



Universiteit
Leiden
The Netherlands

Electrical grid failures: An analysis of rare events

Vasmel, N.

Citation

Vasmel, N. (2019). *Electrical grid failures: An analysis of rare events*.

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository](#)

Downloaded from: <https://hdl.handle.net/1887/3596949>

Note: To cite this publication please use the final published version (if applicable).

Electrical grid failures

An analysis of rare events

Niek Vasmel

June 7, 2019

Contents

Preface	3
1 Introduction	4
1.1 Structure of the power grid	4
1.2 Operating the grid	7
1.3 Thesis overview	7
2 A Mathematical Model of the Power Grid	9
2.1 AC circuits	9
2.1.1 Single-phase AC power supply	10
2.1.2 Electrical components	12
2.1.3 Complex power	13
2.1.4 Three-phase AC power supply	15
2.1.5 Load-flow equations	16
2.2 DC-approximation	18
2.3 Feasible Power Flows and Power Injections	20
3 Monte Carlo Simulation and Importance Sampling	21
3.1 Monte Carlo simulation	21
3.2 Importance sampling	23
3.3 Exponential tilting	26
3.4 Estimator efficiency and large deviations theory	27
4 The DC-Approximation with Stochastic Power Injections	30
4.1 Preliminaries	31
4.2 Shifted mean	35
4.3 Gaussian in the exterior of a polyhedron	40
4.4 Simulation	45
4.4.1 Model	45
4.4.2 Results	47
5 Distflow	50
5.1 Distflow	50
5.2 Telescoping series	51
5.3 Simulation results	53
6 A Continuous-Time Approach to the 2-Bus System	54
6.1 Wiener processes and the Heat equation	54
6.2 Wiener process exiting an interval	58

6.3	Wiener process exiting an elliptical domain	62
6.4	An example	68
A	Matlab code	74
A.1	MATLAB code used in Section 4.4	74

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics. It is based on my own research performed at both Leiden University and CWI (Centrum Wiskunde & Informatica). I wish to thank my supervisors Floske Spieksma (Leiden) and Bert Zwart (CWI) for their guidance in establishing this thesis. Furthermore, I would also like to express my gratitude to Chang-Han Rhee and Tomasso Nesti for their help and feedback.

Chapter 1

Introduction

The electrical grid is essential to modern society distributing power from power plants to consumers. Blackouts and also local failures in the grid are hurtful to the economy and cause discomfort to affected people. To avoid this, grid operators are to a certain degree enabled to actively influence the conditions of the power grid. They may activate capacitors or ask power plants to increase or decrease their power generation. The grid operators aim to keep the characteristics of the grid, such as frequency and voltage at a relatively steady level. Furthermore, they also have to try and avoid an abundance of electrical power flowing through transmission lines, which could cause an overload.

Due to the emergence of renewable energy sources the power supply has become far less predictable. Sudden surges in generated energy will more often occur as a result of weather phenomena. Increased solar power or strong winds will be the main cause of this. Where the steady power supply of fossil fuel or nuclear power plants leads to more constant conditions within the grid, the highly volatile nature of renewable energy sources leads to more rapidly changing conditions. As a consequence, the grid operators have to be able to anticipate faster to the challenge that changing conditions in the power grid bring. Their means of control over the power grid vary with different sources of energy.

In order to avoid unfavourable operating conditions, it becomes more and more necessary to not only react to changing conditions in the grid but also to predict them and act proactively. In this thesis, we will use the tools of *rare event analysis* to determine how small the probability of some specific unfavourable operating conditions occurring in a certain time period will be. That way the grid operator will be made aware of the risk and be enabled to prevent any undesired scenario.

Naturally, it is desirable for this probability to be extremely small. We will also see what happens when we consider the limiting case where this probability becomes infinitesimally small. This will lead us to some concepts seen in *large deviations theory*.

1.1 Structure of the power grid

The power grid connects power stations to industrial, commercial and domestic consumers via a large network of transmission and distribution lines. In general, the power grid is considered to consist of two main components, the transmission grid and the distribution grid, see Figure 1.1.

Generator stations are connected to the aluminium alloy transmission lines at transformer stations (or substations). At this transformer station, the voltage is increased and the current is decreased. This is done to reduce power losses during transport, as the

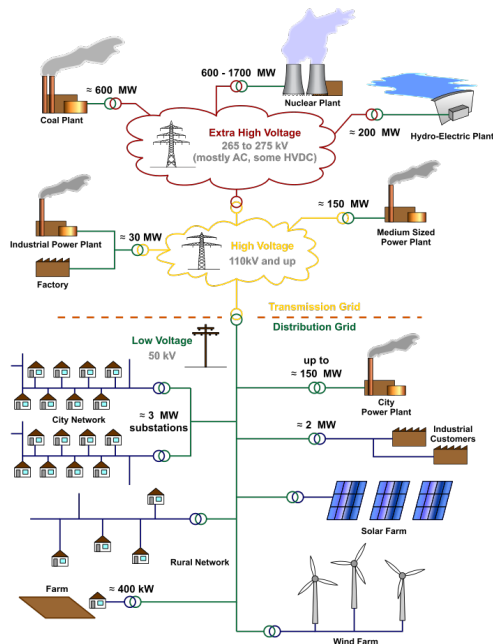


Figure 1.1: A schematic representation of the transmission and the distribution grid and their connected components. At some of the substations a rough indication of the power being transformed is also given, either in megawatts (MW) or kilowatts (kW).
 Source: https://upload.wikimedia.org/wikipedia/commons/9/90/Electricity_Grid_Schematic_English.svg.

power loss is directly related to the square of the current. The transmission lines carry high voltage electricity, typically 110 kilovolts or higher. In 2015 the transmission line losses in different European countries varied from 0.89% to 2.77% of the total power [1]. The transmission lines can be either overhead lines, well known by the supporting electricity pylons decorating many a landscape, or underground power cables. Transmission lines transporting different voltages may be interconnected at other substations. As an example of a transmission grid, the grid of the Netherlands is included as Figure 1.2.

Transmission lines then transport the high-voltage electricity (typically 110 kilovolts and higher) over long distances to other transformer stations that scale the voltage down to connect to the distribution grid.

Due to the lower voltages, the losses in the distribution network are higher. In 2015 the distribution line losses in different European countries varied from 2.24% to 10.44% of the total power [1]. The distribution network consists of the power lines that connect the transformers to factories, businesses and homes. Small local power plants may also be connected to the distribution network. Often there are several transformers in the distribution network to bring the voltage down to the level used in households.

It is good to note that while in geographical regions there essentially is one single interconnected transmission grid, there may be many distribution grids with each city having its own individual grid. These distribution grids also generally have fewer interconnections.

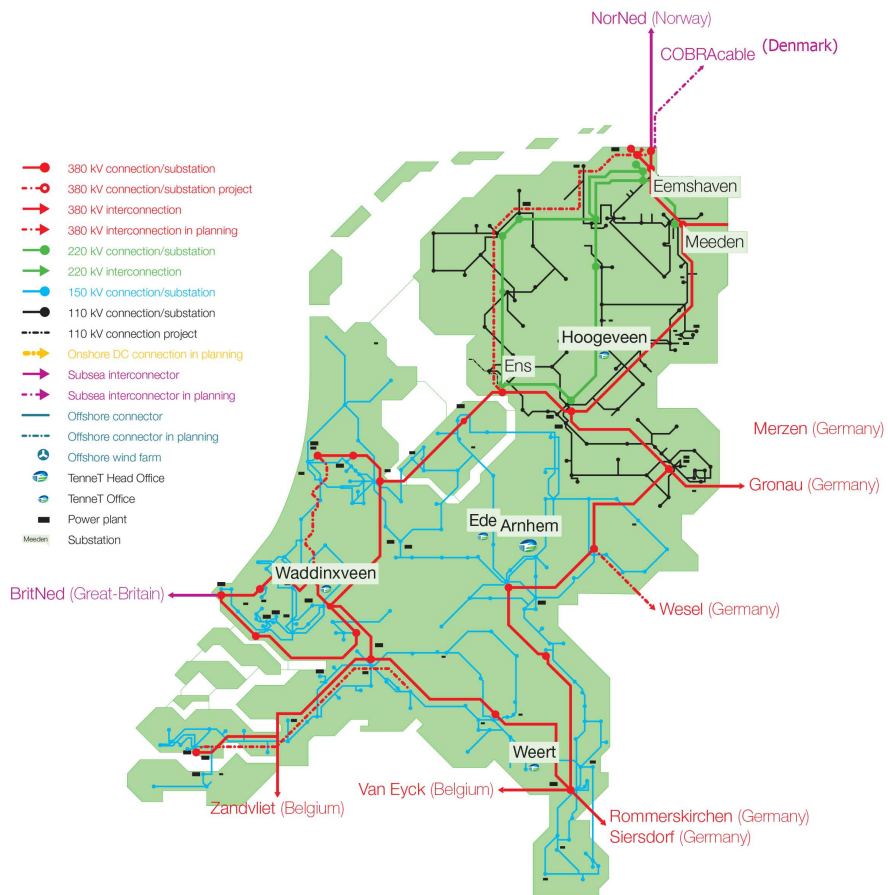


Figure 1.2: A schematic map of the Dutch transmission grid including the locations of power plants and substations. Source: adapted from the publicly accessible map on <https://www.tennet.eu/company/news-and-press/press-room/grid-maps/>.

1.2 Operating the grid

The main task of grid operators is to match the electric power generation to the demand within a certain control area. Currently, there are 43 grid operators active within the European Network of Transmission System Operators, each controlling a different part of the grid. There are still many more distribution grid operators. The grid operators have the authority to instruct power plant operators to increase or decrease their power generation as required. When the demand increases, the grid operators will generally prefer to start or increase generation at the plants where the associated costs are lowest. They may divert from this strategy to ensure the reliability of the grid.

Frequency regulation [2] is used to adjust the power generated to changes in power demand within a matter of minutes. This additional power is generated by *spinning-reserve* power plants, which are power plants that do operate, but not at full capacity, such that they may rapidly adjust their power output. When the spinning-reserves are addressed, the grid operator may also activate *replacement reserve* generators, which require a longer start-up time. This will allow the spinning-reserve operators to return to their previous level of power generation.

Grid operators also have to maintain a certain level of reserve capacity to match forecasts in demand. For example, on a daily base power generators are kept in reserve to activate when the workday commences in order to match the sudden demand of factories starting production. Furthermore, most generators have an automatic response to deal with sudden losses of supply in the network, also known as *frequency-response*. When a sudden loss of power in the system occurs, the remaining power being generated is less than the power being demanded. This would slow the generators down, lowering their operating frequency, were it not that the generators have an inbuilt system that governs their frequency. This response also results in increasing the power generated at these power stations.

In some cases, grid operators may, instead of increasing the power demand, request a major consumer to decrease their power consumption based on mutual agreements.

Grid operators have limited control over power originating from renewable energy sources such as wind and solar, since it is dependent on the current weather conditions. There exist means to deal with part of the fluctuations of renewable energy by integrating battery energy storage systems into the network [3]. Preferably, these storage systems are closely located to the renewable energy power plants. However, generally, renewable energy is simply used as it becomes available, since the operating costs of renewable energy generators are very low.

1.3 Thesis overview

Chapters 2 and 3 form the theoretical framework of this thesis. They are a review of existing literature and scientific articles.

First, in Chapter 2 we will introduce a mathematical model of the power grid. This chapter aims to show how mathematical equations governing the power grid naturally arise as a result of the physical laws of electrodynamics governing the flow of power. Furthermore, we will also give a frequently applied approximation (DC-approximation) of the model, which allows for the linearization of the equations governing the flow of power within the model. In the last part of this chapter, we will discuss how grid operating conditions relate to restrictions we can put on the mathematical model.

In Chapter 3 we will introduce the theory of importance sampling. We will later on use simulation to determine the probability of ending up in unfavourable operating conditions. This probability is typically very small and classified as a *rare event*. Ordinary sampling techniques such as Monte Carlo simulation are not well equipped to deal with rare events, but importance sampling is.

Subsequently, in Chapter 4 we will apply the techniques of importance sampling to the model of the power grid using the DC-approximation. A stochastic model will be used that models the power generation and demand at all of the generators and consumers as a single multivariate Gaussian distribution. In the development of this model we will prove two new theorems, 4.2.2 and 4.3.1. The results of these simulations will be discussed.

In Chapter 5 we will present the equations governing the power grid in the specific case of line networks. These form an alternative version of Distflow equations [4, 5]. They can be used to recursively determine the state of the power grid. This will allow us to present a novel method to determine whether voltages at certain connections along a line network will be within the desired range.

In Chapter 6 we will study the most elementary power grid: a single power generator connected to a single consumer. In this system, we can investigate an aspect of the power grid that is usually ignored. In determining the state of the power grid after a certain time interval, there are usually no restrictions based on the state within this interval. A Wiener process will be used to model the power generation and demand. The results will be compared to the situation of a Gaussian distribution modelling the power at the substations.

Chapter 2

A Mathematical Model of the Power Grid

This chapter is addressed to mathematicians wishing to familiarise themselves with the power grid. We model the power grid by a graph $G = (V, A)$, where the n nodes represent the *buses* and the m arbitrarily directed edges represent the power lines. A bus can be a site generating power, e.g. a power plant, or demanding power, e.g. a substation connected to a city's distribution network. In this chapter, we will build up this graph model as well as describe solution methods for a system of equations describing the behaviour of electricity within the power grid.

Although the power grid is in actuality an AC (Alternating Current) network, it is common practice to approach it as if it is a DC (Direct Current) network. This is because in an AC network the power flowing through the transmission lines is described by differential equations, while in a DC network it is described by linear equations making for easier and faster calculation. We will introduce both models and highlight the mathematical differences.

2.1 AC circuits

Most modern-day generators generate electrical energy by conversion of mechanical energy. This is done by rotating a magnetic field next to a field coil wound around a metal core to induce a current, see Figure 2.1. This type of generator is appropriately named

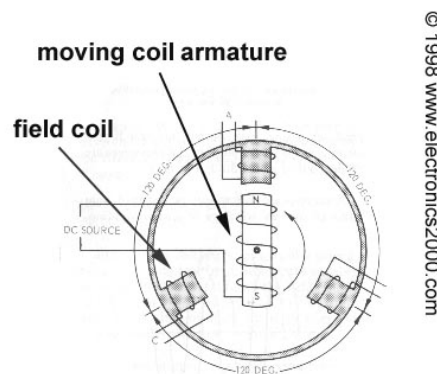


Figure 2.1: A representation of a three-phase generator. A source power rotates the central magnet inducing alternating currents in the three surrounding field coils.

an *alternator*. Many other power sources that, unlike turbines, are not directly connected to mechanical energy, also use alternators to generate electrical energy. For example, in fossil fuel power plants, chemical energy is first converted into mechanical energy via a combustion engine, after which an alternator is used.

As a consequence of their rotating nature, these alternators induce currents and voltages that are not time-independent, but instead of a sinusoidal nature.

A notable exception of a power source that does not use alternators, is solar energy. Solar panels provide far steadier currents. However, using a power inverter, the outgoing current can be transformed into a sinusoidal signal too. This is done to be able to connect to the power grid.

2.1.1 Single-phase AC power supply

Let us consider an alternating power source that supplies a voltage $V(t)$ and a current $I(t)$ with angular frequency $\omega = 2\pi f$ in radians per second and a phase difference between them of π . Here f is the frequency in Hertz, describing the number of cycles per second. The utility frequency in Europe is 50 Hz and in the United States, it is 60 Hz. The amplitudes of the voltage and the current are V_{\max} and I_{\max} respectively,

$$V(t) = V_{\max} \sin(\omega t), \quad (2.1)$$

$$I(t) = I_{\max} \sin(\omega t - \theta). \quad (2.2)$$

An overview of all relevant physical quantities with their accompanying standard units can be found in Table 2.1. This kind of generator is called a *single-phase generator*, the reason for which will become apparent later on, when we will consider the *three-phase generator*.

In the power grid, the transmission of electrical energy is continual. For this reason, we are more interested in the power $P(t)$, i.e. the rate at which energy is generated, sent and consumed, than in the total energy $E(t)$ produced, sent and consumed,

$$P(t) = \frac{dE(t)}{dt}.$$

Additional motivation for considering power rather than energy comes from the fact that physical limitations on the transfer of energy in an electrical circuit are also mostly based on the rate at which energy travels throughout a circuit.

Quantity	Symbol	Standard unit
Voltage	V	V (Volt)
Electric current	I	A (Ampère)
Electrical resistance	R	Ω (Ohm)
Energy	E	J (Joule)
Power	P	W (Watt)
Angular frequency	ω	rad s ⁻¹ (radians per second)
Magnetic flux	Φ	F (Farad)
Electric charge	Q	C (Coulomb)

Table 2.1: Physical quantities and their standard units.

The instantaneous power at any point in a circuit is the product of voltage and current,

$$P(t) = V(t) \cdot I(t). \quad (2.3)$$

This means, that the power supplied by the alternating power source is,

$$\begin{aligned} P(t) &= V_{\max} I_{\max} \sin(\omega t) \sin(\omega t - \theta) \\ &= \frac{1}{2} V_{\max} I_{\max} (\cos(\theta) - \cos(2\omega t - \theta)). \end{aligned}$$

At this point, we introduce the *effective* or root-mean-square value of the voltage and current,

$$\begin{aligned} |V| &= \sqrt{f \int_0^{1/f} V(t)^2 dt} = \frac{1}{\sqrt{2}} V_{\max}, \\ |I| &= \sqrt{f \int_0^{1/f} I(t)^2 dt} = \frac{1}{\sqrt{2}} I_{\max}. \end{aligned}$$

This allows us to rewrite the power in the following form,

$$\begin{aligned} P(t) &= |V||I| \cos(\theta) - |V||I| \cos(2\omega t - \theta) \\ &= |V||I| \cos(\theta)(1 - \cos(2\omega t)) - |V||I| \sin(\theta) \sin(2\omega t). \end{aligned}$$

By introducing two new variables,

$$P = |V||I| \cos(\theta), \quad (2.4)$$

and

$$Q = |V||I| \sin(\theta), \quad (2.5)$$

we can put the instantaneous power in a more compact form,

$$P(t) = P(1 - \cos(2\omega t)) - Q \sin(2\omega t).$$

Now we have accomplished splitting the instantaneous power into two distinct oscillating parts,

$$P(1 - \cos(2\omega t)), \text{ and } -Q \sin(2\omega t).$$

The instantaneous power $P(t)$ fluctuates around the average power, P , also called the *active power*. Conversely, Q is called the *reactive power*, since on average it does not contribute to the power transfer. Both P and Q are measured in standard units in Watts. However, to emphasise the distinction between the active and reactive power, it is the convention in electrical circuit engineering, to express the reactive power in Volt-Ampère-reactive (VAr), which is the same unit as a Watt (W), only by an other name.

The practical use of the newly defined variables P and Q will readily become apparent when we, later on, introduce the notion of *complex power*.

Instead of using the standard units, see Table 2.1, it is common practice in electrical engineering to introduce a unit system specifically based on the network of interest. Some unit values for complex power, S_{base} , and voltage, V_{base} , can be chosen freely. This then defines the unit value of the current via Equation 2.3 as

$$I_{\text{base}} = S_{\text{base}}/V_{\text{base}}. \quad (2.6)$$

This new unit system is also known as the *per-unit system*. In a suitable per-unit system, quantities are the same on both ends of a transformer. This simplifies calculations. It is important to be aware of the chosen base-values in the per-unit system, since in electrical engineering they are usually omitted from equations. The unit value for power, S_{base} , is referred to as the *base MVA* of the system, as it is typically expressed in MVA (mega-volt-ampère).

2.1.2 Electrical components

To create an electrical circuit, a closed loop of conductors has to be formed containing the power source. There are three main components in electrical circuits that do generally appear: *ideal resistors*, *inductors*, and *capacitors*; their graphical representations can be seen in Figure 2.2. Most loads are formed by combinations of these three main components. An ideal resistor R follows Ohm's law,

$$V(t) = R \cdot I(t). \quad (2.7)$$

This ensures that the voltage and current are in phase, i.e. $\theta = 0$, and the power is purely active. The effective voltage and current are related by,

$$|V| = R \cdot |I|.$$

Physically, resistance is related to how freely electrons may move within a material.

An ideal inductor operates on the following principle. By Ampère's circuital law, when a current runs through a conductor, it generates a magnetic field surrounding this wire. The total strength of this magnetic field surrounding the wire is called the *magnetic flux*, $\Phi(t)$. The magnetic flux is directly related to the current by a constant specific to the inductor L ,

$$L \cdot I(t) = \Phi(t).$$

The current $I(t)$, however, is sinusoidally alternating, and therefore the magnetic flux follows the same pattern. Now Faraday's law of induction states, that a changing magnetic field, in turn, induces a voltage difference,

$$V(t) = \frac{d\Phi(t)}{dt} = L \frac{dI(t)}{dt}.$$

Combining Faraday's law of induction with the equations for the voltage and current profile in Equations 2.1 and 2.2, we find,

$$|V| \sin(\omega t) = L \cdot |I| \cdot \omega \cos(\omega t - \theta).$$

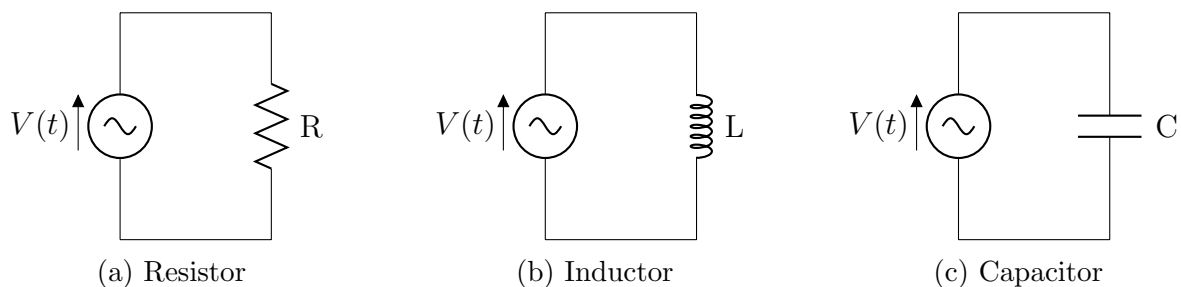


Figure 2.2: Closed electrical circuits containing an alternating voltage source and one of the three main circuit elements.

This causes the current to lag a phase of $\theta = \frac{\pi}{2}$ behind the voltage in case of a purely inductive load. The effective voltage and current are related by,

$$|V| = |I|L\omega. \quad (2.8)$$

An ideal capacitor can be regarded as the dual of an inductor. Physically, a capacitor corresponds to two conducting plates separated by a dielectric medium. The current $I(t)$ at any point in the circuit is defined as the rate of flow of an electric charge $Q(t)$ passing through it,

$$I(t) = \frac{dQ(t)}{dt}.$$

However, charge-carrying electrons cannot actually freely traverse the dielectric medium between the conducting plates of the capacitor. For each electron that accumulates on one plate, a different electron disperses from the opposite plate. This results in a surplus of electrons on one plate and a deficit on the other. Respectively, we now have a relatively positive charge on one plate and a negative charge on the other. The potential electric field between the two plates is directly related to the number of built-up electrons over time on the plates, i.e. the voltage is related to the integral of the current by some constant C specific to the capacitor,

$$V(t) - V(0) = \int_0^t I(s)ds.$$

Plugging the equations for the sinusoidal voltage and current profile into Equations 2.1 and 2.2, we find,

$$|V| \sin(\omega t) = -\frac{|I|}{C\omega} \cos(\omega t - \theta).$$

This causes the current to lead the voltage by a phase of $-\theta = \frac{\pi}{2}$ in case of a purely capacitive load. The effective voltage and current are related by,

$$|V| = \frac{|I|}{C\omega}.$$

We may conclude that the active- and reactive power, P and Q , attain the following values in the case of a simple AC circuit with only one component, either an ideal resistor (R), an ideal inductor (L), or an ideal capacitor (C):

$$\begin{aligned} P_R &= |V|^2/R, & P_L &= 0, & P_C &= 0, \\ Q_R &= 0, & Q_L &= |V|^2/(L\omega), & Q_C &= -|V|^2 \cdot C\omega. \end{aligned}$$

2.1.3 Complex power

Ideal resistors only influence the active power, and conversely ideal inductors and capacitors only influence reactive power. Therefore, when studying transmission lines we wish to be able to clearly describe the separate influence on the active and reactive power. It is possible to write the active and the reactive power respectively as the real and imaginary part of some complex power S ,

$$S = P + jQ.$$

We choose j as the notation for the imaginary unit to fall in line with the notation most commonly used in electrical engineering.

First, we introduce the notion of complex form for the voltage and current profiles, $\tilde{V}(t)$ and $\tilde{I}(t)$ respectively. These are consistent with Equations 2.1 and 2.2 via the relation $\text{Im} [\tilde{V}(t)] = V(t)$ and $\text{Im} [\tilde{I}(t)] = I(t)$,

$$\begin{aligned}\tilde{V}(t) &= |V| \exp(j\omega t), \\ \tilde{I}(t) &= |I| \exp(j(\omega t - \theta)).\end{aligned}$$

Now we find that we can simply write

$$S(t) = \tilde{V}(t) \cdot \tilde{I}^*(t), \quad (2.9)$$

where $\tilde{I}^*(t)$ denotes the complex conjugate of $\tilde{I}(t)$. For notational convenience we will from now on dispense with writing down the time dependency of the complex voltage and current, i.e. $S := S(t)$, $\tilde{V} := \tilde{V}(t)$, and $\tilde{I} := \tilde{I}(t)$.

Next, we also wish to introduce some notation that encompasses the influence on the complex power by the electrical components. This is called the *impedance* of the (not ideal) resistor,

$$Z = R + jX,$$

with R the *resistance* and X the *reactance*.

For an ideal resistor with resistance (R), an ideal inductor with inductance (L), and an ideal capacitor with capacitance C , we respectively define the following impedances $Z_R = R$, $Z_L = j\omega L$, and $Z_C = -j\frac{1}{\omega C}$. In this way, we ensure that the complex version of Ohm's law, Equation 2.7, is also valid,

$$\tilde{V} = Z \cdot \tilde{I}. \quad (2.10)$$

Note that we forego on writing the dependence of the impedance of a resistor on the angular frequency ω , as we will study systems where it is constant.

The per-unit system, see Equation 2.6, can alternatively be based on unit values for power, voltage, and impedance. By combining Equations 2.9 and 2.10, we find the relation,

$$Z_{\text{base}} = S_{\text{base}}/V_{\text{base}}^2.$$

When there are multiple resistors connected in series, see Figure 2.3a, each with impedance $Z_i, i = 1, \dots, n$ we can treat them as a single resistor with impedance

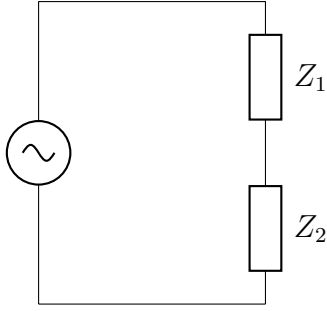
$$Z_{\text{series}} = \sum_{i=1}^n Z_i.$$

This is because the effective currents $|I_i|, i = 1, \dots, n$ passing through the resistors are all equal, and so the total complex power lost is,

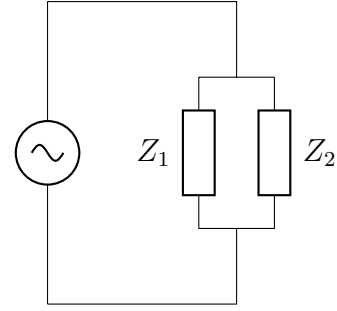
$$S = \sum_{i=1}^n \tilde{V}_i \cdot \tilde{I}_i^* = |I|^2 \sum_{i=1}^n Z_i.$$

We arrive at a similar result for resistors connected in parallel, see Figure 2.3b. The voltage differences $\tilde{V}_i, i = 1, \dots, n$ over all resistors are equal. Thus the total complex power lost is,

$$S = \sum_{i=1}^n \tilde{V}_i \cdot \tilde{I}_i^* = |V|^2 \sum_{i=1}^n \left(\frac{1}{Z_i} \right)^*.$$



(a) Two resistors in series



(b) Two resistors in parallel

Figure 2.3: Closed electrical circuits containing an alternating voltage source and one or more resistors.

