them, although simple, are under patent laws (that also apply if you rediscover the idea on your own).

Suggested Readings. There is a lot of information about dithering in the internet. One of the most informative sites that I have found is `http://caca.zoy.org/study/`. Do not get wrong with the software under lavatory names hosted in the main page, it is done by serious programmers.

2.2 Discrete Fourier analysis

2.2.1 Some discrete transforms

We had already seen in (1.32) the discrete analogue of the Fourier series and integrals. In fact it motivated them. We also saw it later in the more general setting of finite abelian groups (1.99) where (1.32) corresponds to $G = \mathbb{Z}_N$, the cyclic group of N elements. We recover here the definition with some minor notational changes to match part of the literature.

Given a vector $\vec{x} = (x_0, x_1, \dots, x_{N-1}) \in \mathbb{C}^N$, we define its *discrete Fourier transform*, abbreviated DFT, as the vector $(\widehat{x}_0, \widehat{x}_1, \dots, \widehat{x}_{N-1})$ where $\widehat{x}_n$ are the "Fourier coefficients"

$$\widehat{x}_n = \sum_{m=0}^{N-1} x_m e(-nm/N). \tag{2.51}$$

Given these coefficients, one can recover the original vector through the *Fourier inversion formula*

$$x_n = \frac{1}{N} \sum_{m=0}^{N-1} \widehat{x}_m e(nm/N). \tag{2.52}$$

The relation with (1.99) is clear: The vector $\vec{x}$ corresponds to the function $f : \mathbb{Z}_N \longrightarrow \mathbb{C}$ with $f(\overline{n}) = x_n$ for $0 \le n < N$.

An elegant way of introducing the orthogonality relations (1.33) is defining

$$U = (u_{kl})_{k,l=1}^{N} \in \mathrm{U}(N) \qquad \text{with} \quad u_{kl} = \frac{1}{\sqrt{N}} e\Big(-\frac{(k-1)(l-1)}{N} \Big). \tag{2.53}$$

As usual, $\mathrm{U}(N)$ means the $N \times N$ *unitary matrices*, verifying $A^\dagger A = \mathrm{Id}$ where $A^\dagger$ is the conjugate transpose. With this notation, (2.51) and (2.52) are respectively

$$\vec{f} = \sqrt{N} U \vec{x} \qquad \text{and} \qquad \vec{x} = \frac{1}{\sqrt{N}} U^\dagger \vec{f} \tag{2.54}$$

with $\vec{f} = (\widehat{x}_0, \widehat{x}_1, \dots, \widehat{x}_{N-1})$. In this way the inversion formula is a direct consequence of $U \in \mathrm{U}(N)$. The unitary matrices $\mathrm{U}(N)$ preserve the standard scalar product, indeed this property is the reason to introduce them. One deduces at once the *Parseval identities*

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |\widehat{x}_n|^2 \qquad \text{and} \qquad \sum_{n=0}^{N-1} \overline{x}_n y_n = \frac{1}{N} \sum_{n=0}^{N-1} \overline{\widehat{x}}_n \widehat{y}_n. \tag{2.55}$$

In the context of locally compact abelian groups there exists a concept of convolution that in our simple setting is the obvious discretization of (1.76). Given $\vec{x}, \vec{y} \in \mathbb{C}^N$, its convolution $\vec{x} * \vec{y}$ is

$$\vec{z} = \vec{x} * \vec{y} = (z_0, z_1, \dots, z_{N-1}) \qquad \text{with} \quad z_n = \sum_{\substack{k=0 \\ l \equiv n-k}}^{N-1} \sum_{\substack{l=0 \\ (\text{mod } N)}}^{N-1} x_k y_l. \tag{2.56}$$

