1 Quantum mechanics crash course

In 1927, Heisenberg was thinking about quantum mechanics, which was the major scientific breakthrough of the time. By working through some physical ideas, he came up with the following amazing assertion: in quantum mechanics, it is impossible to simultaneously determine a particle's position and momentum to arbitrary precision.

Crucially, this is not a statement about the accuracy of our measurement devices or anything like that. Instead, it's a statement about the physical universe: even with infinitely good measurement devices, the pair (particle's position, particle's momentum) cannot be determined to precision better than some fixed absolute quantity. In other words, our implicit assumption that a particle *has* a well-defined position and momentum is false.

Shortly after Heisenberg put forward this idea, Kennard and Weyl independently found a mathematical formalism for it. Physicists write this as

$$\sigma_x \sigma_p \geq \frac{h}{4\pi} \tag{1}$$

where σ_x, σ_p are the standard deviations (loosely, the error) in the measurement of position and momentum respectively, and h is Planck's constant (a physical constant roughly equal to 6.6×10^{-34} Joule-seconds).

Amazingly, the physical result stems from a purely mathematical theorem, which we will shortly state and prove. But to understand it, we first need to understand something about the formal definitions of position and momentum in quantum mechanics. For simplicity, for the rest of the class, we will only be dealing with particles moving around in one dimension, so that both the position x and the momentum p should be thought of real numbers (rather than vectors).

In quantum mechanics, a particle doesn't have a well-defined position. Instead, there is a function, called the *wave function*, $f : \mathbb{R} \to \mathbb{C}$. The point of the wave function is to define the probability that the particle is at some point x: this probability is postulated to be $|f(x)|^2$. In order to make this make sense, physicists always assume that the wave function f is L^2 *normalized*, meaing that $\int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x = 1$; this just implies $|f(x)|^2$ does define a probability distribution.

Another important idea in quantum mechanics, which I will describe in a somewhat sketchy manner, is that the *momentum* of a particle is also encoded by its wave function. Namely, there is something called the *de Broglie hypothesis* (also known as wave-particle duality) which says that a particle behaves like a wave, and that its momentum p is given by hk, where h is Planck's constant and k is its wavelength. Roughly, what this means is that the momentum of a particle should be determined by how much its wave function oscillates. The formal mathematical way to measure this is with the Fourier transform, which is defined by

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} \, \mathrm{d}x.$$

Then the important idea in quantum mechanics is that $\widehat{f}$ is the wave function for the momentum of the particle. In other words, the probability that the particle has some momentum

p is given by $|\widehat{f}(p)|^2$. There are (at least) two things to say about this. The first is that this probability is well-defined, because of an important fact called *Plancherel's formula*, which says that $\int_{-\infty}^{\infty}|\widehat{f}(\xi)|^2\,\mathrm{d}\xi = \int_{-\infty}^{\infty}|f(x)|^2\,\mathrm{d}x$. The second is that there is a completely separate derivation of this fact—that momentum is given by a Fourier transform—based on ideas related to symmetries of space, preserved quantities, and even Assaf's colloquium on why exponentiating the derivative operator shifts a function.

But in any case, the formal statement of the uncertainty principle is now the following mathematical statement.

Theorem 1.1. *Suppose $f : \mathbb{R} \to \mathbb{C}$ is a function with $\int_{-\infty}^{\infty}|f(x)|^2\,\mathrm{d}x = 1$. Then*

$$\left(\int_{-\infty}^{\infty} x^2|f(x)|^2\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} \xi^2|\widehat{f}(\xi)|^2\,\mathrm{d}\xi\right) \geq \frac{1}{16\pi^2}.$$

Note that if we think of $|f(x)|^2$ as a probability, then $\int_{-\infty}^{\infty} x^2|f(x)|^2\,\mathrm{d}x$ denotes the *variance* of this probability distribution. In other words, the first term measures the variance of the position, and the second measures the variance of the momentum. By taking square roots, we precisely get (1), because the standard deviation of a random variable is defined to be the square root of its variance. Moreover, everything I've done is in "natural units" where $h = 1$, so even the constant $1/16\pi^2$ gets square-rooted to the constant $h/4\pi$ in (1).

