1.2.2 Some properties

A simple and still useful property is that we can integrate by parts to get the Fourier coefficients or the Fourier transforms of the k-derivative of a function. Namely for $f \in C^k(\mathbb{T})$ and for $f \in C^k(\mathbb{R})$ with $f^{(k)}$ integrable, we have respectively

$$(1.69) \qquad a_n = (2\pi i n)^{-k} \int_{\mathbb{T}} f^{(k)}(x) e(-nx)\, dx \qquad \text{and} \qquad \widehat{f}(\xi) = (2\pi i \xi)^{-k} \widehat{f^{(k)}}(\xi)$$

for $n \neq 0$, $\xi \neq 0$, where a_n are the Fourier coefficients of f.

These formulas allow to extend our list of explicit examples. The first one justifies in general term by term integration of the Fourier series of a zero average function. As an example of the second, the Fourier transform of $-2\pi x e^{-\pi x^2}$ is $2\pi i \xi e^{-\pi \xi^2}$ because the former is the derivative of $e^{-\pi x^2}$ whose Fourier transform is itself by (1.64).

An interesting theoretical and not so theoretical consequence of (1.69) is that, under the stated hypotheses, we have

$$(1.70) \qquad a_n = O\big(|n|^{-k}\big) \qquad \text{and} \qquad \widehat{f}(\xi) = O\big(|\xi|^{-k}\big)$$

when $n \to \infty$ and $\xi \to \infty$, respectively. In this way, the Fourier coefficients of $f \in C^\infty(\mathbb{T})$ decay quicker than any negative power and this is good news for numerical methods because we can approximate f with few terms of the Fourier series[5]. In the case of the Fourier transform, if f is in the Schwartz class then $\widehat{f}$ is also there. Actually this is the property that gives relevance to this space of functions, it provides a simple environment to work which is preserved by Fourier transforms.

In $\mathbb{T}$ and $\mathbb{R}$, the translations $x \mapsto x + \beta$ preserve the homogeneous structure of the space, whatever it means. If f and g are related by $f(x) = g(x + \beta)$ then we have

$$(1.71) \qquad a_n = e(\beta n) b_n \qquad \text{and} \qquad \widehat{f}(\xi) = e(\beta \xi) \widehat{g}(\xi),$$

where a_n and b_n are, respectively, the Fourier coefficients of f and g.

An arbitrary dilation in general does not preserve 1-periodicity, then we only consider the case of the Fourier transform. We have

$$(1.72) \qquad g(x) = f(x/\delta) \qquad \text{implies} \qquad \widehat{g}(\xi) = \delta \widehat{f}(\delta\xi) \qquad \text{for } \delta > 0.$$

This inverse scaling under Fourier transform is something to keep in mind for the next section.

The functions $\{e(nx)\}_{n\in\mathbb{Z}}$ are orthonormal with respect to the natural functional scalar product $\langle f, g\rangle = \int \bar{f} g$. For regular functions Theorem 1.2.1 and (1.70) assure that we can

[5] As an aside, this is very relevant for calculations in celestial mechanics because periodic models or combinations of periodic models are quite accurate. Think for instance in Kepler laws.

express this scalar product in terms of the Fourier coefficients integrating term by term. Namely, if $\{a_n\}_{n\in\mathbb{Z}}$ and $\{b_n\}_{n\in\mathbb{Z}}$ are the Fourier coefficients of f and g, we have

$$(1.73)\qquad \int_{\mathbb{T}} |f|^2 = \sum_{n\in\mathbb{Z}} |a_n|^2 \qquad \text{and} \qquad \int_{\mathbb{T}} \bar{f}g = \sum_{n\in\mathbb{Z}} \overline{a}_n b_n.$$

This has a tremendous significance in the theoretical side because $L^2(\mathbb{T})$ is a *Hilbert space* with the scalar product $\langle f, g\rangle$, it means that we can take limits (it is complete) and the same happens with $\ell^2(\mathbb{Z})$, the bilateral sequences with bounded square norm. Each function in $L^2(\mathbb{T})$ can be approximated by regular functions, hence (1.73) holds true for any square integrable function. The convergence of the Fourier series arises in $L^2(\mathbb{T})$ without further conditions.

Theorem 1.2.7. *If $f \in L^2(\mathbb{T})$ then $S_N f \to f$ in $L^2(\mathbb{T})$ i.e., $\lim_{N\to\infty} \|S_N f - f\|_2 = 0$.*

In this way, the L^2 theory becomes natural and easy. As well as their theoretical importance, the formulas (1.73) are the source of many impressive identities. For instance, when we apply the first to the sawtooth wave (1.61), we obtain $\int_0^1 (x - 1/2)^2\, dx = \sum_{n\neq 0} (2\pi n)^{-2}$ that gives readily $\sum_{n=1}^{\infty} n^{-2} = \pi^2/6$, a result that (proved in a quite different way) boosted the fame of L. Euler when he was 28 years old.

