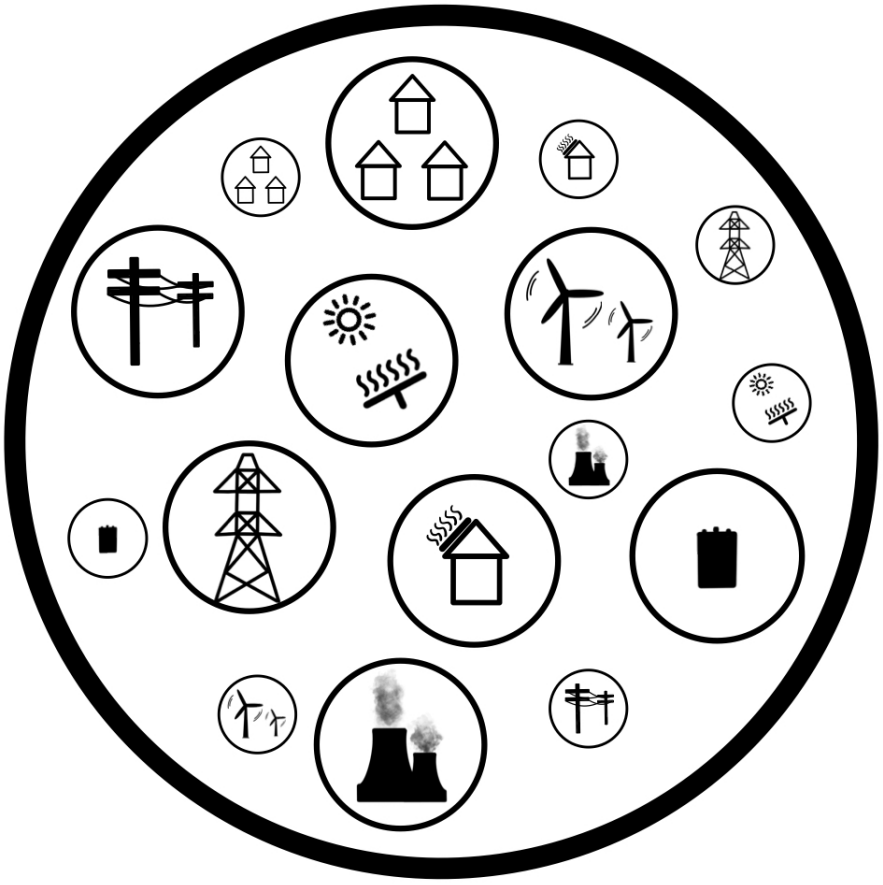
Estimation of Voltages and Currents in an Electric Grid

P5 Project
MatTek5 4.117c



Aalborg University
Mathematics-Technology

This report is written in the typesetting language LaTeX using text editor Overleaf. Experiments and data processing are conducted using Python 5.3.3. Illustrations are created using the TikZ package in LaTeX, Python 5.3.3 or drawn by hand. References are cited using the Vancouver method.

Department of Mathematical Sciences
Skjernvej 4A
DK - 9220 Aalborg
http://www.math.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Estimation of Voltages and Currents in an Electric Grid

Theme:
Statistical Analysis Applied in Technical Systems

Project Period:
Autumn Semester 2022

Project Group:
MatTek5 4.117c

Participants:
Julie Timmermann Werge
Christian Dausel Jensen
Stine Byrjalsen

Supervisors:
Jakob Gulddahl Rasmussen
Rasmus Løvenstein Olsen

Page Numbers: 61

Date of Completion:
December 16, 2022

Abstract:

In this project, the objective has been to derive a model that estimates voltages and currents in an electric distribution grid. This is desirable, as smart meters are expensive to be installed at all locations. To estimate the voltages and currents, a general linear model fulfilling Kirchhoff's and Ohm's laws was derived using the theories of circuit analysis, statistics, complex normal distribution, optimisation, and complex differentiation. The voltages and currents were simulated for a small and large tree. The experiments for the small tree proved the robustness and limits of our script. The large tree experiments showed that when removing a smart meter, its position contains an impact of how well the model estimates voltages and currents. It was also investigated that removing a smart meter at the initial node or terminating sibling nodes would crash the script due to the specifications. Lastly, it was possible to remove approximately half of the smart meters from each tree, without encountering a major change in the estimates.

Preface

This project is written by group 4.117c on the 5th semester 2022, studying Mathematics-Technology at the Department of Mathematical Sciences, Aalborg University. The project is written in the period September 2, 2022 - December 16, 2022. It is expected that the reader has some basic knowledge of linear algebra, calculus and probability. The simulation is written in Python using Networkx 2.8.4, Numpy 1.21.5, matplotlib.pyplot 3.5.3, Scipy 1.21.5 and Statsmodels 0.13.2, the code can be found in the attached zip file. The group wishes to thank our supervisors Jakob Gulddahl Rasmussen and Rasmus Løvenstein Olsen for their guidance and support throughout the project period.

Note that the first 4 chapters are written in collaboration with Tanja Kortsen Bugajski before the group chose to split up in October.

Aalborg University, December 16, 2022

Christian Dausel Jensen
cdje20@student.aau.dk

Julie Timmermann Werge
jwerge20@student.aau.dk

Stine Byrjalsen
sbyrja20@student.aau.dk

Nomenclature

Acronym	Description
KCL	Kirchhoff's Current Law
KVL	Kirchhoff's Voltage Law
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimate
LSE	Least Squares Estimate
GLM	General Linear Model
pdf	Probability Density Function
KKT	Karush-Kuhn-Tucker
SM	Smart Meter

Notation	Description
$\mathbf{v}$	Vector
$\mathbf{V}$	Stochastic vector
$\mathrm{Cov}[X, Y]$	Covariance matrix of X and Y
$\mathrm{Var}[\cdot]$	Variance
$\mathrm{E}[\cdot]$	Expected value
C^n	n-times continuously differentiable
O	Null-matrix
$\mathbb{1}$	Indicator function
$\xrightarrow{D}$	Tends to distribution
N_n	n-dimensional multivariate normal distribution
j	Complex number
$\overline{(\cdot)}$	Complex conjugate
$(\cdot)^T$	Transpose
$(\cdot)^H$	Conjugate transpose
$\underline{(\cdot)}$	Complex augmented vector
$\widetilde{\mathrm{Cov}}[\cdot, \cdot]$	Pseudo-covariance matrix
$N_{\mathbb{C}}$	Complex normal distribution
$(\cdot)^*$	Minimiser
F	Feasible region
$\triangleq$	Equal by definition

Contents

1 Introduction

In Denmark, electricity is used everyday. A functioning electric grid is essential to supply our households with the necessary electricity. We depend on the electricity production matching the electricity demand. However, the demand for electricity is increasing, which could potentially lead to overloads in the electric grid. Due to the green transition, Denmark's total electricity consumption is expected to double from 2019 to 2030 [8]. Households are replacing their old petrol cars with electric cars to reduce greenhouse gas emissions. Furthermore, gas furnaces are replaced with heat pumps running on electricity. Due to increasing gas prices because of the war in Ukraine, the aforementioned process is sped up, leading to massive loads on the electric grid [7]. Overload on the electric grid can lead to blackouts, which happened on Christmas Eve 2021 in Fredericia. The electric company Trefor explained that the blackout was probably due to 57 newly installed heat pumps in the area connected to the same phase [15]. To support the rapid development in power consumption, it is essential to be able to efficiently measure voltages and currents in an electric grid. This is accomplished with smart meters, however, these are expensive to install, and therefore it would be interesting to use statistics to estimate voltages and currents as a supplement. It would also be interesting to investigate if the placement of the smart meters in the grid has any influence on the estimates.

In order to present a method to estimate voltages and currents in an electric grid, some basic knowledge of the aforementioned grid and smart meters must be introduced.

1.1 Electric Grid

An electric power system is a network consisting of electrical components with the main purpose of supplying, transferring, and consuming electric power. An electric power system that supplies power to homes, commercial buildings, and industries is called an electric grid. This grid can be divided into three parts consisting of the power generation, transmission, and distribution system [3, p. 13]. These parts will be explained based on a traditional electric grid from 1985, as today's electric grid is more complicated. Figure 1.1 shows the structure of a traditional electric grid.

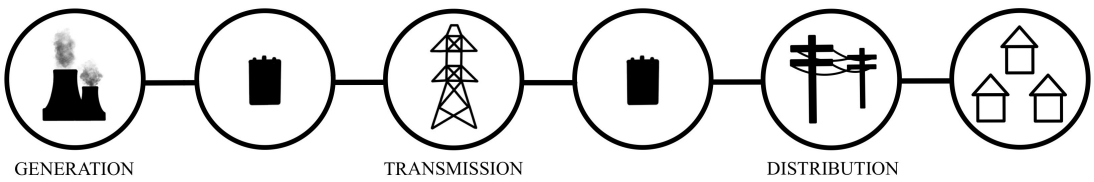


Figure 1.1: The structure of a traditional electric grid, where the link between generation, transmission and distribution are substations. The houses represent any type of customer, e.g. residential, commercial, or industrial.

Generation

In the traditional electric grid, the electricity was first produced and supplied by power generation stations. There are multiple types of power generation stations, depending on the energy source used to generate power. It can be wind turbines, solar panels, gas turbines, or nuclear power plants. The biggest source of power generation in the traditional electric grid in Denmark came from central power stations, which use fossil fuels such as coal, natural gas, or oil. However, renewable energy sources, such as wind or solar, began to contribute more.

The cost of power depends on the source used to generate it. It is the price of the last and most expensive power needed to meet the demand that determines the price. Therefore, the cheapest available power is consumed first. Usually this is power from wind turbines, solar cells, or hydroelectric power stations. If more power is needed, power from gas or coal power plants is used. However, these are more expensive. Therefore, it is beneficial to avoid overproduction of power to reduce costs [3, pp. 15–16].

Transmission

Since most power generation stations in the traditional electric grid were located away from populated areas, the electricity needed to be efficiently transported. Therefore, the next step is to transfer the electricity from the power generation stations to the distribution substations through high-voltage transmission lines. The transmission lines can be overhead, underground, or submarine, however, in Denmark, the transmission lines are mainly overhead and underground. Overhead lines are suspended from masts, thus impractical in urban areas, where underground cables are primarily used. Underground cables are more reliable than overhead lines since they are less exposed to the natural elements. However, underground cables are more expensive at high voltages due to loss and thereby cable heating, and more difficult to maintain than overhead lines due to the difficulty of locating faulty cables. Therefore, transmission networks are commonly overhead lines. In contrast, the distribution system is largely cabled underground [3, p. 73].

To reduce the energy loss that occurs by transmitting power over long distances, the voltage from the power generation stations is first increased at the substations. A substation is a location where the transmission lines are connected. The connections of various devices, such as transmission lines and transformers, are called buses. Transformers are used in substations to step the voltages up and down between facilities operating at different voltage levels. For instance, when the voltage power is transferred from the high-voltage transmission lines to the distribution substations, the voltage level is decreased [3, pp. 76–77].

In Denmark, the high-voltage level in the transmission network is between 50 kV to 400 kV. However, some large energy-demanding industries can be connected to high voltages between 132 kV to 150 kV with the option of reducing the voltages themselves [17].

Distribution

The distribution system is the last part of the electric grid. When the electric power is transferred from the transmission system, transformers at the distribution substations decrease the electric power to a more secure and practical voltage level before entering the distribution system. In Denmark, the distribution system voltage level ranges from 380 V to 50 kV [17]. The power is then delivered to consumers through power lines called feeders. Some consumers draw power directly from the transmission lines, but most are served from feeders. The customer usage is typically referred to as "the load". Primary feeders carry the voltages to the distribution transformers located near the premises of the customers. The voltages are here reduced to a low voltage level used by the consumers of the system. In Denmark, residential customers are connected to the low-voltage grid by a three-phase meter at 380 V. It is common that one transformer supplies several customers at the same time by using the secondary feeders [3, pp. 85–86].

1.2 Smart Grid

Since 1985, the electric grid has undergone an increasing development as the need for more electricity increases. This can be seen in [6], where the central power stations have largely been replaced by several small stations, called decentralised power stations. These stations use gas or alternative energy sources such as biomass or waste. More wind farms have also been added, and in general, the use of renewable energy has increased. Nowadays, the largest source of electricity in Denmark is renewable energy. However, Denmark is not self-sufficient, hence a part of the electricity supply also comes from other European countries [5]. Therefore, today's electric grid, also called smart grid, is more complicated than the traditional electric grid. This can be seen in Figure 1.2, however, notice that the figure is an abstraction of the real smart grid.

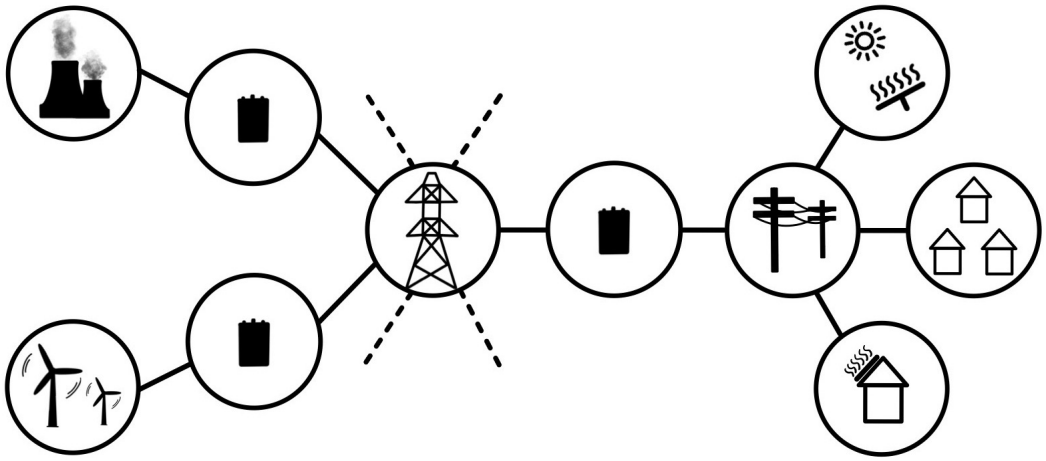


Figure 1.2: Structure of a modern electric grid. The dotted lines indicate several distribution networks connected to the transmission network.

As seen in the figure, the distribution of energy production is more scattered around, where even households produce electricity, e.g. with solar panels on the roofs. This means that the electricity does not only flow in one direction, as in the traditional electric grid, but that the consumers also supply electricity to the electric grid. The consumers are therefore no longer passive customers who only pay for the power they use but also participate in the electric grid. However, the high demand for electric power for additional loads such as electrification of transport, building heating, and the increasing use of renewable energy sources, which leads to the injection of power from the households to the grid, challenge the distribution system increasingly. It can result in an imbalance between production and demand, which can lead to power blackouts, increased electricity costs, or cause damage to connected equipment at the consumer's end. Therefore, it is important to match energy production with energy demand. Smart meters are a solution to the imbalance problem because they assist in the data-driven decisions that are necessary to sustain balance in the electric grid [11, pp. 8–9].

1.3 Smart Meters

A smart meter is an electronic device that communicates via radio waves and transmits the measurements to a server at an electric company. Compared to traditional meters, where you had to read the measurements manually, smart meters are capable of automatically providing near-real-time information about the consumption of electrical energy, voltage levels, currents, and powers in the electric power system. Therefore, smart meters enable two-way communication, which allows request and response interactions between the meter and the supplier. This makes it possible for the consumer to monitor their own electricity consumption, thereby adjusting their consumption. For the electricity company the smart meters are used to check that the electric supply network is up-to-date, so as to prevent that the electric network is not overloaded and that the losses in the network are as small as possible. The smart meters are therefore an important part of today's electric system since they are capable of providing information about supply and demand in the electric grid. This contributes to green energy being used when it is produced, thus reducing electricity costs [11, pp. 8–9].

This brings us to the scope of the project.

1.4 Problem Statement

How well can we estimate the voltages and currents of an electric grid?

2 Circuit Analysis

An electric grid can be described with a mathematical model. Circuit analysis is used for solving such a circuit by determining the voltage and current at each component. In an electric grid, the voltages and currents alternate with time, hence denoted $v(t)$ and $i(t)$, respectively, and follow a sinusoidal wave. Therefore, the voltage can be described as

$$v(t) = A\cos(2\pi ft + \theta),$$

where A is the amplitude of the voltage measured in meters, f is the frequency measured in Hertz, and θ is the phase angle measured in degrees. The voltage can be expressed in the complex form as

$$v(t) = \mathrm{Re}\left[\mathbf{V}_p e^{j(2\pi ft)}\right],$$

where the phasor $\mathbf{V}_p = Ae^{j\theta}$ is a vector in the complex plane with length A and angle θ. The phasor contains both the amplitude and the phase angle,

$$|\mathbf{V}_p| = A \quad \text{and} \quad \angle\mathbf{V}_p = \theta.$$

Identically, the current $i(t)$ can also be expressed using phasor notation $\mathbf{I}_p$ [9, pp. 328, 333–335].

2.1 Ohm's Law

A fundamental law of electric circuit analysis is Ohm's law.

Definition 2.1 (Ohm's Law)
The voltage across a resistor is directly proportional to the current flowing through it

$$v(t) = Z \cdot i(t), \quad Z \geq 0,$$

where the voltage $v(t)$ is measured in volt, the impedance Z is measured in ohm, and the current $i(t)$ is measured in amps. [18, p. 1071]

The impedance of an electric grid can be measured in the cable. A cable is a metallic conductor and is usually made of copper or aluminium. A non-idealised conductor contains opposition which is dependent on its own length, material, and area [18, pp. 851–853]. The impedance of a conductor is defined in Definition 2.2.

Definition 2.2 (Impedance of a Conductor)
The impedance of any particular conductor is related to the resistivity p of its material by

$$Z = \frac{pL}{A},$$

where p is measured in ohm-meter, L is the length of the conductor in meters, and A is the area of the uniform cross section measured in square meters. [18, p. 853].

Remark: The values p, L, and A are dependent on the temperature of the cable, which increases as a high current is running through it, therefore changing the value of Z. This impact will be neglected in this project.

2.2 Kirchhoff's Laws

Two additional fundamental laws of circuit analysis are Kirchhoff's laws, which explain the overall behaviour of a circuit composed of branches, nodes, and loops. The first law is Kirchhoff's current law (KCL).

Definition 2.3 (Kirchhoff's Current Law)
The algebraic sum of the currents entering any node can be expressed by

$$\sum_{j=1}^{N} i_j(t) = 0,$$

where N is the total amount of the j'th branches connected to a node. [18, p. 887]

The second law is Kirchhoff's voltage law (KVL).

Definition 2.4 (Kirchhoff's Voltage Law)
The algebraic sum of the voltages drop around any loop can be expressed by

$$\sum_{j=1}^{N} v_j(t) = 0,$$

where $v_j(t)$ is the voltage across the jth branch in a loop containing N voltages.

[18, p. 887]

The two definitions can be demonstrated with the following example.

Example 2.5 (Kirchhoff's Laws)
In Figure 2.1, a node and a loop is presented to demonstrate Kirchhoff's laws.

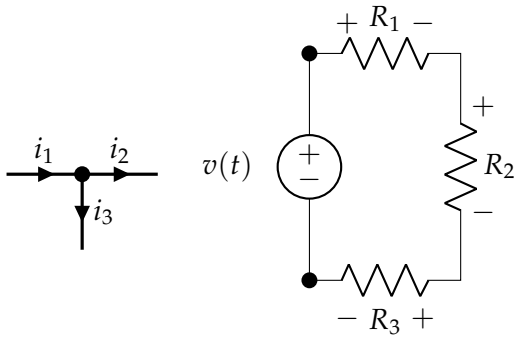


Figure 2.1: a) is a node with three branches connected, and b) is a circuit with 3 resistors and a voltage source.