Perhaps the most important thing to say about Theorem 1.1 is that it is a purely mathematical statement, simply about the behavior of functions $\mathbb{R} \to \mathbb{C}$ and their Fourier transforms. Moreover, the purely mathematical statement has many other physical implications! Because it turns out that there are many other pairs of "canonically conjugate variables" in quantum mechanics; these are pairs of quantities that are related by the Fourier transform. Examples include angular position and angular momentum, electric charge and voltage, and even time and energy. For every such pair, the mathematical result implies an uncertainty principle for that pair.

1.1 Other uncertainty principles

Since Heisenberg and Kennard–Weyl, people have proved very many other uncertainty principles. In general, an *uncertainty principle* refers to a result of the following type. For some notion of "spread" or "non-localization", and for any non-zero $f : \mathbb{R} \to \mathbb{C}$, at least one of f or $\widehat{f}$ has large "spread" under this notion. In Heisenberg's uncertainty principle, the notion of spread is the variance, which we discussed above. Here are just a few examples.

- There is an uncertainty principle for the *support* of a function, which is the set of points where it is non-zero. Namely, if f is non-zero outside of some interval, then $\widehat{f}$ is not: it is non-zero for all but a countable set of $\xi \in \mathbb{R}$.

- Hardy proved an uncertainty principle for the notion of spread which is "decay at infinity". Informally, he proved that it is not possible for both f and $\widehat{f}$ to decay faster than e^{-x^2} as $x \to \infty$.

- Beckner proved an uncertainy principle for a notion of *entropy* of a function. Entropy is an important information-theoretic notion of spread; a highly localized function has very small (in fact, negative) entropy, while a spread-out function has large entropy.

2 The Fourier transform and L^p norms

Recall that the Fourier transform of a function $f : \mathbb{R} \to \mathbb{C}$ is given by

$$\widehat{f}(\xi) \coloneqq \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}\,\mathrm{d}x.$$

The only property we will need of the Fourier transform, besides its definition, is the *Fourier inversion formula*, which says that we can recover f from $\widehat{f}$ with a very similar formula:

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(\xi)e^{2\pi i x\xi}\,\mathrm{d}\xi.$$

In order to understand the uncertainty principle, we will need to understand L^p norms, which are ways of understanding "how large" a function is.

Definition 2.1. Given a real number $p \geq 1$ and a function $g : \mathbb{R} \to \mathbb{C}$, we define

$$\|g\|_p \coloneqq \left(\int_{-\infty}^{\infty} |g(x)|^p\,\mathrm{d}x\right)^{1/p}.$$

Also, for $p = \infty$, we define

$$\|g\|_\infty \coloneqq \max_{x\in\mathbb{R}} |g(x)|.$$

One can check that for nice (e.g. continuous) functions g, we have that

$$\|g\|_\infty = \lim_{p\to\infty} \|g\|_p,$$

which is why $\|g\|_\infty$ is defined as it is. The point is that, as $p \to \infty$, the norm $\|g\|_p$ is picking up more and more behavior from the "peaks" of g, since we are raising $|g(x)|$ to higher and higher powers. As $p \to \infty$, the contribution from all "non-peaks" tends to 0, and thus $\|g\|_p$ converges to the maximum value of $|g|$.

Notice that ratios of L^p norms give us notions of spread of a function. For example, consider the ratio $\|g\|_1/\|g\|_\infty$. If g is very spread out (e.g. it equals 1 on a very long interval then tapers to 0), then $\|g\|_1$ will be much bigger than $\|g\|_\infty$. On the other hand, if g is very sharply localized, then $\|g\|_1$ will be much smaller than $\|g\|_\infty$. So their ratio is a natural notion of how spread out, or non-localized, g is.

For similar reasons (though a little harder to see visually), ratios like $\|g\|_1/\|g\|_2$ are also notions of spread. Basically, if g has a peak, then the contribution of that peak will be more strongly accentuated in $\|g\|_2$ than it is in $\|g\|_1$. In general, $\|g\|_p/\|g\|_1$ is a measure of spread for any $p > 1$.

