Chapter 1

Basics

Important measurement questions

- How can a measurement be devised so that the measurement provides unambiguous information we seek?
- How can a measurement system be used so that the engineer can easily interpret the measured data and be confident in their meaning?

The measurement & measurement system

- Measurement: Act of assigning a specific value to a physical variable; the measured variable.
- Measurement system: Tool used to quantify the measured variable.

Sensors and Transducers

A complete measurement system consists of both a sensor and a transducer.

- Sensor: the physical element to sense the variable being measured.
- Transducer: converts the sensed information into a detectable signal.

Subsequently, the *signal-conditioning stage* takes the signal from the transducer and modifies it to a desired magnitude. The *output stage* indicates or records the measured value (e.g. a display or a marked scale).

Additionally, a measurement system might include a *feedback-control stage*, which makes a decision regarding actions required to control the process.

Experimental test plan

1. Parameter design plan: what do you want to answer, what parameters need to be measured, and which parameters influence the result?
2. System & tolerance design plan: select the equipment (i.e. what measurement device) and how accurate should the result be?
3. Data reduction design plan: How do you interpret the data, how will it answer your question, and how *good* is this solution?

Variables

A *variable* is a physical quantity whose value influence the result. Which is either discrete or continuous in nature. There are variables that influence each other, dependent variables. And those that do not, independent variables. Furthermore, some variables are controlled, meaning they are kept fixed. *Extraneous variables* are variables that are not purposely manipulated or controlled during measurement but that do affect the test result.

- Dependent: variable affected by changes in the value of one or more other variables. These variables are functions of independent variables.
- Independent: variable that does not influence other variables. These are generally kept constant or purposefully changed.

- Extraneous: variables that are not controlled during measurement, but do affect the result.

Randomization

Refers to test strategies that apply the changes in independent variables in some random order. This is to neutralize effects that may not be accounted for in the test plan and therefore attempts to reduce or eliminate bias as much as possible.

Noise & interference

Noise is the random variation in the volume of the measured signal, which increases data scatter. And, *interference* (an extraneous signal) imposes an undesirable/false trend on the measured signal.

Repetition & Replication

- Repetition: Repeated measurements of the same variable during any single test run or on a single batch. With the goal of improving the estimate of a measured variable.
- Replication: intended duplication of a set of measurements or a test. With the goal of quantifying the variation in the measured variable as it occurs between duplicate tests under similar conditions.

Calibration

Calibration applies a known input value to a measurement system for the purpose of observing the system output value. These known values are called standards.

Static calibration

An input is applied to the system under calibration, and the output is recorded. The input value does not vary with time or space. Or an average is used.

We record a functional relationship between the input and output, this correlation can be used to establish an unknown input value based on the output.

We can define the *static sensitivity*, K, for any static input value x_i,

$$K(x_i) = \left.\frac{dy}{dx}\right|_{x=x_i}$$

The static sensitivity is a measure relating the change in the indicated output associated with a given change in a static input.

Dynamic calibration

The input value are time or space dependent. Again, we want to determine a relation between the input and output, however, now the input is dynamic (e.g. sinusoidal).

Range & Span

- The range of an instrument is specified by the lower and upper limit in which the measurement system is designed to operate.
- The span is the difference between that upper and lower limit. $r_i = x_{\max} - x_{\min}$ is the input span and $r_o = y_{\max} - y_{\min}$ is the output span.

Resolution

The resolution is the smallest increment in the measured value that can be measured.

Accuracy and error

The accuracy of a measurement is the closeness between the measured and true value. The accuracy is influenced by many factors, called errors.

We can quantify those errors by,

$$e = \text{measured value} - \text{true value}$$

or in a relative sense as,

$$A = \frac{|e|}{\text{reference value}} \times 100\%$$

- Random errors: random variations in the measured value.
- Systematic errors: creates an offset between the mean and true value of the data set.

Uncertainty

The uncertainty is a numerical estimate of the possible range of the error in the measured value. The uncertainty is caused by system calibration, measurement techniques, and the statistics of the dataset.

Hysteresis error

The difference between the measured value between an upscale sequential test and a downscale sequential test.

$$u_h = y_{\text{up}} - y_{\text{down}}$$

Hysteresis is the dependence of the state of a system on its history, i.e. when the next measurement depends on the previous measurement.

Linearity error

Many instruments are designed to achieve a linear relation between an applied static input and the output value. The linearity error describes how well a linear relation is achieved.

$$u_L = y(x) - y_L(x)$$

where $y_L(x)$ describes a linear fit.

Sensitivity & zero error

The sensitivity error $u_K(x)$ is a statistical measure of random errors in the estimate of the slope of the calibration curve. The zero error $u_z(x)$ is a shift of the calibration curve.

Verification & Validation

Verification and validation determines if the product complies with their specifications and whether they are fit for their intended use.

- Verification: Establishing the truth of correspondence between a work product and its specification. *Are we building the product right?*
- Validation: Refers to ensuring that the experimental model used is itself correct.

Significant Digits

In writing numbers, the leftmost nonzero digit is the *most significant digit*. The rightmost is the *least significant digit*. All nonzero digits in a number are significant. All leading zeros are not significant, they only refer to the magnitude. Zeros between nonzero digits are significant. And trailing zeros are also significant.

Rounding is a process in which the number of digits is reduced and the remaining least significant digit(s) adjusted appropriately. In reducing the number of digits, (1) if the digits to be discarded begin with a digit less than 5,the digit preceding the 5 is not changed; (2) if the digits to be discarded begin with a 5 and at least one of the following digits is greater than 0, the digit preceding the 5 is increased by 1; (3) if the digits to be discarded begin with a 5 and all of the following digits are 0, the digit preceding the5 is unchanged if it is an even digit but increased by 1 if it is an odd digit.

Some rules for numerical operations are:

1. In addition or subtraction, the number of digits following the decimal in the reported result should not be greater than the least number of digits found following a decimal in any of the data points used.
2. In other operations, the significant digits in the result should not be greater than the least number of significant digits in any of the data points or the operand used.
3. The number of significant digits in an exact count (i.e. the number of samples N) is not considered when establishing the number of significant digits to be reported.
4. Round your final result, but do not round intermediate calculations.

Chapter 2

Signal analysis

We can consider multiple types of signals:

- Analog signal: a signal continuous in time. Any value within the operating range is available.

- Discrete (time) signals: information about the magnitude of the signal is available only at discrete points in time.
- Digital signal: signals only exist at discrete values in time. The magnitude is also discrete (i.e. the magnitude is quantized).

An Analog-to-Digital (ADC or A/D) is a solid-state device that converts an analog voltage signal to a binary number representation. Note, however, that the ADC has a limited resolution. The range of voltages creates the quantization levels and establishes the range. The quantization assigns a single number to represent a range of magnitudes of an analog signal.

We can classify signals as either static or dynamic, i.e. not varying with time or varying with time. Often a signal is periodic. A steady periodic signal repeats in regular intervals in time. Simple periodic functions contain only a single frequency and amplitude, whereas complex periodic signals contain a multitude of frequencies and amplitudes.

Mean & RMS Values

The mean $\bar{y}$ describes the DC component of the signal, and is defined as,

$$\bar{y} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} y(t)\, dt$$

The RMS value of a signal on the other hand tells something about the AC component of the signal, it is defined as

$$y_{\text{rms}}^2 = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} y(t)^2\, dt$$

or for a discrete signal as,

$$y_{\text{rms}}^2 = \frac{1}{N} \sum_i y_i^2$$

By analogy, concerning currents and voltages:

> The RMS value is the effective value of a varying voltage or current. It is the equivalent steady DC value which gives the same effect. For example, a lamp connected to a 6V RMS AC supply will shine with the same brightness when connected to a steady 6V DC supply.

