Statistical Modeling of Real Inductor Impedances

Duan, Jianhang

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Undergraduate of Statistics

of

University College London.

Department of Statistical Science

University College London

August 7, 2023

Abstract

Inductors are electronic components whose impedance is a function of frequency is challenging to model. This study compares a sequence of six models of increasing complexity. Starting from the RLC model which adds a resistor (R) and a capacitor (C) to an ideal indcutor to approximate an imperfect inductor, it uses Bayesian optimisation for maximum likelihood parameter estimation. While the model fits qualitative trends well, residuals are clearly structured, and fit is poor near the self-resonance frequency. In addition, model-based statistical hypothesis testing was utilised in this research to evaluate the influence of the coupling effect between series-connected inductors.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr Yvo Pokern, whose patient guidance and vast expertise have been instrumental to me. In addition, I am very appreciative of his ability to keep my attention on course material by employing a variety of engaging teaching techniques. I am also extremely grateful to Dr Yvo Pokern for his assistance in acquiring physics knowledge, as well as his patient explanations and constant support.

Contents

Chapter 1

Introduction

Inductors serve multiple purposes and are one of the most prevalent and widely used electronic components in daily life. The most prevalent application of inductor is as a signal filter. Apart from that, When multiple inductors are combined in the electric systems, transformer that altering the power of transmission can be formed due to magnetic coupling effect. Almost all electronic circuits contain inductors, whose primary function is to store energy in the form of magnetic field. Therefore, modeling an inductor is important and necessary when employing it. The behaviour of the inductor and whole electric circuit under various inductance becomes predictable, thereby preventing situations where they overheat or fail to function.

The objective of this project is to find an appropriate model that can simulate the inductor's impedance properly. In this study, I will begin with a rudimentary conceivable model that taken into account different types of inductor and frequency ν, and then applying statistical optimization (Quassi-Newton and Bayesian Optimisation method) to find the maximum log-likelihood estimate. Subsequently, I will continuously update the current model based on residual images from previous models. After attempting several non-linear models, the optimal outcome is determined based on the information criterion and significance results.

Traditionally, it was common to create an actual circuit in the physical world and simulate the behaviour of the intended electronic component, such as the inductance or the impedance. Simulating the behaviour of electronic components can help us learn how the electronic component respond at different frequencies.

However, inductor simulation is notoriously difficult in physics due to inductor resonance and magnetic saturation. Kolster [27] began modelling inductors in the early 1910s, the predictive power of the model that contains a pair of parallel inductor and capacitor was critical. In order to model the impedance of inductors with varying degrees of inductance, numerous models must be tested. Using conceivable non-linear models to imitate the actual electronic components is of great research interest in order to save a substantial amount of circuit construction time.

Inductors store energy in a magnetic field, but the field generated by one inductor may also pass through all neighbouring inductors, resulting in magnetic coupling that can have a negative impact on the operation of a circuit.

At the end of the result section of this report, I will use the model-based log-likelihood ratio test to determine whether series-connected inductor coupling effect exists in our measured data. Which simply tests whether series-connected inductors, such as 1 μH and 4 μH, produce a theoretical inductance of 5 μH in an electric circuit. If the existence of the coupling effect can be verified, what is the extent of the coupling effect, and if multiple inductors are modeled, can their electromagnetic field interactions be ignored?

The report has the following format: Chapter 2 provides an introduction to the fundamental concepts of electronics and physics. The focus of Chapter 3 was on the statistical optimization techniques and their implementation, which will be used to enhance the non-linear models by finding the maximum log-likelihood estimate. The fourth chapter is an explanation of the data sets and the rationale for our research topics. Chapter 5 covered the reason and process of getting the current 6 candidate models, and numerous statistical results, including residual plots and kolmogorov-smirnov test are employed to help me evaluate the model's quality. Chapter 6 focuses mainly on an investigation into whether the impedance of two series-connected inductors would behave as anticipated. Chapter 7 summarised the previous statistical observations, and drew the final conclusion, and provided suggestions for future work.

Chapter 2

Introduction to Electronics for Statisticians

This part intends to offer a basic introduction to the physics theory of circuitry, which is required for the subsequent statistical analysis. It is intended to be a summary with the objective of providing an introduction that statisticians can understand. The construction of the circuit and its constituent electronic components are described in this chapter.

2.1 Alternating Current & Direct Current

Direct current (DC) is a current of constant magnitude[46]. The direction of the direct current in the circuit will not change over time. Typical direct current sources include batteries, solar cells, etc.

Alternating current(AC) is the exact opposite of direct current, which implies that the direction of the current changes regularly throughout time. Figure 2.1 represents the alternating current as a sinusoidal curve that fluctuates regularly along the time axis[41].

2.2 Electric Components

We often categorise electronic components into two primary types. The first category consists of active electronic components, which can be thought of as power transmitters. In contrast to active components, passive components do not convey

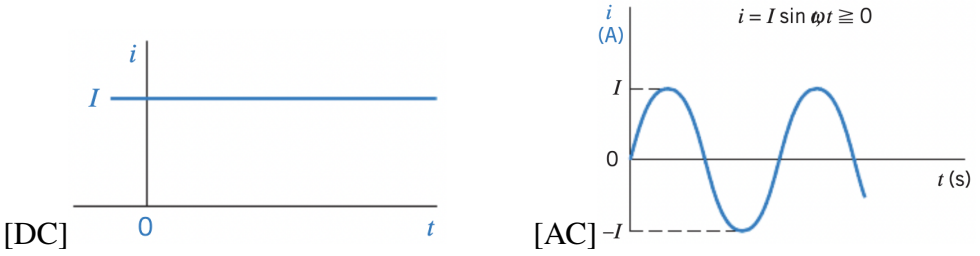


Figure 2.1: Electric current against Time [41]

power to our circuits [42].

- Active components need a source of energy, such as an alternating current (AC) or direct current (DC) power supply, in order to function and deliver current or other forms of energy to our circuits. A generator is a common active component that supplies electricity to our circuits.

- The passive component in our circuit does not provide any power. For instance, there are resistors, inductors, and capacitors.

2.3 Ohm's Law

Only passive components will be considered for this project. Typically, these components fall into one of three categories: resistors, capacitors, and inductors, each of which will be discussed in succession. Ohm's law, which is one of the basic rules of electricity, was found by Georg Simon Ohm (1789-1854)[25], a German physicist, and plays a significant role in the practical electrical applications and it will help us comprehend resistors.

Ohm's law states that if we have a conductor and precisely measure the voltage (V) and current (A) at both ends, we will discover that they are proportional to one another[18]. The ampere (A) is the standard unit of current, and its symbol is I. Resistance (R) and Voltage (V) are measured in Ohms and Volts, represented respectively by the Greek letter Omega and the letter V.

$$V\,(V) = I\,(A)\,R\,(\Omega) \tag{2.1}$$

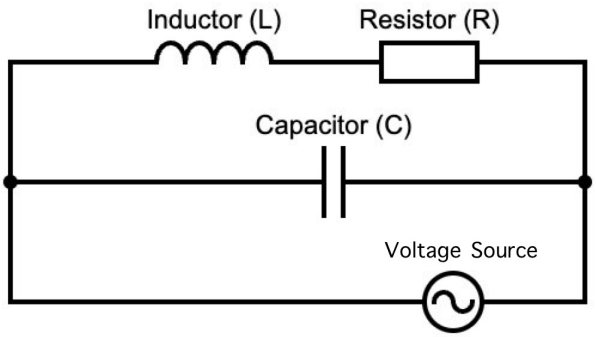


Figure 2.2: Simple Electric Circuit (RLC)

More generally,

$$R = V/I. \tag{2.2}$$

Figure 2.2 depicts a circuit that can be viewed as a route through which an electrical current flows, with connected electronic components. This specific combination of electronic components is referred to as an RLC circuit, which consists of a resistor, an inductor, and a capacitor.

As shown in Figure 2.2, the resistance symbol in the circuit is a small rectangle, whose main function is to provide resistance to the circuit. We can conceive of resistance in a circuit as an impediment to the flow of electric current. The greater the resistance, the greater the degree to which the current is prevented from passing.

2.4 Insulator & Conductor

By categorising a material's resistance, we typically arrive at two broad categories: conductor and insulator[23].

- A conductor is a substance with low resistance that readily transmits electricity. A significant number of charge carriers, which are charged particles that move freely, exist in the conductor. Typically, metals are excellent conductors, with silver being the finest.

- A insulator is a substance with extremely high resistance. The tremendous resistance makes it difficult for electricity to get through. Common insulators include rubber, wood, air, ceramic, etc.

The resistance of the resistor in the electric circuit can be easily calculated by applying Ohm's law. We can use the electrical instruments, such as voltmeter and ammeter to measure the potential difference and current of the terminals of the resistor, and then we can get the resistance by substituting the measured data into the formula of Ohm's law (Equation 2.2) .

2.5 Electric field

Everything is composed of atoms, which consist of electrons and positively charged nuclei. Letter q represent the electric charge of both the electron and the proton; $q = 1.60217663410 \times 10^{-19} C$, where C is the unit of charge known as the Coulomb. In contrast to the electron, which has a relative charge of -1, the proton has a relative charge of 1.

According to a basic aspect of physics, unlike charges attract and like charges repel. Positively charged particles are surrounded by their own electric field. If two positively charged particles are brought together, their electric fields will interact and cause displacement. Therefore, the electric field can be regarded as the interaction force of charged particles.

In physics, force is a vector, thus we may comprehend that electric fields are physical fields with direction and magnitude. This force, represented by $\mathscr{F}$, is defined by [44]

$$\mathscr{F} = qF, \tag{2.3}$$

where F is the electric field strength and defined by

$$F = \frac{Q}{4\pi\varepsilon_0 r^2}. \tag{2.4}$$

Here, ε_0 represent the permittivity of the environment (often air with a permittivity of 1.00051) and r is the distance between positively charged particles and the location of another negatively charged particle[24].

Figure 2.3 demonstrates the electric field around two charged particles, and it is clear that the positively charged particle's electric field is directed toward the

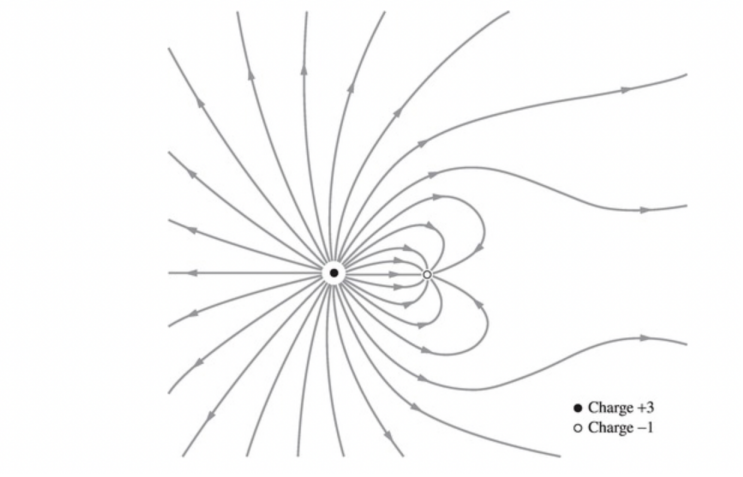


Figure 2.3: Electric field[43]

negatively charged particle; as the two charged particles carry opposite charges, they are attracted to one another.

2.6 Capacitor

The components of a capacitor are a non-conducting material placed between two metal plates. The medium between the two metal plates is called dielectric material. Since dielectric material has an extremely low electrical conductivity, it is also an insulator.

The electrical sign for capacitor in RLC circuit (Figure 2.2) consists of two parallel vertical lines, which also symbolise the capacitor's constituent metal plates. When a capacitor is connected to a power source, current flows into it, charging its plates. One of the metal plates is positively charged while the another one is negatively charged. Therefore, when the plates are charged, the moving electrons between the two plates formed an electric field, which store the electric charge. Moreover, the electric field between the metal plates does not dissipate when the power supply is disconnected, therefore, capacitor can continue to support the energy to the electric circuit by using the charge that stored between the two plates.

The relationship between capacitance (C), current (I) and voltage (V) can be represented as [19]

$$I = C\frac{dV}{dt}. \tag{2.5}$$

2.6.1 Capacitive reactance

Reactance (X) is the obstructive effect of an inductor and capacitor on the alternating current in a circuit [47]. Examples of reactive components are inductor and capacitor. Later we shall explore inductive reactance, but for now we will simply consider capacitive reactance (X_C). The property of reactance is identical to that of resistance, whereas resistance is a real number, reactance is a complex number. We can derive the formula of capacitive reactance using Ohm's law (Equation 2.2), which we mensioned earlier.

All voltages and currents in a linear circuit have the same angular frequency w as the power source. Complex voltage can be expressed in a way that includes both the phase and amplitude of a voltage in a convenient expression. Such voltage (V) can be expressed in Euler's formula

$$V = e^{jwt} = cos(wt) + jsin(wt), \tag{2.6}$$

where j represents the imaginary unit in complex numbers and angular frequency w is defined as

$$w = 2\pi f. \tag{2.7}$$

Letter f appeared in this equation is denoted as the frequency of the alternating current.

When we substitute the voltage in Euler's form into Equation 2.5, we obtain

$$I = C\frac{de^{jwt}}{dt} = C(jwe^{jwt}) = CjwV. \tag{2.8}$$

According to Ohm's law (Equation 2.2), Equation 2.8 can be rearranged to give the capacitive reactance (X_C).

$$\frac{V}{I} = \frac{1}{Cjw} = X_C \tag{2.9}$$

2.6.2 Q-factor of capacitors

Quality factor is abbreviated as Q-factor which indicates the efficiency of capacitors. A capacitor is similar to a battery in that it stores electrical energy. The energy intake

and output of a capacitor with a high level of efficiency should be relatively similar. Nevertheless, if the difference between the two values is too large, which means the majority of the energy stored in the capacitor will be lost as heat.

When considering a capacitor in series with a resistor, the following formula is used to compute the Q-factor of the capacitor. By substituting the definition of capacitive reactance in Equation 2.9 we get

$$Q = \frac{|X_c|}{R_S} = \frac{1}{2\pi f C R_S}. \tag{2.10}$$

X_L denotes capacitive reactance, whereas R_S represents series resistance.

2.7 Magnetic Field

Any magnetic object has two poles, denoted by N for the North Pole and S for the South Pole. These two poles are located at either end of the magnetic item and are the most magnetic locations on the magnet or magnetic object as a whole[9].

By spinning or orbiting an atomic nucleus, charged particles that are in motion generate a magnetic field. Electrons in non-magnetic objects are also in motion, but their motion is primarily random, so the magnetic fields between the electrons cancel out and the item is not magnetic. The reason that our bar magnet has a magnetic field is because its electrons are moving in unison.

Figure 2.4 depicts a bar magnet, where the curve represents the magnet's magnetic field and the arrow represents the magnetic field's direction. We can observed that each magnetic field lines emanated from the North pole of the bar magnet towards the South pole. Moreover, the closer the magnetic field was to the magnetic pole of the magnet, the stronger it was.