Because of this intuition, the following result is an uncertainty principle. In fact, as we will soon see, it is *the* uncertainty principle: from it, we will be able to directly deduce other uncertainty principles such as Heisenberg's.

Theorem 2.2 (Primary uncertainty principle)**.** *For any non-zero $f : \mathbb{R} \to \mathbb{C}$,*

$$\frac{\|f\|_1}{\|f\|_\infty} \cdot \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_\infty} \geq 1. \tag{$*$}$$

Proof. We first claim that

$$\|\widehat{f}\|_\infty \leq \|f\|_1.$$

Indeed, to see this, we note that for any $\xi \in \mathbb{R}$, we have that

$$|\widehat{f}(\xi)| = \left| \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} \, \mathrm{d}x \right| \leq \int_{-\infty}^{\infty} \left| f(x) e^{-2\pi i x \xi} \right| \, \mathrm{d}x = \int_{-\infty}^{\infty} |f(x)| \, \mathrm{d}x = \|f\|_1,$$

using the fact that $\left| e^{-2\pi i x \xi} \right| = 1$ for all $x, \xi \in \mathbb{R}$. Since this bound holds for all $\xi \in \mathbb{R}$, we see that

$$\|\widehat{f}\|_\infty = \max_{\xi \in \mathbb{R}} |\widehat{f}(\xi)| \leq \|f\|_1,$$

as claimed.

For exactly the same reason (plus the Fourier inversion formula), we have that

$$|f(x)| = \left| \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{2\pi i x \xi} \, \mathrm{d}\xi \right| \leq \int_{-\infty}^{\infty} |\widehat{f}(\xi)| \, \mathrm{d}\xi = \|\widehat{f}\|_1$$

which implies that

$$\|f\|_\infty \leq \|\widehat{f}\|_1.$$

Multiplying together our two inequalities, we see that

$$\|f\|_1 \|\widehat{f}\|_1 \geq \|f\|_\infty \|\widehat{f}\|_\infty,$$

which implies the claimed uncertainty principle by dividing both sides by $\|f\|_\infty \|\widehat{f}\|_\infty$. □

3 Deriving other uncertainty principles

Now that we have the primary uncertainty principle ($*$), we can derive other uncertainty principles. The general framework is as follows. Let S be some notion of "spread", which we should think of as an object that takes in a function g and returns a number $S(g)$ which measures how spread g is. Suppose we prove that for all non-zero $g : \mathbb{R} \to \mathbb{C}$, we have

$$S(g) \geq a \left(\frac{\|g\|_1}{\|g\|_\infty} \right)^b$$

for some positive real numbers a, b. Then we can plug this into $(*)$ for $g = f$ and $g = \widehat{f}$, and we conclude that

$$S(f)S(\widehat{f}) \geq \left[a\left(\frac{\|f\|_1}{\|f\|_\infty}\right)^b\right]\left[a\left(\frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_\infty}\right)^b\right] = a^2\left(\frac{\|f\|_1\|\widehat{f}\|_1}{\|f\|_\infty\|\widehat{f}\|_\infty}\right)^b \geq a^2 \cdot 1^b = a^2.$$

So we get an uncertainty principle for S! We conclude that it's not possible for both f and $\widehat{f}$ to have small values of S, or in other words that they can't both be sharply localized.

The crucial thing to observe about this framework is that all the "uncertainty" (or really, the connection between f and $\widehat{f}$) comes from the primary uncertainty principle $(*)$. After that, we only argue about a *single* general function g.

As an example of this technique in action, let's prove an uncertainty principle for the measure of spread $\|g\|_1/\|g\|_p$.

