THE DISCRETE FOURIER TRANSFORM

LARRY SUSANKA

Contents

Date: June 27, 2018.

1. Records

We create **records**, each of which consists of a sequence of N numerical measurements taken from consecutive observations of something outside the window.

Specifically, a record X is an ordered listing of numbers

$$X(0),\ X(1),\ \ldots,\ X(N-1).$$

Each of these numbers is called a **sample**. We assume the samples are taken at regular time intervals, every Δt seconds, starting at time t_0. The number Δt is called the **sample interval**.

The **sample times** t consist of the ordered list

$$t(0),\ t(1),\ \ldots,\ t(N-1) \quad \text{where} \quad t(k) = t_0 + k\Delta t.$$

So sample $X(k)$ is taken at sample time $t(k)$ for each k.

Our goal is to discern periodicities or other repetitive aspects in this record.

We will need to be precise about what we mean by "periodic" in a finite string of numbers. We must specify what *type* of periodicities we hope to find. We also will consider the inherent limitions in the types of periodicities we can distinguish—the record might conflate contributions of different frequencies, and we will need to know how and when this can happen.

The distinction between an ordered list of N numbers and a vector in $\mathbb{R}^N$ or $\mathbb{C}^N$ is primarily one of "point of view,' and we will use vector ideas in handling such lists wherever that is convenient.

The total time to gather the samples for a record is $T = N\,\Delta t$ seconds. T is called the **sample length**. The number $f_s = 1/\Delta t$ samples/second is called the **sample frequency**.

Listing these together, we have identified so far:

$$N\ \frac{\text{samples}}{\text{record}} \qquad \Delta t\ \frac{\text{seconds}}{\text{sample}} \qquad T = N\,\Delta t\ \frac{\text{seconds}}{\text{record}} \qquad f_s = \frac{1}{\Delta t}\,\frac{\text{samples}}{\text{second}}.$$

We present two final definitions, features of the record, which will be useful later. The quantity

$$\Delta f = f_{\min} = \frac{f_s}{N} = \frac{1}{N\,\Delta t} = \frac{1}{T}\ \frac{1}{\text{second}}$$

is called the **frequency increment**. It is the lowest frequency we could possibly detect in the data, corresponding to a **maximum period** of $\mathbf{P_{max}} = T = N\Delta t$ seconds. If you need to decrease this frequency you must increase the sample length.

$$f_{\text{Nyquist}} = f_{\max} = \frac{f_s}{2} = \frac{1}{2\,\Delta t}\ \frac{1}{\text{second}}$$

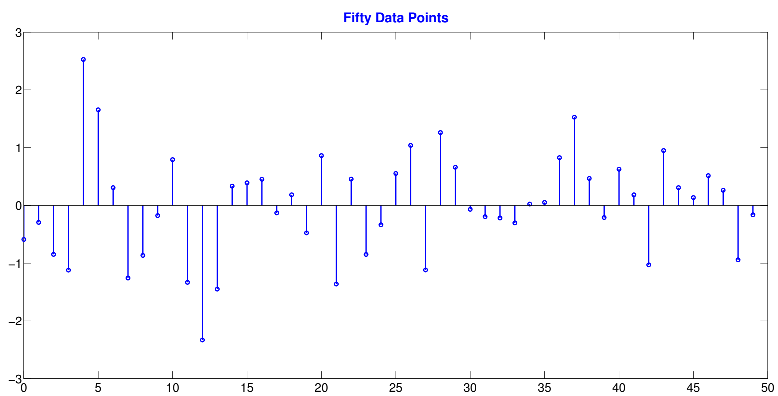
is called the **Nyquist frequency**. It is the highest frequency we will be able to detect directly in the data, corresponding to a **minimum period** of $\mathbf{P_{min}} = 2\,\Delta t$ seconds.

Any periodicities in the phenomenon which we are observing which have frequencies that are integer multiples of f_s, including f_s itself, would be lumped together and would look constant, a common number added to each of the N sample values. Understanding how other high frequencies leave a trace in the record of directly visible frequencies will be a concern.

$$\text{Directly detectable frequencies: } \frac{1}{k\Delta t} \qquad \text{and periods: } k\Delta t \qquad 2 \le k \le N.$$

2. The Setup

Our underlying assumption is that the numbers in each record are taken from combined observed values of some aspect of various simpler phenomena. The records of these simpler phenomena are assumed to be periodic, and to have a particular form which may come from the physical theory of the phenomena in question.



$$\begin{array}{ll}
\text{Phenomenon 1:} & Y_1(0),\ Y_1(1),\ \ldots,\ Y_1(N-1) \\
\text{Phenomenon 2:} & Y_2(0),\ Y_2(1),\ \ldots,\ Y_2(N-1) \\
\text{Phenomenon 3:} & Y_3(0),\ Y_3(1),\ \ldots,\ Y_3(N-1) \\
\vdots & \vdots \quad \vdots \qquad \vdots \qquad \vdots \\
\text{Add the above to get:} & X(0),\ X(1),\ \ldots,\ X(N-1)
\end{array}$$

For the discrete Fourier transform, a pristine prototypical periodic record would have sample values

$$Y(k) = a\,\cos(\beta\,t(k) + s) \quad k = 0, \ldots, N-1$$

for certain numbers a, β and s.