In Figure 2.1a the current i_1 is entering the node and the two other currents are leaving, therefore by Definition 2.3 the sum of the currents can be calculated as

$$\sum_{j=1}^{3} i_j(t) = i_1 - i_2 - i_3 = 0.$$

If we want to know the value of a specific branch e.g. i_3

$$i_1 - i_2 = i_3.$$

In Figure 2.1b, the total voltage drop in the loop can be calculated by Definition 2.4

$$\sum_{j=1}^{4} v_j(t) = v_{R1}(t) + v_{R2}(t) + v_{R3}(t) - v(t) = 0,$$

such that

$$v_{R1}(t) + v_{R2}(t) + v_{R3}(t) = v(t).$$

The basic laws of a circuit can now be applied to describe an electric grid.

3 Statistics

This chapter introduces the neccessary statistics used for estimating the unknown voltages and currents in the electric grid.

3.1 Preliminaries

In order to prove some theorems in this chapter, certain preliminaries need to be introduced. Beginning by defining the covariance.

Definition 3.1 (Covariance)
Let X, Y be stochastic variables. The covariance of X and Y is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}\Big[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\Big].$$

[13, p. 197]

Notice that $\mathrm{Cov}[X, X] = \mathrm{Var}[X]$. The definition of covariance can also be extended to stochastic vectors.

Definition 3.2 (Covariance Matrix)
Let $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^T$ be a stochastic vector. Then the covariance matrix of $\mathbf{X}$ is defined as

$$\mathrm{Cov}[\mathbf{X}, \mathbf{X}] = \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_2, X_1] & \mathrm{Var}[X_2] & \cdots & \mathrm{Cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_n, X_1] & \mathrm{Cov}[X_n, X_2] & \cdots & \mathrm{Var}[X_n] \end{bmatrix}.$$

[10, p. 258]

Next, a property of some matrices is defined.

Definition 3.3 (Positive Semi-Definite Matrix)
Let $\mathbf{x} \in \mathbb{R}^n$ be non-zero and let $A \in \mathbb{R}^{n \times n}$. Then A is said to be positive semi-definite if

$$\mathbf{x}^T A \mathbf{x} \geq 0.$$

This is denoted $A \geq 0$. [10, p. 258]

It then follows that any covariance matrix is positive semi-definite.

Theorem 3.4 (Covariance Matrix is Positive Semi-Definite)
For a stochastic vector $\mathbf{X} \in \mathbb{R}^n$ the covariance matrix is positive semi-definite.
[10, p. 258]

The proof can be found in [10, pp. 258–259].

With the necessary concepts defined, we can introduce the theory of parameter estimation called the likelihood principle.

3.2 The Likelihood Principle

Given a data set $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$, a parametric statistical model with stochastic variable $\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix}^T$ is defined by a joint density $f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta})$, which depends on a real parameter $\boldsymbol{\theta}$ in the parameter space $\Theta^k \subseteq \mathbb{R}^k$. The parameter $\boldsymbol{\theta}$ is usually unknown and therefore we want to estimate it. To achieve this we first want to find the estimator $\hat{\boldsymbol{\theta}}(\mathbf{Y})$, for which we get an estimate $\hat{\boldsymbol{\theta}}(\mathbf{y})$ of the parameter $\boldsymbol{\theta}$. Notice that the estimate is a function of some data $\mathbf{y}$ and gives a specific number, whereas the estimator is a stochastic variable [12, pp. 10–11]. To obtain a suitable estimator close to the parameter, certain properties of the estimator will be defined.

Definition 3.5 (Unbiased Estimator)
An estimator $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is unbiased if $\mathrm{E}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] = \boldsymbol{\theta}, \quad \forall \boldsymbol{\theta} \in \Theta^k$. [12, p. 11]

Definition 3.6 (Consistent Estimator)
An estimator $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is consistent if

$$\mathrm{P}\left(\left\|\hat{\boldsymbol{\theta}}_n(\mathbf{Y}) - \boldsymbol{\theta}\right\| > \epsilon\right) \xrightarrow[n\to\infty]{} 0, \quad \forall\, \boldsymbol{\theta} \in \Theta^k,\ \epsilon > 0,$$

where $\hat{\boldsymbol{\theta}}_n(\mathbf{Y})$ is the sequence of estimators for $n \in \mathbb{N}$ observations. [12, p. 11]

For a consistent estimator a more accurate estimation is expected for $n \to \infty$.

When establishing an estimator, the variance determines the efficiency of said estimator. A low variance results in an efficient estimator and is therefore desirable. In order to evaluate the variance, a lower bound for the covariance matrix is needed. However, before we can define this property, we introduce the likelihood function.

Definition 3.7 (Likelihood Function)
Let $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$ be a parametric density, where $\boldsymbol{\theta} \in \Theta^k$, for the data $\mathbf{y}$. Then the likelihood function for $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta};\mathbf{y}) = c(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}),$$

where $c(\mathbf{y}) > 0$ does not depend on $\boldsymbol{\theta}$. [12, p. 14]

The likelihood function is only meaningful up to an arbitrary constant, therefore, terms not involving the parameter can be ignored. It is often more convenient to work with the log-likelihood function, which is given as

$$\ell(\boldsymbol{\theta};\mathbf{y}) = \log L(\boldsymbol{\theta};\mathbf{y}). \tag{3.1}$$

The first derivative of the log-likelihood function is called the score function.

Definition 3.8 (Score Function)
Let $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \theta_2 & \cdots & \theta_k\end{bmatrix}^T \in \Theta^k$. Assume that Θ^k is an open subspace of $\mathbb{R}^k$ and that $\ell(\boldsymbol{\theta};\mathbf{y}) \in C^1$. Then the function

$$S(\boldsymbol{\theta};\mathbf{y}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\theta};\mathbf{y}) = \begin{bmatrix} \frac{\partial}{\partial \theta_1}\ell(\boldsymbol{\theta};\mathbf{y}) \\ \vdots \\ \frac{\partial}{\partial \theta_k}\ell(\boldsymbol{\theta};\mathbf{y}) \end{bmatrix}$$

is called the score function. [12, p. 17]

The score function has the following property.

Theorem 3.9
Under the regularity condition of interchanging derivatives and integration, the following holds

$$\mathrm{E}[S(\boldsymbol{\theta};\mathbf{Y})] = \mathbf{0}.$$

[12, p. 18]

Proof
First the mean of the score function is calculated

$$\mathrm{E}[S(\boldsymbol{\theta};\mathbf{Y})] = \int \frac{\partial}{\partial \boldsymbol{\theta}^T} \left(\log f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})\right) f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \mathrm{d}y,$$

using the chain rule, we obtain

$$= \int \frac{\partial}{\partial \boldsymbol{\theta}^T} f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \mathrm{d}y.$$

The operations of integration and differentiation are interchanged, which yields

$$\begin{aligned}\mathrm{E}[S(\boldsymbol{\theta};\mathbf{Y})] &= \frac{\partial}{\partial \boldsymbol{\theta}^T}\int f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})\mathrm{d}y \\ &= \frac{\partial}{\partial \boldsymbol{\theta}^T}1 \\ &= \mathbf{0}.\end{aligned}$$

■

The second derivative of the log-likelihood function is defined as follows.

Definition 3.10 (Information Matrix)
Let $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \theta_2 & \cdots & \theta_k\end{bmatrix}^T \in \Theta^k$. Assume that Θ^k is an open subspace of $\mathbb{R}^k$ and that $\ell(\boldsymbol{\theta};\mathbf{y}) \in C^2$. Then the observed information matrix is given as

$$j(\boldsymbol{\theta};\mathbf{y}) = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\ell(\boldsymbol{\theta};\mathbf{y}),$$

and the expected information matrix is given as

$$i(\boldsymbol{\theta}) = \mathrm{E}[j(\boldsymbol{\theta};\mathbf{Y})].$$

[12, p. 18]

The expected information matrix can under some regularity conditions be written as

$$i(\boldsymbol{\theta}) = \mathrm{E}\left[S(\boldsymbol{\theta};\mathbf{Y})\left(S(\boldsymbol{\theta};\mathbf{Y})\right)^T\right],$$

which is called the Fisher information matrix.

Using the above concepts the lower bound is given by the Cramer-Rao inequality.

Theorem 3.11 (Cramer-Rao Inequality)
Let $f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$ be a parametric density with continuous parameter $\boldsymbol{\theta}$, where $\boldsymbol{\theta} \in \Theta^k$. Then the covariance matrix of any unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ of $\boldsymbol{\theta}$ satisfies the following inequality

$$\mathrm{Var}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] \geq i^{-1}(\boldsymbol{\theta}),$$

where $i(\boldsymbol{\theta})$ is the Fisher information matrix. [12, pp. 12–13]

Proof
Assume that the stochastic vector $\mathbf{Y}$ is continuous. Since $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is unbiased, then

$$\mathrm{E}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] = \int \hat{\boldsymbol{\theta}}(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \mathrm{d}y = \boldsymbol{\theta}.$$

First, the following mean is calculated

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y}) \frac{\partial}{\partial \boldsymbol{\theta}^T} \log f_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\theta})\right] = \int \hat{\boldsymbol{\theta}}(\mathbf{y}) \frac{\partial}{\partial \boldsymbol{\theta}^T} \left(\log f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})\right) f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \mathrm{d}y,$$

using the chain rule, we obtain

$$= \int \hat{\boldsymbol{\theta}}(\mathbf{y}) \frac{\partial}{\partial \boldsymbol{\theta}^T} f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \mathrm{d}y.$$

The operations of integration with respect to $\mathbf{y}$ and differentiation with respect to $\boldsymbol{\theta}$ can be interchanged, thus

$$\begin{aligned}\mathrm{E}\left[\hat{\boldsymbol{\theta}}(\mathbf{Y}) \frac{\partial}{\partial \boldsymbol{\theta}^T} \log f_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\theta})\right] &= \frac{\partial}{\partial \boldsymbol{\theta}^T} \int \hat{\boldsymbol{\theta}}(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \mathrm{d}y \\ &= \frac{\partial}{\partial \boldsymbol{\theta}^T} \mathrm{E}[\hat{\boldsymbol{\theta}}(\mathbf{Y})],\end{aligned}$$

using that the estimator is unbiased, we get

$$\begin{aligned}&= \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\theta} \\ &= I.\end{aligned}$$

Here I is the $k \times k$ identity matrix. Additionally,

$$\mathrm{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}^T} \log f_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\theta})\right] = \mathbf{0}^T,$$

as concluded in Theorem 3.9. Here $\mathbf{0}^T$ is the transpose of the $k-$dimensional null-vector. Next, the covariance matrix is found

$$\begin{aligned}\mathrm{Var}\begin{bmatrix}\hat{\boldsymbol{\theta}}(\mathbf{Y}) \\ \frac{\partial}{\partial \boldsymbol{\theta}^T} \log f_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\theta})\end{bmatrix} &= \mathrm{E}\left[\begin{bmatrix}\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta} \\ \frac{\partial}{\partial \boldsymbol{\theta}^T} \log f_{\mathbf{Y}}(\mathbf{Y};\boldsymbol{\theta})\end{bmatrix} \left[(\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta})^T \left(\frac{\partial}{\partial \boldsymbol{\theta}^T} \log f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})\right)^T\right]\right] \\ &= \begin{bmatrix}\mathrm{Var}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] & I \\ I & i(\boldsymbol{\theta})\end{bmatrix}.\end{aligned}$$

Since this is a covariance matrix, it is positive semi-definite by Theorem 3.4. This means that

$$\begin{bmatrix}\mathrm{Var}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] & I \\ I & i(\boldsymbol{\theta})\end{bmatrix} \geq O,$$

where $O \in \mathbb{R}^{2k \times 2k}$ is a null-matrix. If a matrix is positive semi-definite, the definition for positive semi-definite matrices can be used together with the condition that the Fisher information matrix $i(\boldsymbol{\theta})$ is invertible. This yields

$$\begin{aligned} O &\leq \begin{bmatrix} I & -i^{-1}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \mathrm{Var}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] & I \\ I & i(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} I \\ -i^{-1}(\boldsymbol{\theta}) \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Var}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] - i^{-1}(\boldsymbol{\theta}) & 0 \end{bmatrix} \begin{bmatrix} I \\ -i^{-1}(\boldsymbol{\theta}) \end{bmatrix} \\ &= \mathrm{Var}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] - i^{-1}(\boldsymbol{\theta}). \end{aligned}$$

■

The proof for the discrete case is obtained by changing the integrals to summations. The Cramer-Rao lower bound leads to the property of an efficient estimator.

Definition 3.12 (Efficient Estimator)
An unbiased estimator is efficient if its covariance matrix is equal to the Cramer-Rao lower bound. [12, p. 13]

Now with the properties of the estimator defined, we will introduce a method to estimate the parameters. This method is called maximum likelihood estimation.

3.3 Maximum Likelihood Estimation

Maximum likelihood estimation estimates the parameter for a given distribution by maximising the likelihood function. The point in the parameter space that maximises the likelihood function is called the maximum likelihood estimate (MLE), and the estimator estimating the parameter is called the maximum likelihood estimator (ML estimator).

Definition 3.13 (Maximum Likelihood Estimation)
Given the data $\mathbf{y} = \begin{bmatrix} y_1 & y_1 & \cdots & y_n \end{bmatrix}^T$ the maximum likelihood estimate $\hat{\boldsymbol{\theta}}(\mathbf{y})$ exists if $L(\boldsymbol{\theta}; \mathbf{y})$ has a unique maximum, which satisfies

$$L(\hat{\boldsymbol{\theta}}; \mathbf{y}) = \sup_{\boldsymbol{\theta} \in \Theta^k} L(\boldsymbol{\theta}; \mathbf{y}).$$

If $\mathrm{P}(\mathbf{Y} \in \{\mathbf{y} : \hat{\boldsymbol{\theta}}(\mathbf{y}) \text{ exists}\}) = 1$, $\boldsymbol{\theta} \in \Theta^k$, then $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is called the maximum likelihood estimator. [12, p. 21]

The MLE can be found by solving the equation

$$S(\boldsymbol{\theta}; \mathbf{y}) = \mathbf{0}, \tag{3.2}$$

assuming the supremum is found in the interior of Θ^k. To check if the solution to (3.2) is in fact a maximum point, the observed information matrix is considered. If (3.2) has only one solution such that

$$j(\boldsymbol{\theta}; \mathbf{y}) \leq O,$$

where $O \in \mathbb{R}^{k \times k}$, then $\hat{\boldsymbol{\theta}}(\mathbf{y})$ is the MLE. Otherwise the point is a minimum or a saddle point. The maximum likelihood estimation is illustrated in the following example.

Example 3.14 (Maximum Likelihood Estimation)
An unfair coin is flipped 13 times and lands on heads 9 times. What is the probability of the coin showing heads, θ? This is a binomial experiment with independent observations, therefore $n = 13$ and $y = 9$. The likelihood function can be derived from Definition 3.7

$$L(\theta; y) = f_Y(y; \theta) = P_\theta(Y = y) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

We find that $\binom{13}{9} = 330$, hence evaluating the likelihood function for several values of θ yields Table 3.1.

θ	Likelihood
0.5	0.087
0.6	0.18
0.7	0.23
0.8	0.15

Table 3.1

The value $\theta = 0.7$ is far more likely than any other θ, but this is a decimal limited guess. To obtain a better approximation we will find the maximum of the likelihood function. As stated in (3.1) it is more convenient to work with the log-likelihood function, thus we have

$$\ell(\theta; y) = \log L(\theta; y) = y \log \theta + (n - y) \log(1 - \theta) + c,$$

where c is a term that does not depend on θ. Having obtained $\ell(\theta; y)$, the MLE can be found using (3.2)

$$S(\theta; y) = \frac{y}{\theta} - \frac{n-y}{1-\theta} = 0.$$

Thus,

$$\begin{aligned}
\frac{y}{\theta} - \frac{n-y}{1-\theta} &= 0 \\
\frac{y}{\theta} &= \frac{n-y}{1-\theta} \\
n\theta - y\theta &= y - \theta y \\
n\theta &= y \\
\hat{\theta}(y) &= \frac{y}{n}.
\end{aligned}$$

The MLE can now be calculated

$$\hat{\theta}(y) = \frac{9}{13} = 0.69.$$

Hence our logical guess ($\theta = 0.7$) was very close to the MLE of θ. The MLE is plotted in Figure 3.1.

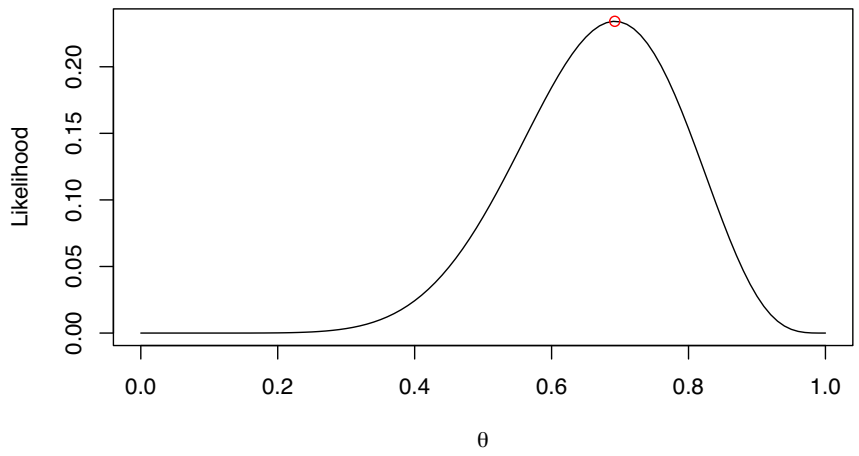


Figure 3.1: Likelihood function of the success probability θ in a binomial experiment, with $n = 13$, $y = 9$, and $\hat{\theta} = 0.69$ as the red circle.

The following theorem gives some properties of the ML estimator.

Theorem 3.15 (Distribution of the Maximum Likelihood Estimator)
Let $E = \{\mathbf{y} : \hat{\boldsymbol{\theta}}(\mathbf{y}) \text{ exists}\}$ and I the $k \times k$ identity matrix. Then, under certain regularity conditions the following holds for $n \to \infty$.

a) $\mathrm{P}(\mathbf{Y} \in E) \to 1$.

b) $\mathrm{P}(\mathbf{Y} \in E, \left\|\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta}\right\| < \epsilon) \to 1$, for any $\epsilon > 0$.

c) $\mathbb{1}[\mathbf{Y} \in E] i(\hat{\boldsymbol{\theta}}(\mathbf{Y}))^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_k(0, I)$ and
$\mathbb{1}[\mathbf{Y} \in E] j(\hat{\boldsymbol{\theta}}(\mathbf{Y}); \mathbf{Y})^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_k(0, I)$.

d) $\mathbb{1}[\mathbf{Y} \in E] i(\boldsymbol{\theta})^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_k(0, I)$.

[12, pp. 22–23]

The proof of this theorem will not be examined, since it is beyond the scope of this project. Property (b) is also referred to as asymptotic consistency of the ML estimator. Property (c) is a result most often used in practice, while property (d) is the theoretical result.

Further analysis of property (d) yields the distribution of the ML estimator. Assume that $\mathbf{Y} \in E$, so the MLE exists. Then for large n,

$$\begin{aligned} i(\boldsymbol{\theta})^{\frac{1}{2}} (\hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta}) &\approx N_k(0, I) \\ \hat{\boldsymbol{\theta}}(\mathbf{Y}) - \boldsymbol{\theta} &\approx N_k\left(0, i(\boldsymbol{\theta})^{-\frac{1}{2}} \; I \; (i(\boldsymbol{\theta})^{-\frac{1}{2}})^T\right) \\ \hat{\boldsymbol{\theta}}(\mathbf{Y}) &\approx N_k\left(\boldsymbol{\theta}, i(\boldsymbol{\theta})^{-1}\right). \end{aligned}$$

Property (c) can similarly be expressed as

$$\hat{\boldsymbol{\theta}}(\mathbf{Y}) \approx N_k\left(\boldsymbol{\theta}, i(\hat{\boldsymbol{\theta}}(\mathbf{Y}))^{-1}\right), \quad \hat{\boldsymbol{\theta}}(\mathbf{Y}) \approx N_k\left(\boldsymbol{\theta}, j(\hat{\boldsymbol{\theta}}(\mathbf{Y}); \mathbf{Y})^{-1}\right).$$

Hence, $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ is asymptotically unbiased and efficient since

$$\mathrm{E}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] \to \boldsymbol{\theta} \;\text{ and }\; \mathrm{Var}[\hat{\boldsymbol{\theta}}(\mathbf{Y})] \to i(\boldsymbol{\theta})^{-1}.$$

After a parameter has been estimated using the maximum likelihood method, we wish to test whether a given assumption about the parameter holds.