Theorem 3.1. *For any real $p > 1$ and any non-zero function $g : \mathbb{R} \to \mathbb{C}$, we have*

$$\frac{\|g\|_1}{\|g\|_p} \geq \left(\frac{\|g\|_1}{\|g\|_\infty}\right)^{\frac{p-1}{p}}.$$

As a consequence, we have the uncertainty principle

$$\frac{\|f\|_1}{\|f\|_p} \cdot \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_p} \geq 1.$$

Proof. By definition, we have that

$$\|g\|_p^p = \int_{-\infty}^{\infty} |g(x)|^p \,\mathrm{d}x = \int_{-\infty}^{\infty} |g(x)|^{p-1} \cdot |g(x)| \,\mathrm{d}x.$$

By the definition of $\|g\|_\infty$, we have $|g(x)| \leq \|g\|_\infty$ for all $x \in \mathbb{R}$. Therefore, we can bound $|g(x)|^{p-1}$ by $\|g\|_\infty^{p-1}$, and we see that

$$\|g\|_p^p \leq \int_{-\infty}^{\infty} \|g\|_\infty^{p-1} \cdot |g(x)| \,\mathrm{d}x = \|g\|_\infty^{p-1} \int_{-\infty}^{\infty} |g(x)| \,\mathrm{d}x = \|g\|_\infty^{p-1}\|g\|_1.$$

By multiplying both sides by $\|g\|_1^{p-1}$, this is equivalent to

$$\|g\|_1^{p-1}\|g\|_p^p \leq \|g\|_\infty^{p-1}\|g\|_1^p$$

which is in turn equivalent to

$$\frac{\|g\|_1}{\|g\|_p} \geq \left(\frac{\|g\|_1}{\|g\|_\infty}\right)^{\frac{p-1}{p}}.$$

To conclude the uncertainty principle, we follow the template described above. For any non-zero function $f : \mathbb{R} \to \mathbb{C}$, plugging in the above and the primary uncertainty principle $(*)$, we get

$$\frac{\|f\|_1}{\|f\|_p} \cdot \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_p} \geq \left(\frac{\|f\|_1}{\|f\|_\infty}\right)^{\frac{p-1}{p}} \left(\frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_\infty}\right)^{\frac{p-1}{p}} \geq 1 \cdot 1 = 1. \qquad \square$$

4 Heisenberg's uncertainty principle

For a function $g : \mathbb{R} \to \mathbb{C}$, let

$$V(g) = \int_{-\infty}^{\infty} x^2 |g(x)|^2 \, \mathrm{d}x.$$

As discussed, if we think of $|g|^2$ as a probability distribution (in the case that $\int |g|^2 = 1$), then this exactly measures the variance of that probability distribution. Let's restate Heisenberg's uncertainty principle, Theorem 1.1, without the assumption that $\int |f|^2 = 1$.

Theorem 4.1. *There exists an absolute constant $C > 0$ such that for any $f : \mathbb{R} \to \mathbb{C}$, we have*

$$V(f) \cdot V(\widehat{f}) \geq C \|f\|_2^2 \|\widehat{f}\|_2^2.$$

The original proof gives the constant $C = 1/(16\pi^2)$; we will obtain the somewhat worse constant $C = 1/4096$. We will prove this using the template discussed above, by deriving it from the primary uncertainty principle. In fact, it'll be a little more convenient to derive this from the uncertainty principle proven in Theorem 3.1, that

$$\frac{\|f\|_1}{\|f\|_2} \cdot \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2} \geq 1.$$

Specifically, Heisenberg's uncertainty principle follows from the following lemma.

Lemma 4.2. *For any non-zero $g : \mathbb{R} \to \mathbb{C}$, we have*

$$\frac{V(g)}{\|g\|_2^2} \geq \frac{1}{64} \left(\frac{\|g\|_1}{\|g\|_2} \right)^4.$$

Given this lemma, Theorem 4.1 follows immediately. Indeed, by Theorem 3.1, we have that

$$\frac{V(f)}{\|f\|_2^2} \frac{V(\widehat{f})}{\|\widehat{f}\|_2^2} \geq \frac{1}{64^2} \left(\frac{\|f\|_1}{\|f\|_2} \cdot \frac{\|\widehat{f}\|_1}{\|\widehat{f}\|_2} \right)^4 \geq \frac{1}{4096}.$$

The one fact that we will need for the proof of Lemma 4.2 is the Cauchy–Schwarz inequality for integrals. It says that for any two functions $g, h : \mathbb{R} \to \mathbb{C}$ and any $I \subseteq \mathbb{R}$, we have

$$\int_I |g(x)h(x)| \, \mathrm{d}x \leq \left(\int_I |g(x)|^2 \, \mathrm{d}x \right)^{1/2} \left(\int_I |h(x)|^2 \, \mathrm{d}x \right)^{1/2} \leq \|g\|_2 \|h\|_2. \tag{2}$$