Other transforms may posit different "elemental" shapes, but the idea—and those parts of the discussion not dependent on specific properties of the cosine function—would be the same.

Our first goal is to determine the way in which a particular underlying periodic Y of this type would reveal itself in a record $X(0),\ X(1),\ \ldots,\ X(N-1)$. Second, we want to estimate in some optimal way the values of a, β and s when the record carries information to make that possible.

There may be noise in our measurement process so we don't expect perfect repetitions. And we don't expect to be able to detect all patterns: for instance, our record might be too short to reveal a pattern that has a very long period. We just want to do the best job we can.

Even though we usually have many records, one after another, we deal with the records one at a time (rather than as one huge record) because we hope to use our analysis to detect changes in the underlying phenomena "on-the-fly." Of course surrounding records might well contribute useful information about *our* record that would be a mistake to ignore, but in this "first-pass" analysis our aims are modest and we *will* ignore other records and see what we can do.

First, we need to examine what we mean by "periodicity" in a finite record.

In its raw, most simplistic form, a record is defined to be periodic if, when these numbers are repeated in a block, forever, as

$$\dots, X(N-2), X(N-1), X(0), X(1), \dots, X(N-1), X(0), X(1), \dots$$

we could find an integer k so that if we shift the whole doubly-infinite sequence to the right by k places[1] the sequence looks just the same: all the same numbers would be in all the same places.

An integer k of this kind is called "**a period**" of the sequence. N is automatically a period of the sequence, but that is just an artifact of our sample length and uninteresting. There might be others—there *will* be others in the basic periodic records we will find useful as building blocks.

We will denote the doubly infinite sequence we have created by $\widetilde{X}$.

$$\widetilde{X}(k) = X(j) \quad \text{where } j \text{ is the remainder of integer } k \text{ after division by } N.$$

The doubly infinite sequence of times $\widetilde{t}$ is defined for integer k by $\widetilde{t}(k) = t_0 + k\Delta t$.

We will not actually *do* anything with $\widetilde{X}$ and $\widetilde{t}$ that we are not doing with X and t. The unbounded sequences $\widetilde{X}$ and $\widetilde{t}$ simply make explicit how we are *thinking* of X and t when we look for periodic phenomena which combine to produce X.

By placing copies of the record one after (and before) another, we are encroaching on the territory of other records, and these copies may well be different from what was, or will be, actually seen in the samples of those records. The pattern we are trying to find in our particular record is the pattern we could detect *if the record were* just a chunk of an infinite sequence of this type, as if all the contributing phenomena were frozen to be just what they were during those particular N sample times, forever past and future.

Another point that bears mentioning is that the periods (other than the constant sequence) that might be directly and exactly perceived in one copy of the data, namely

$$2\Delta t,\ 3\Delta t,\ \dots,\ N\Delta t$$

[1] A negative right shift is interpreted as a left shift.

are dependent on the sample length and sample frequency, and are not likely to be *exactly* present in the underlying phenomenon, any more than any other $N-1$ random choices of wavelength between $2\Delta t$ and $N\Delta t$.

So when we get down to looking for "elemental" periodic contributions we will not confine ourselves to those periods, but instead examine a variety of periods of the form

$$\text{Frequencies: } \frac{1}{k\Delta t} \qquad \text{Periods: } k\Delta t \qquad 2 \le k \le N$$

where k is *rational* but explicitly not necessarily an *integer.* So it might take many copies of the underlying record to reveal a whole number of periods of that contribution.

3. Aliasing. What could we distinguish?

We now make a linear time-change, purely for convenience in labeling, so that

$$t_0 = 0 \quad \text{and} \quad \Delta t = 1.$$

This means that $\widetilde{t}(k) = k$ for all integers k. Also $f_s = 1$ and $T = N$.

This amounts to nothing more than measuring time in units of Δt, and we will always use these (dramatic) simplifications in formulas, but when drawing conclusions we will often phrase things as multiples of Δt and f_s, to aid translation back to the original time units, which would be present in an application.

(i) Suppose we had identified by some means two contributions to X which were of the "elemental" variety, having the same frequency but unequal shift.

Specifically, we assume that both $a_1\cos(\beta\, k + s_1)$ and $a_2\cos(\beta\, k + s_2)$ are contributors to our record X.

Their combined contribution to X is

$$\begin{aligned} a_1\cos(\beta\, k + s_1) &+ a_2\cos(\beta\, k + s_2) \\ &= a_1\cos(\beta\, k)\cos(s_1) - a_1\sin(\beta\, k)\sin(s_1) \\ &\quad + a_2\cos(\beta\, k)\cos(s_2) - a_2\sin(\beta\, k)\sin(s_2) \\ &= A\,\cos(\beta\, k) - B\,\sin(\beta\, k) \end{aligned}$$

where $A = a_1\cos(s_1) + a_2\cos(s_2)$ and $B = a_1\sin(s_1) + a_2\sin(s_2)$.

Defining $a_3 = \sqrt{A^2+B^2}$ and letting s_3 be a value for which

$$\cos(s_3) = \frac{A}{\sqrt{A^2+B^2}} \quad \text{and} \quad \sin(s_3) = \frac{B}{\sqrt{A^2+B^2}}$$

the last line above becomes $a_3\cos(\beta\, k + s_3)$.

