The DFT and Digital Filtering

Daksh Aggarwal October 20, 2020

1 Introduction

In this short article, we wish to introduce the Discrete Fourier Transform (DFT) and explore digital filtering as one of its archetypal applications. The DFT has a rich history, beautiful underlying theory, and numerous applications. As a result, there is no canonical way to motivate the DFT. For instance, the DFT may be viewed as a discrete approximation to the Fourier transform. Or it may be seen as an approximation to coefficients in a Fourier series. We will motivate the DFT through one of its most basic applications: modelling a discrete set of periodic data. Our hope is to convey the simplicity of the DFT while illustrating how it can be utilised for sophisticated applications through the example of digital filtering.

The DFT depends critically upon nice properties of the Nth roots of unity. So, let us begin with a review of these properties.

1.1 The Nth Roots of Unity

Let $N \geq 1$ be a natural number. Recall that, if we let $\omega_N = e^{2\pi i/N} = \cos(\frac{2\pi}{N}) + i\sin(\frac{2\pi}{N})$, then the N complex numbers in $\mathcal{U}_N = \{\omega_N^0, \omega_N^1, \omega_N^2, \ldots, \omega_N^{N-1}\}$ are all distinct roots of the equation $z^N - 1 = 0$. These numbers are known as the *Nth roots of unity*. The Nth roots of unity have a special underlying structure: they form a multiplicative group of order N.

Lemma 1. *Let $N \in \mathbb{N}^+$. Then the Nth roots of unity, $\mathcal{U}_N = \{\omega_N^0, \omega_N^1, \ldots, \omega_N^{N-1}\}$, form a multiplicative group of order N.*

Proof. First, we see that $\omega_N^0 = 1$ serves as the multiplicative identity. Also, $\mathcal{U}_N$ inherits associativity of multiplication from $\mathbb{C}$ because $\mathcal{U}_N \subset \mathbb{C}$. Next, to verify closure, take any two Nth roots $\omega_N^m, \omega_N^n \in \mathcal{U}_N$. By the Division algorithm, we can find nonnegative integers q and $r \leq N-1$ such that

$$m + n = Nq + r.$$

Therefore,

$$\begin{aligned}\omega_N^m \cdot \omega_N^n &= \omega_N^{m+n}\\ &= \omega_N^{Nq+r}\\ &= (\omega_N^N)^q \cdot \omega_N^r\\ &= \omega_N^r \qquad\qquad [\omega_N^N = 1].\end{aligned}$$

Since r was chosen to be less than N, we see that $\omega_N^m \cdot \omega_N^n = \omega_N^r \in \mathcal{U}_N$, and so $\mathcal{U}_N$ is closed under multiplication. Finally, to verify existence of inverses, take any element $\omega_N^k \in \mathcal{U}_N$. Then, note that

$$\omega_N^k \cdot \omega_N^{N-k} = \omega_N^N = 1.$$

As $0 \le k \le N-1$, we have $1 \le N-k \le N$. Therefore, since $\omega_N^N = 1 = \omega_N^0 \in \mathcal{U}_N$, we see that $\omega_N^{N-k} \in \mathcal{U}_N$, showing that $\mathcal{U}_N$ is closed under inverses. This verifies the group structure of $\mathcal{U}_N$. $\square$

The group structure distinguishes the roots of unity from the remaining complex numbers since it can be shown that the Nth roots of unity, for each N, are the only finite multiplicative groups in $\mathbb{C}$. Indeed, the group structure of $\mathcal{U}_N$ may be viewed as a statement about its periodicity, and already begins to indicate why the roots of unity are essential to the DFT.

Next, observe that since

$$z^N - 1 = (z-1)(z^{N-1} + z^{N-2} + \cdots + z^0),$$

and $\omega_N \neq 1$ is a root of $z^N - 1$, we must have

$$\omega_N^{N-1} + \omega_N^{N-2} + \cdots + \omega_N^0 = 0.$$

Thus, the sum of the Nth roots of unity is zero. What about the sum of their squares? Cubes? It is a beautiful fact that they too are all zero (assuming $N > 3$), as we now show. We note that our proof is not the most efficient but it illustrates smaller properties along the way that will be useful when we study the DFT. The proof also gives a flavour of how the roots of unity play an important role in number theory and algebra, indicative of the utility of Fourier transforms in those areas.

Theorem 2. *Let $N \in \mathbb{N}^+$. Then, for $0 \le j \le N-1$,*

$$\sum_{k=0}^{N-1} \omega_N^{kj} = \begin{cases} 0 & j \neq 0 \\ N & j = 0 \end{cases}.$$

Proof. Observe that the result trivially holds for $j = 0$, and so we will be concerned with $1 \le j \le N-1$. We will proceed by induction on N. Note that the base case for $N = 1$ also holds. So, suppose the result holds for all positive integers less than some $N \in \mathbb{N}_{\ge 2}$.