Thus, we can treat parallel resistors as a single resistor with impedance,

$$Z_{\text{parallel}} = \frac{1}{\sum_i^n \frac{1}{Z_i}}.$$

As an alternative to working with the impedance Z , we will sometimes prefer to use its inverse,

$$Y = Z^{-1} = \frac{R - jX}{R^2 + X^2} = G + jB, \quad (2.11)$$

which is called the *admittance*. Its real part G is called the *conductance* and its imaginary part B is called the *susceptance*.

2.1.4 Three-phase AC power supply

A big shortcoming of the single-phase generator is, that the delivered power is not constant over time. A machine connected to the power source will receive power in pulses, reducing its operating efficiency. This can be avoided by supplying power onto three separate parallel lines, each with a phase difference of $2\pi/3$, see Figure 2.4,

$$\begin{aligned} \tilde{V}_a &= |V| \exp(j\omega t), \\ \tilde{V}_b &= |V| \exp(j(\omega t - 2\pi/3)), \\ \tilde{V}_c &= |V| \exp(j(\omega t - 4\pi/3)), \end{aligned}$$

where $|V|$ is the effective voltage supplied to each of the three lines.

If the applied load on each of the three parallel lines is equal, then it follows from symmetry that the currents will be symmetrical too, i.e. for some θ we have,

$$\begin{aligned} \tilde{I}_a &= |I| \exp(j(\omega t - \theta)), \\ \tilde{I}_b &= |I| \exp(j(\omega t - \theta - 2\pi/3)), \\ \tilde{I}_c &= |I| \exp(j(\omega t - \theta - 4\pi/3)), \end{aligned}$$

where $|I|$ is the effective current on each of the three transmission lines.

Now the combined instantaneous power supplied to the load is $P_a(t) + P_b(t) + P_c(t) = 3|V||I| \cos(\theta)$. Note that this is constant. Thanks to the symmetry between the three phases it suffices to determine the power, voltage and current via the relations $\tilde{V} = Z \cdot \tilde{I}$ and $\tilde{V} \cdot \tilde{I}^*$ for a single phase to solve the entire system.

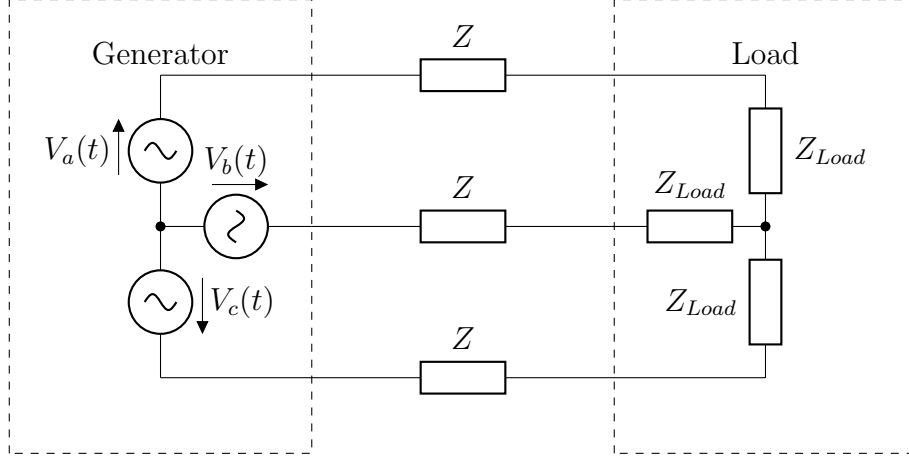


Figure 2.4: A three-phase generator connected to a load through transmission lines, each with equal impedance Z .

2.1.5 Load-flow equations

We are allowed to treat each interconnecting 3-phase transmission line between two buses i and k as a single one with admittance, say y_{ik} . Furthermore, at each bus i , power may not only be produced or consumed, but also be lost. This occurs due to losses in the transformer stations, mainly the inductive leakage. This loss is captured in the *shunt impedance* y_i of the bus. Combining Ohm's law $\tilde{V} = Z \cdot \tilde{I}$ with Kirchhoff's circuit laws yields a system of equations relating the voltages and currents at the junctions in an n -bus system.

Theorem 2.1.1 (Kirchhoff's Current Law). *For any bus i in the circuit, the sum of the currents \tilde{I}_{ik} of the lines flowing out of that bus is equal to the current injection \tilde{I}_i at that particular bus,*

$$\sum_{k \neq i} \tilde{I}_{ik} = \tilde{I}_i.$$

Current injection is the result of the supply or demand of power at a certain bus. The injected current is further also related to the voltage via Equation 2.3.

Theorem 2.1.2 (Kirchhoff's Voltage Law). *For any directed loop in the circuit the sum of the potential differences \tilde{V}_k between every two sequentially connected buses is zero,*

$$\sum_k \tilde{V}_k = 0.$$

Combining both of Kirchhoff's circuit laws we thus find for any single bus i ,

$$\tilde{I}_i = y_i(\tilde{V}_i - 0) + \sum_{k \neq i} y_{ik}(\tilde{V}_i - \tilde{V}_k).$$

For all buses combined this can compactly be written in matrix form,

$$\begin{pmatrix} \tilde{I}_1 \\ \tilde{I}_2 \\ \vdots \\ \tilde{I}_n \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & \dots & Y_{1n} \\ Y_{21} & Y_{22} & \dots & Y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{nn} \end{pmatrix} \begin{pmatrix} \tilde{V}_1 \\ \tilde{V}_2 \\ \vdots \\ \tilde{V}_n \end{pmatrix}, \quad (2.12)$$

where (Y_{ik}) is the *nodal admittance matrix*,

$$Y_{ik} = \begin{cases} y_i + \sum_{k \neq i} y_{ik}, & \text{if } i = k, \\ -y_{ik}, & \text{if } i \neq k. \end{cases}$$

Instead of considering the voltages provided at the buses, it turns out to be useful to consider the contribution of each bus to the power of the network. We speak of the *injection* of real and reactive power into the system at each bus. A negative real power injection corresponds to consuming power and a positive (resp. negative) reactive power injection corresponds to a backward (resp. forward) time phase shift of the current relative to the voltage.

For all i, k let $|Y_{ik}|$ be the magnitude of Y_{ik} and let α_{ik} be its argument, such that we can write $Y_{ik} = |Y_{ik}| \exp(j\alpha_{ik})$. Now by combining Equation 2.12 with Ohm's law, we find for any bus i ,

$$S_i = V_i \exp(j\theta_i) \left(\sum_{k=1}^n |Y_{ik}| V_k \exp(j(\theta_k + \alpha_{ik})) \right)^*. \quad (2.13)$$

We can split this into the real and the reactive parts

$$P_i = \sum_{k=1}^n V_i V_k |Y_{ik}| \cos(\theta_i - \theta_k - \alpha_{ik}), \quad (2.14)$$

$$Q_i = \sum_{k=1}^n V_i V_k |Y_{ik}| \sin(\theta_i - \theta_k - \alpha_{ik}). \quad (2.15)$$

These equations are known as the basic Load-flow equations. The power required at a substation or lost in transmission lines is sometimes referred to as a *load*. For an n -bus system there are $2n$ basic Load-flow equations with $4n$ variables: P_i, Q_i, V_i and θ_i . To solve the equations, we need to specify $2n$ variables. We classify each of the buses in the system to one of three categories.

PQ bus At these buses, loads are connected and therefore these buses are also called *load buses*. Generally, the values of both real and reactive loads connected at these buses are known. Therefore for a PQ bus i the values P_i and Q_i are specified and V_i and θ_i need to be calculated.

PV Bus Physically, these buses are generators. In general, the real power supplied by the generator is known and also, the magnitude of the voltage of the generator is maintained constant at a specified value. Therefore for a PV bus i the values P_i and V_i are specified and Q_i and θ_i need to be calculated.

Slack Bus To calculate the angles θ_i , a reference angle $\theta_i = 0$ is needed. Furthermore, physically, the total power supplied by all the generators must be equal to the sum of all loads in the system plus the system power loss. However, the system loss cannot be computed before the Load-flow equations are solved. Thus the real power output of all the generators in the system cannot be pre-specified. There should be at least one generator in the system that supplies the loss (plus its share of the loads). Thus, for this generator, the real power output cant be pre-specified. However, the voltage magnitude V can still be maintained at a constant specified level. Therefore for a slack bus i the values V_i and θ_i are specified and P_i and Q_i need to be calculated. Usually, the largest generator in the system is designated as the slack bus.

There are some iterative numerical methods to find solutions to the basic Load-flow equations. The most notable among these is the Newton-Raphson method [6]. However, solving these equations does in general not ensure a unique solution [7]. Omitting the time dependence of the system has led to this. For a time-dependent model, it is necessary to study the network as a system of interconnected mutually influencing oscillators. This can be encapsulated by a Kuramoto-like model known as the synchronous machine model [8]. The solutions to the Load-flow equations coincide with those of the synchronous machine model, though not all of them are stable. However, most solutions are very distinct and the simulation software packet MATPOWER [9] which we will use, has a strong preference to converge to the stable solution of interest.

Unfortunately, the required computing power for solving the Load-flow equations using the Newton-Raphson method grows extensively large as the grid size increases. Alternatively, a widely deployed method in practice is to approximate the Load-flow equations by a linearised version: the DC-approximation.

2.2 DC-approximation

Instead of using numerical methods to solve the basic Load-flow equations, we can also try to make some approximations that will result in a closed form solution. In this section, we will describe the popular DC-approximation scheme. It derives its name from the fact that the resulting system of equations looks very similar to that of direct current networks. The DC-approximation is based on the observation that for all transmission lines in the network the resistance is small compared to the reactance, $R < X$. Now, we approximate Equation 2.11 by,

$$y_{ik} \approx b_{ik}j,$$

where $b_{ik} = \text{Im}(y_{ik})$ is the susceptance of the transmission line between buses i and k , and $b_i = \text{Im}(y_i)$ is the shunt susceptance at bus i . This means that the nodal admittance matrix can be approximated by the nodal susceptance matrix,

$$B_{ik} = \begin{cases} b_i + \sum_{k \neq i} b_{ik}, & \text{if } i = k, \\ -b_{ik}, & \text{if } i \neq k. \end{cases}$$

Next, we can approximate the real Load-flow Equations 2.14 to

$$P_i = \sum_{k=1}^n V_i V_k B_{ik} \sin(\theta_i - \theta_k).$$

The reactive Load-flow Equations are ignored for the time being. Due to some of the assumptions we will make, the reactive power injections are completely determined by the real power injections.

Another observation that can be made, is that the angular difference between the voltage phasors of two connected buses is reasonably small. Typically, the difference is less than 15 degrees. Therefore, we approximate the sine to linear order, i.e. $\sin(x) \approx x$. This transforms the Load-flow Equations into a linear system of equations in the voltage

angles θ ,

$$\begin{aligned} P_i &= \sum_{k=1}^n V_i V_k B_{ik} (\theta_i - \theta_k) \\ &= \sum_{k \neq i} V_i V_k B_{ik} (\theta_i - \theta_k). \end{aligned}$$

We make one last observation: the voltage magnitudes throughout the system only vary slightly. Typically, the relative largest difference between any two buses in the system is less than 10 per cent. We therefore consider the voltage magnitude to be constant among all the buses. This allows us to set all the voltages to 1 in the per-unit system yielding,

$$P_i = \sum_{k \neq i} B_{ik} (\theta_i - \theta_k).$$

If the real power injections P_i are known, then we can find values for θ , that satisfy these equations. We notice that this system has the same solutions as

$$P_i = - \sum_{k \neq i} b_{ik} (\theta_i - \theta_k).$$

In matrix form this becomes

$$P = B'\theta,$$

where

$$B'_{ik} = \begin{cases} - \sum_{k \neq i} b_{ik}, & \text{if } i = k, \\ b_{ik}, & \text{if } i \neq k. \end{cases}$$

This matrix B' is the Laplacian of a weighted graph. In a connected network a Laplacian matrix has only 1 zero-eigenvalue [10]. We choose a reference angle which we set to zero, $\theta_1 = 0$. We remove θ_1 from θ yielding θ' , we remove P_1 from P yielding P' , and we remove the first column and row from B' yielding B'' . This matrix B'' is invertible. Thus we find,

$$\theta' = (B'')^{-1} P'.$$

Let A be the node-arc incidence matrix and let D be the diagonal matrix of the susceptances of each line, then the line flow of power is given by,

$$P_{\text{flow}} = -DA\theta.$$

Using the identities $\theta = (0, \theta')^\top$ and $P = (0, P')^\top$, these last steps can also be combined to find a single matrix V , the Power Transmission Distribution Factor (PTDF), that maps the power injections to the line flow of power,

$$P_{\text{flow}} = VP.$$

This method of approximating the Load-flow equations has the major advantage of resulting in a linear system of equations, which is easily solved.

There have been several analyses of the resulting differences between using the true alternating current model and the DC-approximation, see for example [11].

2.3 Feasible Power Flows and Power Injections

In our analysis of the power grid, we are mainly interested in whether or not the state of the network satisfies certain operating conditions. If the power flowing through transmission lines gets too high, they might overheat and burn through. If the voltage becomes too high or too low at a certain point in the network, then consumer devices may over-clock or underclock inadvertently. In a worse scenario, they could stop working or break down. Similar problems arise when the frequency in the network gets too high or low. An interesting related news fact: At the beginning of 2018 the frequency in southeastern Europe got too low, leading people to notice how their digital clocks started running behind, as their speed is designed to follow the frequency of the network [12]. Another problem that may occur is when a bus demands or generates too much power for itself to cope with, though this requires less focus on network analysis and more on the local bus.

In the DC-approximation the constraints are mainly based on the maximum amount of power flow through the power lines. Let n be the number of buses and m the number of transmission lines. Then we can use the Power Transmission Distribution Factor, an $m \times n$ matrix, to write

$$P_l = VP_b,$$

where P_l is the vector of power flowing through the lines, P_b is the vector of power injected into the buses, and V is the PTDF. This gives constraints $P_{\max} \geq |P_l|$ for feasible line flows. This is equivalent to restrictions on the injected power $|VP_b| \leq P_{\max}$.

Now, we have $2m$ linear restrictions on the injected power based on the maximum permissible power flow,

$$\begin{aligned} V_{i,\bullet}P_b &\leq (P_{\max})_i, & i = 1, \dots, m, \\ -V_{i-m,\bullet}P_b &\leq (P_{\max})_{i-m}, & i = m + 1, \dots, 2m. \end{aligned} \quad (2.16)$$

Here $V_{i,\bullet}$ denotes the i^{th} row of V . In Chapter 4 we will not treat the power injections as given values, but instead as Gaussian random variables, reflecting the uncertainty and fluctuations in consumer demand and power generation. These restrictions will then relate to the problem of determining the probability of a multivariate Gaussian sample lying in the exterior of some polyhedron $\{\mathbf{x} \in \mathbb{R}^{n-1} : \mathbf{x}^\top \mathbf{a}_i \leq b_i, i = 1, \dots, 2m\}$, where

$$\begin{aligned} \mathbf{a}_i &= V_{i,\bullet}, & b_i &= (P_{\max})_i, & i &= 1, \dots, m \\ \mathbf{a}_i &= -V_{i-m,\bullet}, & b_i &= (P_{\max})_{i-m}, & i &= m + 1, \dots, 2m. \end{aligned} \quad (2.17)$$

It is of course also possible to put direct limits on the maximum power consumption or generation at any of the buses. This would lead to additional linear constraints. By the assumption in the DC-approximation that all voltage magnitudes are equal, it is impossible to apply any restrictions based on the voltage heights. Similarly, restrictions based on frequency are impossible to apply, since direct current does not have a frequency.

Restrictions in alternating current networks will not translate into as nice linear restrictions on the power injections. In Chapter 5 we will mainly focus on restrictions based on the voltage magnitude. Restrictions based on the frequency are again impossible. Those can only be applied to the synchronous machine model.

Chapter 3

Monte Carlo Simulation and Importance Sampling

For most models, it is impossible to analytically evaluate the exact probability of specific events. Therefore, we are directed towards simulation methods. In this chapter, we will discuss the theory for the most widespread simulation method: *Monte Carlo simulation*. Furthermore, we will describe an adaptation of Monte Carlo integration called *importance sampling*. Importance sampling is specifically well suited to the simulation of rare events, whereas traditional Monte Carlo integration typically falls short. Application of the theory will be left to Chapter 4.

3.1 Monte Carlo simulation

Monte Carlo simulation [13] is a widely used tool to approximate the expectation of a function with respect to a probability distribution. In this section we will sketch the outlines of this method. Let $\mathbf{p} = \mathbb{E}_p[f(X)]$ be the expectation of the function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with respect to a probability distribution with density function $p : \mathbb{R}^d \rightarrow \mathbb{R}$. If both f and p are Riemann-integrable then,

$$\mathbf{p} = \int_{\mathbb{R}^d} f(\mathbf{x})p(\mathbf{x})d\mathbf{x}.$$

For more complicated functions f or probability densities p , this integral can sometimes not be analytically evaluated. In that case, we try to approximate it via numerical methods. An estimator $\hat{\mathbf{p}}$ for \mathbf{p} can be found by averaging over a large number of samples X_i , $i = 1, \dots, N$ independently drawn from the distribution with density function p ,

$$\hat{\mathbf{p}} = \frac{1}{N} \sum_{i=1}^N f(X_i), \quad X_i \sim p.$$

By a slight abuse of conventional notation, we write $X_i \sim p$ to indicate that X_i is a random variable drawn from a probability distribution with density function p .

This estimator $\hat{\mathbf{p}}$ is an unbiased estimator, since

$$\begin{aligned}
\mathbb{E}_p[\hat{\mathbf{p}}] &= \mathbb{E}_p \left[\frac{1}{N} \sum_{i=1}^N f(X_i) \right] \\
&= \frac{1}{N} \sum_{i=1}^N \mathbb{E}_p[f(X_i)] \\
&= \frac{1}{N} \sum_{i=1}^N \mathbf{p} \\
&= \mathbf{p}.
\end{aligned} \tag{3.1}$$

We would also like to know the variance of the estimator, since that would provide us with knowledge on the accuracy of the estimator. A lower variance corresponds to a more accurate estimator. The variance of the estimator $\hat{\mathbf{p}}$ is given by,

$$\begin{aligned}
\text{Var}_p[\hat{\mathbf{p}}] &= \text{Var}_p \left[\frac{1}{N} \sum_{i=1}^N f(X_i) \right] \\
&= \frac{1}{N^2} \sum_{i=1}^N \text{Var}_p[f(X_i)] \\
&= \frac{1}{N} \text{Var}_p[f(X_1)] \\
&= \frac{1}{N} \left(\int_{\mathbb{R}^d} f(\mathbf{x})^2 p(\mathbf{x}) d\mathbf{x} - \mathbf{p}^2 \right).
\end{aligned} \tag{3.2}$$

When we have trouble evaluating $\mathbf{p} = \mathbb{E}_p[f(X)]$, it seems highly unlikely that we will be able to properly evaluate $\mathbb{E}_p[f(X)^2]$. Therefore, we again look for an estimator, this time for the *estimator variance* $\sigma^2 := \text{Var}_p[f(X_1)]$. The sample variance,

$$S_N^2 = \frac{1}{N-1} \sum_{i=1}^N (f(X_i) - \hat{\mathbf{p}})^2, \quad X_i \sim p,$$

is a natural estimator for the estimator variance σ^2 , since it is unbiased, as we derive next,

$$\begin{aligned}
\mathbb{E}_p(S_N^2) &= \frac{1}{N-1} \mathbb{E}_p \left[\sum_{i=1}^N (f(X_i) - \hat{\mathbf{p}})^2 \right] \\
&= \frac{1}{N-1} \mathbb{E}_p \left[\sum_{i=1}^N (f(X_i)^2 - 2\hat{\mathbf{p}}f(X_i) + \hat{\mathbf{p}}^2) \right] \\
&= \frac{1}{N-1} \sum_{i=1}^N (\mathbb{E}_p[f(X_i)^2] - \mathbb{E}_p[\hat{\mathbf{p}}^2]) \\
&= \frac{N}{N-1} \left(\text{Var}_p[f(X_1)] + \mathbf{p}^2 - \left(\frac{1}{N} \text{Var}_p[f(X_1)] + \mathbf{p}^2 \right) \right) \\
&= \text{Var}_p[f(X_1)].
\end{aligned}$$

At this point, we wish to concretise the semantics of the sentence: ‘A lower variance corresponds to a more accurate estimator.’ We introduce some new notation in order to

use the previous results for constructing an asymptotic confidence interval for \mathbf{p} . Let ξ_α denote the upper α -quantile of the standard normal distribution $\mathcal{N}(0, 1)$ i.e.

$$\int_{\xi_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = 1 - \alpha.$$

Theorem 3.1.1. *An asymptotic confidence interval of significance level α for \mathbf{p} is given by,*

$$\left[\hat{\mathbf{p}} - \frac{S_N}{\sqrt{N}} \xi_{\alpha/2}, \hat{\mathbf{p}} + \frac{S_N}{\sqrt{N}} \xi_{\alpha/2} \right]. \quad (3.3)$$

Proof. This proof is taken from [14].

By the weak law of large numbers

$$S_N^2 \xrightarrow{\text{Prob.}} \text{Var}_p[f(X)], \quad \text{as } N \rightarrow \infty,$$

where $\xrightarrow{\text{Prob.}}$ denotes convergence in probability. Consequently, we find by the continuous mapping theorem that,

$$S_N \xrightarrow{\text{Prob.}} \sqrt{\text{Var}_p[f(X)]}, \quad \text{as } N \rightarrow \infty,$$

where $\xrightarrow{\text{Dist.}}$ denotes convergence in distribution. The central limit theorem states that,

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow{\text{Dist.}} \mathcal{N}(0, \text{Var}_p[f(X)]), \quad \text{as } N \rightarrow \infty.$$

By Slutsky's lemma we now find

$$\sqrt{N} \frac{\hat{\mathbf{p}} - \mathbf{p}}{S_N} \xrightarrow{\text{Dist.}} \mathcal{N}(0, 1), \quad \text{as } N \rightarrow \infty.$$

Thus $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})/S_N$ is asymptotically standard normally distributed and we find that,

$$\mathbb{P} \left[-\xi_{\alpha/2} \leq \sqrt{N} \frac{\hat{\mathbf{p}} - \mathbf{p}}{S_N} \leq \xi_{\alpha/2} \right] \rightarrow 1 - \alpha, \quad \text{as } N \rightarrow \infty.$$

□

From this theorem we can finally see what we meant by ‘A lower variance corresponds to a more accurate estimator.’ Namely, as the sample variance decreases, so does the confidence interval of the estimator become narrower. This will give us a good grip on how useful (or useless) Monte Carlo integration will be in different situations. In the next section, we will see how, even when Monte Carlo simulation seemingly falls short, it can still be adapted to be useful.

3.2 Importance sampling

In our study of the power grid, we are essentially performing a form of risk analysis in which we are interested in a specific type of function f . Namely, an indicator function that has the value 1 when something dramatic happens, e.g. a power line failure, and 0

when there are no problems at all. For some Borel-measurable set $A \subseteq \mathbb{R}^d$, the indicator function $\mathbb{1}_A$ is,

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

For this function f , the expectation of f with respect to a probability density p essentially translates into the probability of an event A occurring, $\mathbb{E}_p[\mathbb{1}_A(X)] = \mathbb{P}_p(X \in A)$.

When studying rare events, a good criterion for the accuracy of the estimator $\hat{\mathbf{p}}$ found by Monte Carlo integration is to consider the relative standard deviation (RSD) rather than the variance,

$$\text{RSD}_p[\mathbf{p}] = \frac{\sqrt{\text{Var}_p[f(X)]}}{\mathbf{p}}.$$

Note that by twice applying the weak law of large numbers and by thereafter using the continuous mapping theorem, we find that $S_N/\hat{\mathbf{p}}$ is an asymptotically consistent estimator for $\text{RSD}_p[\mathbf{p}]$, i.e. $S_N/\hat{\mathbf{p}} \xrightarrow{\text{Prob.}} \text{RSD}_p[\mathbf{p}]$.

We consider the relative standard deviation, because the asymptotic confidence interval given in Theorem 3.1.1 may have a nonsensical lower bound, $\hat{\mathbf{p}} - \xi_{\alpha/2}S_N/\sqrt{N} < 0$. Since \mathbf{p} is a probability, it has to be positive. Requiring this bound to be positive, roughly corresponds to requiring that $\text{RSD}_p[\mathbf{p}] \leq \sqrt{N}/\xi_{\alpha/2}$. Furthermore, we would like for the orders of magnitude between which \mathbf{p} could vary, to be as small as possible, or equivalently for the relative standard deviation to be as small as possible.

The estimator variance $\sigma^2 = \text{Var}_p[f(\mathbf{X})]$ reduces in the case $f = \mathbb{1}_A$ to

$$\begin{aligned} \text{Var}_p[f(\mathbf{X})] &= \text{Var}_p[\mathbb{1}_A(\mathbf{X})] \\ &= \mathbb{E}_p[\mathbb{1}_A(\mathbf{X})^2] - \mathbb{E}_p[\mathbb{1}_A(\mathbf{X})]^2 \\ &= \mathbb{E}_p[\mathbb{1}_A(\mathbf{X})] - \mathbb{E}_p[\mathbb{1}_A(\mathbf{X})]^2 \\ &= \mathbf{p} - \mathbf{p}^2. \end{aligned}$$

If we would use this form of Monte Carlo Integration to approximate the probability of a transmission line failure due to fluctuations in local supply and demand of power, then we would need a huge amount of computational resources. The probability of a transmission line failure in a certain time interval is very small $\mathbf{p} \ll 1$ and it is classified as a *rare event*. To find a single combination of power injections at the nodes we on average have to draw \mathbf{p}^{-1} samples. Specifically, the requirement on the lower bound forces us to take an enormous amount of samples, namely

$$\sqrt{N} \geq \xi_{\alpha/2}\text{RSD}_p[\mathbf{p}] = \xi_{\alpha/2}\sqrt{1 - 1/\mathbf{p}} \approx \xi_{\alpha/2}\mathbf{p}^{-1/2}.$$

Therefore, the amount of samples needed for a reliable result is approximately linearly related to the inverse of the probability of a failure, $N \propto \mathbf{p}^{-1}$. To draw this many samples for rare events and solve the power-flow equations for each one of them is simply impracticable without the resource of an astonishingly large amount of processing power.

