The Lecture Contains:

Data Analysis

- Classification of Data
- Analysis of Random Signals
- Fourier Transform Technique
- Probability Density Function Approach

Previous Next

Classification of data

Data received by an observer from an experimental setup can be classified as in Figure 1.7. Methods of analyzing deterministic data are well-established because the data is already in a form, from which integral measures can be extracted. When periodic signals are encountered it is a conventional practice to present results for sinusoidal signals alone. This is because results for a general periodic signal can be constructed from those for harmonic signals using Fourier decomposition of the form

$$u(t) = \sum_{j=1}^{\infty} A_j e^{2\pi i j t/T}, \quad i = \sqrt{-1}$$

Here T is the time period of the signal u and the Fourier coefficients A_j satisfy the condition $\lim_{j\to\infty} |A_j| \to 0$. The coefficients can be determined from the formula

$$A_j = \frac{1}{T}\int_0^T u(t) e^{-2\pi i j t/T} dt$$

As an example, pressure drop in a pipe carrying pulsatile flow can be determined as a weighted average of the individual pressure drops occurring in sinusoidally varying flows whose frequencies are integer multiples of that of the real problem.

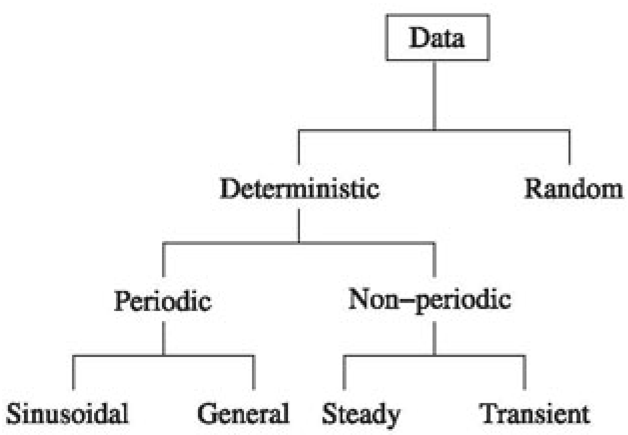

Figure 1.7: Classification of Data

When the data available to the observer is random, one is forced to use statistical techniques. This is because even when a mean value is determinable, one requires prior knowledge of the length of the signal to be considered for averaging. This mean value can subsequently be used for deterministic analysis. However in many applications information the randomness itself may be desired and statistical measures of the signal will have to be calculated.

Previous Next

Analysis of Random Signals

Consider a random signal sensed by a probe and recorded by a measuring system as shown in Figure 1.8. Though it is impossible to collect this signal manually, modern instruments can collect such a signal with considerable amount of accuracy in both magnitude and time. The simplest quantities that must be determined from this signal are the mean, the RMS value and the cross-correlation with a second signal $v(t)$. The relationships between the signal and the reduced quantities are given by:

Mean: $\bar{u} = \lim_{T\to\infty} \frac{1}{T}\int_o^T u(t)dt$

RMS: $u' = \lim_{T\to\infty} [\frac{1}{T}\int_o^T (u-\bar{u})^2 dt]^{1/2}$

Cross-correlation: $\overline{-u'v'} = -\lim_{T\to\infty} [\frac{1}{T}\int_o^T (u-\bar{u})(v-\bar{v})dt]$

If the signals are directly stored in the memory of a computer in digital form, the integrals appearing above can be evaluated using computer programs.

For example, the numerical evaluation of an integral proceeds as

$$\frac{1}{b-a}\int_a^b f(t)dt = \frac{1}{N}(f_1 + f_2 \dots f_N)$$

or,

$$\frac{1}{b-a}\int_a^b f(t)dt = \frac{1}{2(N-1)}(f_1 + f_N + 2(f_2 + f_3 \dots f_{N-1}))$$

In (i) and (ii) $f_i = f(t_i)$ where

$$t_i = a + (\frac{b-a}{N-1})(i-1)$$

Previous Next

Equation (i) is called the ensemble average of f. Equation (ii) is based on a linear fit for f between successive points. Equation (i) is usually preferred over (ii) because it under-predicts the value of the integral and de-emphasizes the occurrence of isolated peak values.