In other words, two (or any number of) contributions from different phenomena which produce "elemental" records with the same frequency but (possibly) different shifts cannot be distinguished in the record X from a single contribution with that frequency and a new shift. It is too much to expect that we will be able to identify different contributions of the same frequency.

(ii) Suppose we had an elemental contribution of the form

$$a\,\cos\left(2\pi\,C\,k + s\right)$$

for some positive integer C. In other words, the frequency of the contributing phenomenon is an integer multiple of the sampling frequency. This is $a\cos(s)$ for every k. This contribution cannot be distinguished from a constant based on the evidence in the record.

(iii) Select a number $P > 0$ with $CP \neq 1$ for any integer C.

Suppose that a phenomenon contributing to the sample measurements is of "elemental form" and has true period P.

This contribution to the record has the form

$$a\,\cos\left(2\pi\,\frac{1}{P}\,k + s\right).$$

Given these conditions on P there is an integer B with $B \geq 0$ and number ε with $0 < \varepsilon < P$ for which $1 = BP + \varepsilon$. In fact, if $[\,\cdot\,]$ denotes the "greatest integer function" then $B = [1/P]$.

ε/P is the fraction of a period P between time BP and time 1. At time 1 our phenomenon is "ahead" of its last full cycle, by that fraction of its cycle. So at time P/ε these fractions will have added up to one extra full cycle.

We note that if $B = 0$ then $\varepsilon = 1$ and the analysis does nothing for us: this just corresponds to periods $P > 1$ and then $P/\varepsilon = P$.

In any case, time P/ε may not (probably *won't* be) at a sample time but (except for the $B = 0$ case above) it is a different period. At sample times a phenomenon with that period will be indistinguishable from one with period P.

Reasoning in a similar fashion, there is an integer C with $C \geq 1$ and number ε with $0 < \varepsilon < P$ for which $1 = CP - \varepsilon$.

ε/P is the fraction of a period P between time 1 and time CP. At time 1 our phenomenon is "behind" by that fraction of its cycle. So at time P/ε these fractions will have added up and we will have lost one full cycle. The record cannot distinguish a phenomenon with periodicity P/ε from one with periodicity P.

The situation where P exceeds 1, uninteresting in our first case, here produces a period $P/(P-1)$ which must differ from P unless $P = 2$. This case corresponds to the Nyquist frequency.

Letting A_a stand for the "ahead" calculated period and A_b the "behind" calculated period we have for any P with $P > 0$ and $kP \neq 1$ for any integer k, and for B equal to the greatest integer in $1/P$ and $C = B + 1$

$$A_a = \frac{P}{1 - BP}, \text{ (note } B \geq 0) \quad \text{and} \quad A_b = \frac{P}{CP - 1}, \text{ (note } C \geq 1).$$

Note: our conditions on P forbid any A_a or A_b of the form $1/n$ for integer $n > 2$.

Let's turn this around and focus on a target period A and the various P, B and C combinations which could produce it.

Solving the first and second cases gives

$$P = \frac{A}{BA+1} \text{ (integer } B \geq 0) \quad \text{and} \quad P = \frac{A}{CA-1} \text{ (integer } C \geq 1)$$

corresponding to frequencies

$$f = B + \frac{1}{A} \text{ (integer } B \geq 0) \quad \text{and} \quad f = C - \frac{1}{A} \text{ (integer } C \geq 1)$$

The record cannot distinguish between positive and negative periods, so if you want to consider all possible periods which might be conflated, you might as well throw "plus or minus" in front of all the periods and frequencies considered above. Then if you allow negative integers as well, the "ahead" and "behind" cases can be subsumed into the single relationship

$$P = \pm\frac{A}{BA+1} \quad \text{and} \quad f = \pm\left(B + \frac{1}{A}\right) \text{ (any integer } B).$$

With a bit of switching around negative periods and negative integers, you see that all the same periods are associated, and there is complete symmetry between $\pm A$ and $\pm P$. Specifically, if

$$P_1 = \pm\frac{A}{B_1A+1} \quad \text{and} \quad P_2 = \pm\frac{A}{B_2A+1}$$

you can solve the first one for A in terms of P_1 (two ways, actually) and substitute that into the second equation producing

$$P_2 = \pm\frac{P_1}{B_3P_1+1}.$$

Since A itself is an example of such a P_2 (the case $B_2 = 0$) the collective "output" of this relation for any A would be the same if A were to be replaced by any of the periods with which it could be confused in the record.

This effect, in which phenomena appear to have periods other than their true periods, is called **aliasing** and is an unavoidable feature of discrete data collection. It is a "bad thing" in the sense that we cannot know from the data which of the possible frequencies corresponds to the real phenomenon, and if you recall that was our goal.

We will now consider how many "large periods" might be in an **alias class**, that is, a group of periods which might be confused with each other in the record.

Thought of as a function of a real variable B, the relation $P = \frac{A}{BA+1}$ has 0 as the large-B asymptote and a vertical asymptote at $B = -1/A$. It gets smaller in magnitude as you move away from there, so the largest values will be near $B = -1/A$. Suppose, purely for convenience, that $A > 0$ and let $\varepsilon = 1/A - [1/A]$. So with our conditions on A we have $0 < \varepsilon < 1$.