3.4 Hypothesis Tests and Confidence Regions

In hypothesis tests an assumption is called the null hypothesis, defined as

$$H_0 : \boldsymbol{\theta} \in \Theta_0,$$

where $\Theta_0 \subset \Theta^k$. An alternative hypothesis, defined as

$$H_1 : \boldsymbol{\theta} \in \Theta^k \backslash \Theta_0,$$

will be accepted in the case where the null hypothesis is rejected [12, p. 25]. To compare the two hypothesis, the test statistic is introduced.

Definition 3.16 (Test Statistics)
Let $\mathbf{Y}$ be a stochastic variable and $T(\mathbf{Y})$ a function with known distribution under H_0. Then $T(\mathbf{Y})$ is the test statistic and $T(\mathbf{y})$ is the observed test statistic. [13, p. 329]

As seen in Figure 3.2, the more extreme $T(\mathbf{y})$ is, the less likely it is that the data could have been generated under the null hypothesis.

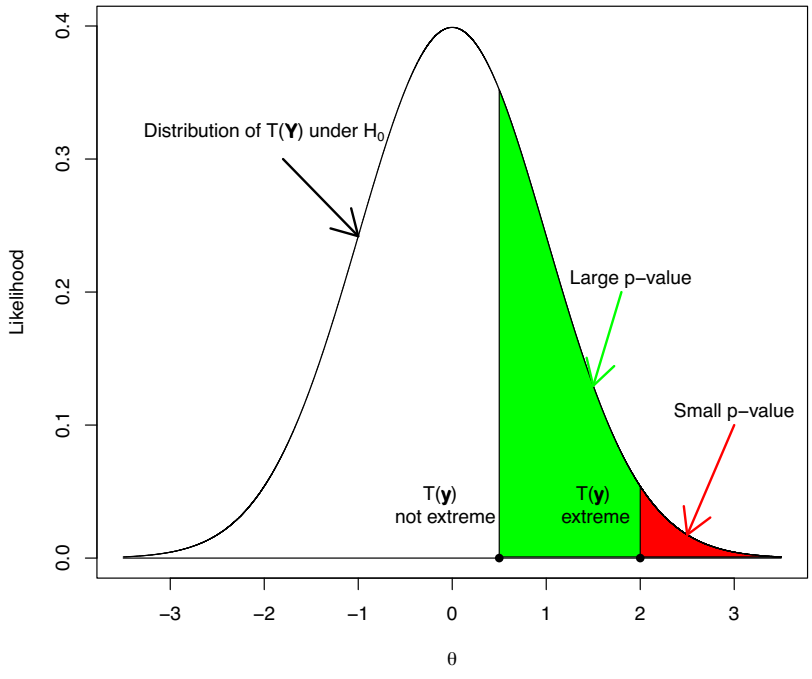


Figure 3.2: Illustration of the test statistic and the p-value.

The test statistics are used to calculate the p-value, which can be used to decide whether to reject or accept the null hypothesis. The p-value is the probability of observing a test

statistic that is more extreme than the observed test statistic $T(\mathbf{y})$, assuming H_0 is true. If the p-value is small, the less likely it is that the test statistic occurred under H_0, and the null hypothesis can be rejected. If the p-value is large, H_0 can be accepted. This is also illustrated in Figure 3.2. However, H_0 can be rejected despite actually being true. This is a type 1 error. Keeping H_0 when H_1 is true is called a type 2 error. The probability for a type 1 error is called the significance level, noted α. This level gives a boundary for small and large p-values. To reject the null hypothesis, the p-value must be less than the significance level, i.e. $p < \alpha$. Typically, α is chosen to be 0.05, though it can be any value between 0 and 1 [12, pp. 25–26].

After estimating a parameter, it can be relevant to see how confident we are in the chosen estimate. Therefore, another way to either reject or accept the null hypothesis is by using confidence regions and intervals.

Definition 3.17 (Confidence Regions)
Let $\alpha \in (0,1)$. Then a $(1-\alpha)-$ confidence region is a subset $A(\mathbf{y}) \subseteq \Theta^k$ that satisfies

$$\mathrm{P}(\boldsymbol{\theta} \in A(\mathbf{Y})) = 1-\alpha, \quad \forall\, \boldsymbol{\theta} \in \Theta^k.$$

[13, pp. 304–305]

Note that for $k = 1$, $A(\mathbf{y})$ is called a confidence interval. For $\alpha = 0.05$, $A(\mathbf{y})$ is a 95%-confidence region (or interval if $k = 1$). The limits of the confidence region is called the critical region. If the observed test statistic is in the critical region, then H_0 is rejected, whereas the hypothesis is accepted if $T(\mathbf{y})$ is in the confidence region [12, p. 26].

To find an approximate $(1-\alpha)$-confidence interval, the following theorem is introduced.

Theorem 3.18 (Confidence Interval Based on Asymptotics for MLE)
Let $\alpha \in (0,1)$ and let $A(\mathbf{y}) \subseteq \Theta^1$. Then for $\hat{\boldsymbol{\theta}}(\mathbf{Y}) \approx N_k\left(\boldsymbol{\theta}, i(\hat{\boldsymbol{\theta}}(\mathbf{Y}))^{-1}\right)$ the $(1-\alpha)$-confidence interval is given as

$$\boldsymbol{\theta} \in \left[\hat{\boldsymbol{\theta}}(\mathbf{y}) + \Phi_{\alpha/2}\sqrt{i(\hat{\boldsymbol{\theta}}(\mathbf{y}))^{-1}}, \hat{\boldsymbol{\theta}}(\mathbf{y}) + \Phi_{1-\alpha/2}\sqrt{i(\hat{\boldsymbol{\theta}}(\mathbf{y}))^{-1}}\right].$$

[13, p. 308]

If $\alpha = 0.05$, then

$$\boldsymbol{\theta} \in \left[\hat{\boldsymbol{\theta}}(\mathbf{y}) - 1.96\sqrt{i(\hat{\boldsymbol{\theta}}(\mathbf{y}))^{-1}}, \hat{\boldsymbol{\theta}}(\mathbf{y}) + 1.96\sqrt{i(\hat{\boldsymbol{\theta}}(\mathbf{y}))^{-1}}\right]. \tag{3.3}$$

Theorem 3.18 is illustrated in the following example.

Example 3.19 (Confidence Interval Based on Asymptotics for MLE)
This example will use the data from Example 3.14, hence $Y \sim \text{Binom}(n, \theta)$, $n = 13$, $y = 9$, and $\hat{\theta}(y) = 0.69$. We must find the asymptotic distribution of $\hat{\theta}(y)$ since the confidence intervals are based on the asymptotics of the MLE. By Theorem 3.15 when $n \to \infty$ and θ is fixed, the distribution tends to be the normal distribution $N\left(\theta, i(\hat{\theta}(y))^{-1}\right)$. We can find the expected value and the variance, using the fact that the estimator is unbiased, hence

$$\mathrm{E}\left[\frac{Y}{n}\right] = \frac{\mathrm{E}[Y]}{n} = \frac{n\theta}{n} = \theta,$$

and

$$\mathrm{Var}\left[\hat{\theta}(Y)\right] = \mathrm{Var}\left[\frac{Y}{n}\right] = \frac{\mathrm{Var}[Y]}{n^2} = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n} = i(\hat{\theta}(y))^{-1}.$$

We have $\hat{\theta}(Y) \sim N(\theta, \frac{\theta(1-\theta)}{n})$. Next, we will find the confidence interval by (3.3), by using $i(\hat{\theta}(y))^{-1}$ to find the asymptotic variance of the ML estimator

$$i^{-1}(\hat{\theta}(y)) = \frac{0.69(1-0.69)}{13} = 0.016.$$

The confidence interval can now be calculated by remembering $\hat{\theta}(y)$ from Example 3.14,

$$\frac{9}{13} - 1.96\sqrt{0.016} = 0.44,$$
$$\frac{9}{13} + 1.96\sqrt{0.016} = 0.94,$$

therefore $\theta \in [0.4414; 0.9432]$. The interval is quite large due to the small sample size, hence having a larger amount of samples the interval would be smaller.

To estimate the voltages and currents in the electric grid, it is convenient to also introduce general linear models.

4 General Linear Models

In statistics a general linear model (GLM) is used to describe variations of experiments with continuous variables.

4.1 Preliminaries

In this section some necessary definitions for working with GLMs is presented. Starting with the multivariate normal distribution is defined.

Definition 4.1 (Multivariate Normal Distribution)
Let $\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix}^T$ be a stochastic vector with $Y_i \overset{i.i.d.}{\sim} N(0,1)$ for $i = 1, \ldots, n$. If Z has the same distribution as $A\mathbf{Y} + \mathbf{b}$ for some n, some $k \times n$ matrix A, and some k-dimensional vector $\mathbf{b}$, then Z is said to have a k-dimensional multivariate normal distribution. The multivariate normal distribution is written as

$$Z \sim N_k\left(\mathbf{b}, AA^T\right).$$

[12, p. 42]

Observe from Definition 4.1 that $\mathrm{E}[Z] = \mathbf{b}$ and $\mathrm{Var}[Z] = AA^T$. Assuming that the covariance matrix is known apart from a constant factor σ^2, a symmetric and positive definite weight matrix Σ is used. Resulting in $\mathrm{Var}[Z] = \sigma^2\Sigma$. Note that if $\Sigma = I$, the case is considered an unweighted problem. Having introduced the matrix Σ it is relevant to define the inner product and norm.

Definition 4.2 (Inner Product and Norm)
Let $\mathbf{y_1}, \mathbf{y_2} \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ be symmetric and positive definite, then

$$\delta_\Sigma(\mathbf{y_1}, \mathbf{y_2}) = \mathbf{y_1}^T\Sigma^{-1}\mathbf{y_2}$$

is an inner product between $\mathbf{y_1}$ and $\mathbf{y_2}$ with respect to Σ. If

$$\delta_\Sigma(\mathbf{y_1}, \mathbf{y_2}) = 0,$$

$\mathbf{y_1}$ and $\mathbf{y_2}$ are orthogonal. The norm is defined as

$$\|\mathbf{y}\|_\Sigma = \sqrt{\delta_\Sigma(\mathbf{y}, \mathbf{y})}.$$

[12, p. 43]

Another useful definition for normal distributed variables is the deviance.

Definition 4.3 (Deviance for Normal Distributed Variables)
The quadratic norm of the vector $(\mathbf{y} - \boldsymbol{\mu})$ corresponding to the inner product defined by Σ^{-1} is defined as the deviance

$$D(\mathbf{y}; \boldsymbol{\mu}) = \|(\mathbf{y} - \boldsymbol{\mu})\|_{\Sigma}^2 = (\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}).$$

[12, p. 43]

Having introduced some fundamental definitions for working with GLMs, the next section will focus on statistical models.

4.2 Statistical Models

A statistical model is used to describe the variation of the observations mathematically. Consider the n observations $\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix}^T$ and assume that

$$\mathbf{Y} \sim N_n \left(\boldsymbol{\mu}, \sigma^2 \Sigma \right) \tag{4.1}$$

is a statistical model. The likelihood function for (4.1) is

$$L(\boldsymbol{\mu}, \sigma^2; \mathbf{y}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n \sqrt{\det(\Sigma)}} \mathrm{e}^{-\frac{1}{2\sigma^2} D(\mathbf{y}; \boldsymbol{\mu})},$$

where $D(\mathbf{y}; \boldsymbol{\mu})$ is the deviance as defined in Definition 4.3. Disregarding the constants, the log-likelihood function is obtained as

$$\ell(\boldsymbol{\mu}, \sigma^2; \mathbf{y}) = -\left(\frac{n}{2}\right) \log\left(\sigma^2\right) - \frac{1}{2\sigma^2} D(\mathbf{y}; \boldsymbol{\mu}).$$

The score function with respect to $\boldsymbol{\mu}$ is

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ell(\boldsymbol{\mu}, \sigma^2; \mathbf{y}) = \frac{1}{\sigma^2} \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}), \tag{4.2}$$

and the observed information with respect to $\boldsymbol{\mu}$ is

$$j(\boldsymbol{\mu}; \mathbf{y}) = \frac{1}{\sigma^2} \Sigma^{-1}.$$

Since the observed information does not depend on the observations $\mathbf{y}$, the expected information is

$$i(\boldsymbol{\mu}) = \frac{1}{\sigma^2} \Sigma^{-1}.$$

Depending on whether $\boldsymbol{\mu}$ varies freely in $\mathbb{R}^n$ or $\mathbb{R}$ it is either defined as a full model or a null model.

Definition 4.4 (The Full Model)
In the full model $\boldsymbol{\mu}$ varies freely in $\mathbb{R}^n$ and $\hat{\boldsymbol{\mu}} = \mathbf{y}$. [12, p. 44]

The null model is more restricted than the full model, since it only describes the variations of the observations by a common mean value for all observations.

Definition 4.5 (The Null Model)
In the null model $\boldsymbol{\mu}$ varies in a one dimensional subspace of $\mathbb{R}^n$ spanned by the vector $\begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T$ and the dimension of the model-subspace is one.

[12, p. 46]

Typically the variations is described by a model somewhere between the full and the null model. Consider a model for the n observations $\begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix}^T$

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim N_n\left(0, \sigma^2\Sigma\right)$. A GLM for $\mathbf{Y}$ is then defined.

Definition 4.6 (General Linear Model)
Assume $\mathbf{Y} \sim N_n\left(\boldsymbol{\mu}, \sigma^2\Sigma\right)$. If $\boldsymbol{\mu} \in \Omega_0$, where Ω_0 is a linear subspace of $\mathbb{R}^n$, then $\mathbf{Y}$ is called a general linear model with dimension $\dim(\Omega_0)$. [12, p. 46]

In connection with a GLM, $\boldsymbol{\mu}$ can be rewritten by using a design matrix.

Definition 4.7 (Design Matrix)
Assume $\Omega_0 = \text{span}\{x_1, \dots, x_k\}$ where $k \leq n$ and let X be a $n \times k$ full rank matrix with columns $\mathbf{x}_1, \dots, \mathbf{x}_k$. The matrix X is called the design matrix with column space Ω_0 and

$$\boldsymbol{\mu} = X\boldsymbol{\beta},$$

where $\boldsymbol{\beta} \in \mathbb{R}^k$ is called the parameter-vector. [12, pp. 46–47]

The GLM can now be written as

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{4.3}$$

The MLE for a GLM is obtained by the following theorem.

Theorem 4.8 (Maximum Likelihood Estimates of Mean Value Parameters)
For a general linear model with $\boldsymbol{\mu} = X\boldsymbol{\beta}$ the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ is a solution to the normal equation

$$X^T\Sigma^{-1}\mathbf{y} = X^T\Sigma^{-1}X\hat{\boldsymbol{\beta}}.$$

If X has full rank, the solution is uniquely given by

$$\hat{\boldsymbol{\beta}} = \left(X^T\Sigma^{-1}X\right)^{-1}X^T\Sigma^{-1}\mathbf{y}.$$

[12, pp. 48–49]

Proof
Solve $S(\boldsymbol{\beta};\mathbf{y}) = \mathbf{0}$, where $\boldsymbol{\mu}(\boldsymbol{\beta}) = X\boldsymbol{\beta}$. Using the chain rule we obtain

$$S(\boldsymbol{\beta};\mathbf{y}) = \left(\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\mu}(\boldsymbol{\beta})\right)^T S\left(\boldsymbol{\mu}(\boldsymbol{\beta});\mathbf{y}\right).$$

From (4.2) the expression yields

$$\begin{aligned}
&= \left(\frac{\partial}{\partial\boldsymbol{\beta}}X\boldsymbol{\beta}\right)^T\frac{1}{\sigma^2}\Sigma^{-1}(\mathbf{y} - X\boldsymbol{\beta}) \\
&= X^T\frac{1}{\sigma^2}\Sigma^{-1}(\mathbf{y} - X\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\left(X^T\Sigma^{-1}\mathbf{y} - X^T\Sigma^{-1}X\boldsymbol{\beta}\right) \\
&= \mathbf{0} \\
\Rightarrow X^T\Sigma^{-1}\mathbf{y} &= X^T\Sigma^{-1}X\hat{\boldsymbol{\beta}}.
\end{aligned}$$

If X has full rank, then

$$\hat{\boldsymbol{\beta}} = \left(X^T\Sigma^{-1}X\right)^{-1}X^T\Sigma^{-1}\mathbf{y},$$

which concludes the proof. ■

The MLE is equivalent with the least squares estimate (LSE).

Definition 4.9 (Least Squares Estimate)
The maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ found by minimizing the distance $\|\mathbf{y} - \boldsymbol{\mu}\|_\Sigma^2$ is called the least squares estimate. [12, p. 50]

For a GLM (4.3) when $\boldsymbol{\epsilon}$ is normally distributed the MLE and LSE for $\boldsymbol{\beta}$ are identical.

Theorem 4.10 (Distribution of the Maximum Likelihood Estimator)
If X has full rank, then

$$\hat{\boldsymbol{\beta}} \sim N_k \left(\boldsymbol{\beta}, \sigma^2 \left(X^T \Sigma^{-1} X \right)^{-1} \right).$$

[12, p. 49]

Proof
Assume X has full rank. The MLE $\hat{\boldsymbol{\beta}}$ being normally distributed follows from $\mathbf{Y}$ being normally distributed. Recall $\boldsymbol{\mu} = X\boldsymbol{\beta}$ as defined in Definition 4.7. The expected value $\mathrm{E}[\hat{\boldsymbol{\beta}}]$ is

$$\begin{aligned}
\mathrm{E}[\hat{\boldsymbol{\beta}}] &= \mathrm{E}\left[\left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} \mathbf{Y} \right] \\
&= \left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} \mathrm{E}[\mathbf{Y}] \\
&= \left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} X \boldsymbol{\beta} \\
&= \boldsymbol{\beta}.
\end{aligned}$$

The variance $\mathrm{Var}[\hat{\boldsymbol{\beta}}]$ is then calculated, note that $\mathrm{Var}[A\mathbf{Y}] = A\mathrm{Var}[\mathbf{Y}]A^T$ and recall $\mathrm{Var}[\mathbf{Y}] = \sigma^2\Sigma$

$$\begin{aligned}
\mathrm{Var}[\hat{\boldsymbol{\beta}}] &= \mathrm{Var}\left[\left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} \mathbf{Y} \right] \\
&= \left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} \mathrm{Var}[\mathbf{Y}] \left(\left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} \right)^T \\
&= \left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} \sigma^2 \Sigma \left(\left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} \right)^T \\
&= \sigma^2 \left(X^T \Sigma^{-1} X \right)^{-1} X^T \Sigma^{-1} X \left(X^T \Sigma^{-1} X \right)^{-1} \\
&= \sigma^2 \left(X^T \Sigma^{-1} X \right)^{-1}.
\end{aligned}$$

The distribution of $\hat{\boldsymbol{\beta}}$ can then be written as $\hat{\boldsymbol{\beta}} \sim N_k \left(\boldsymbol{\beta}, \sigma^2 \left(X^T \Sigma^{-1} X \right)^{-1} \right)$. ■

A weighted problem can be expressed as an unweighted problem where $\Sigma = I$. Since Σ is symmetric and positive definite it is possible to find an invertible matrix $\Sigma^{1/2} \in \mathbb{R}^{n \times n}$ such that $\Sigma = \Sigma^{1/2} \left(\Sigma^{1/2} \right)^T$. Then for the GLM

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N_n \left(0, \sigma^2 \Sigma \right),$$

we obtain the following equation by multiplying with $\left(\Sigma^{1/2} \right)^{-1}$

$$\left(\Sigma^{1/2} \right)^{-1} \mathbf{Y} = \left(\Sigma^{1/2} \right)^{-1} X\boldsymbol{\beta} + \left(\Sigma^{1/2} \right)^{-1} \boldsymbol{\epsilon} \quad \Rightarrow \quad \tilde{\mathbf{Y}} = \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}},$$

where $\tilde{\boldsymbol{\epsilon}} \sim N_n\left(0, \sigma^2 I\right)$, since

$$\begin{aligned}
\text{Var}[\tilde{\boldsymbol{\epsilon}}] &= \text{Var}\left[\left(\Sigma^{1/2}\right)^{-1}\boldsymbol{\epsilon}\right] \\
&= \left(\Sigma^{1/2}\right)^{-1}\text{Var}[\boldsymbol{\epsilon}]\left(\left(\Sigma^{1/2}\right)^{-1}\right)^T \\
&= \left(\Sigma^{1/2}\right)^{-1}\sigma^2\Sigma^{1/2}\left(\Sigma^{1/2}\right)^T\left(\left(\Sigma^{1/2}\right)^T\right)^{-1} \\
&= \sigma^2 I.
\end{aligned}$$

Henceforth, it will be assumed that $\Sigma = I$. The MLE $\hat{\boldsymbol{\beta}}$ from Theorem 4.8 can then be rewritten as

$$\hat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T \mathbf{y}.$$

Leading to the following expression for the MLE

$$\hat{\boldsymbol{\mu}} = X\hat{\boldsymbol{\beta}} = X\left(X^T X\right)^{-1} X^T \mathbf{y}. \tag{4.4}$$

The MLE $\hat{\boldsymbol{\mu}}$ can be found as the projection of $\mathbf{y}$ on Ω_0.