To tackle this hindrance we use *Importance Sampling*, a form of *Variance Reduction* [15]. Henceforth, we will refer to Monte Carlo simulation without the use of additional techniques by calling it *crude* Monte Carlo simulation. Importance sampling is typically applied in situations where $f(\mathbf{x})$ is approximately zero for \mathbf{x} in the exterior of some *important region* $A \subset \mathbb{R}^d$ for which $\mathbb{P}(X \in A : X \sim p)$ is very small. This corresponds well to our model of the power grid, where $f(\mathbf{x}) = 0$ when the power injection configuration \mathbf{x} does not lead to any problems in the power grid, and $f(\mathbf{x}) = 1$ when it does.

In Importance Sampling, instead of drawing samples X_i according to the probability density function p , we draw them according to some auxiliary probability density function q , the *importance sampling distribution*. It is required that $q(\mathbf{x}) > 0$, whenever $f(\mathbf{x})p(\mathbf{x}) \neq 0$. The importance sampling distribution is chosen such, that many samples are found in the important region, for which $f(\mathbf{x}) \not\approx 0$. We have to adjust the overemphasis of q on the important region by adding weights $w(\mathbf{x}) = p(\mathbf{x})/q(\mathbf{x})$ to the *importance sampling estimator*,

$$\hat{\mathbf{p}}_q = \frac{1}{N} \sum_{i=1}^N w(X_i) f(X_i), \quad X_i \sim q.$$

This estimator, $\hat{\mathbf{p}}_q$, is also an unbiased estimator for \mathbf{p} , since

$$\begin{aligned} \mathbb{E}_q[\hat{\mathbf{p}}_q] &= \mathbb{E}_q \left[\frac{1}{N} \sum_{i=1}^N w(X_i) f(X_i) \right] \\ &= \frac{1}{N} \sum_{i=1}^N E_q [w(X) f(X)] \\ &= \int_{\mathbb{R}^d \setminus \{\mathbf{x} \in \mathbb{R}^d : q(\mathbf{x})=0\}} f(\mathbf{x}) \cdot \frac{p(\mathbf{x})}{q(\mathbf{x})} \cdot q(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d \setminus \{\mathbf{x} \in \mathbb{R}^d : q(\mathbf{x})=0\}} f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \\ &= \mathbb{E}_p[f(X)] \\ &= \mathbf{p}. \end{aligned}$$

The estimator variance $\sigma_q^2 := \text{Var}_q [w(\mathbf{X})f(\mathbf{X})]$ of $\hat{\mathbf{p}}_q$ does differ from that of $\hat{\mathbf{p}}$ and is given by,

$$\begin{aligned} \sigma_q^2 &= N \text{Var}_q [\hat{\mathbf{p}}_q] \\ &= N \text{Var}_q \left[\frac{1}{N} \sum_{i=1}^N w(X_i) f(X_i) \right] \\ &= \frac{1}{N} \sum_{i=1}^N \text{Var}_q [w(X) f(X)] \tag{3.4} \\ &= \int_{\mathbb{R}^d \setminus \{q=0\}} (w(\mathbf{x}) f(\mathbf{x}))^2 \cdot q(\mathbf{x}) d\mathbf{x} - E_q [w(X) f(X)]^2 \\ &= \int_{\mathbb{R}^d \setminus \{q=0\}} \frac{(f(\mathbf{x}) p(\mathbf{x}))^2}{q(\mathbf{x})} d\mathbf{x} - \mathbf{p}^2. \end{aligned}$$

For an appropriate choice of importance sampling distribution, this becomes significantly lower than crude Monte Carlo integration. For a positive function f , the optimal choice of q is,

$$q(\mathbf{x}) = \frac{f(\mathbf{x}) p(\mathbf{x})}{\int_{\mathbb{R}^d} f(\mathbf{y}) p(\mathbf{y}) d\mathbf{y}} = \mathbf{p}^{-1} f(\mathbf{x}) p(\mathbf{x}).$$

This yields zero estimator variance,

$$\sigma_q^2 = \mathbf{p} \int_{\mathbb{R}^d \setminus \{q=0\}} \frac{(f(\mathbf{x})p(\mathbf{x}))^2}{f(\mathbf{x})p(\mathbf{x})} d\mathbf{x} - \mathbf{p}^2 = 0.$$

If we would be able to sample from this particular distribution, then we would already know the value of \mathbf{p} , namely $\mathbf{p} = \int f(\mathbf{x})p(\mathbf{x})/q(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d \setminus \{p=0\}$. This would render importance sampling redundant. So in every situation, where we wish to apply importance sampling, this optimal distribution will, unfortunately, be out of reach. However, it does still provide us with a goal to strive for. If we are able to find a q for a positive function f such that $q(\mathbf{x}) \approx f(\mathbf{x})p(\mathbf{x})/\mathbf{p}$, then we expect a low sample variance. For instance, suppose that $c \cdot q(\mathbf{x}) \geq f(\mathbf{x})p(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$ and some $c \geq \mathbf{p}$, then the estimator variance is upper bounded by $\sigma_q^2 \leq c^2 - \mathbf{p}^2$. In the next section we will discuss techniques for finding useful importance sampling distributions.

3.3 Exponential tilting

It is sometimes difficult to think of what distribution to use for importance sampling. An often applied choice is exponential tilting. In exponential tilting, we restrict ourselves to a certain family of possible importance sampling distributions and try to minimise the variance over this family.

Definition 3.3.1. *Given a random variable X in a sample space χ with probability measure \mathbb{P} and moment generating function $M_X(\theta) = \mathbb{E}[e^{\theta^\top X}] < \infty$, the exponentially tilted measure \mathbb{P}_θ is,*

$$\mathbb{P}_\theta(X \in d\mathbf{x}) = \frac{\mathbb{E} \left[e^{\theta^\top X} \mathbb{1}[X \in d\mathbf{x}] \right]}{M_X(\theta)} = \exp(\theta^\top \mathbf{x} - \Lambda(\theta)) \mathbb{P}(X \in d\mathbf{x}).$$

Here Λ is the cumulant generating function $\Lambda(\theta) = \log \mathbb{E}[e^{\theta^\top X}]$.

As a consequence of this method, the density function of the exponentially tilted distribution f_θ closely resembles the density function of the original distribution f , $f_\theta(\mathbf{x}) = \exp(\theta^\top \mathbf{x} - \Lambda(\theta))f(\mathbf{x})$. This can be utilised in certain circumstances to allow for well manageable importance sampling distributions.

One of the simplest examples to illustrate exponential tilting is by using the exponential distribution. Consider an exponential random variable X with rate parameter $\lambda > 0$. The moment generating function is,

$$\mathbb{E} [e^{\theta X}] = \lambda \int_0^\infty e^{\theta x} e^{-\lambda x} dx = \left[\frac{\lambda}{\theta - \lambda} e^{(\theta - \lambda)x} \right]_0^\infty = \frac{\lambda}{\lambda - \theta}, \quad \text{for } -\infty < \theta < \lambda.$$

The exponentially tilted measure is,

$$\mathbb{P}_\theta(X \in dx) = \begin{cases} e^{\theta x} \frac{\lambda - \theta}{\lambda} \cdot \lambda e^{-\lambda x} = (\lambda - \theta) e^{-(\lambda - \theta)x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

We recognise this as simply being again an exponential distribution, but this time with rate parameter $\lambda - \theta$.

This nice relation between the original distribution and the exponentially tilted measure is something that is typical to distributions that belong to the exponential family. Exponential tilting is also possible for any other distribution family with a finite moment generating function, but it does become more complicated.

Definition 3.3.2. Let $\Theta \subset \mathbb{R}^k$ be a parameter space and let $\mathcal{M} = \{P_\theta : \theta \in \Theta\}$ be a family of probability distributions on a sample space χ . \mathcal{M} is an exponential family if there exist some functions $\eta : \Theta \rightarrow \mathbb{R}^k$, $T : \chi \rightarrow \mathbb{R}^k$, $A : \chi \rightarrow \mathbb{R}$ and $B : \Theta \rightarrow \mathbb{R}$ such that the probability density function can be expressed as

$$p(\mathbf{x}, \theta) = \exp(\eta(\theta)^\top T(\mathbf{x}) - A(\mathbf{x}) - B(\theta)), \quad \theta \in \Theta.$$

Arguably, the most famous example of this family is the Gaussian distribution. In this case $\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$ and there are functions η, T, A and B ,

$$\begin{aligned} \eta : \Theta &\rightarrow \mathbb{R}^2 : (\mu, \sigma) \rightarrow \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^\top, \\ T : \mathbb{R} &\rightarrow \mathbb{R}^2 : x \rightarrow (x, x^2)^\top, \\ A : \mathbb{R} &\rightarrow \mathbb{R} : x \rightarrow \frac{1}{2} \ln(2\pi), \\ B : \Theta &\rightarrow \mathbb{R} : (\mu, \sigma) \rightarrow \ln(\sigma) + \frac{\mu^2}{2\sigma^2}, \end{aligned}$$

that yield $p(x, (\mu, \sigma)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$.

In Section 4.2 we will apply exponential tilting to the multivariate Gaussian distribution to find a useful importance sampling distribution.

3.4 Estimator efficiency and large deviations theory

An unfortunate side-effect that may occur in importance sampling, when the sampling distribution is chosen without due deliberation, is that the variance of the estimator with respect to the importance sampling distribution can become very large, or even infinite. On the other hand, a desirable quality in choosing the importance sampling distribution is that the relative standard deviation of the importance sampling estimator is small. This is more difficult for rare events. For example, as we saw in Section 3.2, when we use crude Monte Carlo simulation for determining the probability \mathbf{p} of an event occurring, then the estimator variance will be $\sigma^2 = \mathbf{p} - \mathbf{p}^2$ per sample. This will converge to zero as the probability \mathbf{p} tends to zero. However, the relative error will tend towards infinity as the probability tends to zero,

$$\frac{\sigma}{\mathbf{p}} = \sqrt{\frac{1}{\mathbf{p}} - 1} \sim \frac{1}{\sqrt{\mathbf{p}}} \rightarrow \infty, \quad \text{as } \mathbf{p} \downarrow 0.$$

Thus, as discussed earlier, we would need increasingly many samples for smaller probabilities to achieve the same relative error of the estimator. However, for the right choice of importance sampling estimator, this increase in the need for many samples as the probability decreases, may be very slow. In an ideal case, it can even be non-existent. We can speak of the increase of relative error for a series of estimators for different probabilities

to have *bounded relative error* or to be *logarithmic efficient* [16].

Let $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ be a family of probability measures on a single measurable space (Ω, \mathcal{A}) and let $\{A_n\}_{n \in \mathbb{N}}$ be a family of events on this space, where $n \in \mathbb{N}$ is the rarity parameter such that $\mathbf{p}_n := \mathbb{P}_n(A_n) \rightarrow 0$ as $n \rightarrow \infty$. For each $n \in \mathbb{N}$ let $\hat{\mathbf{p}}_n$ be an unbiased estimator of \mathbf{p}_n with estimator variance σ_n^2 , that tends to zero as n tends to infinity. The family of estimators $\{\hat{\mathbf{p}}_n\}_{n \in \mathbb{N}}$ is said to possess bounded relative error if,

$$\limsup_{n \rightarrow \infty} \frac{\sigma_n^2}{\mathbf{p}_n^2} < \infty.$$

The family of estimators $\{\hat{\mathbf{p}}_n\}_{n \in \mathbb{N}}$ is called logarithmically efficient if,

$$\forall \epsilon > 0 \quad \limsup_{n \rightarrow \infty} \frac{\sigma_n^2}{\mathbf{p}_n^{2-\epsilon}} = 0,$$

or equivalently,

$$\liminf_{n \rightarrow \infty} \frac{\log \sigma_n^2}{\log \mathbf{p}_n^2} \geq 1. \quad (3.5)$$

While bounded relative error is a slightly stronger property than logarithmic efficiency, it is more common to work with logarithmically efficient estimators.

Practically, the increase in number of samples needed for smaller probabilities with logarithmically efficient estimators grows slowly enough to not result in an insurmountable amount of simulation effort. We do not have to take increasingly many samples for decreasing variations in the power injections to reach a similar relative level of confidence in the accuracy of the estimator. Furthermore, in some situations, estimators with bounded relative error do not exist [17] or have not been discovered. In Chapter 4 we will construct several logarithmically efficient estimators. Other examples can be found in [16].

For now, let us consider the example of determining the probabilities $\mathbf{p}_n = \mathbb{P}(X_n > 1)$, where X_n is an exponential random variable with rate n , using the exponential distribution with rate 1 as the importance sampling distribution. Clearly, we have $\mathbf{p}_n = e^{-n}$. Using Equation 3.4, we find for all $n \in \mathbb{N}$ the estimator variance,

$$\sigma_n^2 = \int_1^\infty \frac{n^2 e^{-2nx}}{e^{-x}} dx - \mathbf{p}_n^2 = \frac{n^2}{2n-1} e^{-2n+1} - e^{-2n}.$$

Thus, the set of estimators $\hat{\mathbf{p}}_n$ does not have bounded relative error,

$$\limsup_{n \rightarrow \infty} \frac{\sigma_n^2}{\mathbf{p}_n^2} = \limsup_{n \rightarrow \infty} \frac{n^2}{2n-1} e - 1 = \infty,$$

but it is logarithmically efficient,

$$\forall \epsilon > 0, \quad \limsup_{n \rightarrow \infty} \frac{\sigma_n^2}{\mathbf{p}_n^{2-\epsilon}} = \limsup_{n \rightarrow \infty} e^{-n\epsilon} \left(\frac{n^2}{2n-1} e - 1 \right) = 0.$$

The probabilities \mathbf{p}_n are needed to determine whether a family of estimators has bounded relative error or is logarithmically efficient. These are unknown in the example, otherwise sampling would be unnecessary. It is, however, possible to know the behaviour of \mathbf{p}_n as $n \rightarrow \infty$, i.e. the rate at which \mathbf{p}_n converges to 0. In *Large Deviations Theory* this is studied by using the *rate function*. We use the definition of a rate function as given in [18].

Definition 3.4.1. A rate function I is a lower semi-continuous mapping $I : \Omega \rightarrow [0, \infty]$ such that for all $\alpha \in [0, \infty)$, the level set $\Psi_I(\alpha) = \{x \in \Omega : I(x) \leq \alpha\}$ is a closed subset of Ω . A good rate function is a rate function for which all the level sets $\Psi_I(\alpha)$ are compact subsets of Ω .

Definition 3.4.2. A family of probability measures $\{\mathbb{P}_\alpha\}_{\alpha>0}$ on $(\Omega, \mathcal{B}(\Omega))$ satisfies the large deviations principle with rate function $I : \Omega \rightarrow [0, \infty]$ if, for all $A \in \mathcal{B}(\Omega)$,

$$-\inf_{x \in A^\circ} I(x) \leq \liminf_{\alpha \rightarrow \infty} \frac{1}{\alpha} \log \mathbb{P}_\alpha(A) \leq \limsup_{\alpha \rightarrow \infty} \frac{1}{\alpha} \log \mathbb{P}_\alpha(A) \leq -\inf_{x \in A} I(x).$$

The notation $\mathcal{B}(\Omega)$ is used to represent the Borel sigma-algebra on the set Ω .

Thanks to the lower semi-continuity of the rate function, these inequalities can be turned into equalities in some cases. If a family of probability measures $\{\mathbb{P}_\alpha\}_{\alpha>0}$ on $(\Omega, \mathcal{B}(\Omega))$ satisfies the large deviations principle with a rate function $I : \Omega \rightarrow [0, \infty]$, then for all regular open or regular closed sets $A \in \mathcal{B}(\Omega)$ we find $\inf_{x \in A^\circ} I(x) = \inf_{x \in \bar{A}} I(x)$, and

$$-\inf_{x \in A} I(x) = \lim_{\alpha \rightarrow \infty} \frac{1}{\alpha} \log \mathbb{P}_\alpha(A).$$

Definition 3.4.3. A set $A \in \mathcal{B}(\Omega)$ is called regular open if it is the interior of its own closure,

$$A = \text{int}(cl(A)).$$

A set $A \in \mathcal{B}(\Omega)$ is called regular closed if its the closure of its own interior,

$$A = cl(\text{int}(A)).$$

There is a famous theorem that stood at the inception of large deviations theory known as Cramér's theorem. This theorem yields the rate function for a family of probability measures that satisfies the large deviations principle.

Theorem 3.4.1 (Cramér's theorem). Let $(\mathbf{X}_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with bounded cumulant generating function

$$\Lambda(\lambda) = \log \mathbb{E} [\exp(\lambda^\top \mathbf{X})], \quad \text{for all } \lambda \in \mathbb{R}^d.$$

Let Λ^* be the Legendre-transform of Λ ,

$$\Lambda^*(\mathbf{x}) = \sup_{\lambda \in \mathbb{R}^d} \{\lambda^\top \mathbf{x} - \Lambda(\lambda)\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Then the family of probability measures $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$, where \mathbb{P}_n is the probability measure of the random variable $\frac{1}{n} \sum_{i=1}^n X_i$ for all $n \in \mathbb{N}$, satisfies the large deviations principle with good rate function

$$I(\mathbf{x}) = \Lambda^*(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d.$$

Proof. A proof for probability measures on \mathbb{R} can be found in any book on large deviation theory. A proof specifically for probability measures on the multidimensional \mathbb{R}^d can be found in [18]. \square

We can utilise Cramér's theorem when the studied family of probability measures $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ can somehow be related to the averaged sum of i.i.d. variables. In Section 4.2 we will do this for probability measures of the Gaussian distribution.

Chapter 4

The DC-Approximation with Stochastic Power Injections

In Chapter 2, we introduced the DC-approximation as an approximate model for the power flow in the power grid given the power injections. However, while power demand and generation are to a certain degree predictable, they are by no means deterministic. In general, power generation and demand are higher in the daytime and lower at night. Deviations in power generation arise due to the influence of atmospheric conditions on renewable energy sources. Similarly, the demand for power fluctuates as a consequence of the weather and simply due to fickleness in human behaviour. In Figure 4.1 the predicted and the realised production and load in the Dutch power grid over a period of two weeks are shown.

The actual distribution of the power generation and demand is very difficult to state, especially in general terms. In fact, it appears that there is no standard probabilistic distribution that fully describes the power injections. It is common to use a Gaussian distribution, e.g. [19], [20], or [21]. Several analyses have been performed in specific grids. For example, in [22] the distribution of solar and wind power was studied for short time scales, seconds up to minutes. In [23] a similar analysis was performed on a longer timescale, hours up to a day. Both found that the distributions of forecast errors have kurtosis and skewness that slightly differ from the Gaussian distribution, but they were unimodal.

Despite these differences, we do choose to employ the Gaussian distribution to model the power injections as others did before us. An added benefit of the Gaussian distribution is that it is very easy to include correlation between the power injections at different buses.

For the Gaussian distribution of power injections, we will subsequently determine the probability of violating the power flow constraints as introduced in Section 2.3 using the theory of importance sampling introduced in Chapter 3.

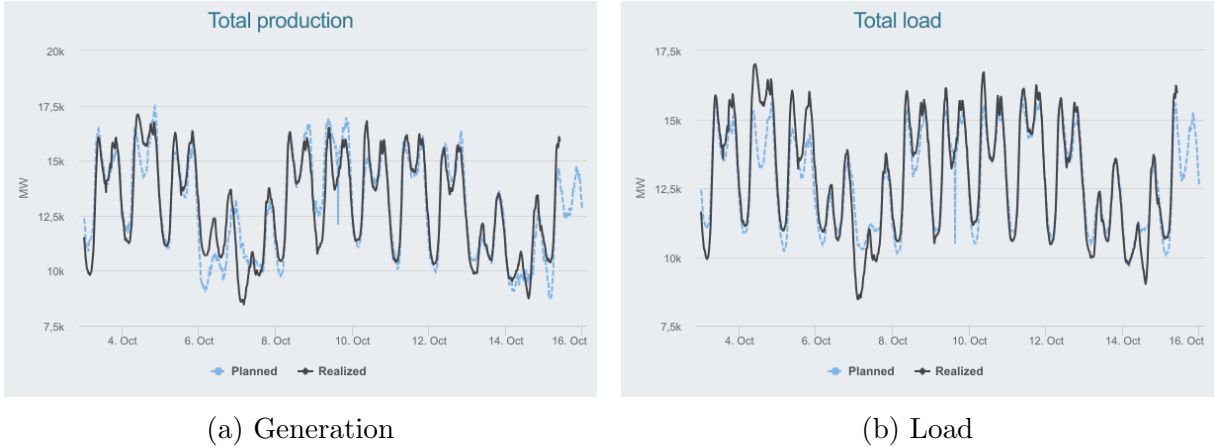


Figure 4.1: The planned and the realised production and load in the Dutch power grid over a two week period in 2018. This data is imported from the publicly accessible website of the Dutch grid manager Tennet, <https://www.tennet.eu/electricity-market/data-dashboard/>. Differences between generation and demand are dealt with in import or export of power with Germany and Belgium via land, and with Great Britain and Norway via undersea high voltage direct current cables [24].

4.1 Preliminaries

In mathematical terms, our goal is to use simulation to evaluate the probability that a random variable \mathbf{X} drawn from a d -dimensional multivariate Gaussian distribution with mean μ and covariance matrix Σ lies in the exterior of some polyhedron

$$\{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \geq \mathbf{b}\}, \mathbf{A} \in \mathbb{R}^{m \times d}, \mathbf{b} \in \mathbb{R}^m.$$

This is directly related to the restrictions for feasible power injections in the DC-approximation discussed in Section 2.3. A polyhedron is defined as the intersection of multiple half-spaces. Therefore, by taking the complement we find that the exterior of a polyhedron is the union of half-spaces. A logical first step towards evaluating this problem is to evaluate the probability that a randomly drawn sample lies in a single half-space.

This can be done both by simulation and by analytic means. In this section, we will focus on the analytic approach. We will also relate the multivariate Gaussian distribution to the chi-squared distribution. This will be useful in studying the probability of a Gaussian random variable lying in some more general set than the exterior of a polyhedron. First, we need to define the Gaussian distribution. We give the probability density function and the cumulative distribution.

Definition 4.1.1. *The probability density function p of the d -dimensional multivariate Gaussian distribution $\mathcal{N}_d(\mu, \Sigma)$, with $\mu \in \mathbb{R}^d$ and Σ a $d \times d$ positive definite matrix, is given by,*

$$p(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right), \quad \mathbf{x} \in \mu + \text{span}(\Sigma).$$

We will continue the use of notation in this definition. When we say that a random variable \mathbf{X} is $\mathcal{N}_d(\mu, \Sigma)$ distributed, it is to be read as a random variable \mathbf{X} that is distributed as a d -dimensional multivariate Gaussian distribution with mean μ and covariance matrix Σ . In formulae, we will write even more concisely, $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$.

Definition 4.1.2. The cumulative distribution function Φ of the 1-dimensional standard normal distribution $\mathcal{N}_1(0, 1)$ is,

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt, \quad x \in \mathbb{R}.$$

Instead of $\Phi(x)$ we will usually prefer to consider its complement $\Phi^C(x) = 1 - \Phi(x)$ instead.

A particularly nice property of the Gaussian distribution is that its conditional distribution is again a Gaussian distribution.

Theorem 4.1.1. Let $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$ be partitioned as,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

where \mathbf{X}_1 consist of the first $k < d$ elements of \mathbf{X} . Let the mean and covariance matrix be correspondingly partitioned,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_{22} \end{pmatrix}.$$

The conditional distribution $\mathbf{X}_1 | \mathbf{X}_2$ is a Gaussian distribution with,

$$\begin{aligned} \mathbb{E}(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) &= \mu_1 + \Sigma_{12} \Sigma_{22}^\dagger (\mathbf{x}_2 - \mu_2), \quad \text{and,} \\ \text{Var}(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) &= \Sigma_{11} - \Sigma_{12} \Sigma_{22}^\dagger \Sigma_{12}^\top, \end{aligned}$$

where Σ_{22}^\dagger is the generalized inverse of Σ_{22} .

Proof. A proof can be found in [25]. □

The covariance matrix Σ of a non-degenerate distribution is positive definite. This is of importance as it both ensures the existence of the inverse as well as that of a unique positive-definite square root.

Definition 4.1.3. Let A be a square matrix. If there exists a matrix B such that $B^2 = A$, then B is called a square root of A .

Lemma 4.1.1. The **principal (square) root** of a positive definite matrix $\Sigma = U^\top D U$, with U an orthogonal and D a diagonal matrix, is uniquely defined as the positive definite matrix $\sqrt{\Sigma} := U^\top \sqrt{D} U$, where \sqrt{D} is the matrix with the elementwise positive square roots of D ,

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}, \quad \sqrt{D} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\lambda_n} \end{pmatrix}.$$

Proof. The proof of this lemma is based on [26]. Clearly, $\sqrt{\Sigma}$ is a square root of Σ , as

$$\sqrt{\Sigma}^2 = U^\top \sqrt{D} U U^\top \sqrt{D} U = U^\top \sqrt{D} \sqrt{D} U = U^\top D U = \Sigma.$$

Since Σ is positive definite, $\lambda_i, i = 1, \dots, n$ are all strictly positive. Thus $\sqrt{\lambda_i}, i = 1, \dots, n$ are all strictly positive too, and hence $\sqrt{\Sigma} = U^\top \sqrt[3]{D} U$ is positive definite. This concludes the proof of the existence of the principal root. What remains, is to prove the uniqueness.

Let A be a positive definite matrix such that $A^2 = \Sigma$. Let P be the interpolating polynomial such that $P(\lambda_i) = \sqrt{\lambda_i}$ for all $i = 1, \dots, n$. For this polynomial it holds that

$$P(\Sigma) = P(U^\top D U) = U^\top P(D) U = U^\top \sqrt[3]{D} U = \sqrt{\Sigma}.$$

Now we use this to find that A and $\sqrt{\Sigma}$ commute,

$$A\sqrt{\Sigma} = AP(\Sigma) = AP(A^2) = P(A^2)A = P(\Sigma)A = \sqrt{\Sigma}A.$$

Thus A and $\sqrt{\Sigma}$ are simultaneously diagonalizable, i.e. there exists a basis that is an eigenbasis for both A and $\sqrt{\Sigma}$. Thus, there is some diagonal matrix \tilde{D} such that $A = U^\top \tilde{D} U$. Since taking the element-wise positive square root of positive numbers is a unique operation and $D^2 = U\sqrt{\Sigma}^2 U^\top = UA^2 U^\top = \tilde{D}^2$, we find that $\tilde{D} = D$ and consequently $A = \sqrt{\Sigma}$. \square

Note that as a corollary of this lemma, we find that for a positive definite matrix $\Sigma = U^\top D U$ the (generalized) inverse of the principal root is equal to the principal root of the (generalized) inverse,

$$\sqrt{\Sigma}^{-1} = U^\top \sqrt[3]{D^{-1}} U = U^\top \sqrt[3]{D^{-1}} U = \sqrt{\Sigma^{-1}}.$$

With the above lemma, we now have all the tools we need to analytically determine the probability of sampling a random Gaussian variable in a half-space.

