

Lecture 17: The Uncertainty Principle

*Prof. V.M. Gadre, EE, IIT Bombay*

## 1 Introduction

We build in this lecture a very important principle. In fact in some sense the principle that lays at the heart of the subject of Wavelets and Time frequency methods namely the uncertainty principle. We shall devote the whole lecture to a discussion of uncertainty principle laying the foundation of what the uncertainty means first, and then proceeding to obtain certain numerical bounds in two domains simultaneously.

## 2 Non-formal introduction to the idea of containment

As we discussed in previous lecture, there is of course a very tight or strong notion of containment. Is it possible to have compact support in both time domain and frequency domain? So both in the time and in frequency, we demand the function to be nonzero only over a finite part of the independent variable or the real axis. This is a very strong demand and of course we mentioned in the previous lecture that it cannot be met ever. It related to the fact that if we noted that the function was finitely supported(compactly supported) in the real axis, there was certain properties of that function, specifically the existence of an infinite number of derivatives, makes it impossible that the function to be compactly supported or nonzero only over a finite interval of the independent variable in the natural domain. Natural domain can mean time, space or whatever.

But we had asked whether a weaker notion of containment could be admitted. So to speak in some sense, be on the finite interval of the independent variable which index it and simultaneously in the transformed domain, i.e. the frequency domain, we insists that most of the content be in a finite interval of the frequency axis. This seems like more reasonable requirement and to a certain extent this requirement can be met.

We are finally going to come out with certain bounds on how much we can contain in the two domains simultaneously. So there are several steps to reach this destination:

The first step is to put down in a non diffused, in a non formal way, what do we mean by containment? What do we mean by most of the content being in certain finite range? We had also hinted at the approach that we would take to do this briefly in the previous lecture. We had said that there are two ways of doing this. We should think on the magnitude squared of the function and the magnitude squared of the Fourier transform as a one dimensional object and then we could talk about the centre of that object or “centre of mass”. We could talk about the spread of the object around the centre of mass, using the notion of “radius of gyration” or probability density, built from the squared magnitude of the function and another density built from the squared magnitude of the Fourier transform. We could then look at the “mean” of these densities and “variance” of these densities. The variances are indicative of the spread. So this was a non formal introduction.

### 3 Formalization of the idea of containment

Now we need to formalize this. We are going to work in  $L_2(\mathbb{R})$ . It is always going to be the space of square integrable functions. In fact, we must mention that sometimes we are actually going to work in the intersection of the space of square integrable functions and absolutely integrable functions.

$$x(t) \in L_2(\mathbb{R}) \cap L_1(\mathbb{R})$$

Now, as the function belongs to  $L_2(\mathbb{R})$ , its Fourier transform also belongs to  $L_2(\mathbb{R})$ . Let,  $x(t)$  have the fourier transform  $\hat{x}(\Omega)$ . Then,  $\hat{x}(\Omega) \in L_2(\mathbb{R})$  as well. So, we first define a density or a “one dimensional mass”.

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \|x(t)\|_2^2$$

which is finite. Therefore, we define the density as,

$$p_x(t) = \frac{|x(t)|^2}{\|x(t)\|_2^2}$$

which is a probability density, because of the following reasons:

1.  $p_x(t) \geq 0 \forall t$  (It is a density in  $t$ ).
2.  $\int_{-\infty}^{\infty} p_x(t) dt = 1$  (from definition).

Similarly, let us define a density in the angular frequency domain.

$$p_{\hat{x}}(\Omega) = \frac{|\hat{x}(\Omega)|^2}{\|\hat{x}(\Omega)\|_2^2}$$

This is also a probability density, because of the following reasons:

1.  $p_{\hat{x}}(\Omega) \geq 0 \forall \Omega$  (It is a density in  $\Omega$ ).
2.  $\int_{-\infty}^{\infty} p_{\hat{x}}(\Omega) d\Omega = 1$  (from definition).