Periodic Signals

A periodic function is a function $y(t + T) = y(t)$ for some value T. A general form of such signals is,

$$y = A \cos \omega t + B \sin \omega t$$

which is a signal with period $T = 2\pi/\omega = 1/f$ and frequency $f = \omega/2\pi$. We can convert this to either a sine or cosine function,

$$y = C\sin(\omega t + \phi^*)$$

or

$$y = C\cos(\omega t - \phi)$$

where

$$C^2 = A^2 + B^2, \quad \phi = \arctan(B/A) \quad \text{and} \quad \phi^* = \arctan(A/B)$$

Fourier Series

If we consider periodic functions, i.e. $y(t+T) = y(t)$, we can represent them as a sum of sines and cosines.

$$y(t) = A_0 + \sum_{n=1}^{\infty}\left[A_n\cos n\omega t + B_n \sin n\omega t\right]$$

Note that the terms corresponding to $n = 1$ are related to the fundamental frequency. The others ($n = 2, 3, 4, \ldots$) are called harmonics.

The coefficients are given by

$$A_0 = \frac{1}{T}\int_{-T/2}^{T/2} y(t)\,dt, \quad A_n = \frac{2}{T}\int_{-T/2}^{T/2} y(t)\cos n\omega t\,dt, \quad B_n = \frac{2}{T}\int_{-T/2}^{T/2} y(t)\sin n\omega t\,dt$$

Which, similarly as for the lone sine and cosine terms can be written solely in terms of cosine or sine terms, but with an added phase term. Note, also, that if the function we consider if even, i.e. $f(-t) = f(t)$, then $B_n = 0$ for all n. Conversely, if we consider an odd function, i.e. $f(-t) = -f(t)$, then $A_n = 0$.

Fourier Transform

In the limit that $T \to \infty$, the sum becomes an integral, and we obtain the so-called Fourier transform:

$$Y(\omega) = \int_{-\infty}^{\infty} y(t)e^{-i\omega t}\,dt$$

or rather in terms of the (cyclical) frequency f, using the substitution $\omega = 2\pi f$. We consider the Fourier transform as a decomposition of $y(t)$ into amplitude versus frequency information. Conversely, we have the inverse Fourier transform,

$$y(t) = \int_{-\infty}^{\infty} Y(\omega) e^{i\omega t}\, dt$$

Note, $Y(f)$ can be thought of as having a magnitude and a phase:

$$Y(f) = |Y(f)|e^{i\phi(f)} = A(f) - iB(f)$$

where

$$|Y(f)| = \sqrt{\Re[Y(f)]^2 + \Im[Y(f)]^2} \quad \text{and} \quad \phi(f) = \arctan \frac{\Im[Y(f)]}{\Re[Y(f)]}$$

The amplitude expressed in frequency,

$$C(f) = \sqrt{A(f)^2 + B(f)^2}$$

is called the amplitude spectrum. And

$$\phi(f) = \arctan \frac{B(f)}{A(f)}$$

is called the phase spectrum.

Discrete Fourier Transform

Generally the signal is sampled N times every δt seconds. Such that measurements are represented as $y(t) \to \{y(r\delta t)\}$, for $r = 0, 1, \ldots, N-1$. And instead of a continuous Fourier transform, we would perform a discrete Fourier transform, defined as,

$$Y(f_k) = \frac{2}{N} \sum_{r=0}^{N-1} y(r\delta t) e^{-i2\pi rk/N}$$

where $f_k = k\delta f$ (for $k = 0, 1, 2, \ldots, N/2 - 1$) and $\delta f = 1/N\delta t$ is the frequency resolution. For instance, at $f = 0$ Hz, when $k = 0$,

$$Y(f_0) = \frac{1}{N} \sum_{r=0}^{N-1} y(r\delta t)$$

Only for $k = 0$ we use $1/N$, which yields the DC value at 0 Hz. Note that this is the so-called half-transform, yielding $N/2$ discrete values of the Fourier transform.

Chapter 3

Measurement System Modeling

We consider an input signal $F(t)$ and a measurement system operation with initial conditions $y(0)$, which produces an output $y(t)$.

A general model consists of an nth-order linear ordinary differential equation in terms of a general output signal represented by $y(t)$ and subjected to a general input, represented by the forcing function $F(t)$.

$$a_n \frac{d^n y}{dt^n} + \cdots + a_0 y = F(t)$$

where

$$F(t) = b_m \frac{d^m x}{dt^m} + \cdots + a_0 x$$

The coefficients represent the physical system parameters, with generally $m \leq n$.

Zero-order Systems

Zero-order systems are represented by,

$$a_0 y = F(t)$$

I.e. the output and forcing function are linearly related,

$$y(t) = KF(t) \quad \text{where } K = \frac{1}{a_0}$$

Noting that K is the static sensitivity, or rather the steady gain of the system. It is the slope of the static calibration curve, i.e. $K = dy/dx|_x$.

In practical scenarios, the zero-order system concept is used to model non-time-dependent measurement system responses to static inputs. It models appropriately any system during a static calibration.

When dynamic inputs are involved, the zero-order model is only valid at static equilibrium.

First-order Systems

First-order systems are modeled by,

$$a_1 \dot{y} + a_0 y = F(t)$$

These could, for instance, be measurement systems that contain storage elements, they do not respond instantaneously to changes in input. We generally write the differential equation in terms of a time constant and static gain,

$$\tau\dot{y} + y = KF(t)$$

where $\tau = a_1/a_0$ and $K = 1/a_0$.

Step function input

The step function is defined to be zero ($AU(t) = 0$) below some time $t < 0$ and for $t \geq 0$, $AU(t) = A$. A step function describes a sudden change in the input signal.

If we consider $y(0) = 0$, i.e. zero initial conditions, and the force function $F(t) = AU(t)$, we can solve the differential equation to obtain:

$$y(t) = KA + (y_0 - KA)e^{-t/\tau}$$

where the first term describes the steady-state response and the latter the transient response. The steady-state response is such that for $t \to \infty$, $y(t) \to KA$.

Error fraction

The error fraction of the output signal is defined as $\Gamma(t)$,

$$\Gamma(t) = \frac{y(t) - y_\infty}{y_0 - y_\infty}$$

Which for a step input is,

$$\Gamma(t) = e^{-t/\tau}$$

It decreases from a value of 1 and approaches a value of 0 with increasing t/τ. A 100% error fraction implies that the system has responded 0% to the input. So an error fraction of 0% implies that the system has responded fully to the input change.

We can also define the *rise time*, as the moment that the system has responded to 90% of the input, i.e. $\Gamma(t) = 0.1$.

The time constant

The time constant τ is a measure of how quickly a first-order measurement system responds to a change in input. A smaller time constant implies a shorter time between the application of the input and the full response.

This constant can be determined from the relation $\ln\Gamma = -t/\tau$, i.e. in plotting $\ln\Gamma$ as a function of the time t, with slope $-1/\tau$.

Periodic input signals

When periodic inputs are applied to the first-order system, the input signal frequency has an important influence on measuring the system time response and it affects the output signal.

