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

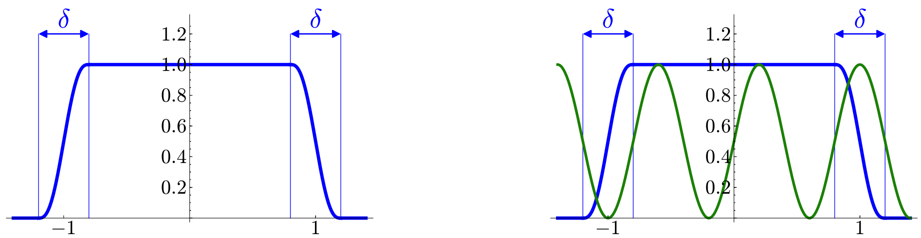
In few words, the essence of the *uncertainty principle* is that with waves of frequencies less than ν one misses details of size much smaller than ν^{-1}.

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.

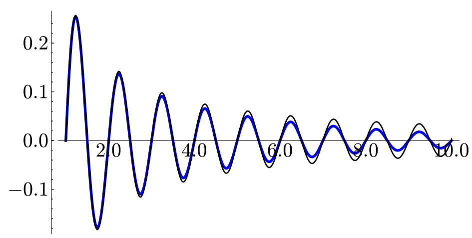
Let us focus on regular functions, for instance in the Schwartz space, and in Fourier integrals instead of Fourier series. The "inverse law" is linked to the simple scaling property (1.72). Say that we have a function $F = F(x)$ and we add a "detail" of size δ not modifying the mass of F. We can model 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

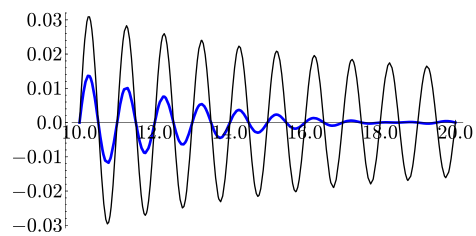
and f_δ are very close, for instance in L^2 norm. The drawback is that a wavelength greater than δ skips the intervals and no change is noticed.



If we do not employ frequencies greater than δ^{-1}, $\widehat{f}_\delta - \widehat{f}_{\delta'}$ with $\delta' < \delta$ is like the zero function. If we think in f as f_0, the conclusion is that using Fourier transforms we cannot distinguish f from its regularization until we do not reach high frequencies. This is the uncertainty in this case. In particular, we cannot profit from the quick decay of $\widehat{f}_\delta$ until ξ is very large. These are actual graphics for $\delta = 0.1$ (using a C^3 regularization instead of C_0^∞ to ease some computational aspects).



$\widehat{\mathsf{f}}$ and $\widehat{\mathsf{f}}_\delta$ in $[1,10]$

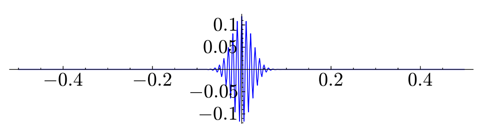


$\widehat{\mathsf{f}}$ and $\widehat{\mathsf{f}}_\delta$ in $[10,20]$

Note that both Fourier transforms are quite similar when $|\xi|$ is much less than δ^{-1} and we only notice the expected quick decay of $\widehat{f}_\delta$ when we have past δ^{-1}.

Let us see in an example with Fourier series that we do not have a "certainty" on the range of frequencies, it depends on the case. Consider the 1-periodic function

$$f(x) = \sum_{k=1}^{\infty} e^{-5-(k-100)^2/100}\cos(2\pi k x) \tag{1.78}$$



Its plot in $[-0.5, 0.5]$ shows that it is essentially the zero function except in an interval of length like 0.1. One can guess that a truncation of the Fourier series to $|n| \le 10$ or so is enough to show that there is something at this scale. This guess is wrong because the definition of f shows readily that its Fourier coefficients are given by $a_0 = 0$ and

$$a_n = \frac{1}{2} e^{-5-(|n|-100)^2/100} \qquad \text{for } n \in \mathbb{Z} - \{0\}. \tag{1.79}$$

This is infinitesimally small when $|n| \le 10$, being its maximal value in this range less than $5\cdot 10^{-38}$. If we extend the truncated Fourier series to $|n| \le 20$ or $|n| \le 30$, still we get a near to zero function, we do not see anything suggesting the previous plot. In fact (1.79) shows that the Fourier coefficients are only noticeable where $|n|$ differs from 100 in something comparable to 10. We need these frequencies to recover the aspect of the function. This is due to the internal oscillations of f. There is a similar situation with the Fourier transform: If we multiply a function by $e(\beta x)$, its Fourier transform is shifted by β by (1.71). Adding several of this multiplications we can force the Fourier transform to live in many intervals.