2.8 Inductor

The inductor is the final and most crucial electronic component that needs to be modelled for this research.

Figure 2.5 depicts a circuit with only an inductor, and from this figure we can see that the inductor is made up of wire that has been wrapped around the core

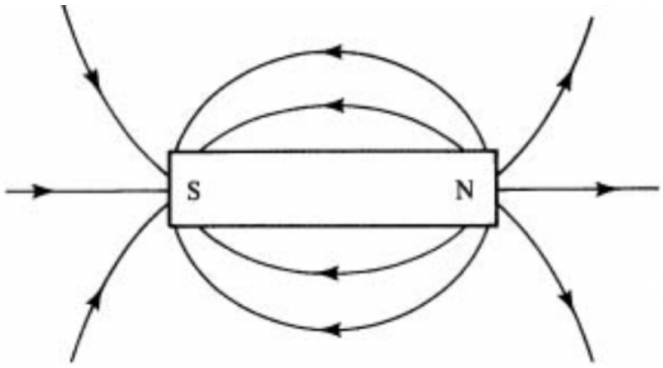


Figure 2.4: Bar Magnet[9]

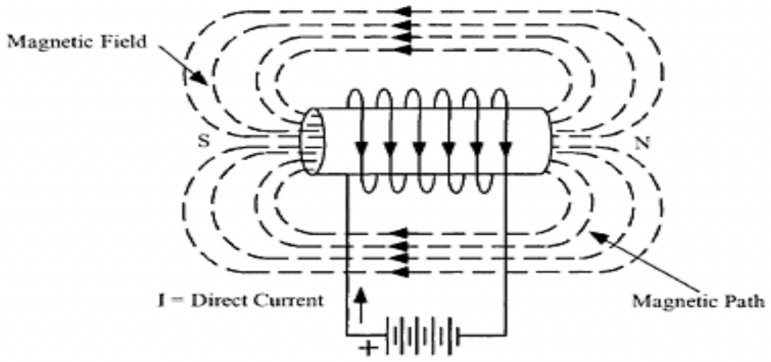


Figure 2.5: Iron Core Inductor[35]

in a cylindrical shape. When current I flows through our coil, a magnetic field is produced (Dotted line surrounding the coil in Figure 2.5), and the inductor stores energy within the magnetic field. This phenomenon is referred to as inductance, and the unit of inductance is Henry (H). Inductors are also known as coils. The electric symbol of an inductor is given in Figure 2.2, it is just like a string of bent wire.

The capacitor utilised in this study differs slightly from the one introduced previously. The capacitor utilised in this instance is known as an Inter-winding Capacitor. The metal plate is replaced with a single coil without the core. Inter-winding capacitance is created between neighboring turns of the wire (Shown in Figure 2.6).

The core of an inductor is composed of a variety of materials, including iron, air, copper, etc. However, the inductance of various materials varies. The vast majority of substances are poor conductors of magnetic flux; their permeability

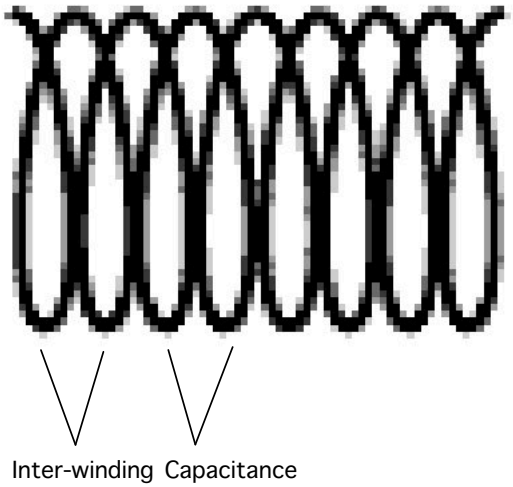
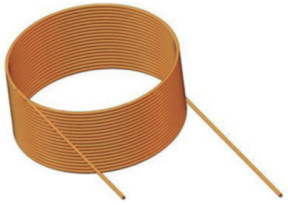
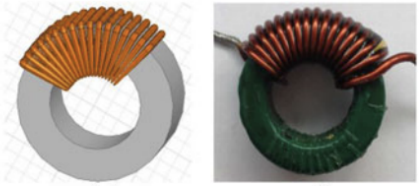


Figure 2.6: Inter-winding Capacitance [26]



2.71[30]



2.72[39]

Figure 2.7: Comparison of Cylindrical inductor and Toroidal inductor

is very low. Air, paper, and copper, which are all nonmagnetic, have the same relative permeability as vacuum. Each has an extremely low permeability of 1.0. Nevertheless, there are ferromagnetic materials, such as iron and cobalt, with a high permeability of up to tens of millions. In this research, iron powder is the primary material utilised for the toroidal core which is built of coils encircled by circular toroid (Figure 2.72).

Air-coil (Cylindrical inductor without the core) inductor shown in Figure 2.71 are not wrapped in any ferromagnetic material; they are supported mechanically by the wires themselves. Unlike cylindrical inductor, the coils on the toroid are firmly connected and form a circle, there is no gap between the coils and the magnetic field is concentrated on the toroid. In general, toroidal inductors are more efficient than cylindrical inductors because their Q-factor is greater.

Magnetic flux is a vector that represents the magnetic field's direction and

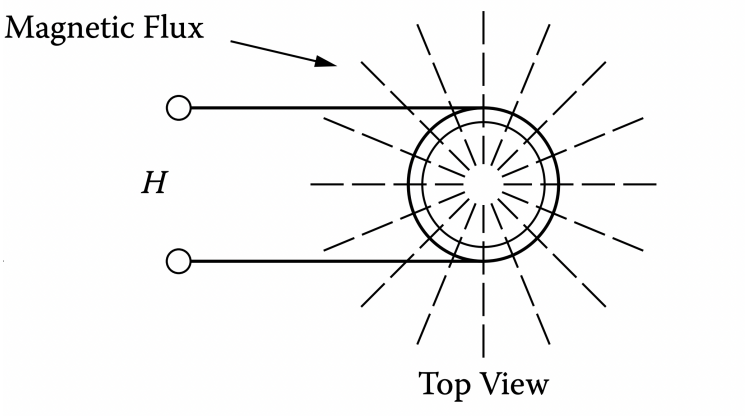


Figure 2.8: Magnetic flux from top vies.[34]

strength. When a magnetic object is placed in a magnetic field, the force on the object can be determined by studying the magnetic flux. Figure 2.8 illustrates the magnetic flux lines from top view of the inductor in Figure 2.5.

Magnetic flux (Φ) of the electromagnet is defined by

$$\Phi = \frac{NI}{\frac{l_{mag}}{\mu_0+\mu_r S} + \frac{d}{\mu_0 S}}. \tag{2.11}$$

S,N,l_{mag},d,μ_0,μ_r represent, respectively, the cross-sectional area of the coil, the number of turns of the coil, the length of the magnetic core, the length of the air gap, the relative permeability of air (Typically assumed to be 1), and the relative permeability of the magnetic core [52].

There are numerous ways to express inductance, but the expression related to magnetic flux is

$$L = \frac{N\Phi}{I} = \frac{N(BS)}{I}. \tag{2.12}$$

As shown by Equation 2.11, when ferromagnetic materials such as iron is utilized as magnetic cores, the magnetic flux will increase, thereby increasing the inductance of the inductor. This is because the magnetic flux is positively proportional to the inductance in the expression of Equation 2.12. In contrast, if we use materials with low permeability, both the magnetic flux and the inductance will decrease.

In addition to the core material, we can see that there are a number of other variables that can possibly influence magnetic flux shown in Equation 2.11, such as

the cross-sectional area of the coils and the air gap. Both variables have a positive correlation with the magnetic flux. Air gap usually represent the distance between the neighbouring coils on the cylindrical core. If coils are too far apart, the strength of the induced magnetic flux will be undermined.

Electromotive force (EMF) and Voltage both represent the generated potential difference, so we assume they are equivalent. The relationship between output voltage and the time derivative of the total magnetic flux can be explained by Faraday's law which is one of the fundamental law of inductors. We can calculate the output voltage from inductor using Faraday's law which is defined as [20]

$$\varepsilon = -\frac{d\Phi_b}{dt}, \tag{2.13}$$

where ε represent the output voltage or electromotive force, Φ_b is the magnetic flux, and t represent the time. According to Faraday's law of induction, When the current through the coil changes, the resulting magnetic field changes over time, inducing an electromagnetic force, which then creates the voltage in the conductor. So we can also understand electromotive force as the rate of change of the magnetic flux.

The magnetic flux, which was previously defined in Equation 2.11, also has a shorter form, which is Φ_b=BS. In this expression, B represent the Magnetic Flux Density, and S represent the cross sectional area of the coil [29]. During our research we consider these two parameters to be constant. Thus, we can simplify the formula further.

$$\varepsilon = -NS\frac{dB}{dt} \tag{2.14}$$

Inductive reactance can be found by the formula

$$X_L = jwL. \tag{2.15}$$

Inductance is represented by the letter L in this formula, while w and j have already been defined in subsection 2.61.

2.8.1 Quality-factor of Inductor

Q-factor of in inductor is the ratio between inductive reactance (X_L) and series resistance (R_s). When using alternating current with a high frequency, the performance of inductors with a high Q-factor is closer to our ideal circumstance and their performance is more stable.

$$Q = \frac{wL}{R_s} = \frac{2\pi fL}{R_s} \tag{2.16}$$

When we use ferromagnetic material as the inductor's core or increase the cross-sectional area of the coil, the inductance will increase, thereby increasing the Q-factor, and the inductor's quality can be enhanced.

In reality, there is both inductance and resistance of an inductor. Resistance is the primary cause of energy loss in inductors. Due to winding resistance, the passing speed of electrons will be hindered when they attempt to pass through the wire. As electrons collide, kinetic energy is transformed into heat, resulting in a loss of energy. The iron core's heat loss is more than that of the air core. The changing magnetic field created by alternating current is the primary cause of the core loss of Iron power core, which ultimately results in greater heat loss [12].

2.9 Parallel Circuit & Series Circuit

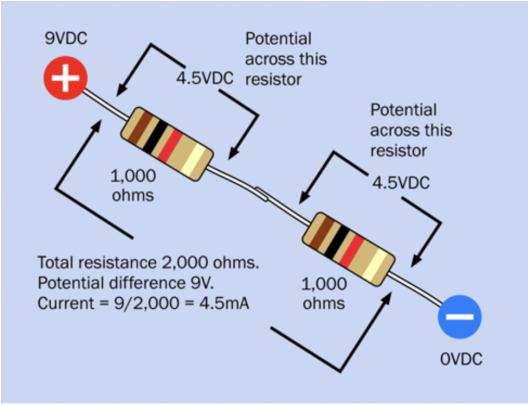


Figure 2.9: Electric Circuit in series[40]

The most fundamental two circuit types, series circuit and parallel circuit, are used in applications ranging from simple circuits to complicated circuits.

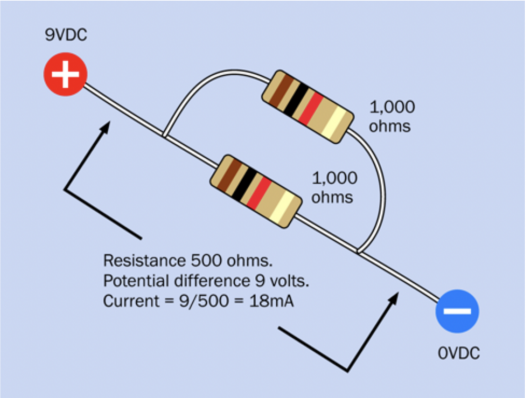


Figure 2.10: Electric Circuit in series[40]

- Electric components in series circuit are connected one follow the other. It is shown in Figure2.9
- Electric components in parallel circuit are connected side by side. It is shown in Figure2.10

The voltage will be divided by the series circuit, and the current will be constant through out. In contrast, in a parallel circuit, the voltage is uniform throughout and the total circuit current is the sum of the branch currents.

Considering series circuit. Assume number N resistors connected in series, then we can find the total resistance and voltage of the circuit.

$$R_{total} = (R_1 + R_2 + R_3 ..... + R_N) \tag{2.17}$$

The voltage of an electric circuit is the sum of the voltages measured at each component's two terminals$(V_1, V_2, ..., V_N)$.

$$V = (V_1 + V_1 + V_3 ..... + V_N) \tag{2.18}$$

Current in the circuit can be calculated using Ohm's Law (Equation 2.2).

$$I = \frac{V}{(R_1 + R_2 + R_3 ..... + R_n)} \tag{2.19}$$

Figure 2.10 shows a parallel circuit with two resistors connected side by side.

Current in the electric circuit shown by Figure 2.10 has two routes to run. The combined current is equal to the sum of the currents in each branch.

$$I_{total} = (I_1 + I_2) \tag{2.20}$$

Each branch's voltage is the same as the supply voltage.

$$V_{supply} = V_{branch1} = V_{branch2} \tag{2.21}$$

The total resistance for the parallel circuit is

$$\frac{1}{R_{total}} = \left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{2.22}$$

Using the same strategy, the reactance of a circuit can be computed by substituting the resistances with the impedance of inductors and capacitors in formulas 2.22 and 2.17.

2.10 Reactance & Impedance

When inductor and capacitor are connected in a series circuit, the total electrical reactance of the entire circuit X can be calculated and expressed as

$$X = X_L + X_c = jwL - \frac{j}{wC}. \tag{2.23}$$

Impedance is the opposition to the alternating current in a circuit that results from the interaction of resistance and reactance [1].

The Impedance itself is a complex number that consists of both real and imaginary parts. Impedance is defined as

$$Z = R + jX. \tag{2.24}$$

The real part of the impedance is the resistance, the imaginary part represent the total reactance(X). In general, a component whose impedance contains a non-zero

imaginary part is referred to be ”reactive”.

2.11 Resonance

In RLC circuits Figure 2.2, for instance, electrical resonance happened when the imaginary part of the impedance between the inductor and capacitor cancel out, and the reactance (X) of the circuit is equal to zero. The impedance(Z) of the RLC circuit (Figure 2.2) can be expressed using the formula below.

$$Z = R + (2\pi fL - \frac{1}{2\pi fC})j \tag{2.25}$$

Resonant frequency is the frequency when

$$| 2\pi fL | = | -\frac{1}{2\pi fC} | . \tag{2.26}$$

Resonant frequency can be calculated by solving the Equation 2.26.

$$f_{resonant} = \frac{1}{2\pi\sqrt{LC}} \tag{2.27}$$

Impedance of the resonant circuit now become

$$Z = R + (0)j. \tag{2.28}$$

Chapter 3

Statistical Methodologies

In this section, we will discuss statistical methods required for later analysis, including optimization methods to obtain the optimal value we want and evaluation methods to select the most appropriate model balancing goodness of fit and number of parameters.