Definition 4.11 (Projection Matrix)
A matrix H is a projection matrix if and only if $H^T = H$ and $H^2 = H$, i.e. H is symmetric and idempotent. [12, p. 50]

If the matrix $X\left(X^T X\right)^{-1} X^T$ from (4.4) fulfils Definition 4.11, it is a projection matrix, and is denoted H, leading to

$$\hat{\boldsymbol{\mu}} = H\mathbf{y}.$$

The difference between the estimated values and the observed values is called residuals.

Definition 4.12 (Observed Residuals)
Observed residuals are defined as

$$\mathbf{r} = \mathbf{y} - X\hat{\boldsymbol{\beta}} = (I - H)\mathbf{y}.$$

[12, p. 51]

The residuals are normally distributed but does not necessarily have the same variance. The variance of r_i is

$$\text{Var}[r_i] = \sigma^2(1 - h_{ii}),$$

where h_{ii} denotes the i'th diagonal element in the projection matrix H [12, p. 73]. A residual is illustrated in Example 4.13.

Example 4.13 (Residuals)
We consider a built in dataset from **R** called "cars". The data points are measurements of the stopping distance for a car measured in ft as a function of the speed in mph. The GLM for the data set is plotted as a black line in Figure 4.1.

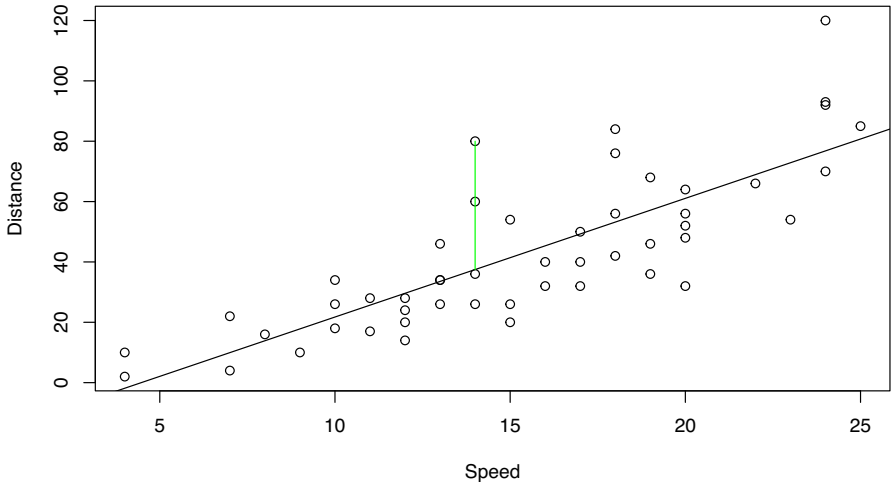


Figure 4.1: Residual from data set "cars" in **R**.

The residual is illustrated as a green line and shows the distance between the observed and estimated value.

To analyse how well a GLM fits the data, compared to the null model, the coefficient of determination is defined.

Definition 4.14 (Coefficient of Determination)
Given a linear model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, the coefficient of determination R^2 is defined as

$$R^2 = \frac{\|H_1\mathbf{y} - H_{null}\mathbf{y}\|^2}{\|\mathbf{y} - H_{null}\mathbf{y}\|^2} = 1 - \frac{\|\mathbf{y} - H_1\mathbf{y}\|^2}{\|\mathbf{y} - H_{null}\mathbf{y}\|^2},$$

where H_1 and H_{null} are projections of the observations $\mathbf{y}$ on the subspaces Ω_1 and Ω_{null}, respectively, and $R^2 \in [0, 1]$. [12, p. 57]

The distance between an observation and the associated estimation obtained by the null model is determined by $\|\mathbf{y} - H_{null}\mathbf{y}\|^2$. While the variation accounted for by the model 1 is given by $\|H_1\mathbf{y} - H_{null}\mathbf{y}\|^2$. If these are close it means that model 1 accounts for more of the variance compared to the null model. Therefore, if $R^2 \approx 1$ the model is considered a good model. The interpretation of the coefficient of determination is illustrated in Figure 4.2.

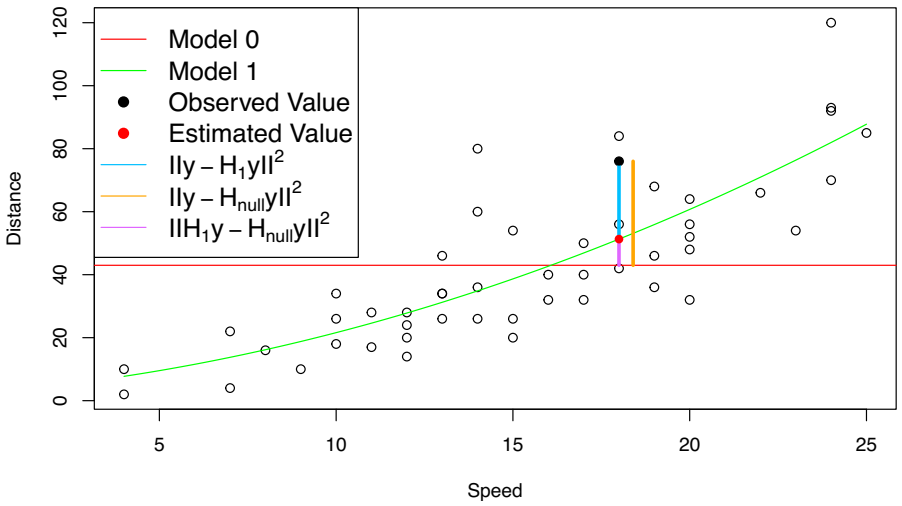


Figure 4.2: Comparison of model 0, and model 1, by the coefficient of determination.

Having presented all the necessary theory from statistics and GLMs to estimate voltages and currents in an electric grid, the next step is to introduce the complex normal distribution.

5 Complex Normal Distribution

The measurement errors from the smart meters are assumed to be complex normal distributed. Therefore, some necessary theory to derive the pdf for the complex normal distribution is introduced in this chapter.

5.1 Preliminaries

Before introducing the complex normal distribution, some basic knowledge of complex vectors is presented. Let Ω be the sample space of a random experiment and define two stochastic vectors $\mathbf{U}, \mathbf{V} : \Omega \to \mathbb{R}^n$. Then a real composite stochastic vector $\mathbf{Z} : \Omega \to \mathbb{R}^{2n}$ is defined as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}.$$

The vectors $\mathbf{U}$ and $\mathbf{V}$ compose a real and imaginary part to define a complex stochastic vector $\mathbf{X} : \Omega \to \mathbb{C}^n$ as

$$\mathbf{X} = \mathbf{U} + j\mathbf{V}.$$

A complex augmented vector $\underline{\mathbf{X}} : \Omega \to \underline{\mathbb{C}}^{2n}$ is defined by

$$\underline{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \overline{\mathbf{X}} \end{bmatrix},$$

where $\overline{\mathbf{X}}$ is the complex conjugate of $\mathbf{X}$ and $\underline{\mathbb{C}}^{2n}$ is the space of complex augmented vectors. To define the relation between augmented and composite vectors the real-to-complex transformation $T \in \mathbb{C}^{2n \times 2n}$ is introduced as

$$T = \begin{bmatrix} I & jI \\ I & -jI \end{bmatrix}. \tag{5.1}$$

The conjugated transpose of T is denoted by T^H and have the relation

$$TT^H = T^H T = 2I.$$

Having introduced the real-to-complex transformation, the relation between the complex augmented and real composite vector can be shown as

$$\underline{\mathbf{X}} = T\mathbf{Z} \quad \Leftrightarrow \quad \mathbf{Z} = \frac{1}{2} T^H \underline{\mathbf{X}} \tag{5.2}$$

[14, p. 31]. In order to define the complex normal distribution, the mean and covariance matrix for $\mathbf{Z}$ must be derived.

5.2 Mean and Covariance

The mean vector for the real composite stochastic vector $\mathbf{Z}$ is

$$\boldsymbol{\mu}_{\mathbf{Z}} = \mathrm{E}[\mathbf{Z}] = \begin{bmatrix} \mathrm{E}[\mathbf{U}] \\ \mathrm{E}[\mathbf{V}] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{U}} \\ \boldsymbol{\mu}_{\mathbf{V}} \end{bmatrix}.$$

The covariance matrix $\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]$ is found by Definition 3.2

$$\begin{aligned} \mathrm{Cov}[\mathbf{Z}, \mathbf{Z}] &= \mathrm{E}[(\mathbf{Z} - \boldsymbol{\mu}_{\mathbf{Z}})(\mathbf{Z} - \boldsymbol{\mu}_{\mathbf{Z}})^T] \\ &= \begin{bmatrix} \mathrm{E}[(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})^T] & \mathrm{E}[(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})^T] \\ \mathrm{E}[(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}})(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})^T]^T & \mathrm{E}[(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})(\mathbf{V} - \boldsymbol{\mu}_{\mathbf{V}})^T] \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Cov}[\mathbf{U}, \mathbf{U}] & \mathrm{Cov}[\mathbf{U}, \mathbf{V}] \\ \mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T & \mathrm{Cov}[\mathbf{V}, \mathbf{V}] \end{bmatrix}. \end{aligned}$$

Having derived $\boldsymbol{\mu}_{\mathbf{Z}}$, the augmented mean vector of $\underline{\mathbf{X}}$ is found as

$$\boldsymbol{\mu}_{\underline{\mathbf{X}}} = \mathrm{E}[\underline{\mathbf{X}}] = T\boldsymbol{\mu}_{\mathbf{Z}} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{U}} + j\boldsymbol{\mu}_{\mathbf{V}} \\ \boldsymbol{\mu}_{\mathbf{U}} - j\boldsymbol{\mu}_{\mathbf{V}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \overline{\boldsymbol{\mu}}_{\mathbf{X}} \end{bmatrix}.$$

Lastly, the augmented covariance matrix of $\underline{\mathbf{X}}$ is found by Definition 3.2

$$\begin{aligned} \mathrm{Cov}[\underline{\mathbf{X}}, \underline{\mathbf{X}}] &= \mathrm{E}[(\underline{\mathbf{X}} - \boldsymbol{\mu}_{\underline{\mathbf{X}}})(\underline{\mathbf{X}} - \boldsymbol{\mu}_{\underline{\mathbf{X}}})^H] \\ &= \mathrm{E}[(T\mathbf{Z} - T\boldsymbol{\mu}_{\mathbf{Z}})(T\mathbf{Z} - T\boldsymbol{\mu}_{\mathbf{Z}})^H] \\ &= T\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]T^H \\ &= \begin{bmatrix} I & jI \\ I & -jI \end{bmatrix} \begin{bmatrix} \mathrm{Cov}[\mathbf{U}, \mathbf{U}] & \mathrm{Cov}[\mathbf{U}, \mathbf{V}] \\ \mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T & \mathrm{Cov}[\mathbf{V}, \mathbf{V}] \end{bmatrix} \begin{bmatrix} I & I \\ -jI & jI \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Cov}[\mathbf{U}, \mathbf{U}] + j\mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T & \mathrm{Cov}[\mathbf{U}, \mathbf{V}] + j\mathrm{Cov}[\mathbf{V}, \mathbf{V}] \\ \mathrm{Cov}[\mathbf{U}, \mathbf{U}] - j\mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T & \mathrm{Cov}[\mathbf{U}, \mathbf{V}] - j\mathrm{Cov}[\mathbf{V}, \mathbf{V}] \end{bmatrix} \begin{bmatrix} I & I \\ -jI & jI \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Cov}[\mathbf{X}, \mathbf{X}] & \widetilde{\mathrm{Cov}}[\mathbf{X}, \mathbf{X}] \\ \overline{\widetilde{\mathrm{Cov}}}[\mathbf{X}, \mathbf{X}] & \overline{\mathrm{Cov}}[\mathbf{X}, \mathbf{X}] \end{bmatrix} \end{aligned} \tag{5.3}$$

where

$$\begin{aligned} \mathrm{Cov}[\mathbf{X}, \mathbf{X}] &= \mathrm{Cov}[\mathbf{U}, \mathbf{U}] + \mathrm{Cov}[\mathbf{V}, \mathbf{V}] + j(\mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T - \mathrm{Cov}[\mathbf{U}, \mathbf{V}]), \\ \overline{\mathrm{Cov}}[\mathbf{X}, \mathbf{X}] &= \mathrm{Cov}[\mathbf{U}, \mathbf{U}] + \mathrm{Cov}[\mathbf{V}, \mathbf{V}] - j(\mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T - \mathrm{Cov}[\mathbf{U}, \mathbf{V}]), \end{aligned}$$

and $\widetilde{\mathrm{Cov}}[\mathbf{X}, \mathbf{X}]$ is the pseudo-covariance matrix defined as

$$\begin{aligned} \widetilde{\mathrm{Cov}}[\mathbf{X}, \mathbf{X}] &= \mathrm{Cov}[\mathbf{U}, \mathbf{U}] - \mathrm{Cov}[\mathbf{V}, \mathbf{V}] + j(\mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T + \mathrm{Cov}[\mathbf{U}, \mathbf{V}]), \\ \overline{\widetilde{\mathrm{Cov}}}[\mathbf{X}, \mathbf{X}] &= \mathrm{Cov}[\mathbf{U}, \mathbf{U}] - \mathrm{Cov}[\mathbf{V}, \mathbf{V}] - j(\mathrm{Cov}[\mathbf{U}, \mathbf{V}]^T + \mathrm{Cov}[\mathbf{U}, \mathbf{V}]) \end{aligned}$$

[14, p. 34]. Having presented the necessary theory, the complex normal distribution can be derived.

5.3 The Complex Normal Distribution

The pdf for the multivariate normal distribution is

$$f(\mathbf{Z}) = \frac{1}{(2\pi)^{2n/2}\sqrt{\det(\mathrm{Cov}[\mathbf{Z},\mathbf{Z}])}} \mathrm{e}^{-\frac{1}{2}(\mathbf{Z}-\boldsymbol{\mu}_\mathbf{Z})^T \mathrm{Cov}[\mathbf{Z},\mathbf{Z}]^{-1}(\mathbf{Z}-\boldsymbol{\mu}_\mathbf{Z})}. \tag{5.4}$$

From (5.3) it can be shown that

$$\mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}] = T\mathrm{Cov}[\mathbf{Z},\mathbf{Z}]T^H \quad \Rightarrow \quad \mathrm{Cov}[\mathbf{Z},\mathbf{Z}]^{-1} = T^H \mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}]^{-1} T.$$

Furthermore, by (5.1), Theorem A.1, and Theorem A.2, the determinant of $\mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}]$ can be found as

$$\begin{aligned}
\det(\mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}]) &= \det(T)\det(\mathrm{Cov}[\mathbf{Z},\mathbf{Z}])\det(T^H) \\
&= \det(I(-j)I - IjI)\det(\mathrm{Cov}[\mathbf{Z},\mathbf{Z}])\det(IjI - (-j)II) \\
&= (-2j)^n \det(\mathrm{Cov}[\mathbf{Z},\mathbf{Z}])(2j)^n \\
&= 2^{2n}\det(\mathrm{Cov}[\mathbf{Z},\mathbf{Z}]),
\end{aligned}$$

which implies that

$$\det(\mathrm{Cov}[\mathbf{Z},\mathbf{Z}]) = 2^{-2n}\det(\mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}]).$$

Now (5.4) can be rewritten as

$$\begin{aligned}
f(\mathbf{X}) &= \frac{1}{(2\pi)^{2n/2}\sqrt{2^{-2n}\det(\mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}])}} \mathrm{e}^{-\frac{1}{2}(\mathbf{Z}-\boldsymbol{\mu}_\mathbf{Z})^T T^H \mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}]^{-1} T(\mathbf{Z}-\boldsymbol{\mu}_\mathbf{Z})} \\
&= \frac{1}{(\pi^n)\sqrt{\mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}])}} \mathrm{e}^{-\frac{1}{2}(\mathbf{Z}-\boldsymbol{\mu}_\mathbf{Z})^T T^H \mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}]^{-1} T(\mathbf{Z}-\boldsymbol{\mu}_\mathbf{Z})}
\end{aligned}$$

[14, p. 39]. Remember by (5.2) that $\underline{\mathbf{X}} = T\mathbf{Z}$. The complex normal distribution can now be defined.

Definition 5.1 (Complex Normal Distribution)
If a complex stochastic vector $\mathbf{X} : \Omega \to \mathbb{C}^n$ has pdf

$$f(\mathbf{X}) = \frac{1}{(\pi^n)\sqrt{\det(\mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}])}} \mathrm{e}^{-\frac{1}{2}(\underline{\mathbf{X}}-\boldsymbol{\mu}_{\underline{\mathbf{X}}})^H \mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}]^{-1}(\underline{\mathbf{X}}-\boldsymbol{\mu}_{\underline{\mathbf{X}}})}$$

it is said to be complex normal distributed, denoted $\mathbf{X} \sim N_{\mathbb{C}}(\boldsymbol{\mu}_{\underline{\mathbf{X}}}, \mathrm{Cov}[\underline{\mathbf{X}},\underline{\mathbf{X}}])$. [14, p. 39]

The derivation of the pdf for the complex normal distribution concludes this chapter. To gain unique estimates for the voltages and currents in an electric grid, some optimisation theory is needed.

6 Optimisation

In this chapter, optimisation theory will be introduced to derive a model of the grid and find the MLE as presented in chapter 3. The voltages and currents in the model must obey Kirchhoff's- and Ohm's laws, which therefore formulate equality constraints for our model.

Consider the most basic optimisation problem, which is unconstrained, given an objective function f and decision variables $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$. The aim is to adjust the decision variables in such a way that $f(\mathbf{x})$ is minimised, which is stated mathematically as

$$\underset{\mathbf{x}}{\text{minimise}} \quad f(\mathbf{x}).$$

The following section expands the optimisation problem by introducing constraints [1, p. 4].