If one defines $y_{l+mN} = y_l$ for $m \in \mathbb{Z}$, the convolution can be defined with the lighter formula $z_n = \sum_{k=0}^{N-1} x_k y_{n-k}$. The analogue of (1.77) is

$$\widehat{z}_n = \widehat{x}_n \widehat{y}_n. \tag{2.57}$$

The underlying idea in the previous definition is to mimic the formulas in the theory of Fourier series substituting functions by vectors formed by sampled values. As we have seen, for regular 1-periodic functions few terms of the Fourier series give a good approximation. We would like to capture a similar concept in the sampled values. A qualitative way of understanding the situation is through the following combinatorial result that is close to the first formula in (1.69)

Lemma 2.2.1. *For $n \neq 0$ we have*

$$\widehat{x}_n = x_{N-1} - x_0 + \sum_{k=0}^{N-2} \frac{x_{k+1} - x_k}{1 - e(-n/N)} e(-n(k+1)/N). \tag{2.58}$$

Proof (sketch). Use the elementary identity

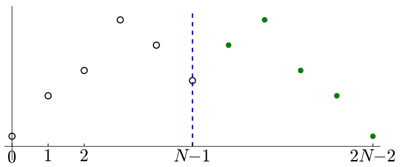
$$\sum_{k=0}^{N-1} x_k y_k = \sum_{k=0}^{N-2} (x_k - x_{k+1}) \sum_{l=0}^{k} y_l + x_{N-1} \sum_{l=0}^{N-1} y_l \tag{2.59}$$

with $y_k = e(-kn/N)$. □

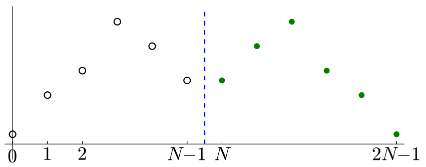
If $\vec{x}$ corresponds to sampled values of a smooth varying periodic function, then $x_{k+1} - x_k$ and $x_{N-1} - x_0$ are small. If n is not close to 0 or N the denominator $1 - e(-n/N)$ is far apart from zero and the corresponding Fourier coefficients $\widehat{x}_n$ are less important to recover the signal with the inversion formula.

In many applications the condition $x_0 \approx x_{N-1}$ resembling periodicity is not assured out of the box and Lemma 2.2.1 suggests that under $x_k \approx x_{k+1}$ the middle Fourier coefficients are equally important for the reconstruction: We cannot compress the signal forgetting some of them. To avoid this problem, it is introduced the *discrete cosine transform*, abbreviated DCT, which has the extra technical advantage of using only real values when the input signal is real. The idea is very simple: If we symmetrize the values of a smooth discrete signal, we force the first and the last value to be close. In the continuous setting it would correspond to the even periodic extension.

There are two typical situations called DCT-I and DCT-II which correspond to perform the symmetry around $N-1$ or around $N-1/2$. The second one is more natural and pleasant from the numerical point of view because it doubles the number of points. The difference is illustrated with the following pictures in which the original signal is marked with hollow points.



DCT-I



DCT-II

Hereafter we only focus on DCT-II and we call it simply DCT as usual.

The duplication of the signal sampled points and the symmetry through $N-1/2$ imply that in the analysis with the DFT we get terms $e\big(n(m+1/2)/2N\big)+e\big(-n(m+1/2)/2N\big)$ that give rise to cosines (see the proof below). The inversion formula loses a part of its symmetry but, as pointed before, we avoid problems with the lack of periodicity and the calculations are with real numbers for real signals. The concrete definition of the DCT and its inversion is contained in the following result

Proposition 2.2.2. *Given* $\vec{x}=(x_n)_{n=0}^{N-1}\in\mathbb{C}^N$ *we define its* discrete cosine transform DCT *as*

$$\widehat{x}_n^c=\sum_{m=0}^{N-1}x_m\cos\big(\frac{\pi n}{N}(m+\frac{1}{2})\big). \tag{2.60}$$

Then we have the Fourier inversion formula

$$x_m=\frac{\widehat{x}_0^c}{N}+\frac{2}{N}\sum_{n=1}^{N-1}\widehat{x}_n^c\cos\big(\frac{\pi n}{N}(m+\frac{1}{2})\big). \tag{2.61}$$

As suggested before, a way of proving this result is to apply the inversion formula (2.52) for the DFT to the symmetrized signal. For illustration we follow this approach although it is not the simplest. An ultra-quick proof motivated by the discretization of certain ODE is included in [Str99]. In this interesting paper it is claimed that the DCT was not discovered until 1974 [ANR74].