We consider an input $F(t) = A\sin(\omega t)$,

$$\tau\dot{y} + y = KA\sin(\omega t)$$

The general solution is,

$$y(t) = Ce^{-t/\tau} + \frac{KA}{\sqrt{1+(\omega\tau)^2}}\sin\left[\omega t - \arctan(\omega\tau)\right]$$

The left most term is the transient periodic response and the second term is the steady periodic response. So when $t \to \infty$,

$$y(t) \to \frac{KA}{\sqrt{1+(\omega\tau)^2}}\sin\left[\omega t - \arctan(\omega\tau)\right]$$

Note that the output frequency is the same as the input frequency, however, the output amplitude now depends on the input signal. We define,

$$B(\omega) = \frac{KA}{\sqrt{1+(\omega\tau)^2}}$$

and

$$\phi(\omega) = -\arctan(\omega\tau)$$

Such that the solution can be written as,

$$y(t) = Ce^{-t/\tau} + B(\omega)\sin\left[\omega t + \phi(\omega)\right]$$

Furthermore, we can define a time delay related to the phase shift, $\beta = \phi/\omega$. Note that β is generally negative, indicating a delay between the output and input. In this manner we can discuss the frequency response of the output response.

The magnitude ratio $M(\omega)$ is the defined as the ratio between the output and input signal,

$$M(\omega) = \frac{B}{KA} = \frac{1}{\sqrt{1+(\omega\tau)^2}}$$

When the magnitude ratio is one, the phase approaches zero and time delay as approaches zero as well. Conversely, if $\omega\tau$ is larger, the amplitude is small and the time delay is large as well.

Note that we also define a *frequency bandwidth*, the frequency band over which,

$$M(\omega) \geq 0.707$$

or in decibels,

$$-3 \text{ dB} \leq M(\omega) \leq 0 \text{ dB}$$

We define the dynamic error,

$$\delta(\omega) = M(\omega) - 1$$

It is a measure of the inability of a system to adequately reconstruct the amplitude of the input signal for a particular input frequency. To minimize the dynamic error we try to obtain a magnitude ratio close to unity. However, this is essentially not possible in reality. The frequency response of a measurement system is found physically by a dynamic calibration.

Second-order Systems

Examples of 2nd order systems include systems with some acceleration, for instance in case of a pressure inducer, present in e.g. microphones.

The system is modeled by,

$$a_2\ddot{y} + a_1\dot{y} + a_0 y = F(t)$$

where the parameters are physical parameters describing the system. Note that we define the parameters,

$$\omega_n = \sqrt{\frac{a_0}{a_2}}$$

the natural frequency and the damping ratio (damping is the property of the system that enables it to dissipate energy internally),

$$\zeta = \frac{a_1}{2\sqrt{a_0 a_2}}$$

and the static sensitivity,

$$K = \frac{1}{\omega_0}$$

Such that the differential equation can be written as,

$$\frac{1}{\omega_n^2}\ddot{y} + \frac{2\zeta}{\omega_n}\dot{y} + y = KF(t)$$

Homogeneous solution

The solutions to the characteristic equation of the ODE is,

$$\lambda_{1,2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

Therefore, three solutions are possible:

1. Underdamped ($0 \leq \zeta < 1$): the transient response will be oscillatory,

$$y_h(t) = Ce^{-\zeta\omega_n t}\sin\left(\omega_n\sqrt{1-\zeta^2}t + \phi\right)$$

2. Critically damped ($\zeta = 1$): the demarcation between a oscillatory and non-oscillatory transient response,

$$y_h(t) = C_1e^{\lambda_1 t} + C_2te^{\lambda_2 t}$$

3. Overdamped ($\zeta > 1$): the transient response will not oscillate,

$$y_h(t) = C_1e^{\lambda_1 t} + C_2e^{\lambda_2 t}$$

Note: these homogeneous solutions determine the transient response of a system.

Step function response

A step function can be applied to determine the general behaviour and speed at which the system will respond to a change in input. Therefore, considering $F(t) = AU(t)$, we obtain three solutions for an underdamped, critically damped, and an overdamped system:

$$y(t) = KA - KAe^{-\zeta\omega_n t}\left[\frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\omega_n t\sqrt{1-\zeta^2}\right) + \cos\left(\omega_n t\sqrt{1-\zeta^2}\right)\right],$$

$$y(t) = KA - KA\big(1-\omega_n t\big)e^{-\omega_n t},$$

$$y(t) = KA - KA\left[\frac{\zeta+\sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}e^{(-\zeta+\sqrt{\zeta^2-1})\omega_n t} - \frac{\zeta-\sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}}e^{(-\zeta-\sqrt{\zeta^2-1})\omega_n t}\right]$$

where we used zero initial conditions. Note that for an underdamped system, the transient response is oscillatory. A critically damped system shows both oscillatory and non-oscillatory behaviour. In an overdamped system no oscillatory behaviour is present.

For the underdamped system we define a period T_d, of the oscillations of the system,

$$T_d = \frac{2\pi}{\omega_d}$$

where

$$\omega_d = \omega_n\sqrt{1-\zeta^2}$$

is called the ringing frequency (free oscillation frequency of a system's displacement from its equilibrium). Note that the ringing frequency is independent of the input signal. *(The experimental determina- tion of the ringing frequency associated with underdamped systems is performed by applying a step input to the second-order measurement system and measuring the response.)*

The duration of the transient response is controlled by the time constant, now defined as,

$$\tau = \frac{1}{\zeta\omega_n}$$

The system reaches the steady-state solution sooner when the time constant is smaller.

For the second order system we can also define the rise time, to be 90% of the time necessary to reach $(KA - y_0)$. Decreasing the damping ratio reduces the rise time. Severe ringing frequencies associated with very lightly damped systems delays the time to achieve a steady value compared to systems with higher damping.

The time for a measurement system's oscillations to settle within $\pm 10\%$ of the steady-state value is defined as the settling time.

Periodic Response

The output due to a sinusoidal input $F(t) = A\sin(\omega t)$ is given by,

$$y(t) = y_h(t) + \frac{KA\sin(\omega t + \phi(\omega))}{\sqrt{\left[1 - (\omega/\omega_n)^2\right]^2 + \left[2\zeta\omega/\omega_n\right]^2}}$$

where

$$\phi(\omega) = \arctan\left[-\frac{2\zeta\omega/\omega_n}{1 - (\omega/\omega_n)^2}\right]$$

Note that the steady-state solution is frequency dependent and given by,

$$y_\infty = B(\omega)\sin[\omega t + \phi(\omega)]$$

We can define the magnitude ratio for a 2nd order system as,

$$M(\omega) = \frac{1}{\sqrt{\left[1 - (\omega/\omega_n)^2\right]^2 + \left[2\zeta\omega/\omega_n\right]^2}}$$

For an ideal system the magnitude ratio approaches unity and the phase shift goes to zero. In reality generally, the magnitude ratio tends to approach zero if the frequency increases, and the phase shift goes to $-\pi$.

For a system without damping, $\zeta = 1$, which don't exist in reality, $M \to \infty$ and $\phi \to -\pi$ when $\omega \geq \omega_n$.

Peak resonance occurs in underdamped systems when the input frequency reaches the natural frequency.

$$\omega_R = \omega_n \sqrt{1 - 2\zeta^2}$$

Resonance behaviour results in a considerable phase shift, is very non-linear, and results in distortion of the signal. Operating at frequencies in the resonance band is confusing and could damage the system. Note that systems where $\zeta > 1/\sqrt{2}$, do not resonate. When $\omega/\omega_n \ll 1$, the magnitude ratio is close to 1 and the phase shift is close to 0. So when the input frequency is much smaller than the output frequency the output frequency the signal is passed with very little alterations. This is the case for,

$$-3\text{dB} \leq M(\omega) \leq 3 \text{ dB}$$

the so-called transmission bandwidth. We can also consider the filter band, for which $M(\omega) \leq -3$ dB. In this case $M \approx 0$ and $\phi \approx -\pi$. This filter band essentially removes undesirable features from a desirable product. I.e. amplitudes of the input signal corresponding to these frequencies are surpressed.