The largest-magnitude values of P occur for integers B near $-[1/A]$.

$$P = \frac{A}{BA+1} = \frac{1}{B + \frac{1}{A}}$$

yields the corresponding values

$$\cdots \frac{1}{\varepsilon-3} \quad \frac{1}{\varepsilon-2} \quad \frac{1}{\varepsilon-1} \quad \frac{1}{\varepsilon} \quad \frac{1}{\varepsilon+1} \quad \frac{1}{\varepsilon+2} \quad \frac{1}{\varepsilon+3} \quad \cdots.$$

The case $\varepsilon = 1/2$, corresponding to the Nyquist frequency, is exceptional. There, $1/2$ is the maximum magnitude frequency on this list.

In all other cases exactly two frequencies exceed 1 in magnitude, and one of these will exceed 2 in magnitude. The other is strictly between 1 and 2 in magnitude.

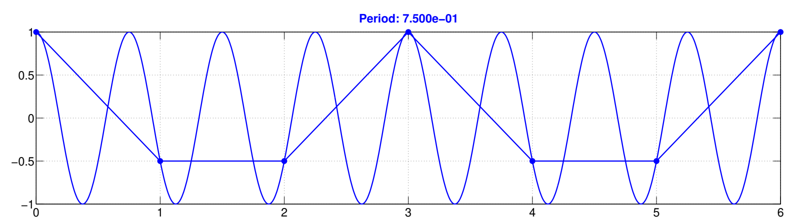
Replacing reference to Δt, the periods $P\Delta t$ and $A\Delta t$ are aliased exactly when

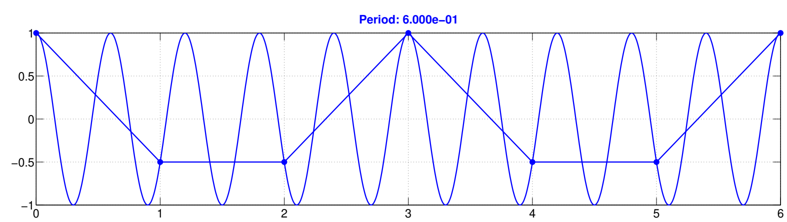
$$P\Delta t = \pm \frac{A}{BA+1}\Delta t = \pm \frac{1}{B+\frac{1}{A}}\Delta t \qquad \text{(for some integer } B\text{)}.$$

In terms of frequencies, this is:

$$f_P \;= f_A \;\pm B f_s \quad \text{ or } \quad f_P \;= -f_A \;\pm B f_s \text{ (for some integer } B\text{)}.$$

The point is that in any alias class (except the Nyquist frequency class itself) exactly one period exceeds $2\Delta t$ and so has frequency less than the Nyquist frequency, and might be directly visible. Another is between the sample frequency and the Nyquist frequency, and all the rest have even higher frequencies.





This type of aliasing is of course a feature of the specific sample interval, and will be *actually* visible whenever the sample length is long enough to reveal it.

For instance frequencies in underlying phenomena between $\frac{N-1}{N}f_s$ and $\frac{N+1}{N}f_s$ may be present and contribute "noise" to the record but they will not be conflated with a frequency that could be detected in the record. The record is not long enough for that because the corresponding large period to which it is aliased exceeds $N\Delta t$.

Making the record longer aids in detecting low frequencies but causes more aliasing with high frequencies to become visible. Making Δt smaller, on the other hand, allows for the detection of these higher frequencies directly.

Aliasing is responsible for many of the artifacts visible or audible in compressed or recorded pictures or sound, features not present in the original. Great efforts are spent reducing or controlling these errors. They cannot be eliminated entirely.

Aliasing can also be used to advantage; it is, for instance, behind techniques of frequency-shifting used in radio and other communication technologies.

4. The Elemental Records

We are now going to focus on those specific elemental records used in the **Discrete Fourier Transform** (usually abbreviated **DFT**) which have the form

$$\boxed{Y_m(k) = \cos\left(\frac{2\pi\, m}{N}\, k + s_m\right).}$$

Note we could just as easily incorporated this shifted cosine as an explicit shift-dependent combination of "shiftless" terms

$$\cos\left(\frac{2\pi k m}{N}\right) \quad \text{and} \quad \sin\left(\frac{2\pi k m}{N}\right).$$

These functions have period $\frac{N}{m}\Delta t$ (for nonzero m) and frequencies $\frac{m}{N} f_s$.

Although defined for any m the DFT uses integer m between 0 and $N-1$.

Also, note that $Y_{m+N}(k) = Y_m(k) = Y_m(k+N)$ for all m and k.

Our goal is to find a way to write our record vector X as a linear combination

$$X = c_0 Y_0 + c_1 Y_1 + \cdots + c_{N-1} Y_{N-1}.$$

If we can we will consider if this choice is optimal in some way, or unique.