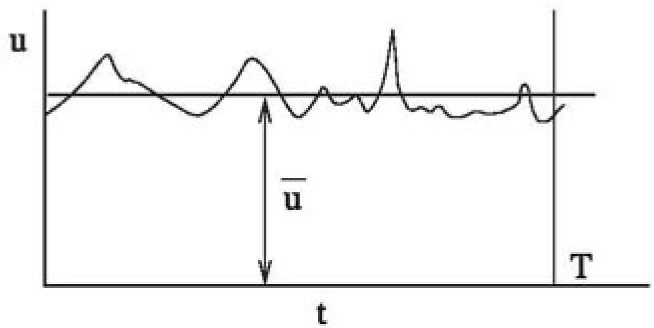

Figure 1.8: Sketch of a Random Signal.

The integrals appearing above are quite sensitive to the choice of T (and hence N), namely the total time of data collection. While the desired value of T can be estimated by increasing the length of the signal till the mean and RMS values converge to values that are independent of T , this is not very convenient. One can estimate the value of T from the physics of the problem being studied. Consider atmospheric flow as an example. There will be fluctuations arising from the motion of eddies of a variety of sizes that are transported by the flow. The fluctuation due to a typical eddy is schematically shown in Figure 1.9.

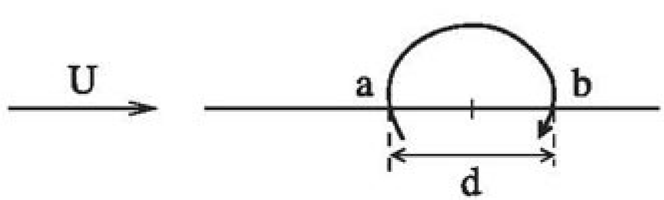

Figure 1.9: Model of an Eddy to Define Time Scales in the Flow.

Previous Next

At a given instant, point `a' moves up while point `b' moves down. After a time interval d/U these directions are reversed. Hence $2d/U$ is a measure of a time period of fluctuation of an eddy of size of d and $U/2d$ is the associated frequency. A distribution of eddy sizes now means that there exists a distribution of frequencies as well. In a boundary-layer, $d \leq \delta$ where δ is the boundary-layer thickness and the largest value of the time period can be estimated conservatively as $2\delta/U$. In flow past a cylinder d may be chosen as the cylinder diameter; in flow past a mesh the grid size or the wire diameter whichever is larger can be used as an estimate of d. As a rule of thumb the integration time T should be 5 to 10 times the characteristic time period $2d/U$.

Other quantities that are frequently required in the study of stationary random signals with a zero mean value are the autocorrelation and power spectrum. These are defined below.

Autocorrelation: $Ru(\tau) = \lim_{T\to\infty} \int_o^T u(t)u(t+\tau)dt$

Power spectrum: $E(\omega)$: $E(\omega)d\omega$ is the fraction of the kinetic energy present in the frequency interval $(\omega, \omega + d\omega)$.

The largest value of Ru occurs when $\tau = 0$. For larger values of τ, u is only partly correlated with itself and in general as $\tau = 0$, $Ru \to 0$. Signals for which $\lim_{\tau\to\infty} Ru \to$ finite and non-zero are said to be coherent since two widely separated events on the time scale continue to bear a relationship to each other. The quantity

$$Tu = \int_0^{\infty} \frac{Ru(\tau)}{(u')^2}\, d\tau$$

is called the integral time scale and is a measure of the time period over which the signal is correlated with itself. The total time T for which the signal is acquired should be larger than Tu, so that the statistics are meaningfully evaluated.