3.1 Non-linear Models

We utilise the nonlinear model to predict the current observation. General relationship between Y and observation X in non-linear model can be expressed as a non-linear function of X, $f : R \to R$, and error term ε [22]. The dependent variable Y is the output of our model, which is the measured impedance, whereas X is the measured frequency of the alternating current. Non-linear model is used to simulate impedance at various frequencies. To ensure that the output value of model is not significantly different from the measured impedance data at a given frequency, we want to minimize the model's sum of squares.

$$Y = f(X;\theta) + \varepsilon_i \tag{3.1}$$

white noise ε_i is a collection of independent, identically distributed (iid) random variables $\varepsilon_1, \varepsilon_2, \varepsilon_3, ..., \varepsilon_N$, in which N represent the number of observations. And $\varepsilon_i \sim N(0, \sigma^2)$, where $\sigma^2 \in R^+$. $\theta \in \Theta$ is a k dimensional parameter (i.e. A k $\times$ 1 vector) with parameter space Θ, such that $\theta = (\theta_1, ..., \theta_k)^T$. Here, Letter T represents the transpose of the vector.

Models can contain several electronic components;therefore, the dimension of the model's parameters is dependent on the number of electronic components in the modeling circuit. Parameter has significant effect on the predictive performance of the model; therefore, optimization is performed to identify the model parameter that best fits.

3.2 Optimisation

Before we look at statistical optimization methodologies we need to know what is optimization and why we should use optimization.

Optimization is the task of finding $\hat{\theta}$ such that the objective function f : in terms of k-dimensional parameter θ is as large or as small as it can be. The process of locating the maximum of an objective function $f(\theta)$ does not differ from that of locating its minimum, since

$$\theta^* = arg\max_{\theta\in\Theta} f(\theta) = arg\min_{\theta\in\Theta}\big(-f(\theta)\big). \tag{3.2}$$

Numerous facets of academic fields can be improved through optimization. In the financial industry, for example, if we want to develop different products, X and Y, and I know my cost function, I may use optimization to determine the best quantity of X and Y that minimises my cost. Clearly, this is the most elementary example. In the real world, we must additionally consider the available resources, the profitability of the products, etc.

3.3 Newton method

Newton method is a prevalent optimization technique utilised in numerous fields. Utilizing Taylor expansion and the iterative method to approximate the ideal value of the entire function is its central concept. Newton's approach has several advantages, including the ability to utilise an iterative solution when it is difficult to obtain the root, its speed, and its ability to precisely locate the optimal value.

When we consider one dimension, and we are trying to optimise the twice differentiable function $f : R \to R$. The first step is to carry out the quadratic Taylor

polynomial of the function $f : X$ and supposing they are approximating [10].

$$f(\theta) \approx f(\theta_n) + (\theta - \theta_n) f^{'}(\theta_n) + (\theta - \theta_n)^2 \frac{1}{2} f^{''}(\theta_n)^2 \tag{3.3}$$

Consider that θ_n in this equation represents a sequence of numbers,$\{\ \theta_0, \theta_1, ..., \theta_n\}$. If we take the derivative of the Equation 3.3 and suppose that it is equal to zero, we will have the following formula, which is our iteration formula.

$$\theta_{n+1} = \theta_n - \frac{f(\theta_n)}{f^{'}(\theta_n)} \tag{3.4}$$

$\theta_0 \in R$ is the initial guess of the function root. Then, we use the Equation 3.4 as the iteration formula to fill in θ_0, and we repeat the iteration process to get $\theta_1, \theta_2, \theta_3, ...$ that is closer to optimal value according to degree of convergence.

When applying Newton's method in higher dimensions, our iterative formula can be represented as

$$\Theta_{n+1} = \Theta_n - [Hf(\Theta_n)]^{-1} \nabla f(\Theta_n), \tag{3.5}$$

in which $\Theta_n = (\theta_1^n, \theta_2^n, ..., \theta_k^n)^T$, k is the number of dimensions. $Hf(\Theta_n)$ represent the Hessian matrix of the function $f(\Theta_n)$. The formula for iteration can alternatively be stated compactly in matrix form.

$$\begin{bmatrix} \theta_1^{n+1} \\ \theta_2^{n+1} \\ \vdots \\ \theta_k^{n+1} \end{bmatrix} = \begin{bmatrix} \theta_1^{n} \\ \theta_2^{n} \\ \vdots \\ \theta_k^{n} \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 f(\Theta_n)}{\partial \theta_1^2} & \frac{\partial^2 f(\Theta_n)}{\partial \theta_1 \theta_2} & \cdots & \frac{\partial^2 f(\Theta_n)}{\partial \theta_1 \theta_k} \\ \frac{\partial^2 f(\Theta_n)}{\partial \theta_2 \theta_1} & \frac{\partial^2 f(\Theta_n)}{\partial \theta_2^2} & \cdots & \frac{\partial^2 f(\Theta_n)}{\partial \theta_2 \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\Theta_n)}{\partial \theta_k \theta_1} & \frac{\partial^2 f(\Theta_n)}{\partial \theta_k \theta_2} & \cdots & \frac{\partial^2 f(\Theta_n)}{\partial \theta_k^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f(\Theta_n)}{\partial \theta_1} \\ \frac{\partial f(\Theta_n)}{\partial \theta_2} \\ \vdots \\ \frac{\partial f(\Theta_n)}{\partial \theta_n} \end{bmatrix} \tag{3.6}$$

In this instance, the Hessian matrix is a k-by-k square matrix. When k is large, it will be costly to calculate the hessian matrix. Due to the complexity caused by the Hessian matrix, the application of the Newton method in high dimensions is no longer appropriate. Quasi-Newton method, on the other hand, is an alternative

optimization approach based on Newton's method that solves the problem by not requiring the actual computation of the Hessian matrix. Rather, the gradient vector is utilised to continuously update an approximation of the Hessian matrix.

3.4 Broyden–Fletcher–Goldfarb–Shanno Method

The Broyden–Fletcher–Goldfarb–Shanno (BFGS) method is a class of Quasi-Newton algorithm; it is named after its discoverers Broyden, Fletcher, Goldfarb, and Shanno [38]. It is an iterative and gradient-based method for solving unconstrained non-linear optimization problems in numerical optimization [14].

Suppose we are attempting to optimise differentiable function $f(\Theta)$in dimension k, where Θ at n iterations can be expressed as $\Theta_n = (\theta_1^n, \theta_2^n, ..., \theta_k^n)^T$. In later formulations, the derivative of the function at Θ_n will be represented by $\nabla f(\Theta_n)$ and $f'(\Theta_n)$.

In the BFGS approach, the aim of our iterative formula is to continuously update B_{Θ_n} which is a $k \times k$ symmetric matrix that approximates the Hessian matrix. Before taking the iteration, we need to confirm the search direction p_Θ, that is, the moving direction of the algorithm, which can be written as

$$p_{\Theta_n} = -B_{\Theta_n}^{-1} \nabla f(\Theta_n). \tag{3.7}$$

When we define the direction of our search, we can find a point along that direction, so our iteration formula can be defined as

$$\theta_{n+1} = \theta_n + \alpha_n(-H_n f'(n)). \tag{3.8}$$

In the iteration formula, α_n is referred to as step length which is very important that determine the rate of convergence. If the step length is too large, then the optimisation process may converged quickly with a incorrect result. In contrast, if the step length is too small, then the optimisation process may require more iterations, which can be time-consuming. Therefore, choosing a appropriate step length is a key factor in the gradient-based optimisation techniques. In the BFGS method, we

are aiming to find α_n that

$$\min_{\alpha}(f(\theta_n + \alpha p_{\Theta_n})). \tag{3.9}$$

Wolfe conditions [31] is a common line search conditions, and it can be utilised as a standard to determine a suitable α value in our text, in order to make later optimisation process efficient.

Similarly, using the same procedure as Newton method, assume the second-order Taylor polynomial obtained by Taylor expansion is approximating to our objective function and we will get

$$f(\theta) \approx f(\theta_{n+1}) + f'(\theta_{n+1})(\theta - \theta_{n+1}) + \frac{1}{2} f''(\theta_{n+1})(\theta - \theta_{n+1})^2. \tag{3.10}$$

When we differentiate this expression, we get

$$f'(\theta) = \nabla f(\theta_{n+1}) + H_{n+1}(\theta - \theta_{n+1}), \tag{3.11}$$

where H_{n+1} represents the Hessian matrix, and since we assume that the previously defined matrix B approximates the genuine Hessian matrix H, therefore $H_{n+1} = B_{n+1}$. And we can rewrite the Equation 3.11 and obtain

$$B_{n+1} s_n = y_n. \tag{3.12}$$

with

$$s_n = (\theta_{n+1} - \theta_n) \tag{3.13}$$

$$y_n = (\nabla f(n+1) - \nabla f(n)) \tag{3.14}$$

According to the addition rule for matrices, we suppose that B_{n+1} matrix is equal to B_n plus an unknown k $\times$ k matrix at n iterations which is defined as $\varpi_n = \alpha z_n z_n^T + \beta q_n q_n^T$. And z_n, q_n are k $\times$ 1 matrix.

$$B_{n+1} = B_n + \varpi_n \tag{3.15}$$

When $z_n = y_n$ and $q_n = s_n$ we get the formula [13]

$$(B_n + \alpha y_n y_n^T + \beta s_n s_n^T) s^T = y_n \tag{3.16}$$

After solving the equation we get the expression for α and β.

$$\begin{aligned} \alpha &= \frac{1}{y_n^T s_n} \\ \beta &= -\frac{1}{s_n^T y_n^T s_n} \end{aligned} \tag{3.17}$$

Finally, we can get the iteration formula

$$B_{n+1} = B_n - \frac{B_n s_n s_n^T B_n}{s_n^T B_n s_n} + \frac{y_n y_n^T}{y_n^T s_n^T}. \tag{3.18}$$

We use the result obtained here for k+1 as the input in Equation 3.8 and thus continue iterations until X_n is converging to a specific number.

Finite difference approximation is utilised to approximates the derivative of the objective function at point θ and it is defined as[51]

$$f(\theta) \approx \frac{f(\theta - \nabla\theta) - f(\theta)}{\nabla\theta} + \varepsilon. \tag{3.19}$$

The approximation of the objective function's derivative will improve for smaller $\nabla\theta$. However, this is not always the case, and when $\nabla\theta$ becomes extremely small, we must consider the floating point error. When $\nabla\theta$ is equal to $1x10^{-20}$, which is a very small value, we would expect to obtain a highly accurate approximation of the derivative. However, when we calculate using computer programming software, the error of the result is greater since θ surpasses the upper limit of significant numbers that can be recognised by software, so it is calculated as 0 in the process of calculation. Therefore, it generates an abrupt increase in errors.

Equation 3.19 represent the finite difference approximation with an additional error term ε, due to the presence of the error term, we may anticipate that the objective function $f(x)$ will not only be twice differentiable, but also higher-order differ-

entiable. BFGS can be utilised for large dimensions, but the objective has always been to locate the local maximum after setting the step length and search direction; this is the restriction of BFGS method. Consequently, the BFGS approach can only handle the converging objective function. Since the purpose of optimisation is to discover the global maximum of the objective function, it is inappropriate and costly to continue using the BFGS approach when the objective function has many local maximum. The subsequent optimization technique is known as Bayesian Optimisation, which is not dependent on Newton's method and seeks global optimal values.

3.5 Bayesian Optimization

Bayesian optimisation employs the analogous concept of using the iteration process to discover the optimal solution as the previously discussed Newton method. However, Bayesian optimisation is more suitable in dealing with the black-box function[50], which is generally unknown and hard to evaluate.

Additionally, Bayesian optimisation incorporates the prior knowledge and introduce a new optimisation problem, that is, optimising the acquisition function, which is less expensive in addressing relatively lengthy objective function. In higher dimensions, Bayesian optimisations tolerates stochastic noise in function evaluations, and is most effective for optimization that fewer than 20 dimensions [16] .

Assume we are now attempting to optimise K dimensional $f : \Theta \to R$, $\Theta \in (\theta_1, \theta_2, .., \theta_k)$, via Bayesian optimisation. Red line shown in Figure 3.1 reflects our objective function. Initially, we do not know what our objective function would look like (e.g., it's shape and trends), thus we apply the surrogate function to represent the behaviour of the target function. The surrogate function is basically the Gaussian process, and it achieves the simulation by collecting n limited points at random and then fitting through them [36]. We can combine all the objective function's output $f(\Theta)$ into a vector form, for each limited random point $\Theta = (\theta_1, \theta_2, ..., \theta_n)$

Because of its versatility, the Gaussian process prior is selected as the surrogate function and function's prior. Therefore, the distribution of one objective function's

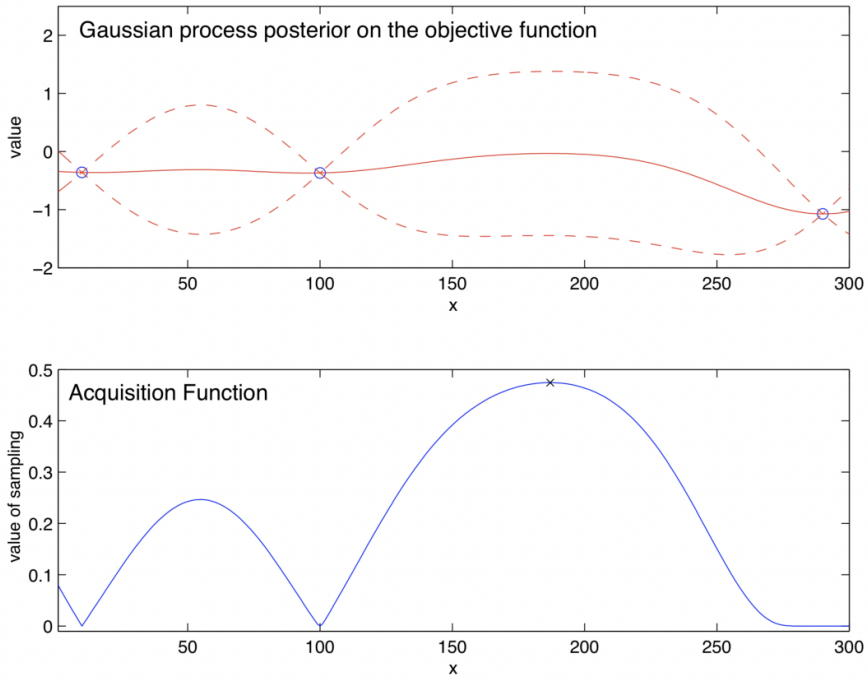


Figure 3.1: Bayesian Optimisation[16]

value $f(\theta_n)$ is expressed as

$$f(\theta_n) \sim N(\mu_0(\theta_n), \sigma^2(\theta_n)). \tag{3.20}$$

Then, we must identify the acquisition function, whose purpose is to integrate newly observed data with prior information by using Bayesian inference in order to determine the next estimation point [45]. Furthermore, another of the primary goals of the acquisition function is to strike a balance between exploration and exploitation; we will discuss this later on.