Now we have taken the probability density perspective, but we could as well take the so called one dimensional mass perspective, i.e., we could think of the  $p_x(t)$  as a one dimensional mass in  $t$  and similarly  $p_{\hat{x}}(\Omega)$  as one dimensional mass in  $\Omega$ . So, here, we have a simplified situation. We have a mass in one dimensional space. That one dimensional space can be the space of  $t$  or the space of  $\Omega$ .

If we choose the “mass” perspective, consider the “center of mass” and the “spread around the center”. Spread around the center in mechanics can be measured by a quantity called the “Radius of gyration”. If we choose the “probability density” perspective, consider the “mean” and “variance”.

Now we must assume that these quantities can be calculated and we shall do that. It is possible that the variance can be infinity. So we are not always guaranteed of a finite variance. We are trying to find a lower limit to where these quantities go in the two domains simultaneously.

Considering the function  $x(t)$ , we prefer to take the probability density perspective. So we think of  $p_x(t)$  and  $p_{\hat{x}}(\Omega)$  as the probability densities and now we shall write down the “mean”. Let,  $p_x(t)$  have the mean  $t_0$ .

$$t_0 = \int_{-\infty}^{\infty} t p_x(t) dt$$

We will of course recognize the definition to hold good for the “center of mass” here. Essentially, we are calculating the movement by choosing the fulcrum to be zero and therefore getting a

different fulcrum or a point at which the movements are balanced. Similarly, let  $p_{\hat{x}}(\Omega)$  have the mean  $\Omega_0$ .

$$\Omega_0 = \int_{-\infty}^{\infty} \Omega p_{\hat{x}}(\Omega) d\Omega$$

Once, if have the mean, we assume the “means” are finite and normally they should be. In some pathological situations, we may have a problem. We are not looking at those pathological situations. So, assuming these “means” are finite, let us look at the “variances”.

So, the variance in  $t$  is defined as,

$$\sigma_t^2 = \int_{-\infty}^{\infty} (t - t_0)^2 p_x(t) dt$$

And similarly, the variance in angular frequency is defined as,

$$\sigma_{\Omega}^2 = \int_{-\infty}^{\infty} (\Omega - \Omega_0)^2 p_{\hat{x}}(\Omega) d\Omega$$

Once again, we are assuming the variances to be finite. In any case, here we don't have such a problem. Even if the variances are infinite, we will accept it. If we choose to think these as one dimensional masses, it is very clear that the variance is an indication of the spread. So larger the variance, the more the density said to have spread around the “mean” and the smaller the variance, the more the density or the mass is said to be concentrated. So, now we have a formal way to define containment.

We can say that the containment in a given domain refers to the variance in that domain. So containment in time is eventually  $\sigma_t^2$  and containment in angular frequency domain is essentially  $\sigma_{\Omega}^2$  quantity. How small can we make any one of these quantities for a valid function? In a few minutes, we will be convinced that there is no limit for this!

In fact we will take the Haar scaling function as an example and calculate it's the variance.

**Example:** Calculate mean and variance for the Haar scaling function.

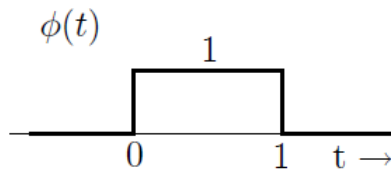


Figure 1: Haar scaling function

We can see that the Haar scaling function  $\phi(t)$  is one between zero and one and zero elsewhere. Its probability density is given as,

$$p_{\phi}(t) = \frac{|\phi(t)|^2}{\|\phi(t)\|_2^2}$$

Now,

$$\|\phi(t)\|_2^2 = \int_{-\infty}^{\infty} |\phi(t)|^2 dt = \int_0^1 1 dt = 1$$

Hence,  $p_{\phi}(t)$  is drawn as, Now, we will find the “mean”. In fact, even before finding the mean