First, suppose first that $1 \le j \le N-1$ is such that $\omega_N^{mj} \neq \omega_N^{nj}$ for any $0 \le m < n \le N-1$. This is equivalent to saying the function $f_j : \mathcal{U}_N \to \mathcal{U}_N$ defined by $f_j(a) = a^j$ is injective (f_j in indeed a function into $\mathcal{U}_n$ because $\mathcal{U}_n$ is closed under multiplication). Since the set $\mathcal{U}_N$ is finite, f_j must be a bijection – it permutes $\mathcal{U}_N$. Therefore, in this case

$$\sum_{k=0}^{N-1} \omega_N^{kj} = \sum_{k=0}^{N-1} f_j(\omega_N^k) = \sum_{k=0}^{N-1} \omega_N^k = 0.$$

So, we need to check that the result holds when $1 \le j \le N-1$ is such that f_j is not injective, i.e., when there exist $0 \le m < n \le N-1$ such that $\omega_N^{mj} = \omega_N^{nj}$, or $\omega_N^{(n-m)j} = 1$. By the well-ordering principle, we choose m, n such that $D = n - m$ is positive and minimal among all pairs (m, n) such that $f_j(m) = f_j(n)$. Thus, our choice ensures that D is the smallest integer in $\{1, 2, \dots, N-1\}$ such that $\omega_N^{Dj} = 1$. In other words, D is the order of the element ω_N^j in the group $\mathcal{U}_N$ of order N (Lemma (1)). By Lagrange's theorem[1], we must have that $D \mid N$. Fix $Q \in \mathbb{N}$ such that $N = DQ$. Then, using the Division algorithm, we get

$$\begin{aligned} \sum_{k=0}^{N-1} \omega_N^{kj} &= \sum_{q=0}^{Q-1} \sum_{r=0}^{D-1} \omega_N^{(Dq+r)j} && \text{[Reindexing sum using } k = Dq + r\text{]} \\ &= \sum_{q=0}^{Q-1} \sum_{r=0}^{D-1} \omega_N^{rj} && \text{[Since } \omega_N^{Dj} = 1\text{]} \end{aligned}$$

[1]Recall Lagrange's theorem states that in a finite group, the order of a subgroup must divide the order of the group. Considering the cyclic subgroup generated by a certain element, we conclude that the order of each element must divide the order of the group.

$$= Q \sum_{r=0}^{D-1} \omega_N^{rj}. \tag{1}$$

However, the relation $\omega_N^{Dj} = 1$ can also be viewed as saying that multiplying ω_N Dj times yields 1. Since N is the order of ω_N, we must have $N \mid Dj$. So, fix $L \in \mathbb{N}^+$ such that $NL = Dj$; observe that since $Dj < DN$, we must have $L < D$. Substituting $j = NL/D$ into (1), we obtain

$$\begin{aligned} \sum_{k=0}^{N-1} \omega_N^{kj} &= Q \sum_{r=0}^{D-1} \omega_N^{rNL/D} \\ &= Q \sum_{r=0}^{D-1} \omega_D^{rL}, \end{aligned}$$

where in the last step we used the observation that $\omega_N^{N/D} = (e^{2\pi i/N})^{N/D} = e^{2\pi i/D} = \omega_D$. Since $1 \leq L \leq D-1$, by the inductive hypothesis, we conclude that $\sum_{r=0}^{D-1} \omega_D^{rL} = 0$, and thus the result follows. □

Having gained familiarity with the Nth roots of unity, we now move to understanding the DFT.

2 Modelling Data using Trigonometric Polynomials[2]

Suppose we want to model a set of periodic data $S = \{(x_k, y_k)\}_{0 \leq k \leq N-1}$ where the N points x_k are evenly spaced in an interval $[a, b]$. So, if $L = b - a$, then $x_k = k \cdot \Delta x$ where $\Delta x = \frac{L}{N}$. Such a set of data might be obtained, for instance, by measuring a certain quantity at uniform time intervals. Now, we have at least two basic ways to model the data S: fitting it to a least squares model or interpolating it. We will first try the least squares method, introduce DFT along the way, and then turn to interpolation to reveal a beautiful property of the DFT. Of course, we have yet to choose a model to which to fit S. Let us try fitting it to the trigonometric polynomial model

$$T(x) = \sum_{k=0}^{N-1} \alpha_k e^{\frac{2\pi i k}{L} x},$$

[2]Our treatment of DFT is inspired by Problems 26 & 28 in [1, Ch. 2].