Transfer Functions

The transfer function defined the mathematical operation that the measurement system performs on the input signal $F(t)$ to yield the response of the system. If we consider a first-order system, with input $F(t)$, initial values $y_0 = y(t = 0)$, being operated on by some function $G(s)$ (in Laplace space) then the Laplace transform of the system is,

$$Y(s) = \frac{KF(t) + y_0}{\tau s + 1}$$

with is written as $Y(s) = G(s)[KF(s) + Q(s)]$, with $G(s) = (\tau s + 1)^{-1}$ and $Q(s)$ being the initial state of the system.

We can find the system frequency response by evaluating $G(s = i\omega)$, noting that,

$$G(i\omega) = M(\omega)e^{i\phi(\omega)}$$

Multiple-function Inputs

When using models that are linear, such as ordinary differential equations subjected to inputs that are linear in terms of the dependent variable, the principle of superposition of linear systems applies in the solution of these equations. The principle of superposition states that a linear combination of input signals applied to a linear measurement system produces an output signal that is simply the linear addition of the separate output signals that would result if each input term had been applied separately. Because the form of the transient response is not affected by the input function, we can focus on the steady response. In general, we can write that if the forcing function of a form,

$$F(t) = A_0 + \sum_{k=1}^{\infty} A_k \sin \omega_k t$$

is applied to a system, then the combined steady response will have the form,

$$y_{\text{ss}} = KA_0 + \sum_{k=1}^{\infty} B(\omega_k) \sin \left[\omega_k t + \phi(\omega_k)\right]$$

where $B(\omega_k) = KA_k M(\omega_k)$.

Coupled Systems

The overall static sensitivity of a coupled system, containing H interconnected devices, is given by,

$$K = K_1 K_2 \cdots K_H$$

similarly the overall system magnitude is the product,

$$M(\omega) = M_1(\omega) M_2(\omega) \cdots M_H(\omega)$$

with an overall phase shift of,

$$\phi(\omega) = \phi_1(\omega) + \cdots + \phi_H(\omega)$$

Chapter 4

Sampling, Statistics, and Systematic Errors

Sampling is obtaining a data set through repeated measurements of a variable under fixed operating conditions.

The statistical error is the (unknown) difference between the retained value and the true value:

$$x' = \bar{x} \pm u_{\bar{x}} \ (P\%)$$

where from left to right we have the true value, the most probable estimate, the uncertainty interval of the estimated value $\bar{x}$, and the probability level P.

Uncertainties are numbers that quantify the possible range of the effects of the errors. Systematic errors do not vary with repeated measurements and do not affect the statistics of the measurement.

Histogram

A graphical way of grouping data is different ranges (bins), the height shows how many data points fall in the range.

The total number of measurements N is the sum of all the data in the bins,

$$N = \sum_{j=1}^{K} n_j$$

where n_j is the number of times that a measured value assumes a value within a certain interval. Note we can convert a histogram to a frequency distribution via $f_j = n_j/N$. Note that the area under a frequency distribution (what we could consider as the PDF) is 100%, i.e. $\sum_j f_j = 1$.

The interval number K (i.e. answering the question: *how many bins should one use?*) is given by,

$$K = 1.87(N-1)^{2/5} + 1$$

For large N, we can use $K \approx \sqrt{N}$.

Standard probability Distributions

- The normal distribution,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-x'}{\sigma}\right)^2 \right]$$

  where x' is the true value, x the measured value, and σ^2 the true variance of x. 68.26% of the data falls within 1σ, 95.45% within 2σ, and 99.73% within 3σ.
- The Poisson distribution, for random events in time, describing observing x events in time t,

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

  where λ is given by x'.
- The binomial distribution, which described the number of occurrences, n, of a particular outcome during N independent tests, each having the same probability P,

$$p(n) = \left[\frac{N!}{(N-n)!n!} P^n (1-P)^{N-n} \right]$$

Given a PDF $p(x)$, the true value is given by,

$$x' = \int_{-\infty}^{\infty} x p(x)\, dx$$

and the true variance σ^2 is given by,

$$\sigma^2 = \int_{-\infty}^{\infty} (x - x')^2 p(x)\, dx$$

for a continuous variable x. For discrete data,

$$x' = \lim_{N\to\infty} \frac{1}{N} \sum_i x_i$$

and

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum_i (x_i - x')^2$$

Population Statistics

We consider the data to be normally distributed. The probability that we find a certain value on a certain interval is given by,

$$P(x' - \delta x \leq x \leq x' + \delta x) = \int_{x'-\delta x}^{x'+\delta x} p(x)\, dx$$

We define a transformation $\beta = (x - x')/\sigma$, with which we define the statistic z_1,

$$z_1 = \frac{x_1 - x'}{\sigma}$$

Such that we write,

$$P(-z_1 \leq \beta \leq z_1) = \frac{1}{\sqrt{2\pi}} \int_{-z_1}^{z_1} e^{-\beta^2/2}\, d\beta = 2\text{erf}(z_1)$$

Using these definitions we find that the probability that some ith measured value of x will have a value $x' \pm z_1 \times$ something is $2P(z_1) \times 100\%$. Which is written as,

$$x_i = x' \pm z_1\sigma \quad (P\%)$$

Finite-sized Datasets Statistics

Suppose we have N measurements x_i, the sample mean is then given by,

$$\bar{x} = \frac{1}{N} \sum x_i$$

and the sample variance is given by,

$$s_x^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2$$

For a normal distribution we can state that,

$$x_i = \bar{x} \pm t_{\nu,P} s_x \quad (P\%)$$

where $t_{\nu,P}$ is a coverage factor (replacing the z-statistic) for finite length datasets and $\nu = N - 1$ is the degrees of freedom in the sample variance (the -1 coming from the sample means). Note that the degrees of freedom is defined as the number of data points minus the number of previously determined statistical estimator (it indicates the number of independent values that can vary in an analysis without breaking any constraints). The t variable is defined as,

$$t = \frac{\bar{x} - x'}{s_{\bar{x}}}$$

where $s_{\bar{x}} = s_x/\sqrt{N}$ is the standard deviation of the means. Which provides a measure of how well a measured mean value from a small sample size represents the true mean of the entire population. Such that we find that the range of possible values of the mean is given by

$$x' = \bar{x} \pm t_{\nu,P} s_{\bar{x}} \quad (P\%)$$

where P is some probability level and $t_{\nu,P} s_{\bar{x}}$ is the confidence interval about the mean value with coverage factor t.

Pooled Data

Suppose we have N measurements of a large population under fixed conditions, and we duplicate this process M times. Then the sample mean and variance will be slightly different each time.

Hypothesis Testing

A hypothesis is a statement that something is believed true. A hypothesis test uses the tools of statistics to test it. E.g. we want to test whether a measured sample mean is a reasonable proxy for the population mean comparing information of the observed scatter in the data set against the assumed distribution of the population. Then, the null hypothesis is expressed as $H_0 \ : \ x = x_0$.