Gaussian process regression is used to produce Bayseian credible intervals which measures the uncertainty around the estimated parameter values. It is shown by the dotted red line in Figure 3.1. Because we simulate our goal function using a Gaussian process regression, therefore, we can express Bayseian credible intervals in terms of the mean and standard deviation of the Gaussian process regression. The Bayesian Credible Interval (BCI) shows the magnitude of our forecast as well as the

confidence interval. BCI can be expressed as

$$BCI = \mu(\theta) \pm k\sigma(\theta), \tag{3.21}$$

in which k represent the $Z_{0.975}$= 1.96.

Bayesian optimization is searching global optimum, hence the acquisition function cannot solely evaluate the present observed lowest or maximum value, which is known as Exploitation. But also investigate unexplored regions where global optimum may exist, that is, the exploration.

Acquisition function is denoted by

$$u : \Theta \to R^{+} \tag{3.22}$$

$$u(\Theta) = u(\Theta|\{\theta_n, y_n\}) \tag{3.23}$$

$$y_n \sim N(f(\theta_n), \sigma^2) \tag{3.24}$$

σ^2 represent the variance of the noise. θ_n, y_n represent the previous observation.

If the findings of the previous iterations were too distinct from the value of objective function, then this step should be skipped and not investigated. Alternatively, if a candidate solution of θ is noticed based on historical data and it may be the global optimum, we would carry out "Exploitation" which deploy additional points and searching for the global optimum near the best point identified so far.

By analyzing the previous data, we can make new judgments about optimal value. The posterior distribution after given the observation now follow the normal distribution with mean $\mu_n(\theta)$ and variance $\sigma_n^2(\theta)$ [45].

$$f(\theta)|f(\theta_{1:n}) \sim N(\mu_n(x), \sigma_n^2(\theta)) \tag{3.25}$$

According to the Gaussian process regression, the distribution of our function value is now known, allowing us to forecast the predicted function output in any unknown location.

The most common and most practical acquisition function is EI (Expected improvement), that is, the improvement of the objective function from the best point identified so far to $f(\theta)$ and it can be evaluated by the following equation.

$$EI_{\theta}(\theta) = \sigma_n(\theta)\varphi\left(\frac{\mu_n(\theta) - f(\theta*)}{\sigma_n(\theta)}\right) - (\mu_n(\theta) - f(\theta*) - \beta)\Phi\left(\frac{\mu_n(\theta) - f(\theta*)}{\sigma_n(\theta)}\right) \tag{3.26}$$

$f(\theta^*)$is formerly observed greatest value and θ^* is the corresponds x-point. It can be expressed as θ^*=$arg\max f(\theta)$, where n is the number of times the function has been evaluated. $\Phi(x)$ is the standard normal cumulative distribution function, and $\varphi(x)$ is the probability density distribution of the standard Gaussian process. β represent learning-rate of the function which is used to balance the trade-off between exploration and exploitation [49].

The expected improvement algorithm then find the point with the greatest expected improvement[16].

$$X_{NextEvaluation} = arg\max EI_{\theta}(\theta) \tag{3.27}$$

The initial optimisation problem, which consisted of maximising the objective function f, is replaced by a second optimisation problem (maximising the function of the expected improvement). Comparing the expected improvement function to the original objective function, it is comparatively less expensive, and relatively simple since it can be evaluated regarding the derivatives by using second-optimisation methods[16]. In addition, EI also considers the uncertainty of the estimated parameter values, which is represented by the posterior variance, and the best predicted value, that is, the posterior mean[32]. Therefore, maximising regard EI rather than original function would make the process more efficient since it learning from the historical findings.

The Bayesian Optimisation approach is employed before the BFGS. Bayesian Optimization can therefore assist in locating the global maximum for the entire likelihood function. Since BFGS method converges more quickly, so it employed to find the maximum Likelihood estimates. And based on the results of Bayesian

optimisation, a precise log-likelihood estimate can be determined. This sequence makes up for the fact that our BFGS algorithm finds the local maximum as opposed to the global maximum.

In this research, I utilised the R package [53] in performing Bayesian Optimisation of my objective function.

3.6 Evaluation Methods

3.6.1 Log-likelihood function

Suppose we have a random sample with n observations $X = (x_1, x_2, x_3, ..., x_n)$ with probability density(mass) function $f_{X_i}(x_i;\theta)$ with parameter(s) θ. The likelihood function for θ is defined as [33]

$$L(\theta;x) = \prod_{i=1}^{n} f_{X_i}(x_i;\theta). \tag{3.28}$$

The likelihood function is just the joint probability density(mass) function of X and the log-likelihood function for θ is defined as [2]

$$\ell(\theta;x) = logL(\theta;x). \tag{3.29}$$

Maximum likelihood estimate (MLE) was first proposed in 1920 by R.A. Fisher. In the previous definition, we assumed that log-likelihood is actually a probability density (mass) function, therefore when we attempt to make inferences about parameter theta, it is natural for us to select the data's highest probability of occurrence, hence the concept of maximum likelihood estimation [37].

Maximum likelihood estimate $\hat{\theta}$ refers to the parameter that maximise the likelihood function. The calculation process is same as the optimisation problem shown in Equation 3.2 after substituting the log-likelihood function (L) to objective function $f(\theta)$.

3.6.2 Akaike Information Criterion

Considering that we have some data, our goal is to determine the optimal model for predicting these variables in order to reduce information loss. The information criteria used to quantify the performance of the candidate model in this project is Akaike information criterion, abbreviated as AIC, which is typically the initial model selection criterion that should be applied in practice[15].

AIC provides a straightforward and trustworthy estimate of the result of the out-of-sample testing. [33]. Out-of-sample deviation refers to the reminder of the original data that was not utilised for training the model. The result of our out-of-sample testing is more convincing because we have a clear realistic understanding of our model's performance when predicting a new set of data. There are numerous common methods for validating the model's predictive performance, such as observing the residual-related plots, performing a significance test of residual-normality, or analyze the quantity of log-likelihood function.

The definition of AIC is described below[2] :

$$AIC = -2Log(L) + 2p, \tag{3.30}$$

Where L is the Likelihood of the model and p is the number of the parameters in the model.

AIC will seek the lowest negative likelihood of the models and choose the candidate model with the best performance[48], taking into account the number of parameters. The penalty term "2p" shown in Formula 3.30 preventing the occurrence of over-fitting. That is, using more parameters than required to describe the data set.

AIC was used in this project to evaluate the predictive impedance performance of the model, which was then utilised for model selection. When comparing two nonlinear models, the model with the smaller AIC value is always preferred.

The Bayesian information criterion (BIC) is a model selection method that

differs from the Akaike information criterion (AIC) by the second term.

$$BIC = -2Log(L) + pLog(n) \tag{3.31}$$

letter n in the above formula represented the number of data sets.

We may discover that the Bayesian information criterion (BIC) quite similar to the Akaike information criterion (AIC), as both are based on the likelihood function. Hence, the quantity of AIC and BIC has close to each other and have little influence on the model selection. The reason that I ultimately choose AIC is because it have distinct implications when compared with BIC. When using AIC, we believe that the candidate is not the genuine model. But I'd like to utilise the best "wrong" model for predicting the inductor impedance. However, the BIC technique suggests that the true model is contained within the candidate model, and I'll eventually find it after collecting more samples[3]. AIC is preferable due to the complexities of modelling impedance that caused by non-ideal behaviour of the inductor. AIC has a straightforward estimation of the out-of-sample deviation. Nevertheless, we cannot apply the out-of-sample testing in our research. Due to Monte Carlo error, comparisons conducted only a few times are unreliable[17]. Therefore, we might need to compare model predictions with observed values repetitively, which is costly.

Chapter 4

Data &Research Questions

The data consist of measurements of the impedance of 31 different coils across a range of frequencies from 1.8 MHz to 30 MHz. The measured impedance are expressed as reflection coefficients relative to a system impedance of 50 Ohms. The link between reflection coefficients and impedance is given via (4.1).

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{4.1}$$

In this formula, Z_L represents the complex load impedance, whereas Z_0 represents the characteristic impedance, which defaults to being purely real (50 ohm) for subsequent operations in order to obtain the load impedance in the electric circuit.

The 31 inductances were actually obtained as serial circuits of five fixed inductors as displayed in Figure 4.1. Since each inductor is connected in parallel to a switch, the switch can control each inductor, allowing us to create an electric circuit with different inductors. Moreover, the impedance varies for each inductor in the circuit, ranging from low to high from $0.25\mu H$ to $4\mu H$ from the first inductor to the fifth indcutor. Inductors with inductances ranging from $0.25\mu H$ to $2\mu H$ are cylindrical air-cored inductor(Shown in 2.71). However, they are distinct from the fifth inductor with inductance of 4 μH, toroidal inductor(Shown in 2.72) with iron powder core.

Observing the differences between the imaginary part of the impedance and frequency of 31 sets of data (Figure4.2) reveals that the impedance of 31 sets of data

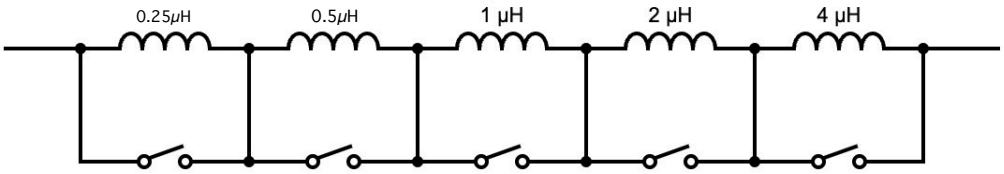


Figure 4.1: This graph illustrates the circuit diagram of five inductors connected in series.

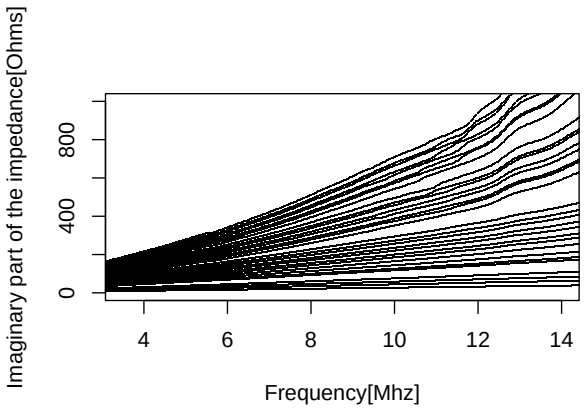


Figure 4.2: Imaginary part of the Impedance against Frequency graph

increases as the AC frequency rises. Because the inductive reactance is proportional to the frequency of alternating current (Shown in Equation2.15). Considering that only inductors are connected in series, the total impedance of the circuit is equal to the sum of the impedance of each inductor. When using a circuit with a relatively higher impedance, the rate of increase of impedance is greater than a circuit with a lower impedance.

4.1 Research Questions

Our objective is to develop the most precise models to simulate impedance (X) at various frequencies (ν). The object we modelled is the inductors shown in Figure 4.1.

This statistical model resembles an unidentified electrical circuit. The purpose of statistical analysis is to compare the model's impedance to the actual measured value. The comparison results would aid me in optimising my initial model and determining the optimal model by comparing various models using statistical method-

ologies.

Chapter 5

Results

Figure 2.2 depicts the simple electric circuit (RLC circuit) that served as the starting point of our modelling.

Chapter 2.10 introduces the formula and content that express the impedance of inductor and capacitor respectively. Combining this with the series and parallel calculation methods described in Chapter 2.9 allows for the expression of the impedance of the RLC model to be derived. The Expression below is my initial nonlinear-model consists of inductance, capacitance, and resistance.

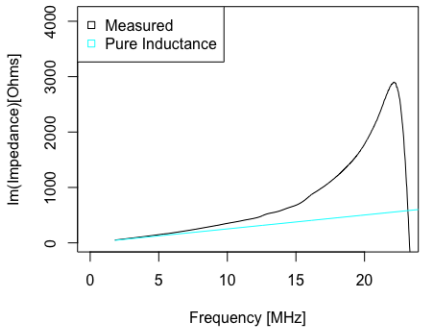
Expression for the impedance of the RLC circuit is described below:

$$Z = (\frac{1}{R + 2\pi fLj} + 2\pi fCj)^{-1}. \tag{5.1}$$

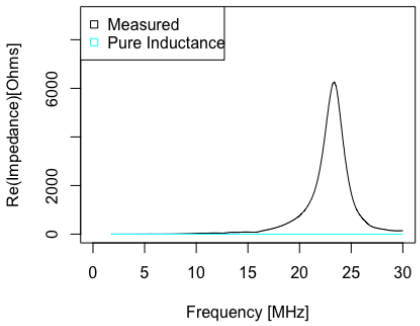
I will start in simulating the corresponding data set for the fifth inductor(4μH) demonstrated in Figure 4.1. In order to evaluate the predictive performance of the candidate model, I will compare the model's prediction with the measured impedance as well as its information criterion.

5.1 Simulation of 4 Microhenry Toroidal Inductor

Figure 5.1 depicts the imaginary and real part of the Impedance versus Frequency graph. The black curve can be seen represents the impedance measured in practise. I conduct exploratory data analysis, as we initially do not know which parameters for our initial model, the RLC model, are appropriate. The blue line represents the simplest conceivable model, i.e. The RLC model that consist of ideal 4μH inductor



5.11



5.12

Figure 5.1: The measured impedance of 4μH datasets is compared to the impedance predicted by conceivable model which has only a ideal 4μH indcutor

only. Other parameters such as capacitance and resistance are therefore null. There will be a sequence of models of increasing complexity and better fit to the data. The model with inductance only is at the extremely simple end of this sequence.

Observing the blue line in the image of imaginary part of the impedance against frequency graph in this instance reveals that our model fits well when the frequency is low (Around 2Mhz). Nonetheless, as frequency increases, the discrepancy between model's prediction and actual measurement results grows. It is obvious that this model is inappropriate since it is not simulating the data set at all at higher frequency.

However, the blue line depicts in the Figure 5.11 is the tangent to the black curve (measured impedance) at about 4 Mhz, indicating that selecting the inductance of 4μH of the inductor to simulate the 4μH data set is an good decision.

One possible explanation for the inaccuracy of model predictions is that, by default, only inductance is available for our circuit, so the predicting power for capacitor and resistor are lost. Figure 5.12 depicts the real part of impedance versus frequency. We can see that current model is incapable of predicting the actual resistance (Real part of the impedance). The RLC model with only the inductance and other parameters set to zero, such as capacitance and resistance, does not appear to be ideal for model fitting. This is an example of erroneous parameters, which will diminish our model's capacity for precise prediction. In order to find the optimal parameter for our model, we endeavour to optimise these three parameters which

can be determined by finding the maximum log-likelihood estimate.

Loss represents the difference between our model's output and actual measurement findings. If the predictive capacity of model is very accurate, then loss will be reduced and close to 0, however if the model's fit is not ideal, then the quantity of loss will be large.