6.1 Constraints

A constrained optimisation problem is when the objective function is subject to some variables known as constraints, which are either defined as an inequality or equality. Hence, a general constrained minimisation problem is expressed as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimise}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & c_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, q, \\ & a_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p. \end{aligned}$$

The right hand side of the inequalities are, without loss of generality, all equal to zero, hence the minimisation problem

$$\underset{\mathbf{x}}{\text{minimise}} \quad f(\mathbf{x})$$

is equivalent to

$$-\underset{\mathbf{x}}{\text{maximise}} \quad (-f(\mathbf{x})).$$

A solution to the constrained optimisation problem, called the minimiser $\mathbf{x}^*$, must be within the feasible region [1, pp. 4, 17].

Definition 6.1 (Feasible Region)
Any point $\mathbf{x}$ that satisfies both the equality as well as the inequality is said to be a feasible point of the optimisation problem. The set of all points that satisfy the constraints constitutes the feasible domain region of $f(\mathbf{x})$

$$F = \{\mathbf{x} : a_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p \quad \text{and} \quad c_j(\mathbf{x}) \leq 0, \quad j = 1, \ldots, q\}.$$

[1, p. 286]

The feasible region contains both the feasible point $\mathbf{x}$ and the minimiser $\mathbf{x}^*$. Therefore, we can define a weak and local minimiser to be the extrema of the objective function.

Definition 6.2 (Weak Local Minimiser)
A point $\mathbf{x}^* \in F$ is said to be a weak local minimiser of $f(\mathbf{x})$ if there exists a distance $\epsilon > 0$ such that

$$f(\mathbf{x}) \geq f(\mathbf{x}^*), \forall\ \mathbf{x} \in F \tag{6.1}$$

if

$$||\mathbf{x} - \mathbf{x}^*|| < \epsilon. \tag{6.2}$$

[1, p. 31]

The difference between a weak and strong minimiser is that the inequality in equation (6.1) becomes a strict inequality. Furthermore, the local minimiser becomes global if inequality (6.2) is not required.

Another addition to the feasible region is a vector that can relocate a feasible point with a small step without leaving the feasible region, called a feasible direction.

Definition 6.3 (Feasible Direction)
Let $\delta = \alpha\mathbf{d}$ be a change in $\mathbf{x}$ where α is a positive constant then $\mathbf{d}$ is a feasible direction at $\mathbf{x}$, if

$$\exists \delta > 0 : \mathbf{x} + \alpha\mathbf{d} \in F, \quad \forall\, \alpha \in (0, \delta]$$

holds. [1, p. 33]

When working with constrained optimisation problems, the concept of Lagrange multipliers and Karush-Kuhn-Tucker (KKT) conditions is sufficient for a point $\mathbf{x}^*$ to be a solution. The Lagrange multipliers will be introduced in the next section through application of an equality constrained minimisation problem.

6.2 Equality constraints

Consider an equality constrained minimisation problem such as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimise}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & a_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p. \end{aligned}$$

The idea is to show the relationship between the gradient of the objective function and the gradient of the equality constraint, which leads to a reformulation of the minimisation

problem. Let $\mathbf{x}^*$ be the local minimiser of the equality constrained problem. By using the Taylor expansion of the constraint in some direction $||\mathbf{s}|| < 1$, we obtain

$$\begin{aligned} a_i(\mathbf{x}^* + \mathbf{s}) &= a_i(\mathbf{x}^*) + \mathbf{s}^T \nabla a_i(\mathbf{x}^*) + o(||\mathbf{s}||) \\ &= \mathbf{s}^T \nabla a_i(\mathbf{x}^*) + o(||\mathbf{s}||). \end{aligned}$$

Since $a_i(\mathbf{x}^*) = 0$ then $a_i(\mathbf{x}^* + \mathbf{s}) = 0$ which indicates that at the minimum, the gradient of the equality constraint is zero or $\mathbf{s}$ is orthogonal to the gradient. Therefore

$$\mathbf{s}^T \nabla a_i(\mathbf{x}^*) = 0,$$

which means that $\mathbf{s}$ is a feasible direction if and only if it is orthogonal to $\nabla a_i(\mathbf{x}^*)$. Now project $\nabla f(\mathbf{x}^*)$ onto $\{\nabla a_i(\mathbf{x}^*) | i = 1...p\}$ and call the projection

$$\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*),$$

where λ is called the Lagrangian multiplier. The gradient of the objective function evaluated at the minimiser is

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \mathbf{d}, \tag{6.3}$$

where $\mathbf{d}$ is orthogonal to the gradients of the constraint functions.

Next we will show that if $\mathbf{x}^*$ is a local minimiser, then $\mathbf{d}$ is zero by using contradiction. Therefore, if $\mathbf{x}^*$ is a local minimiser assume that $\mathbf{d} \neq \mathbf{0}$ and let $\mathbf{s} = -\mathbf{d}$. The equation (6.3) still holds when $\mathbf{d}$ is replaced, $\mathbf{s}$ is a feasible direction at $\mathbf{x}^*$. By multiplying $\mathbf{s}^T$ with equation (6.3) we have that

$$\mathbf{s}^T \nabla f(\mathbf{x}^*) = \mathbf{s}^T \left(\sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \mathbf{d} \right) = -||\mathbf{d}||^2 < 0, \tag{6.4}$$

because of orthogonality. Hence equation (6.4) implies that $\mathbf{s}$ directs $f(\mathbf{x}^*)$ towards an even smaller value

$$f(\mathbf{x}^* + \mathbf{s}) \leq f(\mathbf{x}^*),$$

which by Definition 6.2 contradicts the fact that $\mathbf{x}^*$ is a minimiser. Hence, $\mathbf{d}$ must be zero and (6.3) becomes

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*). \tag{6.5}$$

The functionality of Lagrange multipliers can also be interpreted by the Lagrangian, which is a modified objective function and is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i a_i(\mathbf{x}). \tag{6.6}$$

To intertwine the Lagrangian with (6.5), we can write the gradient of the Lagrangian as

$$\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*),$$

where $\mathbf{x}^*$ is a local minimiser and $\boldsymbol{\lambda}^*$ is the equality Lagrange multipliers. The solution set $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is referred to as the minimiser set. Furthermore, the minimiser set of the Lagrangian can be expanded for inequality constraints [1, pp. 305–307].

6.3 The Lagrangian

If we consider the general constrained minimisation problem, the inequality constraints $c_j(\mathbf{x}) \leq 0$ are active at $\mathbf{x}^*$ when $c_j(\mathbf{x}^*) = 0$, and inactive when $c_j(\mathbf{x}^*) < 0$. Being active implies that the condition for a minimum involves the constraint and a non-zero Lagrange multiplier appears. Therefore the Lagrangian can be modified to

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i a_i(\mathbf{x}) + \sum_{j=1}^{q} \mu_j c_j(\mathbf{x}),$$

where $\boldsymbol{\lambda}, \boldsymbol{\mu}$ are Lagrangian multipliers. The extraction of the additional Lagrangian multiplier is omitted. The following definition, known as the KKT conditions, summarises the conditions needed such that the Lagrangian can be used to find the minimiser [1, p. 314].

Definition 6.4 (Karush-Kuhn-Tucker Conditions)
If $\mathbf{x}^*$ is a local minimiser for the general constrained minimisation problem. Then,

a) $a_i(\mathbf{x}^*) = 0$ for $1 \leq i \leq p$,

b) $c_j(\mathbf{x}^*) \leq 0$ for $1 \leq j \leq q$,

c) there exists Lagrange multipliers λ_i^* for $1 \leq i \leq p$ and μ_j^* for $1 \leq j \leq q$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \sum_{j=1}^{q} \mu_j^* \nabla c_j(\mathbf{x}^*) = 0$$

d) $\lambda_i^* a_i(\mathbf{x}^*) = 0$ for $1 \leq i \leq p$,
$\mu_j^* c_j(\mathbf{x}^*) = 0$ for $1 \leq j \leq q$,

e) and $\mu_j^* \geq 0$ for $1 \leq j \leq q$.

[1, p. 316]

The conditions a) and b) mean that $\mathbf{x}^*$ must be a feasible point. Condition c) states that the gradient of the Lagrangian is zero, meaning that the $\mathbf{x}^*$ is the minimum. Condition d) is referred to as the complementary KKT conditions, meaning that the Lagrangian multiplier and the constraints can not be non-zero simultaneously. Therefore, the inactive constraints make the Lagrangian multiplier zero and are not considered in the Lagrangian. The last condition e) implies that the inequality constraints must be non-negative, which ensures that the gradient points toward the feasible region [1, p. 317].

To properly calculate the Lagrangian related to the currents and voltages of this project's model, complex differentiation is needed.

7 Complex Differentiation

To realise the model for this project, we need to obtain the gradient of the Lagrangian. The Lagrangian is often a real valued, non-negative function that will take a complex argument, such as the voltages and currents for our model. To differentiate a real valued function with respect to a complex argument, the necessary conditions for a function to be complex-differentiable must be defined. Let the vectors $\mathbf{X}$ and $\overline{\mathbf{X}}$ be defined as in Section 5.1. Let x and $\overline{x}$ be the first entries in $\mathbf{X}$ and $\overline{\mathbf{X}}$, respectively.

Consider the classical derivative of a function evaluated at x_0 defined as

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

The limit exists if it is independent of the direction, when x goes towards x_0, and only then f is called differentiable at x_0. Furthermore a differentiable complex function can be defined as holomorphic.

Definition 7.1 (Holomorphic)
A function f on an open domain $A \in \mathbb{C}$ is called holomorphic if it is complex-differentiable for every $x \in A$. [14, p. 280]

When f is holomorphic the generalised complex derivatives, known as Wirtinger derivatives, exist. The idea is to regard f as a function of two independent complex variables, x and $\overline{x}$. The generalised complex derivative is defined by partial differentiation.

Definition 7.2 (Wirtinger Derivatives)
The generalised complex differential operator for $f(x)$ is defined as

$$\frac{\partial}{\partial x} \triangleq \frac{1}{2}\left(\frac{\partial}{\partial u} - j\frac{\partial}{\partial v}\right),$$

and the conjugate generalised complex differential operator as

$$\frac{\partial}{\partial \overline{x}} \triangleq \frac{1}{2}\left(\frac{\partial}{\partial u} + j\frac{\partial}{\partial v}\right).$$

[14, pp. 278–279]

An application of the Wirtinger derivatives can be seen in the following example.

Example 7.3 (Application of Wirtinger Derivatives)
Consider the function $f(x) = |x|^2 = x\overline{x}$. We want to show that the Wirtinger derivatives yields

$$\frac{\partial}{\partial x} f(x) = \overline{x} \quad \text{and} \quad \frac{\partial}{\partial \overline{x}} f(\overline{x}) = x.$$

The function can be rewritten as $f(x) = f(u, v) = u^2 + v^2$. Use Wirtingers generalised complex-differential operator

$$\frac{1}{2}\left(\frac{\partial}{\partial u} - j\frac{\partial}{\partial v}\right) f(u, v) = \frac{1}{2}(2u - j2v) = u - jv = \overline{x},$$

then the conjugate generalised complex differential operator

$$\frac{1}{2}\left(\frac{\partial}{\partial u} + j\frac{\partial}{\partial v}\right) f(u, v) = \frac{1}{2}(2u + j2v) = u + jv = x.$$

The Wirtinger derivatives can be expanded to vector valued functions, thus the following rules applies for Wirtinger derivatives.

Theorem 7.4
Let $\mathbf{f} : \mathbb{C}^n \to \mathbb{R}$ be vector valued functions.

1.
$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x} = \mathbf{I} \quad \text{and} \quad \frac{\partial}{\partial \overline{\mathbf{x}}} \overline{\mathbf{x}} = \mathbf{I},$$

2.
$$\frac{\partial}{\partial \overline{\mathbf{x}}} \mathbf{x} = \mathbf{0} \quad \text{and} \quad \frac{\partial}{\partial \mathbf{x}} \overline{\mathbf{x}} = \mathbf{0},$$

3.
$$\frac{\partial}{\partial \overline{\mathbf{x}}} \overline{\mathbf{f}} = \overline{\left(\frac{\partial}{\partial \mathbf{x}} \mathbf{f}\right)} \quad \text{and} \quad \frac{\partial}{\partial \overline{\mathbf{x}}} \mathbf{f} = \overline{\left(\frac{\partial}{\partial \mathbf{x}} \overline{\mathbf{f}}\right)}. \tag{7.1}$$

[14, p. 281]

The first two rules in Theorem 7.4 explain the idea behind Wirtinger derivatives. The third rule shows two properties for the complex conjugate.

With the necessary theory presented, the next step is to derive a model for the problem.

8 Model

To construct an experiment that estimates voltages and currents in an electric grid, a model must first be derived.

8.1 Description of the Model

An electric grid can be represented by a tree structure, as seen in Figure 8.1. The nodes represent the voltages, and the edges represent the conductors, each containing a current and an impedance. There are a total of n nodes and $n-1$ edges, which leads to a total of $N = 2n-1$ elements in the tree structure. The current between node j and k is denoted by $i_{j,k}$, where j is called the parent node and k is the child. All j's children are called siblings. Notice that node v_0 is different compared to the other nodes in the grid. The node v_0 represents a slack bus, which is the connection to the local transformer station on the secondary side of the coil. The voltage at v_0 is known, and we assume that the voltage does not change.

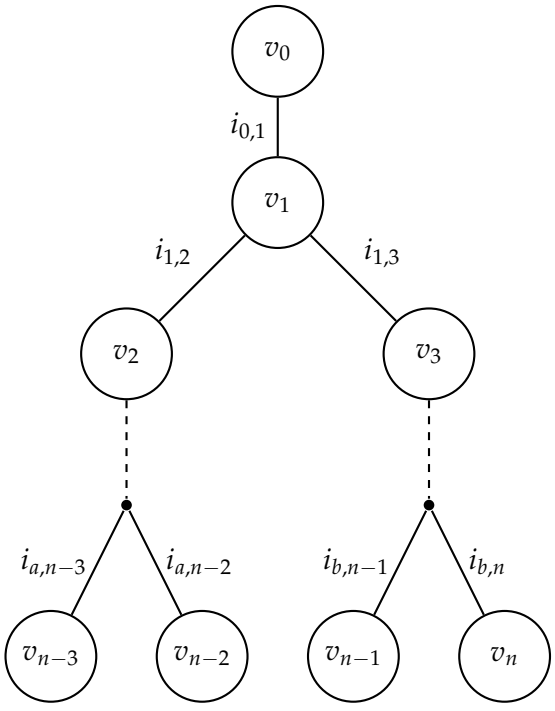


Figure 8.1: A tree structure representing an electrical grid. The dashed lines symbolise that the tree could have many different structures between the nodes.

The voltages and currents in the electrical grid must satisfy KCL, KVL and, Ohm's law as presented in chapter 2.

This results in two linear equations. The first equation is defined by KCL

$$i_{parent(j),j} - \sum_{k \in children(j)} i_{j,k} = 0. \tag{8.1}$$

The second equation is defined by KVL and Ohm's law

$$v_j - v_k - Z_{j,k} i_{j,k} = 0, \tag{8.2}$$

where $Z_{j,k}$ is the impedance of the conductor from node j to k and is defined by Definition 2.2. The linear equations (8.1) and (8.2) can be seen as constraints for an optimisation problem. Let q denote the number of linear equations. The two linear equations can be summarised by the matrix vector product

$$C\mathbf{x} = \mathbf{0}, \tag{8.3}$$

where $\mathbf{x} \in \mathbb{C}^N$ is a vector of the true voltages and currents in the grid, and $C \in \mathbb{C}^{q \times N}$ is a matrix that summarises the linear equations obtained by KCL, KVL and, Ohm's law. The true voltages and currents must equal the estimated voltages and currents with additional measurement errors ϵ. Therefore, the problem can be modelled as a GLM

$$\mathbf{d} = D\mathbf{x} + \boldsymbol{\epsilon}, \tag{8.4}$$

where $\mathbf{d} \in \mathbb{C}^K$ is a vector of the K measurements obtained by smart meters and $D \in \mathbb{C}^{K \times N}$ is a design matrix that relates the obtained measurements with their position in the grid. All of the voltages and currents are complex valued and the measurement errors are therefore assumed to be complex normal distributed

$$\boldsymbol{\epsilon} \sim N_{\mathbb{C}}(0, \mathrm{Cov}[\underline{\boldsymbol{\epsilon}}, \underline{\boldsymbol{\epsilon}}]),$$

where

$$\mathrm{Cov}[\underline{\boldsymbol{\epsilon}}, \underline{\boldsymbol{\epsilon}}]) = \begin{bmatrix} \mathrm{Cov}[\boldsymbol{\epsilon}, \boldsymbol{\epsilon}] & \widetilde{\mathrm{Cov}}[\boldsymbol{\epsilon}, \boldsymbol{\epsilon}] \\ \overline{\widetilde{\mathrm{Cov}}}[\boldsymbol{\epsilon}, \boldsymbol{\epsilon}] & \overline{\mathrm{Cov}}[\boldsymbol{\epsilon}, \boldsymbol{\epsilon}] \end{bmatrix},$$

which was derived in Section 5.2. The vector $\mathbf{x}$ is unknown, hence we need to estimate it.

