

Chapter 3 Measurement Systems, Statistical and Error Analysis

3.1 Statistics

3.1.1 Background of Statistics

3.1.1.1 Probability

General Concept

Major Parameters of Probability Distributions

3.1.1.2 Parameter Estimation and Hypothesis Testing

Point Estimation

Hypothesis Testing Using Confidence Intervals

3.1.1.3 Random Process

Concept of Random Process

Ergodic Process and Averaging

3.1.1.3 Correlation Analysis

Correlation Constants and Functions

Transfer Functions

3.1.2 Regression

3.1.2.1 General Methods, Linear Regression with the Least Square

Simple Regression

Multiple Regression

3.1.2.2 Nonlinear Regression

Linearity

Nonlinear Models

Non-linear Least Squares, Higher Order Functions

3.1.2.3 Empirical Formulation

Concept of Empirical Equation

Similitude Analysis vs. Regression

Iteration

Optimization on Parameter Estimation

3.1.2.4 Orthogonal Principles

Single Variable

Multiple Variables

General Orthogonality Principle

3.2 Error Analysis

3.2.1 General Concept of Errors

3.2.1.1. Random and Uncertainty

3.2.1.2 Sources of Error

3.2.1.3 Definition of Error

3.2.2. Error Estimate & Criteria

3.2.2.1 Systematic Error and Averaging

3.2.2.2 Measurement Bias

3.2.2.3 Propagation of Uncertainties

3.2.2.4 Final Uncertainty of Single-Sample Measurement

3.2.2.5 Final Uncertainty of Single-Sample Measurement, RSS Method

3.2.2.6 Step-by-Step Analysis

3.2.3 Error Control

3.2.3.1 Rejection of Questionable Data

3.2.3.2 Random Error with Zero Mean and Averaging

Arithmetic Average

Simplified Arithmetic Average

Geometric Average

Weighted Average

3.2.3.3 Moving Average and Windowing

Moving Average

Windows and Window Functions

3.2.3.4 Separation of Measurement Signals From Systematic Errors

Non-stationary to Station

Filtering

3.1 Statistics

In this chapter we start to introduce data processing in terms of statistical studies and error analysis. There is another important portion of data processing including the time and the frequency domain analyses, which will be discussed in a later chapter. Such an arrangement is tailored for the time schedule of one semester study of college students with the consideration of theoretical studies in the classroom and experimental tests in the lab. The main idea of such an arrangement is to allow the student the earliest possible “hands-on” laboratory experience.

3.1.1 Background of Statistics

In the following, we shall review necessary mathematical background in experimental mechanics.

3.1.1.1 Probability

1. General Concept

Measured values are likely to be random, because they are likely to be influenced by certain random and/or unpredictable factors. A deterministic number + a random number results in a random number.

What is a random variable, and what is the relationship between randomness and predictability?

Randomness and Probability

A specific event (not necessarily a physical quantity) is *random*.

1) the exact value of such quantity is likely unpredictable

- 2) The position of a certain valued variable is unpredictable. That is, it is not known where (in which order or in what position) that variable will appear.
- 3) How often a certain valued quantity can appear, the exact frequency of the variable appearance is unknown.

From the above requirement, it is understandable that random is

- 1) Not necessarily unpredictable
- 2) Not necessarily uncontrollable

Therefore, we must seek a certain rule to handle random variables. However, to mathematically deal with random events, we must denote them by numbers (under a certain rule), we cannot use names, letters, symbols other than numbers for calculation.

Some of the events can be quantified or are already quantified. For example, the measurement of the length of angle steel, such as 20', 20.5', 20.2', 19.6' etc. is already quantified.

Some events, originally, are not quantified. For example, dropping a coin with equal chances of "tails" or "heads"; Eggs in supermarket marked as "large", "medium" or "small"; A political poll on somebody's performance, "excellent", "acceptable", "unacceptable", etc.

However, we need to calculate the probability of these events, it is not convenient to use words "good" or "OK" or "bad", It is better to mark these events with certain numbers. If some of the random events have already been quantified, then we can use them.

Alternatively, we can also mark them by new numbers. For example, length of angle steel: Let the length between 19.5'-19.75' be marked by 1; length between 19.75'-20.0' be marked by 2; length between 20.0'-20.25' be marked by 3; length between 20.25'-20.5' be marked by 4. Length greater than 20.5' be marked by 5.

We call them **Random Variables**, more precisely,

Discrete Random Variables

For example: n random variables, x_1, x_2, \dots, x_n

What we try to study on these random variables is, first, the probability of each variable, or the chance of appearance of a variable x_i . Denoted by $P(x_i)$.

Continuous random number

Numbers that are random but continuous is different from the above-mentioned discrete number.

The following nomenclature are often to use:

Nomenclature

Population comprises the entire collection of objects, measurements, observations, whose properties are under consideration.

A *Sample* is a representative subset of a population on which an experiment is performed and numerical data are obtained.

The set of all possible outcomes of an experiment is called the *Sample space*

A *Distribution function* $P(x)$ is a graphical or mathematical relationship of how often a certain valued random variable will appear, the relationship is used to represent the values of random variable, see figure 3.1-1 (c) and (d).

Probability $p(x)$ is a function of random variable x , *probability density function*. See figure 3.1-1 (a) and (b).

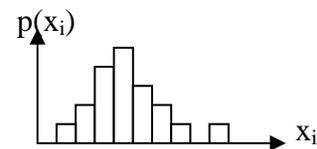


Figure 3.1-1(a)

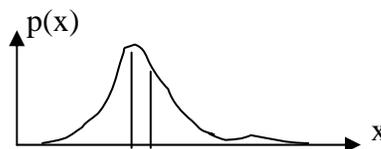


Figure 3.1-1(b)

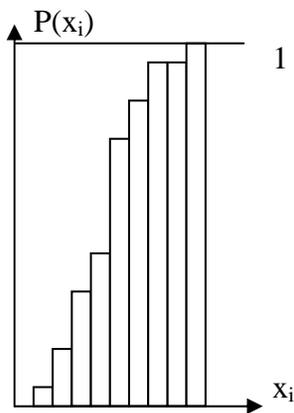


Figure 3.1-1(c)

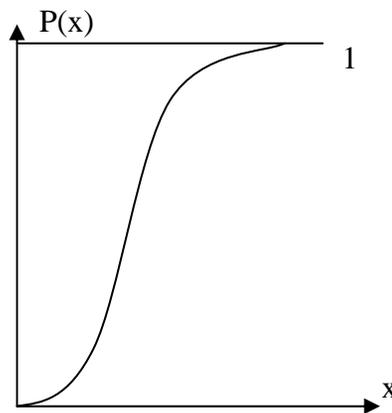


Figure 3.1-1 (d)

$P(x)$, $P(x_i)$ are distribution functions possessing the following properties:

$$P(x) = \int_{-\infty}^x p(x) dx \quad (3.1-1)$$

$$p(x) \geq 0 \quad (3.1-2)$$

$$\int_{-\infty}^{\infty} p(x) dx = 1 \quad (3.1-3)$$

$$P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} p(x) dx = P(x_2) - P(x_1) \quad (3.1-4)$$

A *Parameter* is a numerical attribute of the entire population.

An *Event* is the outcome of a random experiment.

A *Statistic* is a numerical attribute of the sample.

Probability is the chance of occurrence of an event in an experiment.

Uncertainty Affected by unpredictable factors. Estimation of error in a measurement or in a result, usually determined with a certain level of confidence (often 95%)

Confidence Level: Probability that a random variable x lies in a special interval.

Random Process: A sequence of random variables

Correlation Analysis: Measure of how well a curve fits a set of data. A value of 1 indicates perfect relationship and value of 0 indicates no relationship.

2. Major Statistical Parameters for Probability

Human are smart fools. We are fools, because we don't know why and how a random event occurs. Does randomness exist in the real world? Einstein says God does not play with dice. Humans play, however. A man will likely not know many points he will get in the next play, anyway.

Humans, however, are smart. They try to use deterministic numbers to describe random events. That is, they try to use a fixed number to denote a probability distribution. Generally speaking, using knowns to approach unknowns, using deterministic to approach random, (using linear to approach non-linear) are very powerful methodologies, and, in many cases, are the only way to go.

Mathematically, to dig out deterministic properties from random variable rely on a very useful method, finding the mean of random variables. To calculate the mean value is to find a certain average value. That is, it is assumed that although we will not know the features of each individual variable, we will realize the average behavior of them. It is assumed that for a given population, its mean value is deterministic, which is referred to the mathematical expectation. Another important average is how an individual variable is separated from the mean, which is also assumed to be a deterministic value.

The most important deterministic parameters are listed as follows

$$\text{Mean } x_m = E(x) = [\sum_{i=1}^n (x_i)]/n \quad (3.1-5a)$$

Sample mean, x_m . Population mean μ . Mathematical expectation

$$\text{Deviation } d_i = x_i - x_m \quad (3.1.6)$$

$$\text{Variance } \text{Var}(x) = [\sum_{i=1}^n (d_i)^2] / n = E [(x - E(x))^2] \quad (3.1-7a)$$

$$\text{Standard Deviation } \sigma = \sqrt{\text{Var}(x) = [\sum_{i=1}^n (x_i - x_m)^2] / n}^{1/2} \quad (3.1-8a)$$

As mentioned before, the measure of obtaining a mean value is a powerful methodology. Thus, among the above equations, (3.1-5a) is very essential. The simple equation will not only be used to calculate the mean value of a random variable, but for much broader applications. That is, the variable x can be used not only to represent a set of random variables taken from samples from a population, but also to represent any possible values, quantities, events, sequences, process, functions, etc.

For example, in (3.1-7), it represents the value of $(x_i - E(x))^2$. Equation (3.1-5a) can be rewritten as

$$E(x) = \sum_{i=1}^n [1/n (x_i)]$$

Which implies that each variable x_i has an equal chance $1/n$ to be summarized for the measure of averaging. It is seen that the number $1/n$ is the probability of the appearance of variable x_i .

In the case of continuous variables, the formula of calculating the mean value becomes

$$E(x) = \int_{-\infty}^{\infty} p(x) x \, dx \quad (3.1-5b)$$

Which also indicates the way of summarizing variable x with its own chance or probability $p(x)$ of appearance for the measure of averaging.

Again, the continuous variable x can be used not only to represent a set of random variable taken from samples from a population, but also to represent any possible values, quantities, events, sequences, process, functions, etc. In this sense, we have the variance and standard derivation as follows:

$$\text{Var}(x) = \int_{-\infty}^{\infty} p(x) [x - E(x)]^2 \, dx \quad (3.1-7b)$$

$$\sigma = \sqrt{\text{Var}(x)} \quad (3.1-8b)$$

2. Probability Distributions with Engineering Applications

As mentioned before, one of the important ways to handle random variables is to study the relationship between random variables and the probability of how these variables can appear. Such a relationship is assumed to be a deterministic function, the distribution function. In the following, we shall take a look at few most commonly used distributions in engineering applications.

Binomial Distribution:

Discrete variables that can only have two possible outcomes, “0” and “1”, bad, good, fail, success, etc. can have the following distribution, (see figure 3.1-1a).

$$P(x) = [n!/(x! (n-x)!)] p^x(1-p)^{n-x} \quad (3.1-9)$$

Where P, the probability of finding exactly x success in a total n trial
p, probability of success in a given population.

From (3.1-9), we can calculate the mean and the standard deviation as follows:

$$x_m = np \quad (3.1-10)$$

$$\sigma = [np(1-p)]^{1/2} \quad (3.1-11)$$

Gaussian (Normal) Distribution

From figure 3.1-1(b), when $n \rightarrow$ infinity, it can be proven that

$$\lim_{n \rightarrow \infty} P\{a < (x-np)/(np(1-p))^{1/2} \leq b\} = \int_a^b 1/\sqrt{2\pi} \exp(-x^2/2) dx \quad (3.1-12)$$

(3.1-12) defines a certain probability of a random variable in the range of ($a < x \leq b$) such that,

$$P(a < x \leq b) = \int_a^b 1/\sqrt{2\pi} \exp(-x^2/2) dx = \int_a^b f(x) dx \quad (3.1-13)$$

f(x) here is referred to as the probability of the standard normal distribution, with the mean and standard deviation as listed below:

$$x_m = 0, \sigma = 1, \quad (3.1-14)$$

Generally, we have the following mean and standard deviation of a normal distribution:

$$x_m = \mu, \sigma = \sigma, \quad (3.1-15)$$

In this circumstance, we have

$$f(x) = 1/\sqrt{2\pi} \{ \exp[-(x-\mu)^2/2\sigma^2] \} \quad (3.1-16)$$

Figure 3.1-2 shows the normal distribution with different standard distributions:

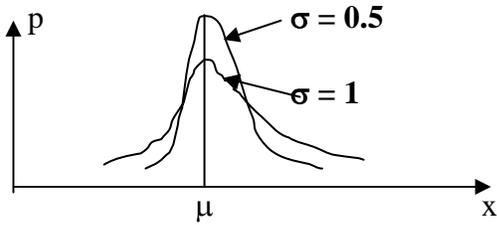


Figure 3.1-2 Normal distribution

Standard Normal Distribution Z-Distribution

It is more convenient to have a standard distribution. To do so, we need to move the mean from μ to zero, we also need to normalize the standard deviation. That is, using the following variable to standardize a general normal distribution

$$Z = (x - \mu) / \sigma \quad (3.1-17)$$

In this case, we have

$$P(a' < z \leq b') = \int_{a'}^{b'} 1/\sqrt{2\pi} \exp(-z^2/2) dz = \int_{a'}^{b'} \phi(z) dz \quad (3.1-18)$$

where

$$\phi(z) = 1/\sqrt{2\pi} \{ \exp[-z^2/2] \} \quad (3.1.19)$$

Normal distribution is very important. The above example of binomial test indicated that the corresponding distribution became normal when the number n became large. This is actually a fundamental idea of the deterministic nature of mean or average value of random populations.

In practice, we have many other types of distribution functions at our disposal, for example, the following:

Student's t Distribution

(Gosset) It is continuous, symmetric, used for analysis of the variation of sample mean value for experimental data with sample size less than 30. For sample sizes greater than 30, student's t approaches normal distribution.

X^2 Distribution

It is continuous, nonsymmetrical distributions, used for analysis of variance of samples in a population, for example, consistency of chemical reaction time are of prime importance in some industrial process, and X^2 Distribution is used for their analysis. This distribution is also used to determine goodness of fit of a distribution for a particular application.

3.1.1.2 Parameter Estimation and Hypothesis Testing

1. Point Estimation

Point estimation is concerned with how to arrive at the parameters that are deterministic numbers. Again, it is essential to understand that the population mean x_m and variance σ^2 are constants.

Assumption: The sample that we have is good (unbiased), we then have

$$E(x_m) = \mu \quad (3.1-20)$$

Standard error of x_m , denoted by S_e , is defined as follows:

$$S_e = \sigma/\sqrt{n} \quad (3.1-21)$$

Unbiased estimate can be described by the following:

U is unbiased estimator of θ if

$$E(U) = \theta \quad (3.1-22)$$

An estimator U is called biased if the above equation does not hold. Then,

$$\text{Bias} = E(U) - \theta \quad (3.1.23)$$

How “good” is an estimator is can be defined by the *Efficient estimator* (minimum variance), which provide how close the estimation to a mean value.

Relative efficiency of two unbiased estimators are explained as follows:

$$\text{Efficiency of V compared to W} = \text{var}(W)/\text{var}(V) \quad (3.1.24)$$

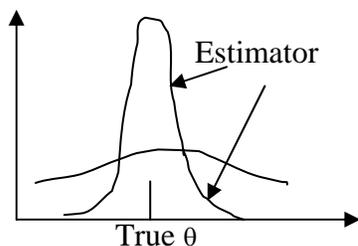


Figure 3.1-3 Estimation

Efficiency of biased and unbiased estimators can be further described by

$$\text{Mean squared error} = E(V-q)^2 = (\text{variance of estimator}) + (\text{its bias})^2 \quad (3.1-25)$$

2. Hypothesis Testing Using Confidence Intervals

An experimenter finds several parameters different from the existing ones while examining his test data. Can the specific set of data represent the true parameters of the population concerned?

An experimenter finds a new phenomenon while examining his test data. Can the specific set of data represent an unknown but realistic relationship?