for some unknowns $\alpha_k \in \mathbb{C}, k = 0, \dots, N-1$. In the least squares method, we seek to minimize the sum of squared errors

$$E = \sum_{k=0}^{N-1} |y_k - T(x_k)|^2 .$$

Note that E is a function of the coefficients $\alpha_0, \dots, \alpha_{N-1}$, and so for E to be minimized, the partial derivative of E with respect to each α_k needs to be zero

$$\frac{\partial E(\alpha_0, \dots, \alpha_{N-1})}{\partial \alpha_k} = 0, \quad \text{for } k = 0, \dots, N-1.$$

Since $|y_k - T(x_k)|^2 = (y_k - T(x_k))\overline{(y_k - T(x_k))}$, for each $k = 0, \dots, N-1$, we obtain

$$\begin{aligned} \frac{\partial E(\alpha_0, \dots, \alpha_{N-1})}{\partial \alpha_k} &= \sum_{j=0}^{N-1} \left(-\frac{\partial T(x_j)}{\partial \alpha_k} \overline{(y_j - T(x_j))} - (y_j - T(x_j)) \frac{\partial \overline{T(x_j)}}{\partial \alpha_k} \right) \\ &= \sum_{j=0}^{N-1} \left(-e^{\frac{2\pi i k}{L} x_j} \overline{(y_j - T(x_j))} - 0 \right), \end{aligned} \tag{2}$$

where we have applied $\frac{\partial \overline{T(x_j)}}{\partial \alpha_k} = 0$ since $\frac{\partial \overline{\alpha_k}}{\partial \alpha_k} = 0$. Using that $\frac{x_j}{L} = \frac{j}{N}$ and setting (2) equal to zero, we obtain

$$\begin{aligned} 0 &= \sum_{j=0}^{N-1} e^{\frac{2\pi i k}{N} j} \overline{(y_j - T(x_j))} \\ &= \sum_{j=0}^{N-1} \overline{\overline{e^{\frac{2\pi i k}{N} j}} (y_j - T(x_j))} \\ &= \overline{\sum_{j=0}^{N-1} e^{-\frac{2\pi i k}{N} j} (y_j - T(x_j))}, \end{aligned}$$

where we used that $\overline{e^{i\theta}} = \overline{\cos\theta + i \sin\theta} = \cos\theta - i \sin\theta = \cos(-\theta) + i \sin(-\theta) = e^{-i\theta}$.

So we conclude that

$$\sum_{j=0}^{N-1} e^{-\frac{2\pi i k}{N} j} (y_j - T(x_j)) = 0,$$

for $k = 0, \dots, N-1$.

Substituting $\omega_N = e^{\frac{2\pi i}{N}}$ and $T(x_j) = \sum_{l=0}^{N-1} \alpha_l \omega_N^{lj}$, we get

$$\begin{aligned}
\sum_{j=0}^{N-1} \omega_N^{-kj} y_j &= \sum_{j=0}^{N-1} \omega_N^{-kj} \sum_{l=0}^{N-1} \alpha_l \omega_N^{lj} \\
&= \sum_{l=0}^{N-1} \alpha_l \sum_{j=0}^{N-1} \omega_N^{j(l-k)} \quad \text{[Rearranging sums]} \\
&= \alpha_k \cdot N,
\end{aligned}$$

with the pleasant simplification in the last step being permitted by Theorem 2: the inner sum is nonzero (and equal to N) only when $l = k$. All in all, we obtain

$$\alpha_k = \frac{1}{N} \sum_{j=0}^{N-1} \omega_N^{-kj} y_j, \quad \text{for } k = 0, \dots, N-1. \tag{3}$$

Thus, this relation gives a possible set of coefficients α_k that might minimize the least squares error E. In fact, given a sequence of points $y_0, \dots, y_{N-1}$, (3) also exactly describes how to obtain its DFT!

Definition 3 (Discrete Fourier Transform)**.** The *Discrete Fourier Transform (DFT)* of a sequence $y = \{y_0, \dots, y_{N-1}\} \subset \mathbb{C}$ is the sequence $Y = \{Y_0, \dots, Y_{N-1}\} \subset \mathbb{C}$ defined by

$$Y_k = \frac{1}{N} \sum_{j=0}^{N-1} \omega_N^{-kj} y_j.$$

We denote the DFT of y by $\mathcal{D}(y) = Y$.

Going back to our least squares model, we have to still check that the coefficients α_k given by (3) indeed minimize E. Let us momentarily put this on hold and attempt to interpolate the data $S = \{(x_k, y_k)\}$ with the trigonometric polynomial model $T(x)$. This amounts to finding coefficients $\widetilde{\alpha}_0, \widetilde{\alpha}_1, \dots, \widetilde{\alpha}_{N-1} \in \mathbb{C}$ such that

$$y_k = T(x_k) = \sum_{j=0}^{N-1} \widetilde{\alpha}_j \omega_N^{jk},$$

for each $k = 0, \dots, N-1$.