8.2 Estimation of the Model

To estimate voltages and currents in the electric grid, the MLE of $\mathbf{x}$ must be found. However, the maximum likelihood estimation of $\mathbf{x}$ can not provide a unique estimate, since there are typically too few measurements compared to unknown values. Therefore, the least squares estimation, from Definition 4.9, will be used instead. The MLE of $\mathbf{x}$ is therefore found by minimising the distance $\|\mathbf{d} - D\mathbf{x}\|_\Sigma^2$ subject to the constraints in (8.3), meaning that the LSE is determined by solving an optimisation problem as

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimise}} \ \|\mathbf{d} - D\mathbf{x}\|_\Sigma^2 \\ &\text{subject to } C\mathbf{x} = \mathbf{0}, \end{aligned}$$

where $\Sigma = \mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon})$. Since the measurement errors are assumed to be complex normal distributed, the covariance matrix is augmented, thus, we have to minimise an augmented version of (8.4). Therefore, the objective function for the complex GLM can be written as

$$\|\underline{\mathbf{d}} - \underline{D}\mathbf{x}\|^2_{\underline{\Sigma}} = \delta_{\underline{\Sigma}}(\underline{\mathbf{d}} - \underline{D}\mathbf{x}) = (\underline{\mathbf{d}} - \underline{D}\mathbf{x})^H \underline{\Sigma}^{-1}(\underline{\mathbf{d}} - \underline{D}\mathbf{x}),$$

which gives the optimisation problem

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimise}}\ (\underline{\mathbf{d}} - \underline{D}\mathbf{x})^H \underline{\Sigma}^{-1}(\underline{\mathbf{d}} - \underline{D}\mathbf{x}) \\ &\text{subject to } C\mathbf{x} = \mathbf{0}. \end{aligned}$$

The constraints can be separated into the real and imaginary parts, yielding

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimise}}\ (\underline{\mathbf{d}} - \underline{D}\mathbf{x})^H \underline{\Sigma}^{-1}(\underline{\mathbf{d}} - \underline{D}\mathbf{x}) \\ &\text{subject to } \mathrm{Re}[C\mathbf{x}] = \mathbf{0} \\ &\qquad\qquad \mathrm{Im}[C\mathbf{x}] = \mathbf{0}. \end{aligned} \tag{8.5}$$

To solve the optimisation problem in (8.5), we use the Lagrangian as defined in (6.6)

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = (\underline{\mathbf{d}} - \underline{D}\mathbf{x})^H \underline{\Sigma}^{-1}(\underline{\mathbf{d}} - \underline{D}\mathbf{x}) + \boldsymbol{\lambda}_1^T \mathrm{Re}[C\mathbf{x}] + \boldsymbol{\lambda}_2^T \mathrm{Im}[C\mathbf{x}],$$

where $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ denotes the complex Lagrange multipliers corresponding to the real and imaginary component, respectively. The real and imaginary parts of a complex number can be extracted using the conjugation, therefore

$$\begin{aligned} \mathrm{Re}[C\mathbf{x}] &= \frac{C\mathbf{x} + \overline{C\mathbf{x}}}{2}, \\ \mathrm{Im}[C\mathbf{x}] &= \frac{C\mathbf{x} - \overline{C\mathbf{x}}}{2j}. \end{aligned}$$

Let

$$k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \boldsymbol{\lambda}_1^T \frac{C\mathbf{x} + \overline{C\mathbf{x}}}{2} + \boldsymbol{\lambda}_2^T \frac{C\mathbf{x} - \overline{C\mathbf{x}}}{2j},$$

then the Lagrangian can be simplified to

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = (\underline{\mathbf{d}} - \underline{D}\mathbf{x})^H \underline{\Sigma}^{-1}(\underline{\mathbf{d}} - \underline{D}\mathbf{x}) + k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2). \tag{8.6}$$

The inverse of

$$\underline{\Sigma} = \begin{bmatrix} \Sigma & \widetilde{\Sigma} \\ \overline{\widetilde{\Sigma}} & \overline{\Sigma} \end{bmatrix}$$

is found by matrix inversion from Theorem A.3, thus

$$\underline{\Sigma}^{-1} = \begin{bmatrix} \Sigma^{-1} + \Sigma^{-1}\widetilde{\Sigma}\left(\overline{\Sigma} - \overline{\widetilde{\Sigma}}\Sigma^{-1}\widetilde{\Sigma}\right)^{-1}\overline{\widetilde{\Sigma}}\Sigma^{-1} & -\Sigma^{-1}\widetilde{\Sigma}\left(\overline{\Sigma} - \overline{\widetilde{\Sigma}}\Sigma^{-1}\widetilde{\Sigma}\right)^{-1} \\ -\left(\overline{\Sigma} - \overline{\widetilde{\Sigma}}\Sigma^{-1}\widetilde{\Sigma}\right)^{-1}\overline{\widetilde{\Sigma}}\Sigma^{-1} & \left(\overline{\Sigma} - \overline{\widetilde{\Sigma}}\Sigma^{-1}\widetilde{\Sigma}\right)^{-1} \end{bmatrix}. \tag{8.7}$$

To ease notation, we introduce $P = (\overline{\Sigma} - \overline{\widetilde{\Sigma}}\Sigma^{-1}\widetilde{\Sigma})^{-1}$ and $W = \overline{\widetilde{\Sigma}}\Sigma^{-1}$, where $W^H = \Sigma^{-1}\widetilde{\Sigma}$. Then (8.7) can be rewritten as

$$\underline{\Sigma}^{-1} = \begin{bmatrix} \Sigma^{-1} + W^H P W & -W^H P \\ -PW & P \end{bmatrix}. \tag{8.8}$$

Before inserting (8.8) in the Lagrangian in (8.6), the remaining augmented expressions will be expressed on vector form

$$\begin{aligned} (\underline{\mathbf{d}} - \underline{D\mathbf{x}})^H &= \begin{bmatrix} (\mathbf{d} - D\mathbf{x})^H & (\mathbf{d} - D\mathbf{x})^T \end{bmatrix}, \\ \underline{\mathbf{d}} - \underline{D\mathbf{x}} &= \begin{bmatrix} \mathbf{d} - D\mathbf{x} \\ \overline{\mathbf{d}} - \overline{D\mathbf{x}} \end{bmatrix}. \end{aligned}$$

Now the Lagrangian can be rewritten as

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) &= \begin{bmatrix} (\mathbf{d} - D\mathbf{x})^H & (\mathbf{d} - D\mathbf{x})^T \end{bmatrix} \begin{bmatrix} \Sigma^{-1} + W^H P W & -W^H P \\ -PW & P \end{bmatrix} \begin{bmatrix} \mathbf{d} - D\mathbf{x} \\ \overline{\mathbf{d}} - \overline{D\mathbf{x}} \end{bmatrix} + k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \\ &= \begin{bmatrix} (\mathbf{d} - D\mathbf{x})^H & (\mathbf{d} - D\mathbf{x})^T \end{bmatrix} \begin{bmatrix} \Sigma^{-1} + W^H P W(\mathbf{d} - D\mathbf{x}) - W^H P \left(\overline{\mathbf{d}} - \overline{D\mathbf{x}}\right) \\ -PW(\mathbf{d} - D\mathbf{x}) + P \left(\overline{\mathbf{d}} - \overline{D\mathbf{x}}\right) \end{bmatrix} + k\left(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\right) \\ &= (\mathbf{d} - D\mathbf{x})^H \left(\Sigma^{-1} + W^H P W(\mathbf{d} - D\mathbf{x}) - W^H P \left(\overline{\mathbf{d}} - \overline{D\mathbf{x}}\right)\right) \\ &\quad + (\mathbf{d} - D\mathbf{x})^T \left(-PW(\mathbf{d} - D\mathbf{x}) + P \left(\overline{\mathbf{d}} - \overline{D\mathbf{x}}\right)\right) + k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2). \end{aligned}$$

To solve the optimisation problem, the derivative of the Lagrangian must first be derived. The reason for this is that it will later be set equal to zero to find an equation to which the MLE of $\mathbf{x}$ is a solution. Therefore, the derivative of the Lagrangian with respect to $\mathbf{x}$, $\overline{\mathbf{x}}$ and $\boldsymbol{\lambda}$ will be derived by applying Wirtinger calculus from Definition 7.2. First the derivative of $k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ with respect to $\mathbf{x}$ is found

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) &= \frac{1}{2}\boldsymbol{\lambda}_1 C^T + \frac{1}{2j}\boldsymbol{\lambda}_2 C^T \\ &= \frac{1}{2} C^T \left(\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2\right). \end{aligned} \tag{8.9}$$

Now the remaining part of the derivative of the Lagrangian with respect to $\mathbf{x}$ is found

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \frac{\partial}{\partial \mathbf{x}} \Bigg(& (\mathbf{d} - D\mathbf{x})^H \Bigg(\underbrace{\left(\Sigma^{-1} + W^H P W\right)}_{A_1} (\mathbf{d} - D\mathbf{x}) - \underbrace{(W^H P)}_{B_2} (\overline{\mathbf{d}} - \overline{D\mathbf{x}}) \Bigg) \\ & + (\mathbf{d} - D\mathbf{x})^T \Bigg(- \underbrace{(PW)}_{C_3} (\mathbf{d} - D\mathbf{x}) + P(\overline{\mathbf{d}} - \overline{D\mathbf{x}}) \Bigg) \Bigg) + \frac{\partial}{\partial \mathbf{x}} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2), \end{aligned}$$

where A_1, B_2 and C_3 are introduced to ease the notation in the following calculations

$$\begin{aligned}
&= \frac{\partial}{\partial \mathbf{x}} \Bigg((\mathbf{d} - D\mathbf{x})^H \left(A_1\mathbf{d} - A_1 D\mathbf{x} - B_2\overline{\mathbf{d}} + B_2\overline{D\mathbf{x}} \right) \\
&\quad + (\mathbf{d} - D\mathbf{x})^T \left(-C_3\mathbf{d} + C_3 D\mathbf{x} + P\overline{\mathbf{d}} - P\overline{D\mathbf{x}} \right) \Bigg) + \frac{\partial}{\partial x} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \\
&= \frac{\partial}{\partial \mathbf{x}} \Bigg(\mathbf{d}^H A_1\mathbf{d} - \mathbf{d}^H A_1 D\mathbf{x} - \mathbf{d}^H B_2\overline{\mathbf{d}} + \mathbf{d}^H B_2\overline{D\mathbf{x}} \\
&\quad - \mathbf{x}^H D^H A_1\mathbf{d} + \mathbf{x}^H D^H A_1 D\mathbf{x} + \mathbf{x}^H D^H B_2\overline{\mathbf{d}} - \mathbf{x}^H D^H B_2\overline{D\mathbf{x}} \\
&\quad - \mathbf{d}^T C_3\mathbf{d} + \mathbf{d}^T C_3 D\mathbf{x} + \mathbf{d}^T P\overline{\mathbf{d}} - \mathbf{d}^T P\overline{D\mathbf{x}} \\
&\quad + \mathbf{x}^T D^T C_3\mathbf{d} - \mathbf{x}^T D^T C_3 D\mathbf{x} - \mathbf{x}^T D^T P\overline{\mathbf{d}} + \mathbf{x}^T D^T P\overline{D\mathbf{x}} \Bigg) \\
&\quad + \frac{\partial}{\partial x} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2). \qquad (8.10)
\end{aligned}$$

The terms not involving $\mathbf{x}$ will be equal to zero when deriving with respect to $\mathbf{x}$, therefore noting that $x^H = \overline{x}^T$, (8.10) can be simplified to

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) &= \frac{\partial}{\partial \mathbf{x}} \Bigg(- \mathbf{d}^H A_1 D\mathbf{x} + \mathbf{x}^H D^H A_1 D\mathbf{x} + \mathbf{d}^T C_3 D\mathbf{x} + \mathbf{x}^T D^T C_3\mathbf{d} \\
&\quad - \mathbf{x}^T D^T C_3 D\mathbf{x} - \mathbf{x}^T D^T P\overline{\mathbf{d}} + \mathbf{x}^T D^T P\overline{D\mathbf{x}} \Bigg) + \frac{\partial}{\partial \mathbf{x}} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2),
\end{aligned}$$

using matrix calculus to derive the Lagrangian with respect to $\mathbf{x}$ yields

$$\begin{aligned}
&= -(\mathbf{d}^H A_1 D)^T + (\mathbf{x}^H D^H A_1 D)^T + (\mathbf{d}^T C_3 D)^T + D^T C_3\mathbf{d} \\
&\quad - D^T C_3 D\mathbf{x} - (D^T C_3 D)^T\mathbf{x} - D^T P\overline{\mathbf{d}} + D^T P\overline{D\mathbf{x}} + \frac{\partial}{\partial x} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \\
&= -D^T (A_1 + P^T)^T\overline{\mathbf{d}} + D^T (A_1 + P^T)^T\overline{D\mathbf{x}} \\
&\quad + D^T (C_3^T + C_3)\mathbf{d} - D^T (C_3 + C_3^T) D\mathbf{x} + \frac{\partial}{\partial \mathbf{x}} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2), \qquad (8.11)
\end{aligned}$$

where

$$\begin{aligned}
G_1 &= D^T (C_3^T + C_3) = D^T (W^T P^T + PW), \\
G_2 &= D^T (A_1 + P^T)^T = D^T (\Sigma^{-1} + W^H PW + P^T)^T.
\end{aligned}$$

With the expressions for G_1 and G_2, inserting (8.9) in (8.11) leads to the following expression

$$\frac{\partial}{\partial \mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = -G_2\overline{\mathbf{d}} + G_2\overline{D\mathbf{x}} + G_1\mathbf{d} - G_1 D\mathbf{x} + \frac{1}{2} C^T (\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2). \qquad (8.12)$$

Then the derivative of the Lagrangian with respect to $\overline{\mathbf{x}}$ is found by (7.1)

$$\begin{aligned}
\frac{\partial}{\partial \overline{\mathbf{x}}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) &= -\overline{G}_2\mathbf{d} + \overline{G}_2 D\mathbf{x} + \overline{G}_1\overline{\mathbf{d}} - \overline{G}_1\overline{D\mathbf{x}} + \frac{\partial}{\partial \overline{\mathbf{x}}} (k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)) \\
&= -\overline{G}_2\mathbf{d} + \overline{G}_2 D\mathbf{x} + \overline{G}_1\overline{\mathbf{d}} - \overline{G}_1\overline{D\mathbf{x}} + \frac{1}{2} C^H (\boldsymbol{\lambda}_1 + j\boldsymbol{\lambda}_2). \qquad (8.13)
\end{aligned}$$

Lastly, the derivative of the Lagrangian with respect to $\boldsymbol{\lambda}$ will be found. Since $k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ is the only term that depends on $\boldsymbol{\lambda}$, we have that

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\lambda}} k(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) &= \frac{1}{2}\mathrm{Re}(C\mathbf{x}) + \frac{1}{2j}\mathrm{Im}(C\mathbf{x}) \\ &= \frac{1}{2}\left(\mathrm{Re}(C\mathbf{x}) - j\ \mathrm{Im}(C\mathbf{x})\right) \\ &= \frac{1}{2}\overline{C\mathbf{x}}.\end{aligned}$$

Now that the derivative of the Lagrangian is derived, it can be set equal to zero to find the solution for the MLE. Hence, (8.12) and (8.13) is combined and set equal to zero

$$\begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda}_2) \\ \frac{\partial}{\partial \overline{\mathbf{x}}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \end{bmatrix} = \begin{bmatrix} G_1 & -G_2 \\ -\overline{G}_2 & \overline{G}_1 \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \overline{\mathbf{d}} \end{bmatrix} - \begin{bmatrix} G_1 D & -G_2 \overline{D} \\ -\overline{G}_2 D & \overline{G}_1 \overline{D} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \overline{\mathbf{x}} \end{bmatrix} + \begin{bmatrix} \frac{1}{2}C^T & O \\ O & \frac{1}{2}C^H \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2 \\ \overline{\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2} \end{bmatrix},$$

expressing it more compactly as

$$= \underline{G\mathbf{d}} - \underline{GD\mathbf{x}} + \underline{C}^T(\underline{\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2}) = \mathbf{0}. \tag{8.14}$$

Solving (8.14) for $\underline{G\mathbf{d}}$ yields

$$\begin{aligned}\underline{G\mathbf{d}} &= \underline{GD\mathbf{x}} - \underline{C}^T(\underline{\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2}) \\ &= \begin{bmatrix} \underline{GD} & -\underline{C}^T \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2} \end{bmatrix}. \end{aligned} \tag{8.15}$$

Combining (8.15) with the constraints in (8.3) on augmented form, we obtain

$$\begin{bmatrix} \underline{G\mathbf{d}} \\ O \end{bmatrix} = \begin{bmatrix} \underline{GD} & -\underline{C}^T \\ \underline{C} & O \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2} \end{bmatrix}, \tag{8.16}$$

where O denotes a zero matrix of appropriate size (here $q \times q$). To find a solution to (8.16), the equation is written as $A\mathbf{x} = \mathbf{b}$ for short notation, hence

$$A = \begin{bmatrix} \underline{GD} & -\underline{C}^T \\ \underline{C} & O \end{bmatrix}.$$

Provided that A is invertible, i.e.

$$A^{-1} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix},$$

the equation $A\mathbf{x} = \mathbf{b}$ has an unique solution given as $\hat{\mathbf{x}} = A^{-1}\mathbf{b}$. Therefore, the MLE is a solution to the equation in (8.16), so multiplying (8.16) with A^{-1} we can isolate $\mathbf{x}$ and find the MLE

$$\begin{bmatrix} \underline{\mathbf{x}} \\ \underline{\boldsymbol{\lambda}_1 - j\boldsymbol{\lambda}_2} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} \underline{G\mathbf{d}} \\ O \end{bmatrix} \quad \Rightarrow \quad \underline{\hat{\mathbf{x}}} = F_{11}\underline{G\mathbf{d}}.$$

The MLE of $\mathbf{x}$ can be found using only the upper left block matrix F_{11} of A^{-1}, assuming A is invertible. Therefore, by the matrix inversion formula in Theorem A.3, we obtain

$$\begin{aligned} F_{11} &= (\underline{GD})^{-1} - (\underline{GD})^{-1}\underline{C}^T(O + \underline{C}(\underline{GD})^{-1}\underline{C}^T)^{-1}\underline{C}(\underline{GD})^{-1} \\ &= (\underline{GD})^{-1}\left(I - \underline{C}^T\left(\underline{C}(\underline{GD})^{-1}\underline{C}^T\right)^{-1}\underline{C}(\underline{GD})^{-1}\right). \end{aligned}$$

Note that the inversion formula only applies if $\underline{GD}$ and $\underline{C}(\underline{GD})^{-1}\underline{C}^T$ is invertible. However, this will not be the case when one or more smart meters are removed. In that case F_{11} is extracted directly from A^{-1}, which is only possible until too many smart meters are removed.

Lastly, we want to achieve the distribution of the ML estimator and the confidence interval, starting with the ML estimator given by

$$\hat{\underline{\mathbf{x}}}(\underline{\mathbf{D}}) = F_{11}\underline{G}\underline{\mathbf{D}}.$$

The mean of the estimator can be calculated as

$$\begin{aligned} \mathrm{E}[\hat{\underline{\mathbf{x}}}(\underline{\mathbf{D}})] &= F_{11}\underline{G}\mathrm{E}[\underline{\mathbf{D}}] \\ &= F_{11}\underline{GD}\mathbf{x}. \end{aligned}$$

To check whether the estimator is unbiased, consider $A^{-1}A$

$$\begin{aligned} A^{-1}A &= \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}\begin{bmatrix} \underline{GD} & -\underline{C}^T \\ \underline{C} & O \end{bmatrix} \\ &= \begin{bmatrix} F_{11}\underline{GD} + F_{12}\underline{C} & -F_{11}\underline{C}^T \\ F_{21}\underline{GD} + F_{22}\underline{C} & -F_{21}\underline{C}^T \end{bmatrix} \\ &= I. \end{aligned}$$

Using that $F_{11}\underline{GD} + F_{12}\underline{C} = I$ and $F_{12}C\mathbf{x} = \mathbf{0}$, then

$$\begin{aligned} \mathrm{E}[\hat{\underline{\mathbf{x}}}(\underline{\mathbf{D}})] &= (F_{11}\underline{GD} + F_{12}\underline{C})\underline{\mathbf{x}} \\ &= I\underline{\mathbf{x}} \\ &= \underline{\mathbf{x}}. \end{aligned}$$

Hence, the ML estimator is unbiased. Then, the variance is calculated as

$$\begin{aligned} \mathrm{Var}[\hat{\underline{\mathbf{x}}}(\underline{\mathbf{D}})] &= F_{11}\underline{G}\mathrm{Var}[\underline{\mathbf{D}}]\underline{G}^H F_{11}^H \\ &= F_{11}\underline{G\Sigma G}^H F_{11}^H. \end{aligned}$$

Then the distribution of the ML estimator is

$$\hat{\underline{\mathbf{x}}}(\underline{\mathbf{D}}) \sim N_{\mathbb{C}}\left(\underline{\mathbf{x}}, F_{11}\underline{G\Sigma G}^H F_{11}^H\right).$$

The confidence intervals can then be found by Theorem 3.18 as

$$\mathrm{Re}(\hat{\mathbf{x}}_k) \pm z_{1-\alpha/2}\sqrt{(F_{11}\underline{G\Sigma G}^H F_{11}^H)_{kk}}, \quad \mathrm{Im}(\hat{\mathbf{x}}_k) \pm z_{1-\alpha/2}\sqrt{(F_{11}\underline{G\Sigma G}^H F_{11}^H)_{kk}},$$

where $k = 1, \ldots, K$. Now that the model has been derived, it will be used to conduct various experiments on various tree structures.

9 Simulation

In this chapter, the derived model based on the previously presented theory will be simulated. We will construct a small and large tree structure to simulate the model. To increase comparability, both trees will be made symmetrical.

9.1 Specifications

Data specifications are required before simulating each tree structure.

- The initial voltage v_0 will always be set to 400 V.
- The terminating currents are simulated randomly between $5 - 15$ A since the main fuse in a detached house are usually 16 A [4].
- The conductors are simplified such that they all contain $0.005\ \Omega$.
- Knowing the initial voltage, the terminating currents, and the impedance of the conductors, the rest of the voltages and currents in the trees are calculated by the linear equations (8.1) and (8.2).
- We assume that the smart meters have a measurement accuracy of 99%. Therefore, the measurement errors ϵ are simulated as varying $\pm 1\%$ for both the voltages and currents.
- We choose to have zero phase resulting in real values for the voltages and currents.
- The confidence region in all of the experiments is a 95%-confidence region.

Having specified the simulation, we will start by constructing a small tree.