Fourier Transform Technique

The complex function $\hat{u}(\omega)$ is defined as the Fourier transform of u and is calculated as

$$\hat{u}(\omega) = \int_{-\infty}^{\infty} u(t) e^{-i\omega t} dt, \quad i = \sqrt{-1}$$

The normalized power spectrum can then be calculated as

$$E(\omega) = (|\hat{u}(\omega)|^2)/(u')^2$$

where u' is the RMS value of $u(t)$. It is possible to show that $Ru(\tau)$ and $E(\omega)$ from a Fourier transform pair, i.e.

$$Ru(\tau) = \int_{-\infty}^{\infty} E(\omega) e^{i\omega t} d\omega$$

$$E(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Ru(\tau) e^{-i\omega t} d\tau.$$

Methods of calculating Fourier transforms are well-established. In particular, the fast Fourier transform (FFT) algorithm has found wide usage both in software and in hardware applications in signal processing. Hence it is to be understood that integrals appearing in the Fourier transforms defined above can be readily determined.

Though the integrals given above are complex-valued, the property $Ru(\tau) = Ru(-\tau)$ guarantees that $E(\omega)$ is purely real. On the other hand, the Fourier integral for Ru is to be interpreted as the real part of the complex function. Typical autocorrelation functions and power spectrum are sketched in Figure 1.10.

Previous Next

For a sinusoidal signal the power spectrum exhibits a peak at the signal frequency f. This suggests a method of measuring frequency of sinusoidal signals and dominant frequencies of non-sinusoidal periodic signals. White noise is defined as a signal whose amplitude at a given instant is purely a random variable within with certain limits. Hence the signal is correlated with itself when $\tau = 0$ and uncorrelated for all $\tau > 0$.

The cross-correlation function $R_{12}(\tau)$ for a pair of signals f_1 and f_2 is defined as

$$R_{12}(\tau) = \int_{-\infty}^{\infty} f_1(t) f_2(t+\tau) p(t) dt$$

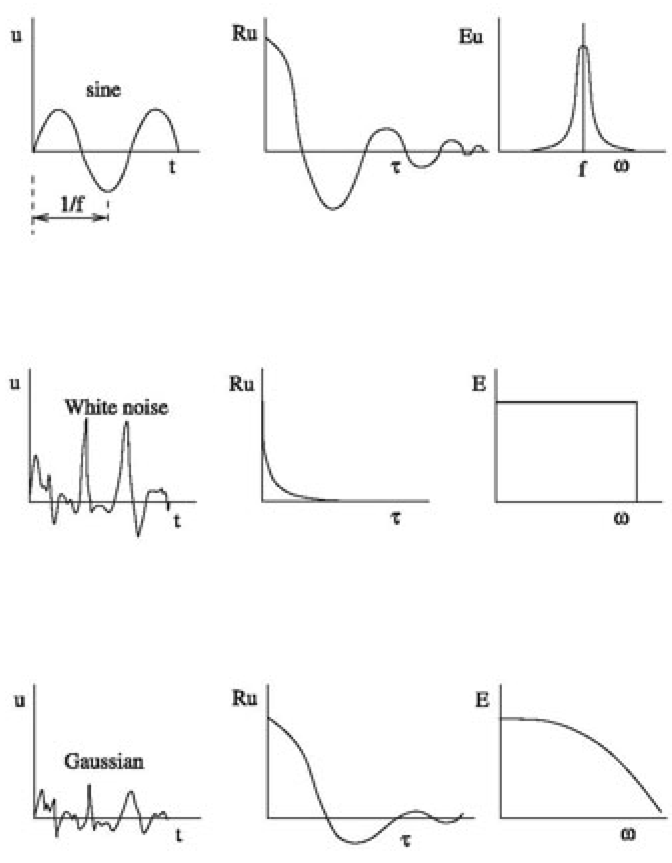

Figure 1.10: Examples of Autocorrelation and Power spectrum.

Here $p(t)$ is a band-limited function that is zero if $|t| > T$ where T is a prescribed large value.

For correlated signals R_{12} has information regarding the phase difference between f_1 and f_2. For example, if f_1 and f_2 are harmonic with phase difference ϕ, R_{12} is also harmonic with a starting value that depends on ϕ. It is thus a convenient measure of the phase difference itself. If f_1 and f_2 are random signals ϕ will be a function of the frequency variable ω. In such problems we work with the cross-spectral function $E_{12}(\omega)$ defined as

$$E_{12}(\omega) = \ FT\ (R_{12}(\tau)).$$

In wave propagation problems, E_{12} has information regarding $\phi(\omega)$ and hence the wave number distribution and the wave speeds as a function of frequency. In nonlinear dynamics this is further interpreted in terms of appearance of coherent structures.