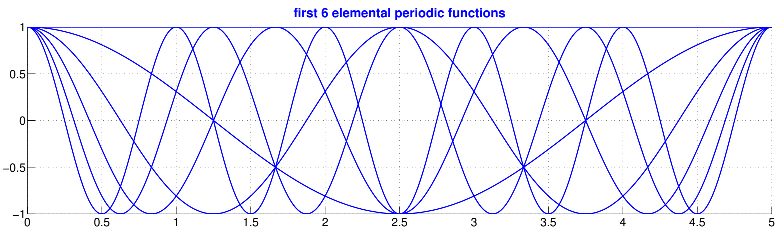
Frequencies

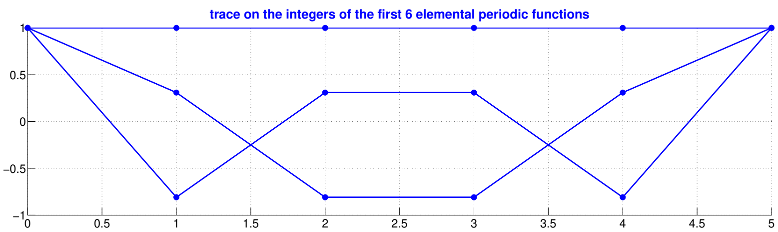
$$\frac{kN \pm m}{N} f_s = k f_s \pm \frac{m}{N} f_s \quad \text{for any integer } k$$

are all, by the calculations in the last section, aliased in the record to $\frac{m}{N} f_s$.

In particular, $\frac{N-m}{N} f_s$ and $\frac{m}{N} f_s$ are aliased.

If N is even and $m = \frac{N}{2}$ then both these frequencies equal f_{Nyquist} but otherwise one is below the Nyquist frequency and so (potentially) directly visible in the record, while the other is between the sample frequency and the Nyquist frequency and *not* directly visible.





On the evidence in the record we could not expect to distinguish between Y_{N-m} and Y_m. So we will not include redundant (and not directly visible anyway) elemental records in our decomposition of X.

Specifically, if N is odd, then the functions

$$Y_0,\ Y_1, \dots, Y_{\frac{N-1}{2}}$$

will be used, while if N is even we will work with

$$Y_0,\ Y_1, \dots, Y_{\frac{N}{2}}.$$

If N is odd, each of these selected Y_m is determined by two coefficients, so there are $N+1$ coefficients among them. However Y_0 is determined by one coefficient (the second is zero), so a total of N coefficients must be found for a decomposition of X into a sum of these elemental functions.

If N is even, then the functions

$$Y_1, \dots, Y_{\frac{N}{2}-1}$$

are determined by $N-2$ independent coefficients, and one more pins down the constant term Y_0 as above. The term with two periods in the record

$$Y_{\frac{N}{2}} = \ a_{\frac{N}{2}} \cos(\pi\, k) + b_{\frac{N}{2}} \sin(\pi\, k)$$

is specified by $a_{\frac{N}{2}}$ alone so, again, we have a total of N parameters to find.

Since there are N pieces of data in X we ought to be able to find these parameters as it stands, using the techniques we know to solve these N linear equations with N unknowns. In the next section we remind you of notation, rephrase the issue in terms of familiar techniques from Linear Algebra, and find a better way.

5. A Particular Basis for $\mathbb{C}^N$

Recall (Euler's formula) that points on the unit circle in the complex plane can be represented as

$$e^{i\theta} = \cos(\theta) + i\,\sin(\theta) \equiv cis(\theta)$$

and also *any* point in the plane can be represented as

$$re^{i\theta} = r\,cis(\theta)$$

for real r and real θ by identifying the first component of the coordinates of a point with the real part of the complex number, and the second component with the complex part.

The numbers r and θ are, essentially, the polar coordinates, and $re^{i\theta}$ is called the **polar form** of the number with **rectangular form** $a + bi$, where $a = r\cos(\theta)$ and $b = r\sin(\theta)$.

If $z = r\,cis(\theta)$ and $w = s\,cis(\mu)$

$$zw = r\,cis(\theta)\,s\,cis(\mu) = re^{i\theta}\,se^{i\mu} = rse^{i(\theta+\mu)} = rs\,cis(\theta + \mu).$$

This implies that if you wish to rotate a point $re^{i\theta}$ counterclockwise around the origin by angle μ you multiply by $e^{i\mu}$, sending $re^{i\theta}$ to $re^{i(\theta+\mu)}$.

It also implies **De Moivre's formula**:

$$(r\,cis(\theta))^n = \left(r\,e^{i\theta}\right)^n = r^n e^{in\theta} = r^n\,cis(n\theta).$$

Finally, we wish to remind the reader that the standard inner product or dot product in $\mathbb{C}^N$ is defined a little differently from that in $\mathbb{R}^N$, although it agrees with the former definition when applied to vectors with real coordinates only.

If $W = (W(0), \ldots, W(N-1))$ and $Z = (Z(0), \ldots, Z(N-1))$ are in $\mathbb{C}^N$ then

$$W \cdot Z = \overline{W}(0)Z(0) + \cdots + \overline{W}(N-1)Z(N-1) = \sum_{m=0}^{N-1} \overline{W}(m)Z(m)$$

where $\overline{W}(m)$ denotes the complex conjugate of $W(m)$.

Magnitudes, projection, orthogonality, the Gram-Schmidt process and orthonormal bases are defined using this dot product in just the same way as in $\mathbb{R}^N$, with this single change[2].