Theorem 4.1.2. *Let $\mathbf{X} = (x_1, x_2, \dots, x_d)$ be $\mathcal{N}_d(\mu, \Sigma)$ distributed and let $\mathbf{a} \in \mathbb{R}^d, b \in \mathbb{R}$, then*

$$\mathbb{P}(\mathbf{X}^\top \mathbf{a} \geq b) = \Phi^C \left(\frac{b - \mu^\top \mathbf{a}}{\|\sqrt{\Sigma} \mathbf{a}\|_2} \right).$$

Proof. We define the half-space $H := \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}^\top \mathbf{a} \geq b\}$ in order to be able to compactly write,

$$\mathbb{P}(\mathbf{X}^\top \mathbf{a} \geq b) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \int_H \exp \left(-\frac{1}{2} (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right) d\mathbf{x}.$$

We apply a change of variables,

$$\mathbf{x}' = \sqrt{\Sigma}^{-1} (\mathbf{x} - \mu), \quad \mathbf{a}' = \sqrt{\Sigma} \mathbf{a}, \quad H' = \{\mathbf{x}' \in \mathbb{R}^d : \mathbf{x}'^\top \mathbf{a}' + \mu^\top \mathbf{a} \geq b\}.$$

The Jacobian matrix of this transformation is $\sqrt{\Sigma}^{-1}$, so its determinant is $|\sqrt{\Sigma}^{-1}|$ and we find $d\mathbf{x}' = \frac{1}{\sqrt{\det \Sigma}} d\mathbf{x}$. Thus, we can write,

$$\mathbb{P}(\mathbf{X}^\top \mathbf{a} \geq b) = \frac{1}{(2\pi)^{d/2}} \int_{H'} \exp \left(-\frac{1}{2} \mathbf{x}'^\top \mathbf{x}' \right) d\mathbf{x}'.$$

Next we find an orthogonal matrix B such that $B\mathbf{a}' = \|\mathbf{a}'\|_2 \cdot e_1$, where e_1 denotes the first standard unit vector, and use it to apply a second change of variables,

$$\mathbf{x}^* = B\mathbf{x}', \quad \mathbf{a}^* = B\mathbf{a}', \quad H^* = \{\mathbf{x}^* : \mathbf{x}^{*\top} \mathbf{a}^* + \mu^\top \mathbf{a} \geq b\}.$$

Since the matrix B is orthogonal, its determinant is $|B| = 1$ and we find $d\mathbf{x}^* = d\mathbf{x}'$.

Now we finally arrive at the stated result of the theorem,

$$\begin{aligned}
\mathbb{P}(\mathbf{X}^\top \mathbf{a} \geq b) &= \frac{1}{(2\pi)^{d/2}} \int_{H^*} \exp\left(-\frac{1}{2} \mathbf{x}^{*\top} \mathbf{x}^*\right) d\mathbf{x}^* \\
&= \frac{1}{(2\pi)^{d/2}} \int_{\frac{b - \mu^\top \mathbf{a}}{\|\mathbf{a}^*\|_2}}^{\infty} \exp\left(-\frac{1}{2} x_1^{*2}\right) dx_1^* \prod_{i=2}^d \left(\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} x_i^{*2}\right) dx_i^* \right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{b - \mu^\top \mathbf{a}}{\|\mathbf{a}^*\|_2}}^{\infty} \exp\left(-\frac{1}{2} x_1^{*2}\right) dx_1^* \\
&= \Phi^C\left(\frac{b - \mu^\top \mathbf{a}}{\|\sqrt{\Sigma} \mathbf{a}\|_2}\right).
\end{aligned}$$

□

We should note that a nice property of the Gaussian distribution has slipped into the proof of the above theorem. If \mathbf{X} is $\mathcal{N}_d(\mu, \Sigma)$ distributed, then $\sqrt{\Sigma}^{-1}(\mathbf{X} - \mu)$ is $\mathcal{N}_d(0, I_n)$ distributed. The converse also holds. If \mathbf{X} is $\mathcal{N}_d(0, I_n)$ distributed, then $\sqrt{\Sigma} \mathbf{X} + \mu$ is $\mathcal{N}_d(\mu, \Sigma)$ distributed.

Later on in this chapter, we will encounter expressions containing products and fractions of the cumulative uni-variate Gaussian distribution. In order to arrive at a better intuitive grasp of how these expressions behave, we provide bounds.

Lemma 4.1.2. For $x \in (0, \infty)$ upper and lower bounds for $\Phi^C(x)$ are given by

$$\begin{aligned}
\Phi^C(x) &\leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{1}{2} x^2\right), \\
\Phi^C(x) &\geq \frac{1}{\sqrt{2\pi}} \frac{x}{1+x^2} \exp\left(-\frac{1}{2} x^2\right).
\end{aligned}$$

Proof. First, we prove the upper bound. Note that for all $t \in [x, \infty)$ we have $\frac{t}{x} \geq 1$. Thus,

$$\begin{aligned}
\Phi^C(x) &\leq \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \frac{t}{x} \exp\left(-\frac{t^2}{2}\right) dt, \\
&= \frac{1}{\sqrt{2\pi}} \left[\frac{-1}{x} \exp\left(-\frac{t^2}{2}\right) \right]_{t=x}^{\infty}, \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{1}{2} x^2\right).
\end{aligned}$$

Next we prove the lower bound. We define the auxiliary function,

$$\Psi(x) = \Phi^C(x) - \frac{1}{\sqrt{2\pi}} \frac{x}{1+x^2} \exp\left(-\frac{1}{2} x^2\right).$$

The derivative of Ψ is strictly negative for $x > 0$,

$$\Psi'(x) = -\frac{1}{\sqrt{2\pi}} \frac{2}{(1+x^2)^2} \exp\left(-\frac{x^2}{2}\right),$$

and in the limit towards infinity we have $\lim_{x \rightarrow \infty} \Psi(x) = 0$. Thus Ψ is non-negative and the lower bound follows. □

As a nice corollary of this lemma we now have a good grasp of how $\Phi^C(x)$ evolves as x becomes large. For any $t > 0$ the complement of the cumulative uni-variate Gaussian distribution $\Phi_t^C(x) : (t, \infty] \rightarrow \mathbb{R} : x \rightarrow \Phi^C(x)$ restricted to the interval $[t, \infty)$ is asymptotically bounded from above and below by a constant times $\frac{1}{x} \exp(-\frac{1}{2}x^2)$. In Landau-notation this is written as $\Phi_t^C(x) = \Theta\left(\frac{1}{x} \exp(-\frac{1}{2}x^2)\right)$.

A secondary interesting bound can be derived from this lemma. Since the natural logarithm is a strictly increasing function, we find

$$-\log \Phi^C(x) \geq \frac{1}{2} \log 2\pi + \log x + \frac{x^2}{2}.$$

Thus, for $x \geq 1/\sqrt{2\pi}$, we have the following bound on x in terms of $\Phi^C(x)$,

$$x \leq \sqrt{-2 \log \Phi^C(x)}. \quad (4.1)$$

4.2 Shifted mean

In this section, we will apply the theory of Chapter 3. First, we will determine the exponentially tilted version of the Gaussian distribution. After that, we will discuss the efficiency of estimators obtained from importance sampling using the exponentially tilted distribution. Under the right circumstances, the exponentially tilted distribution will provide a logarithmically efficient set of estimators for determining $\mathbb{P}(\mathbf{X} \in A : \mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma))$, where $A \in \mathbb{R}^d$ Borel-measurable. Less convenient situations will be dealt with by introducing *mixed importance sampling*.

In Section 3.3 we saw that the univariate Gaussian distribution belongs to the exponential family. The same holds true in the multivariate case. Thus, we expect to be able to find the exponentially tilted distribution without too much inconvenience. The moment generating function of a d -dimensional multivariate Gaussian distribution $\mathcal{N}_d(\mu, \Sigma)$ is,

$$\begin{aligned} M_X(\theta) &= \frac{1}{\sqrt{|2\pi\Sigma|}} \int_{\mathbb{R}^d} \exp(\theta^\top \mathbf{x}) \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right) d\mathbf{x} \\ &= \frac{1}{\sqrt{|2\pi\Sigma|}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu - \Sigma\theta)^\top \Sigma^{-1}(\mathbf{x} - \mu - \Sigma\theta)\right) \exp\left(\mu^\top \theta + \frac{1}{2}\theta^\top \Sigma \theta\right) d\mathbf{x} \\ &= \exp\left(\mu^\top \theta + \frac{1}{2}\theta^\top \Sigma \theta\right). \end{aligned}$$

This leads to the exponentially tilted measure,

$$\mathbb{P}_\theta(X \in d\mathbf{x}) = \exp\left(\theta^\top \mathbf{x} - \mu^\top \theta - \frac{1}{2}\theta^\top \Sigma \theta\right) \mathbb{P}(X \in d\mathbf{x}).$$

The density function of this exponentially tilted measure is,

$$\begin{aligned} p_\theta(\mathbf{x}) &= \exp\left(\theta^\top \mathbf{x} - \mu^\top \theta - \frac{1}{2}\theta^\top \Sigma \theta\right) \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right) \\ &= \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu - \Sigma\theta)^\top \Sigma^{-1}(\mathbf{x} - \mu - \Sigma\theta)\right). \end{aligned}$$

Thus, the exponentially tilted measure of the multivariate Gaussian distribution simply entails a shift of the mean over $\Sigma\theta$. This is therefore also called the *shifted mean distribution*. We will refer to the importance sampling estimator obtained by using the shifted mean distribution as the *shifted mean estimator*. The benefits of the shifted mean distribution are that it both resembles the original distribution and it is quite easy to sample from.

At this point, we introduce a distance function, that is particularly useful in describing the multivariate Gaussian distribution.

Definition 4.2.1. Let $D_\Sigma(\mathbf{x}, \mathbf{y})$ denote the Mahalanobis distance [27] specific to the $\mathcal{N}_d(\mu, \Sigma)$ distribution with covariance matrix Σ between two point \mathbf{x} and \mathbf{y} in \mathbb{R}^d ,

$$D_\Sigma(\mathbf{x}, \mathbf{y}) := \sqrt{(\mathbf{x} - \mathbf{y})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{y})}.$$

When $\mathcal{N}_d(\mu, \Sigma)$ is degenerate, the Mahalanobis distance uses the generalized inverse Σ^\dagger and is only defined on the support of $\mathcal{N}_d(\mu, \Sigma)$,

$$D_\Sigma(\mathbf{x}, \mathbf{y}) := \sqrt{(\mathbf{x} - \mathbf{y})^\top \Sigma^\dagger (\mathbf{x} - \mathbf{y})}, \quad \mathbf{x}, \mathbf{y} \in \text{supp}(\mathcal{N}_d(\mu, \Sigma)).$$

The density $p(\mathbf{x})$ of a $\mathcal{N}_d(\mu, \Sigma)$ distributed random variable can now be written as,

$$p(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}D_\Sigma(\mathbf{x}, \mu)^2\right).$$

Next, we will determine the variance of the shifted mean estimator, but first we need a short lemma that is a slightly more general version of Apollonius' theorem.

Lemma 4.2.1. Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d$ and let Σ be a $d \times d$ positive definite matrix, then

$$D_\Sigma(\mathbf{x}, \mathbf{y})^2 + D_\Sigma(\mathbf{x}, \mathbf{z})^2 = 2D_\Sigma(\mathbf{x}, \frac{\mathbf{y} + \mathbf{z}}{2})^2 + \frac{1}{2}D_\Sigma(\mathbf{y}, \mathbf{z})^2.$$

Proof.

$$\begin{aligned} D_\Sigma(\mathbf{x}, \mathbf{y})^2 + D_\Sigma(\mathbf{x}, \mathbf{z})^2 &= (\mathbf{x} - \mathbf{y})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{y}) + (\mathbf{x} - \mathbf{z})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{z}) \\ &= 2(\mathbf{x} - \frac{\mathbf{y} + \mathbf{z}}{2})^\top \Sigma^{-1} (\mathbf{x} - \frac{\mathbf{y} + \mathbf{z}}{2}) + \frac{1}{2}(\mathbf{y} - \mathbf{z})^\top \Sigma^{-1} (\mathbf{y} - \mathbf{z}) \\ &= 2D_\Sigma(\mathbf{x}, \frac{\mathbf{y} + \mathbf{z}}{2})^2 + \frac{1}{2}D_\Sigma(\mathbf{y}, \mathbf{z})^2. \end{aligned}$$

□

Theorem 4.2.1. Let p and q be respectively distributed as $\mathcal{N}_d(\mu, \Sigma)$ and $\mathcal{N}_d(\nu, \Sigma)$, and let $p(\mathbf{x})$ and $q(\mathbf{x})$ denote their density functions. Let $A \in \mathbb{R}^d$ be Borel-measurable and let $\mathbf{p} = \mathbb{P}_p(X \in A)$. The estimator variance σ^2 using q as importance sampling distribution is given by,

$$\sigma^2 = -\mathbf{p}^2 + \exp(D_\Sigma(\mu, \nu)^2) \mathbb{P}_r(X \in A), \quad (4.2)$$

where r is distributed as $\mathcal{N}_d(2\mu - \nu, \Sigma)$.

Proof. By Equation 3.4 we find,

$$\begin{aligned}\sigma^2 &= -\mathbf{p}^2 + \int_{\mathbb{R}^d} \frac{(\mathbb{1}_A(\mathbf{x})p(\mathbf{x}))^2}{q(\mathbf{x})} d\mathbf{x} \\ &= -\mathbf{p}^2 + \int_A \frac{p(\mathbf{x})^2}{q(\mathbf{x})} d\mathbf{x} \\ &= -\mathbf{p}^2 + \frac{1}{\sqrt{|2\pi\Sigma|}} \int_A \frac{\exp(-D_\Sigma(\mathbf{x}, \mu)^2)}{\exp(-\frac{1}{2}D_\Sigma(\mathbf{x}, \nu)^2)} d\mathbf{x}.\end{aligned}$$

By Lemma 4.2.1, we now find,

$$\sigma^2 = -\mathbf{p}^2 + \exp(D_\Sigma(\mu, \nu)^2) \frac{1}{\sqrt{|2\pi\Sigma|}} \int_A \exp\left(-\frac{1}{2}D_\Sigma(\mathbf{x}, 2\mu - \nu)^2\right) d\mathbf{x},$$

and the stated theorem follows. \square

A natural question that arises following the introduction of the shifted mean is: for which choice of shifted mean ν , will the estimator variance be small? More precisely, we want to know whether it is possible to choose ν such, as to obtain a set of logarithmically efficient estimators. To answer that question, we also need to have knowledge of the asymptotic behaviour of $\mathbb{P}(\mathbf{X}_n \in A)$, where \mathbf{X}_n is a random variable that is somehow dependent on a rarity parameter $n \in \mathbb{N}$.

We consider a family of probability measures $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ that are distributed according to $\mathcal{N}_d(\mu, \Sigma/n)$ and depend on a rarity parameter $n \in \mathbb{N}$. For a Borel-measurable set $A \in \mathbb{R}^d$, we ensure the condition that $\mathbf{p}_n = \mathbb{P}_n(A) \rightarrow 0$ as $n \rightarrow \infty$, by assuming that $\inf_{\mathbf{x} \in A^\circ} \|\mathbf{x} - \mu\|_2 > 0$. This is a very mild assumption, since without it, we would for large n no longer be speaking about rare events.

For each $n \in \mathbb{N}$, the probability measure \mathbb{P}_n follows the law of the averaged sum of n i.i.d. $\mathcal{N}_d(\mu, \Sigma)$ distributed random variables. Thus, we can use Cramér's theorem 3.4.1 to find the rate function.

The cumulant generating function of an $\mathcal{N}_d(\mu, \Sigma)$ distributed random variable \mathbf{X} is given by $\Lambda(\lambda) = \mu^\top + \frac{1}{2}\lambda^\top \Sigma \lambda$. Its Legendre-transform is

$$\Lambda^*(\mathbf{x}) = \sup_{\lambda \in \mathbb{R}^d} \{\lambda^\top \mathbf{x} - \Lambda(\lambda)\}.$$

By taking the derivative with respect to λ ,

$$\frac{\partial}{\partial \lambda} (\lambda^\top \mathbf{x} - \Lambda(\lambda)) = \mathbf{x} - \mu - \Sigma \lambda,$$

we see that $\Lambda^*(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)$ for $\lambda = \Sigma^{-1}(\mathbf{x} - \mu)$. Thus, the family $\{\mathbb{P}_n\}_{n \in \mathbb{N}}$ of $\mathcal{N}_d(\mu, \Sigma/n)$ distributed probability measures satisfies the large deviation principle with good rate function $I(\mathbf{x}) = \frac{1}{2}D_\Sigma(\mathbf{x}, \mu)^2$.

By Definition 3.4.2 this yields,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(A) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(A^\circ) = - \inf_{\mathbf{x} \in A^\circ} \frac{1}{2} D_\Sigma(\mu, \mathbf{x})^2 := -\frac{1}{2} D_\Sigma(\mathbf{x}, A^\circ)^2. \quad (4.3)$$

Remark. We consider the probability of the interior A° , since this is regular open. Therefore, the inequalities in Definition 3.4.2 will become equalities.

Let σ_n^2 denote the estimator variance of the shifted mean estimator of \mathbf{p}_n using mean $\nu \in \mathbb{R}^d$ for all $n \in \mathbb{N}$. Similarly, we find by virtue of Theorem 4.2.1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \sigma_n^2 = -\frac{1}{2} D_\Sigma(2\mu - \nu, A^\circ)^2 + D_\Sigma(\mu, \nu)^2.$$

By our assumption that $\inf_{\mathbf{x} \in A^\circ} \|\mathbf{x} - \mu\|_2 > 0$, we also have $D_\Sigma(\mu, A^\circ)^2 > 0$. Furthermore, we restrict ourselves to consider only those values of $\nu \in \mathbb{R}^d$, for which it holds that $-\frac{1}{2} D_\Sigma(2\mu - \nu, A^\circ)^2 + D_\Sigma(\mu, \nu)^2 \leq 0$. Otherwise, the estimator variance would tend towards infinity as the rarity parameter increases, i.e. $\lim_{n \rightarrow \infty} \sigma_n^2 = \infty$.

Thus, logarithmic efficiency, see Equation 3.5, will be attained when

$$\begin{aligned} 1 &\leq \liminf_{n \rightarrow \infty} \frac{\log \sigma_n^2}{\log \mathbb{P}_n^2} \\ &= \frac{\lim_{n \rightarrow \infty} \frac{1}{n} \log \sigma_n^2}{\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{p}_n^2} \\ &= \frac{\frac{1}{2} D_\Sigma(2\mu - \nu, A^\circ)^2 - D_\Sigma(\mu, \nu)^2}{D_\Sigma(\mu, A^\circ)^2}, \end{aligned}$$

or equivalently,

$$D_\Sigma(\mu, A^\circ)^2 \leq \frac{1}{2} D_\Sigma(2\mu - \nu, A^\circ)^2 - D_\Sigma(\mu, \nu)^2. \quad (4.4)$$

Since D_Σ is a distance function, we may invoke the triangle inequality,

$$D_\Sigma(2\mu - \nu, A^\circ) \leq D_\Sigma(2\mu - \nu, \mu) + D_\Sigma(\mu, A^\circ) = D_\Sigma(\mu, \nu) + D_\Sigma(\mu, A^\circ).$$

Hence we find,

$$D_\Sigma(2\mu - \nu, A^\circ)^2 \leq (D_\Sigma(\mu, \nu) + D_\Sigma(\mu, A^\circ))^2.$$

Combined with Equation 4.4 this yields,

$$0 \leq 2D_\Sigma(\mu, \nu)D_\Sigma(\mu, A^\circ) - D_\Sigma(\mu, \nu)^2 - D_\Sigma(\mu, A^\circ)^2.$$

Thus strict inequality in Equation 4.4 can never be achieved, and equality can only be achieved, when

$$\frac{1}{2} D_\Sigma(2\mu - \nu, A^\circ) = D_\Sigma(\mu, \nu) = D_\Sigma(\mu, A^\circ). \quad (4.5)$$

The only candidate for $\nu \in \mathbb{R}^d$ to satisfy this is,

$$\nu = \arg \inf_{\mathbf{y} \in A^\circ} D_\Sigma(\mu, \mathbf{y}). \quad (4.6)$$

When this value of ν does not satisfy Equation 4.5, it is sometimes still possible to end up with a logarithmically efficient set of estimators by extending the shifted mean importance sampling distribution to incorporate mixed importance sampling [28].

Instead of using a single shifted mean distribution, we may combine several different shifted mean distributions into a single *mixture distribution*.

Definition 4.2.2. Let $q_1(\mathbf{x}), q_2(\mathbf{x}), \dots, q_m(\mathbf{x})$ be probability density functions on \mathbb{R}^d and let w_1, w_2, \dots, w_m be weights such that $\sum_{i=1}^m w_i = 1$ and $w_i > 0$ for all $i = 1, \dots, m$. Then the *mixture distribution* is the distribution with density function,

$$q(\mathbf{x}) = \sum_{i=1}^m w_i q_i(\mathbf{x}).$$

If the original density function p is included in this mixture it is also sometimes referred to as a defensive distribution. We present newly found necessary and sufficient conditions for a mixture distribution of multiple shifted mean distributions to lead to a logarithmically efficient set of estimators.

Theorem 4.2.2. *Let $p^{(n)}$ and $q_1^{(n)}, \dots, q_m^{(n)}$ for $n \in \mathbb{N}$ be respectively distributed as $\mathcal{N}_d(\mu, \Sigma/n)$ and $\mathcal{N}_d(\nu_1, \Sigma/n), \dots, \mathcal{N}_d(\nu_m, \Sigma/n)$, and let $p^{(n)}(\mathbf{x})$ and $q_1^{(n)}(\mathbf{x}), \dots, q_m^{(n)}(\mathbf{x})$ denote their density functions. Let $A \in \mathcal{B}(\mathbb{R}^d)$ be Borel-measurable and let $\mathbf{p}_n = \mathbb{P}_{p^{(n)}}(X \in A)$. Let A_1, \dots, A_m be the regular closed sets,*

$$A_i = \{\mathbf{x} \in \mathbb{R}^d : D_\Sigma(\mu, A^\circ)^2 \leq \frac{1}{2}D_\Sigma(2\mu - \nu_i, \mathbf{x})^2 - D_\Sigma(\mu, \nu_i)^2\}, \quad i = 1, \dots, m.$$

Let $\hat{\mathbf{p}}_n$ be the estimators of \mathbf{p}_n obtained via importance sampling with a mixture of shifted mean distributions with density function $q^{(n)}(\mathbf{x}) = \sum_{i=1}^m w_i q_i^{(n)}(\mathbf{x})$, where the weights w_i satisfy $\sum_{i=1}^m w_i = 1$ and $w_i > 0$ for all $i = 1, \dots, m$. Then the family of estimators $\{\hat{\mathbf{p}}_n\}_{n \in \mathbb{N}}$ is logarithmically efficient if and only if the sets $A_i, i = 1, \dots, m$ form a cover of A° ,

$$A^\circ \subset \bigcup_{i=1}^m A_i.$$

Proof. \Leftarrow : Suppose that the sets $A_i, i = 1, \dots, m$ form a cover of A° . Let σ_n^2 denote the estimator variance of $\hat{\mathbf{p}}_n$. By Equation 3.4 we find,

$$\begin{aligned} \sigma_n^2 &= -\mathbf{p}_n^2 + \int_{A^\circ} \frac{p^{(n)}(\mathbf{x})^2}{q^{(n)}(\mathbf{x})} d\mathbf{x} \\ &\leq -\mathbf{p}_n^2 + \sum_{i=1}^m \int_{A_i} \frac{p^{(n)}(\mathbf{x})^2}{q^{(n)}(\mathbf{x})} d\mathbf{x} \\ &\leq -\mathbf{p}_n^2 + \sum_{i=1}^m \frac{1}{w_i} \int_{A_i} \frac{p^{(n)}(\mathbf{x})^2}{q_i^{(n)}(\mathbf{x})} d\mathbf{x}. \end{aligned}$$

By virtue of the proof of Theorem 4.2.1, we see that,

$$\sigma_n^2 \leq -\mathbf{p}_n^2 + \sum_{i=1}^m \frac{1}{w_i} \exp(nD_\Sigma(\mu, \nu_i)^2) \mathbb{P}_{r_i^{(n)}}(\mathbf{X} \in A_i),$$

where $r_i^{(n)}$ is distributed as $\mathcal{N}_d(2\mu - \nu_i, \Sigma/n)$. Thus, we find by Equation 4.3,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \sigma_n^2 &\leq \max_{i=1, \dots, m} \left\{ D_\Sigma(\mu, \nu_i)^2 - \frac{1}{2}D_\Sigma(2\mu - \nu_i, A_i)^2 \right\} \\ &\leq -D_\Sigma(\mu, A^\circ)^2, \end{aligned}$$

and, as $\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{p}_n = -\frac{1}{2}D_\Sigma(\mu, A^\circ)^2$ by Equation 4.3, logarithmic efficiency is attained,

$$\liminf_{n \rightarrow \infty} \frac{\log \sigma_n^2}{\log \mathbf{p}_n^2} \geq 1.$$

\Rightarrow : Suppose that the sets $A_i, i = 1, \dots, m$ do not form a cover of A° . Then, since $A^\circ \setminus \cup_{i=1}^m A_i$ is open, there is some open set $A' \subset A^\circ \setminus \cup_{i=1}^m A_i$ such that for some $\delta > 0$,

$$D_\Sigma(\mu, A^\circ)^2 \geq \frac{1}{2}D_\Sigma(2\mu - \nu_i, \mathbf{x})^2 - D_\Sigma(\mu, \nu_i)^2 + \delta, \quad \text{for all } \mathbf{x} \in A', i = 1, \dots, m.$$

By Equation 3.4, we find,

$$\begin{aligned}
\sigma_n^2 &= -\mathfrak{p}_n^2 + \int_{A^\circ} \frac{p^{(n)}(\mathbf{x})^2}{q^{(n)}(\mathbf{x})} d\mathbf{x} \\
&\geq -\mathfrak{p}_n^2 + \int_{A'} \frac{p^{(n)}(\mathbf{x})^2}{q^{(n)}(\mathbf{x})} d\mathbf{x} \\
&\geq -\mathfrak{p}_n^2 + \int_{A'} \frac{1}{\sqrt{|2\pi\Sigma/n|}} \min_{i=1,\dots,m} \exp\left(-nD_\Sigma(\mathbf{x}, \mu)^2 + \frac{n}{2}D_\Sigma(\mathbf{x}, \nu_i)^2\right) d\mathbf{x}.
\end{aligned}$$

Lemma 4.2.1 now yields,

$$\begin{aligned}
\sigma_n^2 &\geq -\mathfrak{p}_n^2 + \int_{A'} \frac{1}{\sqrt{|2\pi\Sigma/n|}} \min_{i=1,\dots,m} \exp\left(nD_\Sigma(\mu, \nu_i)^2 - \frac{n}{2}D_\Sigma(\mathbf{x}, 2\mu - \nu_i)^2\right) d\mathbf{x} \\
&\geq -\mathfrak{p}_n^2 + \int_{A'} \frac{1}{\sqrt{|2\pi\Sigma/n|}} \exp\left(-nD_\Sigma(\mu, A^\circ)^2 + n\delta\right) d\mathbf{x} \\
&= -\mathfrak{p}_n^2 + \lambda_d(A') \frac{1}{\sqrt{|2\pi\Sigma/n|}} \exp\left(-nD_\Sigma(\mu, A^\circ)^2 + n\delta\right),
\end{aligned}$$

where $\lambda_d(A') > 0$ is the d -dimensional Lebesgue-measure of the set A' . Thus, we find,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \sigma_n^2 \geq -D_\Sigma(\mu, A^\circ)^2 + \delta,$$

and logarithmic efficiency is not attained,

$$\liminf_{n \rightarrow \infty} \frac{\log \sigma_n^2}{\log \mathfrak{p}_n^2} = \frac{\lim_{n \rightarrow \infty} \frac{1}{n} \log \sigma_n^2}{\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathfrak{p}_n^2} = \frac{-D_\Sigma(\mu, A^\circ)^2 + \delta}{-D_\Sigma(\mu, A^\circ)^2} < 1.$$

□

There exist events $A \in \mathcal{B}(\mathbb{R}^d)$ in combination with distributions $\mathcal{N}_d(\mu, \Sigma)$ for which there is no finite set $\{\nu_1, \dots, \nu_m\} \subset \mathbb{R}^d$, such that A° is covered by the sets,

$$A_i = \{\mathbf{x} \in \mathbb{R}^d : D_\Sigma(\mu, A^\circ)^2 \leq \frac{1}{2}D_\Sigma(2\mu - \nu_i, \mathbf{x})^2 - D_\Sigma(\mu, \nu_i)^2\}, \quad i = 1, \dots, m.$$

However, a logarithmically efficient set of estimators can still be obtained via importance sampling with a mixture of an infinite amount of shifted mean distributions. In [29], it is shown, that this can be done by sampling from a mixed shifted mean distribution, where the shifted mean is chosen uniformly at random on the surface of the ellipsoid $\{\mathbf{x} \in \mathbb{R}^d : D_\Sigma(\mu, \mathbf{x}) = D_\Sigma(\mu, A^\circ)\}$.