Given the current model, the-log likelihood function in the normalized model can be interpreted as the probability that the output of our model closely resembles the measured data. To achieve a reliable model, we must minimise the loss, which is the same concept as minimising the negative log likelihood function. Thus, we assume that the loss function is equivalent to the negative log likelihood function up to a multiplicative parameter that is inverse the error variance. The expression for loss is shown below:

$$\sum_{\nu} |X_{mes}(\nu) - X(\nu,\theta)|^2 = loss = -\ell, \tag{5.2}$$

in which predicted impedance $X(\nu,\theta)$ can be viewed as a function with respect to frequency parameter ν and vector $\theta \in [0,\infty]$. And $X_{mes}(\nu)$ represent the measured Impedance at give frequency ν.

Considering the Equation 5.1, the predicted impedance of RLC model can be expressed in terms of frequency ν and vector θ=[$\theta_1,\theta_2,\theta_3$], in which $\theta_1,\theta_2,\theta_3$ represent Inductance, Capacitance, and Resistance respectively. $X_{RLC}(\nu,\theta)$ can therefore defined as follows:

$$X_{RLC}(\nu,\theta) = \Big(\frac{1}{\theta_3 + 2\pi\nu\theta_1 j} + 2\pi\nu\theta_2 j\Big)^{-1} \tag{5.3}$$

The maximum log-likelihood estimate under normalised model is

$$\hat{\theta} = arg\min_{\theta}\Big(-\ell(\theta)\Big) = arg\max_{\theta}\ell(\theta). \tag{5.4}$$

Model error (ε) is the stochastic error that model the difference between pre-

dicted impedance and the actual measured value.

$$X_{mes}(\nu) = X(\nu,\theta) + \varepsilon \tag{5.5}$$

Impedance is a complex number consisting of both a real and an imaginary component. When modelling the residual, it is natural for us to consider residual as the random vector and the imaginary and real parts as two dependent normal random variables. Therefore, a bivariate normal distribution might be a appropriate choice to model the probability distribution of the error. And it can be defined as:

$$\begin{pmatrix} \varepsilon_{Re} \\ \varepsilon_{Im} \end{pmatrix} = \varepsilon \sim N(0,\Sigma), \Sigma = \begin{bmatrix} {\sigma_r}^2 & \sigma_r\sigma_i \\ \sigma_r\sigma_i & {\sigma_i}^2 \end{bmatrix} \tag{5.6}$$

We can now compute the joint probability density function of ε_{Re} and ε_{Im}. It can be written as

$$f_X(x;\theta) = \prod_{i=1}^{N} \frac{1}{(\sqrt{det(2\pi\Sigma)})} exp\{-\frac{1}{2}(X_{mes}(\nu) - X(\nu,\theta))^T \Sigma^{-1} (X_{mes}(\nu) - X(\nu,\theta))\}. \tag{5.7}$$

The quantity in $\{\cdot\}$ is a quadratic form. The measured impedance and modeled impedance are in the vector form written as:

$$X_{mes} = \begin{bmatrix} Re(X_{mes}) \\ Im(X_{mes}) \end{bmatrix}, X(\nu;\theta) = \begin{bmatrix} Re(X(\nu;\theta)) \\ Im(X(\nu;\theta)) \end{bmatrix}. \tag{5.8}$$

According to the definition of the log-likelihood function expressed in Equation 3.29, we can then take a logarithm of the joint probability distribution and compute the log-likelihood function in terms of the covariance matrix[Covariance matrix in the log-likelihood function is also estimated and it is demonstrated in the Appendix]. The estimated log-likelihood function can be expressed as:

$$\ell(x;\hat{\theta}) = -\frac{1}{2}\sum_{1}^{n} \left| (X_{mes}(\nu) - X(\nu,\hat{\theta}))^T \hat{\Sigma}^{-1} (X_{mes}(\nu) - X(\nu,\hat{\theta})) + log(det(2\pi\hat{\Sigma})) \right| \tag{5.9}$$

In order to determine the maximum log-likelihood estimates, we will now employ the optimization methodologies introduced in Chapter 2 to the log-likelihood function of RLC model. As an illustration, the best estimate obtained by using different optimisation methods is shown below:

Simple Inductance Model	Inductance[H]	Capacitance[F]	Resistance[Ω]	Log-likelihood
Initial Parameters	4	0	0	-17522.83
Parameter after BaysOpt	4.0157	11.5308	55.9391	-11796.56
Parameter afterBaysOpt&BFGS	3.9801	11.6554	53.7760	-11709.51

Table 5.1: A comparison of the quantity of the log-likelihood resulting from NO statistical optimization, the BFGS method alone, and combination of BFGS and Bayesian Optimization together.

Observing the above table visually reveals the trend of improvement of the predictive performance of the RLC model that using different sequence of optimisation method. Initial log-likelihood has been significantly improved from -17522 to -11709. Model optimised by the BFGS method after Bayesian optimisation is preferable to models with no optimisation or Bayesian optimisation alone. Because its log-likelihood is less negative than others.

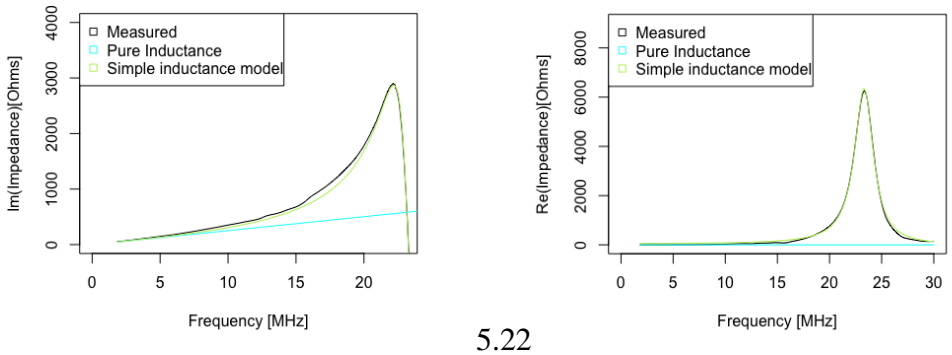
Table demonstrated below represent the iterations and the updating procedure accompanied with the quantity of log-likelihood of Bayesian Optimisation.

Iterations of BaysOpt	Inductance	Capacitance	Resistance	Log-likelihood
1st sampling point	4.551552	28.75003	38.48987	-17733.31
2nd sampling point	5.065576	17.28041	99.26881	-17704.78
3rd sampling point	3.186615	29.44952	43.4751	-17610.38
4th sampling point	3.877136	24.72166	29.4267	-17819.73
5th sampling point	4.760303	35.87559	73.30402	-17587.10
1st iteration	4.162245	10.93917	64.24394	-13736.15
2nd iteration	4.073188	10.47663	61.77328	-16459.45
3rd iteration	4.696899	11.28179	64.37451	-17111.47
4th iteration	4.155704	11.11235	56.73725	-12582.85
5th iteration	4.203839	11.3686	57.3363	-13591.75
⋮	⋮	⋮	⋮	⋮

Table 5.2: Quantity of log-likelihoods and parameters for each iterations and sampling points of Bayesian optimisation

The first five rows of the table indicate five points on the log-likelihood function that were randomly selected by the surrogate function. The remaining rows of

the table indicate the subsequent iteration and updating process obtained by learning the prior randomly sampled points. I have only displayed the first five iterations of Bayesian Optimisation; BayesOpt would require at least 50 iterations to get a reliable estimate. The iteration with the largest value of log-likelihood is already displayed in Table 5.1.



5.21 5.22

Figure 5.2: The measured impedance of 4 μH datasets is compared to the impedance predicted by model 1(RLC Model) after statistical optimization.

The green lines in Figure 5.2 indicate the model's forecast after the optimisation methods was applied. It was evident that the simulation conditions resulting from the optimisation procedure for both the imaginary and real parts of the impedance were vastly superior to those of the original model (Inductance model shown in blue). Our model's forecast nearly matched the measured impedance in the real electric circuit. However, there are still modest deviations from the prediction for frequencies 15–20 MHz and 25–30 MHz for the imaginary and real parts of the impedance, respectively. Y-axis has a large scale, so even a small deviation may cause significant issues.

Looking at Figure 5.22, we can see that at the low frequency range, the real part of the predicted impedance resemble a straight line. This is because the inductor acts as a short circuit in the RLC model (Figure 2.5) , and the capacitor acts as an open circuit (Since an ideal capacitor has infinite resistance, so electricity will not pass through.), leaving only the resistor in the circuit.

In order to have a deeper understanding of the performance of our model, we must consider its residuals. Residuals are utilised in numerous processes aiming

to detect various forms of model-data discrepancies. In conventional usage, we also use the term ”error”. Chesher and Irish’s work from the early 1980s forms the foundation for a significant proportion of the prevalent residual analysis techniques [8].

The Affine relationship between the index and the frequency f_i can be expressed as:

$$f_i = i \times \Delta f + f_0. \tag{5.10}$$

f_0= 1.8 MHz is the frequency at which the measured data begins. Δf is a constant value that equal to 0.028 MHz. Index i$\in$ {1,2,3,....,n }, in which n denotes the number of measurements available for each inductor.

Residuals are the difference between the predicted and measured data and are defined by the following formula, where Y_f and y_f are n $\times$ 1 vectors representing the fitted value and measured data at frequency f_i, respectively [5].

$$Y_f = \begin{bmatrix} Y_{f_0} \\ Y_{f_1} \\ \vdots \\ Y_{f_n} \end{bmatrix}, y_f = \begin{bmatrix} y_{f_0} \\ y_{f_1} \\ \vdots \\ y_{f_n} \end{bmatrix} \tag{5.11}$$

Therefore, the residuals can represented as:

$$Z_f = y_f - Y_f. \tag{5.12}$$

We can also express this more compactly using matrix notation:

$$\begin{bmatrix} Z_{f_0} \\ Z_{f_1} \\ \vdots \\ Z_{f_n} \end{bmatrix} = \begin{bmatrix} y_{f_0} \\ y_{f_1} \\ \vdots \\ y_{f_n} \end{bmatrix} - \begin{bmatrix} Y_{f_0} \\ Y_{f_1} \\ \vdots \\ Y_{f_n} \end{bmatrix} \tag{5.13}$$

On the basis of an idealised theory, we believe that our model error follows a normal distribution with mean zero and constant variance, the Residuals against

Frequency plots should vary around 0 and not exhibit a regular pattern, akin to random noise shown in Figure 5.5.

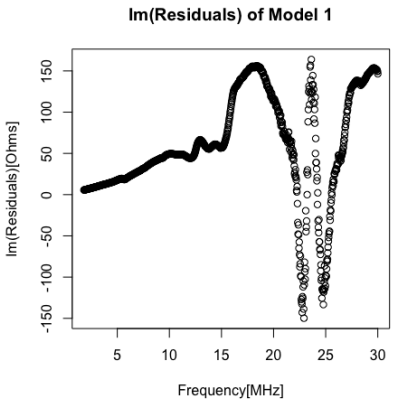


Figure 5.3: Imaginary part of the residuals of Model 1 testing the 4μH data sets

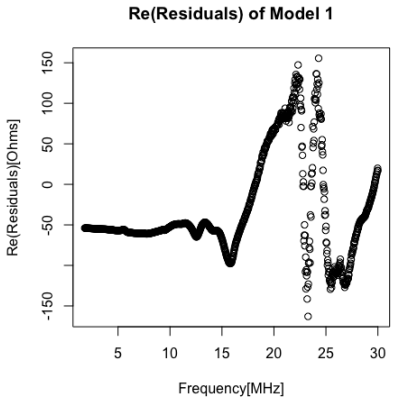


Figure 5.4: Real part of the residuals of Model 1 testing the 4μH data sets

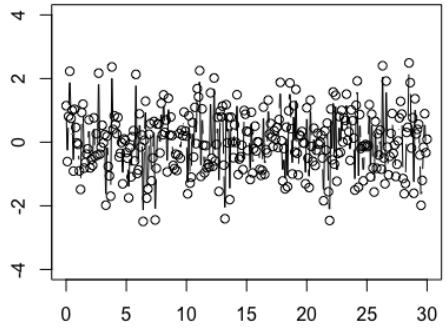


Figure 5.5: Expected plot of the Residual against Frequency

The two pictures of Figure 5.3 and 5.4 depict, respectively, the imaginary and real parts of residuals against with frequency. The preceding images Figure 5.2 led us to conclude that the simulation power of our model after statistical optimization was favourable. However, the result depicted in Figure 5.3 and 5.4 is not. Both imaginary and real part of the residuals are structured and depicts a clear increasing tendency. A significant peak can be observed at the frequency of 17 Mhz and 24 Mhz.

Moreover, the variance of the residual is not constant over time, and variance occurs in clusters, i.e., there are periods of extreme volatility, such as at the resonance frequency of approximately 23 MHz. There are also periods of low volatility between 5Mhz and 15Mhz. Moreover, Figure 5.3 differs significantly from the expected Residual distribution (Figure 5.3). The study of the evident upward trend suggests that the impedance of our current model is significantly lower than the measured impedance, and therefore the current model's predictive performance is insufficient.

Likewise, the real part of the Residual against Frequency plot (Figure 5.4) is structured and therefore not favourable. However, by observing the imperfect residual image, I can attempt to improve my current model by incorporating additional electronic components. When frequency is around 1 MHz, the residual is roughly -50 ohms, and it remains at this value until frequency is increased to 15 MHz. Such an image characteristic indicates that the total resistance in simple electric circuit is greater than the measured data, causing the curve to oscillate at -50 ohms. Upon inspection of the residual plots, it is evident that the residual does not follow the normal distribution.

After examining the imaginary part of the residuals of the first model, in order to combat the rising trend shown in Figure 5.3, we can attempt to increase the impedance of the electric circuit by adding a series-connected inductor. The modifications made to model 1 result in the formation of a new models, model 2 which is given in Figure 5.8.

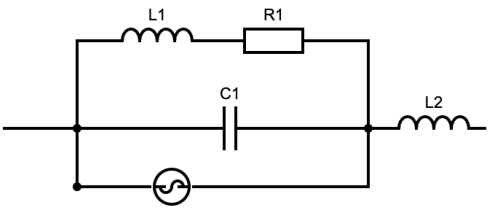
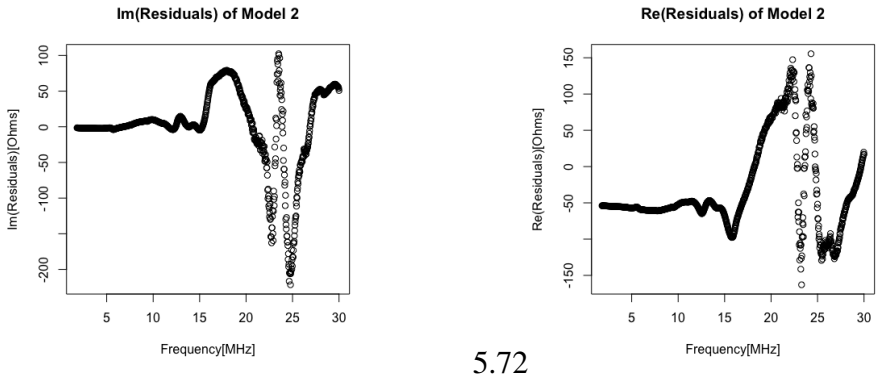


Figure 5.6: Model 2

On the basis of model 1(Figure 2.2), model 2 adds a series-connected inductor, which only affects the imaginary part of the impedance. Hence, we anticipate no



5.71 5.72

Figure 5.7: Residual against Frequency graph of Model 2 testing the 4μH data sets

change in the real part of the residual plot of model 2. Furthermore, as demonstrated in Figure 5.7, Model 2 does resolve the rising trend in Model 1's image, indicating that our model fitting has been improved. Yet, the general appearance of the residual plot of model 2 is still significantly different from what we anticipated(Figure 5.5), and the residual distribution is definitely not normal.