First he makes a certain hypothesis, then he uses statistic tool to test the hypothesis. In this section, we discuss the first case. For examples, some samples are taken from a population with normal distribution, can we make sure that the mean of the population is μ_0 ? Two independent samples are taken from two populations, do we have a way to tell these two populations possess identical means and deviations?

Confidence Interval

Confidence interval, usually taken to be 95% can be explained as follows:

First, a *statistical hypothesis* is a claim about a population that can be put to a test by drawing a random sample. A *confidence interval* may be regarded as just the set of acceptable hypotheses. Recall standard normal distribution: $Z = (x - \mu) / \sigma$. Using the standard normal distribution implies that we assume the point estimated, or the hypothesis we made actually has normal distributions. How valid the hypothesis is can be examined by the probability obtained from this distribution.

Suppose the mean we calculated from a sample is x_m with the standard error S_e . We now can use the Z-distribution

$$Z = (x_m - \mu) / S_e \quad (3.1.26)$$

with a certain confidence level to test our hypothesis.

To do so, first, from the standard normal distribution, we can find the relationship between a variable z and the corresponding probability from the following formula (see figure 3.1-4).

$$P(|x| > z_{a/2}) = a, \quad (3.1.27)$$

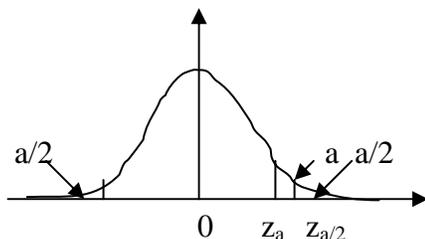


Figure 3.1-4 Confidence interval

For example, see figure 3.1-5,

$$P(|x| > z_{0.05/2}) = 0.05 \rightarrow z_{0.05/2} = 1.96$$

The commonly used confidence level 95% can then be described:

$$P(\mu - 1.96S_e < x_m < \mu + 1.96S_e) = 95\% \quad (3.1.28)$$

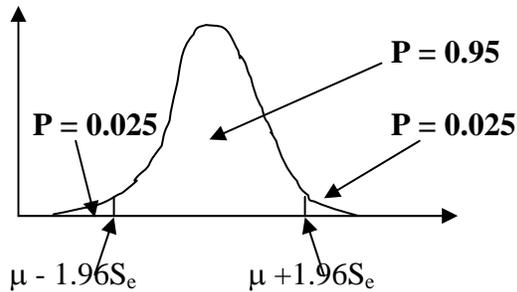


Figure 3.1-5 Confidence level

t Testing

The above procedure to test if a hypothesized mean x_m is a true mean μ is standardized so that the standard normal table can be used. The key statistic of that is

$$Z = (x_m - \mu)/S_e \rightarrow Z = Z = (x_m - \mu)/\text{exact } S_e \quad (3.1-28)$$

which implies that the standard deviation is known. However, quite often, σ is unknown and σ is also estimated with the sample standard deviation s . Then the statistic is called t instead of Z ;

$$t = (x_m - \mu)/\text{estimated } (S_e) \quad (3.1-29)$$

The value of t can be found from existing tables, too.

In a latter section, we shall use the Student's t again to examine measured data.

3.1.1.3 Random Process (Stochastic Process)

In this section, we shall introduce the concept of random process. In certain textbooks, it is also called stochastic process. As with the discussion of random variables, we use deterministic quantities and functions to describe a random sequence. We shall first define the distribution functions, and then shall learn major statistics of a random process. In this sense, the computation of means is still a powerful tool.

In contrast to the study of a random variable, which is a point quantity, a random process describes an entire process that contains many points. Each point may have different statistics from others, that is, the statistics are likely a function of the variable that indicates the position or location or point. In order to cope with this kind of functions, the concept of correlation become quite important, which describes how close two sets of variables are related. However, correlation analysis does not provide how they are related, which is the major topic of regression analysis and will be discussed in the next section.

1. Concept of Random Process

In the above, we have learned properties of random variables. These variables are quantities that define certain values at certain points, or certain states. It may be regarded as static quantities, that is, quantities at point A, or B.

Another kind phenomena are process, that is, a process from A to B along with a certain path. To describe all the events happens along the path, we need first the following definitions:

A sequence of random variables. A good example is the Time sequence, it is really a dynamic process, not just discrete static points.

$X = X(t)$ is random process with time set $(t_1, t_2, \dots t_m)$

At any $t = t_i$, $X(t_i)$ is a random variable

Distribution Function of Random Process

The distribution function of $X(t_1)$, is understandably a function of t_1 . Denoted by

$$P_1(x_1, t_1) = P\{X(t_1) \leq x_1\} \quad (3.1-30)$$

If there exists a function $p_1(x_1, t_1)$ such that

$$P_1(x_1, t_1) = \int_{-\infty}^{x_1} p_1(x_1, t_1) dx_1 \quad (3.1-31)$$

then, $p_1(x_1, t_1)$ is the 1^{st} order probability density function of random process $X(t)$

Similarly, we can define the distribution function with order m as follows:

$$P_m(x_1, x_2, \dots x_m, t_1, t_2, \dots t_m)$$

And probability density function with order m is

$$p_m(x_1, x_2, \dots x_m, t_1, t_2, \dots t_m)$$

Then, we have

$$P_m(x_1, x_2, \dots, x_m, t_1, t_2, \dots, t_m) =$$

$$\int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_m} p_m(x_1, x_2, \dots, x_m, t_1, t_2, \dots, t_m) dx_1 \dots dx_m$$

2. Statistics of Random Process

Similar treatment as for random events, we attempt to find deterministic parameters for random process, such as the mean, variance, etc. Note that, they are generally functions of time t .

Mean:

$$\mu_X(t_1) = E[X(t_1)] = \int_{-\infty}^{\infty} x_1 p_1(x_1) dx_1 \quad (3.1-32)$$

It is important to understand that the averaging treatment $E[X(t_1)]$ is the average of all the samples of $X(t)$ at time t , it is referred to as the *set average*, which is not the *time average*

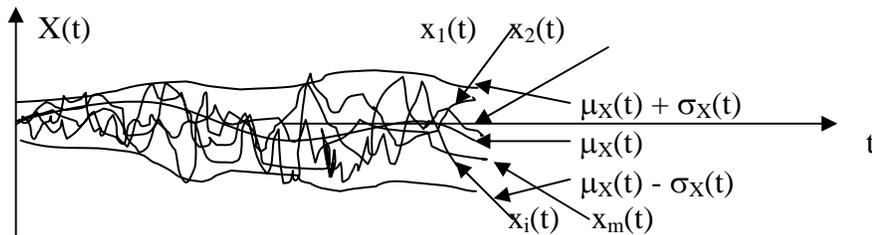


Figure 3.1-6 Set average

The following are a set of definitions based on auto- and cross- correlation functions:

Autocorrelation:

The *autocorrelation* $R(t_1, t_2)$ of $X(t)$ is the expected value of the product $X(t_1)X(t_2)$:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_2(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (3.1-33)$$

The *average power* of $X(t)$ is the value of $R(t_1, t_2)$ when $t_1 = t_2 = t$, that is

$$\Psi_X^2(t) = E[X(t)^2] = R_X(t, t) \quad (3.1-34)$$

The *auto-covariance* $C_X(t_1, t_2)$ of $X(t)$ is the covariance of $X(t_1)$ and $X(t_2)$:

$$C_X(t_1, t_2) = E\{[X(t_1) - \mu_X(t_1)][X(t_2) - \mu_X(t_2)]\} \quad (3.1-35)$$

then

$$C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1) \mu_X(t_2) \quad (3.1-36)$$

Variance $\sigma_X^2(t)$

$$\sigma_X^2(t) = C_X(t, t) = E\{ [X(t) - \mu_X(t)]^2 \} = R_X(t, t) - \mu_X(t)^2 \quad (3.1-37)$$

The *standard deviation* is

$$\sigma_X(t) = (\sigma_X^2(t))^{1/2} \quad (3.1-38)$$

Cross Correlation Function:

Two random process $X(t)$, $Y(t)$ with respectively time sets (t_1, t_2, \dots, t_m) and $(t_1', t_2', \dots, t_n')$

The order of $m + n$ random variables $[X(t_1), X(t_2), \dots, X(t_m); Y(t_1'), Y(t_2'), \dots, Y(t_n')]$

have joint distribution function with order $m + n$:

$$P_{m,n}(x_1, x_2, \dots, x_m, t_1, t_2, \dots, t_m; y_1, y_2, \dots, y_n, t_1', t_2', \dots, t_n')$$

And joint probability density function with order $m + n$:

$$p_{m,n}(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n, t_1, t_2, \dots, t_m, t_1', t_2', \dots, t_n')$$

Second order of $p_{m,n}$ has important applications, denoted by $p_{xy}(x; y; t_1, t_2)$

$$P_{XY}(x, y, t_1, t_2) = P\{ X(t_1) \leq x, Y(t_2) \leq y \} = \int_{-\infty}^x \int_{-\infty}^y p_{xy}(x, y, t_1, t_2) dx dy$$

The following figure shows conceptually the probability of $P\{ X(t_1) \leq x, Y(t_2) \leq y \}$

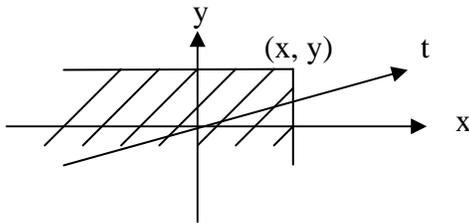


Figure 3.1-7 Joint Probability

We then have *cross-correlation function* defined as follows:

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y p_{xy}(x, y; t_1, t_2) dx dy \quad (3.1-39)$$

And *cross-covariance* $C_{XY}(t_1, t_2)$ for $X(t)$ $Y(t)$ is

$$C_{XY}(t_1, t_2) = E \{ [X(t_1) - \mu_X(t_1)] [Y(t_2) - \mu_Y(t_2)] \} \quad (3.1-40)$$

If for any t_1 and t_2

$$C_{XY}(t_1, t_2) = 0 \quad (3.1-41)$$

$X(t)$ and $Y(t)$ are not correlated.

From the above set of definitions, it is understandable that the correlation is nothing but a certain mean of product of two sets of variables. Thus, to find correlation is a procedure of average. On the average, if these two sets of variables are somewhat related, or not closely related, the value of correlation functions should have different indications.

3. Ergodic Process and Averaging

One of the random processes, called stationary process, is important and commonly found in engineering applications, which can be described as follows:

Generally, time t can be taken at different moments t_1, t_2, \dots, t_n , The random variable with dimension n

$[X(t_1), X(t_2), \dots, X(t_n)]$ has distribution function denoted by

$$P_n(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\} \quad (3.1-42)$$

Which is the n^{th} order distribution function of random process $X(t)$.

1st order distribution can only describe the phenomenon at each discrete point in time, it cannot represent the relationship between different time points.

The higher the order of the distribution function, the better will be the results of description of different time states.

Stationary Process at any n of t_1, t_2, \dots, t_n and any time duration, ε

$$P_n(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) = P_n(x_1, x_2, \dots, x_n, t_1 + \varepsilon, t_2 + \varepsilon, \dots, t_n + \varepsilon) \quad (3.1-43)$$

It is difficult to examine if a process is stationary by the above definition. But, it is easy to see in practical experiments

Now, let us consider the 1st order probability density of a stationary process

let $\varepsilon = -t_1$ then

$$p_1(x_1, t_1) = p_1(x_1, t_1 + \varepsilon) = p_1(x_1, 0) \quad (3.1-44)$$

That is, the 1st order probability density of a stationary process is independent to time t. We can denote it by

$$p_1(x_1)$$

It implies that the mean of X(t) must be constant, denoted by μ_X , we can have

$$E[X(t)] = \int_{-\infty}^{\infty} x_1 p_1(x_1) dx_1 = \mu_X \quad (3.1-45)$$

Similarly, it is understandable that the variance and standard deviation of X(t) are also constant:

So, the characteristics of stationary process:

Mean is constant,

Variance and standard deviation are also constant.

Correlation function has single variable $\tau = t_2 - t_1$

Ergodic Process

It is seen that, from above discussion, to find the statistics of a stationary process through the aforementioned equations, we must know the corresponding probability density, which is very difficult to determine.

If we try to approximate the mean and correlation function by using summations, we still need a lot of samples $x_i(t)$, from a stationary process by means of repeatedly observe the process. It is still quite difficult. Note that, statistic parameters are generally functions of time t, that is, they depend on t.

However, for a stationary process, its mean, variance and standard deviation are also constant, they do not depend on t. Can we than use time average, instead of set average? If so, we can have the statistics from only ONE sample of time history. This is why the concept of ergodicity is so important

Before, introduction of ergodic process, let us examine the concept of

Integration for a Random Process

Let X(e, t) be a random process in the duration $a \leq t < b$, where the parameter e is a random variable

If a certain e_i is given, then the integration

$$Y(e_i) = \int_a^b X(e_i, t) dt \quad (3.1-46)$$

is fully deterministic.

However, for different e_i , the value of $Y(e_i)$ will be different, thus,

$Y(e)$ is a random variable.

We can extend this concept to a weighted integration as follows:

$$Y(t) = \int_a^b X(\lambda) h(\lambda, t) d\lambda \quad (3.1-47)$$

Where the weighting function $h(\lambda, t)$ is a deterministic function. In this way, we have a new random process. The mean of Y is

$$E[Y] = E\left[\int_a^b X(t) dt\right] = \int_a^b E[X(t)] dt \quad (3.1-48)$$

Note that, exchanging the order of calculations of integration and finding the mean can usually be done in many cases, generally speaking.

Similarly

$$E\left[\int_a^b X(\lambda) h(\lambda, t) d\lambda\right] = \int_a^b E[X(\lambda)] h(\lambda, t) d\lambda \quad (3.1-49)$$

We now define *time average*

$$\langle X(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T X(t) dt \quad (3.1-50)$$

and define *time correlation function*

$$\langle X(t) X(t + \tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T X(t) X(t + \tau) dt \quad (3.1-51)$$

With the help of above definitions, we further define the ergodic process:

$$a) \text{ If } \langle X(t) \rangle = E[X(t)] = \mu_X, \quad (3.1-52)$$

The mean of $X(t)$ is ergodic

$$b) \text{ If } \langle X(t) X(t + \tau) \rangle = E[X(t) X(t + \tau)] = R_X(\tau) \quad (3.1-53)$$

The auto-correlation function of $X(t)$ is ergodic. Especially, when $\tau = 0$, the average power is ergodic

Most stationary process are ergodic. But not all of them are.

If a random process is not stationary, it is not ergodic. Then, we cannot use time average to replace the set average. Otherwise, in most times, it will introduce large errors.

One practical way to judge if we can do it is to examine if the average can converge to a certain value. In a later section, we shall introduce a way to change non-stationary processes into stationary ones, which can then be treated as an ergodic process.

3.1.1.4 Correlation Analysis

Correlation analysis is a good means of studying stationary, especially ergodic process, in terms of finding their statistics, as mentioned above.

Correlation analysis is a powerful tool to separate useful signals from measurement noises.

The Fourier and Laplace Transforms of correlation functions are various power spectra, which are essential in signal processing.

In the following, we shall first discuss the nature of correlation functions, then, learn the power spectra for stationary process and finally apply this knowledge to an investigation of the relationship between the input and output signals of linear systems, i.e., the transfer functions.

Correlation Constants and Functions

a) Correlation coefficient

Suppose one has measured two random variables x and y in a set of n data pairs:

$$[(x_i, y_i), i = 1, \dots, n]$$

The *correlation coefficient* (is a constant number used to determine if there in fact exists a functional relationship between the two measured variables x and y . The linear correlation coefficient is defined as follows:

$$\rho_{xy} = [\sum_{i=1}^n (x_i - x_m)(y_i - y_m)] / \{[\sum_{i=1}^n (x_i - x_m)^2 \sum_{i=1}^n (y_i - y_m)^2]\}^{1/2} \quad (3.1-54)$$

$$\rho_{xy} = E[(x-x_m)(y-y_m)] / [\sigma_x \sigma_y] \quad (3.1-55)$$

Where σ_x and σ_y are standard deviations, (recall equation (3.1-8))

The value of ρ_{xy} will lie in between -1 and $+1$. The value of $+1$ indicates a perfect linear relationship between y and x with a positive slope. The value of -1 also indicates a perfect linear relationship between y and x but with a negative slope. The value of 0 indicates no linear relationship between y and x at all, mathematically (see figure 3.1-6). In real measurement, however, even there is no linear relationship between y and x , the value of ρ_{xy} will very unlikely be exactly zero, although it will be very small.

b) Characteristics of Correlation Functions:

The correlation functions, unlike the correlation coefficient, which is a constant number

used to measure the relationship between the two random variables, are used to measure the statistics of random process.