It should be noted, that despite our focus on logarithmic efficiency in this section, it should not be viewed as the only indicator of the quality of a set of estimators. In the next section, we will use Theorem 4.2.2 to find a logarithmically efficient set of estimators when the event $A \in \mathcal{B}(\mathbb{R}^d)$ is the exterior of a polyhedron. Furthermore, we will also provide bounds on the relative standard deviation.

4.3 Gaussian in the exterior of a polyhedron

In the previous section, we discovered conditions that will ensure logarithmically efficient estimators for shifted mean distributions and mixtures thereof for the Gaussian

distribution and some general event $A \in \mathcal{B}(\mathbb{R}^d)$. In this section, we will construct these logarithmically efficient estimators for the model of the DC-approximation with stochastic power injections.

In Section 2.3, we relate the power injections to feasibility constraints within the network. We consider a network with d load-buses and generator-buses and one reference bus. The real power injections of the d non-reference buses are modeled with a Gaussian random vector $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma/n)$. Its probability measure is given by \mathbb{P}_n . A configuration of power injections \mathbf{X} satisfies the feasibility constraints if and only if it lies inside some polyhedron $A^C = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}_i^\top \mathbf{x} \leq b_i, i = 1, \dots, m\}$, where $\mathbf{a}_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$.

The exterior of this polyhedron, A , is a union of half-spaces,

$$H_i = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}_i^\top \mathbf{x} > b_i\}, \quad i = 1, \dots, m.$$

First, we will find a logarithmically efficient set of shifted mean estimators for the probabilities $\mathbf{p}_n := \mathbb{P}_n(\mathbf{X} \in H)$ of a Gaussian random variable $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma/n)$ lying in a single half-space $H = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}^\top \mathbf{x} > b\}$, $\mathbf{a} \in \mathbb{R}^d$, $b \in \mathbb{R}$. After that, we will make the extension to the union of half-spaces using a mixture of shifted mean distributions.

By virtue of Theorem 4.1.2 we are able to give an explicit expression for \mathbf{p}_n in terms of the complement of the univariate Gaussian cumulative distribution Φ^C ,

$$\mathbf{p}_n = \Phi^C \left(\sqrt{n} \frac{b - \mu^\top \mathbf{a}}{\|\sqrt{\Sigma} \mathbf{a}\|_2} \right).$$

In spite of this, we also wish to find an importance sampling distribution, that we can use to determine this probability.

We choose to use shifted mean importance sampling with as mean the unique point satisfying Equation 4.6,

$$\nu = \arg \inf_{y \in H} D_\Sigma(\mu, \mathbf{y}).$$

This will lead to a logarithmically efficient set of estimators $\{\hat{\mathbf{p}}_n\}_{n \in \mathbb{N}}$ for $\{\mathbb{P}_n(H)\}_{n \in \mathbb{N}}$.

We have to find which value of ν satisfies this. We will show, that the constant

$$\beta := \frac{b - \mu^\top \mathbf{a}}{\|\sqrt{\Sigma} \mathbf{a}\|_2},$$

equals the Mahalanobis distance $D_\Sigma(\mu, H)$, and that,

$$\nu = D_\Sigma(\mu, H) = \mu + \frac{\beta}{\|\sqrt{\Sigma} \mathbf{a}\|_2} \Sigma \mathbf{a}.$$

To show this, we use an affine transformation $L : \mathbb{R}^d \rightarrow \mathbb{R}^d : \mathbf{x} \rightarrow \sqrt{\Sigma}^{-1}(\mathbf{x} - \mu)$ on the probability space and on H . This gives us the equivalent problem of determining the point in the half-space $L(H) = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x}^\top \sqrt{\Sigma} \mathbf{a} + \mu^\top \mathbf{a} \geq b\}$ that has the highest probability density for the distribution $\mathcal{N}_d(0, I_n)$. Thanks to the radial property of the density function of the standard normal distribution, this is the closest point in $L(H)$ to the origin, namely $\frac{b - \mu^\top \mathbf{a}}{\|\sqrt{\Sigma} \mathbf{a}\|_2} \sqrt{\Sigma} \mathbf{a} = \frac{\beta}{\|\sqrt{\Sigma} \mathbf{a}\|_2} \sqrt{\Sigma} \mathbf{a}$. Using the inverse transformation, we then find

$$\nu = L^{-1} \left(\frac{\beta}{\|\sqrt{\Sigma} \mathbf{a}\|_2} \sqrt{\Sigma} \mathbf{a} \right) = \mu + \frac{\beta}{\|\sqrt{\Sigma} \mathbf{a}\|_2} \Sigma \mathbf{a}. \quad (4.7)$$

This demonstrates that β is simply the Mahalanobis distance between the mean μ of the distribution and the half-space H ,

$$D_{\Sigma}(\mu, H) = D_{\Sigma}(\mu, \nu) = \sqrt{(\mu - (\mu + \frac{\beta}{\|\sqrt{\Sigma}\mathbf{a}\|_2}\Sigma\mathbf{a}))^{\top}\Sigma^{-1}(\mu - (\mu + \frac{\beta}{\|\sqrt{\Sigma}\mathbf{a}\|_2}\Sigma\mathbf{a}))} = \beta.$$

We take a closer look at the estimator variance and the relative standard variation. For all $n \in \mathbb{N}$ we find by Theorem 4.1.2 that $\mathbf{p}_n = \mathbb{P}_n(H) = \Phi^C(\sqrt{n}\beta)$. Inserting ν into Theorem 4.2.1, we find for all $n \in \mathbb{N}$ the estimator variance,

$$\sigma_n^2 = -\mathbf{p}_n^2 + \exp(n\beta^2) \Phi^C(2\sqrt{n}\beta).$$

To present a clearer picture of the evolution of σ_n^2 as n increases, we apply the bounds of Lemma 4.1.2. First, applying the upper bound on Φ^C we find,

$$\sigma_n^2 \leq -\mathbf{p}_n^2 + \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{n}\beta} \exp(-n\beta^2).$$

The lower bound on $\mathbf{p}_n = \Phi^C(\sqrt{n}\beta)$ can be used to give an upper bound on $\exp(-\frac{n}{2}\beta^2)$,

$$\exp(-\frac{n}{2}\beta^2) \leq \sqrt{2\pi} \frac{1+n\beta^2}{\sqrt{n}\beta} \mathbf{p}_n.$$

Combining these expressions, we find a single bound on the estimator variance in terms of the probability \mathbf{p}_n ,

$$\sigma_n^2 \leq -\mathbf{p}_n^2 + \mathbf{p}_n^2 \sqrt{2\pi} \frac{(1+n\beta^2)^2}{2n\sqrt{n}\beta^3}. \quad (4.8)$$

Likewise, for all $n \in \mathbb{N}$ we get the bound on the relative standard deviation,

$$\text{RSD}[\hat{\mathbf{p}}_n] \leq \sqrt{\sqrt{2\pi} \frac{(1+n\beta^2)^2}{2n\sqrt{n}\beta^3} - 1}.$$

For $\sqrt{n}\beta \geq 1$, we can apply the inequality in Equation 4.1, $\sqrt{n}\beta \leq \sqrt{-2 \log \Phi^C(\sqrt{n}\beta)}$, to express the relative standard deviation in terms of the probability $\mathbf{p}_n = \Phi^C(\sqrt{n}\beta)$,

$$\text{RSD}[\hat{\mathbf{p}}_n] \leq \sqrt{2\sqrt{2\pi}\sqrt{n}\beta^2 - 1} \leq \sqrt{4\sqrt{\pi}\sqrt{-\log \mathbf{p}_n} - 1}.$$

For arbitrarily large $n \in \mathbb{N}$, this bound can be tightened to,

$$\text{RSD}[\hat{\mathbf{p}}_n] \leq \sqrt{\frac{1}{2}\sqrt{2\pi}\sqrt{n}\beta^2 - 1} \leq \sqrt{\sqrt{\pi}\sqrt{-\log \mathbf{p}_n} - 1}.$$

Now, we move our attention to A , the exterior of a polyhedron. As mentioned earlier, we consider the set A to be the union of half-spaces $H_i = \{x \in \mathbb{R}^d : \mathbf{x}^{\top} \mathbf{a}_i > b_i\}$, where $\mathbf{a}_i \in \mathbb{R}^d$, $b_i \in \mathbb{R}$, $i = 1, \dots, m$. Let for all $n \in \mathbb{N}$ p_n be the density function of $\mathcal{N}_d(\mu, \Sigma/n)$, where $\mu \notin A$. For convenience, we define as before the constant,

$$\beta_i := \frac{b_i - \mu^{\top} \mathbf{a}_i}{\|\sqrt{\Sigma} \mathbf{a}_i\|_2}, \quad (4.9)$$

which is the Mahalanobis distance $D_\Sigma(\mu, H_i)$ between μ and the half-space H_i for each $i = 1, \dots, m$. Now the probability of sampling in any of the half-spaces H_i is $\mathbf{p}_{n,i} = \mathbb{P}_{p_n}(H_i) = \Phi^C(\sqrt{n}\beta_i)$. We consider as possible mean ν_i the most likely point on each of the half-spaces, similarly to Equation 4.7, namely,

$$\nu_i = \mu + \frac{\beta_i}{\|\sqrt{\Sigma}\mathbf{a}_i\|_2} \Sigma \mathbf{a}_i, \quad (4.10)$$

for all $i = 1, \dots, m$.

Now for any choice of weights $w_i, i = 1, \dots, m$, we find by Theorem 4.2.2 the following bound on the estimator variance obtained by means of importance sampling for all $n \in \mathbb{N}$,

$$\sigma_n^2 \leq -\mathbf{p}_n^2 + \sum_{i=1}^m \frac{1}{w_i} \exp(n\beta_i^2) \Phi^C(2\sqrt{n}\beta_i). \quad (4.11)$$

Similarly to the case of a single half-space H , we find for all $i = 1, \dots, m$,

$$\exp\left(-\frac{n}{2}\beta_i^2\right) \leq \sqrt{2\pi} \frac{1 + n\beta_i^2}{\sqrt{n}\beta_i} \mathbb{P}_n(H_i) \leq \sqrt{2\pi} \frac{1 + n\beta_i^2}{\sqrt{n}\beta_i} \mathbf{p}_n.$$

Thus, we find the bound,

$$\text{RSD}[\hat{\mathbf{p}}_n] \leq \sqrt{\sqrt{2\pi} \sum_{i=1}^m \frac{1}{w_i} \frac{(1 + n\beta_i^2)^2}{2n\sqrt{n}\beta_i^3} - 1}.$$

When $\sqrt{n}\beta_i \geq 1$ for all $i = 1, \dots, m$, this reduces to,

$$\text{RSD}[\hat{\mathbf{p}}_n] \leq \sqrt{2\sqrt{2\pi} \sum_{i=1}^m \frac{1}{w_i} \sqrt{n}\beta_i - 1}.$$

To keep the bound on the relative standard deviation low, it is necessary to choose weights $w_i, i = 1, \dots, m$ such that they keep $\sum_{i=1}^m \frac{1}{w_i} \sqrt{n}\beta_i$ small. What constitutes a good choice for the weights, depends on the rarity parameter n .

Instead of fixing the weights w_i , we can also let them depend on the rarity parameter $n \in \mathbb{N}$. We use the notation $w_i(n)$ for this. This requires a slightly different version of Theorem 4.2.2 to prove logarithmic efficiency.

Theorem 4.3.1. *Let, for $n \in \mathbb{N}$, $p^{(n)}$ and $q_1^{(n)}, \dots, q_m^{(n)}$ be respectively distributed as $\mathcal{N}_d(\mu, \Sigma/n)$ and $\mathcal{N}_d(\nu_1, \Sigma/n), \dots, \mathcal{N}_d(\nu_m, \Sigma/n)$, and let $p^{(n)}(\mathbf{x})$ and $q_1^{(n)}(\mathbf{x}), \dots, q_m^{(n)}(\mathbf{x})$ denote their density functions. Let, for $n \in \mathbb{N}$, $w_i(n), i = 1, \dots, m$ be weights such that $w_i(n) > 0$, $\sum_{i=1}^m w_i(n) = 1$, and the limits $\lim_{n \rightarrow \infty} \frac{1}{n} \log w_i(n) = W_i$ exist. Let $A \in \mathcal{B}(\mathbb{R}^d)$ be Borel-measurable and let $\mathbf{p}_n = \mathbb{P}_{p^{(n)}}(X \in A)$. Let A_1, \dots, A_m be the regular closed sets,*

$$A_i = \{\mathbf{x} \in \mathbb{R}^d : D_\Sigma(\mu, A^\circ)^2 \leq \frac{1}{2} D_\Sigma(2\mu - \nu_i, \mathbf{x})^2 - D_\Sigma(\mu, \nu_i)^2 + W_i\}, \quad i = 1, \dots, m.$$

The estimators $\hat{\mathbf{p}}_n$ of \mathbf{p}_n obtained via importance sampling with a mixture of shifted mean distributions with density function $q^{(n)}(\mathbf{x}) = \sum_{i=1}^m w_i(n) q_i^{(n)}(\mathbf{x})$ is logarithmically efficient if and only if the sets $A_i, i = 1, \dots, m$ form a cover of A° ,

$$A^\circ \subset \bigcup_{i=1}^m A_i.$$

Proof. The proof follows along the same lines as the proof of Theorem 4.2.2. \square

The weights $w_i(n)$ that minimize the sum in Equation 4.11 can be found by introducing a Lagrange multiplier λ and studying the Lagrange function,

$$\mathcal{L} = \sum_{i=1}^m \frac{1}{w_i(n)} \exp(n\beta_i^2) \Phi^C(2\sqrt{n}\beta_i) - \lambda \left(\sum_{i=1}^m w_i(n) - 1 \right).$$

We find the derivatives,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_i(n)} &= -\frac{1}{w_i(n)^2} \exp(n\beta_i^2) \Phi^C(2\sqrt{n}\beta_i) - \lambda, \quad i = 1, \dots, m, \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 1 - \sum_{i=1}^m w_i(n). \end{aligned}$$

Setting both derivatives to zero, we find the weights that minimize the bound on the estimator variance,

$$w_i(n) = \frac{\exp\left(\frac{1}{2}n\beta_i^2\right) \sqrt{\Phi^C(2\sqrt{n}\beta_i)}}{\sum_{j=1}^m \exp\left(\frac{1}{2}n\beta_j^2\right) \sqrt{\Phi^C(2\sqrt{n}\beta_j)}}, \quad i = 1, \dots, m. \quad (4.12)$$

We check to see, that logarithmic efficiency will be attained. For the weights we find,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log w_i(n) = -\frac{1}{2} D_{\Sigma}(\mu, \nu_i)^2 + \frac{1}{2} D_{\Sigma}(\mu, A^\circ)^2.$$

Thus, for $i = 1, \dots, m$, the sets A_i are given by,

$$A_i = \{\mathbf{x} \in \mathbb{R}^d : D_{\Sigma}(\mu, A^\circ)^2 \leq D_{\Sigma}(2\mu - \nu_i, \mathbf{x})^2 - 3D_{\Sigma}(\mu, \nu_i)^2\},$$

while for all $\mathbf{x} \in H_i$,

$$\begin{aligned} D_{\Sigma}(2\mu - \nu_i, \mathbf{x})^2 &\geq D_{\Sigma}(2\mu - \nu_i, \nu_i)^2 \\ &= 4D_{\Sigma}(\mu, \nu_i)^2 \\ &\geq 3D_{\Sigma}(\mu, \nu_i)^2 + \min_{j=1, \dots, m} D_{\Sigma}(\mu, \nu_j)^2 \\ &= 3D_{\Sigma}(\mu, \nu_i)^2 + D_{\Sigma}(\mu, A^\circ)^2. \end{aligned}$$

Thus, $H_i \subset A_i$ for all $i = 1, \dots, m$ and the obtained set of estimators $\{\hat{\mathbf{p}}_n\}_{n \in \mathbb{N}}$ are logarithmically efficient by Theorem 4.3.1.

We again provide a bound on the estimator variance. Inserting the weights from Equation 4.12 into Equation 4.11 yields,

$$\sigma_n^2 \leq -\mathbf{p}_n^2 + \left(\sum_{i=1}^m \exp\left(\frac{1}{2}n\beta_i^2\right) \sqrt{\Phi^C(2\sqrt{n}\beta_i)} \right)^2.$$

We can now find a bound on the estimator variance in terms of the probabilities $\mathbf{p}_{n,i}$ by applying the bounds of Lemma 4.1.2,

$$\sigma_n^2 \leq -\mathbf{p}_n^2 + \sqrt{\frac{\pi}{2}} \left(\sum_{i=1}^m \mathbf{p}_{n,i} \frac{1 + n\beta_i^2}{n^{3/4} \beta_i^{3/2}} \right)^2.$$

If $\min_{i=1,\dots,m} \Phi^C(\sqrt{n}\beta_i) = \min_{i=1,\dots,m} \mathbf{p}_{n,i} \leq \mathbf{p}_n \leq \Phi^C(1) \approx .1587$, then $\sqrt{n}\beta_i \geq 1$ for all $i = 1, \dots, m$ and we can apply the inequality in Equation 4.1 to find,

$$\sigma_n^2 \leq -\mathbf{p}_n^2 + 4\sqrt{\pi} \left(\sum_{i=1}^m \mathbf{p}_{n,i} \sqrt[4]{-\log \mathbf{p}_{n,i}} \right)^2.$$

Let $\bar{\mathbf{p}}_n = \sum_{i=1}^m \mathbf{p}_{n,i}$ be the union bound, then the estimator variance and relative standard deviation satisfy,

$$\begin{aligned} \sigma_n^2 &\leq -\mathbf{p}_n^2 + 4\bar{\mathbf{p}}_n^2 \sqrt{\pi} \sqrt{-\log \mathbf{p}_n}, \\ \text{RSD}[\hat{\mathbf{p}}_N] &\leq 4 \frac{\bar{\mathbf{p}}_n^2}{\mathbf{p}_n^2} \sqrt{\pi} \sqrt{-\log \mathbf{p}_n} - 1. \end{aligned}$$

For arbitrarily large values of the rarity parameter $n \in \mathbb{N}$, this can be tightened to,

$$\text{RSD}[\hat{\mathbf{p}}_n] \leq \frac{\bar{\mathbf{p}}_n^2}{\mathbf{p}_n^2} \sqrt{\pi} \sqrt{-\log \mathbf{p}_n} - 1.$$

In the next section, we will run simulations to find the probability of infeasible configurations of power injections. We will use the data on power grids of several IEEE test-systems.

4.4 Simulation

Thus far in this chapter, we have constructed a method for finding a logarithmically efficient set of estimators for the probability of an infeasible configuration of power injections in the power grid. In this section, we will use this method to perform several simulations.

First, we extract data on the power grids present in the IEEE test system using the MATPOWER software package [9] and use it to find a realistic Gaussian distribution to model the power injections. The data does not include the distribution of the power injections, but instead only a single configuration of power injections at a specific moment of observation.

4.4.1 Model

Caution is required when constructing the probability distribution. In the DC-approximation the transmission lines are assumed to be lossless. Therefore, it becomes necessary to impose the condition that the total power injection in the network is zero.

We initially choose the mean power injections in a $d + 1$ -bus system, $\mu \in \mathbb{R}^{d+1}$, to be the power injections at the moment of observation of the data. We choose the power injection to be independent and we let the standard deviation be proportional to the mean power injection. This leads to an initial covariance matrix $\Sigma = \text{diag}(\mu)^2$.

Next, we impose the condition that the total power injection is zero. We wish to study the distribution of $\mathbf{X} | \mathbf{1}_{d+1}^\top \mathbf{X} = 0$, where $\mathbf{X} \sim \mathcal{N}_{d+1}(\mu, \Sigma)$ and $\mathbf{1} = \underbrace{(1, 1, \dots, 1)^\top}_{d+1\text{-times}}$.

By virtue of Theorem 4.1.1, this is a Gaussian distribution whose mean and the covariance matrix are given by,

$$\mathbb{E}(\mathbf{X}|\mathbf{1}^\top \mathbf{X} = 0) = \mu - \frac{\Sigma \mathbf{1} \mathbf{1}^\top \mu}{\mathbf{1}^\top \Sigma \mathbf{1}}, \text{ and,}$$

$$\text{Var}(\mathbf{X}|\mathbf{1}^\top \mathbf{X} = 0) = \Sigma - \frac{\Sigma \mathbf{1} \mathbf{1}^\top \Sigma}{\mathbf{1}^\top \Sigma \mathbf{1}}.$$

This distribution is degenerate. However, this forms no problem. We remove one element from \mathbf{X} , which we now call \mathbf{X}' , and construct the Power Transmission Distribution Factor V , see Section 2.3, such that the column corresponding to the removed element is a zero-column. We remove this column too and call the resulting matrix V' .

\mathbf{X}' is a non-degenerate Gaussian distribution. By letting b_i be the maximum power allowed to flow through branch i for all $i = 1, \dots, m$ branches, we find the infeasible half-spaces $H_i = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}_i^\top \mathbf{x} > b_i\}$ and $H_{m+i} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}_i^\top \mathbf{x} < -b_i\}$, where $\mathbf{a}_i = \mathbf{a}_{m+i} = V'_{i,\bullet}$.

We now summarize our method developed throughout this chapter in a step-by-step plan for a given Gaussian distribution $\mathcal{N}_d(\mu, \Sigma)$ with density function p and an infeasible set of configurations $A = \cup_{i=1}^m H_i$, that is a union of half-spaces.

- First, we find the Mahalanobis distance of the mean of the power injections μ to each of the half-spaces H_i ,

$$\beta_i = \inf_{\mathbf{x} \in H_i} \{D_\Sigma(\mu, \mathbf{x})\}.$$

- Second, we find the most likely point ν_i on each of the half-spaces H_i ,

$$\nu_i = \mu + \frac{\beta_i}{\|\sqrt{\Sigma} \mathbf{a}_i\|_2} \Sigma \mathbf{a}_i.$$

- Next, for some selection of rarity parameters $n \in \mathbb{N}$, we do the following:

- we determine the weights,

$$w_i(n) = \frac{\exp\left(\frac{1}{2}n\beta_i^2\right) \sqrt{\Phi^C(2\sqrt{n}\beta_i)}}{\sum_{j=1}^m \exp\left(\frac{1}{2}n\beta_j^2\right) \sqrt{\Phi^C(2\sqrt{n}\beta_j)}}, \quad i = 1, \dots, m,$$

- we draw samples $\mathbf{X}_1, \dots, \mathbf{X}_N$ from the mixed shifted mean distribution with density function,

$$q(\mathbf{x}) = \frac{n^{d/2}}{|\sqrt{2\pi}\Sigma|} \sum_{i=1}^m w_i(n) \exp\left(-\frac{n}{2}(\mathbf{x} - \nu_i)^\top \Sigma^{-1}(\mathbf{x} - \nu_i)\right),$$

- and we calculate the estimators,

$$\hat{\mathbf{p}}_n = \frac{1}{N} \sum_{i=1}^N \frac{p(\mathbf{X}_i)}{q(\mathbf{X}_i)} \mathbb{1}_A(\mathbf{X}_i),$$

and sample variances,

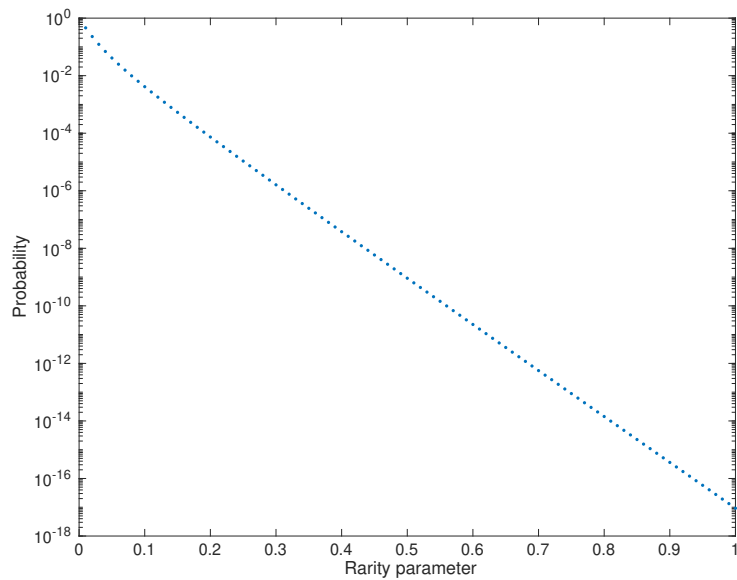
$$S_N^2 = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{p(\mathbf{X}_i)}{q(\mathbf{X}_i)} \mathbb{1}_A(\mathbf{X}_i) - \hat{\mathbf{p}}_n \right)^2.$$

In Appendix A the MATLAB code is presented, that performs simulations based on this method. We use the MATPOWER software package [9] to extract data on the power grids present in the IEEE test systems. We also use MATPOWER to solve the power flow equations.

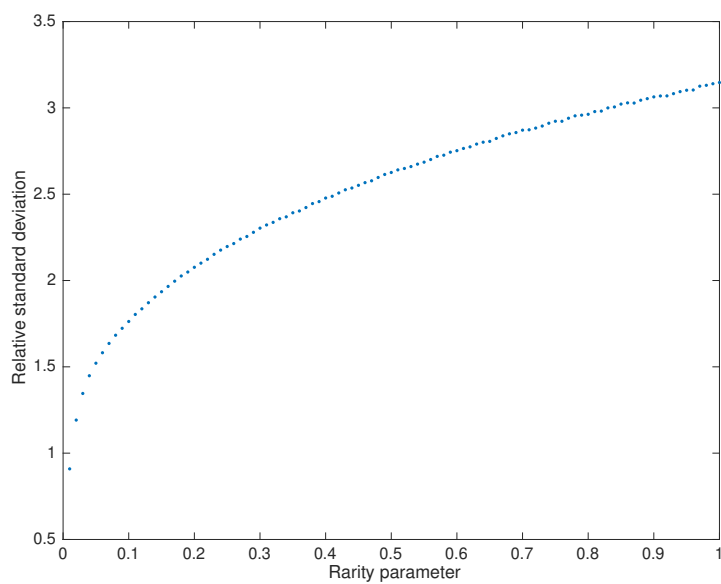
4.4.2 Results

We applied the methods to both the 3-bus network ‘case3bus_P6_6’ and the 30-bus network ‘case30bus’ in Matpower. We notice, that the probability seems to follow the exponential trend quite quickly for both networks. Both in Figure 4.2a and Figure 4.3a the logarithm of the probability seems almost a linear function of the rarity parameter once it drops below 10^{-2} .