The following results can be easily derived.

If $f_1 = \sin\omega_o t$ and $f_2 = \sin(\omega_o t \pm \phi)$, $R_{12}(\tau) = 1/2\cos(\omega_o \tau \pm \phi)$ and $E_{12}(\omega)$ is a delta function centred at $\omega = \omega_o$. One can calculate ϕ from $R_{12}(0)$. In a travelling wave problem, f_1 and f_2 may be two signals obtained from probes separated by a distance X. The wave number is then given as $k = \phi/X$. If f_1 is white noise and $f_2(t) = f_1(t+T)$, i.e. f_2 is a time shifted form of f_1, one can show that $R_{12}(\tau) = 2\pi S_o \delta(\tau - T)$ where S_o is the power spectrum of f_1 (and a constant if f_1 is white noise). Subsequently it is easy to show that $E_{12} = S_o \exp(-i\omega T)$ and $|E_{12}| = S_o$.

Previous Next

Probability Density Function Approach

The probability density function (PDF) of a signal $u(t)$ denoted by the symbol Bu, is defined as follows: $Bu(\bar{u})du$ is the fraction of the total time spent by $u(t)$ between the levels $\bar{u}$ and $\bar{u} + du$. See Figure 1.11

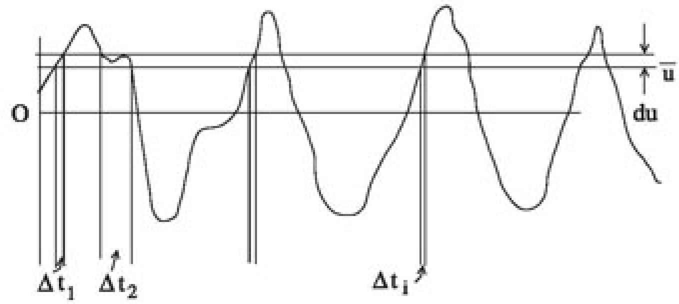

Figure 1.11: Definition of Probability Density Function.

Thus,

$$Bu(u) = \lim_{T \to \infty_{\Delta u \to 0}} \frac{1}{T\Delta u} \sum_{i=1}^{N} \Delta t_i$$

Clearly $Bu \geq 0$ and $\int_{-\infty}^{\infty} Budu = 1$. The quantity $Bu(0)$ is called *zero crossing probability.* It indicates the time spent by the signal around time $t = 0$.

A Gaussian signal is one whose probability density function has a Gaussian profile. Such signals have a finite range of values of time lag over which the autocorrelation Ru is non-zero. Additionally, the power spectrum, interpreted as the harmonically decomposed kinetic energy, is spread over a range of frequencies. The *central limit theorem* of probability theory is worth recalling in this connection. This theorem states that a large number of identically distributed independent variables will together have a Gaussian probability density function regardless of the shape of the density of the variables themselves. Signals in homogeneous, stationary turbulent flow that exhibit equilibrium between energy production and dissipation are known to exhibit a Gaussain probability density function . Hence deviation from Gaussain behaviour can be used as a measure of deviation from equilibrium itself.

The shape of a Gaussian PDF for a zero-mean signal $u(t)$ is given by the formula

$$Bu(u) = \frac{1}{\sqrt{2\pi}\,u'} e^{-u^2/2(u')^2}$$

and is sketched in Figure 1.12. Here, u' is the RMS value of $u(t)$

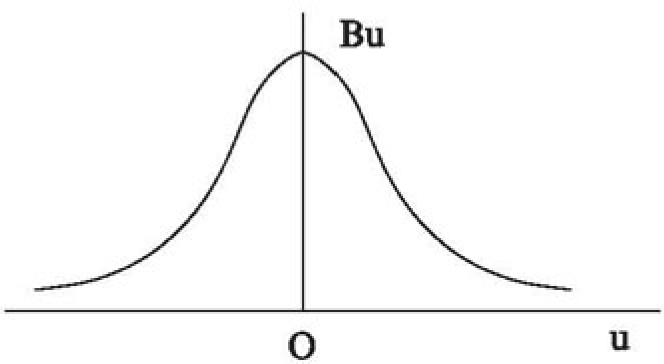

Figure 1.12: An Example of Gaussian PDF.