9.2 Small Tree

The purpose of simulating a small tree is to ensure a well functioning script, which can be expanded to simulate larger trees. The small tree can be seen in Figure 9.1.

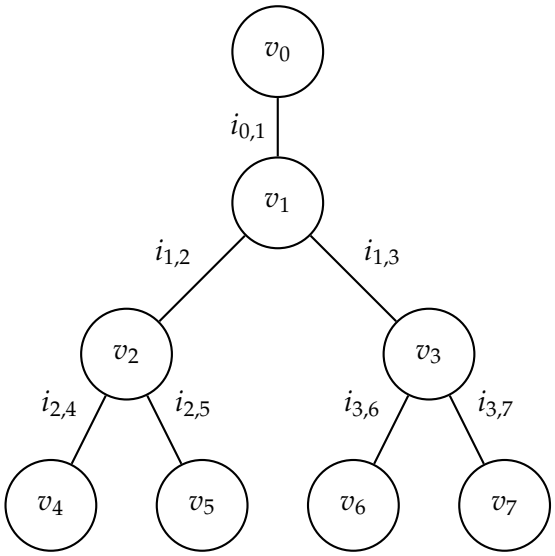


Figure 9.1: A small tree structure with 8 voltages and 7 currents.

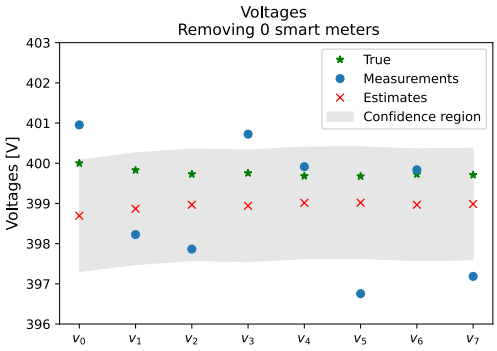
The script generates the C matrix as

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.005 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

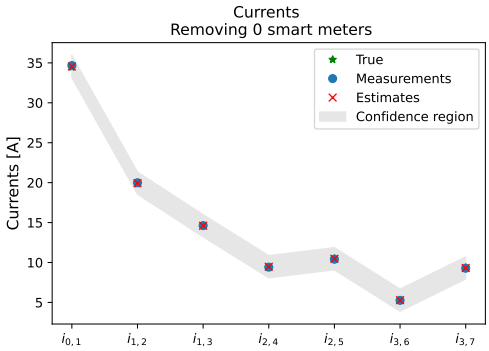
The corresponding $\mathbf{x}$ is

$$\mathbf{x} = \begin{bmatrix} v_0 & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & i_{0,1} & i_{1,2} & i_{1,3} & i_{2,4} & i_{2,5} & i_{3,6} & i_{3,7} \end{bmatrix}^T.$$

The first experiment simulates the small tree without removing any smart meters. The resulting plots can be seen in Figure 9.2.



(a) True values, measurements, estimates, and confidence region for the 8 voltages.



(b) True values, measurements, estimates, and confidence region for the 7 currents.

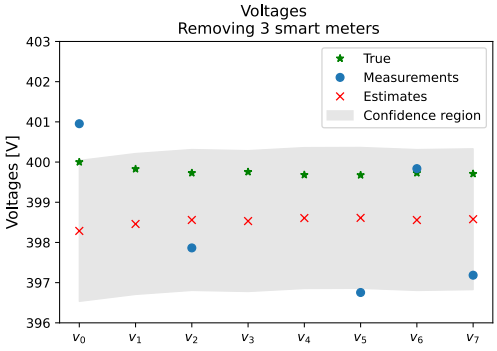
Figure 9.2: Voltages and currents for the small tree.

In Figure 9.2a, all of the true values and estimates are inside the confidence region, while some of the measurements are outside the confidence region. However, this is expected due to the added measurement errors. In Figure 9.2b, the true values, estimates, and measurements are inside the confidence region. Furthermore, they are all approximately on top of each other, with a very small difference. Since the currents are smaller values than the voltages, the measurement errors are corresponding smaller for the currents. Therefore, all of the measurements are within the confidence region for the currents.

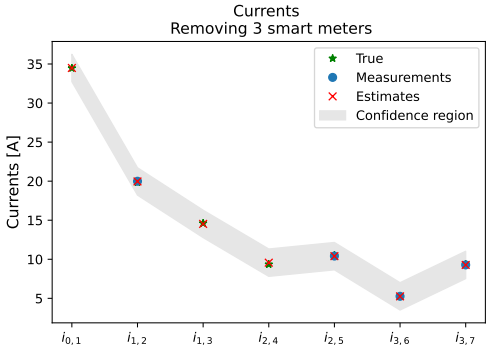
The coefficient of determination for the model is

$$R^2 \approx 0.99,$$

for both the voltages and currents, meaning that our model estimates the true values well. To test how robust our model is, we tried removing smart meters. Each smart meter contains information about a voltage and a current. In Figure 9.3, we remove 3 smart meters from the small tree.



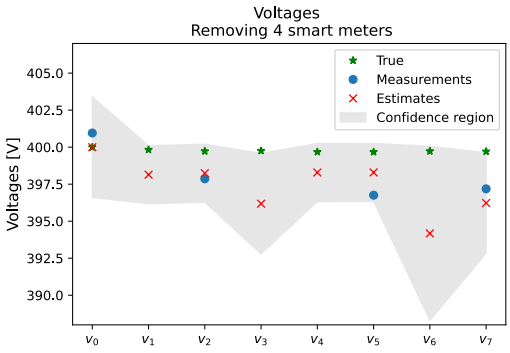
(a) True values, measurements, estimates, and confidence region for the voltages with 3 smart meters removed.



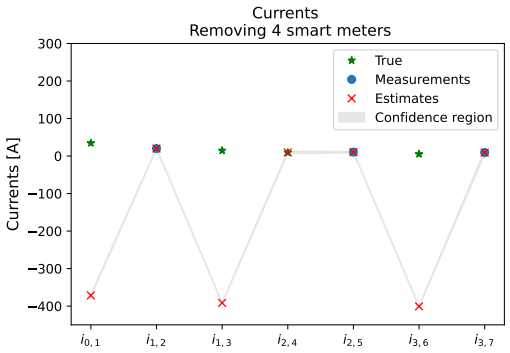
(b) True values, measurements, estimates, and confidence region for the currents with 3 smart meters removed.

Figure 9.3: Voltages and currents for the small tree with 3 smart meters removed.

The measurements of the voltages v_1, v_3 and v_4 with the corresponding currents $i_{0,1}, i_{1,3}$ and $i_{2,4}$ are removed in Figure 9.3a, and Figure 9.3b. The true values and estimates are still within the confidence region, however, it can be seen that the confidence region is wider than before when 0 smart meters were removed. If we remove 4 smart meters we obtain the plots in Figure 9.4.



(a) True values, measurements, estimates, and confidence region for the voltages with 4 smart meters removed.

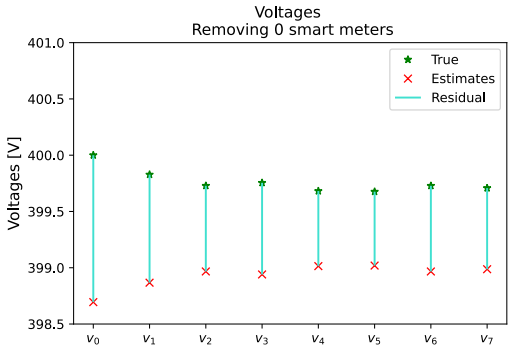


(b) True values, measurements, estimates, and confidence region for the currents with 4 smart meters removed.

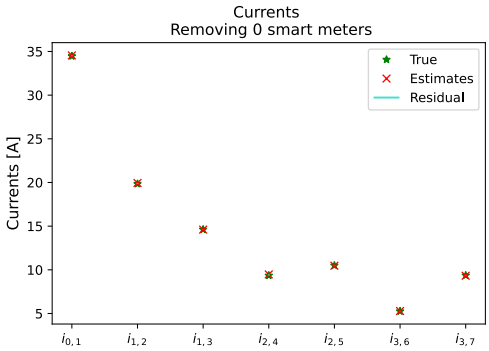
Figure 9.4: Voltages and currents for the small tree with 4 smart meters removed.

The measurements of the voltages v_1, v_3, v_4 and v_6 are removed in Figure 9.4a. This results in 2 of the true values falling outside the confidence region. Furthermore, the confidence region is wider and more asymmetrical as for the case when 3 smart meters were removed. The corresponding measurements of the currents $i_{0,1}$, $i_{1,3}$, $i_{2,4}$ and $i_{3,6}$ are simultaneously removed. In Figure 9.4b, it can be seen that 3 of the true values are far outside the confidence region. Furthermore, the confidence region is very asymmetric, and the model becomes very unlikely to predict future values.

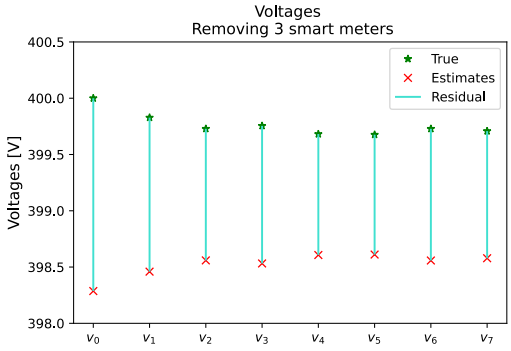
By removing different smart meters from the script, we observed that it was not possible to remove measurements from the nodes v_4 and v_5 or v_6 and v_7 simultaneously, since they are the only 2 children of a parent to a terminating node. This resulted in a singular matrix, and then the model does not work. To further analyse our model, the residuals for all three experiments are plotted in Figure 9.5.



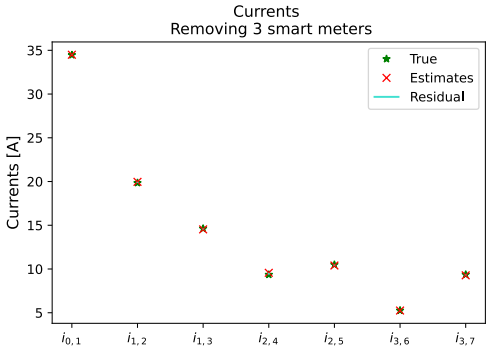
(a) Residuals for the voltages when removing 0 smart meters.



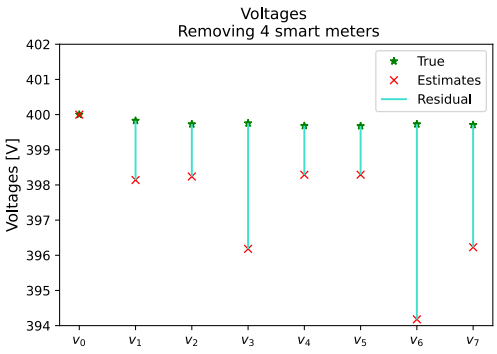
(b) Residuals for the currents when removing 0 smart meters.



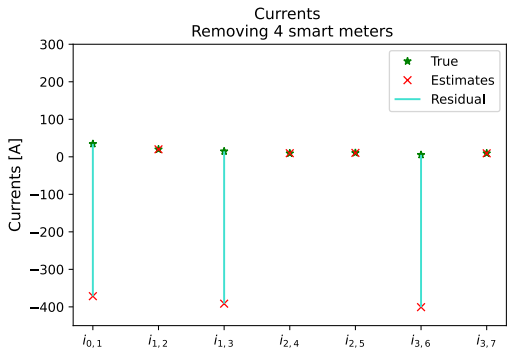
(c) Residuals for the voltages when removing the measurements v_1, v_3 and v_4.



(d) Residuals for the currents when removing the measurements $i_{0,1}, i_{1,3}$ and $i_{2,4}$.



(e) Residuals for the voltages when removing the measurements v_1, v_3, v_4 and v_6.



(f) Residuals for the currents when removing the measurements $i_{0,1}, i_{1,3}, i_{2,4}$ and $i_{3,6}$.

Figure 9.5: Plotted residuals.

The residuals between the true values and the estimates become larger as more smart meters are removed. However, the difference between the residuals in Figure 9.5a, and Figure 9.5c is relatively small compared to the residuals in Figure 9.5e. The residuals for the currents are almost impossible to spot on Figure 9.5b, and Figure 9.5d, because our estimates are very close to the true values. Removing 4 smart meters yields Figure 9.5f, where the residuals suddenly becomes very large at 3 out of 4 removed measurements. The absolute values for the residuals can be seen in Table 9.1.

Residuals			
Voltages	Removing 0 SMs	Removing 3 SMs	Removing 4 SMs
v_1	1.31	1.71	0.0026
v_2	0.96	1.37	1.69
v_3	0.76	1.17	1.49
v_4	0.81	1.22	3.57
v_5	0.67	1.07	1.39
v_6	0.66	1.06	1.38
v_7	0.76	1.17	5.55
v_8	0.72	1.13	3.48
Average	0.83	1.24	2.32
Currents	Removing 0 SMs	Removing 3 SMs	Removing 4 SMs
$i_{0,1}$	0.077	0.024	405.97
$i_{1,2}$	0.092	0.11	0.13
$i_{1,3}$	0.016	0.090	406.09
$i_{2,4}$	0.12	0.21	0.22
$i_{2,5}$	0.027	0.098	0.096
$i_{3,6}$	0.024	0.0071	406.02
$i_{3,7}$	0.041	0.083	0.077
Average	0.057	0.078	174.19

Table 9.1: Absolute values of the residuals between the true values and the estimates in the small tree.

The table verifies the observations from the plots, and now it is possible to observe that the residuals also increase for the currents between removing 0 and 3 smart meters. Furthermore, it can be seen that for the currents, the residuals increase by more than 400 for 3 of the removed measurements when removing 4 smart meters.

9.3 Large Tree

Having successfully tested the script for the small tree, a large tree is constructed. The large tree can be seen in Figure 9.6 and contains 128 voltages and 127 currents. The large tree can represent an electric grid in a small city or area.

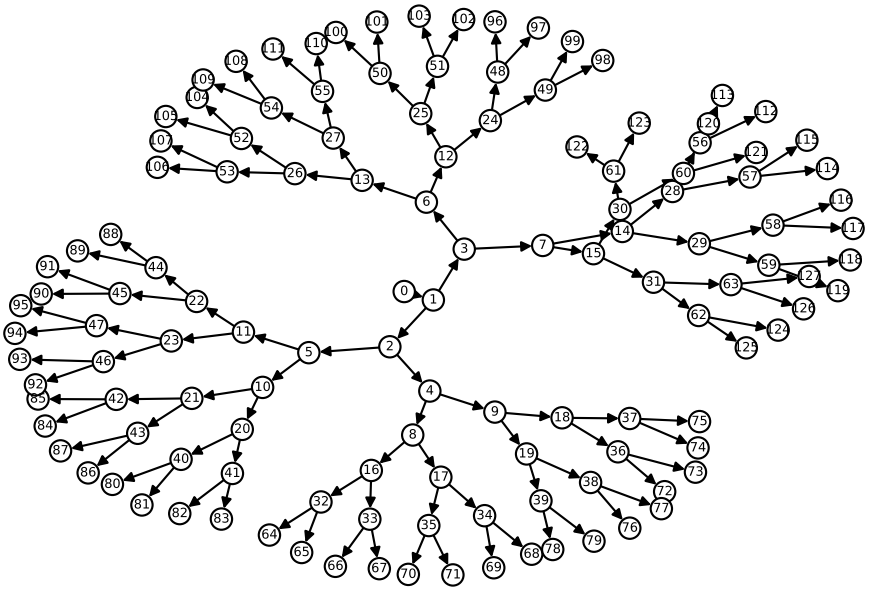
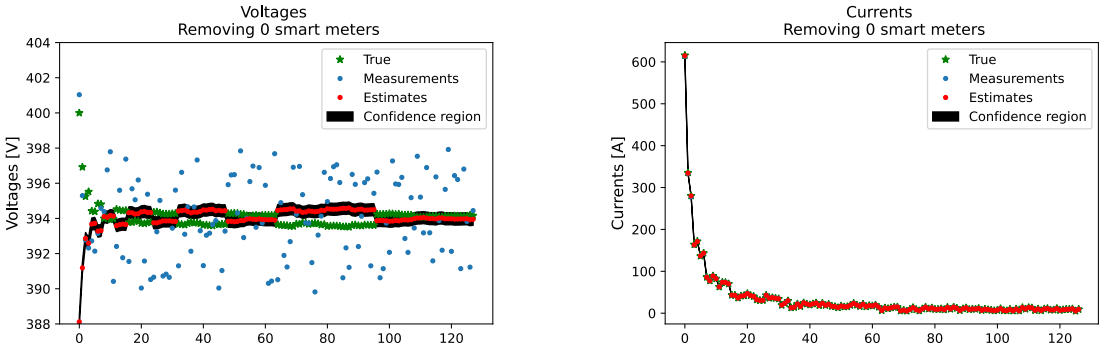


Figure 9.6: Large tree.

Having obtained measurements from all smart meters, we obtain the following plots in Figure 9.7.



(a) True values, measurements, estimates, and confidence region for the 128 voltages.

(b) True values, measurements, estimates, and confidence region for the 127 currents.

Figure 9.7: Voltages and currents for the large tree.

In Figure 9.7a, the true values and estimates are far apart at the first 4 estimates, however, afterwards they come closer to each other within ± 1 V. Most of the measurements are outside the confidence region, which is expected for the same reason as in the small tree.

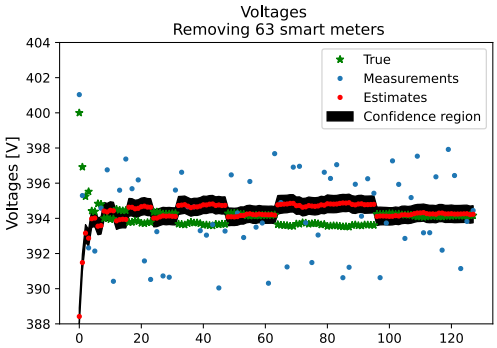
Not all of the true values are within the confidence region, but they are still close, apart from the first 4 values. In Figure 9.7b, all of the true values, measurements, and estimates are on top of each other and inside the confidence region. The coefficient of determination for the large tree is

$$R^2 \approx 0.99,$$

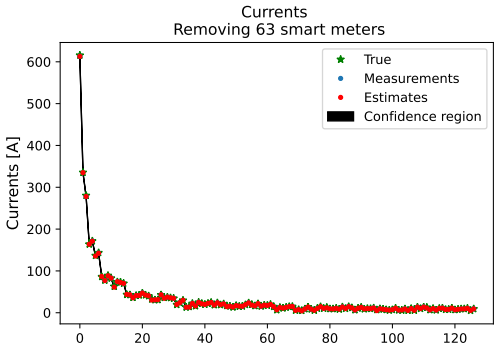
meaning that our model still estimates the true values well.

Removing 63 Smart Meters

To investigate the impact of removing certain smart meters from the large tree, we first wanted to remove a large amount of smart meters while maintaining decent estimates for the voltage and currents. Removing 63 smart meters results in Figure 9.8.



(a) True values, measurements, estimates, and confidence region for the voltages with 63 smart meters removed.



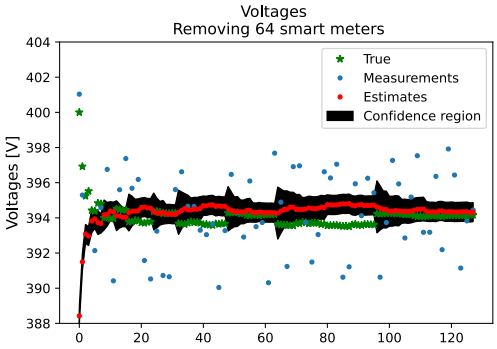
(b) True values, measurements, estimates, and confidence region for the currents with 63 smart meters removed.

Figure 9.8: Voltages and currents for the large tree with 63 smart meters removed.