Definition of a hypothesis test

1. Establish the null hypothesis.
2. Assign a level of significance, α, to determine the critical values; this sets the rejection region.
3. Calculate the observed value of the test statistic.
4. Compare the observed test statistic to the critical values. If the observed statistic is in the rejection region, reject the null hypothesis; otherwise, do not reject.

z-test

The z-test can be used when the population variance σ^2 is known. As test statistic we use the z-variable with $x' = x_0$,

$$z_0 = \frac{\bar{x} - x_0}{\sigma/\sqrt{N}}$$

This statistic is evaluated against critical values of z at a desired level of significance, α, where $P(z) = 1 - \alpha$. For a two-tailed test the acceptance region is defined by $P(-z_{\alpha/2} \leq z \leq z_{\alpha/2})$, for one-tailed tests we use $P(-z_\alpha \leq z_0)$ or $P(z_0 \leq z_\alpha)$. We then reject the null hypothesis if the value lies outside the acceptance region. Most often we find $\alpha = 0.05$ is adequate.

t-test

Similar to the z-test, but now we don't know the population variance and therefore we use the t-variable.

p-value

The p-value reports the probability that the difference between two tested values is due only to random chance. It is the observed level of significance. A very small p-value indicates that effects other than random chance are influencing the variations in a sampling. To calculate the p-value, assume that the null hypothesis is true and compute the observed level of significance corresponding to the test statistic. If this p-value is less than α, we reject H_0, otherwise we do not reject H_0.

Chi-squared Distribution

The χ^2-distribution defines the distribution of the sample standard deviation for many datasets, each with N data points. For the normal distribution $\chi^2 = \nu s_x^2/\sigma^2$, where $\nu = N - 1$. We can use this distribution to test how well the sample variance estimates the population variance.

We define $P(\chi^2_\alpha) = 1 - \alpha$ and by computing χ^2 given the available information and a χ^2_α table we find the precision interval.

Goodness-of-fit test

To test how well a set of measurements follow an assumed distribution function. To do this we apply the χ^2-test, which would provide a measure of the discrepancy between the measured variation of a dataset and that due to chance as predicted by the density function.

Construct a histogram of K intervals a from a dataset of N measurements. The number of data points in each jth interval is n_j. Calculate the degrees of freedom in the variance for the dataset, $\nu = N - m$, where m is the number of restrictions imposed. Lastly, estimate the number of occurrences n'_j to be expected from the distribution function.

$$\chi^2 = \frac{\sum_j (n_j - n'_j)^2}{n'_j}$$

for $j = 1, 2, \ldots, K$. The better a dataset fits the proposed distribution the lower the χ^2 value, ideally it approaches unity. Too large a value will imply a rather dubious fit. Similarly, a value much smaller than unity is also not desirable. These problems might be due to too small error bars or measurement errors.

Regression Analysis

Regression analysis can be used to establish a functional relationship between the dependent and independent variable. Generally assuming that the variation found in the measured variable is normally distributed.

Least-Squares Regression

We want to fit an mth order polynomial,

$$y_c = a_0 + a_1 x + \cdots + a_m x^m$$

for $m \leq n-1$, where n is the number of different values of the independent variables included in the analysis. In LS fitting we attempt to minimize the squared deviation,

$$D = \sum_{i=1}^{N} (y_i - y_{c_i})^2$$

The standard deviation, considering the pairs (x_i, y_i), is given by,

$$s_{yx} = \sqrt{\frac{\sum (y_i - y_{c,i})^2}{\nu}}$$

where $\nu = N - (m+1)$. Note s_{yx} is called the standard error of the fit, which is related to how closely a polynomial fits the data. Generally, the best order of polynomial fit to apply is one that makes logical physical sense. The confidence interval of a fit is defined (and approximated) as,

$$y_c \pm t_{\nu,P} \frac{s_{yx}}{\sqrt{N}}$$

This confidence interval is also called the *Scheffe band*.

Correlation

We can define the correlation coefficient r, which provides a measure of the association between x and y as predicted by the form of the curve fit equation. Perfect correlation is marked by $r = \pm 1$, for $\pm 0.9 \leq r \leq \pm 1$ the fit can still be considered reliable.

The coefficient of determination r^2, indicates how well the variance in y is account for by the fit.

Outlier Detection/Rejection

Outliers may be the result of simple measurement glitches, such as a random spike in operating condition, or may reflect a more fundamental problem with test variable controls. If outliers are removed from a data set, the statistics are recomputed using the remaining data.

We can detect outliers using Chauvenet's criterion, which identifies outliers having less than a $1/2N$ probability to occur. Consider the test statistic $z_0 = |x_i - \bar{x}|/s_x$, then the data point is an outlier if,

$$1 - 2P(z_0) < \frac{1}{2N}$$

The Number of Required Measurements

To find how many measurements N are required to reduce the estimated value for random error in the sample mean to an acceptable level. We find,

$$N \approx \left(\frac{t_{\nu,P} s_x}{d}\right)^2 \quad (P\%)$$

where d is the precision value, defined as,

$$d = \frac{t_{\nu,P} s_x}{\sqrt{N}}$$

Monte Carlo Simulations / Methods

Monte Carlo experiments are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. Using randomness to solve deterministic problems.

These methods are mainly used for optimization, numerical integration, and generating samples from a PDF. In physics, e.g. simulating systems with many coupled degrees of freedom.

Consider a result R, a function of variables x and y (drawn from $p(x)$ and $p(y)$). Each iteration we randomly draw values x_i and y_i, and compute R_i. After $\sim 10^5$ iterations be result approaches the true distribution of R. We generally continue until the standard deviation of R converges to an asymptotic value, within for instance 1-5%.

The uncertainty of Monte Carlo results depends on the number of *histories* (samples) N, with the error scaling as $1/\sqrt{N}$.

Chapter 5

General Definitions

Errors are effects, and uncertainties are numbers. Specifically,

- Errors are a property of the measurement: which causes a difference in the measured and true value of the population of the variable measured. They come from individual instrument calibrations, finite data set statistic, or the measurement method/approach.
- Uncertainty is a property of the result: it describes an interval about the measured value within the true value falls.

In short: the errors are *effects* and the uncertainties are *assigned numerical values*.

In uncertainty analysis we assume,

1. The test objectives are known and the measurement itself is a clearly defined process.
2. Any known corrections for systematic error have been applied to the data set, in which case the systematic uncertainty assigned to that error is the uncertainty of the correction.
3. Except where stated otherwise, we assume a normal distribution of errors and reporting of uncertainties.
4. Unless stated otherwise and for simplicity, the errors are assumed to be independent (uncorrelated) of each other.

We often begin with a *general* analysis, in which no distinction is made between random and systematic errors. In a general analysis, we simply assign uncertainty values to errors as indicated by a value u. This is particularly useful when we have little information about the basis of each uncertainty value or its effect on the test outcome or are making a quick analysis perhaps to guide further test decisions.

When we have sufficient information to distinguish between systematic and random errors, so we can assign uncertainty values separately for each error, enabling a *detailed* analysis.

Design-Stage Uncertainty Analysis

This refers to an analysis performed early in the formulation of a test. To obtain a quick estimate of the minimum uncertainty that is expected.

The zero-order uncertainty of an instrument, u_0, attempts to estimate the expected variation in the measured values. As a minimum, this variation will be an amount on the level of the instrument resolution (i.e., interpolation error). Assign a numerical value to u_0 of either one-half of the analog instrument resolution or its digital least count. This value will reasonably represent the uncertainty interval on either side of the reading with a probability of 95%. Then,

$$u_0 = \frac{1}{2}\text{resolution} = 1 \text{ LSD}$$

where LSD refers to the least significant digit of the readout (its least count).

Combining Errors

Each individual measurement error interacts with other errors to affect the overall uncertainty of a measurement. Consider a measurement of x that is subject to some K elements of error, each of uncertainty u_k, where $k = 1, 2 \ldots, K$. Then the total uncertainty can be given by,

$$u_x = \sqrt{\sum_{k=1}^{K} u_k^2} \quad (P\%)$$

The RSS method of combining uncertainties is based on the assumption that the square of an uncertainty is a measure of the variance assigned to an error, and the summation of these variances yields a probable estimate of the total uncertainty.