Another way of thinking about the *uncertainty principle* and its most popular formulation is that one cannot localize simultaneously a function and its Fourier transform. There is a nice result due to G.H. Hardy [Har33] quantifying this property when the decay of f is controlled by a Gaussian function.

Theorem 1.2.8 (Hardy)**.** *If $f(x) = O\big(e^{-\alpha\pi x^2}\big)$ and $\widehat{f}(x) = O\big(e^{-\beta\pi x^2}\big)$ with $\alpha, \beta > 0$ and f is not identically zero, then $\alpha\beta \le 1$. Moreover the equality is reached if and only if f is a constant multiple of $e^{-\alpha\pi x^2}$.*

The proof is very short if one applies a (false) result of complex variables waving hands. To cover the loose ends see [Tao] or [DM72].

Proof (with a gap). We can assume $\alpha = 1$ by (1.72). In this case, it is enough to prove that if $f(x)$ and $\widehat{f}(x)$ are $O\big(e^{-\pi x^2}\big)$, then f is a constant multiple of $e^{-\pi x^2}$.

Under $f(x) = O\big(e^{-\pi x^2}\big)$, the function

$$F(z) = e^{\pi z^2} \int_{-\infty}^{\infty} f(t) e(-tz)\, dt \qquad \text{with} \qquad z = x + iy \in \mathbb{C}, \tag{1.80}$$

defines an entire function because the function under the integral is $O\big(e^{-\pi t^2 + 2\pi t y}\big)$.

If $y = 0$, using $\widehat{f}(x) = O\big(e^{-\pi x^2}\big)$ we have $F(z) = e^{\pi x^2} O\big(e^{-\pi x^2}\big) = O(1)$. On the other hand, if $x = 0$, $F(z) = e^{-\pi y^2} O\Big(\int_{-\infty}^{\infty} e^{-\pi t^2 + 2\pi t y}\, dt\Big) = O(1)$, just changing variables $t \mapsto t + y$.

Let us say (or dream) that maximum modulus principle can be applied in each quadrant of $\mathbb{C}$ (here it is the gap). Then as F is bounded in the real and imaginary axes, we have that F is bounded in each quadrant and consequently it is constant (Liouville's theorem). For each $z = x$ real, we have const. $= e^{\pi x^2} \widehat{f}(x)$ and eliminating $\widehat{f}$ and taking the inverse transform, $f(x) = \text{const.}\ e^{-\pi x^2}$. □

If you are curious about the gap but not so curious to go to the bibliography you will like to know that maximum modulus principle is not true in full generality for a quadrant but it is true if we assume that the growth is under control. The proxy of the maximum modulus principle for unbounded regions is the *Pragmén-Lindelöf principle*.

Even third grade kids know that uncertainty principle is something of quantum physics with philosophical consequences. How is that? Does the third grade syllabus include now Fourier transforms? You can find a big pile of books with the nonsensical things said in the media about quantum physics and repeated by everybody but here they have a point.

The epitome of the uncertainty principle is *Heisenberg's uncertainty principle* and somebody can accuse engineers and mathematicians of hijacking the term. In quantum physics the space of moments is related to the Fourier transform and one of the mathematical statements of this principle is the inequality in the following result. To be precise, actually W. Heisenberg did not state it in this way, he obtained the idea working with Fourier series of an anharmonic oscillator [SR01] and presented it as a lack of commutativity [Hei27], as we shall do later.

Theorem 1.2.9 (Heisenberg inequality). *For any $a, b \in \mathbb{R}$ and $f \in L^2(\mathbb{R})$*

$$16\pi^2 \int_{-\infty}^{\infty} (x-a)^2 |f(x)|^2\, dx \cdot \int_{-\infty}^{\infty} (\xi - b)^2 |\widehat{f}(\xi)|^2\, d\xi \ge \|f\|_2^4 \tag{1.81}$$

if the integrals exist. Moreover the equality is reached if and only if f is a constant multiple of $e(bx)e^{-c(x-a)^2}$ for some $c > 0$.