We let ω denote the complex number

$$\omega \equiv e^{\frac{2\pi i}{N}} = cis\left(\frac{2\pi}{N}\right)$$

[2]Be aware that, often, mathematicians use a dot product in which conjugation is applied to the *second* factor rather than the first. Engineers and physicists *usually* use our convention.

which is, by the way, on the unit circle, as are its additive and multiplicative inverses, all its powers and their conjugates.

It is a fact, used below, that $\omega^N = 1$ and also the conjugate $\overline{\omega^k}$ of ω^k is ω^{-k}. Further,

$$\omega^{kN} = 1 \quad \text{so} \quad \omega^{k(N-m)}\omega^{km} = 1 \quad \text{and then} \quad \omega^{k(N-m)} = \omega^{-km} = \overline{\omega^{km}}.$$

With these facts in hand, we rephrase our search for periodicity using complex records.

Create, for each integer k, the following list of complex numbers:

$$W_k(m) = e^{i\frac{2\pi}{N}km} = cis\left(\frac{2\pi}{N}km\right) = \omega^{km} \qquad m = 0, \dots, N-1.$$

For each k the vector $W_k = (W_k(0), \dots, W_k(N-1))$ is a member of $\mathbb{C}^N$.

By the above calculation we see that for $1 \le m < N$ we have $W_{N-m} = \overline{W}_m$.

$$W_k \cdot W_k = \sum_{m=0}^{N-1} \overline{W}_k(m)W_k(m) = \sum_{m=0}^{N-1} \overline{\omega}^{km}\,\omega^{km} = \sum_{m=0}^{N-1} 1 = N.$$

So each of these vectors is of length $\sqrt{N}$.

If k and l are unequal integers between 0 and $N-1$ we find

$$\begin{aligned} W_l \cdot W_k &= \sum_{m=0}^{N-1} \overline{W}_l(m)W_k(m) = \sum_{m=0}^{N-1} \overline{\omega}^{lm}\,\omega^{km} = \sum_{m=0}^{N-1} \omega^{lm-km} \\ &= \sum_{m=0}^{N-1} \left(\omega^{l-k}\right)^m = \frac{1-\left(\omega^{l-k}\right)^N}{1-\omega^{l-k}} = \frac{1-\left(\omega^{N}\right)^{l-k}}{1-\omega^{l-k}} = \frac{1-1}{1-\omega^{l-k}} = 0. \end{aligned}$$

So these vectors are perpendicular to each other. Therefore the vectors

$$W_0,\, W_1, \dots,\, W_{N-1}$$

form an orthogonal ordered basis of $\mathbb{C}^{\boldsymbol{N}}$, and any member of $\mathbb{C}^N$, including our record X, can be written in a unique way as a complex linear combination involving these complex vectors.

The vectors $W_0/\sqrt{N},\, W_1/\sqrt{N}, \dots,\, W_{N-1}/\sqrt{N}$ constitute an orthonormal basis, and some of the formulae below would be more symmetrical if we used this basis rather than the W_m, but there are practical computational advantages to *not* doing this, and it is customary to avoid the final normalization step.

Defining the vector projection $Proj_V$ onto nonzero vector V by

$$Proj_V(W) = \frac{V \cdot W}{V \cdot V}\; V$$

We have

$$X = \sum_{k=0}^{N-1} Proj_{W_k}(X) = \sum_{k=0}^{N-1} \frac{W_k \cdot X}{W_k \cdot W_k}\ W_k = \frac{1}{N}\sum_{k=0}^{N-1} (W_k \cdot X)\ \ W_k$$
$$= \frac{1}{N}\sum_{k=0}^{N-1} q_k\ W_k \quad \text{where for each } k, \quad q_k = W_k \cdot X.$$

Because X is a real vector, *and only because of that*, we have for $1 \le k < N$

$$\overline{q_k} = \overline{W_k \cdot X} = \overline{W}_k \cdot X = W_{N-k} \cdot X = q_{N-k}.$$

In other words, because X is a real vector the coefficients q_k corresponding to conjugate basis vectors W_k must be, themselves, conjugate.

We will let X be a column vector whose kth entry is sample value $X(k-1)$, define Q to be the column vector with kth entry q_{k-1} and define W to be the matrix whose kth row is the transpose of W_{k-1}. The above equation then has matrix form

$$X = \frac{1}{N} WQ = \frac{1}{N}\begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \dots & \omega^{2(N-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \dots & \omega^{3(N-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N-1} & \omega^{(N-1)2} & \omega^{(N-1)3} & \dots & \omega^{(N-1)(N-1)} \end{pmatrix} Q.$$

The inverse of an orthogonal matrix (i.e. its columns form an orthonormal basis) is the conjugate transpose of that matrix. The columns of W form an orthogonal basis of vectors with common length $\sqrt{N}$, and a calculation shows that its inverse W^{-1} is the conjugate transpose of W divided by N. Note that W is a symmetric matrix. So

$$Q = \overline{W^t} X = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \omega^{-3} & \dots & \omega^{-(N-1)} \\ 1 & \omega^{-2} & \omega^{-4} & \omega^{-6} & \dots & \omega^{-2(N-1)} \\ 1 & \omega^{-3} & \omega^{-6} & \omega^{-9} & \dots & \omega^{-3(N-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{-(N-1)} & \omega^{-(N-1)2} & \omega^{-(N-1)3} & \dots & \omega^{-(N-1)(N-1)} \end{pmatrix} X.$$