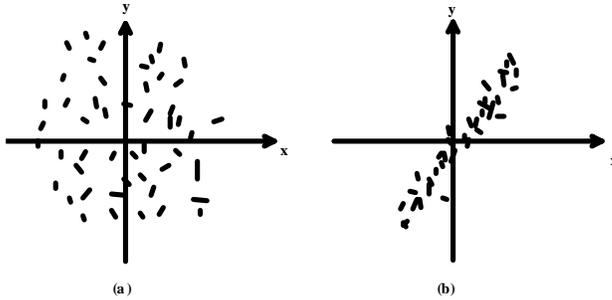


Figure 3.1-7 Correlation between y and x

At the beginning of this section, we have already defined these correlation functions. In the following, we discuss some important features of them.

Suppose $X(t)$ and $Y(t)$ are stationary processes. $R_X(\tau)$, $R_Y(\tau)$ and $R_{XY}(\tau)$ are respectively the auto-correlation and cross-correlation functions of X and Y . Here, since the correlation function has just a single variable of time, we have $\tau = t_2 - t_1$. We now have:

$$1) R_X(0) = E[X^2(t)] = \psi_X^2 = \sigma_X^2 + \mu_X^2 \geq 0 \quad (3.1-56)$$

$R_X(0)$ represents the average power for a stationary processes

$$2) R_X(-\tau) = R_X(\tau), \quad (3.1-57)$$

That is, $R_X(\tau)$ is an even function of τ , see figures 3.1-8 and 3.1-9.

$$3) R_{XY}(-\tau) = R_{YX}(\tau) \quad (3.1-57)$$

Note that $R_{XY}(\tau)$ is neither an even nor an odd function

$$4) |R_X(\tau)| \leq R_X(0) \quad (3.1-58)$$

$$|C_X(\tau)| \leq C_X(0) = \sigma_X^2 \quad (3.1.59)$$

Both become maximum at $\tau = 0$; (a certain auto-correlation function can become maximum only at $\tau = 0$)

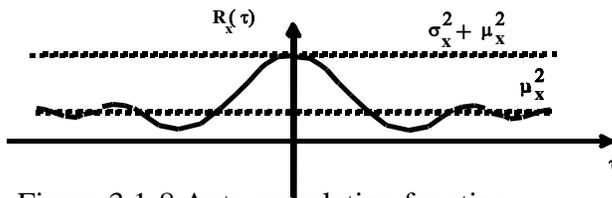


Figure 3.1-8 Auto-correlation function

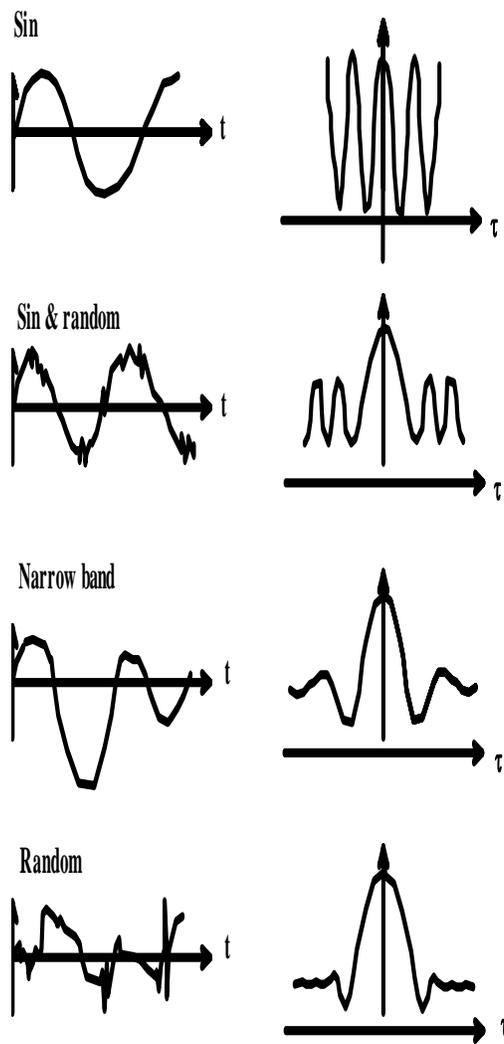


Figure 3.1-9 Auto-correlation functions of different type signals

Similarly, we have

$$|R_{XY}(\tau)|^2 \leq R_X(0) R_Y(0) \quad (3.1-60)$$

$$|C_{XY}(\tau)|^2 \leq C_X(0) C_Y(0) \quad (3.1-61)$$

Define *standard auto-covariance function*

$$\rho_X(\tau) = C_X(\tau)/C_X(0) \quad (3.1-62)$$

$$|\rho_X(\tau)| \leq 1 \quad (3.1-63)$$

Define *standard cross-covariance function*

$$\rho_{XY}(\tau) = C_{XY}(\tau) / [C_X(0) C_Y(0)] \quad (3.1-64)$$

$$|\rho_{XY}(\tau)| \leq 1 \quad (3.1-65)$$

Note that, when $\rho_{XY}(\tau) = 0$, the random process X and Y are not correlated, they are *independent*.

For example, denote records of earthquake ground motions measured at different locations, and/or different earthquakes at the same location, to be X and Y. Then one can determine the correlation between the two records.

$$5) R_X(\tau) \geq 0 \quad (3.1-66)$$

Where the symbol “ \geq ” stands for positive semi-definite, that is, to any time set (t_1, t_2, \dots, t_n) and any function $g(t)$, the following are always true:

$$\sum_{i,j=1}^n R_X(t_i - t_j) g(t_i) g(t_j) \geq 0$$

This equation can be rewritten in the form of matrix productions, that is we can have

$$\mathbf{g}^T \mathbf{R} \mathbf{g} \geq 0$$

where

$$\mathbf{R} = \begin{bmatrix} R_X(t_1 - t_1) & R_X(t_1 - t_2) & \dots & R_X(t_1 - t_n) \\ R_X(t_2 - t_1) & R_X(t_2 - t_2) & \dots & R_X(t_2 - t_n) \\ \dots & \dots & \dots & \dots \\ R_X(t_n - t_1) & R_X(t_n - t_2) & \dots & R_X(t_n - t_n) \end{bmatrix}$$

It is also equivalent to a positive semi-definite matrix \mathbf{R} . From the theory of matrix algebra, it is seen that $\mathbf{R} \geq 0$, provided all the eigenvectors of \mathbf{R} is greater than or equal to zero.

This is an essential property of stationary process. (See figure 3.1-9). In fact, any continuous function, if it is positive semi-definite, it must be an auto-correlation function of certain stationary process.

6) If a stationary process X(t) satisfies the following:

$$X(t) = X(t + T), \quad (3.1-67)$$

it is periodic. T is the period

7) If X(t) and X(t+ τ) becomes independent and $E(X(t)) = 0$, when $|\tau| \rightarrow \infty$ then

$$\lim_{|\tau| \rightarrow \infty} R_X(\tau) = 0 \quad (3.1-68)$$

A measurement noise with zero means is denoted by $X(t)$. When τ become large, $X(t)$ and $X(t+\tau)$ becomes independent and the corresponding auto-correlation function becomes zero.

8) The formula $\tau = t_2 - t_1$ implies that:
for a stationary process, by using correlation functions, we lose information of the initial time (phase).

c) Spectral Analysis and Transfer Functions

In most dynamic measurement, we use Fourier (Laplace) Transforms to analyze the frequency spectrum of a time history. In this section, we discuss how to use this power tool to analyze the frequency structure of a random process by means of the power density spectrum.

Power Density Function of a Stationary Process

Generally, it is well known that if a function $x(t)$ satisfies the the following condition,

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty \quad (3.1-69)$$

then, signal $x(t)$ has its frequency spectrum $F_x(\omega)$ and

$$F_x(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (3.1-70)$$

Generally speaking, spectrum $F_x(\omega)$ is complex valued,

$$[F_x(-\omega)]^* = F_x(\omega) \quad (3.1-71)$$

where the superscript * stands for complex conjugate.

Parseval Equation

$$\int_{-\infty}^{\infty} x(t)^2 dt = 1/2\pi \int_{-\infty}^{\infty} |F_x(\omega)|^2 d\omega \quad (3.1-72)$$

In the above Parseval equation, the left hand side stands for the total energy of $x(t)$ in the range $-\infty < t < \infty$. On the right hand side, $|F_x(\omega)|^2$ is called the energy density spectrum. So the right hand side stands for the spectrum of the total energy.

However, the total amounts of energy in many signals commonly seen in engineering practice are infinite. Many of them do not satisfy equation (3.1-69) either. We then try to study the average power, denoted by

$$\lim_{T \rightarrow \infty} 1/2T \int_{-T}^T x(t)^2 dt$$

In most cases, the above term exists and is finite. For example, let a new function be defined as

$$x_T(t) = \begin{cases} x(t), & |t| \leq T \\ 0 & |t| > T \end{cases}$$

When we measure a signal, we do not measure forever, but just within a certain duration.

So

$$F_x(\omega, T) = \int_{-\infty}^{\infty} x_T(t) e^{-j\omega t} dt = \int_{-T}^T x(t) e^{-j\omega t} dt \quad (3.1-73)$$

The corresponding Parseval Equation becomes

$$\int_{-\infty}^{\infty} x_T(t)^2 dt = 1/2\pi \int_{-\infty}^{\infty} |F_x(\omega, T)|^2 d\omega$$

Dividing both side of the above equation by $2T$, and let $T \rightarrow \infty$, we can have

$$\text{Lim}_{T \rightarrow \infty} 1/2T \int_{-T}^T x(t)^2 dt = 1/2\pi \int_{-\infty}^{\infty} \text{Lim}_{T \rightarrow \infty} \{1/2T |F_x(\omega, T)|^2\} d\omega \quad (3.1-74)$$

Equation (3.1-74) represents the *average power* in the range $-\infty < t < \infty$.

Denote

$$S_X(\omega) = \text{Lim}_{T \rightarrow \infty} \{1/2T |F_x(\omega, T)|^2\} \quad (3.1-75)$$

which is called the *average power density spectrum*. Or *auto-power density spectrum*; sometimes, *power density spectrum*. (In some textbooks, even: *power spectrum*)

Extend the concept to stationary process, $x(t) \rightarrow X(t)$

Called $\text{Lim}_{T \rightarrow \infty} E\{1/2T \int_{-T}^T x(t)^2 dt\}$ the average power of the stationary process, recall equation (3.1-34):

$$\Psi_X^2(t) = E[X(t)^2] = R_X(t, t)$$

$$\text{Lim}_{T \rightarrow \infty} E\{1/2T \int_{-T}^T x(t)^2 dt\} = \text{Lim}_{T \rightarrow \infty} 1/2T \int_{-T}^T E\{x(t)^2\} dt = \Psi_X^2(t) \quad (3.1-76)$$

The above equation means that the average power of a stationary process is equal to its mean square value. Combine equation (3.1-74) and (3.1-76). We have

$$\Psi_X^2(t) = 1/2\pi \int_{-\infty}^{\infty} \text{Lim}_{T \rightarrow \infty} \{1/2T E[|F_x(\omega, T)|^2]\} d\omega \quad (3.1-77)$$

Denoting

$$S_X(\omega) = \lim_{T \rightarrow \infty} \{1/2T E [| F_x(\omega, T) |^2]\}, \quad (3.1-78)$$

we have

$$\Psi_X^2(t) = 1/2\pi \int_{-\infty}^{\infty} S_X(\omega) d\omega \quad (3.1-79)$$

which is the average power spectrum: how the average power is distributed in the frequency domain.

In real measurement, t cannot be negative, we thus define single sided power density

$$G_X(\omega) = \begin{cases} 2 \lim_{T \rightarrow \infty} \{1/T E [| F_x(\omega, T) |^2]\} & \omega \geq 0 \\ 0 & \omega < 0 \end{cases} \quad (3.1-80)$$

because $S_X(\omega)$ is an even function.

Characteristics of Power Density Functions

1) $S_X(\omega)$ is real-valued, positive semi-definite function of ω

$$2) S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau \quad (3.1-81)$$

$$R_X(\tau) = 1/2\pi \int_{-\infty}^{\infty} S_X(\omega) e^{j\omega\tau} d\omega \quad (3.1-82)$$

That is, $S_X(\omega)$ and $R_X(\tau)$ are *Fourier Transform Pair*
Denoted by

$$S_X(\omega) \Leftrightarrow R_X(\tau) \quad (3.1-83)$$

Advantage of using (3.1-83), instead of using

$$X(\omega) \Leftrightarrow x(t)$$

is that, after the calculation of the auto-correlation function $R_X(\tau)$, the interested frequency components become more emphasized. The signal-to-noise-ratio becomes higher.

FFT stands for the fast Fourier transform. IFFT stands for the inverse of the last Fourier transform. We can use the FFT algorithm to calculate the above equations. That is,

$$R_X(\tau) = \text{FFT}(S_X(\omega)) \quad (3.1-84)$$

$$S_X(\omega) = \text{IFFT}(R_X(\tau)) \quad (3.1-85)$$

Equations (3.1-81) and (3.1-82) are called Wiener-Khintchine equations

Since both $S_X(\omega)$ and $R_X(\tau)$ are even functions, the following are true

$$S_X(\omega) = 2 \int_0^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau \quad (3.1-86)$$

$$R_X(\tau) = 1/\pi \int_0^{\infty} S_X(\omega) e^{j\omega\tau} d\omega \quad (3.1-87)$$

3) Let $X(t)$ and $Y(t)$ be two stationary processes

Define *cross density functions*

$$S_{XY}(\omega) = \lim_{T \rightarrow \infty} \{ 1/2T E [F_X(-\omega, T) F_Y(\omega, T)] \} \quad (3.1-89)$$

where the definitions of $F_X(-\omega, T)$ and $F_Y(\omega, T)$ are the same described in (3.1-73)

Note that, $S_{XY}(\omega)$ is no longer real-valued. However, we can have

$$S_{XY}(\omega) = [S_{YX}(\omega)]^* \quad (3.1-90)$$

$$4) S_{XY}(\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau \quad (3.1-91)$$

$$R_{XY}(\tau) = 1/2\pi \int_{-\infty}^{\infty} S_{XY}(\omega) e^{j\omega\tau} d\omega \quad (3.1-92)$$

$$\text{That is, } S_{XY}(\omega) \Leftrightarrow R_{XY}(\tau) \quad (3.1-93)$$

And

$$R_{XY}(\tau) = \text{FFT}(S_{XY}(\omega)) \quad (3.1-94)$$

$$S_{XY}(\omega) = \text{IFFT } R_{XY}(\tau) \quad (3.1-95)$$

5) $\text{Re}[S_{XY}(\omega)]$ and $\text{Re}[S_{YX}(\omega)]$ are even functions of frequency ω

$\text{Im}[S_{XY}(\omega)]$ and $\text{Im}[S_{YX}(\omega)]$ are odd functions of frequency ω

$$6) |S_{XY}(\omega)| \leq S_X(\omega)S_Y(\omega)$$

Transfer Functions of Linear Systems

A linear system is defined by the following ordinary differential equations:

$$b_n y^{(n)} + b_{n-1} y^{(n-1)} + \dots + b_0 y = a_m x^{(m)} + a_{m-1} y^{(m-1)} + \dots + a_0 x \quad (3.1-96)$$

where b_i 's and a_j 's are constant, the superscript (k) of variable (.) stands for

$$(.)^{(k)} = d^k (.) / dt^k$$

For example, SDOF vibration system has

$$m y'' + c y' + k y = -m x''_g$$

Now, let us solve equation (3.1-96) by using the Laplace transformation, assuming zero initial conditions.

Multiplying e^{-st} on both sides of equation (3.1-96) and taking the integrals of each term over the range $[-\infty, \infty]$ we have

$$(b_n s^n + b_{n-1} s^{n-1} + \dots b_0) Y(s) = (a_m s^m + a_{m-1} s^{m-1} + \dots a_0) X(s) \quad (3.1-97)$$

where

$$Y(s) = \int_{-\infty}^{\infty} y(t) e^{-st} dt \quad (3.1-98)$$

$$X(s) = \int_{-\infty}^{\infty} x(t) e^{-st} dt \quad (3.1-99)$$

are called the Laplace transform, also denoted by

$$Y(s) = L[y(t)], \quad (3.1-100)$$

$$X(s) = L[x(t)] \quad (3.1-101)$$

$$\text{And } s = \eta + j \omega \quad (3.1-102)$$

is called the Laplace variable.