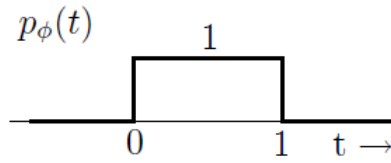


Figure 2: Probability density function in time

formally, we can find it graphically. The mean is going to be at the centre of 0 and 1 i.e, at  $\frac{1}{2}$ . Let us do it formally,

$$\begin{aligned}
 t_0 &= \int_{-\infty}^{\infty} t p_{\phi}(t) dt \\
 &= \int_0^1 t dt \\
 &= \frac{t^2}{2} \Big|_0^1 \\
 &= \frac{1}{2}
 \end{aligned}$$

Hence, mean is shown as,

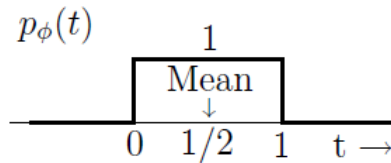


Figure 3: Mean of  $p_{\phi}(t)$

Now, we will find the “variance”.

$$\begin{aligned}
 \sigma_t^2 &= \int_{-\infty}^{\infty} (t - t_0)^2 p_{\phi}(t) dt \\
 &= \int_{-\infty}^{\infty} (t - \frac{1}{2})^2 p_{\phi}(t) dt \\
 &= \int_0^1 (t - \frac{1}{2})^2 dt
 \end{aligned}$$

Let,  $t - \frac{1}{2} = \lambda$ ,  
 $\Rightarrow dt = d\lambda$ .

As, limits of  $t$  are 0 to 1, we get, limits of  $\lambda$  are  $\frac{-1}{2}$  to  $\frac{1}{2}$ . Hence, integral becomes,

$$\sigma_t^2 = \int_{\frac{-1}{2}}^{\frac{1}{2}} \lambda^2 d\lambda$$

$$\begin{aligned}
&= \frac{\lambda^3}{3} \left| \frac{1}{2} \right|^{-\frac{1}{2}} \\
&= \frac{1}{3} \left( \frac{1}{8} + \frac{1}{8} \right) \\
&= \frac{1}{12}
\end{aligned}$$

Therefore, taking positive square root, we get,

$$\sigma_t = \frac{1}{2\sqrt{3}}$$

As we can be seen that,  $\sigma_t$  is less than  $\frac{1}{2}$ . In a certain sense, we don't really use the number half to denote the spread of  $\phi(t)$  around its mean. The variance doesn't say it goes all the way to half. It says the spread is a number slightly less than half. Most of the energy is contained in that region around the mean captured by the variance. In fact, if we are very specific, the fraction of the energy contained here would be, i.e. the energy contained in  $[t_0 - \sigma_t, t_0 + \sigma_t]$  would eventually be given by,

$$\int_{t_0 - \sigma_t}^{t_0 + \sigma_t} p_\phi(t) dt$$

We are not looking for 100%. We are considering the significant part of it. Now we will calculate this value. Substituting the values, the integral becomes,

$$\begin{aligned}
\int_{t_0 - \sigma_t}^{t_0 + \sigma_t} p_\phi(t) dt &= \int_{\frac{1}{2} - \frac{1}{2\sqrt{3}}}^{\frac{1}{2} + \frac{1}{2\sqrt{3}}} 1 dt \\
&= \left( \frac{1}{2} + \frac{1}{2\sqrt{3}} \right) - \left( \frac{1}{2} - \frac{1}{2\sqrt{3}} \right) \\
&= \frac{1}{\sqrt{3}} \\
&= 0.577
\end{aligned}$$

It is certainly not a large fraction like 90%, but it is more than 50%. This fraction is not going to be the same for all functions. It depends on the density. Hence, we can say that, the variance is one accepted measure of spread. And very often the variance actually tells us where most of the function is concentrated. Even in this case, if we look at it carefully, quite a bit of this function is between  $(\frac{1}{2} - \frac{1}{2\sqrt{3}})$  and  $(\frac{1}{2} + \frac{1}{2\sqrt{3}})$ .