In other words,

$$\underbrace{\begin{bmatrix} \omega_N^0 & \omega_N^0 & \omega_N^0 & \dots & \omega_N^0 \\ \omega_N^0 & \omega_N^1 & \omega_N^2 & \dots & \omega_N^{(N-1)} \\ \omega_N^0 & \omega_N^2 & \omega_N^4 & \dots & \omega_N^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega_N^0 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \dots & \omega_N^{(N-1)^2} \end{bmatrix}}_{T} \begin{bmatrix} \widetilde{\alpha}_0 \\ \widetilde{\alpha}_1 \\ \widetilde{\alpha}_2 \\ \vdots \\ \widetilde{\alpha}_{N-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix}. \tag{4}$$

Thus, the coefficients $\widetilde{\alpha}_k$ will exist if and only if the matrix T is invertible. Further, *if* the coefficients $\widetilde{\alpha}_k$ do exist, then they will also automatically minimize the error E (namely, make it zero). Now, note that (3) can also be expressed in matrix form as

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{N-1} \end{bmatrix} = \frac{1}{N} \underbrace{\begin{bmatrix} \omega_N^0 & \omega_N^0 & \omega_N^0 & \dots & \omega_N^0 \\ \omega_N^0 & \omega_N^{-1} & \omega_N^{-2} & \dots & \omega_N^{-(N-1)} \\ \omega_N^0 & \omega_N^{-2} & \omega_N^{-4} & \dots & \omega_N^{-2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega_N^0 & \omega_N^{-N-1} & \omega_N^{-2(N-1)} & \dots & \omega_N^{-(N-1)^2} \end{bmatrix}}_{\mathcal{D}} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix}. \tag{5}$$

Engaging in some wishful thinking, assume the coefficients α_k given by (5) do minimize E (we have not verified this). So, if the interpolating coefficients $\widetilde{\alpha}_k$ exist, they must be given by (5). Notice that this speculation is true if and only if $T^{-1} = \mathcal{D}$, which we now verify.

Lemma 4. *Let T and $\mathcal{D}$ be defined as above. Then $T^{-1} = \mathcal{D}$.*

Proof. We check that $(T\mathcal{D})_{m,n} = \delta_{m,n}$ for $0 \le m, n \le N-1$.[3] First, suppose $m = n$. Then,

$$(T\mathcal{D})_{m,m} = \sum_{k=0}^{N-1} \omega_N^{mk} \cdot \frac{\omega_N^{-mk}}{N} = \frac{1}{N} \sum_{k=0}^{N-1} \omega_N^0 = 1.$$

[3]As usual, δ is the Kronecker delta: $\delta_{m,n} = 1$ if $m = n$ and 0 otherwise.

For $m \neq n$, we get

$$(T\mathcal{D})_{m,n} = \frac{1}{N}\sum_{k=0}^{N-1} \omega_N^{mk} \cdot \omega_N^{-nk} = \frac{1}{N}\sum_{k=0}^{N-1} \omega_N^{k(m-n)} = 0,$$

by an application of our handy Theorem 2. Thus, $T^{-1} = \mathcal{D}$. □

Therefore, the interpolation coefficients $\widetilde{\alpha}_k$ are indeed given by (5), and the points $\{(x_n, y_n)\}$ can be interpolated using $T(x)$. This is one of the most basic applications of the DFT: we can find the interpolating coefficients $\widetilde{\alpha}_k$ for the trigonometric polynomial $T(x)$ through a set of evenly-spaced points $\{(x_n, y_n)\}$ by simply applying the DFT operator $\mathcal{D}$ to the sequence $\{y_0, \ldots, y_{N-1}\}$. Moreover, Lemma 4 implies that $\mathcal{D}$ has an inverse T^{-1}, which can be applied to the coefficients $\{\widetilde{\alpha}_0, \ldots, \widetilde{\alpha}_{N-1}\}$ to recover the data y_n.

x_n	y_n	k	$Y_k = \mathcal{D}(y)_k$
0	9.96518	0	$5.28565 + 0i$
1	6.91185	1	$1.49404 - 0.03851i$
2	4.53458	2	$0.48084 - 0.00997i$
3	3.50964	3	$0.30528 + 0.00721i$
4	2.76787	4	$0.11917 + 0i$
5	3.46101	5	$0.30528 - 0.00721i$
6	4.35168	6	$0.48084 + 0.00997i$
7	6.78341	7	$1.49404 + 0.03851i$