One specific consequence of this is if we take h to be the indicator function of some interval $[a, b]$, i.e. the function

$$h(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

Plugging this in to (2), we see that

$$\int_a^b |g(x)|\,\mathrm{d}x \le \sqrt{b-a}\|g\|_2. \tag{3}$$

With these preliminaries, we can prove Lemma 4.2, and thus Theorem 4.1.

Proof of Lemma 4.2 (and thus Heisenberg's uncertainty principle). Fix a function $g : \mathbb{R} \to \mathbb{C}$. Our first goal is to find a real number $T > 0$ so that "most" of the function g "lives outside" the interval $[-T, T]$. Concretely, we want to pick T so that

$$\int_{-T}^{T} |g(x)|\,\mathrm{d}x \le \frac{1}{2}\int_{-\infty}^{\infty} |g(x)|\,\mathrm{d}x,$$

meaning that at least half of the "mass" of g is outside the interval $[-T, T]$.

To do this, we define

$$T = \frac{1}{8}\left(\frac{\|g\|_1}{\|g\|_2}\right)^2.$$

To see that this satisfies the property we want, we use (3), which tells us that

$$\int_{-T}^{T} |g(x)|\,\mathrm{d}x \le \sqrt{2T}\|g\|_2 = \left(\frac{1}{2}\frac{\|g\|_1}{\|g\|_2}\right)\|g\|_2 = \frac{1}{2}\|g\|_1 = \frac{1}{2}\int_{-\infty}^{\infty} |g(x)|\,\mathrm{d}x,$$

where the first equality uses the definition of T.

Using this, we see that

$$\int_{|x|>T} |g(x)|\,\mathrm{d}x = \int_{-\infty}^{\infty} |g(x)|\,\mathrm{d}x - \int_{-T}^{T} |g(x)|\,\mathrm{d}x \ge \frac{1}{2}\int_{-\infty}^{\infty} |g(x)|\,\mathrm{d}x = \frac{1}{2}\|g\|_1.$$

Finally, we will see the point of picking T. The point is that we want to multiply and divide by x in the integral, and then use the Cauchy–Schwarz inequality (2) to split these up, and then use the choice of T to be able to bound the resulting integral of $1/x^2$ that arises. More precisely, continuing our final computation, we see that

$$\begin{aligned}
\frac{1}{2}\|g\|_1 &\le \int_{|x|>T} |g(x)|\,\mathrm{d}x \\
&= \int_{|x|>T} \frac{1}{x}\cdot x|g(x)|\,\mathrm{d}x \\
&\le \left(\int_{|x|>T} \frac{1}{x^2}\,\mathrm{d}x\right)^{1/2} \left(\int_{|x|>T} x^2|g(x)|^2\,\mathrm{d}x\right)^{1/2} \\
&\le \left(\frac{2}{T}\right)^{1/2} \left(\int_{-\infty}^{\infty} x^2|g(x)|^2\right)^{1/2} \\
&= \frac{4\|g\|_2}{\|g\|_1}\cdot\sqrt{V(g)}.
\end{aligned}$$

Rearranging, we find that

$$V(g) \geq \frac{1}{64} \frac{\|g\|_1^4}{\|g\|_2^2},$$

which is equivalent to the desired result by dividing by $\|g\|_2^2$. □

5 Final remarks

This is not the original proof of Heisenberg's uncertainty principle! In fact, it's rather different. The original proof uses certain important and special properties of the Fourier transform. However, the proof we showed is much more general: it implies that the uncertainty principle holds not just for the Fourier transform, but for *any* operator A with the property that $\|Af\|_\infty \leq \|f\|_1$ and $\|f\|_\infty \leq \|Af\|_1$. This proof, as well as the proofs of many other uncertainty principles using the same general framework, comes from the paper "The uncertainty principle: variations on a theme" by Wigderson and Wigderson.