Now we can have

$$Y(s) = H(s) X(s) \quad (3.1-103)$$

$$\text{Where } H(s) = (a_m s^m + a_{m-1} s^{m-1} + \dots a_0) / (b_n s^n + b_{n-1} s^{n-1} + \dots b_0) \quad (3.1-104)$$

is called as the *transfer function* .

The relationship described by equation (3.1-103) implies much broader information. Any linear relationship between two sets of variables X and Y, can be described by equation (3.1-103).

The following is a special case with a useful function:

δ -function and its Laplace tranform, Impulse Excitation

$$L[\delta] = 1 \quad (3.1-105)$$

So, if the input is taken to be a δ function, the output is called the impulse response function, denoted by $h(t)$. Denote the Laplace transform of $h(t)$ be $Y_h(s)$, that is,

$$L[h(t)] = Y_h(s), \quad (3.1-106)$$

We then have

$$Y_h(s) = H(s) 1 \quad (3.1-107)$$

Combining equation (3.1-106) and (3.1-107), we have

$$H(s) = L[h(t)] = \int_{-\infty}^{\infty} h(t) e^{-st} dt \quad (3.1-108)$$

That is, the transfer function equals the Laplace transform of the impulse response. It is often referred to as the *Impulse Response Spectrum*.

Taking the inverse Laplace transform, according to the theory of convolution, we have

$$y(t) = \int_{-\infty}^{\infty} x(t - \lambda) h(\lambda) d\lambda \quad (3.1-109)$$

This equation implies that the output function $y(t)$ is the convolution of the input function $x(t)$ and impulse function $h(t)$.

Sinusoidal Excitation

When the input is sinusoidal and the output reaches its steady state, we can let

$$s = j\omega \quad (3.1-110)$$

Otherwise, we generally cannot have the above equation. However, ISO has a certain standard for sweep-sine*.

The above Laplace transform described in equation (3.1-108) now becomes the Fourier transform:

$$H(j\omega) = \int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt = \int_0^{\infty} h(t) e^{-j\omega t} dt \quad (3.1-111)$$

It is called the *Frequency Response Function* .

From (3.1-103), let $s = j\omega$, we can have

$$H(j\omega) = Y(j\omega) / X(j\omega) \quad (3.1-112)$$

Virtually, for a linear system, the transfer function, the impulse response function and the frequency response function are the same, which describe the eigen-parameters or modal parameters of the system, including natural frequencies, damping ratios and mode shapes. In a latter chapter we shall introduce the concept of modal testing and modal analysis in detail.

Linear System with Random Input (Random Excitation)

When the input $X(t)$ is taken to be a random process, according to the knowledge of random integration, the output

$$Y(t) = \int_{-\infty}^{\infty} X(t - \lambda) h(\lambda) d\lambda$$

is a random process, too. It's mean is

$$E[Y(t)] = E \left[\int_{-\infty}^{\infty} X(t - \lambda) h(\lambda) d\lambda \right] \quad (3.1-113a)$$

Exchange the order of integration and calculation of the mean, we have

$$E[Y(t)] = \int_{-\infty}^{\infty} E[X(t - \lambda)] h(\lambda) d\lambda \quad (3.1-113b)$$

Since the input is stationary,

$$E[X(t - \lambda)] = \mu_X = \text{constant} \quad (3.1-114)$$

Thus the output also has constant mean:

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(\lambda) d\lambda \quad (3.1-115)$$

It is also can be proven that the auto-correlation function of the output is also the function of a single variable $\tau = t_2 - t_1$. That is,

$$R_Y(t_1, t_2) = R_Y(\tau)$$

Since $S_Y(\omega) \Leftrightarrow R_Y(\tau)$, it can also be proven that the *auto-power density spectrum* of the output is

$$S_Y(\omega) = |H(j\omega)|^2 S_X(\omega) \quad (3.1-116)$$

That is, the power density function of output is the product of power density function of input with the “dynamic magnification factor” $|H(j\omega)|^2$.

In this way,

$$\begin{aligned} R_Y(\tau) &= 1/2\pi \int_{-\infty}^{\infty} S_Y(\omega) e^{j\omega\tau} d\omega \\ &= 1/2\pi \int_{-\infty}^{\infty} |H(j\omega)|^2 S_X(\omega) e^{j\omega\tau} d\omega \end{aligned} \quad (3.1-117)$$

And the average power is

$$\psi_Y^2 = 1/2\pi \int_{-\infty}^{\infty} |H(j\omega)|^2 S_x(\omega) d\omega \quad (3.1-118)$$

The *Cross-Power Density Spectrum* and the cross correlation function of the output vs. input has the following relationship:

$$S_{XY}(\omega) \Leftrightarrow R_{XY}(\tau)$$

Then, we have

$$S_{XY}(\omega) = H(j\omega) S_X(\omega) \quad (3.1-119)$$

The cross power density spectrum between the input and the output is equal to the product of the auto-power spectrum and the frequency response function.

The *coherence function*

$$\gamma^2(\omega) = |S_{XY}(\omega)|^2 / [S_X(\omega) S_Y(\omega)] \quad (3.1-120)$$

can be used to check the relation between the input X(t) and the output Y(t).

We can have different type of transfer functions by means of using different power density spectra, namely, we can have

$$H_1(\omega) = S_{XY}(\omega) / S_X(\omega) \quad (3.1-121)$$

Which can be directly obtained from equation (3.1-119)

And also

$$H_2(\omega) = S_Y(\omega) / S_{YX}(\omega) \quad (3.1-122)$$

The coherence function can then be expressed as

$$\gamma^2(\omega) = H_1(\omega) / H_2(\omega) \quad (3.1-123)$$



Figure 3.1-10 Measurement noises

Now, let us find an alternative formula for the transfer function, that can be easily realized from the response functions of input and output signals.

From (3.1-112), for simplicity, we neglect j and then have

$$H(\omega) = Y(\omega) / X(\omega)$$

Note that in this circumstance, the input is X and output is Y .

Furthermore, we can have

$$E[H(\omega)] = E[Y(\omega) X(\omega)^* / X(\omega) X(\omega)^*]$$

Rewrite the definition of auto- and cross-power density spectra:

$$S_X(\omega) = \lim_{T \rightarrow \infty} \{ 1/2T E [| F_X(\omega, T) |^2] \},$$

$$S_{XY}(\omega) = \lim_{T \rightarrow \infty} \{ 1/2T E [F_X(-\omega, T) F_Y(\omega, T)] \}$$

In real measurement, we approximate the functions $\{ \lim_{T \rightarrow \infty} F_X(\omega, T) \}$ and $\{ \lim_{T \rightarrow \infty} F_Y(\omega, T) \}$ by $X(\omega)$ and $Y(\omega)$, because our measurements have limited length. Therefore, we have

$$S_X(\omega) = E [| X(\omega) |^2],$$

And

$$S_{XY}(\omega) = E [X(-\omega) Y(\omega)]$$

We thus have

$$E[H(\omega)] = H(\omega) = E[Y(\omega)X(\omega)^*] / E[X(\omega)X(\omega)^*] = S_{XY}(\omega) / S_X(\omega)$$

$$H_1(\omega) = E[Y(\omega)X(\omega)^*] / E[X(\omega)X(\omega)^*] = S_{XY}(\omega) / S_X(\omega)$$

$$= E[Y(\omega)] E[X(\omega)^*] / E[X(\omega)] E[X(\omega)^*]$$

which is just equation (3.1-121) and in this case the transfer function is denoted by H_1 . Similarly, we can also prove that

$$H_2(\omega) = E[Y(\omega) Y(\omega)^* / X(\omega) Y(\omega)^*] = E[Y(\omega)] E[Y(\omega)^*] / E[X(\omega)] E[Y(\omega)^*]$$

$$= S_Y(\omega) / S_{YX}(\omega)$$

These two equations tell us that we can use the averaged frequency response functions $E[Y(\omega)]$ and $E[X(\omega)]$ to obtain the transfer functions.

Why should we not find the transfer function directly from its definition formula, because of the measurement noises? Suppose we have a structure under test, see figure 3.1-10 . The Fourier spectrum of the input is denoted by $X(\omega)$, the Fourier spectrum of the output is denoted by $Y(\omega)$. On the input side, we have noise denoted by $m(\omega)$, whereas on the output side, we have noise denoted by $n(\omega)$.

Assume that $S_{Xm}(\omega) = S_{Xn}(\omega) = S_{Ym}(\omega) = S_{Yn}(\omega) = 0$. It is then easy to understand that in this circumstance,

$$H_1(\omega) = S_{XY}(\omega) / [S_X(\omega) + S_m(\omega)] \quad (3.1-124)$$

and

$$H_2(\omega) = [S_Y(\omega) + S_n(\omega)] / S_{YX}(\omega) \quad (3.1-125)$$

It can be seen that if the output has a higher noise level, then the transfer function H_1 can be used to illuminate the output noise. And, if the input has higher noise level then the transfer function H_2 can be used to illuminate the input noise.

Other Excitation Methods

There are many ways to apply forcing functions as excitations. In chapter 4, we shall discuss them in more detail together with feasible actuators and their controls.

Here we just list the most commonly used excitations including the aforementioned methods as follows:

Input Controlled

Sinusoidal Excitation

- Sweep Sine
- Chirp Excitation
- Sine Dwell

Impact Excitation

- Hammer
- Explosion,
- Rocket

White Noise

- Bandpass White Noise

Earthquake Excitation,

Modified Earthquake Excitation,

Recorded Response as Excitations,

Step Function,

- Weight Dropping,
- Cable Cutting

Input Uncontrolled Ambient

3.1.2 Regression

Regression is a methodology applied in situations where there are two or more variables

3.1.2.1 General Methods, Linear Regression with the Least Square

To determine how two or more sets of variables are related, no matter whether linear or nonlinear, is the objective of regression (see figure 3.2-1).

1. Simple Regression

In the following, we first discuss the linear relationship between two sets of variables.

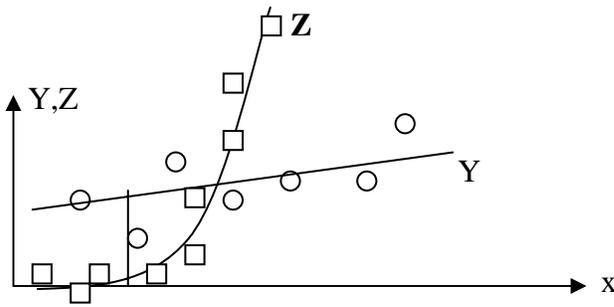


Figure 3.1-11 Variable x and random variable Y.

A set of variables x and random variable Y, are plotted in figure 3.1-11 (see line Y). Suppose the x's are given values and Y is measured from a test correspondingly. Y is taken to be random variables due to measurement errors.

Assume for each value of x_i , we have

$$E(y_i) = a + b x_i \quad (3.1-126a)$$

The relationship can then be written as

$$Y = a + b x + \varepsilon \quad (3.1-126b)$$

Here a and b are parameters to be realized. ε is a certain random error with a normal distribution, that is,

$$E(\varepsilon) = 0; \quad (3.1-127a)$$

$$\text{Var}(\varepsilon) = \sigma^2 \quad (3.1-127b)$$

What we have to do now is to estimate the parameters a and b . The estimators are denoted by a' and b' . For a given x , the estimator of $a + bx$ then denoted by

$$Y' = a' + b'x \quad (3.1-128)$$

Note that the deviation is

$$d = Y - Y' \quad (3.1-129)$$

what we have to do is to

$$\text{Minimize } \sum |d| = \sum |Y - Y'| \quad (3.1-130)$$

This is difficult mathematically, we thus turn to another formula

$$\text{Minimize } \sum d^2 = \sum [Y - Y']^2 \quad (3.1-131)$$

The idea described by (3.1-131) is called the *criterion of ordinary least squares*. And it will select a unique line called the *ordinary-least-squares line*.

Since the goal is actually find a and b by their estimator a' and b' , in the following, for simplicity, we can use a and b to replace a' and b' , with the risk of confusion. Therefore, we substitute Y for $a + b x + \varepsilon$ and $Y' = a' + b'x$ for $a + bx$ in (3.1-131) to have

$$Y - Y' = \varepsilon$$

To realize (3.1-131), arrange the measured data in to n pairs of (x_i, y_i) , $i = 1, 2, \dots, n$;

Now, let us construct a function Q of a and b in each pair (x_i, y_i) instead of the overall relation Y and Y' , that is

$$Q(a, b) = \varepsilon^2 = \sum_{i=1}^n (y_i - a - b x_i)^2 \quad (3.1-132)$$

Let

$$\partial Q / \partial a = 0 \text{ and}$$

$$\partial Q / \partial b = 0$$

we have

$$-2 \sum_{i=1}^n (y_i - a - b x_i) = 0 \text{ and}$$

$$-2 \sum_{i=1}^n (y_i - a - b x_i) x_i = 0$$

Furthermore, we have

$$na + nx_m b = n y_m \quad (3.1-133)$$

$$nx_m a + \sum_{i=1}^n x_i^2 b = \sum_{i=1}^n x_i y_i$$

By solving the above equations, we can have the estimated value of a and b, as denoted before by a' and b' as follows:

$$b' = + \sum_{i=1}^n x_i^2 b = \{ \sum_{i=1}^n (x_i - x_m) (y_i - y_m) \} / \{ \sum_{i=1}^n (x_i - x_m)^2 \} \quad (3.1-134)$$

$$a' = y_m - b' x_m \quad (3.1-135)$$

Under the condition (3.1-1127a) and (3.1-127b), it can be proven that the estimators a' and b' are unbiased with normal distribution. The mean and variance of b are:

$$E(b') = b; \quad (3.1-136)$$

$$\text{Var}(b') = \sigma^2 / \{ \sum_{i=1}^n (x_i - x_m)^2 \} \quad (3.1-137)$$

2. Multiple Regression

The above idea can be extended to the case of multiple linear regressions.

Suppose we have

$$Y = a + b_1 x_1 + b_2 x_2 + \varepsilon \quad (3.1-138)$$

Here x_1 and x_2 are variables; a, b_1 and b_2 are parameters to be realized. ε is a certain random error with the normal distribution described again in (3.1-127a) and (3.1-127b). By means of the above idea of the least squares, we first arrange the measured data in to n pairs of (x_{1i} , x_{2i} , y_i), $i = 1, 2, \dots, n$; To realize (3.1-131), let us construct a function of a and b, that is

$$Q(a, b_1, b_2) = \varepsilon^2 = \sum_{i=1}^n (y_i - a - b_1 x_{1i} - b_2 x_{2i})^2 \quad (3.1-139)$$

Similarly, by letting $\partial Q / \partial a$, $\partial Q / \partial b_1$ and $\partial Q / \partial b_2 = 0$, we have

$$n a + \sum_{i=1}^n x_{1i} b_1 + \sum_{i=1}^n x_{2i} b_2 = \sum_{i=1}^n y_i$$

$$\sum_{i=1}^n x_{1i} a + \sum_{i=1}^n x_{1i}^2 b_1 + \sum_{i=1}^n x_{1i} x_{2i} b_2 = \sum_{i=1}^n x_{1i} y_i \quad (3.1-140)$$

$$\sum_{i=1}^n x_{2i} a + \sum_{i=1}^n x_{1i} x_{2i} b_1 + \sum_{i=1}^n x_{2i}^2 b_2 = \sum_{i=1}^n x_{2i} y_i$$

The solution of the above equations are the required estimation a', b_1' and b_2' .

It can be proven that the estimators of b_i 's are unbiased with normal distributions.

3.1.2.2 Non-linear Regression

1. Linearity

$$y = f(x) \quad (3.1-141)$$

$$f(ax) = a f(x) \quad (3.1-142a)$$

$$f(x_1 + x_2) = f(x_1) + f(x_2) \quad (3.1-142b)$$

Input $\sin(\omega t)$

Output $\sin(\omega t + \theta)$

2. Nonlinear Models

$$Y = g(X, B) + E \quad (3.1-143)$$

$$Y = [y_1, y_2, \dots, y_n]^T \quad (3.1-144)$$

$$X = [x_1, x_2, \dots, x_n]^T \quad (3.1-145)$$

$$B = [b_0, b_1, \dots, b_m]^T \quad (3.1-146)$$

$$E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T \quad (3.1-147)$$

Assume the errors, ε_i have normal distributions, and

$$E(\varepsilon_i) = 0 \quad (3.1-148)$$

$$\text{Var}(\varepsilon_i) = \sigma_i^2 \quad (3.1-149)$$

Another model:

$$Y = g(X, B)a^E \quad (3.1-150)$$

$$\ln(Y) = \ln[g(X, B)] + E \ln(a) \quad (3.1-150)$$

which becomes exactly the model described by (3.1-143). Therefore, in the following, we only briefly discuss the regression of (3.1-143).