Table 1: An evenly spaced sequence in [0,8] along with its DFT

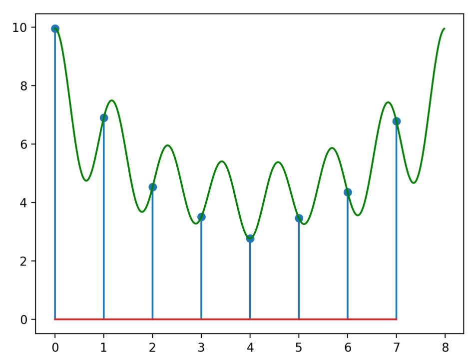
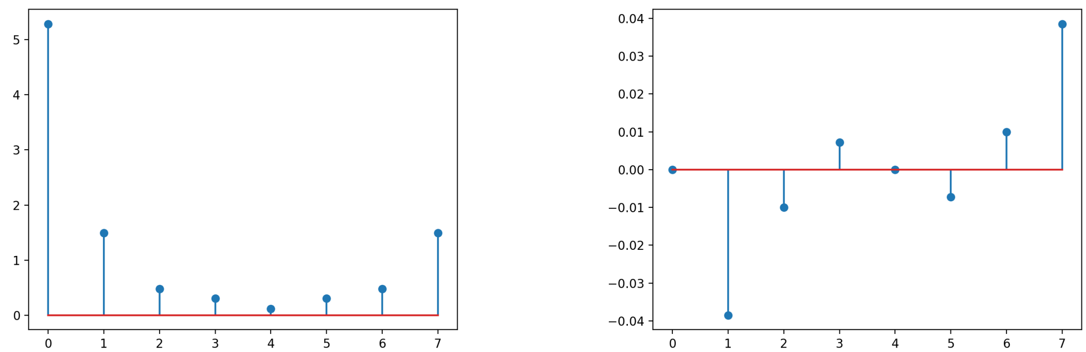


Figure 1: An 8-point signal along with its (real) trigonometric interpolation

To concretely illustrate the DFT, we have computed in Table 1 the DFT of an 8-point sequence sampled at evenly spaced points in the interval $[0, 8]$. We have plotted the points (x_k, y_k) along



(a) Real part of DFT (b) Imaginary part of DFT

Figure 2: The DFT of an 8-point sequence

with the real part of its trigonometric polynomial interpolation in Figure 1, which captures the periodic aspect of the data. We have also plotted the real and imaginary parts of the DFT in Figure 2, which shows some striking aspects of the DFT. Observe that the real part of the DFT is even: for $1 \leq k \leq 7$, $\mathrm{Re}(Y_k) = \mathrm{Re}(Y_{8-k})$. Similarly, the imaginary part is odd: for $1 \leq k \leq 7$, $\mathrm{Im}(Y_k) = -\mathrm{Im}(Y_{8-k})$. These are both general properties of the DFT of any *real* sequence y, as can be readily checked by using the fact that $\omega_N^{-(N-k)j} = \omega_N^{kj} = \overline{\omega_N^{-kj}}$ in Definition 3. Next, notice the relative size of the real parts: the real parts of Y_0 and Y_1 seem to dominate those of $Y_2, Y_3, \ldots, Y_6$. A similar pattern holds for the imaginary parts. What might we infer from this observation? We must keep in mind the basic interpretation of the DFT coefficients. The coefficient Y_k represents the weight of the complex exponential $e^{2\pi ikn/N}$ in the decomposition of y_n as

$$y_n = T(x_n) = \sum_{k=0}^{N-1} Y_k e^{\frac{2\pi ikn}{N}}.$$

In other words, since $e^{2\pi ikn/N} = \cos(2\pi \frac{k}{N} n) + i\sin(2\pi \frac{k}{N} n)$, Y_k can be seen as the contribution of the frequency $f_k = \frac{k}{N}$ to y_n. Thus, we can interpret from the small values of $Y_2, \ldots, Y_6$, that the corresponding higher frequencies have less contribution to the periodic data. Now, suppose we had obtained this data by measuring a certain periodic system. What if we suspected that the small contribution of higher frequencies in the data did not reflect the true nature of the system but was rather an error introduced by imperfections in the measurement process? If this is the case, how might we rid the data of these spurious high frequencies? The DFT provides us with a simple way:

we could set the components of the DFT that we think must be zero (in this case, $Y_2, \dots, Y_6$) and then apply the inverse DFT to recover a "cleaned" version of the data! This is the underlying idea of digital filtering, which we now turn to explore.

3 Digital Filtering

Since the canonical application of digital filtering is to filter sound, it might be useful to think of the sequence $y = \{y_k\}_{0 \le k \le N-1}$ as $N-1$ amplitudes of a continuous piece of sound (such as music) sampled at evenly-spaced intervals of time. It will also be convenient to extend y into an infinite periodic sequence. In other words, let us assume that we are measuring the amplitudes of a periodic piece of music such that $y_n = y_{n \bmod N}$ for any integer n. So, for example, $y_{N+1} = y_1$, $y_{-1} = y_{N-1}$, and $y_{100N} = y_0$.