Except in the trivial case $f = 0$, we can always assume under scaling that $\|f\|_2 = 1$, hence $|f(x)|^2\, dx$ and $|\widehat{f}(\xi)|^2\, d\xi$ are probability measures. The minimum of the left hand side is reached for the expectations $a = \int x|f(x)|^2\, dx$ and $b = \int \xi |\widehat{f}(\xi)|^2\, d\xi$. In this situation, Heisenberg inequality states that the product of the variances is always greater than $1/16\pi^2$. This is close to the quantum mechanics interpretation that we shall briefly consider later. Translating at the same time f and $\widehat{f}$ one can always force $a = b = 0$ (see the proof).

Proof. Define $f(x) = g(x-a)e(bx)$. It is easy to see that $\widehat{f}(\xi) = \widehat{g}(\xi - b)e\big(a(b-\xi)\big)$ and then Heisenberg inequality is equivalent to

$$16\pi^2 \int_{-\infty}^{\infty} x^2 |g(x)|^2\, dx \cdot \int_{-\infty}^{\infty} \xi^2 |\widehat{g}(\xi)|^2\, d\xi \ge \|g\|_2^4, \tag{1.82}$$

i.e., we can restrict ourselves to the case $a = b = 0$.

Assume firstly that g is in the Schwartz space of rapidly decreasing functions to avoid convergence problems. Integrating by parts

$$\int_{-\infty}^{\infty} |g(x)|^2\, dx = -\int_{-\infty}^{\infty} x\big(|g(x)|^2\big)'\, dx = -2\Re \int_{-\infty}^{\infty} x g(x)\overline{g'(x)}\, dx. \tag{1.83}$$

Cauchy-Schwarz inequality and Parseval identity (1.74) prove

$$\|g\|_2^4 \le 4\int_{-\infty}^{\infty} x^2|g(x)|^2\, dx \cdot \int_{-\infty}^{\infty} |g'(x)|^2\, dx = 4\int_{-\infty}^{\infty} x^2|g(x)|^2\, dx \cdot \int_{-\infty}^{\infty} |\widehat{g'}(\xi)|^2\, d\xi \tag{1.84}$$

and it is enough to use (1.69). The inequality saturates (becomes an equality) if and only if xg and g' are proportional and solving a simple ODE this is the same as saying that g is a Gaussian function.

To extend the proof when g is not in the Schwartz class it is enough to approximate by a sequence $\{g_n\}_{n=1}^{\infty}$ in this space in such a way that $\lim \int_{-\infty}^{\infty} (1+\xi^2)|\widehat{g}(\xi) - \widehat{g}_n(\xi)|^2\, d\xi = 0$. See the details in [DM72, §2.8]. ∎

If you talk to your physicist friend and proudly mention you now know that Heisenberg uncertainty in quantum mechanics is a property of the Fourier transform she will reply "Of course it is not, it is about the noncommutativity of two operators" and she can show you tons of books and web pages in which the *commutation relation*

$$[\mathbf{x}, \mathbf{p}] = i\hbar \qquad \text{or} \qquad [\mathbf{x}, \mathbf{p}] = i \tag{1.85}$$

is printed with big types and boxed. We focus on the second which is not other than the first one written in *natural units* in which $\hbar = 1$. After asking and asking your friend, you will figure out that at least in the 1D setting the *position* $\mathbf{x}$ and the *momentum* $\mathbf{p}$ are the operators acting on $\Psi : \mathbb{R} \longrightarrow \mathbb{C}$ given by

$$\mathbf{x} : \Psi \longmapsto x\Psi \qquad \text{and} \qquad \mathbf{p} : \Psi \longmapsto -i\frac{d}{dx}\Psi. \tag{1.86}$$

Strange, isn't it? Then the commutation relation (1.85), with $[\cdot, \cdot]$ the commutator, as usual, turns to be rather easy:

$$[\mathbf{x}, \mathbf{p}]\Psi = \mathbf{x}(\mathbf{p}\Psi) - \mathbf{p}(\mathbf{x}\Psi) = x\Big(- i\frac{d}{dx}\Psi\Big) + i\frac{d}{dx}(x\Psi) = i\Psi. \tag{1.87}$$

It goes without saying that the giant step in quantum mechanics is to model positions and momentum with operators, not checking almost trivial calculations like this.