Note that, the notation of a^E is

$$a^E = [a^{e_1}, a^{e_2}, \dots, a^{e_n}]^T \quad (3.1-151)$$

3. Non-linear Least Squares, Higher Order Functions

Similar to the treatment on the linear regression by using the least squares, we first construct a function Q , such that

$$\begin{aligned} Q(B') &= E^T E = [Y - g(X, B')]^T [Y - g(X, B')] \\ &= \sum_{i=1}^n [y_i - g(x_i, B')]^2 \end{aligned} \quad (3.1-152)$$

Let $\partial Q / \partial B' = 0$, we can have n nonlinear equations as follows,

$$-2 \sum_{i=1}^n [y_i - g(x_i, B')] \partial Q / \partial B' = 0, \quad i = 1, 2, \dots, n \quad (3.1-153)$$

By solving these non-linear equations, we can have the estimated parameters B' . There are several existing algorithms to solve the above non-linear equations. Interested readers can find these methods in standard textbooks in statistical analysis.

Note that, even if each error ε_i has normal distribution with the mean and variance described in (3.1-148) and (3.1-149), generally speaking, the parameters of in B do not have normal distributions. The estimators in B' are no longer unbiased.

3.1.2.3 Empirical Formulation, General Discussion

1. Concept of Empirical Formulation

Statistical Model vs. Analytical Model
Axiom Systems for Mathematics and Mechanics
 Constitutional Laws and Conditions
 Abstraction vs. Multiple Factors
Empirical Formulae

2. Similitude Analysis vs. Regression

We can use similitude analysis for better understanding and more accurate regression, than the method of regression discussed above.

Empirical Equation

 Often we do not consider the dimensions in the first run
 After regression, find right units

Similitude Analysis

 We have to consider the dimensions first \rightarrow more close to physical law
 Finding all the pi-terms \rightarrow can make it easier to extend the conclusions to the prototype

3. Iteration Procedures

In many cases, the procedure of parameter estimation involves iterations. Two issues need to be considered:

- 1) First is the convergence of the iteration,
- 2) Second is that the convergence point must be a good estimation or the corresponding error should be minimized. The second issue will be discussed in more detail in the following section on the principle of orthogonality.

4. Optimization on Parameter Estimation

Parameter estimation can have much more general applications than just finding out certain coefficients or numbers in a mathematical model; For example, designing a cost-effective test setup, using trial-and-error method to fix some parameters, trying to find an optimal location for loading, etc.

General Idea of Optimization on Parameter Estimation

Mathematically, find maximum or minimum values for a function, for a functional. This implies that there must exist a maximum (minimum) value. The function must not be monotonic increasing or decreasing.

As a matter of fact, the above condition is often not known or guaranteed to obtain. Therefore, quite often, the optimal parameters are located at a certain boundary line.

Construct an objective function.

In practice, the objective function cannot be analytically written. The procedures involve linear and nonlinear methods.

Initial Guess

Issues on Convergence

Single Variable, Single Peak

$$y = f(x)$$

Within the range $a \leq x \leq b$, find an x^* such that

$$f(x^*) = \max_{a \leq x \leq b} f(x) \quad (3.2-154a)$$

Or

$$f(x^*) = \min_{a \leq x \leq b} f(x) \quad (3.2-154b)$$

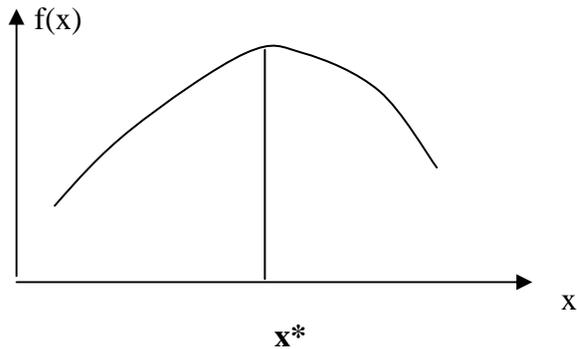


Figure 3.1-12 Single Peak Optimization

Let us discuss equation (3.2-154b) only. Single peak value:

$$\begin{aligned} f(x_1) > f(x_2) > f(x^*), & \text{ when } x_1 < x_2 < x^* \text{ and} \\ f(x_2) < f(x_1) < f(x^*), & \text{ when } x_2 > x_1 > x^* \end{aligned} \quad (3.2-155)$$

Golden Cut Method

Golden Cut method is one of the simplest yet most efficient methods for one-dimensional search for the optimal point, provided there is one and only one optimal point along the line of search.

This method can also be extended to multiple variable search whereas only one optimal point is allowed.

A more rigorous method is called Fraction method by using Fubinius series. The Golden cut method can be seen as an approximation of the fractional method. A more rough but practically quite effective method is just using the factor 0.5, instead of using 0.618. In many cases, it just needs to have few more iterations.

$$t_{k+1} = 0.382 (b_k - a_k) + a_k \quad (3.2-156)$$

$$t_{k+1}' = 0.618 (b_k - a_k) + a_k \quad (3.2-157)$$

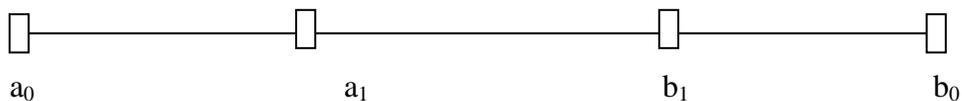


Figure 3.2-13 Golden Cut Optimization

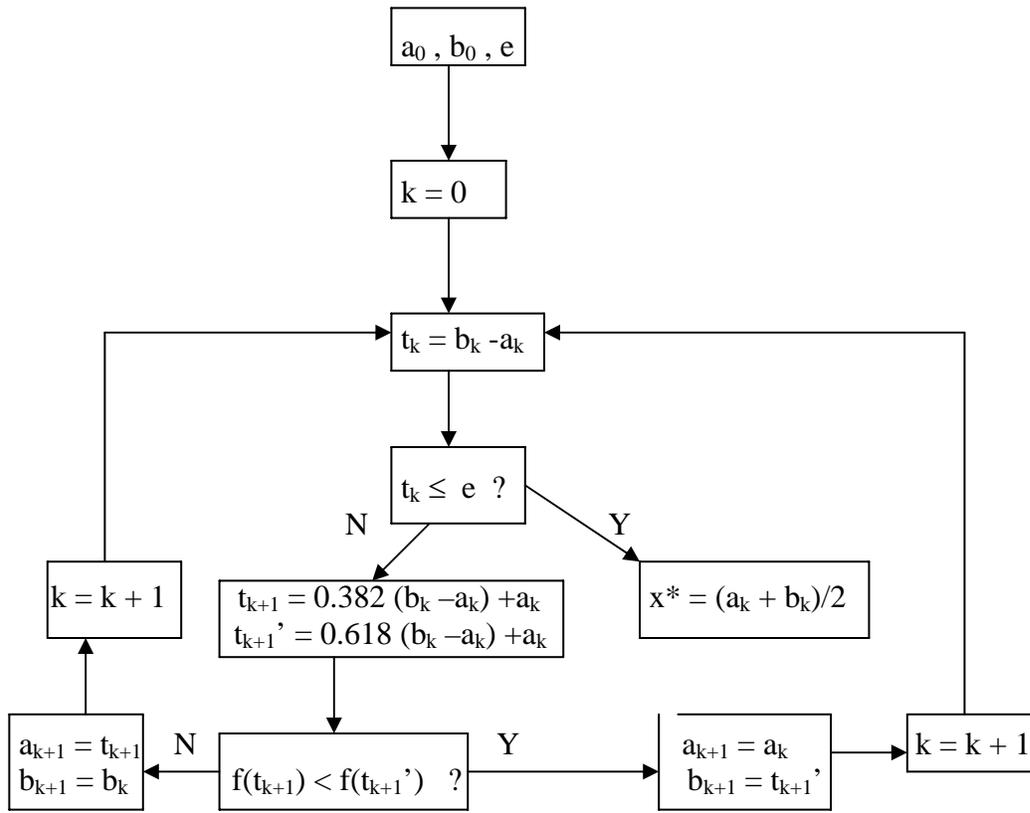


Figure 3.2-14 Block Diagram of Golden Cut Optimization

3.1.2.4 Orthogonal Principles

1. Single Variable

In order to estimate the parameter a and b for linear regression, in equation (3.1-132), we first used a function Q such that

$$Q(a, b) = \varepsilon^2 = \sum_{i=1}^n (y_i - a - b x_i)^2 \quad (3.2-158)$$

And then, we let $\partial Q / \partial b = 0$ to obtain

$$\sum_{i=1}^n (y_i - a - b x_i) x_i = 0 \quad (3.2-159)$$

Dividing n on both sides of above equation, we have

$$E\{[y - (a + bx)] x\} = 0 \quad (3.1-160)$$

This equation implies an important principle, the *Orthogonality Principle*, which states when the error $\varepsilon = y - (a + bx)$ is orthogonal to the variable x , we can obtain the best estimator a and b . Here, it is noted that we use lower case y instead of capital Y in the formula of $\varepsilon = y - (a + bx)$, (see (3.1-126b)). The reason for doing so is to extend the case from random variable Y to broader applications with any variable y .

We can use a geometric interpretation for better understanding the term of orthogonality. For simplicity, let us consider the case of $a = 0$, that is, $y = bx$, or y is estimated by bx . From equation (3.1-160), the best estimation of a is such that

$$e = E\{(y - ax) x\} = 0$$

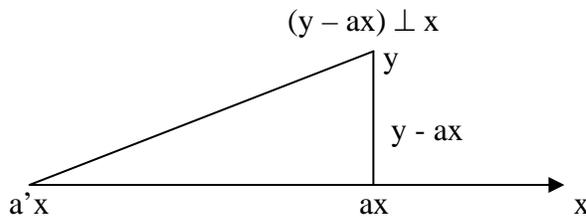


Figure 3.1-12 Orthogonal Projection

From figure 3.1-12, it is seen that the vector $y - ax$ is the vector from point ax on the x line to the point y . The length of that vector is equal to \sqrt{e} . When $y - ax$ is perpendicular to x , the length becomes minimum.

Bases and Spaces
Orthogonality

2. Multiple Variable Analysis

In more general cases, we carry out the mean square estimation for multiple variables. Equation (3.1-139) is used for regression with two variables x_1 and x_2 , which is the simplest case of multiple variable regression. In general, we may define

$$Q(B) = \varepsilon^2 = \sum_{i=1}^n [y_i - (b_0 + \sum_{j=1}^m b_j x_{ji})]^2 \quad (3.1-161)$$

Where $B = [b_1, \dots, b_m]^T$ is the set of parameters to be estimated.

Similarly, it can be proven that the following orthogonality principle hold:

The function Q in equation (3.1-161) is minimum, provided that the error of

$y - (b_0 + \sum_{j=1}^m b_j x_j)$ is orthogonal to the quantities $x_j, j = 1, \dots, n$, that is,

$$E\{ [y - (b_0 + \sum_{j=1}^m b_j x_j)] x_j \} = 0, j = 1, \dots, n \quad (3.1-162)$$

3. General Orthogonality Principle

The principle of orthogonality can be further extended to nonlinear regression, namely we can have

$$E\{ y - g(X) \} w(X) = 0 \quad (3.1-163)$$

Where $g(X)$ is a nonlinear mean square estimation of y . The error of such an estimation $y - g(X)$ is orthogonal to any function $w(X)$, linear or nonlinear, of quantities x_j .

The principle of orthogonality is a powerful tool to check whether an estimation (regression, realization) has the minimum error. In a latter chapter, we shall use the principle to examine the validity of modal parameters from modal testing.

3.2 Error Analysis

Measurement contains errors, which are likely random. In this section, we shall use the knowledge of statistics mentioned above as the major tool to analyze the measurement errors.

3.2.1 General Concept of Errors

In experimental tests, errors occur for the following reasons:

Uncontrollable measurement conditions (environmental or ambient), accidents, human error.

Limited testing capability: Inability to physically measure, limitation of measurement

Degree of measurement accuracy: Instruments

Modeling, Mathematically impossible to model, Erroneous modeling

Unknown factors:

3.2.1.1. Randomness and Uncertainty

The following issues are important in data analysis.

Random

The value of sample is uncertain

The position of a sample of certain value is uncertain

The distribution of the probability is certain

Uncertainty

Uncertain means unpredictable. Specifically, most errors in measurement are unpredictable.

Measurement Repeatability

Data obtained from different tests should be compatible.

Convergence of mean and variance

After a certain number of averages, the resulting mean should converge to a specific value.

Modeling Stability

In many cases, the purpose of experiments is to verify an existing model. It is therefore very important that the model be stable in terms of small variations in its parameters. For a stable model, even if we have a rather large amount of measurement/testing errors, we can still find a good agreement between the test data and the model. On the other hand, we need a very precise measurement and careful testing procedure.

The same is true when we seek to find a new model.

Suppose we have a function in the vector form, see equations (3.1-143) through (3.1-147):

$$Y = f(X, B)$$

where

Y consists of n variables y_1, \dots, y_n and

X consists of x_1, \dots, x_n

B is a certain set of parameters (coefficients)

When a certain perturbation Δx_i occurs such that

$$x_i^P = x_i + \Delta x_i \quad (3.2-1)$$

The values of Y changes accordingly:

$$Y^P(X, B) = Y + \Delta Y = [y_1 + \Delta y_1, \dots, y_n + \Delta y_n]^T \quad (3.2-2)$$

Modeling stability to variable x_i is defined as follows:

$$m_{si} = |\Delta Y| / |\Delta x_i| \quad (3.2-3)$$

It is thus seen that m_{si} is a function of the change in x_i , that is,

$$m_{si} = f(\Delta x_i) \quad (3.2-4)$$

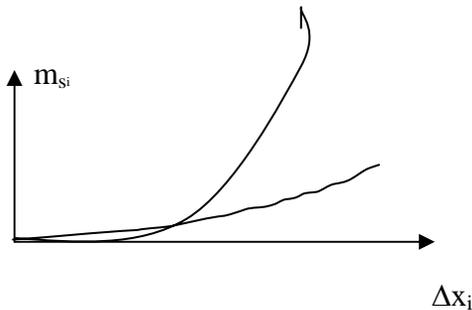


Figure 3.2-1 Stability of Modeling

Seen in figure 3.2-1, the above function can be very close to a constant in a certain range, which implies that in such a range, the modeling is quite stable.

The above function can also become exponentially increasing function, which indicates that when such a phenomenon occurs, the modeling become unstable.

A good modeling can have a large range of stable m_{si} 's.

3.2.1.2 Sources of Error

Sources of Errors in Experimental Testing, General Description

Instrumental Errors

Construction, Variability, temperature effects (Not known to experimentalist)

Systematic Errors (Fixed)

Consistent form that result from conditions or procedures that are correctable (zero shift, calibration factor, etc)

Random Errors

These are accidental errors that occur in all measurement. They are characterized by their inconsistent nature and their origin cannot be determined in the measurement process (corrected by statistical analysis) fluctuations, noise. Etc.

Sources of Elemental Error

Large numbers of errors have their source in the measurement system itself. Errors in measurement systems are often referred to as elemental errors, including:

1. *Calibration Error* also include known but uncorrected calibration errors such as hysteresis and non-linearity
2. *Data Acquisition Error*
3. *Data Reduction Error*, interpolation, curve fits, differentiation of data curves, erroneous modeling

3.2.1.3 Definition of Error

In the following, let us first define the measurement errors.