Now we will calculate the variance in the frequency domain of this same function. So, let us look at  $\hat{\phi}(\Omega)$ . Actually, we are interested in  $|\hat{\phi}(\Omega)|^2$ . And that is of the form,

$$|\hat{\phi}(\Omega)|^2 = \left| \frac{\sin(\frac{\Omega}{2})}{(\frac{\Omega}{2})} \right|^2$$

We could integrate this. In deed we know that,

$$\|\phi(t)\|_2^2 = \int_{-\infty}^{\infty} |\phi(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{\phi}(\Omega)|^2 d\Omega$$

which is equal to be 1. Hence,

$$\|\hat{\phi}(\Omega)\|_2^2 = \int_{-\infty}^{\infty} |\hat{\phi}(\Omega)|^2 d\Omega = 2\pi$$

Hence,  $p_{\hat{\phi}}(\Omega)$  is given as,

$$p_{\hat{\phi}}(\Omega) = \frac{|\hat{\phi}(\Omega)|^2}{\|\hat{\phi}(\Omega)\|_2^2} = \frac{|\hat{\phi}(\Omega)|^2}{2\pi}$$

It has an appearance like,

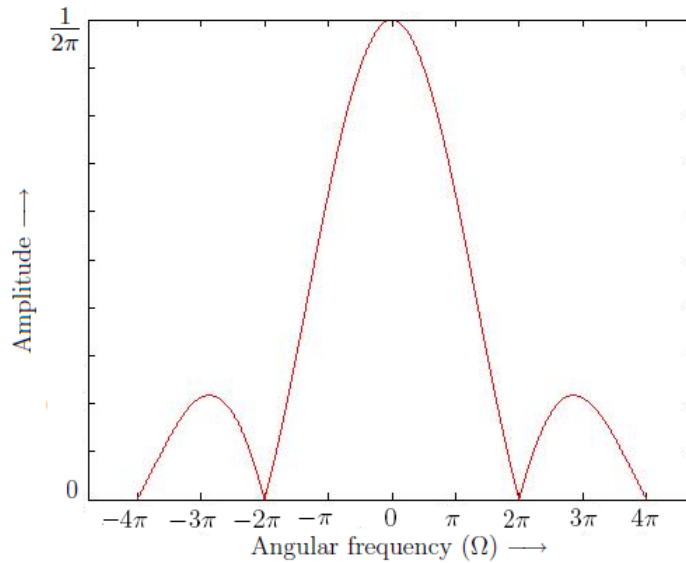


Figure 4: Waveform of  $p_{\hat{\phi}}(\Omega)$

Now, it is very easy to see that the mean of this function is zero. This function is symmetrical around  $\Omega = 0$ . For all real functions  $x(t)$ ,  $\hat{x}(\Omega)$  is magnitude symmetric. Therefore the mean  $\Omega_0 = 0$ .

Now, let us find the variance. So, the variance of  $\hat{\phi}(\Omega)$  is given as,

$$\begin{aligned} \sigma_{\Omega}^2 &= \int_{-\infty}^{\infty} (\Omega - \Omega_0)^2 p_{\hat{\phi}}(\Omega) d\Omega \\ &= \int_{-\infty}^{\infty} \Omega^2 \frac{|\hat{\phi}(\Omega)|^2}{2\pi} d\Omega \\ &= \int_{-\infty}^{\infty} \frac{\Omega^2}{2\pi} \left| \frac{\sin(\frac{\Omega}{2})}{(\frac{\Omega}{2})} \right|^2 d\Omega \\ &= \int_{-\infty}^{\infty} \frac{4}{2\pi} \left| \sin\left(\frac{\Omega}{2}\right) \right|^2 d\Omega \end{aligned}$$

Here, we are in serious trouble. The constant  $\frac{4}{2\pi}$  is not important, but  $|\sin(\frac{\Omega}{2})|^2$  is very important. We are trying to integrate the  $|\sin(\frac{\Omega}{2})|^2$  function.