A *filter* is any function g we apply to y to obtain a new sequence of amplitudes $y' = \{y'_k\}_{0 \le k \le N-1}$. Usually, as the name suggests, the filter g modifies y. The modification, which is determined by the choice of the filter, might represent noise reduction (as in our example above), the enhancement of certain frequencies, or any of the variety of subtle alterations, for instance, a musician might require.

For a simple example of a filter, consider the function g which acts on a sequence y as

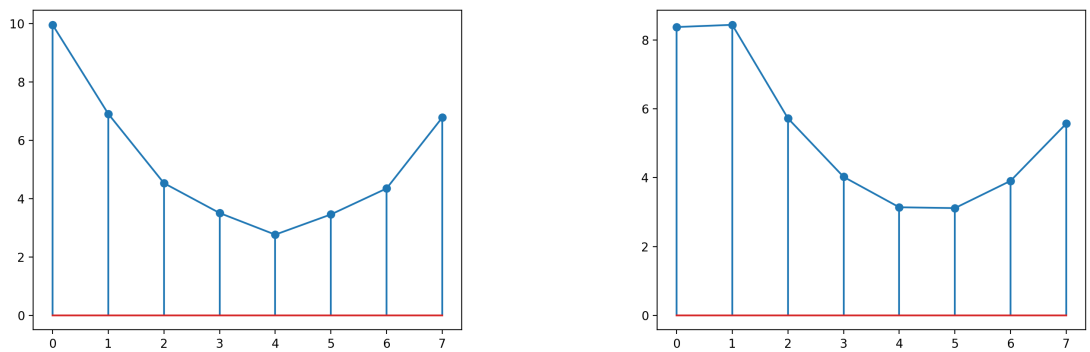
$$y'_n = g(y_n) = \frac{1}{2} y_{n-1} + \frac{1}{2} y_n, \quad \text{for each } n \in \mathbb{Z}. \tag{6}$$

So, this filter assigns the average of y_{n-1} and y_n to y'_n. For example, using this filter on y from Table 1, we obtain the modified sequence in Table 2. Compare the signals y and y' plotted in Figure 3 – arguably, the filtered signal seems to be slightly smoother. Also, compare the real part of their DFTs: the contribution of higher frequencies has diminished, with Y'_4 being zero. Intuitively, by averaging adjacent amplitudes, we have amplified the dominant frequencies while reducing the effect of aberrant ones. Of course, the imaginary parts of higher frequencies have increased, but we can only expect so much from such a simple filter!

A powerful way to represent the action of many filters is provided by the operation of convolution.

y_n	$Y_k = \mathcal{D}(y)_k$	y'_n	$Y'_k = \mathcal{D}(y')_k$
9.96518	$5.28565 + 0i$	8.37430	$5.28565 + 0i$
6.91185	$1.49404 - 0.03851i$	8.43852	$1.26163 - 0.56110i$
4.53458	$0.48084 - 0.00997i$	5.72322	$0.23544 - 0.24541i$
3.50964	$0.30528 + 0.00721i$	4.02211	$0.04726 - 0.10688i$
2.76787	$0.11917 + 0i$	3.13876	0
3.46101	$0.30528 - 0.00721i$	3.11445	$0.04726 + 0.10688i$
4.35168	$0.48084 + 0.00997i$	3.90635	$0.23544 + 0.24541i$
6.78341	$1.49404 + 0.03851i$	5.56755	$1.26163 + 0.56110i$

Table 2: Filtered sequence along with its DFT



(a) Original signal

(b) Filtered signal

Figure 3: Illustrating the action of the filter in (6)

Definition 5. (Discrete Convolution) Let $N \in \mathbb{N}$ and let $y = \{y_k\}_{k\in\mathbb{Z}}$ and $g = \{g_k\}_{k\in\mathbb{Z}}$ be two infinite complex sequences with period N. Then the *convolution* of y and g produces the infinite sequence $y' = \{y'_k\}_{k\in\mathbb{Z}}$ defined by

$$y'_n = \sum_{k=0}^{N-1} y_k g_{n-k}, \quad \text{for each } k \in \mathbb{Z}.$$

We denote the convolution of y and g by $y * g$.