Magnetic saturation, which is the collapse of the magnetic field, thermal phenomena, and inductor resonance are three explanations for model's poor predictive ability. We are using iron power toroidal inductor with low signal which is unlikely to to reach magnetic saturation. Blanchard [6] and Detka [11] explain the underlying physics behind the difficulty of simulating a resonant inductor, and why thermal phenomena must be taken into consideration with toroidal inductors in particular. These demand a deeper understanding of physics, and fitting models like this goes slightly beyond the scope of the research.

Model 1 and Model 2, their imaginary part of the residual plot are both not satisfying, since they are oscillating around -50 Ohms and we expect they are oscillating around 0. On the basis of model 1, model 3 (shown in Figure 5.8) is presented, which, by replacing the series connection between the resistor and inductor with a parallel connection, may possibly alleviate the problem of the current model's excessive resistance.

Compare the Residual plot of model 3(Figure 5.9) with model 1 (Figure 5.3 and Figure 5.4). Consider initially the imaginary part of residuals. It is worth noting that, Model 1 has a maximum absolute residual of approximately 100 at its resonant

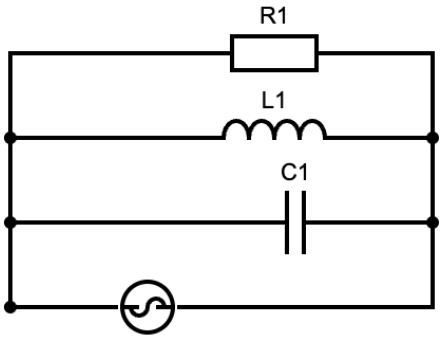
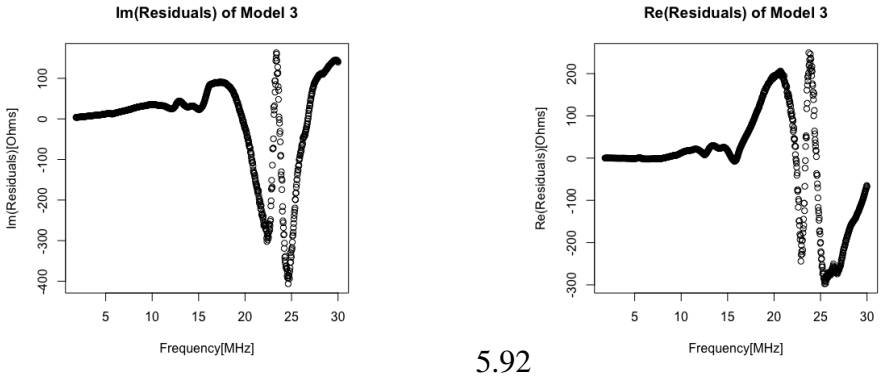


Figure 5.8: Model 3



5.91 5.92

Figure 5.9: Residual against Frequency graph of Model 3 testing the 4μH data sets

frequency, whereas model 3 reaches a shocking 450. This indicates that model 3's ability to forecast at resonant frequency is inferior to model 1's. In addition, When we consider the real part of residual. Model 3's maximum absolute residual at the resonant frequency is 300 ohms, which is roughly double that of Model 1 (150 Ohms). Consequently, in this example based on the image, model 3 is inferior to model 1.

After optimising the parameters of each model and computing the log-likelihood, the AIC of each model can be derived using the Chapter 2 definition of AIC (Equation 3.30).

	Model 1	Model 2	Model 3
Akaike Information Criterion	23561.17	22213.40	24695.37

Table 5.3

In our context, the number of Model's parameters is not excessively high which

is approximately around 5. Hence, the error of the model, as indicated by the model's likelihood function, would be the primary factor that influence AIC. So, we may intuitively assess our model's predictive performance based on the quantity of AIC. Table 5.3 data reveals that Model 2 has the smallest AIC, hence we may consider model 2 to be the best model at present. Predictive performance has been greatly enhanced by the addition of an inductor to Model 1. Additionally, the AIC of Model 3 is significantly higher than that of the previous model, which confirms our preliminary judgment based on the residual plot of Model 3.

Model 2 is the superior model. Yet, although model 3's predictive accuracy is quite weak, but its real part of the residuals no longer fluctuate around -50 Ohms. Hence, we attempt to integrate them in two distinct ways to create two new models, model 4 and model 5, depicted in Figure 5.10.

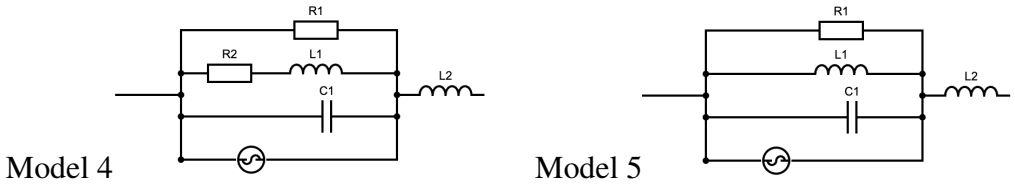


Figure 5.10: Two new candidate models, Models 4 and 5, will be evaluated for their predictive performance.

The residual plot for Model 5 is shown in Figure 5.15, which reveals that the residual plot of Model 4 is analogous to that of Model 2. Both fluctuation and kurtosis are essentially unchanged.

	Model 2	Model 4
AIC	22213.40	22246.96

Table 5.4

Table 5.4 shows the AIC of Model 2 and Model 4. Model 4's AIC is only slightly greater than that of model 2. Consequently, we can infer that the addition of a parallel resistor does not appear to greatly enhance our model's simulation performance.

Table 5.5 represents the log-likelihood estimate ($\hat{\theta}$) of Model 5 that was calculated using the Bayesian Optimisation and BFGS methods. 'Inductance(2)' ap-

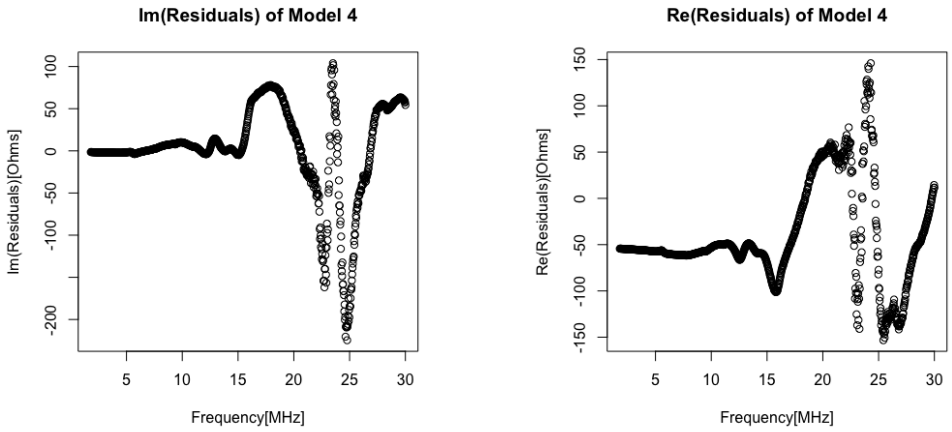


Figure 5.11: Residual against Frequency graph of Model 4 testing the 4μH data sets

peared in the Table 5.5 corresponds to the newly added inductor labeld with 'L2' shown in Figure 5.10. The extremely low value of Inductance(2) indicates that the L2 inductor contributes almost nothing towards the prediction of the impedance. Hence, the predictive ability of model 5 is essentially identical to that of its nested model, model 3.

Model 5	Inductance[H]	Capacitance[F]	Resistance[Ω]	Inductance(2)[H]
Parameters	4.1065	11.4026	6350.8349	2.2×10^{-9}

Table 5.5: Values of Model 4's parameters after applying the BFGS and Bayesian optimisation.

Table 5.6 demonstrates that the Akike information criteria for Models 3 and 5 are similar, confirming our hypothesis that Model 5 is comparable to Model 3.

	Model 3	Model 5
AIC	24695.37	24701.43

Table 5.6

Initial attempted model, Model 1 and Model 2, have performed exceptionally well in model predictive performance to $4\mu H$ data set. Based on the AIC values of models 4 and 5, we determined that adding or repositioning electronic components had a negligible effect on the simulation of $4\mu H$ data. It is worth exploring the scenario in which connecting Model 1 and Model 2 in parallel and then to deter-

mine whether combining the two models would considerably enhance the model's predicted accuracy. I anticipated Figure 5.12 to depict a general circuit illustration.

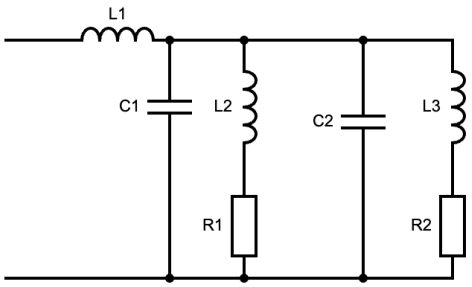


Figure 5.12: The speculated model that combine Model 1 and Model 2.

The combination of electric components of L1, C1, L2, and R1 in Figure 5.12 is Model 2, and C2,L3, and R2, which represent Model 1. Nevertheless, circuit diagram shown in Figure 5.12 cannot be utilised straightly as model 6. Because the two capacitors (C1 and C2) are non-identifiable in the electric circuit since they are observationally equivalent [21]. There are distinct capacitance values of capacitors that result in the same data distribution.

As illustrated in Figure 5.13, two parallel capacitors each with capacitance C1 and C2 respectively, is identical to the capacitance (C1+C2) of a single parallel capacitor. Similarly, the L2 and R1 branch in Figure 5.12 are not identifiable from the

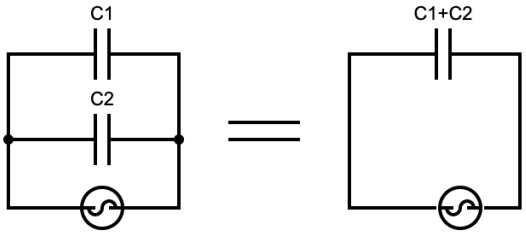


Figure 5.13

L3 and R2 branch. Nonetheless, for our new model, retaining a set of unidentifiable branches and eliminating the unnecessary capacitor is acceptable. The ultimate version of Model 6 is illustrated in Figure 5.14.

When comparing the residual plot of model 6 to the residual plot of the image of model 2, the best model at present (Figure 5.7), the difference between them is

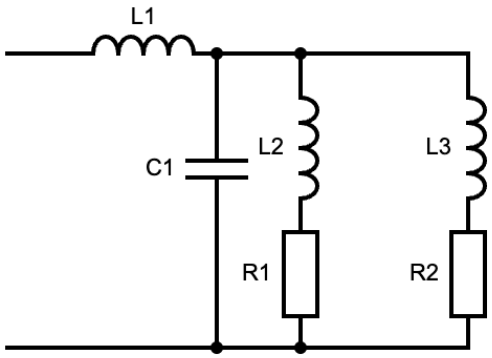


Figure 5.14: Model 6

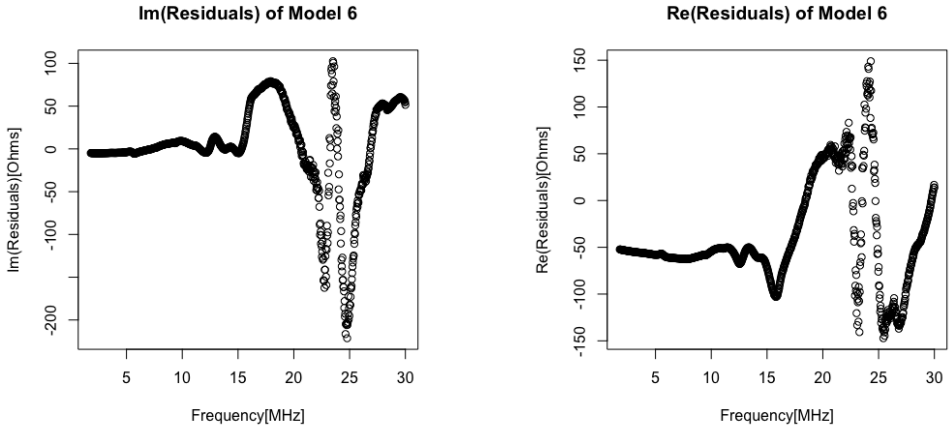


Figure 5.15: Residual against Frequency graph of Model 6 testing the 4μH data sets

negligible, and their images are nearly identical.

Table 5.7 and displays the AIC of the all candidate models.

At this stage, it is possible to performing the Kolmogorov-Smirnov test to check the normality of the residuals for current 6 model. The p-value of the residuals' real and imaginary part are compared for uniformity in the Table 5.8. The significance results for all existing models are less tan 2.2×10^{-16}. which means there are substantial evidence to deny the hypothesis that the residuals followed the normal distribution.

However, even if the final p-value is very little, I expected model 6's signifi-

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
AIC	23561.17	22213.40	24695.37	22246.96	24701.41	22208.03

Table 5.7

Kolmogorov-Smirnov Test	Residual	Residual
Model 1: P- value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$
Model 2: P- value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$
Model 3: P- value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$
Model 4: P- value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$
Model 5: P- value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$
Model 6: P- value	$<2.2 \times 10^{-16}$	$<2.2 \times 10^{-16}$

Table 5.8

cance results to be slightly larger than the p-values of the other models. The fact that the p-value of all models is tiny and hence equal to 2.2×10^{-16} indicates that the predictive performance improvement gain from our model updates is not statistically significant.

We may anticipate that there will be a difference between the value predicted by our model and the actual measured data, as this error is inherent in the model's prediction. Residual's degree of volatility is determined by its variance, which is also known as a stochastic error. Nonetheless, according to Table 5.8, none of the six models currently under consideration satisfy the normality assumption. This also indicates that the systematic error of our model is significantly greater than the expected stochastic error.

Model 6 has the smallest AIC of the six tested models, with a value of 22264.71. Due to the fact that the AIC of all models are quite high, the AIC values of models 6, 2, and 4 are not significantly different (Difference of AIC is around 2000).