Proof. Let y_n the symmetrized signal. For convenience we consider the indexes of y modulo $2N$. In this way, $y_n=x_n$ if $0\le n<N$ and $y_n=x_{-1-n}$ for $-N\le n<0$. The definition (2.51) applied to y reads

$$\widehat{y}_n=\sum_{m=0}^{N-1}x_m e\Big(-\frac{nm}{2N}\Big)+\sum_{m=-N}^{-1}x_{-1-m}e\Big(-\frac{nm}{2N}\Big)\qquad\text{for}\quad |n|\le N. \tag{2.62}$$

Changing m into $-m-1$ in the last sum, we have

$$\widehat{y}_n=\sum_{m=0}^{N-1}x_m\Big(e\Big(-\frac{nm}{2N}\Big)+e\Big(\frac{nm}{2N}\Big)e\Big(\frac{n}{2N}\Big)\Big). \tag{2.63}$$

The big parenthesis is $2e(n/4N)\cos\big(\pi n(m+1/2)/N\big)$ that vanishes for $n=N$. Then

$$\widehat{y}_n = 2e\Big(\frac{n}{4N}\Big)\widehat{x}^c_{|n|} \quad \text{for} \quad |n|<N \qquad \text{and} \qquad \widehat{y}_N = 0. \tag{2.64}$$

The inversion formula (2.52) gives for $0 \le n < N$

$$2Nx_n = 2Ny_n = \sum_{m=-N+1}^{N-1} 2e\Big(\frac{m}{4N}\Big)\widehat{x}^c_{|m|}e\Big(\frac{mn}{2N}\Big) \tag{2.65}$$

and grouping the terms m and $-m$ we have the formula in the statement. □

Let us finish introducing two forms of another transform related to the discretization of classic Fourier analysis.

If theoretically we have to our disposal all the sampled values in past and future of a signal $(x_n)_{n=-\infty}^{\infty}$, we may think that

$$\sum_{n=-\infty}^{\infty} x_n e(-n\xi) \tag{2.66}$$

might be an approximation to its Fourier transform. This is called the *discrete time Fourier transform* because it only involves sampled values of the signal at discrete times. The series (2.66), when converges, defines a 1-periodic function then it does not approximate the Fourier transform which must decay by Riemann-Lebesgue lemma. Indeed, if $x_n = f(n)$ with a rapidly decaying f, by Poisson summation formula (2.11)

$$\sum_{n=-\infty}^{\infty} \widehat{f}(\xi+n) = \sum_{n=-\infty}^{\infty} x_n e(-n\xi). \tag{2.67}$$

The *Z-transform* of $(x_n)_{n=-\infty}^{\infty}$ is (2.66) after the complex change of variables $z=e(\xi)$,

$$X(z) = \sum_{n=-\infty}^{\infty} x_n z^{-n}. \tag{2.68}$$

Of course this is just a formal series if we cannot assure the convergence. This apparently unmotivated transform plays an important role for engineers because, as we will see later, it allows to develop the theory about filter design. From the mathematical point of view it is a generating function that is used to solve linear difference equations. The most fundamental property of the Z-transform is that

$$z_n = \sum_{l=-\infty}^{\infty} x_l y_{n-l} \qquad \text{implies} \qquad Z(z) = X(z)Y(z) \tag{2.69}$$

where X, Y and Z are, respectively, the Z-transforms of the sequences x_n, y_n and z_n.

Suggested Readings. Any book on digital signals enters in the definition of several discrete transforms and in the basics of discrete Fourier analysis but if you are a theoretician and you want to read a masterpiece from the point of view of the quality of the exposition, your book is [Ter99].