Design-Stage Uncertainty

We define this uncertainty by $u_d = \sqrt{u_0^2 + u_c^2}$, where u_c is the instrument uncertainty. Where u_c is information available from the manufacturer of the instrument. Assuming a 95% probability level (i.e. a 2σ coverage), assuming that no probability level is given.

Note: this estimate (u_d) is solely used for selecting a measurement device, rather than for a final reported uncertainty.

Error Sources

- Calibration errors: errors in the standard/reference value, instrument/system errors, calibration process errors, curve fit errors, etc.
- Data-acquisition errors: measurement operation condition errors, instrument errors, environmental effects, etc.
- Data reduction errors: curve fit errors, truncation errors, modeling errors, etc.

Again we have systematic errors and random errors:

- Systematic errors remain constant in repeated measurements under fixed conditions. We can estimate the probable range of the systematic uncertainty by the so-called *standard systematic uncertainty* b, such that for $\pm b$ we have 1σ coverage. The systematic uncertainty for some specific confidence level is then $t_{\nu,P}b$. For a 95% probability coverage we assume, unless otherwise states, an uncertainty $B = 2b$ for 2σ coverage.
- Random errors manifest themselves in repeated measurements under fixed conditions, the effect of which can be easily observed. They are introduced through the repeatability of the measurement system components, calibration, and the measurement procedure and technique. The estimate of the probable range of a random error is assigned to a value, the *random standard uncertainty* $s_{\bar{x}}$ (at 1σ confidence). More generally we use $t_{\nu,P}s_{\bar{x}}$.

Note that we sometimes use the terms:

- *Type A* uncertainty: assigned based on statistical analysis of the data set.
- *Type B* uncertainty: assigned through non-statistical means.

These could both be random or systematic of nature.

Multivariable Error Propagation

Assume we consider a result R depending on L independent variables,

$$R = f(x_1, x_2, \ldots, x_L)$$

Each of those variables contains some measure of uncertainty that affects the result. Tee best estimate of the true mean would be,

$$R' = \bar{R} \pm u_R$$

where the sample mean of R is given by,

$$\bar{R} = f(\bar{x}_1, \ldots, \bar{x}_L)$$

and the uncertainty is found from,

$$u_R = f(u_{\bar{x}_1}, \ldots, u_{\bar{x}_L})$$

where each uncertainty in the independent variables represents the uncertainty associated with its best estimate. The total uncertainty in R is then given by the square root of the sum of the squares (RSS) of $\theta_i u_{\bar{x}_i}$ ($\theta_i = \partial R/\partial x_i|_{x=\bar{x}}$ is a sensitivity index, which is a measure of how R is affected by a change in x_i),

$$u_R = \sqrt{\sum_{i=1}^{L} \left(\frac{\partial R}{\partial x_i} u_{\bar{x}_i} \right)^2}$$

Instead of via direct differentiation we can also apply dithering, applying a small perturbation to the mean and approximating the derivative. Consider a function $y = f(x)$, for $x = x_0$ we perturb the function, such that we have $y^{\pm} = x \pm \delta x$,

$$\left. \frac{\partial y}{\partial x} \right|_{x=x_0} = \frac{y^+ - y^-}{2\delta x}$$

Note that δx should be sufficiently small. Additionally, we can use tabulated function data.

Similarly, we can use a method called *sequential perturbation* to compute the uncertainty in a result directly via perturbations.

1. Calculate R_0 based on variables under some fixed operating conditions.
2. Increase each independent variable sequentially, to obtain perturbations $R_i^{\pm} = f(\ldots, x_i \pm u_{x_i}, \ldots)$.
3. Compute, $\delta R_i^{\pm} = R_i^{\pm} - R_0$ for all $i = 1, 2, \ldots, L$.
4. Compute,
$$\delta R_i = \frac{\delta R_i^+ - \delta R_i^-}{2} \approx \theta_i u_i$$
Such that,
$$u_R = \sqrt{\sum_{i=1}^{L} (\delta R_i)^2}$$

Advanced-Stage Uncertainty Analysis

We consider it a method for a thorough uncertainty analysis when a large data set is not available. This is often the case in the early stages of a test program or for certain tests for which repeating measurements may not be possible. Such an advanced-stage analysis, also known as single-measurement uncertainty analysis, can be used (1) in the advanced design stage of a test to estimate the expected uncertainty beyond the initial design stage estimate and (2) to report the results of a test program that involved measurements over a range of one or more parameters but with no or relatively few repeated measurements of the pertinent variables at each test condition. Essentially, the method assesses different aspects of the main test by quantifying potential errors though various well-focused verification tests.

Zero-Order Uncertainty u_0

All variables and parameters that affect the measurement are assumed to be fixed (including time), *except for the act of observation* itself.

Higher-Order Uncertainty

Higher-order uncertainty estimates consider the controllability of the test operating conditions and the variability of all measured variables. For example, at the first-order level, the effect of time as an extraneous variable in the measurement might be considered. A set of data (e.g. $N \geq 30$) would be obtained under some set operating condition. The first-order uncertainty of our ability to estimate the true value of a measured value could be estimated as,

$$u_1 = t_{\nu,P} s_{\bar{x}}$$

For example, at the second level it might be appropriate to assess spatial variations that affect the outcome, such as when a value from a point measurement is assigned to quantify a larger volume.

At the Nth-order uncertainty, instrument calibration characteristics are entered into the scheme through the instrument uncertainty u_c. Such that,

$$u_N = \sqrt{u_c^2 + \sum_{i=1}^{N-1} u_i^2}$$

Uncertainty estimates at the Nth order allow for the direct comparison between results of similar tests obtained either using different instruments or at different test facilities

Multiple-Measurement Uncertainty Analysis

Propagation of Elemental Errors

A scheme for combining multiple measurement uncertainties is depicted in the graph below.

Note that the degrees of freedom ν of respectively the *random measurement uncertainty* and final measurement uncertainty u_x correspond to,

$$\nu = \frac{\left(\sum_{k=1}^{K}(s_{\bar{x}}^2)_k\right)^2}{\sum_{k=1}^{K}(s_{\bar{x}}^4)_k/\nu_k}$$

where $\nu_k = N_k - 1$, and,

$$\nu = \frac{\left(\sum_{k=1}^{K}(s_{\bar{x}}^2)_k + (b_{\bar{x}}^2)_k\right)^2}{\sum_{k=1}^{K}(s_{\bar{x}}^4)_k/\nu_k + \sum_{k=1}^{K}(b_{\bar{x}}^4)_k/\nu_k}$$