How do we represent the action of the filter in (6) as a convolution? We claim that if we define g to be the infinite 8-periodic sequence given by letting

$$\{g_k\}_{0\le k<8} = \{\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0\},$$

then $y' = y * g$ will give us the same sequence as in (6). For example,

$$y'_0 = \sum_{k=0}^{7} y_k g_{-k} = \sum_{k=0}^{7} y_k g_{8-k} = \sum_{k=0}^{7} y_{8-k} g_k = y_8 \cdot \frac{1}{2} + y_7 \cdot \frac{1}{2} = y_0 \cdot \frac{1}{2} + y_{-1} \cdot \frac{1}{2},$$

which matches (6). The reader is encouraged to check why this holds for other n.

Let us again consider Table 2. While the real parts of the higher frequencies decreased, we had wanted to completely rid our data of the frequencies corresponding to $k = 2, 3$, and 4, which would automatically also eliminate those corresponding to $k = 5$ and 6 by the evenness and oddness of the real and imaginary parts respectively. Earlier, we had suggested the seemingly simple-minded approach of deleting these unwanted entries from the DFT of y and then inverting the resulting DFT to obtain a clean version of the data. It is natural to wonder if there actually exists a filter that achieves this effect. More generally, is modifying the DFT of a sequence as we like to obtain a modified sequence a legitimate act of filtering? Put differently, can we always find a filter that achieves a desired effect on the frequencies of its input? We are thus lacking an understanding of how filtering a signal affects its DFT. The following theorem fills in this gap and shows the elegance of convolutions.

Theorem 6 (Discrete Convolution theorem). *Let $N \in \mathbb{N}^+$ and let $y = \{y_k\}_{k\in\mathbb{Z}}$ and $g = \{g_k\}_{k\in\mathbb{Z}}$ be two infinite complex sequences with period N. Let $Y = \mathcal{D}(y)$ and $G = \mathcal{D}(g)$. Then*

$$\mathcal{D}(y * g) = NYG,$$

where, as usual, YG denotes the pointwise product of the sequences Y and G.

Proof. Let $y' = y * g$. The main strategy is to take advantage of the fact that $y_k = \mathcal{D}^{-1}(Y)_k$ (we know $\mathcal{D}^{-1}$ exists by Lemma 4) and similarly for g. We have for each $0 \le n \le N-1$,

$$\begin{aligned} y'_n &= \sum_{k=0}^{N-1} y_k g_{n-k} \\ &= \sum_{k=0}^{N-1} \mathcal{D}^{-1}(Y)_k \mathcal{D}^{-1}(G)_{n-k} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=0}^{N-1} \left(\sum_{l=0}^{N-1} \omega_N^{kl} Y_l \right) \cdot \left(\sum_{j=0}^{N-1} \omega_N^{(n-k)j} G_j \right) \quad \text{[Lemma 4]} \\
&= \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{j=0}^{N-1} \left(\omega_N^{kl+nj-kj} Y_l G_j \right) \\
&= \sum_{j=0}^{N-1} \omega_N^{nj} G_j \left(\sum_{l=0}^{N-1} Y_l \cdot \sum_{k=0}^{N-1} \omega_N^{k(l-j)} \right) \quad \text{[Switching first and last sums]} \\
&= \sum_{j=0}^{N-1} \omega_N^{nj} G_j (Y_j \cdot N) \quad \text{[By Theorem 2]} \\
&= \mathcal{D}^{-1}(NYG)_n.
\end{aligned}
$$

Thus, $y' = \mathcal{D}^{-1}(NYG)$, and so applying $\mathcal{D}$, we obtain $\mathcal{D}(y') = NYG$. □

Thus, the DFT of a filtered sequence is equal to the pointwise product of the DFTs of the original sequence and the filter (with a scaling factor of N). We could not expect this relationship to be any simpler! To illustrate the immense utility of Theorem 6, we use it to derive a filter that eliminates frequencies in a signal beyond a certain level. Such a filter is known as a *low-pass filter*.

Specifically, with the same notation as above, we require a filter g that satisfies

$$
\mathcal{D}(y * g)_k = \begin{cases} Y_k & 0 \le k \le k_c \text{ or } N - k_c \le k \le N-1 \\ 0 & k_c < k < N - k_c \end{cases}.
$$

The parameter k_c is known as the *cutoff frequency index* and specifies that g must filter out the frequencies represented by $k_c < k < N - k_c$. In our example, we want to remove frequencies represented by $1 < k_c < 7$ and so $k_c = 1$. By Theorem 6, we must then have

$$
G_k = \begin{cases} \frac{1}{N} & 0 \le k \le k_c \text{ or } N - k_c \le k \le N-1 \\ 0 & k_c < k < N - k_c \end{cases}.
$$