Absolute Error

$$\text{Error}_{\text{total}} = \text{measured value} - \text{true value} \quad (3.2-5)$$

$$\text{Precision error} = \text{reading} - \text{average of readings} \quad (3.2-6)$$

$$\text{Bias error} = \text{average of readings} - \text{true value} \quad (3.2-7)$$

Relative Error

$$\text{Relative Error} = (\text{Absolute error} - \text{true value}) / \text{true value} \times 100\% \quad (3.2-8)$$

It is important to know that

Bias Error is a systematic error

Precision Error is a random error

These uncertainties, known as *bias limit*, denoted by B, and *precision limit*, denoted by P. Suppose we have m measured variables: x_1, x_2, \dots, x_m

Precision Index (Propagation of Error) (Also known as *uncertainty/propagation of variation*)

Precision Index denoted by S_X , can be formulated as follows:

$$S_X = \{ [\sum_{i=1}^n (x_i - E(x))^2] / (n-1) \}^{1/2} \quad (3.2-9)$$

Precision limit, for a single measurement x_i , denoted by P_i , can be estimated by using the Student's statistics:

$$P_i = t S_X \quad (3.2-10)$$

Note that, t is a function of confidence level, and the degree of freedom, which is often set to be $n - 1$. The letter n here is the number of measurements. The reason to introduce the function t is as follows: If the size of a sample is small ($n < 30$), we may not use the sample standard deviation to represent the population standard deviation. Due to the uncertainty in the standard deviation, for the same confidence level, it is understandable that the confidence interval will be wider. In such as case, we use the Student's t:

$$t = (x_m - \mu_x) / (S / \sqrt{n})$$

Note that, the standard deviation of the mean is related to the *standard deviation of the measurement* by

$$S_{E(x)} = S_x / (\sqrt{n}) \quad (3.2-11)$$

The *uncertainty in the mean* can be written as

$$P_{E(x)} = t S_{E(x)} \quad (3.2-12)$$

The precision uncertainty interval in the mean is given by $\pm P_{E(x)}$

Figure 3.2-2(a) shows the relation among true value, population mean and Precision limit graphically. Figure 3.2-2(b) shows the relation among true value, population mean and the standard deviation of the measurement.

Relative Frequency

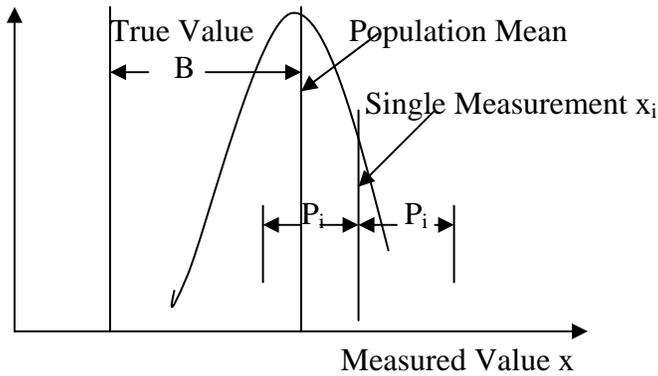


Figure 3.2-2a Graphical description of the true value and the Precision limit

Relative Frequency

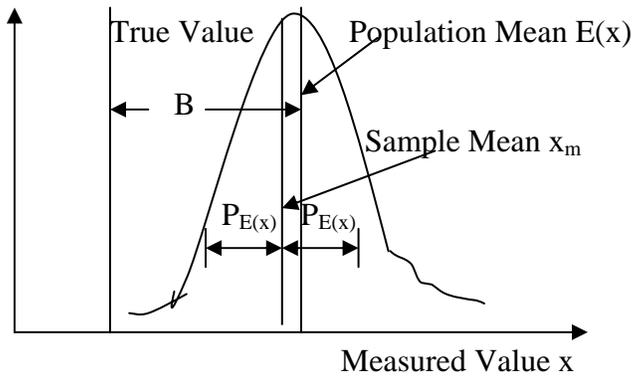


Figure 3.2-2b: Graphical display of precision and bias errors (Coleman and Steele)

The *bias limit*, B, however, will remain constant if the repeated tests are under the same measurement conditions.

To combine both the bias and precision uncertainties, the approach of the *square root of the sum of the sequence* (*RSS*), is used:

$$w = (B^2 + P^2)^{1/2} \quad (3.2-13)$$

where w is called the confidence level in the uncertainty. In ANSI/ASME (1986) it is recommended that a confidence level of 95% be used for uncertainty analysis.

In the above equation, P is precision limit of systematic error, which will be discussed in more detail in the following section.

3.2.2. Error Estimate & Criteria

To realize the maximum possible errors quantitatively is important in measurement. In the following, we use statistics to quantify the errors in details:

3.2.2.1 Random Error

Suppose we have m random errors, whose precision indices are denoted by S_i by using RSS, the combination becomes

$$S = [\sum_{i=1}^m (S_i)^2]^{1/2} \quad (3.2-14)$$

Then, we can further describe the systematic error in terms of precision limits, P,

$$P = tS \quad (3.2-15)$$

Again, t here is the value of Student's t for the degree of freedom r and the appropriate level of confidence.

Using the Student distribution to find the value of S, we must find a special value called the degree-of-freedom, denoted by r first. If the sample sizes are not less than 30, the value of t is only a function of the confidence level. However, if the number of samples is smaller than 30, ANSI/ASME (1986) suggests the use of the Welch-Satterthwaite formula:

$$r = S^4 / \{ \sum_{i=1}^m (S_i^4/r_i) \} \quad (3.2-16)$$

where r_i is the degree of freedom associated with the individual elemental uncertainty, which equals to the number of data values in the sample minus one. It is seen that, r is then the value of the total degree of freedom for the variable x.

3.2.2.2 Measurement (Systematic) Bias

Suppose we have n systematic bias limits, using RSS, the combination becomes

$$B = [\sum_{i=1}^n (B_i)^2]^{1/2} \quad (3.2-17)$$

3.2.2.3 Propagation of Uncertainties

Identify Possible Errors and Associate with Given Data

Often, we can have errors from several sources. Suppose we can determine the range of each individual error source. Our goal now is to estimate the total or resulted error from those error sources. For example, when we attempt to calculate the stress from measured force and the cross area of a specimen, we have the following formulae:

$$\begin{aligned} \sigma &= F/A \\ F &= F_o \pm \Delta F \\ A &= b h \\ b &= b_o \pm \Delta b \\ h &= h_o \pm \Delta h \\ \sigma_o &= F_o / b_o h_o \end{aligned}$$

Now, from the above relations, can we determine the error of $\Delta\sigma$? Where $\pm \Delta F$, $\pm \Delta b$, and $\pm \Delta h$ are random errors, ΔF , Δb , and Δh can be determined from accuracy of instruments, derivation in reading or data transfer, truncation in readings etc. In the following let us introduce several methods to estimate the resulted error:

Initial Error Contribution

Suppose we have

$$f = f_o (x_1, x_2, \dots, x_m)$$

Its total differential is

$$df = \partial f_o / \partial x_1 dx_1 + \partial f_o / \partial x_2 dx_2 + \dots + \partial f_o / \partial x_m dx_m \quad (3.2-18)$$

where dx_1 , dx_2 , and dx_m can be treated as errors; $\partial f_o / \partial x_1$, etc are weighting factors. It is understandable that df represents the resulting error.

Suppose in the above example, we have

$$\begin{aligned} f &= 10 \text{ kips} + 1.0 \text{ kip} \\ b &= 2'' \pm 0.01'' \\ h &= 1'' \pm 0.01'' \end{aligned}$$

$$\sigma_o = F_o / b_o h_o = 10 / (2)(1) = 5 \text{ ksi}$$

Then we can calculate:

$$\begin{aligned} d\sigma &= \partial\sigma/\partial F dF + \partial\sigma/\partial b db + \partial\sigma/\partial h dh \\ &= 1/b h dF + (-F/b^2h) db + (F/b h^2) dh \\ &= (1/(2)(1)) (1.0) + (10/(4)(1) (\pm 0.01) - (10/(2)(1) (\pm 0.01) \\ &= 0.5 + (\pm 0.025) + (\pm 0.05) \\ &\leq 0.575 \text{ ksi (worst possible error)} \end{aligned}$$

Now, the worst possible percentage error is

$$d\sigma / \sigma_o \leq 0.575 / 5 = 11.5\%$$

3.2.2.4 Final Uncertainty of Single-Sample Measurement

In the above section we introduced the concept of initial error contribution. An alternative approach is the concept of final uncertainty.

Again, R is a function of m measured variables, x_1, x_2, \dots, x_m

$$R = f(x_1, x_2, \dots, x_m) \quad (3.2-19)$$

And B_i is the bias uncertainty and S_i is the precision index of the measurement of variable x_i , then the bias uncertainty B_R and the precision index S_R of the result can be calculated from

$$B_R = [\sum_{i=1}^n (B_i \partial f / \partial x_i)^2]^{1/2} \text{ and} \quad (3.2-20)$$

$$S_R = [\sum_{i=1}^n (S_i \partial f / \partial x_i)^2]^{1/2} \quad (3.2-21)$$

Where the partial derivatives, $\partial f / \partial x_i$'s are called sensitivity coefficients B_R and S_R can be combined into the uncertainty in the final result, w_R , with the following formula

$$w_R = [B_R^2 + (tS_R)^2]^{1/2} \quad (3.2-22)$$

where t is the Student's value. To obtain t, we must know the degree of freedom of the precision index S_R .

If all the variables have r-values no less than 30, t is independent of r.

If S for one or more variables is based on a small sample, the appropriate value of r can be determined from the modified Welch-Satterthwaite formula (ANSI/ASME, 1986)

$$r_R = (S_R^2)^2 / \{ \sum_{i=1}^n 1/r_i [(S_i \partial f / \partial x_i)^2]^2 \} \quad (3.2-23)$$

Data Reduction Error

R = calculated quantity

$$R = f(x_1, x_2, \dots, x_m)$$

$$w_R^2 = \sum (\partial R / \partial x_i)^2 w_{x_i}^2 \quad (3.2-24)$$

Where x_i 's are measured independent variables

$$\text{for } R = k x_1^a, x_2^b, \dots, x_m^m \quad (3.2-26)$$

$$(w_R/R)^2 = a^2 (w_{x1}/x_1)^2 + b^2 (w_{x2}/x_2)^2 + \dots + a^2 (w_{xm}/x_m)^2 \quad (3.2-27)$$

Above example again:

$$f = 10 \text{ kips} + 1.0 \text{ kips}$$

$$b = 2'' \pm 0.01''$$

$$h = 1'' \pm 0.01''$$

$$\sigma_o = F_o / b_o h_o = 10 / (2)(1) = 5 \text{ ksi}$$

$$\partial \sigma = [0.5^2 + 0.025^2 + 0.05^2]^{1/2} = 0.503 \text{ ksi}$$

$$\text{so, } \sigma = 5 \pm 0.503 \text{ ksi}$$

3.2.2.5 Final Uncertainty of Single-Sample Measurement, RSS Method

The third approach is *the square root of the sum of the sequence* (RSS) which we have already employed before, that is,

$$\partial f = \{[(\partial f_o / \partial x_1) dx_1]^2 + [(\partial f_o / \partial x_2) dx_2]^2 + \dots + [(\partial f_o / \partial x_m) dx_m]^2\}^{1/2} \quad (3.2-28)$$

Above example again:

$$(\partial \sigma / \sigma)^2 = (\partial F / F)^2 + (\partial b / b)^2 + (\partial h / h)^2 = (1/10)^2 + (0.01/2)^2 + (0.01/1)^2$$

$$\partial \sigma = 25 [0.01 + 0.00025 + 0.0001]^{1/2} = (0.2531)^{1/2} = 0.5031 \text{ ksi}$$

$$\text{so, } \sigma = 5 \pm 0.503 \text{ ksi which is exactly the one calculated by (3.2-27).}$$

3.2.2.6 Step-by-Step Analysis

1. Define the Measurement Process
2. List all of the Elemental Error Sources
3. Estimate the Elemental Errors
4. Calculate the Bias and Precision Error for Each Measurement
5. Propagate the Bias Limits and Precision Indices All the Way to the Results
6. Calculate the Overall Uncertainty of the Results

7. If the Errors are Unacceptable, Try to Replace Certain Procedures/Instruments with Higher precision and Redo from Step 1.

3.2.3 Error Control

3.2.3.1 Rejection of Questionable Data

Modified Thompson τ Technique

Suppose we have m measurements with mean $E(x)$ and standard deviation S . Arrange the data in ascending order x_1, x_2, \dots, x_m . Extreme values (highest and lowest) are suspected outliers. For these suspected points, x_i 's, the deviation σ_i can be calculated as

$$\sigma_i = |x_i - E(x)| \quad (3.2-29)$$

Then select the largest value.

Find a value of τ from a table. The largest value of σ_i should be compared to the product τS . If the value of d exceeds τS , the data can be rejected as an outlier.

Normally, by using this method, only one datum is allowed to be rejected. Note that, the elimination of an outlier is not always positive event. It should be careful in practical tests.

3.2.3.2 Random Error with Zero Mean and Averaging

Performed after data is acquired.

Arithmetic Average

$$x_A = \{ \sum_{i=1}^m x_i \} / m \quad (3.2-30)$$

This is the best estimator of the mean value in the sense of least squares. It is the most commonly used averaging method.

The Standard Deviation for Arithmetic Average is define by equation (3.2-9) and rewrite as follows:

$$\sigma = \{ [\sum_{i=1}^m (x_i - x_A)^2] / (m - 1) \}^{1/2}$$

Average Error for Arithmetic Average is defined in equation (3.2-11)

$$S_{E(x)} = S_x / (\sqrt{n})$$

And rewrite as follows:

$$\delta = |x_i - \mu_X| = \sigma / \sqrt{m} = \{ [\sum_{i=1}^m (x_i - x_A)^2] / [m(m-1)] \}^{1/2} \quad (3.2-31)$$

Simplified Arithmetic Average

This method is used when we have a large number of data.

$$x_S = A + C \{ \sum_{i=1}^m v_i y_v \} / m \quad (3.2-32)$$

Here A is a constant, m is the number of groups, C is the distance between groups. And C is taken to be constant; v_i is the probability of the i^{th} group,

$$\sum_{i=1}^m v_i = n \quad (3.2-33)$$

where n is the total number of observations.

$$y_i = (x_{mi} - A) / C \quad (3.2-34)$$

x_{mi} is the mid number of the data in the i^{th} group.

$$x_{mi} = (I_{\max} - I_{\min}) / 2 \quad (3.2-35)$$

where I_{\max} and I_{\min} are respective the maximum and minimum value of the data in the i^{th} group.

Geometric Average

$$x_G = \{ \prod_{i=1}^m x_i \}^{1/m} \quad (3.2-36)$$

or

$$\lg x_G = \{ \sum_{i=1}^m \lg x_i \} / m \quad (3.2-37)$$

Weighted Average

$$X_W = \{ \sum_{i=1}^m w_i x_i \} / \{ \sum_{i=1}^m w_i \} \quad (3.2-38)$$

The Standard Deviation for Weighted Average is define as follows:

$$\sigma = \{ [\sum_{i=1}^m w_i (x_i - x_W)^2] / [(m - 1) \sum_{i=1}^m w_i] \}^{1/2}$$

Average Error for Weighted Average is defined as follows:

$$\delta = | x_i - \mu_X | = \sigma / \sqrt{m} = \{ [\sum_{i=1}^m w_i (x_i - x_W)^2] / [m (m - 1) \sum_{i=1}^m w_i] \}^{1/2} \quad (3.2-39)$$

3.2.3.3 Moving Averaging and Windowing

1. Moving Averages

$$x_j = \{ \sum_{i=j-(m-1)/2}^{j+(m-1)/2} w_i x_i \} / m \quad (3.2-40)$$

where m = number of points to be averaged (must be odd number)

i.e if 5 point average moving window: then

$$x_6 = \{ \sum_{i=6-(5-1)/2}^{6+(5-1)/2} w_i x_i \} / 5$$

where x_j averaged value at point j

x_i independent variable at point i

w_i weighting factor at point i .

Boxcar Function

$$w_i = \text{constant} \quad (3.2-41)$$

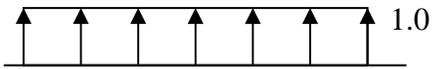
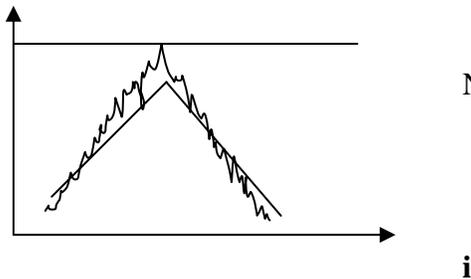


Figure 3.2-3 Constant window.