Q is called the **discrete Fourier transform** of X. Done directly, it takes N^2 complex multiplications (and a bunch of additions) to calculate Q from X. But because of the evident symmetries and relations among the entries in the matrix many of these multiplications are duplicated and by carefully organizing the calculation, if N has many small factors (for instance if N is a power of 2) it can be done with roughly $N\,Log(N)$ multiplications, a huge time saver when N is large. The algorithm that carries this out is called the **fast Fourier transform.**

Remember that Q is the list of coordinates of the vector NX in the basis given by the columns of W. So the entries of Q provide the *one and only* way of writing X as a linear combination of these linearly independent complex periodic functions.

As a final piece of notation, you will commonly see the caret symbol used to denote the application of the discrete Fourier transform:

$$\wedge\colon \mathbb{C}^N \to \mathbb{C}^N \quad \text{given by formula} \quad \hat{X} = \overline{W^t}X$$

and its inverse function, the inverse discrete Fourier transform, is indicated as

$$\vee\colon \mathbb{C}^N \to \mathbb{C}^N \quad \text{given by formula} \quad \overset{\vee}{X} = \frac{1}{N}WX.$$

So for any X in $\mathbb{C}^N$ we have $\overset{\wedge}{\overset{\vee}{X}} = \overset{\vee}{\hat{X}} = X$.

6. Back to Sines and Cosines

Suppose Q is the discrete Fourier transform of real X, as above, and

$$\boxed{q_k = g_k + ih_k \quad \text{for certain real } g_k \text{ and } h_k.}$$

We have already established for $k = 1, \ldots, N-1$ that

$$q_k = \overline{q_{N-k}} \quad \text{and} \quad W_k = \overline{W_{N-k}}$$

which means

$$\overline{q_k W_k} = q_{N-k}W_{N-k} \quad \text{for} \quad k = 1, \ldots, N-1.$$

If N is even, then

$$\begin{aligned}
N\,X &= q_0W_0 + q_1W_1 + q_2W_2 + \cdots + q_{N/2-1}W_{N/2-1} \\
&\quad + q_{N/2}W_{N/2} + q_{N/2+1}W_{N/2+1} + \cdots + q_{N-2}W_{N-2} + q_{N-1}W_{N-1} \\
&= q_0W_0 + q_{N/2}W_{N/2} + (q_1W_1 + q_{N-1}W_{N-1}) + (q_2W_2 + q_{N-2}W_{N-2}) \\
&\quad + \left(q_{N/2-1}W_{N/2-1} + q_{N/2+1}W_{N/2+1}\right) \\
&= q_0W_0 + q_{N/2}W_{N/2} + \sum_{m=1}^{\frac{N}{2}-1} \left(q_mW_m + \overline{q_mW_m}\right).
\end{aligned}$$

In each summand on the right, the complex part is zero, and what remains is twice the real part of q_mW_m. And both q_0W_0 and $q_{N/2}W_{N/2}$ are real vectors. Note that W_0 is filled with ones, and $W_{N/2}$ alternates between 1 and -1.

For each k,

$$q_mW_m(k) = (g_m + ih_m)\left(\cos\left(\frac{2\pi km}{N}\right) + i\sin\left(\frac{2\pi km}{N}\right)\right)$$

and twice the real part of this is

$$2g_m\cos\left(\frac{2\pi km}{N}\right) - 2h_m\sin\left(\frac{2\pi km}{N}\right).$$

Recall from Section 4 we can write a pristine elemental *real* record as a shifted cosine of one of the selected frequencies or a linear combination of sines and cosines with *no* shift of that selected frequency. Our original goal therefore was to find real coefficients a_m and b_m which determine the contribution to the record X of those "periodicities" used in the DFT, as

$$a_m \cos\left(\frac{2\pi km}{N}\right) + b_m \sin\left(\frac{2\pi km}{N}\right)$$

in some kind of unique or optimal way.

We find that for even N

$$a_0 = q_0 = g_0, \quad a_{N/2} = q_{N/2} = g_{N/2} \quad \text{and}$$
$$a_m = 2g_m \quad \text{and} \quad b_m = -2h_m \qquad \text{for} \quad m = 1, \ldots, \frac{N}{2} - 1$$

and the kth sample, the kth entry of X, can be written as

$$NX(k) \; = \; a_0 + (-1)^k a_{N/2} + \sum_{m=1}^{\frac{N}{2}-1} \left(a_m \cos\left(\frac{2\pi km}{N}\right) + b_m \sin\left(\frac{2\pi km}{N}\right)\right).$$

If N is odd, the conjugates are paired off similarly, with

$$\begin{aligned} NX \; &= \; q_0W_0 + q_1W_1 + q_2W_2 + \cdots + q_{(N-1)/2}W_{(N-1)/2} \\ &\quad\; + q_{(N+1)/2}W_{(N+1)/2} + \cdots + q_{N-2}W_{N-2} + q_{N-1}W_{N-1} \\ &= \; q_0W_0 + (q_1W_1 + q_{N-1}W_{N-1}) + (q_2W_2 + q_{N-2}W_{N-2}) \\ &\quad\; + \left(q_{(N-1)/2}W_{(N-1)/2} + q_{(N+1)/2}W_{(N+1)/2}\right) \\ &= \; q_0W_0 + \sum_{m=1}^{\frac{N-1}{2}} \left(q_mW_m + \overline{q_mW_m}\right). \end{aligned}$$

Again, in each summand on the right the complex part is zero, and what remains is twice the real part of q_mW_m.