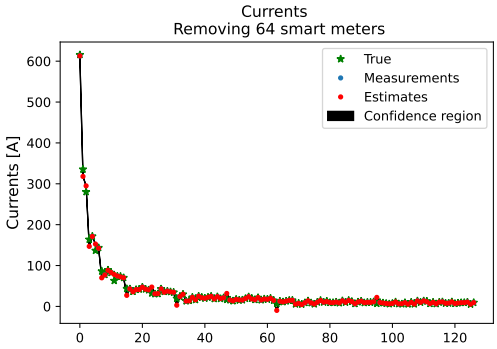
We have removed all the smart meters at the even numbered nodes, except the one at node 0. These plots are very similar to the ones in Figure 9.7, where we removed 0 smart meters. From this it can be seen that it is possible to remove almost half of the smart meters and still be able to obtain estimates close to the true values for both the voltages and currents.

Removing 64 Smart Meters - Version A

Now we will remove one more smart meter and see if it makes any difference where in the tree we choose to remove the smart meter from. Removing 64 smart meters results in Figure 9.9.



(a) True values, measurements, estimates, and confidence region for the voltages with 64 smart meters removed.



(b) True values, measurements, estimates, and confidence region for the currents with 64 smart meters removed.

Figure 9.9: Voltages and currents for the large tree with 64 smart meters removed, version A.

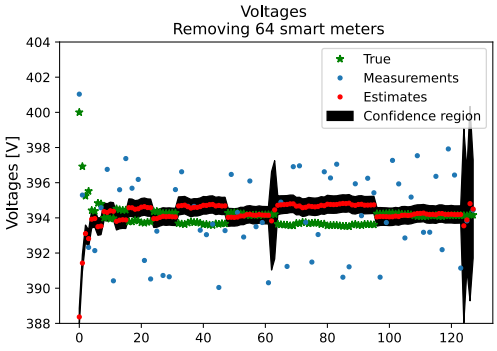
We remove the same smart meters as before plus an additional node close to v_0. Therefore we choose to remove node 3. For both the voltages and currents this results in a wider confidence region with more spikes than before, when we removed 0 smart meters. However, since we have removed half of the total amount of smart meters from the tree, this is expected. In Figure 9.9a, the distance between the true values and estimates are largest for the first 4 estimates. This was also the case when we did not remove any smart meters. After that, the true values and estimates come closer to each other within approximately ± 2 V. Furthermore, most of the true values and all of the estimates are inside the confidence region. In Figure 9.9b, the true values and estimates are still close to each other. Most of them are on top of each other. The confidence region contains most of the true values and all of the estimates. The model still estimates most of the voltages and almost all of the currents very well, even after removing half of the smart meters. The adjusted coefficients of determination are in this case

$$R^2_{Adj} \approx 0.93 \text{ (for the voltages)} \quad \text{and} \quad R^2_{Adj} \approx 0.99 \text{ (for the currents).}$$

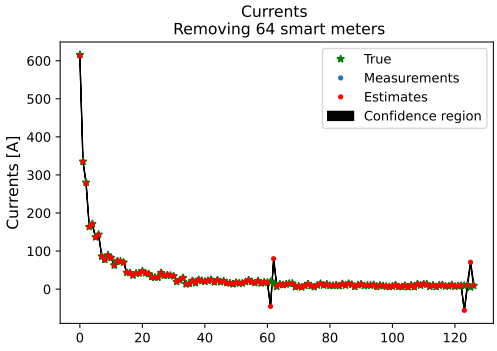
This supports the observation that the currents are better estimated by the model than the voltages, when removing 64 smart meters. The adjusted coefficients of determination are calculated to compare version A of removing 64 smart meters with version B.

Removing 64 Smart Meters - Version *B*

Now we want to remove a smart meter far away from v_0, without it being at a terminating node. Therefore, we remove the smart meter at node 63 instead of node 3. We obtain the plots in Figure 9.10.



(a) True values, measurements, estimates, and confidence region for the voltages with 64 smart meters removed.



(b) True values, measurements, estimates, and confidence region for the currents with 64 smart meters removed.

Figure 9.10: Voltages and currents for the large tree with 64 smart meters removed, version *B*.

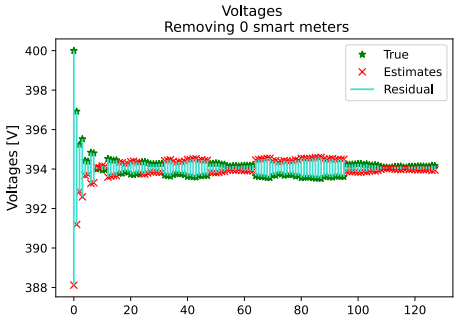
In Figure 9.10a, we obtain a confidence region with fewer but wider spikes. All of the estimates are included in the confidence region. More of the true values are outside the confidence region than before. However, the confidence region is more narrow than before, except at the spikes. In Figure 9.10b, we get a few larger spikes than before. However, the true values and estimates are still very close and more of them are on top of each other than before. All of the true values are inside the confidence region except 4 values corresponding to the estimates at the spikes. The adjusted coefficients of determination are in this case

$$R^2_{Adj} \approx 0.97 \text{ (for the voltages)} \quad \text{and} \quad R^2_{Adj} \approx 0.97 \text{ (for the currents).}$$

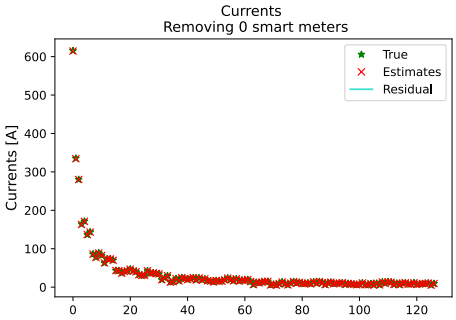
From this, we observe that removing a smart meter closer to the terminating nodes in the tree yields a more precise model to estimate the voltages, while the opposite is the case for the currents. This makes sense since the voltages in the tree are calculated from above, while the currents are calculated from below.

Residuals for the Large Tree

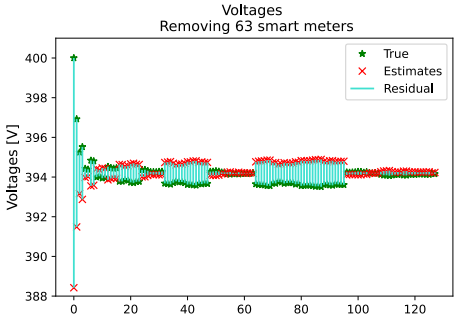
To further analyse our model the residuals for the large tree in all three reviewed cases are plotted in Figure 9.11.



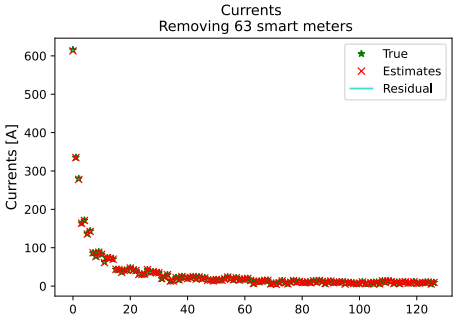
(a) Residuals for the voltages when removing 0 smart meters.



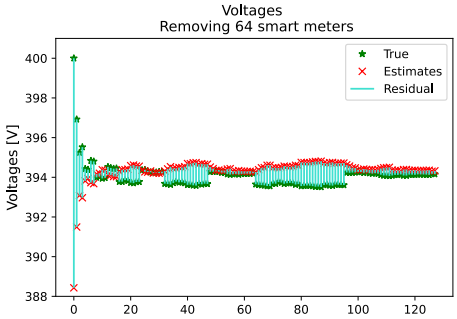
(b) Residuals for the currents when removing 0 smart meters.



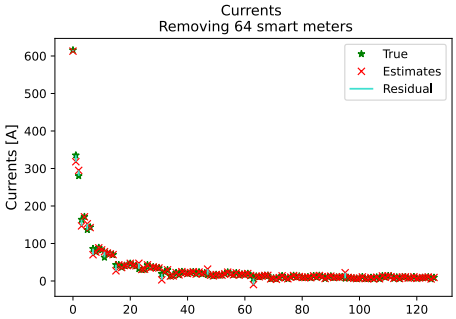
(c) Residuals for the voltages when removing 63 smart meters.



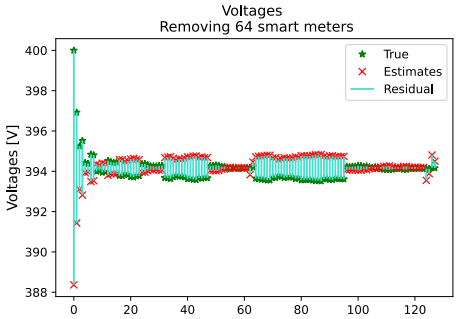
(d) Residuals for the currents when removing 63 smart meters.



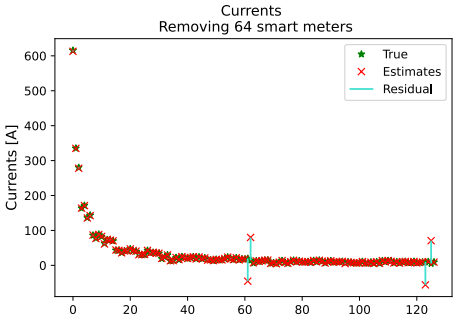
(e) Residuals for the voltages when removing 64 smart meters, version A.



(f) Residuals for the currents when removing 64 smart meters, version A.



(g) Residuals for the voltages when removing 64 smart meters, version B.



(h) Residuals for the voltages when removing 64 smart meters, version B.

Figure 9.11: Plotted residuals.

The residuals becomes larger when removing 63 compared to removing 0 smart meters, as expected. In Figure 9.11a, the residuals are largest at the beginning and then afterwards become smaller and more stable with few and small fluctuations. The residuals in Figure 9.11c, Figure 9.11e, and Figure 9.11g are also larger in the beginning and then become smaller and more stable. However, the fluctuations are bigger than in Figure 9.11a. In Figure 9.11b and Figure 9.11d, the residuals for the currents are almost impossible to spot. When removing 64 smart meters we begin to spot some residuals in Figure 9.11f, however, these are very small. In Figure 9.11h, the residuals are once again impossible to spot except at the four spikes we also observed earlier. The minimum, maximum, and average absolute values of the residuals can be seen in Table 9.2.

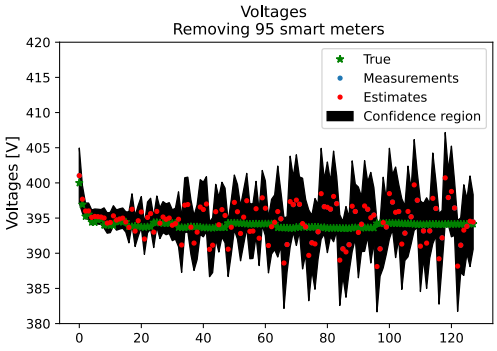
Residuals				
Voltages	Removing 0 SMs	Removing 63 SMs	Removing 64 SMs (A)	Removing 64 SMs (B)
Min value	0.020	0.014	0.0026	0.00071
Max value	11.88	11.58	11.57	11.63
Average	0.76	0.76	0.75	0.70
Currents	Removing 0 SMs	Removing 63 SMs	Removing 64 SMs (A)	Removing 64 SMs (B)
Min value	0.00010	0.0036	0.0039	0.0038
Max value	1.08	2.65	17.32	65.35
Average	0.34	0.34	1.73	2.39

Table 9.2: Minimum, maximum, and average absolute values of the residuals between the true values and estimates in the large tree. Version A of removing 64 smart meters include removing the smart meter from node 3. Version B of removing 64 smart meters include removing the smart meter from node 63.

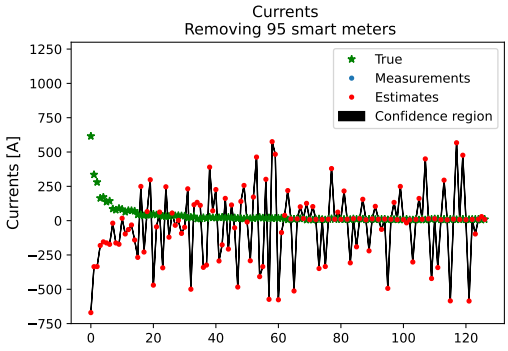
The values in the table support the previous observation, when removing a smart meter closer to a terminating node. This shows that the model has more difficulty estimating the currents, while the opposite is true for the voltages.

Limits of the Model

We wanted to see how many smart meters it was possible to remove from the large tree, while still being able to run the script and simulate data. We observed that we needed to keep a smart meter at node 0 and at one out of two terminating nodes from the same parent, to keep the script running. This leads to the maximum possible number of smart meters, we are able to remove, being 95. We removed the smart meters at all the even nodes and the uneven nodes from node $1 - 63$. The resulting plots can be seen in Figure 9.12.



(a) True values, measurements, estimates, and confidence region for the voltages with 95 smart meters removed.



(b) True values, measurements, estimates, and confidence region for the currents with 95 smart meters removed.

Figure 9.12: Voltages and currents for the large tree with 95 smart meters removed.

It is clearly seen that while the script still generates an output, the model is now very unlikely to be able to predict future values. In Figure 9.12a, this is due to the confidence region being very wide and containing lots of spikes. In Figure 9.12b, the estimates and true values are very far apart at some points. Furthermore, the confidence region also contains a lot of spikes. An estimate for the current that varies up to approximately 1400 A from the true value is definitely not acceptable. Therefore, it is not possible to remove 95 smart meters and still obtain a useful model.

We also tested our script for how large a tree it was possible to simulate. It was possible to simulate a tree with 2048 voltages and 2047 currents. We could also generate a tree structure for 4096 voltages and 4095 currents, however, it was not possible to simulate any data, due to the matrices becoming too large.

Future Work

If we were to continue working on this project, we would develop the script such that it would be more real-life applicable. The first objective would be to expand the values within the model from real to complex. This would achieve a more practical and realistic simulation.

The second objective would be to simulate even larger and non-symmetric tree structures. Various new experiments could be performed on the large non-symmetric tree. Therefore, it would be interesting to investigate the following problems.

1. How many smart meter would we be able to remove without exceeding a certain threshold of tolerance. This would be an optimisation problem, and decrease the total cost of the tree.
2. How well the estimation of voltages and currents would be if there was to be removed multiple smart meters connected alongside each other. This would mean that the tree structure would have a few "dead branches". This is inspired by the black out in Fredericia, as described in the introduction.
3. How to simulate realistic cables, based on their length, material, and heat, such that each node would obtain its own unique impedance and risk of being faulty.
4. How to detect a faulty cable within the tree structure.

Lastly, we would redesign the tree structure such that it no longer would represent a traditional electric grid, but a modern grid as designed in Figure 1.2. This would require extended research into modern electric grids.

10 Conclusion

For an electric grid, Kirchhoff's and Ohm's laws were used in association with statistics, complex normal distribution, optimisation, and complex differentiation to find estimates for the voltages and currents. A model was derived for the grid using MLE and the Lagrangian. Different tree structures were constructed to simulate the model and perform various experiments.

From the simulation, it can be concluded that it was possible to remove almost half of the smart meters from a grid and still obtain estimates very close to the estimates obtained when zero smart meters were removed. Furthermore, it was observed that the currents were better estimated than the voltages in the beginning. However, after removing more than half of the smart meters in the grid, the currents' residuals tend to be larger than the residuals of the voltages. Removing a smart meter placed closer to the terminating nodes in the large tree entailed a more precise model to estimate the voltages, while the opposite was the case for the currents. It was not possible to remove the smart meter at node 0 and two terminating nodes from the same parent in our script. If we removed all removable smart meters, we obtained confidence regions with many spikes and residuals so large that the model was no longer useful for predicting future values.

Overall, our model estimated the voltages and currents in the constructed grid well, up to a unspecified threshold. It would be possible to further expand the script to perform new experiments and add more complex trees than those constructed in this project.

Bibliography

[1] Andreas Antoniou and Wu-Sheng Lu. *Practical optimization*. 1. edition. Springer, 2007. ISBN: 9780387711065.

[2] Sheldon Axler. *Linear Algebra Done Right*. 3. edition. Springer, 2015. ISBN: 3319110799.

[3] Jack Casazza and Frank Delea. *Understanding Electric Power Systems: An Overview of the Technology and the Marketplace*. 1. edition. John Wiley & Sons, 2003. ISBN: 0471446521.

[4] Ny Eltavle. "Hvad er en hovedsikring?" In: (2022). Last accessed 5 December 2022. URL: https://www.ny-eltavle.dk/hvad-er-en-hovedsikring/.

[5] Energinet. "Energisystemet lige nu". In: (2022). Last accessed 16 November 2022. URL: https://energinet.dk/.

[6] Energistyrelsen. "Fra central til decentral elproduktion". In: (2017). Last accessed 15 November 2022. URL: https://ens.dk/sites/ens.dk/files/Statistik/foer_efter.pdf.

[7] Nynne G. Hallander. "Varmepumpe i stedet for gas: Sådan udskifter du fyret". In: (2022). Last accessed 29 September 2022. URL: https://www.3byggetilbud.dk/boligejer/varmepumpe-i-stedet-for-gas/.

[8] Michael B. Hinrichsen. "Det elektriske Danmark". In: (2020). Last accessed 29 September 2022. URL: https://greenpowerdenmark.dk/nyheder/elektriske-danmark.

[9] David J. Irwin and Mark R. Nelms. *Engineering Circuit Analysis*. 11. edition. John Wiley & Sons, 2015. ISBN: 9781118960639.

[10] Steven Kay. *Intuitive Probability and Random Processes using MATLAB®*. 4. edition. Springer Science & Business Media, 2006. ISBN: 9780387241579.

[11] Klima-, Energi- og Bygningsministeriet. "Smart Grid-Strategi - fremtidens intelligente energisystem". In: (2013). Last accessed 22 November 2022. URL: https://ens.dk/sites/ens.dk/files/EnergiKlimapolitik/smart_grid-strategi-2.pdf.

[12] Henrik Madsen and Poul Thyregod. *Introduction to General and Generalized Linear Models*. 1. edition. CRC Press, 2011. ISBN: 9781420091557.

[13] Peter Olofsson and Mikael Andersson. *Probability, statistics, and stochastic processes*. 2. edition. John Wiley & Sons, 2012. ISBN: 9780470889749.

[14] Peter J. Schreier and Louis L. Scharf. *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. 1. edition. Cambridge University Press, 2010. ISBN: 9780511815911.

[15] Peter F. Siggaard. “Elbiler og varmepumper udfordrer elnettet. Der skal investeres milliarder for at undgå strømsvigt”. In: (2022). Last accessed 29 September 2022. URL: https://www.dr.dk/nyheder/regionale/trekanten/elbiler-og-varmepumper-udfordrer-elnettet-der-skal-investeres-milliarder.

[16] John R. Silvester. “Determinants of Block Matrices”. In: (1999). Last accessed 30 November 2022. URL: https://hal.archives-ouvertes.fr/hal-01509379/document.

[17] Birgitte Wistoft, Carl-Erik Lundgren, and Ove W. Dietrich. “Elforsyning”. In: (2014). Last accessed 16 November 2022. URL: https://denstoredanske.lex.dk/elforsyning.

[18] Hugh D. Young and Roger A. Freedman. *University Physics: with Modern Physics*. 12. edition. San Francisco: Addison Wesley, 2008. ISBN: 9780321501219.

A Linear Algebra

Theorem A.1 (Determinant of Block Matrix)
If a block matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where A, B, C and D are $n \times n$ matrices and $CD = DC$, then

$$\det(M) = \det(AD - CB).$$

[16, p. 4]

Theorem A.2 (Multiplicative Determinant)
If A and B are square matrices of the same size, then

$$\det(AB) = \det(BA) = \det(A)\det(B).$$

[2, p. 317]

Theorem A.3 (Matrix Inversion)
Let a block matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

be invertible. If A and $D - CA^{-1}B$ is invertible, then

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}.$$

[14, p. 275]