Therefore, since $g = \mathcal{D}^{-1}(G)$, we obtain

$$
g_n = \sum_{k=0}^{N-1} \omega_N^{nk} G_k
$$

$$
\begin{aligned}
&= \sum_{k=0}^{k_c} \omega_N^{nk} \cdot \frac{1}{N} + \sum_{k=k_c+1}^{N-k_c-1} \omega_N^{nk} \cdot 0 + \sum_{k=N-k_c}^{N-1} \omega_N^{nk} \cdot \frac{1}{N} \\
&= \frac{1}{N} \left(\sum_{k=0}^{k_c} \omega_N^{nk} + \sum_{k=0}^{k_c-1} \omega_N^{n(N-k_c+k)} \right) \\
&= \frac{1}{N} \left(\omega_N^{nk_c} + \sum_{k=0}^{k_c-1} \omega_N^{nk} + \sum_{k=0}^{k_c-1} \omega_N^{-nk_c+nk} \right) \qquad [\omega_N^{nN} = 1] \\
&= \frac{1}{N} \left(\omega_N^{nk_c} + (1 + \omega_N^{-nk_c}) \sum_{k=0}^{k_c-1} \omega_N^{nk} \right) \\
&= \frac{1}{N} \left(\omega_N^{nk_c} + (1 + \omega_N^{-nk_c}) \frac{1 - \omega_N^{nk_c}}{1 - \omega_N^{n}} \right) \qquad [\text{Geometric series, assuming } n \neq 0] \\
&= \frac{1}{N} \cdot \frac{\omega_N^{-nk_c} - \omega_N^{n(k_c+1)}}{1 - \omega_N^{n}}.
\end{aligned} \tag{7}
$$

For $n = 0$, we directly obtain $g_0 = \frac{2k_c+1}{N}$. Inspecting the formula (7) we have obtained for a low-pass filter, we see that it would have been difficult to guess without Theorem 6. For our example, putting $k_c = 1$, we obtain

$$
g_n = \frac{1}{8} \cdot \frac{\omega_8^{-n} - \omega_8^{2n}}{1 - \omega_8^{n}}, \text{ for } 1 \leq n \leq 7,
$$

and $g_0 = 3/8$. The result of applying this filter to the sequence of Table 1 yields the sequence in Table 3.

y_n	$Y_k = \mathcal{D}(y)_k$	y'_n	$Y'_k = \mathcal{D}(y')_k$
9.96518	$5.28565 + 0i$	8.27374	$5.28565 + 0i$
6.91185	$1.49404 - 0.03851i$	7.45302	$1.49404 - 0.03851i$
4.53458	$0.48084 - 0.00997i$	5.36268	0
3.50964	$0.30528 + 0.00721i$	3.22723	0
2.76787	$0.11917 + 0i$	2.29757	0
3.46101	$0.30528 - 0.00721i$	3.11830	0
4.35168	$0.48084 + 0.00997i$	5.20863	0
6.78341	$1.49404 + 0.03851i$	7.34409	$1.49404 + 0.03851i$

Table 3: Result of filtering with a low-pass filter

The power of Theorem 6 also in lies the fact that we could have applied the low-pass filter without explicitly computing g and convolving it with y. Instead, having determined G, we could have computed $y * g$ as $\mathcal{D}^{-1}(NYG)$! Indeed, in practice, filters are usually applied in this way. The reason for this is that a DFT can be efficiently computed by a Fast Fourier Transform (FFT). The FFT cleverly takes advantage of repetitious multiplications in a DFT to compute it in $O(N \log N)$

operations rather than $O(N^2)$ operations as is suggested by (5); for a nice introduction to FFT, the book [1, Ch.12] is recommended. As a result, the convolution $y * g$, whose definition indicates $O(N^2)$ operations, can be computed as $\mathcal{D}^{-1}(N\mathcal{D}(y)\mathcal{D}(g))$, which requires only $O(3N \log N)$ operations.[4] So, Theorem 6 has both theoretical and practical utility and is one of the reasons why DFT has become fundamental to efficient signal processing.

Our exploration concludes here. We have barely scratched the surface of the DFT − both in its theory and applications. We have seen just the simplest examples of filters and mentioned the fascinating FFT only in passing. In the author's (rather limited) experience, the DFT appears to have surprising connections to various parts of mathematics, and the topic seems to offer something for every kind of mathematical taste. We hope this article might inspire the reader to embark on their own exploration of DFT; [1] would be an excellent starting point.

References

[1] Briggs, William L and Henson, Van Emden. *The DFT: An Owner's Manual for the Discrete Fourier Transform*. SIAM, 1995.

[2] Sauer, Timothy. *Numerical Analysis*. Pearson, 2012.

[3] Smith III, Julius O. *Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications*. Accessed October 15, 2020. `https://ccrma.stanford.edu/~jos/mdft/mdft.html`.

[4] Notice that an inverse DFT can be computed using a DFT since $\mathcal{D}^{-1}(y) = N\overline{D}(y) = N\overline{D(\overline{y})}$.