For Fourier transforms of functions in the Schwartz class, we have similar formulas

$$(1.74)\qquad \int_{-\infty}^{\infty} |f|^2 = \int_{-\infty}^{\infty} |\widehat{f}|^2 \qquad \text{and} \qquad \int_{-\infty}^{\infty} \bar{f}g = \int_{-\infty}^{\infty} \overline{\widehat{f}}\widehat{g}.$$

To prove the second (the first corresponds to $f = g$), it is enough to apply the inversion formula (1.36) twice

$$(1.75)\qquad \int_{-\infty}^{\infty} \bar{f}g = \int_{-\infty}^{\infty} \bar{f}(x) \int_{-\infty}^{\infty} \widehat{g}(\xi)e(x\xi)\, dx d\xi = \int_{-\infty}^{\infty} \overline{\widehat{f}}\widehat{g}.$$

Again, (1.74) allows to construct a nice $L^2(\mathbb{R})$ theory of the Fourier transform except for... our definition of $\widehat{f}$ may be nonsensical for $f \in L^2(\mathbb{R})$. We have $L^1(\mathbb{T}) \subset L^2(\mathbb{T})$ and for $f \in L^2(\mathbb{T})$ its Fourier coefficients always exist while in $\mathbb{R}$, due to the lack of compactness, $L^1(\mathbb{R}) \not\subset L^2(\mathbb{R})$. If $f \in L^2(\mathbb{R}) - L^1(\mathbb{R})$, we have to give sense to $\widehat{f}$ and if we want to save (1.74), the procedure is clear: $\widehat{f}$ should be the limit in $L^2(\mathbb{R})$ of $\widehat{f}_n$ with $\|f_n - f\|_2 \to 0$. Three methods to define $\widehat{f}$ in $L^2(\mathbb{R})$ keeping (1.74) and the inversion formula are discussed in [DM72, §2.3-2.5].

The relations (1.73) and (1.74) are called generically *Parseval identity* or *Plancherel identity*. Some authors apply both names indistinctly and some others establish differences.

Arguably the most important property for signal processing is the behavior of the *convolution* $f * g$ under Fourier analysis. In $\mathbb{T}$ and in $\mathbb{R}$ the convolution is defined respectively as

$$(1.76)\qquad (f*g)(x) = \int_{\mathbb{T}} f(t)g(x-t)\, dt \qquad \text{and} \qquad (f*g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)\, dt.$$

Note that it is commutative $f*g = g*f$. If you look up the convergence theorems, you will see that we have already used convolutions without defining them. In signal processing they allow to introduce factors in Fourier series and integrals acting as filters. Mathematically, if a_n, b_n and c_n are the Fourier coefficients of f, g and $f*g$ with $f, g \in L^2(\mathbb{T})$ or if $f, g \in L^2(\mathbb{R})$, we have respectively

$$c_n = a_n b_n \qquad \text{and} \qquad (f*g)\widehat{\ } = \widehat{f}\,\widehat{g}. \tag{1.77}$$

For instance, the convolution with the Dirichlet kernel (1.40) cuts the Fourier series to give the partial sum S_N in (1.42). In general, if we want to select a certain set of frequencies we must consider the convolution with a function such that its Fourier transform vanishes outside this set.

Suggested Readings. The topics discussed here still belong to the basic theory of Fourier series and integrals and then they are covered by the monographs suggested in the previous subsection.

1.2.3 Uncertainty

In few words, the essence of the *uncertainty principle* is that with waves of frequencies less than ν one misses details of size much smaller than ν^{-1}.

For instance, if one wants a good approximation of a function $f = f(x)$ in such a way that its variation in intervals of length δ is well represented, then the truncated Fourier series $\sum_{|n|\le N} a_n e(nx)$ or the truncated Fourier integral $\int_{-N}^{N} \widehat{f}(\xi)e(x\xi)\,d\xi$ are useless to mimic f with this detail if $N\delta$ is small. The range of frequencies is at least the inverse of the required precision.

Let us focus on regular functions, for instance in the Schwartz space, and in Fourier integrals instead of Fourier series. The "inverse law" is linked to the simple scaling property (1.72). Say that we have a function $F = F(x)$ and we add a "detail" of size δ not modifying the mass of F. We can model this detail as adding a function $\varphi(x/\delta)$ with φ compactly supported in an interval of length 1 and $\int \varphi = 0$ to preserve the mass. By (1.72), the Fourier transform of $G(x) = F(x) + \varphi(x/\delta)$ differs from that of F in $\delta\widehat{\varphi}(\delta\xi)$. We know that $\widehat{\varphi}(0) = \int \varphi = 0$ then this term is expected to be negligible when $\delta\xi$ is small and we need larger values of ξ to notice a difference between F and G. If the "detail" has size δ but it is oscillatory, it does not match the model $\varphi(x/\delta)$ of the previous analysis and it might resonate only with large frequencies and it could take even larges values ξ to detect that there is something there. We shall illustrate the situation with a highly oscillatory example later.

Let us expand the idea to cover a common situation. Imagine that we want to use Fourier analysis for $f = \chi_{[-1,1]}$. It is not even continuous and its Fourier transform shows a poor decay like $|\xi|^{-1}$. We decide to regularize it as a function f_δ in the Schwartz space such that $f = f_\delta$ except in the intervals $[-1-\delta/2, -1+\delta/2]$ and $[1-\delta/2, 1+\delta/2]$ and preserving the mass (the integral) in each of them. With an eye to applications, we can think that this is a winning strategy because $\widehat{f_\delta}$ is now rapidly decreasing by (1.70) and f