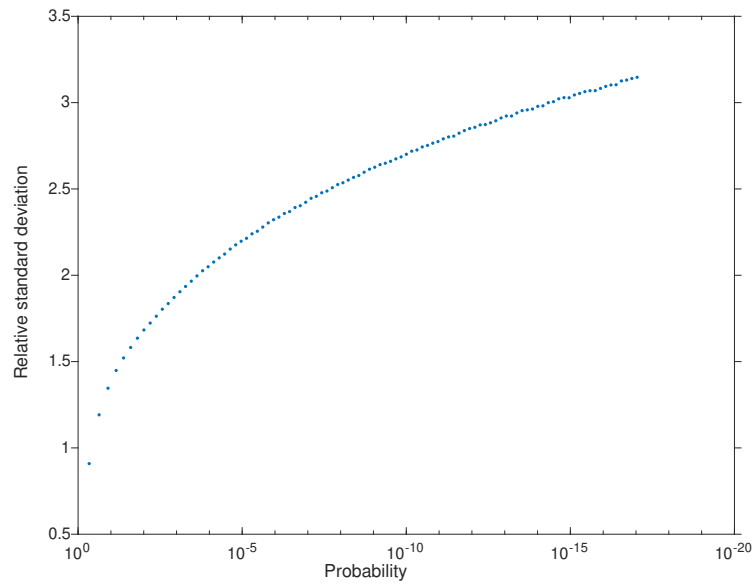
Furthermore, the relative standard deviation remains low for rare events. With a little effort one can spot the bounds on the relative standard deviation in terms of the probability in Figures 4.2c and 4.3c.



(a) Plot of the probability versus the rarity parameter.

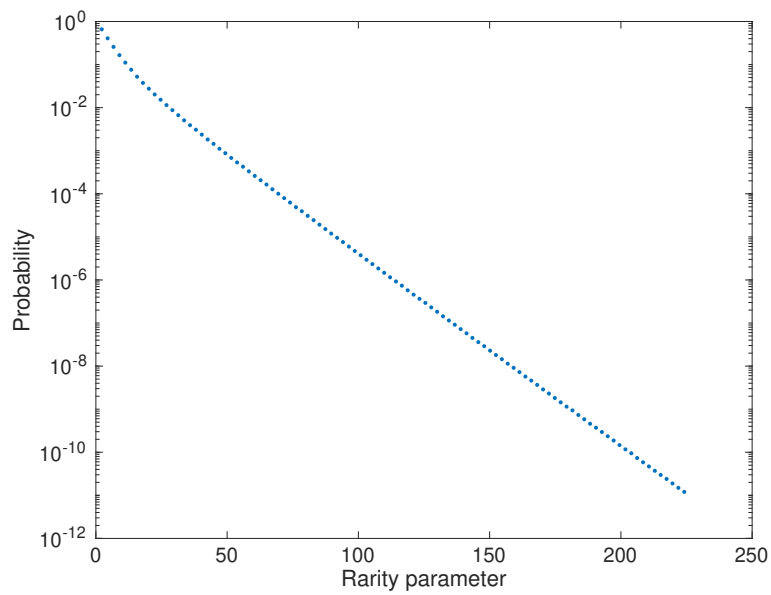


(b) Plot of the relative standard deviation versus the rarity parameter.

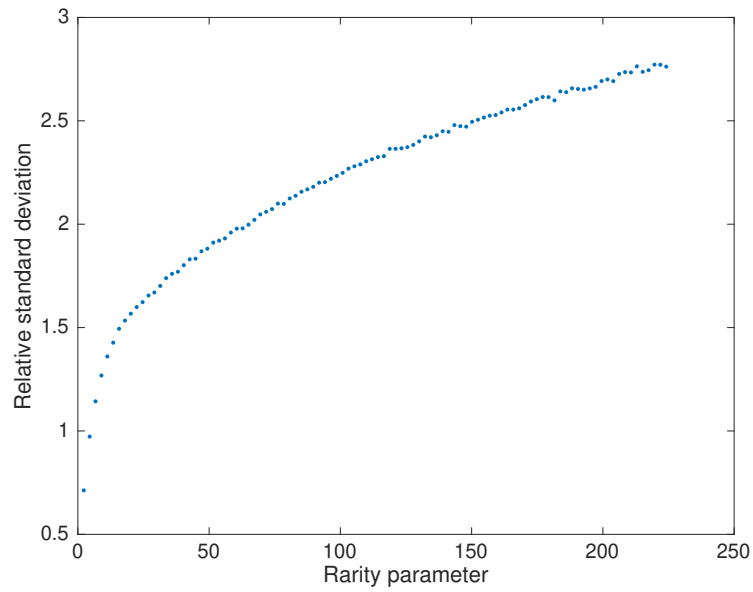


(c) Plot of the relative standard deviation versus the probability.

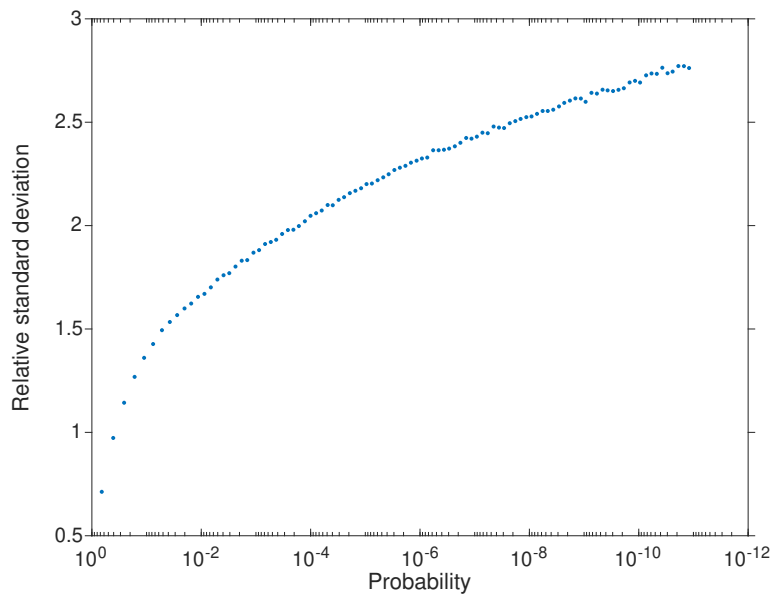
Figure 4.2: Plots of the probability and the relative standard deviation for a 3-bus system, 'case3bus_P6.6' in Matpower. We used $N = 10^6$ samples.



(a) Plot of the probability versus the rarity parameter.



(b) Plot of the relative standard deviation versus the rarity parameter.



(c) Plot of the relative standard deviation versus the probability.

Figure 4.3: Plots of the probability and the relative standard deviation for a 30-bus system, ‘case30bus’ in Matpower. We used $N = 10^6$ samples.

Chapter 5

Distflow

The DC-approximation provides a fast method to approximate the power flowing through the transmission lines. Unfortunately, this approximation can be quite rough, especially when viewed from a rare event point of view. A small difference in allowable power injections could potentially lead to orders of magnitude difference in the probability of the rare event occurring.

A second drawback of the DC-approximation lies in the assumption that all voltage magnitudes are equal. This makes it impossible to analyse which power injections lead to too low or too high a voltage.

In a random network, there are no practical alternatives to using numerical methods. However, there are certain conditions under which radial networks can be exactly studied without employing numerical methods. In this chapter, we will specifically look whether the voltages at buses along a single line will be within a specified range.

5.1 Distflow

In [4] the concept of Distflow was first introduced. Shortly after that, it was expanded upon by the same authors in [5]. The main idea centres around the fact, that the power flows and voltages in the system can be determined recursively when the voltage and power injection at one of the ends of a series of sequentially connected buses are known.

We will derive the Distflow equations in an alternative way to how they were first introduced, leading to a slightly different, but still equivalent set of equations. In the original Distflow papers a first order recurrence relation in three real variables was derived: the active power flow between two subsequent buses, the reactive power flow between two subsequent buses, and the square of the voltage magnitude at each bus. We will find a second order recurrence relation in a single complex variable: the complex voltage at each bus.

We consider a system of $n + 1$ sequentially connected buses without shunt impedances along a transmission line, see Figure 5.1 for a schematic representation. The first n buses are consumers. The final bus is a reference bus.

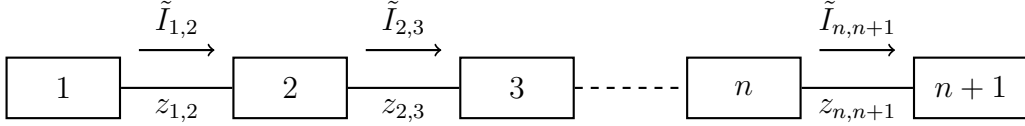


Figure 5.1: A schematic representation of $n + 1$ sequentially connected buses with impedances and currents. There are no shunt impedances.

By Kirchhoff's current law, see Theorem 2.1.1 the injected current at each bus is equal to the sum of outward flowing current from that bus. By the complex power formula, Equation 2.3, the injected current at bus i is $\tilde{I}_i = S_i^*/\tilde{V}_i^*$, while by Ohm's law, Equation 2.10, the current flowing from bus i to bus $i + 1$ is $\tilde{I}_{i,i+1} = \frac{\tilde{V}_i - \tilde{V}_{i+1}}{z_{i,i+1}}$, where $z_{i,i+1}$ is the impedance of the transmission line between buses i and $i + 1$. This results in a system of equations governing the power flow in this network,

$$\begin{aligned} \frac{S_i^*}{\tilde{V}_i^*} &= \frac{\tilde{V}_i - \tilde{V}_{i+1}}{z_{i,i+1}} - \frac{\tilde{V}_{i-1} - \tilde{V}_i}{z_{i-1,i}}, & i = 2, \dots, n, \\ \frac{S_1^*}{\tilde{V}_1^*} &= \frac{\tilde{V}_1 - \tilde{V}_2}{z_{1,2}}, \\ \frac{S_{n+1}^*}{\tilde{V}_{n+1}^*} &= \frac{\tilde{V}_{n+1} - \tilde{V}_n}{z_{n,n+1}}. \end{aligned}$$

Rearranging these terms yields a second order recursive relation,

$$\begin{aligned} \tilde{V}_{i+1} &= \left(1 + \frac{z_{i,i+1}}{z_{i-1,i}}\right) \tilde{V}_i - \frac{z_{i,i+1}}{z_{i-1,i}} \tilde{V}_{i-1} - z_{i,i+1} \frac{S_i^*}{\tilde{V}_i^*}, & i = 2, \dots, n, \\ \tilde{V}_2 &= \tilde{V}_1 - z_{1,2} \frac{S_1^*}{\tilde{V}_1^*}. \end{aligned}$$

In the hypothetical situation where V_1 is known, this is extremely helpful. When V_1 is unknown, it might still prove useful. We will study the situation where the ratio of reactance to resistance in the transmission line equals the ratio of the reactive to the real power demand of consumers.

5.2 Telescoping series

We notice that the injected currents form a telescoping series,

$$\sum_{i=1}^k \frac{S_i^*}{\tilde{V}_i^*} = \frac{\tilde{V}_k - \tilde{V}_{k+1}}{z_{k,k+1}}, \quad k = 1, \dots, n.$$

This telescoping series in turn provides a second telescoping series. We define $z'_{i,j}$ to be impedance of the transmission line between buses i and j . For $i, j \in \{1, 2, \dots, n+1\}$ with $i < j$, we thus have

$$z'_{i,j} = \sum_{k=0}^{j-i-1} z_{i+k,i+k+1}.$$

Now, we find,

$$\sum_{i=1}^l z'_{i,l+1} \frac{S_i^*}{\tilde{V}_i^*} = \sum_{k=1}^l z_{k,k+1} \sum_{i=1}^k \frac{S_i^*}{\tilde{V}_i^*} = \tilde{V}_1 - \tilde{V}_{l+1}, \quad l = 1, \dots, n. \quad (5.1)$$

Typically, transmission lines have positive resistance and reactance, in which case $z_{i,i+1}$ lies in the first quadrant of the complex plane. Consumers do usually indeed consume both active and reactive power. Therefore, the complex power injection lies in the third quadrant of the complex plane. We make the assumptions, that the argument of the impedance in all transmission lines is the same and that the argument of the power injections is exactly opposite to this, so

$$z_{i,i+1} S_j^* \in \mathbb{R}_{\leq 0}, \quad i, j = 1, \dots, n.$$

Although we mainly make this assumption for simplifying the mathematical model, it is not out of the realm of physical possibility.

Under this assumption, we clearly see, that if we choose our reference angle such that the voltage at bus 1 is real and positive, then all the voltages will be real and increasing along the transmission line. Since all voltages are real we will drop the complex notation for the rest of this chapter, meaning that we write V_i instead of \tilde{V}_i and \tilde{V}_i^* .

We wish to determine the probability of the rare event, that the magnitude of the voltage at any consumer falls below a certain threshold or that a stable solution does not even exist at all. We will show how this rare event relates to a system where the voltage at the final consumer is exactly at the threshold. First, we need the following lemma.

Lemma 5.2.1. *Let $V_1 > 0$ be given. If $V_{n+1} \leq 2V_1$, then*

$$0 \leq \frac{dV_k}{dV_1} \leq 1,$$

for all $k = 1, \dots, n + 1$.

Proof. Clearly, this is true for $k = 1$, since $\frac{dV_1}{dV_1} = 1$.

Suppose that there is some $1 \leq K \leq n$ such that $0 \leq \frac{dV_k}{dV_1} \leq 1$ for all $1 \leq k \leq K$, then

$$\begin{aligned} \frac{dV_{K+1}}{dV_1} &= 1 + \sum_{i=1}^K z'_{i,K+1} \frac{S_i^*}{V_i^2} \frac{dV_i}{dV_1} \\ &\geq 1 + \frac{1}{V_1} \sum_{i=1}^K z'_{i,K+1} \frac{S_i^*}{V_i} \frac{dV_i}{dV_1} \\ &\geq 1 + \frac{1}{V_1} \sum_{i=1}^K z'_{i,K+1} \frac{S_i^*}{V_i} \\ &= 1 - \frac{1}{V_1} (V_{K+1} - V_1) \\ &\geq 0, \end{aligned}$$

and

$$\frac{dV_{K+1}}{dV_1} = 1 + \sum_{i=1}^K z'_{i,K+1} \frac{S_i^*}{V_i^2} \frac{dV_i}{dV_1} \leq 1.$$

□

Thanks to this lemma it becomes much easier to find whether a particular rare event we are interested in occurs, namely if the voltage at any bus falls too far below the nominal voltage. Given $V_{n+1} = 1$, is there no solution for V_1 with $V_1 \geq 1 - \alpha$, where $0 \leq \alpha \leq 1/2$? This is because we find an equivalence with a system where not the voltage at bus $n + 1$, but the voltage at bus 1 is known instead. This allows for recursively determining the voltage at the reference bus via the recursive relations.

Theorem 5.2.1. *Let the impedances $z_{i,i+1}$ and the power injections $S_i, i = 1, \dots, n$ be fixed. Let $V_{n+1}(V_1)$ express the voltage at bus $n + 1$ as a function of the voltage at bus 1. The following equivalence holds, for $0 \leq \alpha \leq 1/2$, for the voltages at bus 1 and bus $n + 1$,*

$(\exists x \geq 1 - \alpha \text{ s.t. } V_1 = x \text{ and } V_{n+1}(V_1) = 1)$ if and only if $(V_1 = 1 - \alpha \text{ and } V_{n+1}(V_1) \leq 1)$.

Proof. \Rightarrow : $V_{n+1}(V_1)$ is through composition of continuous recursive functions itself a continuous function of V_1 , when $V_1 > 0$. At $V_1 = 1$ we have $V_{n+1}(1) \geq V_1 = 1$. So, if $V_{n+1}(1 - \alpha) \leq 1$, then we find by the intermediate value theorem that, there exists an $1 - \alpha \leq x \leq 1$ such that $V_{n+1}(x) = 1$.

\Rightarrow : If $V_{n+1}(1 - \alpha) > 1$, then the function V_{n+1} is by lemma 5.2.1 increasing or strictly larger than $2 \cdot (1 - \alpha) \geq 1$ for all $x \geq 1 - \alpha$. Thus there exists no $x \geq 1 - \alpha$ such that $V_{n+1}(x) = 1$. \square

5.3 Simulation results

We consider a system of 20 buses connected on a single line. We work in a per-unit system. The voltage at the reference bus is 1. The resistance between any two subsequent buses is .001. The power demand of all buses is purely real. The demand at each bus is modeled as a half-normal distribution with mean 1/4. The voltage at the last bus is not allowed to drop below .9.

In a simulation with 10^7 samples this lead to 294 cases where the voltage at the end bus was too low, a probability of $2.94 \cdot 10^{-5}$ and relative standard deviation of $\sqrt{1/(2.94 \cdot 10^{-5}) - 1} \approx 184$.

Chapter 6

A Continuous-Time Approach to the 2-Bus System

In this chapter, we will study the most basic network consisting of a single load-bus and a single generator: a 2-bus system. Note that a load-bus could be a substation, that in turn is connected to many different consumers. This situation may thus still be used to model more complex situations than that of a single consumer. Up until now, we have considered the power demand at a load-bus at a certain point in time and investigated whether that would lead to an undesirable situation in the system. With this approach, we can conclude that, if no undesirable situation exists at a certain time T , then there was no undesirable situation at any moment before time T . Though this is a reasonable rule of thumb, it is not necessarily true. We avoid this fallacious reasoning by considering the power demand as a continuous-time stochastic process.

We choose to model the demand of power by a Wiener process, because it is a well tractable continuous-time process, that has Gaussian increments. Therefore, this model of the power injections in the 2-bus system concurs with our earlier model of Gaussian power injections in Chapter 4. We will make a comparison between the continuous-time and discrete-time model for the 2-bus system.

6.1 Wiener processes and the Heat equation

In a general $d+1$ -bus AC-powerflow system, where bus $d+1$ is the reference bus, we model the vector of real and reactive power injections $(P_1, Q_1, P_2, Q_2, \dots, P_d, Q_d)^\top$ by an affine transformation of a \mathbb{R}^{2d} standard Wiener process.

Definition 6.1.1. [30] A **filtration** on a measurable space (Ω, \mathcal{F}) is a collection $\{\mathcal{F}_t\}_{t \geq 0}$ of σ -algebras $\mathcal{F}_t \subset \mathcal{F}$ such that

$$0 \leq s < t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t.$$

A stochastic process $(X_t)_{t \geq 0}$ on a probability space (Ω, \mathcal{F}, P) is called **adapted** if for every $t \geq 0$ the random variable X_t is \mathcal{F}_t -measurable.

Definition 6.1.2. [30] Let (Ω, \mathcal{F}, P) be a probability space with a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ and an \mathbb{R}^d -valued adapted stochastic process $W = (W_t)_{t \geq 0}$. Then W is called an \mathbb{R}^d -valued **standard Wiener process** with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ if the following conditions hold:

1. $W_0 = 0$ almost surely.
2. (independence of increments) $W_t - W_s$ is independent of \mathcal{F}_s for all $s \leq t$.
3. (stationarity of increments) $W_t - W_s \stackrel{Dist.}{=} \mathcal{N}_d(0, (t-s)I_d)$ in distribution.
4. W almost surely has continuous sample paths.

An affine transformation of a standard Wiener process is referred to as a Wiener process. If $W = (W_t)_{t \geq 0}$ is an \mathbb{R}^d -valued standard Wiener process, then $W' = \sqrt{\Sigma}W + \mu$, where Σ is a $d \times d$ positive definite matrix and $\mu \in \mathbb{R}^d$, is a Wiener process. Let $\{\mathcal{F}_t\}_{t \geq 0}$ be its filtration. The process W' satisfies the following conditions:

1. $W'_0 = \mu$ almost surely.
2. (independence of increments) $W'_t - W'_s$ is independent of \mathcal{F}_s for all $s \leq t$.
3. (stationarity of increments) $W'_t - W'_s \stackrel{Dist.}{=} \mathcal{N}_d(0, (t-s)\Sigma)$ in distribution.
4. W' almost surely has continuous sample paths.

The third property (stationarity of increments) holds, because

$$\begin{aligned} \text{Var}(W'_t - W'_s) &= \text{Var}(\sqrt{\Sigma}(W_t - W_s)) = \sqrt{\Sigma}\text{Var}(W_t - W_s)\sqrt{\Sigma}^\top \\ &= \sqrt{\Sigma}(t-s)I_d\sqrt{\Sigma}^\top = (t-s)\Sigma. \end{aligned}$$

Based on restrictions in the network of required voltages at the buses and maximal power flows through the transmission lines, we can define a domain \mathcal{D} of feasible power injections. Let the Wiener process $W = (W_t)_{t \geq 0}$ describe the power injections, then the network is considered to be in an infeasible state at a time t if and only if $W_t \notin \mathcal{D}$. We want to know the probability that the network reaches an infeasible state within a given time $t > 0$.

We make the assumptions that \mathcal{D} is open and that $W_0 \in \mathcal{D}$. The first time we exit the region \mathcal{D} is given by the first exit time,

$$\tau = \inf\{t > 0 : W_t \notin \mathcal{D}\}. \quad (6.1)$$

Definition 6.1.3. Let (Ω, \mathcal{F}, P) be a probability space with a filtration $\{\mathcal{F}_t\}_{t \geq 0}$. A random variable $\tau : \Omega \rightarrow [0, \infty]$ is called a **stopping time** with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ if,

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t, \quad \text{for all } t \geq 0.$$

In loose terms, this definition means, that τ is a stopping time when it only depends on the past of a stochastic process and not on its future.

Lemma 6.1.1. For a stochastic process X_t with right-continuous paths and an open set \mathcal{D} , the first exit time $\tau = \inf\{t > 0 : X_t \notin \mathcal{D}\}$ is a stopping time.

Proof. A proof can be found in [31]. □

Next, we show that the first exit time of a Wiener process from a bounded open set is almost surely finite.

Lemma 6.1.2. *Let $W = (W_t)_{t \geq 0}$ be an \mathbb{R}^d -valued Wiener process on the probability space $(\Omega, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$, whose increments $W_t - W_s$, for $t > s$, are distributed as $\mathcal{N}_d(0, (t - s)\Sigma)$, where Σ is a $d \times d$ positive definite matrix. Let \mathcal{D} be a bounded open set. The first exit time $\tau = \inf\{t > 0 : W_t \notin \mathcal{D}\}$ is almost surely finite.*

Proof. The events $\{\omega \in \Omega : \tau \leq t\}_{t \geq 0}$ are increasing, thus $\mathbb{P}(\lim_{t \rightarrow \infty} \{\omega \in \Omega : \tau \leq t\}) = \lim_{t \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : \tau \leq t\})$. Furthermore, for all $t \geq 0$, we have $\{\omega \in \Omega : W_t \notin \mathcal{D}\} \subset \{\omega \in \Omega : \tau \leq t\}$. Thus, we conclude,

$$\begin{aligned} \mathbb{P}(\lim_{t \rightarrow \infty} \{\omega \in \Omega : \tau \leq t\}) &= \lim_{t \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : \tau \leq t\}) \\ &\geq \lim_{t \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : W_t \notin \mathcal{D}\}) \\ &= \lim_{t \rightarrow \infty} \left[1 - \int_{\mathcal{D}} \frac{1}{\sqrt{|2\pi t \Sigma|}} \exp\left(-\frac{1}{2t} D_{\Sigma}(\mathbf{x}, W_0)^2\right) d\mathbf{x} \right] \\ &\geq 1 - \lim_{t \rightarrow \infty} \int_{\mathcal{D}} \frac{1}{\sqrt{|2\pi t \Sigma|}} d\mathbf{x} \\ &= 1. \end{aligned}$$

□

Next, we introduce a theorem that directly links the transition probability on the stopped stochastic process,

$$W_{t \wedge \tau} = \begin{cases} W_t, & \text{if } t \leq \tau, \\ W_{\tau}, & \text{if } t > \tau, \end{cases}$$

to the solution of a partial differential equation with boundary values and an initial condition. This is more commonly referred to as a Cauchy problem. This translates the stochastic process into a deterministic problem.

Theorem 6.1.1. *Let $W = (W_t)_{t \geq 0}$ be an \mathbb{R}^d -valued standard Wiener process. Let $\mathcal{D} \subset \mathbb{R}^d$ be a bounded open set, and let $\mathbf{x} \in \mathcal{D}$. The solution to the Cauchy problem,*

$$\partial_t u(\mathbf{y}, t) = \frac{1}{2} \Delta_{\mathbf{y}} u(\mathbf{y}, t), \quad \text{for } t > 0, \mathbf{y} \in \mathcal{D},$$

with initial condition,

$$\lim_{t \downarrow 0} u(\mathbf{y}, t) = \delta(\mathbf{x} - \mathbf{y}), \quad \mathbf{y} \in \mathcal{D},$$

and Dirichlet boundary conditions,

$$u(\mathbf{y}, t) = 0, \quad \text{for } \mathbf{y} \in \partial \mathcal{D},$$

is equal to the transition density of hitting a point $\mathbf{y} \in \mathcal{D}$ at time t for the Wiener process $W_t + \mathbf{x}$, i.e.

$$p^{\mathcal{D}}(\mathbf{x}, \mathbf{y}, t) = u(\mathbf{y}, t).$$

Proof. The first exit time $\tau^{\mathbf{x}} = \inf\{t > 0 : B_t + \mathbf{x} \notin \mathcal{D}\}$ from \mathcal{D} is an almost surely finite stopping time by Lemma 6.1.2. Therefore, the stopped stochastic process,

$$X_t = \begin{cases} W_t + \mathbf{x}, & \text{if } t \leq \tau^{\mathbf{x}}, \\ W_{\tau^{\mathbf{x}}} + \mathbf{x}, & \text{if } t > \tau^{\mathbf{x}}, \end{cases}$$

possesses the strong Markov property [32], and consequently it is a Feller-process. From the fact that the generator of the standard Wiener process is a half times the Laplacian and by [33], the generator A of this process is given, for suitable functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$, by

$$Af(\mathbf{y}) = \begin{cases} \frac{1}{2}\Delta_{\mathbf{y}}f(\mathbf{y}), & \mathbf{y} \in \mathcal{D}, \\ 0, & \mathbf{y} \in \partial\mathcal{D}. \end{cases}$$

The Kolmogorov forward equation now yields that $p^{\mathcal{D}}(\mathbf{x}, \mathbf{y}, t) = u(\mathbf{y}, t)$, for $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, where $u(\mathbf{y}, t)$ satisfies,

$$\begin{aligned} \partial_t u(\mathbf{y}, t) &= \begin{cases} \frac{1}{2}\Delta_{\mathbf{y}}u(\mathbf{y}, t), & t > 0, \mathbf{y} \in \mathcal{D}, \\ 0, & t > 0, \mathbf{y} \in \partial\mathcal{D}. \end{cases} \\ \lim_{t \downarrow 0} u(\mathbf{y}, t) &= \delta(\mathbf{x} - \mathbf{y}). \end{aligned}$$

This is equivalent to the posed Cauchy problem. \square

The described partial differential equation is better known as the Heat Equation with diffusion constant 1. We note that the transition probability on \mathcal{D} is *sub-stochastic* for $t > 0$ i.e. $p^{\mathcal{D}}(\mathbf{x}, \mathcal{D}, t) < 1$. This is due to the killing of some of the paths on the boundary $\partial\mathcal{D}$ within any given time $t > 0$. In fact, we have,

$$\mathbb{P}(\tau \leq t) = 1 - p^{\mathcal{D}}(\mathbf{x}, \mathcal{D}, t).$$

Similarly to the model where the power injections were normally distributed, we wish to study the asymptotic behaviour of $\mathbb{P}(\tau \leq t)$ as t tends to zero.