We find for odd N that

$$a_0 = q_0 = g_0, \quad \text{and}$$
$$a_m = 2g_m \quad \text{and} \quad b_m = -2h_m \qquad \text{for} \quad m = 1, \ldots, \frac{N-1}{2}$$

which produces the representation

$$NX(k) \; = \; a_0 + \sum_{m=1}^{\frac{N-1}{2}} \left(a_m \cos\left(\frac{2\pi km}{N}\right) + b_m \sin\left(\frac{2\pi km}{N}\right)\right).$$

As a final remark, the uniqueness of the entries of Q implies that these coefficients provide the *only* way of writing X in terms of these particular sine and cosine functions since, reversing the process above, any alternative would lead to a different Q, which is not possible.

7. Helpful MATLAB m-files

```
function z = elemental(n)

%this will plot the first n+1 elemental periodic functions on [0,n].

t=linspace(0,n,1000);
t1=0:n;
k=0;
figure(12)
clf
hold on
while k < n+1
y=cos(2*pi*k.*t./n);
plot(t,y);
k=k+1;
end
A=num2str(n+1, '%2.0f');
A=['first 'A 'elemental periodic functions'];
title(A);
grid
axis([0 n -1 1])
hold off
k=0;
figure(3)
clf
hold on
while k¡n+1
y2=cos(2*pi*k.*t1./n);
plot(t1,y2,'b.-');
k=k+1;
end
A=num2str(n+1, '%2.0f ');
A=['trace on the integers of the first 'A 'elemental periodic functions'];
title(A);
grid
axis([0 n -1 1])
hold off
```

```
function z = alias(Per1,interv)

%input the period you want to examine and the length of the interval
%(choose 5 or 6 at least) you wish to graph.
%The plot will show the actual graph
%of the cosine with period Per1 and an aliased wave.

A=num2str(Per1, '%7.3e');
A=['Period: 'A];

t1=0:interv;
t2=linspace(0,interv,1000);
y1=cos(2*pi/Per1.*t1);
y2=cos(2*pi/Per1.*t2);

hold on
figure(66)

plot(t1,y1,'b.-');
plot(t2,y2);
title(A);
grid
hold off
```

```
function z = alias2(Per1,Per2,interv)

%This m-file function does the same as the last but for two periods at once.
%This shows if two different periods are aliased to the same visible period

t1=0:interv;
t2=linspace(0,interv,1000);
y1=cos(2*pi/Per1.*t1);
y2=cos(2*pi/Per1.*t2);
y3=cos(2*pi/Per2.*t1);
y4=cos(2*pi/Per2.*t2);

A=num2str(Per1, '%7.3e');
A=['Period: 'A];
B=num2str(Per2, '%7.3e');
B=['Period: 'B];
figure(77)

subplot(2,1,1)
plot(t1,y1,'b.-');
title(A);
hold on
subplot(2,1,1)
plot(t2,y2);
grid
hold off

subplot(2,1,2)
plot(t1,y3,'b.-');
title(B);
hold on
subplot(2,1,2)
plot(t2,y4);
grid
hold off
```

```
function [Per1, P] = aliasall2(Per1,terms)

% This m-file function lists terms (or terms+1) periods aliased to Per1

C=(-floor(terms/2)):(floor(terms/2));
k=1;
while k < length(C)+1
W(k)=1;
k=k+1;
end

P=sort(abs(Per1./(Per1.*C+1)),'descend');

A=['Aliased Periods: 'num2str(Per1, '%7.3e')];
B=['Aliased Periods: Log Scale 'num2str(Per1, '%7.3e')];

figure(34)

subplot(2,1,1)
stem(P,W);
title(A);
grid

subplot(2,1,2)
stem(P,W);
title(B);
ax = axis;
set (gca, 'xscale', 'log')
axis (ax)
grid
hold off
```

```
function z = noisydata(pervec ,coeffvec ,plussvec,interv,standev)

%takes a vector of periods and produces data at the integer
%points between 0 and interv that is the linear combination, with coefficients
%drawn from coeffvec , of cosines with given periods and shift.
%It then adds a bit of normally distributed noise with
%specified standard deviation. The output is the data plus a stem plot.
%Note the first three terms must be vectors of the same length.
% This is a tool to create data to experiment with using fft.

t=0:interv;
m=1;
while m<length(t)+1
k=1;
y(m)=0;
while k<length(pervec )+1
y(m)=y(m)+coeffvec (k)*cos(2*pi/pervec (k)*(m-1)+plussvec(k));
k=k+1;
end
m=m+1;
end
y=y+standev.*randn(1,length(y));

z=y;

stem(t,y)
grid
```