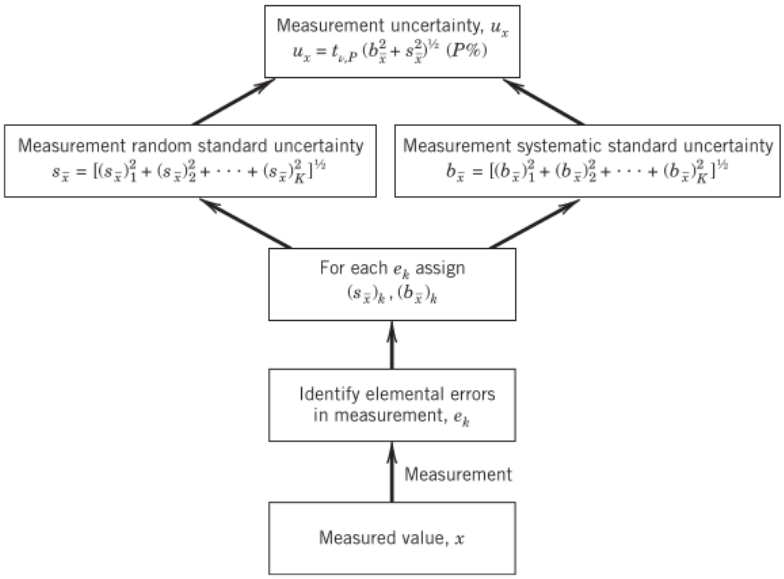

Figure 1: Multiple measurement uncertainty analysis scheme

Propagation of Uncertainty to a Result

Consider the result of some functional relationship as given by,

$$R' = \bar{R} \pm u_R$$

Note,

$$u_R = t_{\nu,P}\sqrt{b_R^2 + s_R^2}$$

where,

$$s_R = \sqrt{\sum_{i=1}^{L} (\theta_i s_{\bar{x}_i})^2}$$

and,

$$b_R = \sqrt{\sum_{i=1}^{L} (\theta_i b_{x_i})^2}$$

Chapter 6

Information from electrical analog devices is often transferred in form of an electrical signal between stages of a measurement system. Analog electrical signals typically originate from measurements of a physical variable using a fundamental electromagnetic or electric phenomenon.

Note that Kirchhoff's laws are,

$$\sum_k E_k = 0 \quad \text{(around a closed loop)}$$

and,

$$\sum_k I_k = 0 \quad \text{(to/from a node)}$$

Measuring the Current

Direct Current

A DC current is the flow of electric charge in a single direction, it is the steady state of a constant voltage circuit.

We can measure DC currents using an analog device that responds to the force exerted on a current-carrying conductor in a magnetic field.

A current loop in a magnetic field experiences a torque,

$$T_\mu = NIAB \sin \alpha$$

when the loop is not aligned with the magnetic field. Here A is the cross-sectional area defined by the perimeter of the current loop and α the angle between the normal to the cross-sectional area of the current loop and the magnetic field.

Lot's of devices used to measure the current contain a so-called D'Arsonval movement, which employs a pointer whose deflection increases with the magnitude of the current applied. This is called a *deflection mode*.

One such sensor-transducer using this method is called a *galvanometer*, which can be used to detect a current flow in a circuit. This employs a mechanism that attempts to put the circuit into a zero current state. This is a so-called *null mode* of operation.

These devices contain hysteresis, repeatability errors, and linearity error. Along with a so-called *loading error*, which is the draining of energy from a signal being measured changing the measured signal.

Alternating Current

AC current periodically reverses direction and changes its magnitude continuously with time.

We can measure AC currents in various ways, for instance,

- We can use a rectifier to convert the AC current to a DC current. This can be achieved via a *rectifier* (an arrangement of diodes).
- Using an *electrodynamometer*, essentially a D'Arsonval movement device but with an electromagnet instead of a permanent magnet.
- We can use a Hall effect probe, which is clamped over the current-carrying wire (conductor) to measure its unknown current. Note that the Hall effect is the voltage developed from a conductor placed perpendicular to a magnetic field. For a known current the magnitude of the voltage depends on the magnitude of the magnetic field. The Hall-effect sensor is a thin conducting semiconductor wafer driven by a known current.

Measuring Voltages

Analog Voltage Meters

We can measure a DC voltage through a circuit which uses a resistor placed in series with a D'Arsonval movement. This would use a certain known fixed resistor which, through Ohm's law $E = RI$, allows us to relate the D'Arsonval movement in terms of voltage.

AC voltages can be measured, similarly, through rectification or through the use of an electromagnet. Note that these instruments are sensitive to the *rms* value of the AC current, and can be calibrated in terms of the voltage.

Oscilloscope

It is a graphical display device, generally displaying the voltage magnitude as a function of time. It is able to measure frequencies well into the GHz range.

Potentiometer

We use a potentiometer for DC voltages in the micro to millivolt range. A potentiometer balances an unknown input voltage against a known internal voltage until both sides are equal. Thus it is a *null balance* instrument, which drives the loading error to essentially zero. Note that the *loading error* is the normalized difference between the output voltage with an infinite meter resistance and that with a finite meter resistance.

A component found in these potentiometers are *voltage divider circuits*. Resistor voltage dividers are commonly used to create reference voltages, or to reduce the magnitude of a

voltage, so it can be measured. Given an input voltage E_i, the output voltage is,

$$E_o = \frac{R_x}{R_T} E_i$$

where R_T is the total resistance and R_x is a variable resistance depending on the position of the sliding contact (thus note that $R_x < R_T$).

An example of such a potentiometer based instrument might use a voltage divider and a galvanometer as shown in the figure below. It uses a *null balance* scheme, i.e. when the voltages $E_{AB} = E_m$ (E_m is the measured voltage) are equal, the galvanometer does not report any current flow. Note that E_i is a known supply voltage.

Note: *the current flow measured by G is the result of imbalance between E_m and E_{AB}. And with a known and constant E_i, the position of A can be calibrated to indicate E_m directly.*

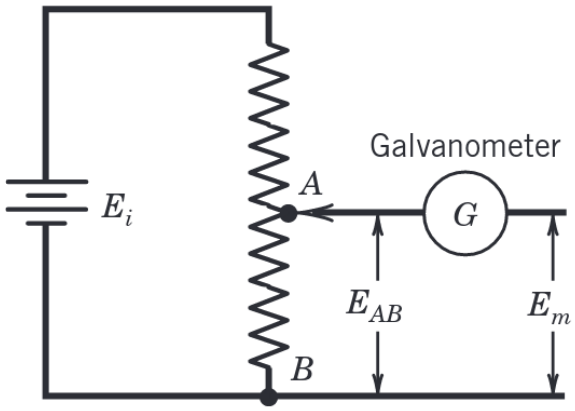

Figure 2: Basic potentiometer circuit

Resistance Measurements

Ohmmeter Circuits

One way to measure resistance is by imposing a voltage across the unknown resistance and measuring the resulting current flow where $R = E/I$. This is the basis of common analog Ohmmeters. They tend to use a D'Arsonval mechanism with a shunt resistor, to limit the flow of current through the meter movement.

Bridge Circuits

Bridge circuits can be used to measure capacitance ($C = Q/E$ or $I = Q\dot{E}$), inductance ($L = E/\dot{I}$), but are most often used for resistance measurements.

A common bridge circuit is the Wheatstone bridge (see the figure below), where R_1 some sensor that experiences a change in resistance associated with a change in some physical variable. Furthermore, a DC voltage is applied across nodes A and D.

When no current flows through the galvanometer, $I_g = 0$, we say that the bridge is in a balanced condition. This is characterized by the following systems of equations,

$$I_1 R_1 - I_3 R_3 = 0$$

$$I_2 R_2 - I_4 R_4 = 0$$

In that case the currents through the arms of the bridge are equal, i.e. $I_1 = I_2$ and $I_3 = I_4$. This implies that,

$$\frac{R_2}{R_1} = \frac{R_4}{R_3}$$

which is a necessary condition for the resistances of the balanced bridge.

This circuit can be used to measure resistance (changes) if the resistor R_1 varies with changes in the measured physical variable, then one of the other arms of the bridge can be adjusted to null the circuit and determine the resistance. Or we use a voltage measuring device to measure the voltage unbalance in the bridge as an indication of the change in resistance.