Note: Reduced peak slightly wider

Figure 3.2-4 Filtering effect

Gaussian Function

$$w_i = \exp(- [(x_i - x_j)/(2\sigma)]^2) / (\sigma\sqrt{2\pi}) \quad (3.2-42)$$

where σ is the variable of smoothing

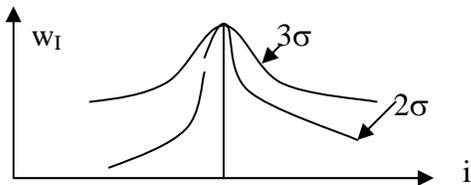


Figure 3.2-5 Gaussian windows

More weight to the center and less weight to the ends of the window

2. Windows and Window Functions

The weighting factor can be varied for a variety of applications. It actually forms windowing functions.

We realize a window in the time domain by using weighting functions for different purpose. They may be described in the time domain, or in the frequency domain.

a. Time Domain Windowing, Noise Reduction

The main purpose of using a window in the time domain is to reduce measurement errors. The aforementioned method of moving average by using Boxcar function and Gaussian function are good examples.

The following exponential window is another example:

Exponential Window

The exponential window is mainly used to reduce the noise from a response of a structure subjected to impulse excitations. From figure 3.2-6, it is understandable that at the beginning of the response, the signal to noise ratio can be quite high because shortly after the impulse excitation, the value of the response is high. However, towards the end, since the response of the free decay vibration is almost died out, what we have may just be noises. In this case, we can use the following windowing function to reduce the noise, that is,

$$w_E(t) = e^{-at} \quad (3.2-43)$$

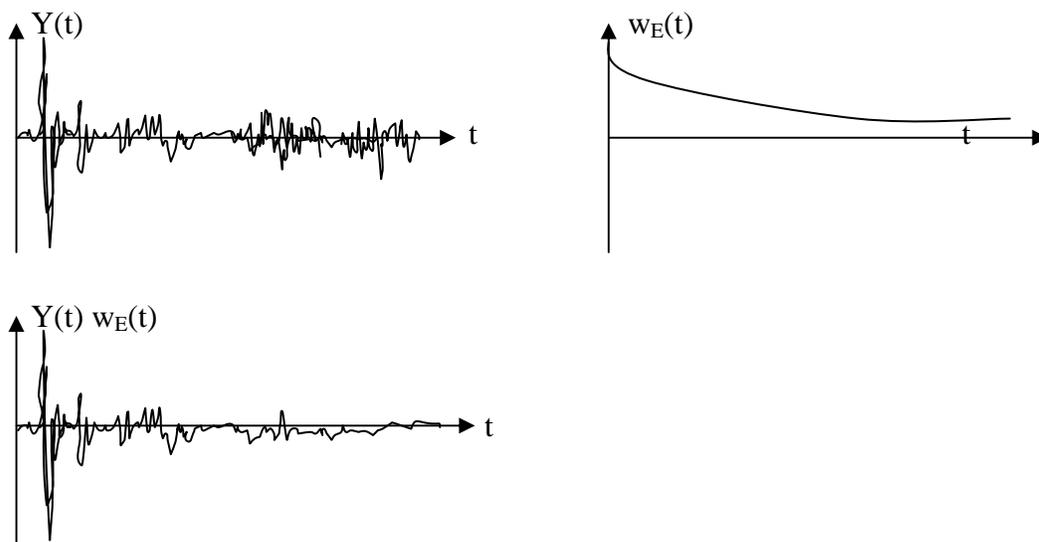


Figure 3.2-6 Exponential window

Note that, exponential window can make the measured damping ratio of a structure larger than the real value.

b. Frequency Domain Windowing, Power Leakage and Compensation

Another type of windowing is used to reduce the power leakage. To understand the power leakage, let us examine the following procedure of sampling.

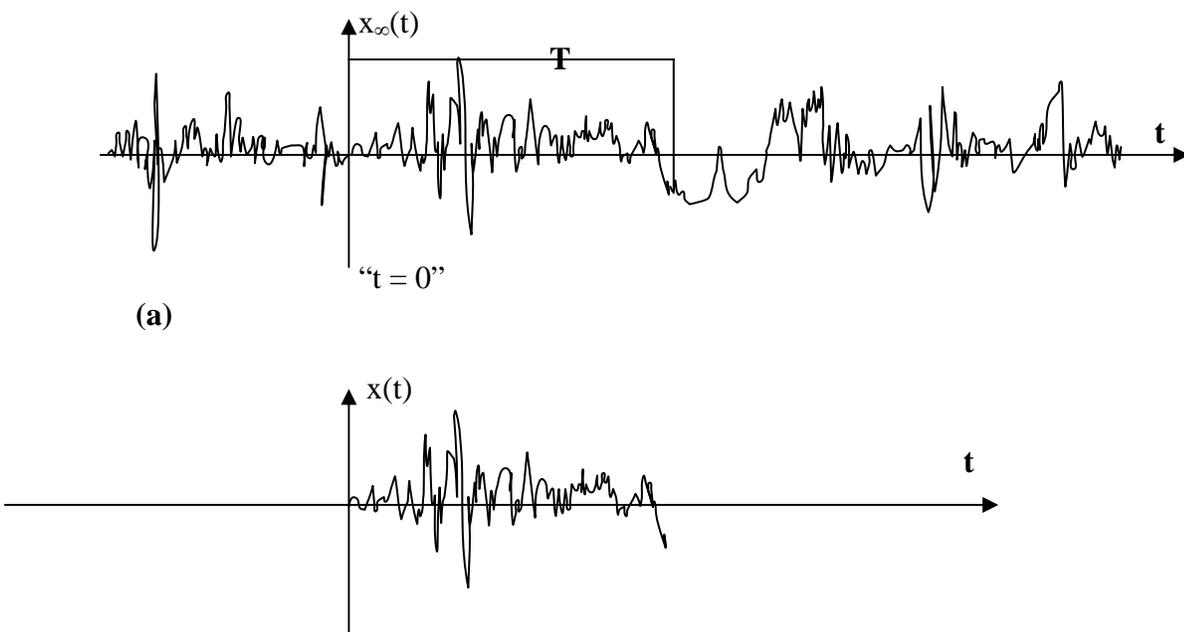
Any measured signal $x(t)$ having limited length T can be seen as a product of a signal $x_\infty(t)$ with infinite length and a boxcar function $w_B(t)$:

$$x(t) = x_\infty(t) w_B(t) \quad (3.2-44)$$

where we rewrite equation (3.2-41) and use $w_B(t)$ to denote the continuous Boxcar function as follows

$$w_B(t/T) = \begin{cases} 1 & 0 \leq t < T \\ 0 & t > T \end{cases} \quad (3.2-45)$$

The result of using the Boxcar function can be learned from figures 3.2-7 and 3.2-8. In figure 3.2-7(a), it is seen that a Boxcar function is used to truncate the signal $x_\infty(t)$, which become $x(t)$ shown in figure 3.2-7(b). However, when we perform the Fourier transform, it is actually equivalent to a periodic function generated by the periodic extension from $x(t)$ shown in figure 3.2-8, since a Fourier transform is an integration in the time range $[-\infty, \infty]$. By comparing the signal shown in figure 3.2-7(a) and figure 3.2-8, it is seen that this new signal can be quite different from the original one.



(b)

Figure 3.2-7 Using Boxcar function



Figure 3.2-8 Periodic Extension

The above explanation in the time domain can also be understood in the frequency domain. It is noted that the Fourier spectrum of a signal production in the time domain will become a convolution in the frequency domain, that is,

$$X(\omega) = \int_{-\infty}^{\infty} X_{\infty}(\lambda)W_B(\omega-\lambda)d\lambda \quad (3.2-46)$$

The Fourier transform of Boxcar function can be proven to be

$$W_B(\omega) = T[\sin(\omega T/2)/(\omega T/2)] \quad (3.2-47)$$

The absolute amplitude of $W_B(\omega)$ can be plotted in figure 3.2-9.

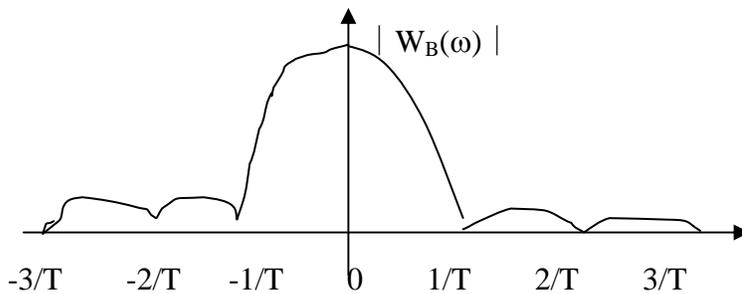


Figure 3.2-9 Fourier Spectrum of Boxcar Function

For simplicity, let $x_{\infty}(t)$ be a sinusoidal signal, that is

$$x_{\infty}(t) = A e^{j\omega_0 t} \quad (3.2-48)$$

Its Fourier transform is

$$X_{\infty}(\omega) = A \delta(\omega - \omega_0) \quad (3.2-49)$$

Substitution of (3.2-47) and (3.2-49) into (3.2-46), we can have

$$X(\omega) = AT [\sin[(\omega - \omega_0)T/2]]/[(\omega - \omega_0)T/2]e^{jT(\omega-\omega_0)/2t} \quad (3.2-50)$$

And its absolute amplitude is

$$|X(\omega)| = AT |\sin[(\omega - \omega_0)T/2]/[(\omega - \omega_0)T/2]| \quad (3.2-51)$$

which can be shown in figure 3.2-10.

From figure 3.2-10, it is seen that the original spectrum, a straight line at ω_0 now becomes a much complicated curve with many side petals. From the energy point of view, the energy has a leakage from the original case concentrated at ω_0 into those side petals. It is noted that the height of the first petal is about 21.7% of the main petal. We thus can realize that the power leakage is a serious problem.

However, for a period function, if the sampling length is exactly equal to the integer number of its period T_p , we will not have the problem of power leakage. That is,

$$\omega_0 = 2\pi/T_p = 2\pi n/T - n \Delta \omega_0 \quad (3.2-52)$$

Note that, the zero point of the Fourier transform of the Boxcar function is at exactly

$$k \Delta \omega_0 = k 2\pi/T, \quad k = 1, 2, \dots \quad (3.2-53)$$

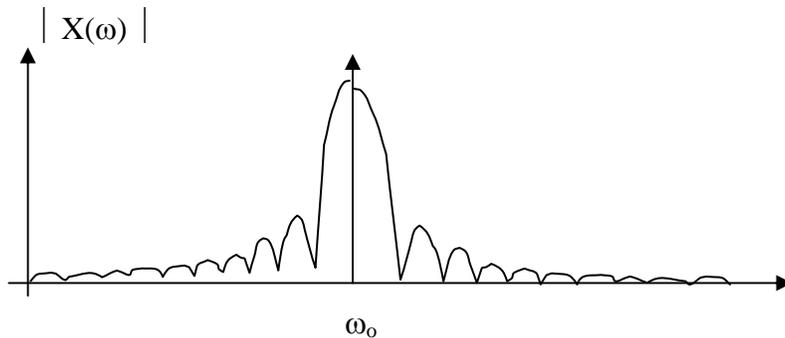


Figure 3.2-10 Fourier Spectrum Truncated Sinusoidal Signal

Therefore, the spectrum of a sinusoidal signal in this case is a line with the amplitude of A and located at $\omega_0 = n \Delta \omega$, which indicates that there is no leakage at all. However, if

$$\omega_0 \neq n \Delta \omega$$

we will witness the leakage. It is understandable that a sample taken from random signals with a limited length must yield leakage.

Window Functions to Reduce Leakage

It is seen that the leakage is caused by limited length of samples, which causes the beginning and the end of the sampled signal to be discontinuous. It is therefore for us to create a certain window functions that make the beginning and the end of the sampled signal more smooth. In other words, these window functions are designed to reduce the amplitudes at the beginning and the end of the sampled signal while keeping the rest unchanged.

Hanning Window

$$w_H(t) = \begin{cases} [1 - \cos(2\pi t/T)]/2 & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (3.2-54)$$

Hanning window can be seen in figure 3.2-11.

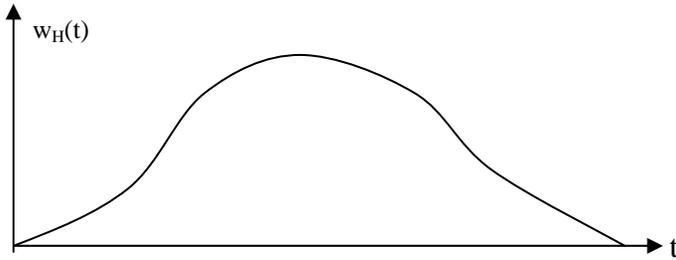


Figure 3.2-11 Hanning Windows

Rewrite the Hanning window as follows:

$$w_H(t) = \frac{1}{2} - \frac{1}{4} e^{j2\pi/T t} - \frac{1}{4} e^{-j2\pi/T t} \quad 0 \leq t \leq T \quad (3.2-55)$$

Its Fourier transform is

$$W_H(\omega) = \frac{1}{2}W(\omega) - \frac{1}{4} W(\omega - 2\pi/T) - \frac{1}{4} W(\omega + 2\pi/T) \quad (3.2-56)$$

Where

$$W(\omega) = T \left\{ \frac{\sin [(\omega T)/2T]}{[(\omega T)/2T]} \right\} e^{-j\omega/2T}$$

$$W(\omega - 2\pi/T) = -T \left\{ \frac{\sin [(\omega - 2\pi/T)/2T]}{[(\omega - 2\pi/T)/2T]} \right\} e^{-j\omega/2T}$$

$$W(\omega - 2\pi/T) = -T \left\{ \sin \left[\frac{(\omega + 2\pi/T)/2T}{[(\omega + 2\pi/T)/2T]} \right] \right\} e^{-j\omega/2T}$$

Therefore,

$$W_H(\omega) = T e^{-j\omega/2T} \left\{ \frac{1}{2} \sin \left[\frac{(\omega T)/2T}{[(\omega T)/2T]} \right] + \frac{1}{4} \left[\frac{(\omega - 2\pi/T)/2T}{[(\omega - 2\pi/T)/2T]} \right] + \frac{1}{4} \left[\frac{(\omega + 2\pi/T)/2T}{[(\omega + 2\pi/T)/2T]} \right] \right\} \quad (3.2-57)$$

From equation (3.2-57), it is seen that the spectrum of Hanning window consists of three function of $\sin x/x$. The amplitude of the second and the third terms are one half of that in the first term. The second term shift $1/T$ towards right from the first term and the third term shift $1/T$ towards left from the first term. In this manner, the second side petal is only 2.67% of the main petal. Note that the second petal of Boxcar function is 21.7% of that in the main petal. Therefore, the Hanning window can effectively reduce the power leakage, see figure 3.2-12.

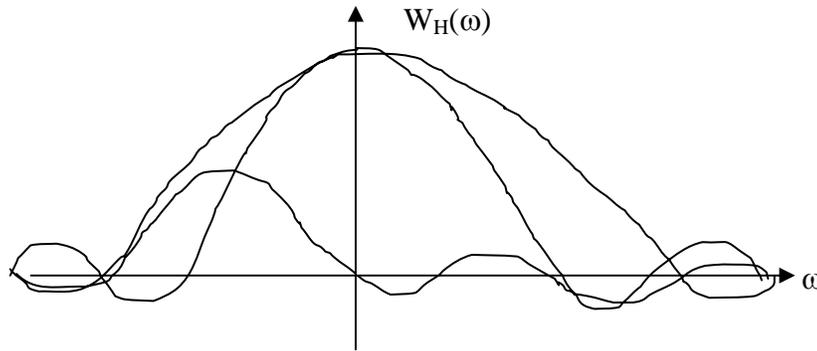


Figure 3.2-12 Spectrum of Hanning window

However, since the width of the petal becomes wider, the frequency resolution is also reduced.

Hamming Window

Hamming window can deal with the side petal at $\omega = 1.25 \pi/T$

$$w_M(t) = a - (1 - a)\cos(2\pi t/T) \quad 0 \leq t \leq T \quad (3.2-58)$$

when $a = 0.54$, it becomes Hanning window.