A mantra for physicists is that for two Hermitian operators A and B a commutation relation of the form $[A, B] = iC$ gives the uncertainty relation

$$\Delta A\, \Delta B \geq \frac{1}{2}|\langle C\rangle|. \tag{1.88}$$

They say, if A and B commute $C = 0$ and you can measure A and B with arbitrarily small *uncertainties* ΔA and ΔB, but if they do not, then there is a limit in the precision of the simultaneous measurement of A and B. For position $\mathbf{x}$ and momentum $\mathbf{p}$ in natural limits the lower limit for $\Delta A\, \Delta B$ is $1/2$ and in "normal" units $\hbar/2$, around $5.27 \cdot 10^{-35} m^2 kg/s$, something infinitesimal for our daily experience but relevant at atomic scale. Wait, wait, wait… we urgently need a Physics-Mathematics translator. What is the uncertainty ΔA of an operator? What are those funny vertical lines and brackets around C in (1.88)?

First of all, physicists apply their Hermitian operators (they say *observables*) to functions very often denoted by Ψ (they say *wave functions*) usually with $|\Psi|^2$ a probability density function, although it is not relevant for (1.88). To fix ideas you can think in $\Psi : \mathbb{R} \longmapsto \mathbb{C}$, as before, or $\Psi : \mathbb{R}^n \longmapsto \mathbb{C}$. Once a wave function (a *state*) is fixed, the average (expectation) of a Hermitian operator A acting on it and its uncertainty are defined by

$$\langle A\rangle = \int \overline{\Psi} A\Psi \qquad \text{and} \qquad \Delta A = \Big(\int \overline{\Psi}(A - \langle A\rangle \mathrm{Id})^2\Psi\Big)^{1/2}. \tag{1.89}$$

Here the square means the square of the operator. At least formally, ΔA is a kind of standard deviation and the name *uncertainty* is not unjustified. The vertical lines in (1.88) are the absolute value (modulus) of the complex number. Physicists usually are not very worried in this context about conditions assuring the existence of $\langle A\rangle$, ΔA, etc. Monkey see, monkey do.

Proposition 1.2.10. *Taking as granted the existence of the quantities appearing in the proof, the relation* (1.88) *holds for* A *and* B *Hermitian operators with* $[A,B]=iC$.

Proof. Changing A and B by $A-\langle A\rangle\mathrm{Id}$ and $B-\langle B\rangle\mathrm{Id}$, we can assume $\langle A\rangle=\langle B\rangle=0$. By Cauchy-Schwarz inequality

$$\Delta A\,\Delta B=\Big(\int|A\Psi|^2\int|B\Psi|^2\Big)^{1/2}\ge\int\overline{A\Psi}B\Psi=\int\overline{\Psi}AB\Psi. \tag{1.90}$$

The first and the last equalities follow because A is Hermitian. Clearly

$$\int\overline{\Psi}AB\Psi=\frac{1}{2}\int\overline{\Psi}(AB+BA)\Psi+\frac{i}{2}\int\overline{\Psi}C\Psi. \tag{1.91}$$

As $AB+BA$ and C are Hermitian, the integrals are real and taking absolute values the result follows. □

Very easy, right? It is nevertheless unclear the relation of (1.88) with "our" Heisenberg inequality in Theorem 1.2.9. Take $A=\mathbf{x}$ and $B=\mathbf{p}$ as in (1.86) and write f instead of Ψ. For the sake of simplicity we assume $\langle A\rangle=\langle B\rangle=0$. Then we have

$$(\Delta A)^2=\int x^2|f(x)|^2\,dx,\qquad\langle C\rangle=\|f\|_2^2 \tag{1.92}$$

and by Parseval identity and the properties of the Fourier transform

$$(\Delta B)^2=-\int\overline{f}f''=-\int\overline{\widehat{f}(\xi)}(2\pi i\xi^2)\widehat{f}(\xi)\,d\xi=4\pi^2\int\xi^2|\widehat{f}(\xi)|^2\,d\xi. \tag{1.93}$$

Then (1.88) gives Theorem 1.2.9 for $a=b=0$. As shown in its proof, a change in the function of the form $f\mapsto f(x-c_1)e(c_2x)$ allows to drop the condition $\langle A\rangle=\langle B\rangle=0$ and put general constants.

Suggested Readings. The idea of that Fourier series truncated to frequencies less than δ^{-1} skip details of size δ very seldom appears in mathematical books because it is difficult to state it as a theorem. It is anyway something important to keep in mind in signal processing. In [DS89] there are some mathematical statements that capture the practical idea. For the quantum mechanics part, there are a lot of basic books. My advice for a mathematical reader is to escape from those that barely contain formulas and keep in mind that quantum mechanics outreach literature is very often misleading. I find a notes written by B. Zwiebach [Zwi13] very interesting (and free). An old and interesting book is [LL58]. Perhaps the most pleasant for a hard-core mathematician is [GP90].