac{N}{2}} & -D_{\frac{N}{2}} \end{bmatrix}\begin{bmatrix} \mathcal{F}_{\frac{N}{2}} & \\ & \mathcal{F}_{\frac{N}{2}} \end{bmatrix}\begin{bmatrix} \text{even-odd} \\ \text{permutation} \end{bmatrix}.$$

Proof. We will prove by induction. The $N = 1$ is trivial.

Consider the $N = 2$ case:

$$\begin{aligned}
\begin{bmatrix} 1 & 1 \\ 1 & \omega_2 \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & \\ & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}
\end{aligned}$$

The $N = 4$ more clearly demonstrates the logic of the proof:

$$
\begin{aligned}
&\begin{bmatrix} I_2 & D_2 \\ I_2 & -D_2 \end{bmatrix} \begin{bmatrix} \mathcal{F}_2 & \\ & \mathcal{F}_2 \end{bmatrix} \begin{bmatrix} \text{even-odd} \\ \text{permutation} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \omega_4 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -\omega_4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & \omega_2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & \omega_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \omega_4 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -\omega_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & \omega_2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & \omega_2 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega_4 & -1 & -\omega_4 \\ 1 & -1 & 1 & -1 \\ 1 & -\omega_4 & -1 & \omega_4 \end{bmatrix} \\
&= \mathcal{F}_4
\end{aligned}
$$

Then consider $N = 2^k$,

$$
\begin{aligned}
\begin{bmatrix} I_{2^k} & D_{2^k} \\ I_{2^k} & -D_{2^k} \end{bmatrix} \begin{bmatrix} \mathcal{F}_{2^{k-1}} & \\ & \mathcal{F}_{2^{k-1}} \end{bmatrix} \begin{bmatrix} \vec{v}_{\text{even}} \\ \vec{v}_{\text{odd}} \end{bmatrix} &= \begin{bmatrix} I_{2^k} & D_{2^k} \\ I_{2^k} & -D_{2^k} \end{bmatrix} \begin{bmatrix} \mathcal{F}_{2^{k-1}}\vec{v}_{\text{even}} \\ \mathcal{F}_{2^{k-1}}\vec{v}_{\text{odd}} \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{F}_{2^{k-1}}\vec{v}_{\text{even}} + D_{2^{k-1}}\mathcal{F}_{2^{k-1}}\vec{v}_{\text{odd}} \\ \mathcal{F}_{2^{k-1}}\vec{v}_{\text{even}} - D_{2^{k-1}}\mathcal{F}_{2^{k-1}}\vec{v}_{\text{odd}} \end{bmatrix}
\end{aligned}
$$

Then using the Danielson-Lanczos Lemma,

$$
\begin{bmatrix} \mathcal{U}_{2^{k-1}}\vec{v}_{\text{even}} + D_{2^{k-1}}\mathcal{U}_{2^{k-1}}\vec{v}_{\text{odd}} \\ \mathcal{U}_{2^{k-1}}\vec{v}_{\text{even}} - D_{2^{k-1}}\mathcal{U}_{2^{k-1}}\vec{v}_{\text{odd}} \end{bmatrix} = \mathcal{U}_{2^k}\vec{v}
$$

□

This factorization my be applied recursively $\log_2 N = k$ times until $[\hat{\vec{v}}]_i$ is computed entirely using scalar multiplication. Doing so requires computing N roots of unity.

Then computing each $[\hat{\vec{v}}]_i$ is order $\mathcal{O}(\log N)$. $\hat{\vec{v}}$ is $N-$dimensional, so computing $\hat{\vec{v}}$ is $\mathcal{O}(N \log N + N) = \mathcal{O}(N \log N)$.

6 Sources

[1] Halmos, P. *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, Masfield Centre, CT: Martino Publishing, 2013. Chp. 1.

[2] Choudhary, B. *The Elements of Complex Analysis: Second Edition*, New Delhi: New Age International Publishers, 2003. Chp. 2.

[3] Melrose, R. *Functional Analysis: Lecture notes for 18.102*. MIT, 2013. Chp. 3

[4] Hernández, E. and Weiss, G. *A First course on Wavelets*, Boca Raton, FL: CRC Press, 1996. Chps. Intro. and 1.

[5] "Complex Fourier Series" May 18, 2020. https://eng.libretexts.org/@go/page/1614.

[6] Weisstein, Eric W. "Discrete Fourier Transform" From MathWorld–A Wolfram Web Resource. https://mathworld.wolfram.com/DiscreteFourierTransform.html

[7] Danielson, G. C., and C. Lanczos, "Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids" *J. Franklin Inst.* (1942).

[8] Weisstein, Eric W. "Fourier Matrix." From MathWorld–A Wolfram Web Resource. https://mathworld.wolfram.com/FourierMatrix.html