Previous Next

In the definitions given in the previous slide for quantities such as u and E the signal $u(t)$ is available in digital form and stored in a computer. Many of these integrals can instead be evaluated in terms of the probability density function Bu of the signal $u(t)$. The advantages of this approach are:

1. Bu can be determined using hardware (instruments) and,

2. Bu is a usually a smooth function of its argument and hence integrals involving Bu can be accurately calculated by high order numerical integration formulas.

However, the difficulty of having a long enough signal for $u(t)$ is now transferred to waiting for a long enough time to determine Bu. The accuracy with which Bu is measured depends on the choice of the window Δu and total time T.

In general, for small values $\Delta u,$ a large value of time T is required for satisfactory convergence of the limit process arising in the definition of Bu. In terms of Bu the mean and RMS values are defined as follows:

$$Mean: \bar{u} = \int_{-\infty}^{\infty} uBu(u)du$$

$$RMS: \qquad u' = [\int_{-\infty}^{\infty} u^2 Bu(u)du]^{1/2}$$

For signals that do not have a zero mean

$$u' = [\int_{-\infty}^{\infty} (u-\bar{u})^2 Bu(u-\bar{u})du]^{1/2}$$

The nth order moment of a signal with a zero mean is defined as

$$u^n = \lim_{T\to 0} \frac{1}{T} \int_0^{\infty} u^n dt = \int_{-\infty}^{\infty} u^n Bu(u)du$$

The second order moment of u'^2, namely the mean square. The third moment $(n=3)$ is called the skewness factor. This $u(t)$ is zero for a Gaussian signal. The fourth moment $(n=4)$ is called flatness factor or Kurtosis. Note that as n increases the accuracy with which Bu is determined for large u becomes critical. Typical examples where the skewness and flatness factors are respectively large as shown in Figure 1.13.

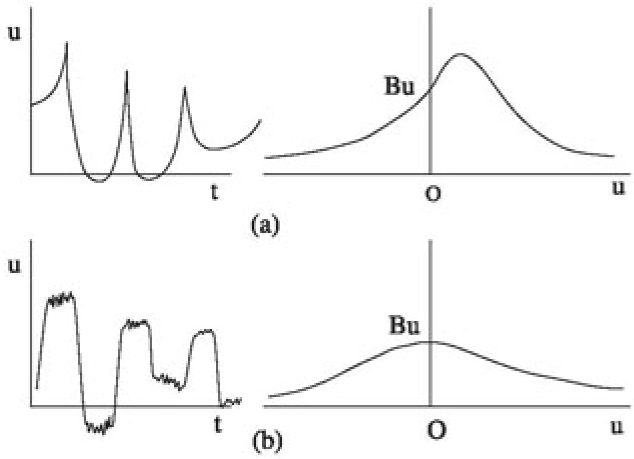

Figure 1.13: Signals with Large Skewness (a) and Large Flatness (b).

The cross correlation $\overline{u'v'}$ is determined in terms of PDF as

$$\overline{u'v'} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uvB(u,v)dudv$$

where $B(u,v)$ is called the joint probability density function. It is defined as the fraction of the time for which $u(t)$ lies between u and $u+du$ and $v(t)$ between v and $v+dv$ simultaneously. Autocorrelation can be determined in terms of $B(u,v)$ by identifying $v(t)$ as $u(t+\tau)$.

Note that the PDF approach evaluates integrals in the amplitude domain alone. In comparison, the autocorrelation function represents time-domain statistics; the spectra are descriptors of the flow field in the frequency domain.