Blackman Window

Blackman Window can further reduce the side petal

$$w_B(t) = \sum_{m=0}^K (-1)^m a_m \cos(2\pi m/T) \quad 0 \leq t \leq T \quad (3.2-59)$$

One of the popular choice $K = 2$, $a_0 = 0.42$, $a_1 = 0.50$ and $a_2 = 0.08$;

Flat Top Window

In order to increase the frequency resolution, keep the main petal flat top.

$$w_F(t) = w_B(t) [a_0 + 2 \sum_{k=1}^3 a_k \cos(2\pi kt)] \quad (3.2-60)$$

$a_0 = 0.99948$, $a_1 = 0.95573$ and $a_2 = 0.091581$;

Half Windows

In earthquake engineering, we often deal with signals with small amplitude at the beginning. Or when we take samples, we can purposely choose such a beginning point whose value close to zero. In these circumstances, we can use Half Windows. For example, we can have the flat top half window:

$$w_{Hf}(t) = \begin{cases} 1 & 0 \leq t \leq T_1 \\ \{1 + \cos[\pi (t - T_1)] / T\} / 2 & T_1 \leq t \leq T_1 + T \\ 0 & \text{elsewhere} \end{cases} \quad (3.2-61)$$

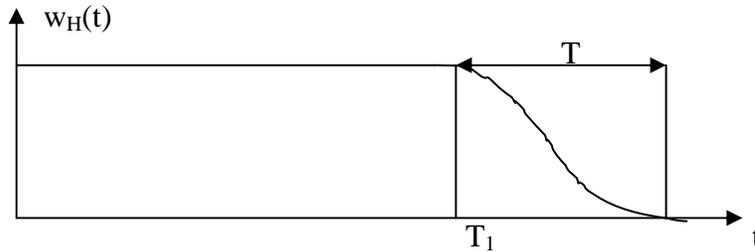


Figure 3.2-13 Half Window

From the author's experience, the half window can effectively reduce the power leakage and remains good frequency resolution.

Another kind of half windows is the cosine half window:

$$w_{Hc}(t) = \begin{cases} \{1 + \cos[\pi t/T]\} / 2 & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (3.2-62)$$

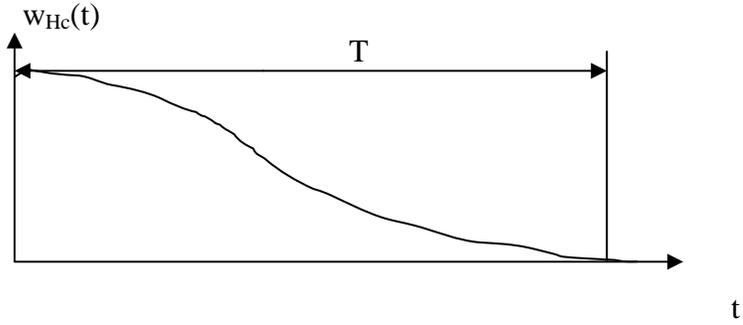


Figure 3.2-14 Cosine half window

The third type of half window is Cosine Flat Top Window

$$w_{Hcf}(t) = \begin{cases} 1 & 0 \leq t \leq T_1 \\ \cos[\pi (t - T_1)/2T] & T_1 \leq t \leq T_1 + T \\ 0 & \text{elsewhere} \end{cases} \quad (3.2-63)$$

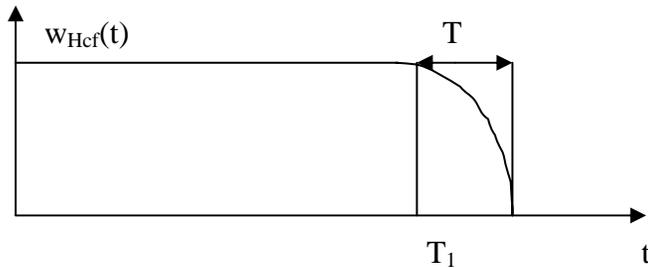


Figure 3.2-1t Cosine flat top half window:

3.2.3.4 Separation of Measurement Signals from Systematic Errors

1. Non-stationary to Stationary

First Order Non-stationary Process

$$Y(t) = f_1(t) + X(t) \quad (3.2-65)$$

Where $f_1(t)$ is a deterministic function, and $Y(t)$ is nonstationary process with time-varying mean $\mu_Y(t)$. However, its auto-correlation is only a function of $\tau = t_2 - t_1$. $X(t)$ is a stationary process with zero mean, i.e., $\mu_X = 0$.

Therefore, we can have

$$\mu_Y(t) = E[f_1(t) + X(t)] = f_1(t) \quad (3.2-66)$$

$$\begin{aligned} R_Y(\tau) &= E\{ [f_1(t) + X(t) - \mu_Y(t)] [f_1(t + \tau) + X(t + \tau) - \mu_Y(t + \tau)] \} \\ &= E[X(t) X(t + \tau)] = R_X(\tau) \end{aligned} \quad (3.2-67)$$

Therefore, we can centralize such a non-stationary process by first finding

$$\mu_Y(t) = f_1(t)$$

and then have

$$X(t) = Y(t) - f_1(t) \quad (3.2-68)$$

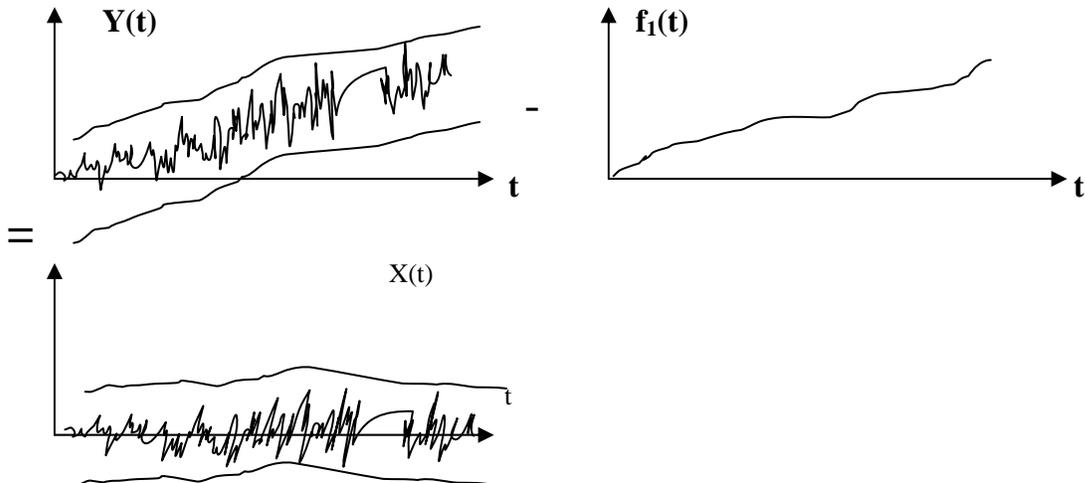


Figure 3.2 – 15 First order non-stationary process

Second Order Non-stationary Process

If the auto-correlation function of the non-stationary process cannot be written as the function of τ only, then it is called the second order non-stationary process.

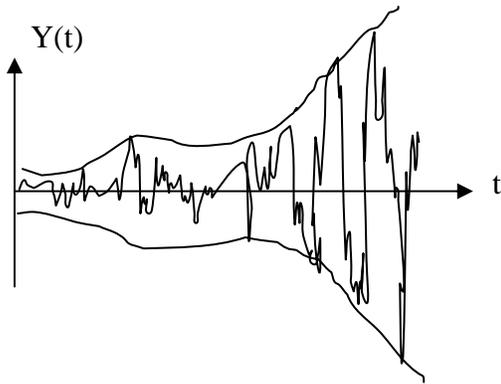


Figure 3.2-16 Second order non-stationary process

$$Y(t) = f_2(t)X(t) \quad (3.2-69)$$

Where $f_2(t)$ is a deterministic function, and $Y(t)$ is a second non-stationary process. $X(t)$ is a stationary process with zero mean, i.e., $\mu_X = 0$, unitary variance, i.e. $\sigma_X = 1$ and its standard deviation is denoted by $\rho_X(\tau)$

First of all

$$\mu_Y(t) = E[f_2(t)X(t)] = f_2(t)E[X(t)] = 0 \quad (3.2-70)$$

Secondly

$$\begin{aligned} R_Y(t, t+\tau) &= E\{ [f_2(t)X(t)][f_2(t+\tau)X(t+\tau)] \} \\ &= f_2(t)f_2(t+\tau)\rho_X(\tau) \end{aligned} \quad (3.2-71)$$

Let $\tau = 0$, we have

$$R_Y(t, t) = \sigma_Y^2(t) = f_2(t)^2$$

Or

$$\sigma_Y(t) = f_2(t) \quad (3.2-72)$$

Therefore, we can find $\sigma_Y(t) = f_2(t)$ and normalize $Y(t)$ by

$$X(t) = Y(t)/f_2(t) \quad (3.2-73)$$

Generally speaking, we need first estimate the value of $\sigma_Y(t)$ from the samples measured in practice.

2 Filtering

Filter has a very general meaning in measurement systems. In fact, it is one of the basic concepts in linear systems. In the following, we shall discuss the fundamental definitions and theories of filters, but leave the issue of how to design a filter to a latter chapter.

Concept of Filter

Transfer function

$$H = \text{Output/Input} \quad (3.2-74)$$

Filter

$$\text{Output} = (\text{Transfer Function of Filter}) \text{Input} \quad (3.2-75)$$

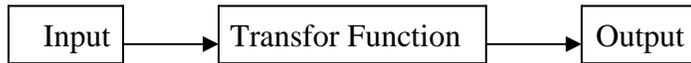


Figure 3.2-17 Concept of Filtering

Purpose of Filter,

1. Remove unwanted

Design in the frequency domain, remove unwanted frequency components

Lowpass

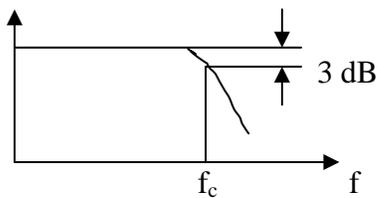


Figure 3.2-18(a) Lowpass Filter

Lowpass filter is the most commonly used filters in measurement systems. It is often used to cut off high frequencies to let the measures signal satisfying sampling theory. From figure 3.2-18(a), we can define the follows:

Passband [0 f_c]

Cutoff frequency f_c , the frequency at 3 dB attenuation. In order to effectively fliter out unwanted signals, we can have the following practical filters

Butterworth filter has maximally flat in the passband. Essentially the gain is constant in the passband. For a lowpass filter, unit gain in the passband.

$$|G(f)| = (1 + (f / f_c)^{2n})^{-1/2} \tag{3.2-76}$$

Where n is the order of the filter:

n =1: -6dB/octave attenuation in amplitude

n =2: -12 dB/octave

That is, generally, we have $-6n$ /oct for a order.

Phase vs. Frequency is another important characters of a filter. The first order filter has the following relationship:

$$\Phi(f) = \tan^{-1} [f / f_c] \tag{3.2-77}$$

Generally we have

$$G(f) = (D(f))^{-1/2} \tag{3.2-78}$$

Where the Denominator D(f) are as follows:

$$D(f)_1 = (j 2\pi f) + 1 \tag{3.2-79}$$

$$D(f)_2 = (j 2\pi f)^2 + 1.4142 (j 2\pi f) + 1 \tag{3.2-80}$$

$$D(f)_3 = (j 2\pi f)^3 + 2(j 2\pi f)^2 + 2(j 2\pi f) + 1 \tag{3.2-81}$$

$$D(f)_4 = (j 2\pi f)^4 + 2.6131(j 2\pi f)^3 + 3.4142(j 2\pi f)^2 + 2.6131(j 2\pi f) + 1 \tag{3.2-82}$$

$$D(f)_5 = (j 2\pi f)^5 + 3.2361(j 2\pi f)^4 + 5.2361(j 2\pi f)^3 + 5.2361(j 2\pi f)^2 + 3.2361(j 2\pi f) + 1 \tag{3.2-83}$$

Integration Effect can be found in low pass filters, that is,

Lowpass filter → integration, remove higher frequency, phase

The Frequency Domain Production can be written as

$$Y(\omega) = H_F(\omega) X(\omega) \tag{3.2-84}$$

The Time domain convolution is

$$Y(t) = \int h_f(\lambda) x(\lambda-t) d\lambda \tag{3.2-85}$$

From figure 3.2-19(a) to 3.2-19 (b), we can conceptually visualize the effect of a low pass filter.

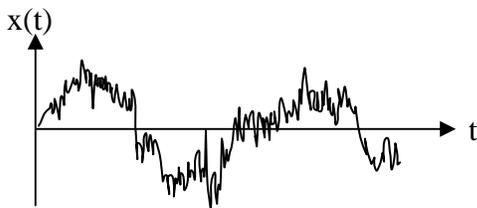


Figure 3.2-19 (a)

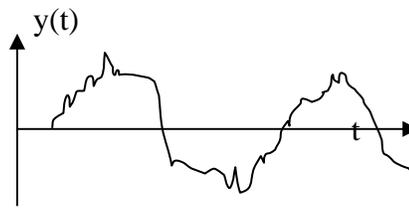


Figure 3.2-19(b)

Highpass

The highpass filter has the frequency response shown in figure 3.2-18(b).

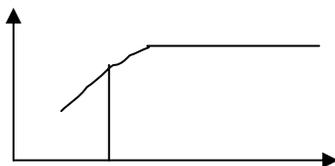


Figure 3.2-18 (b) Highpass Filter

By using a highpass filter, we can have the phenomenon shown in figure 3.2-20(a) to 3.2-20 (b).

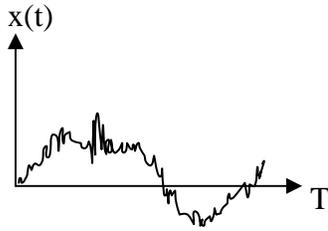


Figure 3.2-20 (a)

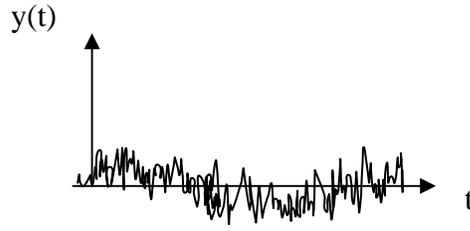


Figure 3.2-20(b)

Bandpass

The bandpass filter has the frequency response shown in figure 3.2-18(c).

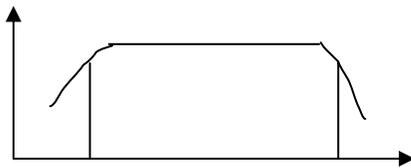


Figure 3.2-18(c) Bandpass Filter

By using a bandpass filter, if the band is too narrow, we can have the phenomenon shown in figure 3.2-21(a) to 3.2-21 (b).

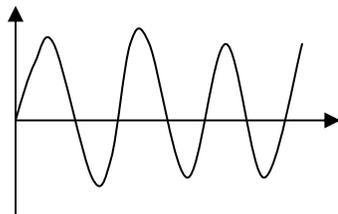


Figure 3.2-21(a)

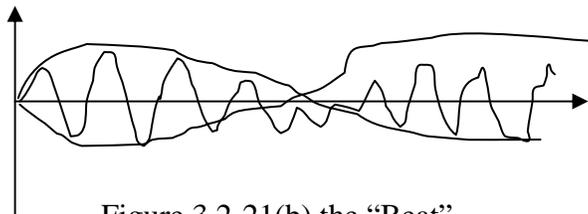


Figure 3.2-21(b) the "Beat"

Bandstop

The bandstop filter has the frequency response shown in figure 3.2-18(d).

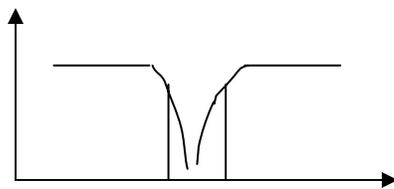


Figure 3.2-18 (d) Bandstop Filter

2. Estimation

Time duration of estimation is denoted by t_e ; Time duration of measurement is denoted by t_m . We can have the following situations,

1) Averaging, Smoothing

$$t_e < t_m \quad (3.2-86)$$

2) Filtering

$$t_e = t_m \quad (3.2-87)$$

3) Predicting

$$t_e > t_m \quad (3.2-88)$$

From equation (3.2-86) through (3.2-88), it is seen that, with the periods of t_e and t_m , we can have different functions of filtering. When $t_e > t_m$, we use the "filter" to predict or the estimate the future behavior of our signals. Usually, in measurement systems, we do not often meet this situation. However, prediction is one of the important measures in control problems.