Large deviations theory and Cramér's theorem 3.4.1 provided us with the result, that for an open set \mathcal{D} and random variables $\mathbf{Y}_n \sim \mathcal{N}_d(\mu, \Sigma/n)$, $n \in \mathbb{N}$, Equation 4.3 holds,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathbf{Y}_n \notin \mathcal{D}) = -\frac{1}{2} \inf_{\mathbf{x} \in \mathcal{D}^c} (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu).$$

In large deviations theory, Schilder's theorem [18, 34] provides a similar result for $\mathbb{P}(\tau \leq t)$.

Theorem 6.1.2 (Schilder). *Let $W = (W_t)_{t \geq 0}$ be an \mathbb{R}^d -valued standard Wiener process. Let, for $\epsilon > 0$, ν_ϵ be the probability measure induced by $W_{\epsilon t}$, $t \in [0, T]$, on the space $C_0([0, T], \mathbb{R}^d)$ of continuous functions $\phi : [0, T] \rightarrow \mathbb{R}^d$ such that $\phi(0) = 0$, equipped with the supremum norm topology.*

The family of probability measures $\{\nu_{\epsilon^{-1}}\}_{\epsilon^{-1} > 0}$ satisfies, in $C_0([0, T], \mathbb{R}^d)$, the large deviations principle with good rate function,

$$I_W(\phi) = \begin{cases} \frac{1}{2} \int_0^T |\dot{\phi}(t)|^2 dt, & \text{if } \phi \text{ is absolutely continuous,} \\ \infty, & \text{otherwise.} \end{cases}$$

Proof. A proof can be found in [18]. \square

As a corollary of Schilder's theorem, we see that for an \mathbb{R}^d -valued standard Wiener process

$W = (W_t)_{t \geq 0}$, a bounded regular open set \mathcal{D} , and some time $T > 0$,

$$\begin{aligned}
\lim_{\epsilon \downarrow 0} \epsilon \log \mathbb{P}(\tau \leq \epsilon T) &= \lim_{\epsilon \downarrow 0} \epsilon \log \mathbb{P}(\exists t \in [0, T] \text{ such that } W_{\epsilon t} \in \mathcal{D}^C) \\
&= -\inf \{I_W(\phi) : \phi \in C_0([0, T], \mathbb{R}^d), \exists t \in [0, T] \text{ such that } \phi(t) \in \mathcal{D}^C\} \\
&= -I_W \left(t \rightarrow \frac{t}{T} \cdot \arg \inf_{\mathbf{x} \in \mathcal{D}^C} \{\|\mathbf{x}\|_2\} \right) \\
&= -\frac{1}{2} \int_0^T \left\| \frac{1}{T} \arg \inf_{\mathbf{x} \in \mathcal{D}^C} \{\|\mathbf{x}\|_2\} \right\|_2^2 dt \\
&= -\frac{1}{2T} \inf_{\mathbf{x} \in \mathcal{D}^C} \|\mathbf{x}\|_2^2.
\end{aligned}$$

Setting $\epsilon = t$ and $T = 1$ yields

$$\lim_{t \downarrow 0} t \log \mathbb{P}(\tau \leq t) = -\frac{1}{2} \inf_{\mathbf{x} \in \mathcal{D}^C} \|\mathbf{x}\|_2^2.$$

More generally, we find for the Wiener process $W' = \sqrt{\Sigma}W + \mu$, where Σ is a $d \times d$ positive definite matrix and $\mu \in \mathbb{R}^d$,

$$\lim_{t \downarrow 0} t \log \mathbb{P}(\tau \leq t) = -\frac{1}{2} \inf_{\mathbf{x} \in \mathcal{D}^C} (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) = -\frac{1}{2} D_\Sigma(\mu, \mathcal{D}^C)^2, \quad (6.2)$$

since, for $t \geq 0$, $(W_t \in \mathcal{D}^C) \Leftrightarrow (W'_t \in \sqrt{\Sigma}^{-1} \mathcal{D}^C - \mu)$. Thus, we have the remarkable result, that the rate function of the probability of having exited \mathcal{D} at some time before t is equal to the rate function of the probability of being outside \mathcal{D} at time t ,

$$\lim_{t \downarrow 0} t \log \mathbb{P}(\inf\{s > 0 : W'_s \in \mathcal{D}^C\} \leq t) = \lim_{t \downarrow 0} t \log \mathbb{P}(W' \in \mathcal{D}^C).$$

6.2 Wiener process exiting an interval

We first study the Wiener-process in a single dimension. Consider a 2-bus system of a reference bus, that operates at a given constant voltage, and a load bus, where the reactive power demand is a function of the real power demand. Through composition of functions, the voltage at the load bus is a function of the real power demand. An allowable interval of real power demand can now be derived from bounds on the allowed range of operating voltage at the load bus.

Similarly, an interval for the real power demand may be based on restrictions on the power losses in the transmission line between the two buses, see [35].

Let us consider a standard Wiener process W_t to model the power demand and an allowable interval $[-a, b]$, where $0 < a, b$.

We are interested in the distribution of the stopping time,

$$\tau = \inf\{t > 0 : W_t \notin [-a, b]\}.$$

We introduce auxiliary stopping times,

$$\begin{aligned}
\tau_a &= \inf\{t > 0 : W_t \leq -a\}, \\
\tau_b &= \inf\{t > 0 : W_t \geq b\}.
\end{aligned}$$

The distribution of this stopping times is relatively easy to find compared to the distribution of τ . By Theorem 6.1.1, the transition density of a stopped standard Wiener process,

$$W_{t \wedge \tau_b} = \begin{cases} W_t, & \text{if } t \leq \tau_b, \\ b, & \text{if } t > \tau_b, \end{cases}$$

is given by the solution to the Cauchy problem,

$$\begin{aligned} \frac{\partial u}{\partial t} &= u_{xx}, & t > 0, x \leq b, \\ u(b, t) &= 0, & t > 0, \\ \lim_{t \downarrow 0} u(x, t) &= \delta(x), & x \leq b. \end{aligned}$$

It can be verified, that this equation is solved by the function

$$u(x, t) = \frac{1}{\sqrt{2\pi t}} \left(\exp\left(-\frac{x^2}{2t}\right) - \exp\left(-\frac{(x-2b)^2}{2t}\right) \right), \quad x < b, t > 0.$$

Consequently, we find,

$$\begin{aligned} \mathbb{P}(\tau_b \leq t) &= 1 - \int_{-\infty}^b \frac{1}{\sqrt{2\pi t}} \left(\exp\left(-\frac{x^2}{2t}\right) - \exp\left(-\frac{(x-2b)^2}{2t}\right) \right) dx \\ &= 1 - \Phi\left(\frac{b}{\sqrt{t}}\right) + \Phi\left(\frac{-b}{\sqrt{t}}\right) \\ &= 2 - 2\Phi\left(\frac{b}{\sqrt{t}}\right), \end{aligned} \tag{6.3}$$

and,

$$\frac{\mathbb{P}(\tau_b \in dt)}{dt} = \frac{b}{\sqrt{2\pi t^3}} \exp\left(-\frac{b^2}{2t}\right). \tag{6.4}$$

The distribution of τ_b can alternatively be derived by considering the reflection principle. By virtue of the strong Markov property, the process

$$W_t^b = \begin{cases} W_t, & \text{if } t \leq \tau_b, \\ b - W_t, & \text{if } t > \tau_b, \end{cases}$$

we obtain by reflecting the standard Wiener process $W = (W_t)_{t \geq 0}$ at the random time τ_b , is again a standard Wiener process, whose paths are equally likely as those of the original process W . Therefore, we find,

$$\mathbb{P}(W_t \geq b | t \geq \tau_b) = \mathbb{P}(W_t^b \geq b | t \geq \tau_b) = \mathbb{P}(W_t \leq b | t \geq \tau_b),$$

and,

$$\mathbb{P}(\tau_b \leq t) = \frac{\mathbb{P}(\tau_b \leq t, W_t \geq b)}{\mathbb{P}(W_t \geq b | \tau_b \leq t)} = \frac{\mathbb{P}(W_t \geq b)}{1/2} = 2 - 2\Phi\left(\frac{b}{\sqrt{t}}\right).$$

We return to the problem of determining the distribution of τ . By Theorem 6.1.1, the transition density on $(-a, b)$ of the stopped standard Wiener process,

$$W_{t \wedge \tau} = \begin{cases} W_t, & \text{if } t \leq \tau, \\ W_\tau, & \text{if } t > \tau, \end{cases}$$

is given by the solution to the Cauchy problem,

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2}u_{xx}, & t > 0, \quad -a \leq x \leq b, \\ u(-a, t) = u(b, t) &= 0, & t > 0, \\ \lim_{t \downarrow 0} u(x, t) &= \delta(x), & -a \leq x \leq b. \end{aligned}$$

This Cauchy problem is often solved using Fourier analysis, e.g. see [36],

$$u(x, t) = \frac{2}{a+b} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi(x+a)}{a+b}\right) \sin\left(\frac{n\pi a}{a+b}\right) \exp\left(-\frac{t}{2} \left(\frac{n\pi}{a+b}\right)^2\right).$$

This expression does not prove particularly helpful when considering rare events, since for small $t > 0$,

$$\begin{aligned} \mathbb{P}(\tau \leq t) &= 1 - \int_{-a}^b u(x, t) dx \\ &= 1 + \sum_{n=1}^{\infty} \frac{2}{n\pi} (\cos(n\pi) - 1) \sin\left(\frac{n\pi a}{a+b}\right) \exp\left(-\frac{t}{2} \left(\frac{n\pi}{a+b}\right)^2\right) \\ &= 1 - \sum_{n=1}^{\infty} \frac{4}{(2n-1)\pi} \sin\left(\frac{(2n-1)\pi a}{a+b}\right) \exp\left(-\frac{t}{2} \left(\frac{(2n-1)\pi}{a+b}\right)^2\right), \end{aligned} \tag{6.5}$$

is rather difficult to evaluate. As t decreases, the sum converges more slowly. Using this expression, it seems impossible to find the asymptotic behaviour of $\mathbb{P}(\tau \leq t)$ as t tends to zero, $\lim_{t \downarrow 0} t \log \mathbb{P}(\tau \leq t)$.

By Equation 6.2, we know that $\lim_{t \downarrow 0} t \log \mathbb{P}(\tau \leq t) = -\frac{1}{2} \min\{a^2, b^2\}$. We opt for a different method to obtain an expression for $\mathbb{P}(\tau \leq t)$, that will show this more clearly.

Applying the inclusion-exclusion principle, we find,

$$\mathbb{P}(\tau \leq t) = \mathbb{P}((\tau_a \wedge \tau_b) \leq t) = \mathbb{P}(\tau_a \leq t) + \mathbb{P}(\tau_b \leq t) - \mathbb{P}((\tau_a \vee \tau_b) \leq t).$$

where ‘ \wedge ’ denotes the minimum of two elements and ‘ \vee ’ denotes the maximum. The term $\mathbb{P}((\tau_a \vee \tau_b) \leq t)$ corresponds to the probability of crossing the boundary of the interval within time t .

We recursively define more auxiliary stopping times,

$$\begin{aligned} \tau_a^{(0)} &= \inf\{t > 0 : W_t \leq -a\}, \\ \tau_b^{(0)} &= \inf\{t > 0 : W_t \geq b\}, \\ \tau_a^{(n+1)} &= \begin{cases} \inf\{t > \tau_a^{(n)} : W_t \geq b\}, & \text{if } n \in \mathbb{N}_0 \text{ is even,} \\ \inf\{t > \tau_a^{(n)} : W_t \leq -a\}, & \text{if } n \in \mathbb{N}_0 \text{ is odd,} \end{cases} \\ \tau_b^{(n+1)} &= \begin{cases} \inf\{t > \tau_b^{(n)} : W_t \leq -a\}, & \text{if } n \in \mathbb{N}_0 \text{ is even,} \\ \inf\{t > \tau_b^{(n)} : W_t \geq b\}, & \text{if } n \in \mathbb{N}_0 \text{ is odd.} \end{cases} \end{aligned}$$

Thanks to the repeated application of the strong Markov property, these are all indeed stopping times. The stopping time $\tau_a^{(n)}$ is the first time, that alternatingly the upper and

lower bounds are crossed n times after the lower bound has been crossed for the first time. A useful equivalence relation for these stopping times is,

$$\left((\tau_a^{(n)} \vee \tau_b^{(n)}) \leq t \right) \Leftrightarrow \left((\tau_a^{(n+1)} \wedge \tau_b^{(n+1)}) \leq t \right), \quad n \in \mathbb{N}_0.$$

The implication to the left follows from the fact that $\tau_a^{(n)}, \tau_b^{(n)} \leq (\tau_a^{(n+1)} \wedge \tau_b^{(n+1)})$.

For the implication to the right, we consider two separate cases. If $(\tau_a^{(n)} \wedge \tau_b^{(n)}) \leq t$ and $\tau_a^{(0)} \leq \tau_b^{(0)}$, then we have $\tau_a^{(n+1)} = \tau_b^{(n)} \leq t$. Alternatively, if $(\tau_a^{(n)} \wedge \tau_b^{(n)}) \leq t$ and $\tau_b^{(0)} \leq \tau_a^{(0)}$, then we have $\tau_b^{(n+1)} = \tau_a^{(n)} \leq t$.

By repeatedly applying both this equivalence and the inclusion-exclusion principle,

$$\mathbb{P}((\tau_a^{(n)} \wedge \tau_b^{(n)}) \leq t) = \mathbb{P}(\tau_a^{(n)} \leq t) + \mathbb{P}(\tau_b^{(n)} \leq t) - \mathbb{P}((\tau_a^{(n)} \vee \tau_b^{(n)}) \leq t), \quad n \in \mathbb{N}_0,$$

we find,

$$\mathbb{P}((\tau_a \wedge \tau_b) \leq t) = \sum_{n \in \mathbb{N}_0} (-1)^n \left[\mathbb{P}(\tau_a^{(n)} \leq t) + \mathbb{P}(\tau_b^{(n)} \leq t) \right]. \quad (6.6)$$

We find the probabilities $\mathbb{P}(\tau_a^{(n)} \leq t)$ and $\mathbb{P}(\tau_b^{(n)} \leq t)$ by relating $\tau_a^{(n)}$ and $\tau_b^{(n)}$ to some other stopping times,

$$\begin{aligned} \tilde{\tau}_a^{(n)} &= \inf\{t > 0 : W_t \geq a + n(a + b)\}, & n \in \mathbb{N}_0, \\ \tilde{\tau}_b^{(n)} &= \inf\{t > 0 : W_t \geq b + n(a + b)\}, & n \in \mathbb{N}_0. \end{aligned}$$

It can be shown by induction on $n \in \mathbb{N}_0$, that

$$\begin{aligned} \mathbb{P}(\tilde{\tau}_a^{(n)} \leq t) &= \mathbb{P}(\tau_a^{(n)} \leq t), & n \in \mathbb{N}_0, \\ \mathbb{P}(\tilde{\tau}_b^{(n)} \leq t) &= \mathbb{P}(\tau_b^{(n)} \leq t), & n \in \mathbb{N}_0. \end{aligned}$$

For $n = 0$, this is trivially true, since $\tau_a^{(0)} = \tilde{\tau}_a^{(0)}$ and $\tau_b^{(0)} = \tilde{\tau}_b^{(0)}$.

Suppose that for some $N \in \mathbb{N}_0$ the induction hypothesis holds for all $n \leq N$, then for N even,

$$\begin{aligned} \mathbb{P}(\tau_a^{(N+1)} \leq t) &= \mathbb{P}(\tau_a^{(N+1)}, \tau_a^{(N)} \leq t) \\ &= \int_{u=0}^t \mathbb{P}(\inf\{s > \tau_a^{(N)} : W_s \geq b\} \leq t | \tau_a^{(N)} \in du) \mathbb{P}(\tau_a^{(N)} \in du) \\ &= \int_{u=0}^t \mathbb{P}(\inf\{s > \tilde{\tau}_a^{(N)} : W_s \geq b + (N+1)(a+b)\} \leq t | \tilde{\tau}_a^{(N)} \in du) \\ &\quad \times \mathbb{P}(\tilde{\tau}_a^{(N)} \in du) \\ &= \int_{u=0}^t \mathbb{P}(\tilde{\tau}_a^{(N+1)} \leq t | \tilde{\tau}_a^{(N)} \in du) \mathbb{P}(\tilde{\tau}_a^{(N)} \in du) \\ &= \mathbb{P}(\tilde{\tau}_a^{(N+1)}, \tilde{\tau}_a^{(N)} \leq t) \\ &= \mathbb{P}(\tilde{\tau}_a^{(N+1)} \leq t). \end{aligned}$$

Similar relations hold when N is odd and for $\tau_b^{(N)}$.

Combining Equation 6.6 with Equations 6.3 and 6.4, we now find,

$$\mathbb{P}(\tau \leq t) = 2 \sum_{n \in \mathbb{N}_0} (-1)^n \left[\Phi^c \left(\frac{a + n(a+b)}{\sqrt{t}} \right) + \Phi^c \left(\frac{b + n(a+b)}{\sqrt{t}} \right) \right], \quad (6.7)$$

and,

$$\frac{\mathbb{P}((\tau_a \wedge \tau_b) \in dt)}{dt} = \frac{2}{\sqrt{2\pi t^3}} \sum_{n \in \mathbb{N}_0} (-1)^n \left[(a + n(a+b)) \exp\left(-\frac{(a+n(a+b))^2}{2t}\right) + (b + n(a+b)) \exp\left(-\frac{(b+n(a+b))^2}{2t}\right) \right].$$

Equation 6.7 is significantly easier to evaluate as t tends to zero than Equation 6.5. Using the bounds provided in Lemma 4.1.2, we find the same result as in Equation 6.2,

$$\lim_{t \downarrow 0} t \log \mathbb{P}(\tau \leq t) = -\frac{1}{2} \min\{a^2, b^2\}.$$

6.3 Wiener process exiting an elliptical domain

In the previous section, we studied the distribution of the exit time of a Wiener process from an interval. This corresponds to a 2-bus system in which the reactive power demand is fully dependent on the real power demand. In this section we will study a 2-bus system where the reactive power demand is not fully dependent on the real power demand.

We consider the 2-bus AC-powerflow system. We have a reference bus, this will be a generator, at which the voltage magnitude is controlled at a constant value of $|V_{\text{gen}}|$. We impose restrictions on the real and reactive power loads, P and Q , at the load-bus based on the voltage magnitude $|V_{\text{load}}|$ at this bus. When power is required at the load-bus, the power injection is negative. We consider the power load tuple (P, Q) to follow an affine transformation of a 2-dimensional standard Wiener process.

The buses are connected by a transmission line with conductance G , susceptance B , and admittance Y . The magnitude of the admittance is $|Y| = \sqrt{G^2 + B^2}$. We assume, that there are no shunt impedances.

There are three possible undesirable situations related to the voltage magnitude: the system has no stable solution, $|V_{\text{load}}|$ is too low, or $|V_{\text{load}}|$ is too high. We denote the lower and upper bounds on $|V_{\text{load}}|$ by $|V_{\text{min}}|$ and $|V_{\text{max}}|$ respectively. We will focus on the violation of the lower bound $|V_{\text{min}}|$, since it is the most probable one under typical operating conditions.

By the complex power formula, Equation 2.9, the injected current at the load bus is $\tilde{I}_{\text{load}} = S_{\text{load}}^* / \tilde{V}_{\text{load}}^*$, where \tilde{V}_{load} and S are the complex voltage and power injection at the load-bus respectively. Combining this with Equation 2.12, we find the relation,

$$\frac{S_{\text{load}}^*}{\tilde{V}_{\text{load}}^*} = Y \left(\tilde{V}_{\text{load}} - \tilde{V}_{\text{gen}} \right),$$

or equivalently,

$$S_{\text{load}}^* - Y|V_{\text{load}}|^2 = -Y\tilde{V}_{\text{gen}}\tilde{V}_{\text{load}}^*.$$

Now we see, that the steady state solution for the voltage at the generator is related to the other given variables via the implicit relation,

$$(P - |V_{\text{load}}|^2 G)^2 + (Q + |V_{\text{load}}|^2 B)^2 = |Y|^2 |V_{\text{gen}}|^2 |V_{\text{load}}|^2.$$

If there is no value for $|V_{\text{load}}|$ that solves this equation, then there is no steady-state solution. This occurs when,

$$\frac{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2}{4|Y|^2} < P^2 + Q^2.$$

For each pair of values of real and reactive power demand (P, Q) at the load-bus with,

$$\frac{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2}{4|Y|^2} > P^2 + Q^2,$$

there are two distinct solutions. We discern between a *high voltage branch*,

$$\begin{aligned} |V_{\text{load}}| = & \frac{1}{2}|V_{\text{gen}}|^2 + \frac{PG - QB}{|Y|^2} \\ & + \frac{\sqrt{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2 - 4|Y|^2(P^2 + Q^2)}}{2|Y|^2}, \end{aligned} \quad (6.8)$$

and a *low voltage branch*,

$$\begin{aligned} |V_{\text{load}}| = & \frac{1}{2}|V_{\text{gen}}|^2 + \frac{PG - QB}{|Y|^2} \\ & - \frac{\sqrt{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2 - 4|Y|^2(P^2 + Q^2)}}{2|Y|^2}. \end{aligned} \quad (6.9)$$

The power losses in the transmission line scale inversely with the difference of the squares of the voltage magnitudes at the buses. Therefore, it is desirable to be operating within the high voltage branch.

At this point, it is important to note that, when the real and reactive power change continuously over time, it is impossible to change from a high voltage branch solution to a low voltage branch solution without having ever had values for P and Q such that there is only one solution. These pairs of (P, Q) lie on the parabola, for which

$$\frac{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2}{4|Y|^2} = P^2 + Q^2.$$

Now the open domain, within which the safe and desirable operating conditions are met, is inscribed within the curves given by the following three implicit curves,

$$\begin{aligned} |V_{\text{min}}|^2 = & \frac{1}{2}|V_{\text{gen}}|^2 + \frac{PG - QB}{|Y|^2} \\ & + \frac{\sqrt{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2 - 4|Y|^2_1(P^2 + Q^2)}}{2|Y|^2}, \end{aligned} \quad (6.10)$$

$$\begin{aligned} |V_{\text{max}}|^2 = & \frac{1}{2}|V_{\text{gen}}|^2 + \frac{PG - QB}{|Y|^2} \\ & - \frac{\sqrt{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2 - 4|Y|^2_1(P^2 + Q^2)}}{2|Y|^2}, \end{aligned} \quad (6.11)$$

$$P^2 + Q^2 = \frac{(2PG - 2QB + |V_{\text{gen}}|^2|Y|^2)^2}{4|Y|^2}. \quad (6.12)$$

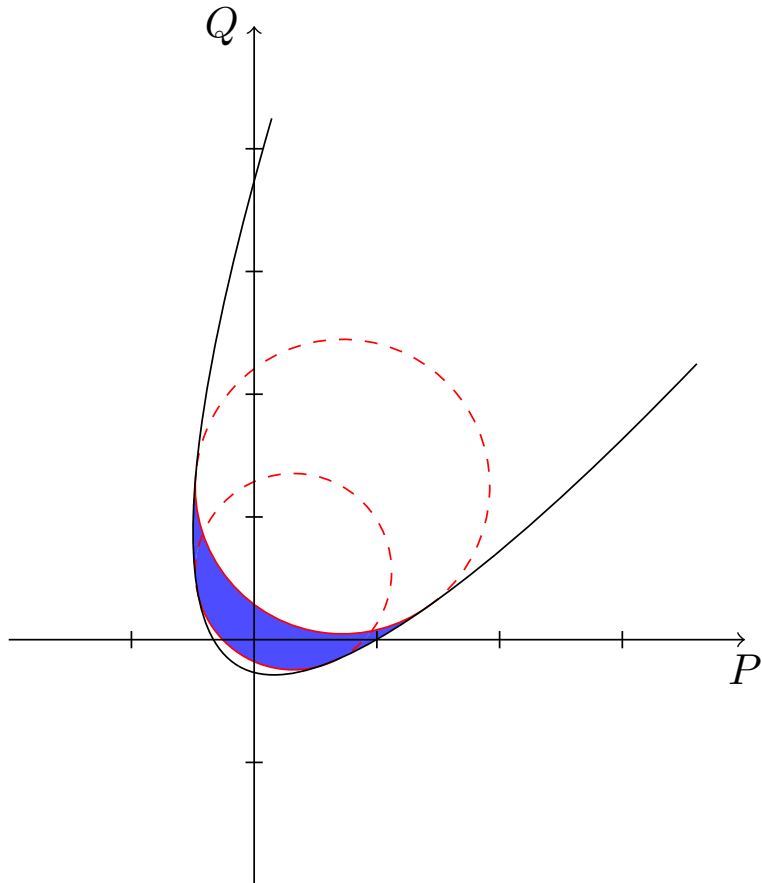


Figure 6.1: A graphical representation of the feasible region in blue. The small circle corresponds to all solutions for $|V_{\min}|$, Equation 6.10, and the large circle corresponds to all solutions for $|V_{\max}|$, Equation 6.11. The parabola, Equation 6.12, is the set of points with exactly one solution. The solid red lines represent the high-voltage branches, Equation 6.8, and the dashed red lines represent the low voltage branches, Equation 6.9.

A graphical representation of this region is given in Figure 6.1.

As mentioned earlier, we will mainly focus on violation of the lower bound on the voltage at the load-bus. The region of power injections at the load-bus, for which this bound is not violated, is

$$\mathcal{D} := \{(P, Q)^\top \in \mathbb{R}^2 : (P - G|V_{\min}|^2)^2 + (Q + B|V_{\min}|^2)^2 < |Y|^2|V_{\text{gen}}|^2|V_{\min}|^2\}. \quad (6.13)$$

This corresponds to the region enclosed by the small circle in Figure 6.1 with boundary equation 6.10.

We wish to know for how long we will stay within this region given the following model for the real and reactive power injection at the load-bus. Let Σ be a positive definite 2×2 matrix and let $W = (W_t)_{t \geq 0}$ be an \mathbb{R}^2 -valued standard Wiener process, then the real and reactive power load at the load-bus at time t are given by $X_t = (P_t, Q_t)^\top$, where

$$X_t = \sqrt{\Sigma}W_t + (P_0, Q_0)^\top,$$

and $(P_0, Q_0)^\top$ is the initial state. Now the first time we exit the region \mathcal{D} is given by the first exit time, $\tau = \inf\{t > 0 : X_t \notin \mathcal{D}\}$. This first exit time corresponds to the first time the network reaches an undesirable state.

Our goal is to find the distribution of τ , such that we can find the probability of violating the lower bound on the operating voltage at the load-bus $\mathbb{P}(\tau \leq T)$ within any given time T .