$|\sin(\frac{\Omega}{2})|^2$  is a periodic function with period  $2\pi$ . We are trying to integrate a periodic function from  $-\infty$  to  $+\infty$ , and obviously that integral is going to diverge. So the fear that we had when

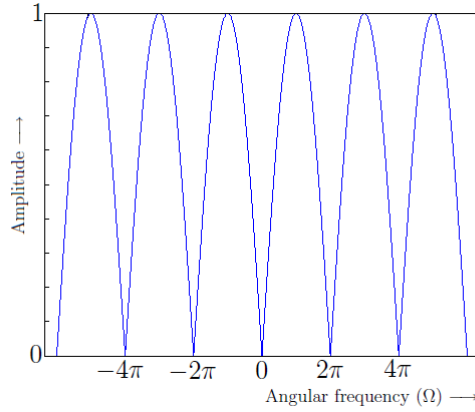


Figure 5: Waveform of  $|\sin(\frac{\Omega}{2})|^2$

we started the discussion comes out to be true right in the very simple case of scaling function that we know. The variance of  $\hat{\phi}(\Omega)$  is infinite! In other words,  $\phi(t)$  is not at all confined in the frequency domain, at least in this sense. All this while in our discussion, when we talked about time and frequency together and so on, in the previous lecture we had been worried about these side lobes. Besides it is all right to look at the main lobe and talk about the presence in the main lobe. But then we have these side lobes and the side lobes are falling off by the factor of  $\frac{1}{\Omega}$  in magnitude. As we can see, the side lobes have created the problem after multiplication of  $\Omega^2$  in the calculation of variance. The side lobe creates a periodic function to be integrated, and we are in trouble.

So, this tells us again why we have to much beyond the Haar. We have been asking again and again, why we can't content with the Haar multi-resolution analysis. Now we have one more formal answer, if we look at the scaling function in the Haar multi-resolution analysis, its variance in the frequency domain analysis is infinite. It is not at all contained in the frequency domain in this sense. Now, it is a natural question to ask, what is it make the variance infinity? Why did we have a divergent variance? In fact we can answer these questions, if we only care to make a slight adjustment of the expression of variance. The variance of  $\hat{\phi}(\Omega)$  is given as,

$$\begin{aligned} \sigma_{\Omega}^2 &= \int_{-\infty}^{\infty} \Omega^2 \frac{|\hat{\phi}(\Omega)|^2}{\|\hat{\phi}(\Omega)\|_2^2} d\Omega \\ &= \frac{1}{\|\hat{\phi}(\Omega)\|_2^2} \int_{-\infty}^{\infty} \Omega^2 |\hat{\phi}(\Omega)|^2 d\Omega \\ &= \frac{1}{\|\hat{\phi}(\Omega)\|_2^2} \int_{-\infty}^{\infty} |j\Omega\hat{\phi}(\Omega)|^2 d\Omega \end{aligned}$$

Now,  $j\Omega\hat{\phi}(\Omega)$  has some meaning. It is essentially the Fourier transform of  $\frac{d\phi(t)}{dt}$ . Hence, energy in the derivative function is given as,

$$\int_{-\infty}^{\infty} \left| \frac{d\phi(t)}{dt} \right|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |j\Omega\hat{\phi}(\Omega)|^2 d\Omega$$

Hence, the variance of  $\hat{\phi}(\Omega)$  is given as,

$$\text{Variance of } \hat{\phi}(\Omega) = \frac{2\pi(\text{Energy in derivative})}{\|\hat{\phi}(\Omega)\|_2^2}$$

$$= \frac{2\pi(\text{Energy in derivative})}{2\pi(\text{Energy in function})}$$

For a real  $x(t)$ , the frequency variance, i.e. the  $\Omega$  variance is,

$$\sigma_{\Omega}^2 = \frac{\text{Energy in } \frac{dx(t)}{dt}}{\text{Energy in function } x(t)} = \frac{\|\frac{dx(t)}{dt}\|_2^2}{\|x(t)\|_2^2}$$

And now we have the answer for the trouble! As we can see,  $\phi(t)$  is discontinuous. So, when its derivative is considered, there are impulses in the derivative. And impulses are not square integrable. So, the numerator diverges. The moment we have a discontinuous function, we have an infinite frequency variance. With this note, we realize that, if we want to get some meaningful uncertainty, some meaningful bound, we must at least consider continuous functions.