5.2 Simulation of 0.5 Microhenry to 2 Microhenry Cylindrical Inductor

The AIC value of the existing model's simulation of 0.5-4 μH data is displayed in Table 5.9. We can see that model 2 simulates 0.5-1 μH data the best. Models 4 and 6 performed optimally while modelling 2 and 4 μH data, respectively. All of our models exhibit great predictive performance when simulating 1 μH data set (Models have a mean AIC value of 3,600), and there is little variation in their

	0.5 μ H dataset	1 μ H dataset	2 μ H dataset	4 μ H dataset
AIC of Model 1	4526.18	3103.84	14615.05	23561.17
AIC of Model 2	2065.27	270.54	11631.38	22213.40
AIC of Model 3	5915.34	4916.83	13511.53	24695.37
AIC of Model 4	2987.33	1543.83	10618.97	22246.96
AIC of Model 5	5917.33	4920.69	11574.61	24801.41
AIC of Model 6	4502.32	1807.58	11632.28	22208.03
Best Model	Model 2	Model 2	Model 4	Model 6

Table 5.9: AIC for six current models testing data sets ranging from 0.5μH to 4μH

AIC values. One probable explanation is that when simulating a 1 μH coil, the number of turns of coils is substantially less than that of the other coils with a larger inductance. As a result, the former coil has a lower inter-winding capacitance (Inter-winding capacitance is illustrated in Figure 2.6), and is therefore easier to simulate. In contrast, the predictive performance of all models simulating 4 μH and 2 μH data are quite constrained.

Model 2 that simulating the 2μH data set has the lowest AIC through out(AIC=270.54). It's residuals graph is shown in Figure 5.16.

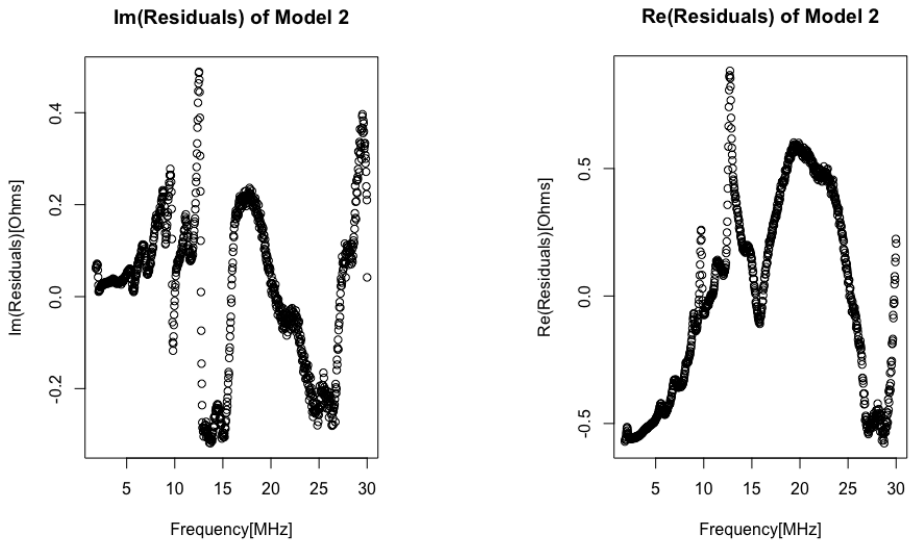


Figure 5.16: Residual against Frequency graph of Model 2 simulating the 2μH data sets

Figure 5.7 depicts the residual plot of Model 2 when simulating the data of 4 μ H. We can observe that the range of both the real and imaginary parts of residual is between 55 and -200 Ohms. Note that we are now simulating the 0.5 μ H data, we

expect the relative error to be within (6.875, -25)[Ohms], which is approximately eight times smaller than the range depicted in Figure 5.7. However, the residual plot degree of volatility depicted in Figure 5.16 is significantly lower than what we anticipated, the range of residual is between 0.5 Ohms and -0.2 Ohms. Yet, they are not a random scattering of points around (0,0), as the jitter of the points follows a very distinct pattern.

We have not yet discovered a model with superior performance. But, it is quite difficult to find the ideal model in real life. As often attributed [4] to George Box, "All models are wrong, but some are useful".

5.3 Modeling the Low Frequency Range

We can infer that the ability to predict inductor in around self-resonant frequency [24Mhz] is really limited by observing residual plots for $4\mu H$ data set. However the prediction ability of models may be enhanced so long as they remain below self-resonance frequency. In order to confirm our claim, we will test the model with the best performance when simulating the varied data set from 0.5 to 4 μH recorded in Table 5.9. Each model will now simulate the low frequency range of 1.8 to 15 MHz only. The result is shown below.

	Model 2 fitting 0.5uh data	Model 2 fitting 1 uh data	Model 4 fitting 2 uh data	Model 6 fitting 4uh data
AIC for [1.8-30Mhz]	2065.27	270.54	10618.97	22208.03
E_{pp}[1.8-30Mhz]	1.02	0.131	5.30	11.1
E_{pp}[1.8-15Mhz]	-0.611	-1.04	1.652	6.4
Resonant Frequency	101Mhz	86Mhz	35Mhz	24Mhz

Table 5.10: Comparison of the Error per point(E_{pp}) for the best model fitting the 0.5-4μH data sets with a frequency range of 1.8 MHz to 15 MHz and the entire frequency range(1.8-30MHz).

There are 1000 data sets for the entire frequency range of 1.8-30Mhz, and 500 data sets for the low frequency range of 1.8-15Mhz. The 2 data set have different length, therefore, I am applying Error per point E_{pp} to evaluate the performance of the same model at different frequency. Since Error per point is a relatively straight-forward statistic for determining the amount of error at each data point. It is defined

as:

$$E_{pp}(\theta) = \frac{1}{2\nu_{max}} \sum_{\nu}^{\nu_{max}} |X_{mes}(\nu) - X(\nu,\theta)|^2. \tag{5.14}$$

ν_{max} represent the maximum of frequency of the choosing data set.

If our model's predictive performance is consistent across the entire frequency range, then we might expect $\ell(\hat{\theta};D1) \approx 2\ell(\hat{\theta};D2)$, where $D1$ and $D2$ represent whole range frequency and low frequency data set respectively and the Error per point E_{pp} should be approximately consistent within the full frequency range.

By observing the Error per point (E_{pp}) in Table 5.10 . When simulating the low frequency range of 0.5 and 1 data, the Error per pint of model 2 becomes negative, indicating that the error is reduced and the model's prediction is more accurate, although the improvement is not very significant. Small improvement can be explained by the skin effect, which essentially indicates that when the frequency of alternating current increases, the conductor's effective resistance increase either and causes variations in alternating current that affect impedance. Ken [28] elaborated in his report on how to simulate the skin effect of an inductor which requires more electromagnetic understanding.

The maximum frequency of data set is 30Mhz, and the resonant frequency of $0.5\mu H$ and $1\mu H$ data is 101Mhz and 86Mhz, respectively. As a result, model 2 does not need to simulate inductor performance at resonant frequency, so the model predictive performance is almost consistent across the entire frequency range.

However, the $2\mu H$ and $4\mu H$ data sets have resonant frequencies of 35Mhz and 24Mhz, respectively. Therefore, model must simulate inductor impedance near resonant frequency, which is extremely difficult to model. So there is a significant improvement in performance when the model simulates low frequency range by considering the quantity of E_{pp}.

Chapter 6

Inductance for 2 inductors connected in series

Inductor can store energy in the form of magnetic field, so when there are multiple inductors in an electric circuit, the magnetic fields of the inductors will inevitably interfere with one another. If the extent of magnetic coupling effect is relatively small, then it may be negligible, but if it is significant, ignoring it could cause problems with the system because the inductor no longer operates as desired. Therefore, understanding whether the coupling effect exists between the series-connected inductors is of great interest of this section.

This section will discuss how accurate it is to assert that measured impedance at 1μH (Denoting by $X_{1\mu H}$) plus measured impedance at 4μH ($X_{4\mu H}$) equal to the measured impedance at 5μH ($X_{5\mu H}$). Given the magnetic coupling effect, the total impedance (X_{total}) of 2 series connected inductors that are close enough to one other may not necessarily be $X_{total} = 2L1$. Moreover, in the case of perfect coupling, where all of the magnetic flux of each inductor points in the same direction, resulting in zero magnetic flux loss, the total impedance could surprisingly be $X_{total} = 4L1$. In contrast with the perfect coupling, perfect anti-coupling effect means all magnetic flux between the inductor cancels out, then the total impedance is only $X_{total} = 0$.

I'm testing the coupling effect between series-connected inductors using the model-based log-likelihood ratio test. The null and alternative hypotheses are:

H_0: $X_{1\mu H} + X_{4\mu H} = X_{5\mu H}$ (Suggesting no coupling effect).

H_1: $X_{1\mu H} + X_{4\mu H} \neq X_{5\mu H}$ (Suggesting an existence of coupling effect).

I will use the RLC model (Shown in Figure 2.2), which is the nested model and the first model I chose for this project to compute the log-likelihood function. For the sake of compact expressions, I will use the term $RLC_{1\mu H}$ to denote RLC model modeling the 1 μH data.

$\ell_0(\theta_0)$ represent the log-likelihood function under the Null hypothesis for the modeling of $RLC_{4\mu H}$ & $RLC_{1\mu H}$. It's expression is shown in the formula below where $\theta_0 \in R^6$, since there are two models with three-dimensional parameters each (which are inductance, capacitance, and resistance).

$$
\begin{aligned}
\ell_0(\theta_0) = -\frac{1}{2}\Bigg[& \sum_{1}^{n} \left| \zeta^T \Sigma_1^{-1} \zeta + log(det(2\pi\Sigma_1)) \right| \\
& + \sum_{\nu} \left| \big(X_{1\mu H}(\nu) - RLC_{1\mu H}(\nu,\theta_0)\big)^T \Sigma_2^{-1} \big(X_{1\mu H}(\nu) - RLC_{1\mu H}(\nu,\theta_0)\big) + log(det(2\pi\Sigma_2)) \right| \\
& + \sum_{\nu} \left| \big(X_{4\mu H}(\nu) - RLC_{4\mu H}(\nu,\theta_0)\big)^T \Sigma_3^{-1} \Big(X_{4\mu H}(\nu) - RLC_{4\mu H}(\nu,\theta_0)\Big) + log(det(2\pi\Sigma_3)) \right| \Bigg] \\
& \textbf{Where} : \zeta = \Big(X_{5\mu H}(\nu) - (RLC_{1\mu H}(\nu,\theta_0) + RLC_{4\mu H}(\nu,\theta_0)\Big)
\end{aligned}
\tag{6.1}
$$

The log-likelihood function of ℓ_0 simply expresses one thing: we are aiming to identify model's parameters that not only maximise the log-likelihood when modelling $1\mu H$ and $4\mu H$ data separately, but also maximise the log-likelihood when they are corporate with each other in modelling the $5\mu H$ data. Note that, the estimated covariance matrix for the modeling of 1μH and 4μH data set $[\hat{\Sigma_2}, \hat{\Sigma_3}]$ is now estimated also based on the residuals when $RLC_{1\mu H}$ and $RLC_{4\mu H}$ incorporate with each other in modeling the 5μH data.

$\ell(\theta)$ represent the sum of the log-likelihood function for the Alternative hypothesis for the modeling for RLC of $1\mu H$, $4\mu H$, and $5\mu H$ data solely. It can be

expressed as:

$$\ell(\theta) = \left[\sum_{\nu} \left| \left(X_{5\mu H}(\nu) - RLC_{5\mu H}(\nu,\theta)\right)^T \Sigma_4^{-1} \left(X_{5\mu H}(\nu) - RLC_{5\mu H}(\nu,\theta)\right) + log(det(2\pi\Sigma_4)) \right| \right.$$
$$+ \sum_{\nu} \left| \left(X_{1\mu H}(\nu) - RLC_{1\mu H}(\nu,\theta)\right)^T \Sigma_5^{-1} \left(X_{1\mu H}(\nu) - RLC_{1\mu H}(\nu,\theta)\right) + log(det(2\pi\Sigma_5)) \right|$$
$$\left. + \sum_{\nu} \left| \left(X_{4\mu H}(\nu) - RLC_{4\mu H}(\nu,\theta)\right)^T \Sigma_6^{-1} \left(X_{4\mu H}(\nu) - RLC_{4\mu H}(\nu,\theta)\right) + log(det(2\pi\Sigma_6)) \right| \right] \tag{6.2}$$

Where $\theta \in R^9$

Under the Null hypothesis, if inductors are uncoupled, then we would expect the log-likelihood function $\ell(\theta)$ and $\ell_0(\theta_0)$ has no significant difference, which indicating that the RLC model modeling the $1\mu H$ and $4\mu H$ inductor are statistically indistinguishable form the RLC model modeling the $5\mu H$ inductor.

Test Statistics T, which is the 2 times of the difference between $\ell(\theta)$ and $\ell_0(\theta)$, is asymptotically under the null hypothesis follows the Chi-square distribution with the degree of freedom of 3 [7].

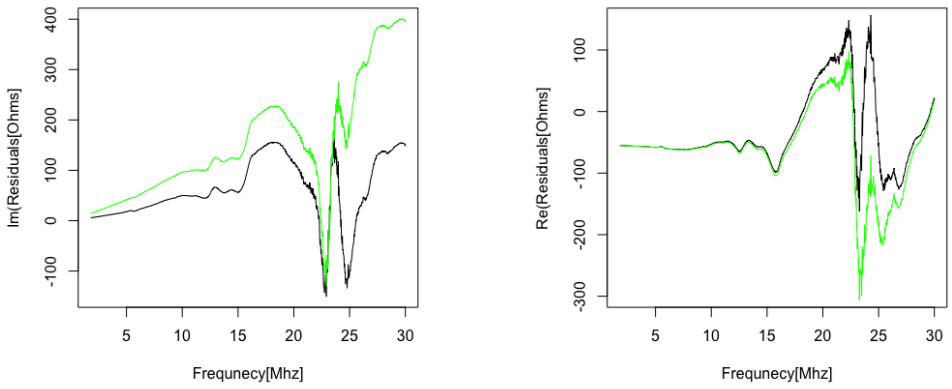
$$T = -2(\ell(\theta) - \ell_0(\theta_0)) \sim \chi^2_{9-6=3} \tag{6.3}$$

By applying the statistical optimisation to the Equation 6.1 and 6.2, we are be able to find the maximum log-likelihood estimate $\hat{\theta}, \hat{\theta}_0$, and hence compute the quantity of $\ell(\hat{\theta})$ and $\ell_0(\hat{\theta}_0)$, which are -25051 and -24271 respectively. High-valued likelihood function implies that the data set is more consistent with the model than with a model with a lower likelihood function. In our context, $\ell(\hat{\theta})$ has a relative larger value compared with $\ell_0(\hat{\theta}_0)$. And the test statistics T=1158 is obviously lies in the rejection region.