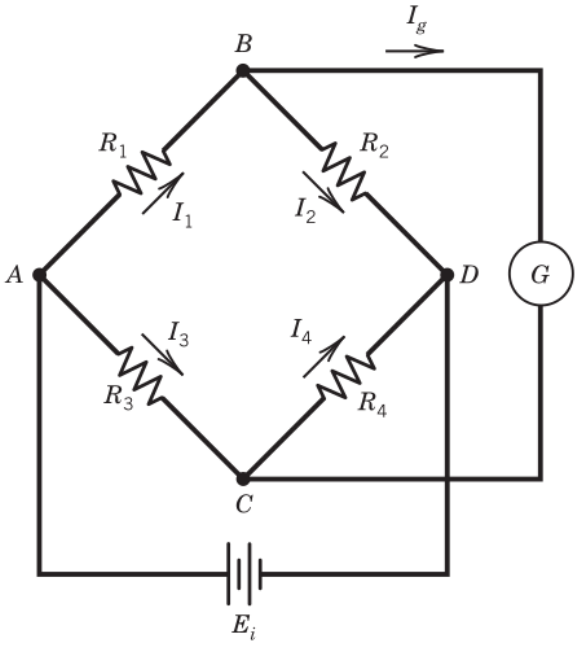

Figure 3: A Wheatstone bridge with a galvanometer

Null Method

Now consider the same figure, however, now with R_2 being a variable resistor. We adjust R_2 so that the bridge is balanced. Note that R_2 should be a calibrated variable resistor, such that any adjustments in R_2 directly indicate the value of R_1.

We can balance manually or through a closed-loop controller. Advantages of the null method are:

1. The applied input voltage need not be known, and changes in the input voltage do not affect the accuracy of the measurement.
2. The current detector or controller need only detect if there is a flow of current, not measure its value.

Deflection Method

In an unbalanced condition, the magnitude of the current or voltage drop for the meter portion of the bridge is a direct indication of the change in resistance in one or more of the arms.

Deflection Under Balanced Conditions: If we consider a meter with infinite internal impedance (no current flow through the meter), the output voltage is given by,

$$E_o = I_1 R_1 - I_3 R_3$$

Such that with the equations for a balanced bridge we have,

$$E_o = E_i \left(\frac{R_1}{R_1 + R_2} - \frac{R_3}{R_3 + R_4} \right)$$

If initially $E_o = 0$ and $R_1 = R_2 = R_3 = R_4 = R$, we find that the change in output voltage is given by,

$$\frac{\delta E_o}{E_i} = \frac{\delta R / R}{4 + 2(\delta R / R)}$$

Note: that in contrast to the null method, the deflection method requires a meter capable of accurately indicating the output voltage, as well as a stable and known input voltage. The bridge should respond to any resistance changes over frequency input. Thus, it is often used for time varying signals.

Deflection Under Unbalanced Conditions: We know have a current sensitive bridge. We assume that the current sensing device has a resistance R_g. We find that the current through the meter I_g is given by,

$$I_g = \frac{E_i (R_1 R_4 - R_2 R_3)}{R_1 R_4 (R_2 + R_3) + R_2 R_3 (R_1 + R_4) + R_g (R_1 + R_2)(R_3 + R_4)}$$

if we consider voltage drops in the path through R_1, R_g, and R_3. Now we have $E_o = I_g R_g$. And in terms of the bridge deflection voltage E_o, the change in resistance R_1 is found by,

$$\frac{\delta R}{R_1} = \frac{(R_3/R_1)\left[E_o/E_i + R_2/(R_2+R_4)\right]}{1 - E_o/E_i - R_2/(R_2+R_4)} - 1$$

Now if we again consider all resistances initially equal, R, and $R_1 = R + \delta R$,

$$I_g = E_i \frac{\delta R/R}{4(R+R_g)}$$

with output voltage,

$$E_o = E_i \frac{\delta R/R}{4(1+R/R_g)}$$

Note: in case we have a voltage source with some internal resistance R_s, and with the effective bridge resistance R_B (see equation 6.23 in the book), we find that the input voltage is slightly different,

$$E_i = \frac{E_s R_B}{R_s + R_B}$$

In a similar manner, the bridge impedance can affect the voltage indicated by the voltage measuring device. For a voltage-measuring device of internal impedance R_g, the actual deflection voltage, relative to the indicated voltage (E_m) is,

$$E_o = \frac{E_m}{R_g}\left(\frac{R_1 R_2}{R_1+R_2} + \frac{R_3 R_4}{R_3+R_4} + R_g\right)$$

The difference between $E_m$and E_o is a loading error, due to the bridge impedance load.

Loading Errors & Impedance Matching

Any effect that affects the measured variable is considered a *loading* that the measurement system exerts on that variable. The *loading error* is the difference between the measurand and the indicated value. Loading occurs somewhere along the signal path. For instance *interstage loading errors* occur due to the output from one stage being affected by the subsequent stage.

Loading Errors for Voltage-Dividers

Consider the voltage divider in figure 6.14. As a sliding contact moves it divides the full-scale deflection resistance R into R_1 and R_2, such that the total resistance is $R_T = R_1 + R_2$. The measured resistance is R_m, and the current flow from the voltage source is,

$$I = \frac{E_i}{R_{\text{eq}}} = \frac{E_i}{R_2 + R_1 R_m/(R_1+R_m)}$$

where $R_{\text{eq}} = R_2 + R_L$ (R_L is the equivalent resistance from the parallel resistors R_m and R_1). Such that the output voltage is given by,

$$E_o = E_i - IR_2 \quad \Rightarrow \quad \frac{E_o}{E_i} = \frac{1}{1 + (R_2/R_1)(R_1/R_m + 1)}$$

Note: if we have $R_m \to \infty$,

$$\frac{E_o}{E_i} = \frac{R_1}{R_1 + R_2}$$

If we define this last expression as the true value value $(E_o/E_i)'$, then the loading error e_I may be given by,

$$e_I = E_i \left[\left(\frac{E_o}{E_i} \right) - \left(\frac{E_o}{E_i} \right)' \right]$$

Thus, as $R_m \to \infty$ we have that $e_I \to 0$.

Interstage Loading Errors

Consider a common situation in the measurement chain in which the output voltage signal from one system device provides the input signal to the following device.

An open circuit voltage E_1 will be present at the output terminal of device 1, with equivalent impedance Z_1 (*impedance extends the concept of resistance to AC circuits; note that for an ideal resistor $Z_R = R$, for a capacitor $Z_C = 1/j\omega C$, and for an inductor $Z_L = j\omega L$*). Then device 2 is attached across the terminals. This is equivalent to placing an impedance Z_m across the terminals. This impedance acts as a load on the first device. Such that the voltage sensed by device 2 is,

$$E_m = IZ_m = \frac{E_1}{1 + Z_1/Z_m}$$

where I is the current that is flowing due to placing Z_m across the terminals. The voltage differs from the original voltage E_1, so the loading error is $e_I = E_m - E_1$. Thus to minimize the loading error, we require that the interstage measuring device has a high input impedance, $Z_m \gg Z_1$.

If the signal is current-driven, maximum current transfer between devices 1 and 2 is desirable. With the extra impedance the current measured by device 2 is,

$$I_m = \frac{E_1}{Z_1 + Z_m}$$

whereas otherwise $I' = E_1/Z_1$. Such that the loading error is $e_I = I_m - I'$. So to reduce the current-driven loading error, $Z_m \ll Z_1$. When such a current loop is used, the measured voltage is,

$$E_m = I' Z_m \left(\frac{Z_1}{Z_1 + Z_m} \right)$$

Amplifiers

An amplifier scales the magnitude of an analog input signal according to,

$$E_o(t) = h\{E_i(t)\}$$

A linear scaling amplifier will have $h\{E_i(t)\} = GE_i(t)$. Amplifiers have a finite frequency response and limited input voltage range.