We can write the covariance-matrix Σ as $\Sigma = U^\top D U$, with U a unitary matrix and $D = \text{diag}(\lambda_1, \lambda_2)$ a diagonal matrix with $0 < \lambda_1 \leq \lambda_2$. When $\lambda_1 = \lambda_2$ we are just studying the exit problem from a sphere, which is less general and already well described in [37]. So, we assume $\lambda_1 < \lambda_2$. We define the affine operation,

$$A : \mathbb{R}^2 \rightarrow \mathbb{R}^2 : (P, Q)^\top \rightarrow \sqrt{D}^{-1}U(P - G|V_{\min}|^2, Q + B|V_{\min}|^2)^\top,$$

and apply it to the Wiener process $X = \sqrt{\Sigma}W + (P_0, Q_0)^\top$. We find,

$$A(X) = UW + \sqrt{D}^{-1}U(P_0 - G|V_{\min}|^2, Q_0 + B|V_{\min}|^2)^\top.$$

Since U is a unitary matrix, the process $A(X)$ is equal in distribution to $Y := W + \mathbf{x}$, where $\mathbf{x} = \sqrt{D}^{-1}U(P_0 - G|V_{\min}|^2, Q_0 + B|V_{\min}|^2)^\top$.

Next, we alternatively to τ define τ' ,

$$\tau' := \inf\{t > 0 : Y_t \notin \mathcal{D}'\},$$

where $\mathcal{D}' = A(\mathcal{D})$. Since the process $A(X)$ is equal in distribution to Y , the distributions of τ and τ' are also equal.

The domain \mathcal{D}' describes an ellipse with foci $(-K, 0)$ and $(K, 0)$, where the focal distance is $K = \sqrt{\lambda_1^{-1} - \lambda_2^{-1}}|Y||V_{\text{gen}}||V_{\text{load}}|$. Therefore, we do decide to introduce the elliptical coordinate system (ξ, η) . For $\xi \geq 0$ and $0 \leq \eta < 2\pi$, the relation to the Cartesian coordinate system is given by,

$$\begin{aligned} x_1 &= K \cosh \xi \cos \eta, \\ x_2 &= K \sinh \xi \sin \eta. \end{aligned}$$

The infinitesimal surface element is,

$$\frac{1}{2}K^2 (\cosh 2\xi - \cos 2\eta).$$

Thus, the domain \mathcal{D}' can be defined in elliptical coordinates,

$$\mathcal{D}' = \{(\xi, \eta) \in \mathbb{R}_{\geq 0} \times 2\pi S^1 : \xi \leq \xi^*\},$$

where $\xi^* = \operatorname{arctanh}(\sqrt{\lambda_1/\lambda_2})$.

This transforms the heat equation into the following form,

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{u_{\xi\xi} + u_{\eta\eta}}{K^2(\cosh 2\xi - \cos 2\eta)}, & t > 0, 0 \leq \xi \leq \xi^* \\ u(\xi, \eta, t) &= 0, & \xi = \xi^*, \\ \lim_{t \downarrow 0} u(\xi, \eta, t) &= \frac{2\delta_{\xi_0}^K(\xi)\delta_{\eta_0}^K(\eta)}{K^2(\cosh(2\xi) - 2\cos(\eta))}, \\ u(\xi, \eta, t) &= u(\xi, \eta + 2\pi, t), \\ u_{\xi}(\xi, \eta, t) &= u_{\xi}(\xi, \eta + 2\pi, t), \\ u_{\eta}(\xi, \eta, t) &= u_{\eta}(\xi, \eta + 2\pi, t). \end{aligned}$$

This partial differential equation can be solved by separation of variables. We can write,

$$u(\xi, \eta, t) = \Xi(\xi)H(\eta)T(t).$$

This leads to an ordinary differential equation,

$$\frac{1}{T(t)} \frac{dT}{dt} = \frac{1}{K^2(\cosh 2\xi - \cos 2\eta)} \left(\frac{1}{\Xi(\xi)} \frac{d^2\Xi}{d\xi^2} + \frac{1}{H(\eta)} \frac{d^2H}{d\eta^2} \right).$$

We introduce separation constants $\lambda > 0$ and a , and we set $h = \frac{1}{2}K\lambda$ to match the notation used in existing problems. This yields three separate ordinary differential equations,

$$\frac{dT}{dt} = -\frac{1}{2}\lambda^2 T(t), \tag{6.14}$$

$$\frac{d^2H}{d\eta^2} = -(\rho - 2h^2 \cos(2\eta))H(\eta), \tag{6.15}$$

$$\frac{d^2\Xi}{d\xi^2} = (\rho - 2h^2 \cosh(2\xi))\Xi(\xi). \tag{6.16}$$

Equation 6.14 allows for solutions of the form,

$$T(t) = \exp\left(-\frac{1}{2}\lambda^2 t\right).$$

The closely related differential equations 6.15 and 6.16 are known as Mathieu's differential equation and Mathieu's modified differential equation respectively. Sometimes, they are also referred to as the angular and the radial Mathieu equation respectively. Note that Equation 6.16 can be obtained by substitution of $\xi = i\eta$ into Equation 6.15. In [38], it is shown that the solutions are given by the (modified) Mathieu cosine and sine functions

of integral order, also called the (modified) cosine elliptic and (modified) sine elliptic functions,

$$\begin{aligned} H(\eta) &= ce_m(\eta, h), & m \geq 0, \\ H(\eta) &= se_m(\eta, h), & m \geq 1, \\ \Xi(\xi) &= Ce_m(\xi, h), & m \geq 0, \\ \Xi(\xi) &= Se_m(\xi, h), & m \geq 1. \end{aligned}$$

The dependency of the elliptic functions on the separation constant ρ has disappeared as a result of the requirement that ρ is such, that the solution is periodic in η with period 2π . In fact, by Sturm-Liouville theory, for fixed values of $h = \frac{1}{2}K\lambda^2$ the separation constant ρ takes different values for each value m and for both the Mathieu sine and Mathieu cosine function. Therefore, we find the following form for the solutions of $\Xi(\xi)H(\eta)$,

$$\begin{aligned} \Xi(\xi)H(\eta) &= Ce_m(\xi, h)ce_m(\eta, h), & m \geq 0, \\ \Xi(\xi)H(\eta) &= Se_m(\xi, h)se_m(\eta, h), & m \geq 1. \end{aligned}$$

The boundary condition $u(\xi, \eta, t) = 0$ at $\xi = \xi^*$ leads to nontrivial solutions when $Ce_m(\xi^*, h) = 0$ or $Se_m(\xi^*, h) = 0$. Call the positive roots of these equations $h_{m,n}$ and $k_{m,n}$, $n = 1, 2, \dots$, respectively. The general solution to the heat equation inside the ellipse with zero boundary condition is given by,

$$\begin{aligned} u(\xi, \eta, t) &= \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} A_{m,n} Ce_m(\xi, h_{m,n}) ce_m(\eta, h_{m,n}) \exp\left(-\frac{2h_{m,n}^2 t}{K^2}\right) \\ &\quad + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} B_{m,n} Se_m(\xi, k_{m,n}) se_m(\eta, k_{m,n}) \exp\left(-\frac{2k_{m,n}^2 t}{K^2}\right). \end{aligned}$$

All the different solutions are orthogonal. The coefficients $A_{m,n}$ and $B_{m,n}$ can be determined as is common in Fourier-series analysis,

$$\begin{aligned} A_{m,n} &= \frac{\int_0^{\xi^*} \int_0^{2\pi} Ce_m(\xi, h_{m,n}) ce_m(\eta, h_{m,n}) \delta_{\xi_0}^K(\xi) \delta_{\eta_0}^K(\eta) d\eta d\xi}{\int_0^{\xi^*} \int_0^{2\pi} Ce_m^2(\xi, h_{m,n}) ce_m^2(\eta, h_{m,n}) \frac{K^2}{2} (\cosh(2\xi) - \cos(2\eta)) d\eta d\xi} \\ &= \frac{2}{K^2} \frac{\int_0^{\xi^*} \int_0^{2\pi} Ce_m^2(\xi, h_{m,n}) ce_m^2(\eta, h_{m,n}) (\cosh(2\xi) - \cos(2\eta)) d\eta d\xi}{\int_0^{\xi_0} \int_0^{2\pi} Ce_m^2(\xi, h_{m,n}) ce_m^2(\eta, h_{m,n}) (\cosh(2\xi) - \cos(2\eta)) d\eta d\xi}, \\ B_{m,n} &= \frac{\int_0^{\xi^*} \int_0^{2\pi} Se_m(\xi, h_{m,n}) se_m(\eta, h_{m,n}) \delta_{\xi_0}^K(\xi) \delta_{\eta_0}^K(\eta) d\eta d\xi}{\int_0^{\xi^*} \int_0^{2\pi} Se_m^2(\xi, h_{m,n}) se_m^2(\eta, h_{m,n}) \frac{K^2}{2} (\cosh(2\xi) - \cos(2\eta)) d\eta d\xi} \\ &= \frac{2}{K^2} \frac{\int_0^{\xi^*} \int_0^{2\pi} Se_m^2(\xi, h_{m,n}) se_m^2(\eta, h_{m,n}) (\cosh(2\xi) - \cos(2\eta)) d\eta d\xi}{\int_0^{\xi_0} \int_0^{2\pi} Se_m^2(\xi, h_{m,n}) se_m^2(\eta, h_{m,n}) (\cosh(2\xi) - \cos(2\eta)) d\eta d\xi}. \end{aligned}$$

Since $\int_0^{2\pi} ce_m(\eta, h_{m,n}) d\eta = \int_0^{2\pi} se_m(\eta, h_{m,n}) d\eta = 0$ for $m \geq 1$, see [38], we find

$$\mathbb{P}(\tau \geq t) = \sum_{n=1}^{\infty} A'_{0,n} \exp\left(-\frac{2h_{0,n}^2 t}{K^2}\right), \quad (6.17)$$

where

$$A'_{0,n} = A_{0,n} \int_0^{\xi^*} Ce_0(\xi, h_{0,n}) d\xi \int_0^{2\pi} ce_0(\eta, h_{0,n}) d\eta, \quad \text{for } n \in \mathbb{N}.$$

6.4 An example

We consider a specific example of a 2-bus network to illustrate the results of this chapter. For this example, we will also compare the situation, where the power injection at the load-bus follows the law of a Wiener process, to the situation, where the power injection is Gaussian distributed.

All of the following variables are expressed in the per-unit system. Let the power demand at the load-bus in a 2-bus system $\{(P_t, Q_t)^\top\}_{t \geq 0}$ follow the law of an \mathbb{R}^2 -valued Wiener process $W = (W_t)_{t \geq 0}$, with $W_0 = (0, 0)^\top$ and whose increments $W_t - W_s$, for $t > s$, are distributed as $\mathcal{N}_2(0, (t - s)\Sigma)$, where

$$\Sigma = \begin{pmatrix} 9 & 2 \\ 2 & 6 \end{pmatrix}.$$

The transmission line has conductance 10 and susceptance -10. The nominal voltage is 1 and the minimum voltage is 0.9.

We wish to find the distribution of the first time the voltage at the load-bus drops below 0.9. This is given by the first exit time, $\tau = \inf\{t > 0 : W_t \notin \mathcal{D}\}$, where following Equation 6.13,

$$\mathcal{D} = \{(P, Q)^\top \in \mathbb{R}^2 : (P - 10 \cdot 0.9^2)^2 + (Q - 10 \cdot 0.9^2)^2 < 200 \cdot 0.9^2\}.$$

We rewrite Σ as $\Sigma = U^\top D U$, where

$$U = U^\top = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \text{and } D = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}.$$

The distribution of τ is equal to that of $\tau' = \inf\{t > 0 : W'_t \notin \mathcal{D}'\}$, where $W' = (W'_t)_{t \geq 0}$ is an \mathbb{R}^2 -valued Wiener process, with $W'_0 = \sqrt{D}^{-1} U (P_0 - 10 \cdot 0.9^2, Q_0 - 10 \cdot 0.9^2)^\top = (-1.62; -2.43\sqrt{2})^\top$ in Cartesian coordinates and whose increments $W'_t - W'_s$, for $t > s$, are distributed as $\mathcal{N}_2(0, (t - s)I_2)$, and

$$\mathcal{D}' = \sqrt{D}^{-1} U (\mathcal{D} - 10 \cdot 0.9^2(1, 1)^\top).$$

In elliptical coordinates with focal distance $K = \sqrt{1/5 - 1/10} \cdot 10\sqrt{2} \cdot 0.9 \cdot 1 = 9\sqrt{5}/5$, the set \mathcal{D}' corresponds to

$$\{(\xi, \eta) \in \mathbb{R}_{\geq 0} \times 2\pi S^1 : \xi \leq \xi^*\},$$

where $\xi^* = \operatorname{arctanh}(1/\sqrt{2})$.

The function $h \rightarrow Ce_0(\xi^*, h)$ is plotted in Figure 6.2. The first 10 roots of this function can be found in Table 6.1.

In Figure 6.3, plots are given of the functions $\xi \rightarrow Ce_0(\xi, h_{0n})$, $n = 1, 2, 3$, where h_{01}, h_{02} , and h_{03} are the first three roots of this function.

The relation of a point (x_1, x_2) in Cartesian coordinates to the same point in elliptic coordinates (ξ, η) can be described in a single complex equation, $x_1 + jx_2 = K \cos(j\xi + \eta)$,

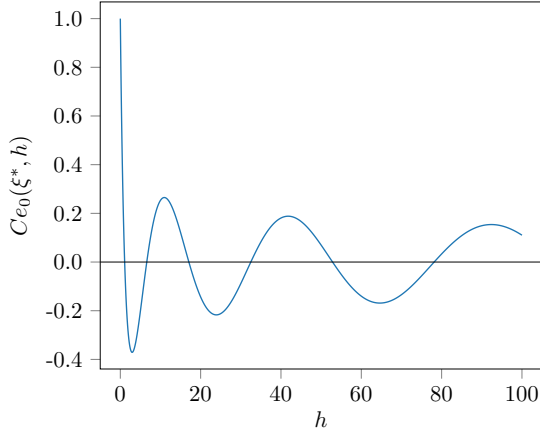


Figure 6.2: A plot of $C_{e_0}(\xi^*, h)$ as a function of h . The roots of this function are called $h_{01}, h_{02}, h_{03}, \dots$

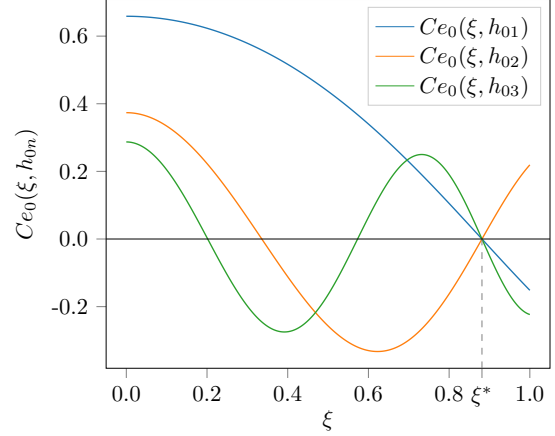


Figure 6.3: A plot of the functions $C_{e_0}(\xi, h_{0n})$ in ξ , for $n = 1, 2, 3$. Note that all functions have a shared root at $\xi = \xi^*$.

where j denotes the imaginary unit. For $x_2 < 0$, this is equivalent to,

$$\begin{aligned} \eta &= 2\pi - \Re[\arccos((x_1 + jx_2)/K)] \\ &= 2\pi - \frac{\arccos((x_1 + jx_2)/K) + \arccos((x_1 - jx_2)/K)}{2}, \quad \text{and,} \\ \xi &= \Im[\arccos((x_1 + jx_2)/K)] \\ &= \frac{\arccos((x_1 + jx_2)/K) - \arccos((x_1 - jx_2)/K)}{2j}. \end{aligned}$$

Thus, after some computation, we find, that W'_0 corresponds in elliptical coordinates to $(\xi_0, \eta_0) = (0.80516; 4.4078)$.

These coordinates can be used to determine the coefficients $A_{0,n}$ and $A_{0,n}$, for $n \in \mathbb{N}$, see Table 6.1 for the first ten values.

Subsequently, we can find the probability of violating the lower voltage bound within some time t , by using Equation 6.17. Notice the terms $\exp\left(-\frac{2h_{0,n}^2}{K^2}t\right)$ in this equation. To accurately approximate this probability for times t on different time scales, we should include all terms n for which $\frac{2(h_{0,n}^2 - h_{0,1}^2)}{K^2}t < C$, for some precision constant $C > 0$. A higher precision constant leads to a more accurate approximation.

n	$h_{0,n}$	$A_{0,n}$	$A'_{0,n}$
1	1.0807	2.1490×10^{-2}	1.8864×10^{-2}
2	6.5889	-1.5084	8.3767×10^{-3}
3	17.071	5.7865×10^1	9.9985×10^{-3}
4	32.490	-1.6544×10^3	9.0812×10^{-3}
5	52.845	3.8713×10^4	7.7751×10^{-3}
6	78.135	-7.6337×10^5	6.1108×10^{-3}
7	108.36	1.2710×10^7	4.3175×10^{-3}
8	143.52	-1.7265×10^8	2.6092×10^{-3}
9	183.62	1.6306×10^9	1.1378×10^{-3}
10	228.65	5.1211×10^8	-1.6998×10^{-5}

Table 6.1: Eigenvalues and coefficient used for determining $\mathbb{P}(\tau < t)$ in Equation 6.17.

Bibliography

- [1] CEER, “CEER report on power losses,” tech. rep., 2017. Ref: C17-EQS-80-03.
- [2] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*, vol. 7. McGraw-hill New York, 1994.
- [3] S. Teleke, M. E. Baran, A. Q. Huang, S. Bhattacharya, and L. Anderson, “Control strategies for battery energy storage for wind farm dispatching,” *IEEE Transactions on Energy Conversion*, vol. 24, no. 3, pp. 725–732, 2009.
- [4] M. Baran and F. F. Wu, “Optimal sizing of capacitors placed on a radial distribution system,” *IEEE Transactions on power Delivery*, vol. 4, no. 1, pp. 735–743, 1989.
- [5] M. E. Baran and F. F. Wu, “Network reconfiguration in distribution systems for loss reduction and load balancing,” *IEEE Transactions on Power delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.
- [6] O. I. Elgerd, “Electric energy systems theory: an introduction,” 1982.
- [7] A. Klos and J. Wojcicka, “Physical aspects of the nonuniqueness of load flow solutions,” *International Journal of Electrical Power & Energy Systems*, vol. 13, no. 5, pp. 268–276, 1991.
- [8] P. M. Anderson and A. A. Fouad, *Power system control and stability*. John Wiley & Sons, 2008.
- [9] R. D. Zimmerman, C. E. Murillo-Sánchez, R. J. Thomas, *et al.*, “Matpower: Steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2011.
- [10] C. Godsil and G. Royle, “Strongly regular graphs,” in *Algebraic graph theory*, pp. 217–247, Springer, 2001.
- [11] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans, “Usefulness of dc power flow for active power flow analysis,” in *Power Engineering Society General Meeting, 2005. IEEE*, pp. 454–459, IEEE, 2005.
- [12] J.-P. Sandvik and A. Årnes, “The reliability of clocks as digital evidence under low voltage conditions,” *Digital Investigation*, vol. 24, pp. S10–S17, 2018.
- [13] C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- [14] A. W. Van der Vaart, *Asymptotic statistics*, vol. 3. Cambridge University Press, 1998.

- [15] J. Blanchet and H. Lam, “State-dependent importance sampling for rare-event simulation: an overview and recent advances,” *Surv. Oper. Res. Manag. Sci.*, vol. 17, no. 1, pp. 38–59, 2012.
- [16] S. Asmussen and P. W. Glynn, *Stochastic simulation: algorithms and analysis*, vol. 57. Springer Science & Business Media, 2007.
- [17] M. K. Nakayama, “General conditions for bounded relative error in simulations of highly reliable markovian systems,” *Advances in Applied Probability*, vol. 28, no. 3, pp. 687–727, 1996.
- [18] A. Dembo and O. Zeitouni, “Large deviations techniques and applications. corrected reprint of the second (1998) edition. stochastic modelling and applied probability, 38,” 2010.
- [19] K. Methaprayoon, C. Yingvivatanapong, W.-J. Lee, and J. R. Liao, “An integration of ann wind power estimation into unit commitment considering the forecasting uncertainty,” *IEEE Transactions on Industry Applications*, vol. 43, no. 6, pp. 1441–1448, 2007.
- [20] E. D. Castronuovo and J. P. Lopes, “On the optimization of the daily operation of a wind-hydro power plant,” *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1599–1606, 2004.
- [21] A. B. Owen, Y. Maximov, and M. Chertkov, “Importance sampling the union of rare events with an application to power systems analysis,” *arXiv preprint arXiv:1710.06965*, 2017.
- [22] M. Anvari, G. Lohmann, M. Wächter, P. Milan, E. Lorenz, D. Heinemann, M. R. R. Tabar, and J. Peinke, “Short term fluctuations of wind and solar power systems,” *New Journal of Physics*, vol. 18, no. 6, p. 063027, 2016.
- [23] J. Zhang, B.-M. Hodge, and A. Florita, “Joint probability distribution and correlation analysis of wind and solar power forecast errors in the western interconnection,” *Journal of Energy Engineering*, vol. 141, no. 1, p. B4014008, 2014.
- [24] J.-E. Skog, K. Koreman, B. Pääjärvi, T. Worzyk, and T. Andersröd, “The norned hvdc cable link—a power transmission highway between norway and the netherlands,” *Proceedings of Energex 2006*, 2006.
- [25] J. Gordesch, “Introduction to multivariate analysis-chatfield, ch.; collins, aj,” *Metrika*, vol. 29, pp. 142–142, 1982.
- [26] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1990.
- [27] P. C. Mahalanobis, “On the generalized distance in statistics,” National Institute of Science of India, 1936.
- [28] T. Hesterberg, “Weighted average importance sampling and defensive mixture distributions,” *Technometrics*, vol. 37, no. 2, pp. 185–194, 1995.
- [29] J. Bucklew, *Introduction to rare event simulation*. Springer Science & Business Media, 2013.

- [30] P. Mörters and Y. Peres, *Brownian motion*, vol. 30. Cambridge University Press, 2010.
- [31] R. L. Schilling and L. Partzsch, *Brownian motion: an introduction to stochastic processes*. Walter de Gruyter GmbH & Co KG, 2014.
- [32] B. Øksendal, *Stochastic differential equations*. Springer, 2003.
- [33] F. Spieksma, “Lecture notes: An introduction to stochastic processes in continuous time,” May 2016.
- [34] M. Schilder, “Some asymptotic formulas for wiener integrals,” *Transactions of the American Mathematical Society*, vol. 125, no. 1, pp. 63–85, 1966.
- [35] B. Zhang and D. Tse, “Geometry of injection regions of power networks,” *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 788–797, 2013.
- [36] D. L. Powers, *Boundary value problems: and partial differential equations*. Academic Press, 2009.
- [37] J. Wendel, “Hitting spheres with brownian motion,” *The annals of probability*, pp. 164–169, 1980.
- [38] N. W. McLachlan, “Theory and application of mathieu functions,” 1951.

Appendix A

Matlab code

A.1 Matlab code used in Section 4.4

```
1 %% INITIALIZATION
2 % Load data and call on functions present in Matpower
3 mpc = loadcase('...');
4
5 % define named indices into bus, gen, branch matrices
6 [PQ, PV, REF, NONE, BUS_I, BUS_TYPE, PD, QD, GS, BS, BUS_AREA, ...
7   VM, VA, BASE_KV, ZONE, VMAX, VMIN, LAM_P, LAM_Q, MU_VMAX, ...
8   MU_VMIN] = idx.bus;
9 [F_BUS, T_BUS, BR_R, BR_X, BR_B, RATE_A, RATE_B, RATE_C, ...
10  TAP, SHIFT, BR_STATUS, PF, QF, PT, QT, MU_SF, MU_ST, ...
11  ANGMIN, ANGMAX, MU_ANGMIN, MU_ANGMAX] = idx.brch;
12 [GEN_BUS, PG, QG, QMAX, QMIN, VG, MBASE, GEN_STATUS, PMAX, ...
13  PMIN, MU_PMAX, MU_PMIN, MU_QMAX, MU_QMIN, PC1, PC2, QC1MIN, ...
14  QC1MAX, QC2MIN, QC2MAX, RAMP_AGC, RAMP_10, RAMP_30, RAMP_Q, ...
15  APF] = idx.gen;
16
17 % convert to internal indexing
18 mpc = ext2int(mpc);
19 [baseMVA, bus, gen, branch] = deal(mpc.baseMVA, mpc.bus, ...
20   mpc.gen, mpc.branch);
21
22 % get bus index lists of each type of bus
23 [ref, pv, pq] = bustypes(bus, gen);
24
25 % compute power injections (generation - load)
26 % adjusted for phase shifters and real shunts
27 [B, Bf, Pbusinj, Pfinj] = makeBdc(baseMVA, bus, branch);
28 Pbus = real(makeSbus(baseMVA, bus, gen)) - Pbusinj - ...
29   bus(:, GS) / baseMVA;
30
31 %% DISTRIBUTION OF THE POWER INJECTIONS
32 % original distribution
33 mu = Pbus;
34 Sigma = diag(mu.^2);
35
36 % condition on power injections summing to zero
37 nbus = size(bus,1); % number of buses
38 mucond = mu - Sigma * ones(nbus) * mu / sum(Sigma, 'all');
```

```

39 Sigmacond = Sigma - Sigma * ones(nbus) * Sigma / sum(Sigma, 'all');
40
41 % remove reference bus
42 mu2 = mucond;
43 mu2(ref) = [];
44 Sigma2 = Sigmacond;
45 Sigma2(ref, :) = [];
46 Sigma2(:, ref) = [];
47
48 %% HALF-SPACES
49 % find the Power Transmission Distribution Factor
50 PTDF = makePTDF(mpc);
51
52 % find the maximum power flow through transmission lines
53 maxflow = branch(:, RATE_A) / baseMVA;
54 % this data is not present in all cases
55 % in absence of data a reasonable choice has to be made
56
57 % define half-spaces
58 a = PTDF;
59 a(:,ref) = [];
60 a = [a;-a];
61 b = [maxflow;maxflow];
62
63 % Mahalanobis distance
64 mdist = (b - a * mu2) ./ vecnorm(a * sqrtm(Sigma2), 2, 2);
65
66 % most likely point per half-space
67 nu = mu2' + (mdist ./ vecnorm(a * sqrtm(Sigma2), 2, 2)) .* a * ...
68     Sigma2;
69
70 %% SIMULATION
71 rarity = ...; % a vector of rarity parameters
72 prob = zeros([length(rarity) 1]);
73 svar = zeros([length(rarity) 1]);
74
75 N = ...; % number of samples
76 for i = 1:length(rarity)
77     mix = exp(mdist.^2 * rarity(i) / 2) .* ...
78         sqrt(normcdf(2*sqrt(rarity(i)) * mdist, 'upper'));
79     dist = gmdistribution(nu(mix > 0, :), Sigma2 / rarity(i), ...
80         mix(mix > 0));
81     X = random(dist, N);
82     p = mvnpdf(X, mu2', Sigma2 / rarity(i));
83     q = pdf(dist, X);
84     w = p ./ q;
85     ind = any(a * X' > b);
86     % indicator function for lying outside the polyhedron
87     prob(i) = sum(w' .* ind)/N;
88     svar(i) = sum((w' .* ind - prob(i)).^2)/N;
89 end

```