In order to make precise statistical inferences, the likelihood ratio test (LRT) would require an accurate likelihood function. In our context, however, estimating the variance of the residuals is challenging, and the predictive performance of the current 6 candidate models in simulating the higher inductance inductor is not reli-

able, so the statistical results cannot be relied upon. Instead, we will only consider LRT as a theoretical possibility in this context.

We may further investigate the existence of the coupling effect by visually observe the graph that shows the deviation of the $RLC_{1\mu H} + RLC_{4\mu H}$ from the measured 5μ H data compare with the deviation of the measured 5μ H data from a fitted $RLC_{5\mu H}$ model.



The black line shown in the graph above represent the residuals of the measured 5μ H data from a fitted $RLC_{5\mu H}$ model. And the green line shown in the graph above is the deviation of the $RLC_{1\mu H} + RLC_{4\mu H}$ from the measured 5μ H data set. The predictive performance of our model $RLC_{1\mu H} + RLC_{4\mu H}$ is estimated to be comparable to that of model $RLC_{5\mu H}$ in the absence of coupling effects. However, through direct observation of images, we can determine that the residuals are quite distinct from one another. Furthermore, in the resonant frequency interval, the deviation between the two residuals curves is most prominent. In summary, we can conclude that a coupling effect does exist between series-connected inductors and the effect is rather obvious. We may have to consider the influence caused by the coupling effect in later modeling.

Chapter 7

Conclusion

In this study, I tested a total of six distinct nonlinear models and performed statistical analysis on data sets ranging from $0.5\mu H$ to $4\mu H$ in the entire frequency range and the low frequency range, respectively. All models did reasonably well in modelling the $0.5\mu H$ and $1\mu H$ data sets. While the predictive performance for current 6 nonlinear models modelling the $2\mu H$ and $4\mu H$ data sets are limited.

Due to the skin effect, the predictive performance of all models in the low frequency range has improved, and the performance of the models that simulate the $2\mu H$ and $4\mu H$ data sets has improved substantially, as they are no longer required to simulate the impedance near the resonant frequency.

The first nest model examined in this study is an RLC, which has only three parameters. As a result, if Bayesian optimization is not necessary at first, the log-likelihood function can be optimised with the BFGS approach alone (with reasonable starting values) to give me a relatively accurate parameter output. Nevertheless, when attempting more sophisticated models with higher dimensions, such as Model 2 and model 6, the role of BFGS becomes very limited, and the maximum log-likelihood estimate cannot be appropriately determined. The Bayesian optimisation method was used to successfully resolve the situation.

However, carrying out the Bayesian optimisation is not a simple procedure. Before formally using the Bayesian optimisation to the objective function, we need to find the constraint for the maximum log-likelihood estimate to ensure that Bayesian optimization is meaningful. This step is important since it make sure the

maximum likelihood can be accurately found in subsequent updates. Some electronic components may not have any predictive power at all in the model, so the corresponding parameter show up as 0 during optimization. Therefore, I know from the prior knowledge that all of my parameters must be greater than or equal to 0. In addition, the exploratory data analysis I conduct at the beginning of the Section 5.1 also helped me in defining the parameter space. The result has showed that when simulating the $4\mu H$ data set, using the model that contain $4\mu H$ inductor would be a good choice.

7.1 Future work

In this project, I simulated individual inductors primarily, but this is insufficient. I have verified that the existence of coupling effect between two inductors using hypothesis test in the preceding Section 6. In this context, it is essential to simulate the inductor coupling effect because real-world appliances such as refrigerators and air conditioners have a more sophisticated electric systems that contain a greater number of inductors. As a result, accurately simulating the coupling effect is important and can be a future aim before modeling multiple inductor.

In addition, a non-parametric approach can be utilised in later inductor modelling to determine if a functional form of inductor response that generalises the response of various inductors at varying frequencies can be attained. For example, maybe the deviations of the $2\mu H$ inductor from the RLC model at frequency 15 MHz looks similar that of the $1\mu H$ inductor from its RLC model at frequency 30 MHz.

Furthermore, because of the structured residual plots, the performance of our model in replicating 0.5 and $1\mu H$ data sets that do not contain resonant frequency cannot be described as outstanding. One likely explanation is that the impedance of wires between electrical components is not considered in our inductor modelling. As a result, I believe we can attempt to model the wire impedance and discover an acceptable model. This model can be appropriately paired with our inductor candidate model in the future inductor modeling.

Appendix A

Appendix

In this project, Rstudio was the primary programming language utilised. The code that follows is the code I mostly utilised for this research for Optimisation problem. And I will add comments to each code to explain its purpose. While the code for the optimisation procedure of other models are in the same format with various parameters, I will just demonstrate the RLC model's code for our basic comprehension.

```
## Creat a function that automatically compute the log-likelihood.

LLH_RLC<-function(x,y,z){
list(

## Firstly, transfer the complex number into 2 x n matrix form,
## where n is the length of the dataset

X_mes_Re<-matrix(c(Re(inductance.table$Z[bit.subset])),nrow = 1),
X_mes_Im<-matrix(c(Im(inductance.table$Z[bit.subset])),nrow = 1),
X_mes<-rbind(X_mes_Re,X_mes_Im),

## Secondly, transfer the predicted impedance into 2 x n matrix form.

X_modeled_Re<-matrix(c(Re(LLH_RLC(inductance.table$f[bit.subset],
c(x,y,z)))),nrow=1),
X_modeled_Im<-matrix(c(Im(LLH_RLC(inductance.table$f[bit.subset],
c(x,y,z)))),nrow=1),
X_modeled<-rbind(X_modeled_Re,X_modeled_Im),
```

```
## Obatin the residuals
residuals<- X_mes-X_modeled,

## Construct the covariance matrix of the residuals
cov<-var(t(matrix(c(residuals[1,],residuals[2,]),nrow=2,
                    ncol = ncol(residuals),byrow = TRUE))),

## Determine the inverse of the covariance matrix
cov_inv<-solve(cov),

## Compute the log-likelihood function.
Score=
log_likerlihood<-(-1/2)*sum(abs(apply(t(X_mes-
X_modeled),1,function(x) x%*%cov_inv%*%x))
+log(det(2*pi*cov))),

predict=c(x,y,z)))
}
```

In the example code that I have shown above. LLH_RLC is the log-likelihood function of RLC model. And I am simply using the Equation 5.9 to find the log-likelihood function. Moreover, the code ”inductance.table$Z [bit.subset]” means choosing the wanted impedance data sets according to the logical command ”[bit.subset]” which simply allocate our different data sets using bit array method.

Now I can carry the Bayesian optimisation to the log-likelihood function.

```
RLC_par<-BayesianOptimization(LLH_RLC,bounds =
list(x = c(0,1),y=c(0,3),z=c(0,5)),
init_points = 20, n_iter = 50, acq = "ucb",
kappa = 2.576, eps = 0.0,verbose = TRUE)
```

Here, I define the log likelihood function of RLC model as an objective function and search for parameters that maximise the log likelihood. In Bayesian optimization, I

specify the interval in which I believe a parameter may exist, as well as the number of sampling and iterations.

After obtain the best estimate from the Bayesian Optimisation. Then, I construct a new function known as "LLH", which is essentially the log-likelihood function with an unknown parameter, so that it can be utilised with the BFGS optimisation method.

```
LLH<-function(parameter){
  nll<-LLH_RLC(parameter[1],parameter[2],parameter[3])$Score
  return(-nll)
}
```

BFGS methods can be employed to more precisely find our maximum Likelihood estimates based on the results of Bayesian optimisation. Therefore, I use the"*RLC_par*$*Best_Par*" as the starting value of BFGS method to locate the maximum Likelihood estimates.

```
opt.result.l2= optim(fn= LLH,par=c(RLC_par$Best_Par),
                     lower=c(0,0,0,0),method="BFGS",
                     control=count)

opt.result.l2
```

Bibliography

[1] Slurzberg; Osterheld (1950). *Essentials of Electricity for Radio and Television. 2nd ed.* McGraw-Hill, 1950.

[2] Henry De-Graft Acquah. *Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship*. University of Cape Coast, 2010.

[3] Henry De-Graft Acquah. *Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship*. University of Cape Coast, 2010.

[4] Chris Anderson. The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7):16–07, 2008.

[5] Francis John Anscombe. *Examination of residuals*. US Army Research Office-Durham, 1961.

[6] Julian Blanchard. The history of electrical resonance. *The Bell System Technical Journal*, 20(4):424–431, 1941.

[7] Berger R L. Casella G. *Statistical inference.* Cengage Learning., 2021.

[8] R Dennis Cook and Sanford Weisberg. *Residuals and influence in regression*. New York: Chapman and Hall, 1982.

[9] Bernard Dennis Cullity and Chad D Graham. *Introduction to magnetic materials*. John Wiley & Sons, 2011.

[10] J.E. Dennis and Robert B. Schnabel. Chapter i a view of unconstrained optimization. `https://www.sciencedirect.com/science/article/pii/S0927050789010029`, 1989.

[11] Kalina Detka, Krzysztof Gorecki, and Janusz Zarebski. Modeling single inductor dc–dc converters with thermal phenomena in the inductor taken into account. *IEEE Transactions on Power Electronics*, 32(9):7025–7033, 2016.

[12] F. Dong Tan, J.L. Vollin, and S.M. Cuk. A practical approach for magnetic core-loss characterization. *IEEE Transactions on Power Electronics*, 10(2):124–130, 1995.

[13] Roger Fletcher. *Practical methods of optimisation*. Chichester ; New York : Wiley, 2 edition, 1987.

[14] Roger Fletcher. *Practical Methods of Optimization (2nd ed.)*. New York: John Wiley Sons, 1987.

[15] Svetlozar T. Rachev Frank J.Fabozzi, Sergio M. Focardi and Bala G. Arshanapalli. *The Basics of Financial Econometric: Tools,Concepts,and Asset Management Applications.* John WileySons, Inc., 2014.

[16] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, page 1, 2018.

[17] Seymour Geisser. *Predictive Inference.* New York, NY: Chapman and Hall., 1993.

[18] Madhu Sudan Gupta. Georg simon ohm and ohm's law. *IEEE Transactions on Education*, 23(3):156–162, 1980.

[19] Roger F Harrington. *Introduction to electromagnetic engineering*. Courier Corporation, 2003.

[20] 1920 Hayt, William Hart. *Engineering electromagnetics(5th ed).* New York : McGraw-Hill, 1989.

[21] Cheng Hsiao. Identification. *Handbook of econometrics*, 1:223–283, 1983.

[22] Kuldeep Kumar. Jan G. De Gooijer. *Some recent developments in non-linear time series modelling, testing, and forecasting,*. International Journal of Forecasting, 1992.

[23] S. L. Kakani. *Electronics Theory and Applications*. New Age International, 1 edition, 2005.

[24] Kwan Chi Kao. *Dielectric phenomena in solids*. Elsevier, 2004.

[25] Nahum Kipnis. A law of physics in the classroom: The case of ohm's law. *Science & Education*, 18(3):349–382, 2009.

[26] David W Knight. The self-resonance and self-capacitance of solenoid coils. *DW Knight*, 2013.

[27] F.A. Kolster. The effects of distributed capacity of coils used in radio-telegraphic circuits. *Proceedings of the Institute of Radio Engineers*, 1(2):19–27, 1913.

[28] Ken Kundert. Modeling skin effect in inductors. *The Designer's Guide*, 2006.

[29] Zhongjie Li, Yong Liu, Peilun Yin, Yan Peng, Jun Luo, Shaorong Xie, and Huayan Pu. Constituting abrupt magnetic flux density change for power density improvement in electromagnetic energy harvesting. *International Journal of Mechanical Sciences*, 198:106363, 2021.

[30] Tingting Lin, Kun Zhou, Yiming Cao, and Ling Wan. A review of air-core coil sensors in surface geophysical exploration. *Measurement*, 188:110554, 2022.

[31] Nocedal J. Liu, D.C. *On the limited memory BFGS method for large scale optimization*. Mathematical Programming 45, 1989.

[32] Daniel James Lizotte. Practical bayesian optimization. page 19, 2008.

[33] Richard McElreath. Statistical rethinking: A bayesian course with examples in r and stan. *CRC Press.*, 2016.

[34] Colonel William T McLyman. *Transformer and inductor design handbook*. Dekker New York, NY, USA, 1988.

[35] C.W.T. (2004) McLyman. *Transformer and Inductor Design Handbook, Third Edition (3rd ed.)*. CRC Press, 2004.

[36] KSS Kumar Moriconi, R and MP Deisenroth. High-dimensional bayesian optimization with projections using quantile gaussian processes. *arXiv preprint arXiv:1807.02811*, page 2, 2019.

[37] In Jae Myung. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1):90–100, 2003.

[38] Jorge. Nocedal and Stephen. Wright. *Numerical Optimization / by Jorge Nocedal, Stephen Wright. 2nd ed.* New York, NY: Springer New York, 2006.

[39] Szymon Wojciech Pasko, Marian K Kazimierczuk, and Bogusław Grzesik. Self-capacitance of coupled toroidal inductors for emi filters. *IEEE Transactions on Electromagnetic Compatibility*, 57(2):216–223, 2015.

[40] Charles Platt. *Electronics (Second edition)*. Maker Media, 2015.

[41] Ray G. Powell. *Introduction to electric circuits*. Arnold, 1995.

[42] Ray G. Powell. *Introduction to electric circuits*. Arnold, 1995.

[43] Edward M. Purcell and David J. Morin. *Electricity and Magnetism*. Cambridge University Press, 3 edition, 2013.

[44] Francis Sears. *University Physics (6th ed.)*. Addison Wesley, 1982.

[45] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

[46] James A Svoboda and Richard C Dorf. Introduction to electric circuits(3rd ed.). *Don Fowley*, page 3, 2013.

[47] Victor F. C. Veley. *The Benchtop Electronics Reference Manual (1st ed.)*. Blue Ridge Summit, PA : Tab Books, 1987.

[48] Eric-Jan Wagenmakers and Simon Farrell. Aic model selection using akaike weights. *Psychonomic bulletin & review*, 11:192–196, 2004.

[49] Hao Wang, Bas van Stein, Michael Emmerich, and Thomas Back. A new acquisition function for bayesian optimization based on the moment-generating function. pages 507–512, 2017.

[50] M.; Hutter F.; Matheson D.; de Freitas N. Wang, Z.; Zoghi. Bayesian optimization in high dimensions via random embeddings. *Palo Alto, Calif.: AAAI Press/International Joint Conferences on Artificial Intelligence*, page 1, 2013.

[51] Paul Wilmott. *The mathematics of financial derivatives*. Cambridge University Press.

[52] Hang Xu, Hiroharu Kamada, Shinichi Nomura, Hirotaka Chikaraishi, Hiroaki Tsutsui, and Takanori Isobe. A simple calculation method for center magnetic flux density of a magnetic core electromagnet with a wide air gap. *IEEE Transactions on Applied Superconductivity*, 32(6):1–6, 2022.

[53] Yachen Yan. rbayesianoptimization: Bayesian optimization of hyperparameters. `https://github.com/yanyachen/rBayesianOptimization`, 2022.